

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 763 551**

51 Int. Cl.:

A61K 48/00 (2006.01)

C07K 14/755 (2006.01)

C12N 9/64 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.04.2016** **PCT/US2016/027931**

87 Fecha y número de publicación internacional: **20.10.2016** **WO16168728**

96 Fecha de presentación y número de la solicitud europea: **15.04.2016** **E 16720253 (0)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019** **EP 3283126**

54 Título: **Promotores y vectores recombinantes para la expresión de proteínas en el hígado y uso de los mismos**

30 Prioridad:

16.04.2015 US 201562148696 P

06.08.2015 US 201562202133 P

01.09.2015 US 201562212634 P

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

29.05.2020

73 Titular/es:

EMORY UNIVERSITY (50.0%)

1599 Clifton Road, NE 4th Floor

Atlanta, GA 30322, US y

CHILDREN'S HEALTHCARE OF ATLANTA, INC. (50.0%)

72 Inventor/es:

DOERING, CHRISTOPHER, B.;

SPENCER, TRENT, H. y

BROWN, HARRISON, C.

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 763 551 T3

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

DESCRIPCIÓN

Promotores y vectores recombinantes para la expresión de proteínas en el hígado y uso de los mismos

5 **Campo de la divulgación**

La presente invención se refiere a promotores y vectores recombinantes de expresión transgénica, así como a moléculas de ácidos nucleicos recombinantes que codifican nuevos factores de coagulación.

10 **Antecedentes**

Las mutaciones en el gen del factor de coagulación VIII (fVIII) dan como resultado una proteína disminuida o defectuosa del factor de coagulación (fVIII) que da lugar a la hemofilia A, que se caracteriza por un sangrado incontrolado. La hemofilia B se asocia de manera similar con el factor de coagulación IX (fIX). El tratamiento de la hemofilia A generalmente implica una infusión intravenosa de varias semanas de por vida de producto de fVIII derivado de plasma humano o recombinante. Debido al alto coste, menos del 30% de la población mundial con hemofilia A recibe esta forma de tratamiento. Además, alrededor del 25 % de los pacientes tratados con productos de reemplazo de fVIII desarrollan anticuerpos neutralizantes que hacen que el tratamiento futuro sea ineficaz. Por lo tanto, existe la necesidad de identificar terapias mejoradas.

Las terapias génicas se basan generalmente en virus genéticamente modificados diseñados para suministrar transgenes funcionales al paciente para que sus propias células puedan biosintetizar las proteínas carentes o defectuosas. Se han realizado avances clínicos usando vectores de virus adenoasociados recombinantes (rAAV) para la expresión de fIX en el hígado; sin embargo, el uso de rAAV para la expresión de fVIII en pacientes con hemofilia A ha sido un desafío debido a la biosíntesis ineficaz de fVIII humano (hfVIII). Los vectores víricos adenoasociados recombinantes (rAAV) producen cápsides que tienen un espacio limitado para encapsular ácidos nucleicos. El fVIII es una glicoproteína grande, y las secuencias de rAAV necesarias para codificar y expresar fVIII generalmente exceden la capacidad de empaquetamiento de la cápside.

30 **Sumario**

En el presente documento se divulgan realizaciones de una nueva molécula de ácido nucleico recombinante que comprende un promotor que se ha optimizado para tener una longitud mínima y promover la expresión de proteínas específicas de tejido. En la invención, el promotor es un promotor específico de hígado que promueve sustancialmente más expresión de proteínas en el hígado y en las células hepáticas que en otros tipos de tejidos. En algunas realizaciones, el promotor puede incluirse en un vector vírico (tal como un vector de virus adenoasociado) en combinación operativa con una secuencia del ácido nucleico heterólogo que codifica una proteína de interés para promover la expresión de la proteína de interés, por ejemplo en tejido y/o células del hígado.

Por lo tanto, la invención proporciona una molécula de ácido nucleico recombinante que comprende un promotor específico de hígado que comprende un primer elemento de respuesta de no más de 160 nucleótidos de longitud que comprende: un sitio de unión al factor de transcripción (TF, por sus siglas en inglés) HNF1a que comprende o que consiste en los nucleótidos 1-12 de la SEQ ID NO: 4, un sitio de unión al TF HNF1-1 que comprende o que consiste en los nucleótidos 16-23 de la SEQ ID NO: 4, un sitio de unión al TF HNF4 que comprende o que consiste en los nucleótidos 26-36 de la SEQ ID NO: 4, un sitio de unión al TF HNF3a que comprende o que consiste en los nucleótidos 39-45 de la SEQ ID NO: 4, un sitio de unión al TF HNF1-2 que comprende o que consiste en los nucleótidos 48-62 de la SEQ ID NO: 4, un sitio de unión al TF HNF3-2 que comprende o que consiste en los nucleótidos 65-71 de la SEQ ID NO: 4, un sitio de unión al TF HP1 que comprende o que consiste en los nucleótidos 75-87 de la SEQ ID NO: 4, una caja TATA que comprende o que consiste en los nucleótidos 108-114 de la SEQ ID NO: 4, y un Sitio de Inicio de la Transcripción que comprende o que consiste en los nucleótidos 116-146 de la SEQ ID NO: 4. El primer elemento de respuesta es de no más de 160 nucleótidos de longitud (tal como no más de 150 nucleótidos de longitud, tal como de 146 nucleótidos de longitud).

En algunas realizaciones, el primer elemento de respuesta comprende, de 5' a 3', el sitio de unión al TF HNF1a, el sitio de unión al TF HNF1-1, el sitio de unión al TF HNF4, el sitio de unión al TF HNF3a, el sitio de unión al TF HNF1-2, el sitio de unión al TF HNF3-2, el sitio de unión al TF HP1, la caja TATA y el Sitio de Inicio de la Transcripción (TSS, por sus siglas en inglés).

En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 4 (HCB), o una secuencia al menos un 90 % idéntica a la misma.

En algunas realizaciones, la molécula de ácido nucleico recombinante puede comprender un promotor que comprende el primer elemento de respuesta de la invención, y puede comprender además un segundo elemento de respuesta. El segundo elemento de respuesta puede comprender, por ejemplo, un elemento de respuesta HSh (por ejemplo, que comprende o que consiste en la secuencia de nucleótidos establecida como la SEQ ID NO: 111, o una secuencia al

menos un 90 % idéntica a la misma), un elemento de respuesta 5'HS (por ejemplo, que comprende o que consiste en la secuencia de nucleótidos establecida como los nucleótidos 6-32 de la SEQ ID NO: 111, o una secuencia al menos un 90 % idéntica a la misma), o un elemento de respuesta 3'HS (por ejemplo, que comprende o que consiste en la secuencia de nucleótidos establecida como los nucleótidos 44-68 de la SEQ ID NO: 111, o una secuencia al menos un 90 % idéntica a la misma).

En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como una de las SEQ ID NO: 102 (HSh-HCB), SEQ ID NO: 104 (5'HSh-HCB), o SEQ ID NO: 103 (3'HSh-HCB), o una secuencia al menos un 90 % idéntica a una de las SEQ ID NO: 102 (HSh-HCB), SEQ ID NO: 104 (5'HSh-HCB), o SEQ ID NO: 103 (3'HSh-HCB). En aspectos adicionales de la presente divulgación, una molécula de ácido nucleico recombinante puede comprender un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como una de las SEQ ID NO: 5 (shortABP-HPI-God-TSS), SEQ ID NO: 7 (ABP-HP1-God-TSS), SEQ ID NO: 105 (HSh-SynO-TSS), SEQ ID NO: 106 (sHS-SynO-TSS), SEQ ID NO: 107 (Agro), SEQ ID NO: 108 (HS-SynO-TSS), o SEQ ID NO: 112 (HNF1-Short-ABPEXact-SynO-TSS-Int), o una secuencia al menos un 90 % idéntica a una de las SEQ ID NO: 5 (shortABP-HP1-God-TSS), SEQ ID NO: 7 (ABP-HP1-God-TSS), SEQ ID NO: 105 (HSh-SynO-TSS), SEQ ID NO: 106 (sHS-SynO-TSS), SEQ ID NO: 107 (Agro), SEQ ID NO: 108 (HS-SynO-TSS), o SEQ ID NO: 112 (HNF1-ShortABPEXact-SynO-TSS-Int).

En algunas realizaciones, el promotor de la invención puede incluirse en un vector, tal como un vector vírico (por ejemplo, un vector de virus adenoasociado). En algunas realizaciones, el promotor se incluye en el vector en combinación operativa con una secuencia del ácido nucleico heterólogo que codifica una proteína de interés para promover la expresión de la proteína de interés, por ejemplo en tejido y/o células del hígado. En algunas realizaciones, la proteína de interés puede ser un factor de coagulación, tal como fVIII o fIX o una variante de los mismos, tal como una variante de fVIII que comprende los dominios A1, A2, A3, C1 y C2 de fVIII, con los dominios A2 y A3 unidos por un enlazador peptídico y delección del dominio B de fVIII. En algunas realizaciones, la proteína de interés puede ser una variante de fVIII y la molécula de ácido nucleico heterólogo puede comprender o consistir en la secuencia del ácido nucleico establecida como la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125, o SEQ ID NO: 126, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125 o SEQ ID NO: 126. En algunas realizaciones, la proteína de interés puede ser un fIX y la molécula de ácido nucleico heterólogo puede comprender o consistir en la secuencia del ácido nucleico establecida como la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124, o SEQ ID NO: 127, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124 o SEQ ID NO: 127.

En algunas realizaciones, el vector puede ser un vector AAV recombinante que comprende un genoma que comprende una molécula de ácido nucleico que codifica cualquiera de los promotores específicos de hígado de la invención (tal como el promotor HCB, SEQ ID NO: 4) operativamente unido a una molécula de ácido nucleico heterólogo que codifica una variante de fVIII, en donde la molécula de ácido nucleico heterólogo comprende o consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125, o SEQ ID NO: 126, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125 o SEQ ID NO: 126. En varias de tales realizaciones, el genoma de AAV recombinante (de 5' a 3' ITR) no tiene más de 5,1, 5,0, 4,9, 4,8, 4,7, 4,6 o 4,5 kb de longitud.

En algunas realizaciones, el vector puede ser un vector AAV recombinante que comprende un genoma que comprende una molécula de ácido nucleico que codifica cualquiera de los promotores específicos del hígado de la invención (tal como el promotor HCB, SEQ ID NO: 4) operativamente unido a una molécula de ácido nucleico heterólogo que codifica una variante de fIX, en donde la molécula de ácido nucleico heterólogo comprende o consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124, o SEQ ID NO: 127, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124 o SEQ ID NO: 127. En varias de tales realizaciones, el genoma de AAV recombinante (de 5' a 3' ITR) no tiene más de 5,1, 5,0, 4,9, 4,8, 4,7, 4,6 o 4,5 kb de longitud.

La invención también proporciona un vector o composición de la invención para su uso en un método para inducir la coagulación de la sangre en un sujeto que lo necesite. El método comprende administrar al sujeto una cantidad terapéuticamente eficaz de un vector (tal como un vector AAV) que codifica un factor de coagulación como se describe en el presente documento. En algunas realizaciones, el sujeto es un sujeto con un trastorno de la coagulación, tal como hemofilia A o hemofilia B. En algunas realizaciones, el trastorno de la coagulación es hemofilia A y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una proteína con actividad de fVIII. En otras realizaciones, el trastorno de la coagulación es hemofilia B y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una proteína con actividad de fIX.

Las características y ventajas anteriores y otras de la presente divulgación serán más evidentes a partir de la siguiente descripción detallada de varias realizaciones que procede con referencia a las figuras adjuntas.

Breve descripción de las figuras

Las FIG. 1A y 1B ilustran una alineación de secuencia de los dominios A1 y A3 para ortólogos humanos y porcinos

de la proteína variante ET3 de fVIII. (FIG. 1A) Alineación de la secuencia del dominio A1 para la variante ET3 humana (secuencia superior, SEQ ID NO: 13) y porcina (secuencia media, SEQ ID NO: 14) de fVIII. La secuencia inferior muestra restos idénticos. Se muestran las alineaciones de la secuencia de aminoácidos para el péptido señal (barra N-terminal), el dominio ácido de cadena pesada (barra C-terminal), el fVIII humano (arriba) y ET3 (abajo). Los enlaces disulfuro se señalan por las líneas que conectan los restos de cisteína. Los lugares donde bien la secuencia humana, ET3 o ambas secuencias codifican un sitio de unión por glicosilación ligada a N (N-X-S/T) se delinean con un recuadro. (FIG. 1B) Alineación de la secuencia del dominio A3 para la variante ET3 humana (secuencia superior, SEQ ID NO: 15) y porcina (secuencia media, SEQ ID NO: 16) de fVIII. La secuencia inferior muestra restos idénticos. Se muestran las alineaciones de la secuencia de aminoácidos para el péptido de activación (barra), fVIII humano (arriba) y ET3 (abajo). Los enlaces disulfuro se señalan por las líneas que conectan los restos de cisteína. Los lugares donde bien el humano, ET3 o ambas secuencias codifican un sitio de unión por glicosilación ligada a N (N-X-S/T) se delinean con un recuadro.

Las FIG. 2A-2C muestra tablas de preferencia codónica específica de hígado, humana completa y mieloide.

La FIG. 3A muestra datos de expresión *in vitro* en células HepG2 que indican que la optimización de codones específicos de hígado mejora la expresión para las variantes HSQ y ET3 de la proteína fVIII.

La FIG. 3B muestra datos *in vivo* para HSQ y ET3 con codones optimizados que indican una mayor expresión de fVIII después de la inyección hidrodinámica de un vector AAV que codifica variantes con codones de hígado optimizados de la proteína HSQ y ET3 fVIII en ratones.

La FIG. 4 ilustra vectores AAV que codifican variantes de fVIII.

La FIG. 5 muestra datos de expresión *in vitro* en células HepG2 que indican que la optimización de codones específicos de hígado, pero no la optimización de codones específicos mieloides, mejora la expresión de las variantes HSQ y ET3 de la proteína fVIII en las células de hígado.

La FIG. 6 muestra datos de expresión *in vitro* en células HepG2 que indican que la optimización de codones específicos de hígado, pero no la optimización de codones específicos mieloides, mejora la expresión de las variantes HSQ y ET3 de la proteína fVIII en las células de hígado.

La FIG. 7 muestra datos de expresión *in vitro* en células HepG2 que indican que la optimización de codones específicos de hígado, pero no la optimización de codones específicos humanos, mejora la expresión de la proteína fIX en células de hígado.

La FIG. 8 ilustra promotores que comprenden un potenciador ABP o un potenciador ABP acortado (shortABP).

Las FIG. 9A y 9B muestran datos sobre el uso de vectores AAV que contienen los promotores indicados para la expresión de fVIII. (FIG. 9A) Datos de expresión *in vitro* en células HepG2. (FIG. 9B) Expresión *in vivo* de fVIII después de la inyección hidrodinámica de un vector AAV.

La FIG. 10 ilustra secuencias de elementos de respuesta y componentes de los elementos de respuesta indicados.

Las FIG. 11A y 11B ilustran secuencias promotoras y componentes de los promotores indicados.

Las FIG. 12A y 12B muestran datos sobre el uso de vectores AAV que contienen los promotores indicados para la expresión de fVIII. (FIG. 12A) Datos de expresión *in vitro* en células HepG2. (FIG. 12B) Expresión *in vivo* de fVIII después de la inyección hidrodinámica de un vector AAV.

La FIG. 13 ilustra secuencias promotoras y componentes de los promotores indicados.

Las FIG. 14A y 14B muestran datos sobre el uso de vectores AAV que contienen los promotores indicados para la expresión de fVIII. (FIG. 14A) Datos de expresión *in vitro* en células HepG2. (FIG. 14B) Expresión *in vivo* de fVIII después de la inyección hidrodinámica de un vector AAV.

Las FIG. 15 y 16 muestran diagramas esquemáticos que ilustran la estructura de potenciadores específicos de hígado (FIG. 15) y promotores que incluyen los potenciadores específicos de hígado (FIG. 16).

La FIG. 17 muestra datos de ensayos *in vitro* relacionados con el uso de vectores AAV que contienen los promotores indicados para la expresión de fVIII en células HepG2.

La FIG. 18 ilustra unos vectores AAV que codifican una variante de fVIII.

La FIG. 19 es un gráfico que muestra que la transducción de ratones con el vector AAV2-HCB-ET3-LCO-NCG-SpA (SEQ ID NO: 130) que codifica ET3 con codones de hígado optimizados con motivos CpG eliminados conduce a un aumento significativo de la actividad de fVIII en ratones transducidos.

Descripción detallada

Existe la necesidad de desarrollar una estrategia de transferencia de genes segura y eficaz para el tratamiento de la hemofilia, tal como hemofilia A y B, y hemofilia adquirida. En el contexto de las terapias génicas para el tratamiento de la hemofilia A, varios obstáculos han dificultado el desarrollo del uso de un vector vírico adenoasociado como vehículo de suministro génico, tal como la capacidad limitada de empaquetamiento de ADN del virus adenoasociado para el transgén de fVIII grande y la biosíntesis ineficaz de fVIII humano. En el presente documento se informa de un sistema de suministro de transgenes basado en AAV que utiliza mejoras para la expresión de fVIII en el contexto de la transferencia de genes por AAV dirigidos al hígado. Estos incluyen: 1) El uso de una secuencia de nucleótidos codificante que tiene una preferencia de uso de codones mejorada para la célula hepática humana en comparación con la secuencia de nucleótidos de origen natural de fVIII; 2) Optimización del uso de codones para eliminar los dinucleótidos 5'-CG-3' y otros motivos de ADN de acción en cis dañinos, por ejemplo, sitios de corte y empalme crípticos, cajas TATA, señal terminal, estructura secundaria de ARNm, señales prematuras de poliA, motivos de inestabilidad de ARN, sitios internos de unión al ribosoma; y 3) promotores dirigidos al hígado de tamaño mínimo para reducir el tamaño del transgén para que pueda usarse en el entorno de tamaño limitado del sistema de vectores víricos adenoasociados. Las mejoras pueden generalizarse para la expresión mejorada de cualquier transgén de AAV. En

algunas realizaciones, el vector AAV suministra una expresión eficaz de fVIII a dosis víricas que no se predice que causen toxicidad en seres humanos.

En algunas realizaciones, estas mejoras también se pueden aplicar a fIX, especialmente para diseños de vectores fIX autocomplementarios. Los diseños autocomplementarios tienen la mitad de la capacidad de empaquetamiento que los diseños monocatenarios, por lo que las limitaciones de tamaño de los vectores (~ 2,4kb) se convierten en una preocupación incluso para fIX.

El trabajo previo sugirió que el tratamiento de ratones con deficiencia en fVIII (hemofilia A) con un vector AAV que codifica una forma modificada de fVIII (dominio B eliminado) denominada ET3 a dosis de vector que varían de 5×10^{11} - 2×10^{13} pv/ kg podría corregir teóricamente su deficiencia de fVIII y fenotipo hemorrágico (véase Brown et al., "Bioengineered Factor FVIII Enables Long-Term Correction of Murine Hemophilia A Following Liver-Directed Adeno-Associated Viral Vector Delivery", Molecular Therapy - Methods and Clinical Development. 1:14036, 2014). Sin embargo, debido al genoma de gran tamaño de ET3, el vector sufría una baja fabricación de títulos y una heterogeneidad sustancial entre partículas. El gran tamaño del genoma ET3-AAV con codones optimizados seguía siendo incompatible con el empaquetamiento eficaz del vector vírico. Para los vectores AAV, se prefiere un tamaño de genoma de AAV de no más de 4,7-5,0 kb para obtener un mayor rendimiento y consistencia que los genomas que exceden de 5,0kb. La secuencia codificante de ET3 con el dominio B eliminado es de 4,4 kb. Sin embargo, con la adición de los elementos de control vírico y regulador necesarios, los genomas de fVIII ET3-AAV excedieron sustancialmente la capacidad de empaquetamiento.

Por primera vez se divulga en el presente documento un genoma fVIII (ET3 u otra variante con dominio B eliminado) -AAV de menos de 5,0 kb de longitud que se desarrolló para permitir tanto la expresión mejorada de fVIII (o variante del mismo) como un empaquetamiento vírico eficaz. Se siguieron múltiples etapas para reducir el tamaño del genoma de AAV a niveles aceptables. Por ejemplo, se utilizó un enfoque de ensamblaje del sitio de unión al factor de transcripción combinatorio para crear un conjunto de promotores específicos de hígado que varían en tamaño. Estos promotores representan una reducción de tamaño del 30-90 % sobre los promotores específicos de hígado utilizados actualmente, tal como HLP y HCR-hAAT, que varían en tamaño de 250 a más de 700 bases. Algunos de estos promotores impulsan niveles de expresión y especificidad transgénica comparables o mejores a los observados con HLP y HCR-hAAT.

Una barrera importante para el desarrollo de terapias clínicas exitosas basadas en la transferencia de genes es la disponibilidad de elementos genéticos de origen natural o sintéticos capaces de expresión funcional, y a menudo dirigida o restringida por el tipo celular, en el contexto de un casete de ácido nucleico suministrado por vectores (véase, por ejemplo, Papadakis et al., "Promoters and control elements: designing expression cassettes for gene therapy", Curr Gene Ther., 4(1):89-113, 2004). En general, se cree que los promotores existentes naturalmente han sido perfeccionados por la evolución para impulsar la expresión finamente sintonizada a través de la combinación de múltiples secuencias reguladoras en cis. En la mayoría de los organismos vivos, y especialmente en eucariotas con grandes tamaños de genoma, no parece haber una fuerza impulsora para limitar el tamaño del promotor y, por lo tanto, la mayoría de los promotores endógenos se extienden en cientos, y más a menudo miles, de pares de bases de ADN (pb). Debido a su tamaño, estos promotores de genes naturales endógenos generalmente no son susceptibles de inclusión en productos de terapia génica debido a limitaciones de tamaño.

Los promotores víricos endógenos, por otro lado, han evolucionado para poseer una eficacia de fuerza y tamaño que los hace atractivos para su uso en tecnologías de transferencia de genes. Los ejemplos más destacados incluyen el promotor temprano inmediato (IE, por sus siglas en inglés) de citomegalovirus (CMV), el promotor tardío principal de adenovirus (Ad), el promotor del virus simio 40 (SV40) y la repetición terminal larga (LTR por sus siglas en inglés) del virus de la leucemia murina de Moloney (MoMLV). Cada uno de estos promotores puede impulsar la transcripción de alto nivel de transgenes heterólogos exógenos en una variedad de tipos de células eucariotas. Sin embargo, de manera no sorprendente, las células eucariotas han desarrollado mecanismos de defensa celular para detectar e inactivar eficazmente (es decir, silenciar) los promotores víricos y, por lo tanto, estos promotores funcionan más eficazmente en sistemas modelo de cultivo celular que las aplicaciones de terapia génica *in vivo*.

Por estas razones, ha habido un interés significativo en el desarrollo de promotores sintéticos, ya sea genéricos (véase, por ejemplo, Juven-Gershon et al., "Rational design of a super core promoter that enhances gene expression", Nat Methods, 3(11):917-22, 2006; Schlabach et al., "Synthetic design of strong promoters". PNAS, 2010;107(6):2538-43, 2010), o adaptados a aplicaciones específicas de terapia génica, incluidas la hemofilia A y B (véase, por ejemplo, McIntosh et al., "Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant", Blood, 121(17):3335-44, 2013; Nair et al., "Computationally designed liver-specific transcriptional modules and hyperactive factor IX improve hepatic gene therapy", Blood, 123(20):3195-9, 2014).

El conocimiento de los promotores y las mejoras sigue siendo limitado y actualmente no es posible diseñar por ordenador un promotor óptimo con ninguna confianza. Estudios como los descritos por Juven-Gershon y Kadonaga han avanzado en la definición de diseños optimizados de promotores centrales, como su Super Core Promoter 1 (SCP1) y son informativos en el campo. Sin embargo, como se muestra en los promotores descritos en los ejemplos en el presente documento, que no contienen una gran similitud con SCP1 en el dominio del promotor central, estas

secuencias no son necesarias para una función promotora fuerte, al menos en el contexto de aplicaciones de terapia génica dirigida al hígado (véase también, Juven-Gershon et al., "Rational design of a super core promoter that enhances gene expression", Nat Methods, 3(11):917-22, 2006).

El desarrollo del promotor más genérico se ha centrado en lograr un equilibrio óptimo de potencia transcripcional con un tamaño mínimo. En el campo del diseño de promotores dirigidos al hígado, el uso de enfoques de diseño racional por McIntosh et al. y Nair et al. (*supra*) condujo a la identificación de promotores para su uso en enfoques de terapia génica con AAV-fVIII y AAV-flX. Sin embargo, a pesar del extenso estudio, ambos grupos describen diseños de promotores que son significativamente más grandes (≥ 252 pb) y no más (o menos) potentes que los descritos en el presente documento, tal como la SEQ ID NO: 4 (HCB). De hecho, dados los intentos previos para optimizar el diseño de promotores, fue particularmente sorprendente identificar promotores tales como los descritos en el presente documento que son más pequeños que los promotores de la técnica anterior (tal como el promotor HLP), pero potencia equivalente o mejorada para impulsar la transcripción, particularmente en el contexto de aplicaciones de terapia génica *in vivo*.

I. Términos

A menos que se indique otra cosa, los términos técnicos se usan según su uso convencional. Pueden encontrarse definiciones de términos comunes sobre biología molecular en Benjamin Lewin, Genes V, publicado por Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al. (eds.), The Encyclopedia of Molecular Biology, publicado por Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); y Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, publicado por VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Para facilitar la revisión de las diversas realizaciones de la divulgación, se proporcionan las siguientes explicaciones de términos específicos:

5' y/o 3': Se dice que las moléculas de ácido nucleico (tal como el ADN y el ARN) tienen "extremos 5'" y "extremos 3'" porque los mononucleótidos se hacen reaccionar para formar polinucleótidos de manera tal que el fosfato 5' de un anillo de pentosa mononucleotídico esté unido al oxígeno 3' de su vecino en una dirección a través de un enlace fosfodiéster. Por lo tanto, un extremo de un polinucleótido lineal se denomina "extremo 5'" cuando su fosfato 5' no está unido al oxígeno 3' de un anillo de pentosa mononucleotídico. El otro extremo de un polinucleótido se denomina "extremo 3'" cuando su oxígeno 3' no está unido a un fosfato 5' de otro anillo de pentosa mononucleotídico. A pesar de que un fosfato 5' de un anillo de pentosa mononucleotídico está unido al oxígeno 3' de su vecino, también se puede decir que una secuencia interna de un ácido nucleico tiene extremos 5' y 3'.

En una molécula de ácido nucleico lineal o circular, los elementos internos discretos se denominan "**cadena arriba**" o 5' de los elementos "**cadena abajo**" o 3'. Con respecto al ADN, esta terminología refleja que la transcripción procede en una dirección de 5' a 3' a lo largo de una cadena de ADN. Los elementos promotores y potenciadores, que transcriben directamente un gen unido, generalmente se localizan 5' o cadena arriba de la región codificante. Sin embargo, los elementos potenciadores pueden ejercer su efecto incluso cuando se localizan 3' del elemento promotor y de la región codificante. Las señales de terminación de la transcripción y de poliadenilación se encuentran a 3' o cadena abajo de la región codificante.

Virus Adenoasociados (AVV): Un virus pequeño, defectuoso en la replicación, sin envoltura que infecta a los seres humanos y algunas otras especies de primates. No se sabe que el AAV cause enfermedad y provoca una respuesta inmunitaria muy leve. Los vectores de terapia génica que utilizan AAV pueden infectar tanto las células en división como las quiescentes y pueden persistir en un estado extracromosómico sin integrarse en el genoma de la célula hospedadora. Estas características hacen que AAV sea un vector vírico atractivo para terapia génica. Actualmente hay 11 serotipos reconocidos de AAV (AAV1-11).

Administración/Administrar: Proporcionar o dar a un sujeto un agente, tal como un agente terapéuticos (por ejemplo, un AVV recombinante), mediante cualquier vía eficaz. Las vías de administración ejemplares incluyen, aunque sin limitación, inyección (tal como subcutánea, intramuscular, intradérmica, intraperitoneal e intravenosa), vía oral, intraductal, sublingual, rectal, transdérmica, intranasal, vaginal y por inhalación.

Ensayo de Tiempo de Sangrado: Un análisis utilizado para medir la cantidad de tiempo que tarda la sangre de un sujeto en coagularse. Se coloca un manguito de presión arterial en la parte superior del brazo y se infla. Se hacen dos incisiones en la parte inferior del brazo. Estas tienen aproximadamente 10 mm (menos de 1/2 pulgada) de largo y 1 mm de profundidad (solo lo suficientemente profundo como para causar un sangrado mínimo). El manguito de presión arterial se desinfla inmediatamente. El papel secante se hace tocar los cortes cada 30 segundos hasta que se detiene el sangrado. Se registra el tiempo que tardan los cortes en detener el sangrado. En individuos normales, no hemofílicos, el sangrado se detiene en aproximadamente uno a diez minutos y puede variar de un laboratorio a otro, dependiendo de cómo se mida el ensayo. Por el contrario, los hemofílicos graves que tienen menos del 1 % de los niveles normales del factor de coagulación apropiado tienen un tiempo de coagulación de la sangre total mayor de 60 minutos. En ratones, el tiempo de sangrado se analiza seccionando la punta de la cola y tocando periódicamente con un papel secante hasta que se forme un coágulo en la punta de la cola. El tiempo normal de sangrado es entre 2-4 minutos. Por el contrario, los ratones hemofílicos que tienen menos del 1 % de los niveles normales del factor de

coagulación apropiado tienen un tiempo de sangrado de más de 15 minutos.

ADNc (AND complementario): Un fragmento de ADN que carece de segmentos internos no codificantes (intrones) y secuencias reguladoras que determinan la transcripción. El ADNc se sintetiza en el laboratorio por transcripción inversa del ARN mensajero extraído de las células. El ADNc también puede contener regiones no traducidas (UTR por sus siglas en inglés) que son responsables del control traduccional en la molécula de ARN correspondiente.

Trastorno de la coagulación: Un término general para una amplia gama de problemas médicos que conducen a una mala coagulación de la sangre y sangrado continuo. Los médicos también se refieren a los trastornos de la coagulación por términos como, por ejemplo, coagulopatía, sangrado anormal y trastornos hemorrágicos. Los trastornos de la coagulación incluyen cualquier defecto congénito, adquirido o inducido que dé como resultado un sangrado anormal (o patológico). Los ejemplos incluyen, aunque sin limitación, trastornos de coagulación insuficiente o hemostasia, tal como hemofilia A (una deficiencia en fVIII), hemofilia B (una deficiencia en fIX), hemofilia C (una deficiencia en el Factor XI), otras deficiencias en factores de coagulación (tal como Factor VII o fXIII), niveles anormales de inhibidores de factores de coagulación, trastornos plaquetarios, trombocitopenia, deficiencia de vitamina K y enfermedad de von Willebrand.

Algunos trastornos de coagulación están presentes al nacer y en algunos casos son trastornos hereditarios. Ejemplos específicos incluyen, pero sin limitación: hemofilia A, hemofilia B, deficiencia de proteína C y enfermedad de Von Willebrand. Algunos trastornos de la coagulación se desarrollan durante determinadas enfermedades (tal como la deficiencia de vitamina K, enfermedad hepática grave) o tratamientos (tal como el uso de fármacos anticoagulantes o el uso prolongado de antibióticos).

Factor de coagulación: Incluye cualquier proteína que promueva la hemostasia adecuada. En una realización, un factor de coagulación es fVIII o fIX, o una variante o fragmento de los mismos que conserva su actividad hemostática, por ejemplo, como se mide usando un ensayo APTT o un ensayo de tiempo de sangrado. En algunas realizaciones, cuando se administra en una cantidad terapéuticamente eficaz, el factor de coagulación aumenta la hemostasia en un sujeto que padece un trastorno de la coagulación, tal como hemofilia.

Factor de Coagulación VIII (fVIII): El fVIII es una proteína necesaria para la coagulación eficaz de la sangre y funciona en la coagulación como cofactor en la activación del factor X por fIX. Una concentración de aproximadamente 100 ng/ml para fVIII en la sangre se considera en el intervalo normal. La deficiencia de fVIII está asociada con la hemofilia A, y pueden producirse formas graves de la enfermedad cuando un sujeto tiene menos del 1 % de la cantidad normal de fVIII (es decir, menos de aproximadamente 1 ng de fVIII por ml de sangre). El fVIII se sintetiza como una proteína precursora de cadena sencilla de 2351 aminoácidos, que se procesa proteolíticamente. El gen del factor VIII humano (186.000 pares de bases) consta de 26 exones que varían en tamaño de 69 a 3.106 pb e intrones de hasta 32,4 kilobases (kb). Ejemplos de secuencias del ácido nucleico y proteína fVIII están disponibles públicamente (por ejemplo, véanse los números de acceso de Genbank: K01740, M14113 y E00527). En el presente documento se proporcionan variantes de fVIII que conservan la actividad de fVIII para la coagulación de la sangre pero se reducen en tamaño, tal como las variantes de fVIII que carecen del dominio B de fVIII. Variantes de fVIII ejemplares incluyen las variantes HSQ y ET3.

Factor de Coagulación IX (fIX): fIX es una proteína dependiente de la vitamina K necesaria para la coagulación eficaz de la sangre, y funciona en la coagulación como un activador del factor X. Una concentración de aproximadamente 1-5 µg/ml de fIX en la sangre se considera en el intervalo normal. La deficiencia de fIX está asociada con la hemofilia B, y se producen casos graves cuando la concentración de fIX es inferior a aproximadamente el 1 % de la concentración normal de fIX (es decir, inferior a aproximadamente 0,01-0,05 µg de fIX por ml de sangre). Las secuencias del ácido nucleico y proteína fIX están disponibles públicamente (por ejemplo, véanse Kurachi et al., 1982. Proc. Natl. Acad. Sci. U.S.A. 79(21):6461-4; Números de acceso de GenBank: J00136, XM045316, K02402, J00137 y M11309).

Codones optimizados: Un ácido nucleico "con codones optimizados" se refiere a una secuencia de ácido nucleico que se ha alterado de tal manera que los codones son óptimos para la expresión en un sistema particular (tal como una especie o grupo de especies particulares). Por ejemplo, una secuencia de ácido nucleico puede optimizarse para la expresión en células de mamífero o en una especie de mamífero particular (tal como las células humanas). La optimización de codones no altera la secuencia de aminoácidos de la proteína codificada.

El término "codones de aminoácidos específicos de hígado" se refiere a codones que se utilizan de manera diferencial representados en genes altamente expresados dentro del hígado humano en comparación con el uso de codones de toda la región codificante del genoma humano. Una estrategia que usa una cantidad máxima de codones de aminoácidos específicos de hígado busca evitar los codones que están subrepresentados, por ejemplo, debido a las bajas cantidades de ARNt con coincidencia de codones en las células del hígado que dan como resultado una traducción de proteínas más lenta.

Control: Un patrón de referencia. El control puede ser una muestra de control negativo obtenida de un paciente sano. En otros casos, el control es una muestra de control positivo obtenida de un paciente diagnosticado con hemofilia. En aún otros casos, el control es un control histórico o un valor de referencia patrón o un intervalo de valores (tal como

una muestra de control previamente probada, tal como un grupo de pacientes con hemofilia A con pronóstico o resultado conocido, o un grupo de muestras que representan valores basales o normales).

Una diferencia entre una muestra de prueba y un control puede ser un aumento o, por el contrario, una disminución. La diferencia puede ser una diferencia cualitativa o una diferencia cuantitativa, por ejemplo, una diferencia estadísticamente significativa. En algunos ejemplos, una diferencia es un aumento o disminución, en relación con un control, de al menos aproximadamente un 5 %, tal como al menos aproximadamente un 10 %, al menos aproximadamente un 20 %, al menos aproximadamente un 30 %, al menos aproximadamente un 40 %, al menos aproximadamente un 50 %, al menos aproximadamente un 60 %, al menos aproximadamente un 70 %, al menos aproximadamente un 80 %, al menos aproximadamente un 90 %, al menos aproximadamente un 100 %, al menos aproximadamente un 150 %, al menos aproximadamente un 200 %, al menos aproximadamente un 250 %, al menos aproximadamente un 300 %, al menos aproximadamente un 350 %, al menos aproximadamente un 400 %, al menos aproximadamente un 500 % o más del 500 %.

ADN (ácido desoxirribonucleico): El ADN es un polímero de cadena larga que comprende el material genético de la mayoría de los organismos vivos (algunos virus tienen genes que comprenden ácido ribonucleico (ARN)). Las unidades que se repiten en los polímeros de ADN son cuatro nucleótidos diferentes, cada uno de los cuales comprende una de las cuatro bases, adenina (A), guanina (G), citosina (C) y timina (T) unidas a un azúcar desoxirribosa al cual se fija un grupo fosfato. Los tripletes de nucleótidos (denominados codones) codifican para cada aminoácido en un polipéptido, o para una señal de parada. El término codón también se usa para las secuencias correspondientes (y complementarias) de tres nucleótidos en el ARNm en el que se transcribe la secuencia de ADN.

A menos que se especifique lo contrario, cualquier referencia a una molécula de ADN está destinada a incluir el complemento inverso de esa molécula de ADN. Excepto cuando el texto en el presente documento necesite una hebra simple, las moléculas de ADN, aunque escritas para representar solamente una hebra simple, abarcan ambas hebras de una molécula de ADN bicatenario. Por lo tanto, una referencia a la molécula de ácido nucleico que codifica una proteína específica, o un fragmento de la misma, abarca tanto la cadena en sentido como su complemento inverso. Por ejemplo, es apropiado generar sondas o cebadores a partir de la secuencia del complemento inverso de las moléculas de ácido nucleico divulgadas.

Potenciador: Una secuencia de ácido nucleico que aumenta la velocidad de transcripción al aumentar la actividad de un promotor.

Flanqueante: Cerca o al lado de, también, incluyendo adyacente, por ejemplo, en un polinucleótido lineal o circular, tal como una molécula de ADN.

Gen: Una secuencia de un ácido nucleico, generalmente una secuencia de ADN, que comprende secuencias de control y codificantes necesarias para la transcripción de un ARN, ya sea un ARNm o no. Por ejemplo, un gen puede comprender un promotor, uno o más potenciadores o silenciadores, una secuencia de un ácido nucleico que codifica un ARN y/o un polipéptido, secuencias reguladoras cadena abajo y, posiblemente, otras secuencias de los ácidos nucleicos involucradas en la regulación de la expresión de un ARNm.

Como es bien conocido en la técnica, la mayoría de los genes eucariotas contienen tanto exones como intrones. El término "exón" se refiere a una secuencia de un ácido nucleico encontrada en el ADN genómico que se predice bioinformáticamente y/o se confirma experimentalmente que contribuye a una secuencia contigua a un transcrito de ARNm maduro. El término "intrón" se refiere a una secuencia de un ácido nucleico encontrada en el ADN genómico que se predice y/o confirma que no contribuye a un transcrito de ARNm maduro, sino que se "corta y empalma" durante el procesamiento de la transcripción.

Terapia génica: La introducción de una molécula de ácido nucleico heterólogo en una o más células receptoras, en donde la expresión del ácido nucleico heterólogo en la célula receptora afecta la función de la célula y produce un efecto terapéutico en un sujeto. Por ejemplo, la molécula heteróloga de ácido nucleico puede codificar una proteína, que afecta la función de la célula receptora.

Hemofilia: Un trastorno de la coagulación de la sangre causado por una actividad deficiente del factor de coagulación, que disminuye la hemostasia. Las formas graves se producen cuando la concentración del factor de coagulación es inferior a aproximadamente el 1 % de la concentración normal del factor de coagulación en un sujeto normal. En algunos sujetos, la hemofilia se debe a una mutación genética que da como resultado una expresión alterada de un factor de coagulación. En otros, la hemofilia es un trastorno autoinmunitario, denominado hemofilia adquirida, en el que los anticuerpos que se generan contra un factor de coagulación en un sujeto provocan una disminución de la hemostasia.

La hemofilia A es el resultado de una deficiencia del fVIII de coagulación funcional, mientras que la hemofilia B es el resultado de una deficiencia del fIX de coagulación funcional. Estas afecciones que se deben a una mutación genética son causadas por un rasgo recesivo hereditario ligado al sexo con el gen defectuoso ubicado en el cromosoma X, y esta enfermedad, por lo tanto, generalmente se encuentra solo en varones. La gravedad de los síntomas puede variar

con esta enfermedad, y las formas graves se hacen evidentes desde el principio. El sangrado es el sello distintivo de la enfermedad y generalmente se produce cuando se circuncida a un bebé varón. Otras manifestaciones hemorrágicas aparecen cuando el bebé se vuelve capaz de desplazarse. Los casos leves pueden pasar desapercibidos hasta más adelante en la vida cuando se producen en respuesta a una cirugía o trauma. El sangrado interno puede producirse en cualquier lugar, y es común el sangrado en las articulaciones.

Hemostasia: Detención de sangre hemorrágica por formación de coágulos sanguíneos. El tiempo de coagulación de la sangre es el tiempo que tarda la sangre periférica en coagularse utilizando un ensayo de tiempo de tromboplastina parcial activado (APTT) o midiendo el tiempo de sangrado. En una realización particular, el tiempo de coagulación de la sangre disminuye al menos un 50 %, por ejemplo, al menos un 60 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 %, al menos un 99 % o incluso aproximadamente un 100 % (es decir, el tiempo de coagulación sanguínea es similar al observado para un sujeto normal) en comparación con el tiempo de coagulación sanguínea del sujeto antes de la administración de un vector terapéutico que codifica el factor de coagulación apropiado como se describe en el presente documento. En otra realización más, el tiempo de coagulación sanguínea en el sujeto afectado se corrige de aproximadamente el 50 % de un sujeto normal, a aproximadamente el 75 % de un sujeto normal, a aproximadamente el 90 % de un sujeto normal, a aproximadamente el 95 %, por ejemplo, aproximadamente el 100 %, después de la administración oral de una cantidad terapéuticamente eficaz del factor de coagulación apropiado. Como se usan en el presente documento, "aproximadamente" se refiere a más o menos el 5 % de un valor de referencia. Por lo tanto, aproximadamente el 50 % se refiere de 47,5 % a 52,5 %.

Intrón: Un tramo de ADN dentro de un gen que no contiene información codificante para una proteína. Los intrones se eliminan antes de la traducción de un ARN mensajero.

Repetición terminal invertida (ITR): Secuencias simétricas del ácido nucleico en el genoma de virus adenoasociados necesarias para una replicación eficaz. Las secuencias ITR se encuentran en cada extremo del genoma de ADN de AAV. Las ITR sirven como los orígenes de replicación para la síntesis de ADN vírico y son componentes en *cis* esenciales para generar vectores integradores de AAV.

Aislado: Un componente biológico "aislado" (tal como una molécula de ácido nucleico, proteína, virus o célula) se ha separado o purificado sustancialmente de otros componentes biológicos en la célula o en el tejido del organismo, o en el propio organismo, en el que el componente se produce naturalmente, tal como otros ADN y ARN cromosómicos y extracromosómicos, proteínas y células. Las moléculas de ácido nucleico y las proteínas que se han "aislado" incluyen los purificados por métodos de purificación convencionales. El término también abarca moléculas de ácido nucleico y proteínas preparados por expresión recombinante en una célula hospedadora, así como moléculas de ácido nucleico y proteínas sintetizados químicamente.

Moléculas de ácido nucleico: Una forma polimérica de nucleótidos, que puede incluir cadenas de ARN, ADNc, ADN genómico tanto en sentido como antisentido y formas sintéticas y polímeros mixtos de los anteriores. Un nucleótido se refiere a un ribonucleótido, desoxinucleótido o una forma modificada de cualquier tipo de nucleótido. La expresión "molécula de ácido nucleico" como se usa en el presente documento es sinónimo de "ácido nucleico" y "polinucleótido". Una molécula de ácido nucleico suele tener al menos 10 bases de longitud, a menos que se especifique otra cosa. El término incluye formas de ADN monocatenarias y bicatenarias. Un polinucleótido puede incluir bien uno o ambos nucleótidos de origen natural y modificados unidos entre sí por enlaces nucleotídicos de origen natural y/o no de origen natural. "ADNc" se refiere a un ADN que es complementario o idéntico a un ARNm, en forma monocatenaria o bicatenaria. "Codificación" se refiere a la propiedad inherente de secuencias específicas de nucleótidos en un polinucleótido, tal como un gen, un ADNc o un ARNm, para servir como plantillas para la síntesis de otros polímeros y macromoléculas en procesos biológicos que tienen una secuencia definida de nucleótidos (es decir, ARNr, ARNt y ARNm) o una secuencia definida de aminoácidos y las propiedades biológicas resultantes de los mismos.

Nucleótido: Este término incluye, aunque sin limitación, un monómero que incluye una base unida a un azúcar, tal como una pirimidina, purina o análogos sintéticos de las mismas, o una base unida a un aminoácido, tal como en un ácido nucleico peptídico (PNA). Un nucleótido es un monómero en un polinucleótido. Una secuencia de nucleótidos se refiere a la secuencia de bases en un polinucleótido.

Unido operativamente: Una primera secuencia de un ácido nucleico está operativamente unida con una segunda secuencia de un ácido nucleico cuando la primera secuencia de un ácido nucleico se coloca en una relación funcional con la segunda secuencia de un ácido nucleico. Por ejemplo, un promotor está unido operativamente a una secuencia codificante si el promotor afecta la transcripción o expresión de la secuencia codificante. Por lo general, las secuencias de ADN unidas operativamente son contiguas y, cuando es necesario unir dos regiones codificantes de proteínas, en el mismo marco de lectura.

ORF (marco de lectura abierto): Una serie de tripletes de nucleótidos (codones) que codifican aminoácidos. Estas secuencias son generalmente traducibles en un péptido.

Transportadores farmacéuticamente aceptables: Los transportadores farmacéuticamente aceptables para su uso

son convencionales. Remington's Pharmaceutical Sciences, por E. W. Martin, Mack Publishing Co., Easton, PA, 19ª edición, 1995, describe composiciones y formulaciones adecuadas para el suministro farmacéutico de los vectores divulgados.

- 5 En general, la naturaleza del transportador dependerá del modo particular de administración que se emplee. Por ejemplo, las formulaciones parenterales generalmente comprenden líquidos inyectables que incluyen líquidos farmacéuticamente y fisiológicamente aceptables tales como agua, solución salina fisiológica, soluciones salinas equilibradas, dextrosa acuosa, glicerol o similares como vehículo. Para las composiciones sólidas (por ejemplo, formas en polvo, píldora, comprimido o cápsula), los transportadores sólidos no tóxicos convencionales pueden incluir, por ejemplo, grados farmacéuticos de manitol, lactosa, almidón y estearato de magnesio. Además de transportadores biológicamente neutros, las composiciones farmacéuticas (tales como las composiciones de vectores) a administrar pueden contener cantidades menores de sustancias auxiliares no tóxicas, tales como agentes humectantes o emulsionantes, conservantes y agentes tamponadores del pH y similares, por ejemplo, acetato de sodio o monolaurato de sorbitán. En realizaciones particulares, el transportador adecuado para la administración a un sujeto puede ser estéril y/o estar suspendido o contenido en una forma de dosificación unitaria que contiene una o más dosis medidas de la composición adecuada para inducir la respuesta inmunitaria deseada. También puede ir acompañado de medicamentos para su uso con fines de tratamiento. La forma de dosificación unitaria puede ser, por ejemplo, en un vial sellado que contiene contenido estéril o una jeringa para inyección en un sujeto, o liofilizado para su posterior solubilización y administración o en una dosificación de liberación controlada o sólida.

20 **Polipéptido:** Cualquier cadena de aminoácidos, independientemente de la longitud o la modificación postraducciona (p. ej., glucosilación o fosforilación). "Polipéptido" se aplica a los polímeros de aminoácidos, incluidos los polímeros de aminoácidos de origen natural y polímeros de aminoácidos de origen no natural, así como en los que uno o más restos de aminoácidos es un aminoácido no natural, por ejemplo, un mimético químico artificial de un aminoácido de origen natural correspondiente. Un "resto" se refiere a un aminoácido o mimético de aminoácido incorporado en un polipéptido por un enlace amida o un mimético de enlace amida. Un polipéptido tiene un extremo amino terminal (N-terminal) y un extremo carboxilo terminal (C-terminal). "Polipéptido" se usa indistintamente con péptido o proteína, y se usa en el presente documento para referirse a un polímero de restos de aminoácidos.

30 **Prevenir, tratar o mejorar una enfermedad:** Prevenir "una enfermedad (tal como la hemofilia) se refiere a inhibir el desarrollo completo de una enfermedad. "Tratar" se refiere a una intervención terapéutica que mejora un signo o síntoma de una enfermedad o afección patológica después de haber comenzado a desarrollarse. Mejorar "se refiere a la reducción en el número o la gravedad de los signos o síntomas de una enfermedad.

35 **Promotor:** Una región de ADN que dirige/inicia la transcripción de un ácido nucleico (por ejemplo, un gen). Un promotor incluye secuencias de un ácido nucleico necesarias cerca del sitio de inicio de la transcripción. Generalmente, los promotores se localizan cerca de los genes que transcriben. Un promotor también incluye opcionalmente elementos potenciadores o represores distales que pueden ubicarse hasta varios miles de pares de bases desde el sitio de inicio de la transcripción. Un promotor específico de tejido es un promotor que dirige/inicia la transcripción principalmente en un solo tipo de tejido o célula. Por ejemplo, un promotor específico de hígado es un promotor que dirige/inicia la transcripción en el tejido hepático en un grado sustancialmente mayor que otros tipos de tejidos.

45 **Proteína:** Una molécula biológica expresada por un gen u otro ácido nucleico codificante (por ejemplo, un ADNc) y compuesta de aminoácidos.

50 **Purificado:** El término "purificado" no requiere pureza absoluta; si no que, se concibe como un término relativo. Por lo tanto, por ejemplo, un péptido, proteína, virus u otro compuesto activo purificado es uno que está aislado total o parcialmente de proteínas y otros contaminantes asociados naturalmente. En determinadas realizaciones, la expresión "sustancialmente purificado" se refiere a un péptido, proteína, virus u otro compuesto activo que ha sido aislado de una célula, medio de cultivo celular u otra preparación en bruto y sometido a fraccionamiento para eliminar varios componentes de la preparación inicial, tales como proteínas, residuos celulares y otros componentes.

55 **Recombinante:** Una molécula de ácido nucleico recombinante es aquella que tiene una secuencia que no es de origen natural, por ejemplo, incluye una o más sustituciones, eliminaciones o inserciones en ácidos nucleicos, y/o tiene una secuencia que está hecha por una combinación artificial de dos segmentos de secuencia separados de otra manera. Esta combinación artificial se puede lograr mediante síntesis química o, más habitualmente, mediante la manipulación artificial de segmentos aislados de ácidos nucleicos, por ejemplo, mediante técnicas de ingeniería genética.

60 Un virus recombinante es uno que incluye un genoma que incluye una molécula de ácido nucleico recombinante. Como se usan en el presente documento, **"AAV recombinante"** se refiere a una partícula de AAV en la que se ha empaquetado una molécula de ácido nucleico recombinante (tal como una molécula de ácido nucleico recombinante que codifica un factor de coagulación).

65 Una proteína recombinante es aquella que tiene una secuencia que no es de origen natural o tiene una secuencia que está hecha por una combinación artificial de dos segmentos de secuencia separados de otra manera. En varias realizaciones, una proteína recombinante está codificada por un ácido nucleico heterólogo (por ejemplo, recombinante)

que se ha introducido en una célula hospedadora, tal como una célula bacteriana o eucariota, o en el genoma de un virus recombinante.

Elemento de respuesta (RE): Una secuencia de ADN incluida en un promotor al que uno o más factores de transcripción pueden unirse y conferir un aspecto de control de la expresión génica.

Identidad de secuencia: La identidad o similitud más secuencias de los ácidos nucleicos o dos o más secuencias de aminoácidos, se expresa en términos de identidad o similitud entre las secuencias. La identidad de secuencia puede medirse en términos de porcentaje de identidad; cuanto mayor sea el porcentaje, más idénticas son las secuencias. La similitud de secuencia puede medirse en términos de porcentaje de similitud (que tiene en cuenta sustituciones de aminoácidos conservativas); cuanto mayor sea el porcentaje, más similares son las secuencias. Los homólogos y ortólogos de secuencias de los ácidos nucleicos o de aminoácidos poseen un grado relativamente alto de identidad/similitud de secuencia cuando se alinean usando métodos convencionales. Esta homología es más significativa cuando las proteínas o ADNc ortólogos proceden de especies que están más estrechamente relacionadas (tales como secuencias humanas y de ratón), en comparación con especímenes más remotamente relacionados (tales como secuencias humanas y de *C. elegans*).

Los métodos de alineación de secuencias para la comparación son bien conocidos en la técnica. Se describen diversos programas y algoritmos de alineación en: Smith & Waterman, Adv. Appl. Math. 2:482, 1981; Needleman & Wunsch, J. Mol. Biol. 48:443, 1970; Pearson & Lipman, Proc. Natl. Acad. Sci. USA 85:2444, 1988; Higgins & Sharp, Gene, 73:237-44, 1988; Higgins & Sharp, CABIOS 5:151-3, 1989; Corpet et al., Nuc. Acids Res. 16:10881-90, 1988; Huang et al. Computer Appls. in the Biosciences 8, 155-65, 1992; y Pearson et al., Meth. Mol. Bio. 24:307-31, 1994. Altschul et al., J. Mol. Biol. 215:403-10, 1990, presenta una consideración detallada de métodos de alineación de secuencias y cálculos de homología.

La Basic Local Alignment Search Tool (BLAST) del NCBI (Altschul y col., J. Mol. Biol. 215:403-10, 1990) está disponible a partir de varias fuentes, entre los que se incluye el Centro Nacional de Información Biológica (NCBI) y en internet, para su uso en conexión con los programas de análisis de secuencias blastp, blastn, blastx, blastn y tblastx. Puede encontrarse información adicional en la página web del NCBI.

Como se usan en el presente documento, referencia a "al menos un 90 % de identidad" se refiere a "al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o incluso al menos un 100 % de identidad" con una secuencia de referencia especificada.

Sujeto: Organismos vertebrados multicelulares vivos, una categoría que incluye mamíferos humanos y no humanos.

Sintético: Producido por medios artificiales en un laboratorio, por ejemplo, un ácido nucleico sintético puede sintetizarse químicamente en un laboratorio.

Caja TATA: Una secuencia de ADN que se encuentra en la región promotora de un gen que puede unirse mediante la proteína de unión a TATA y el factor de transcripción II D durante el desenrollado y unión de ADN por la ARN polimerasa II. Una secuencia de caja TATA generalmente incluye una secuencia TATAAA y a menudo incluye nucleótidos de adenina 3' adicionales. Se proporciona una secuencia de caja TATA ejemplar como los nucleótidos 108-114 de la SEQ ID NO: 4.

Cantidad terapéuticamente eficaz: Una cantidad de un agente farmacéutico o terapéutico especificado (por ejemplo, un AAV recombinante) suficiente para lograr un efecto deseado en un sujeto, o en una célula, que se está tratando con el agente. La cantidad eficaz del agente dependerá de varios factores, incluyendo, pero sin limitación, el sujeto o las células a tratar, y la forma de administración de la composición terapéutica.

Factor de transcripción (TF): Una proteína que se une a secuencias de ADN específicas y, por lo tanto, controla la transferencia (o transcripción) de información genética del ADN al ARN. Los TF realizan esta función solos o con otras proteínas en un complejo, promoviendo (como activador) o bloqueando (como represor) el reclutamiento de ARN polimerasa (la enzima que realiza la transcripción de información genética de ADN a ARN) para genes específicos. Las secuencias de ADN específicas a las que se une un TF se conocen como elemento de respuesta (RE) o elemento regulador. Otros nombres incluyen elemento en cis y elemento regulador transcripcional de acción en cis.

Los factores de transcripción interactúan con sus sitios de unión usando una combinación de fuerzas electrostáticas (de las cuales los enlaces de hidrógeno son un caso especial) y de Van der Waals. Debido a la naturaleza de estas interacciones químicas, la mayoría de los factores de transcripción se unen al ADN de una manera específica de secuencia. Sin embargo, no todas las bases en el sitio de unión al factor de transcripción pueden interactuar realmente con el factor de transcripción. Además, algunas de estas interacciones pueden ser más débiles que otras. Por lo tanto, muchos factores de transcripción no se unen a una sola secuencia, sino que son capaces de unir un subconjunto de secuencias estrechamente relacionadas, cada una con una fuerza de interacción diferente.

Por ejemplo, aunque el sitio de unión consenso para la proteína de unión a TATA (TBP) es TATAAAA; sin embargo, el factor de transcripción de TBP también puede unir secuencias similares como TATATAT o

TATATAA

Los factores de transcripción (TF) se clasifican en función de muchos aspectos. Por ejemplo, las estructuras secundarias, terciarias y cuaternarias de la proteína estructuran la secuencia y las propiedades de unión al ADN, la interacción con la doble hélice del ADN y el metal y otras características de unión. La base de datos JASPAR y TRANSFAC (TRANSFAC® 7.0 Public 2005) son dos bases de datos de factores de transcripción basadas en la web, sus sitios de unión probados experimentalmente y genes regulados.

HNF1a: También denominado caja homeótica A de HNF1 A o HNF1, la proteína HNF1a es un factor de transcripción necesario para la expresión de varios genes específicos de hígado. HNF1a forma un homodímero que se une a secuencias promotoras particulares. Los sitios de unión de TF HNF1a ejemplares incluyen el sitio de unión al TF "HNF1a" proporcionado como los nucleótidos 1-12 de la SEQ ID NO: 4, el sitio de unión al TF "HNF1-1" proporcionado como los nucleótidos 16-23 de la SEQ ID NO: 4, y el sitio de unión al TF "HNF1-2" proporcionado como los nucleótidos 48-62 de la SEQ ID NO: 4. (Véanse, por ejemplo, PubMed Gene ID NO. 6927; Chi et al., "Diabetes mutations delineate an atypical POU domain in HNF-1alpha", Mol. Cell., 10:1129-1137, 2002; y Rose et al., "Structural basis of dimerization, coactivator recognition and MODY3 mutations in HNF-1alpha", Nat. Struct. Biol. 7:744-748, 2000).

HNF3a: Un factor de transcripción necesario para la expresión de varios genes específicos de hígado. HNF3a se une a secuencias promotoras particulares. Se proporciona un sitio de unión al TF HNF3a ejemplar como los nucleótidos 39-45 de la SEQ ID NO: 4. Se proporciona un sitio de unión al TF HNF3-2 ejemplar como los nucleótidos 65-71 de la SEQ ID NO: 4. (véase, por ejemplo, Laganieri et al., "Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response", Proc. Natl. Acad. Sci. U.S.A. 102:11651-11656, 2005; Williamson et al., "BRCA1 and FOXA1 proteins coregulate the expression of the cell cycle-dependent kinase inhibitor p27(Kip1)", Oncogene 25:1391-1399, 2006; Lupien et al., "FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription", Cell 132:958-970, 2008; Song et al., "Role of Foxa1 in regulation of bcl2 expression during oxidative-stress-induced apoptosis in A549 type II pneumocytes", Cell Stress Chaperones, 14:417-425, 2009; and Malik et al., "Histone deacetylase 7 and FoxA1 in estrogen-mediated repression of RPRM", Mol. Cell. Biol. 30:399-412, 2010).

HNF4: Un factor de transcripción necesario para la expresión de varios genes específicos de hígado. HNF4 se une a secuencias promotoras particulares. Se proporciona un sitio de unión al TF HNF4 ejemplar como los nucleótidos 26-36 de la SEQ ID NO: 4. (Véanse, por ejemplo, Wang et al., "Hepatocyte nuclear factor-4α interacts with other hepatocyte nuclear factors in regulating transthyretin gene expression", FEBS J., 277(19):4066-75, 2010).

HP1: Un factor de transcripción necesario para la expresión de varios genes específicos de hígado. HP1 se une a secuencias promotoras particulares. Se proporciona un sitio de unión al TF HP1 ejemplar como los nucleótidos 75-87 de la SEQ ID NO: 4. (Véanse, por ejemplo, Schorpp et al., "Hepatocyte-specific promoter element HP1 of the Xenopus albumin gene interacts with transcriptional factors of mammalian hepatocytes", J Mol Biol., 202(2):307-20, 1988)

Sitio de Inicio de Transcripción: La ubicación donde comienza la transcripción en el extremo 5' de una secuencia génica. Se proporciona un Sitio de Inicio de Transcripción ejemplar como los nucleótidos 116-146 de la SEQ ID NO: 4.

Cantidad terapéuticamente eficaz: La cantidad de agente, tal como un vector AAV recombinante divulgado que codifica un factor de coagulación, que es suficiente para prevenir, tratar (incluida la profilaxis), reducir y/o mejorar los síntomas y/o las causas subyacentes de un trastorno o enfermedad, por ejemplo para prevenir, inhibir y/o tratar la hemofilia. Por ejemplo, esta puede ser la cantidad necesaria para inhibir o prevenir la replicación vírica o para alterar de manera medible los síntomas externos de la enfermedad o afección.

En un ejemplo, una respuesta deseada es reducir el tiempo de coagulación en un sujeto (tal como un sujeto con hemofilia), por ejemplo, medido usando un ensayo de tiempo de sangrado. El tiempo de coagulación no necesita ser completamente restaurado al de los sujetos sanos normales sin hemofilia para que el método sea eficaz. Por ejemplo, la administración de una cantidad terapéuticamente eficaz de un vector (tal como un vector que codifica fVIII) como se divulga en el presente documento puede disminuir el tiempo de coagulación (u otro síntoma de la hemofilia) en una cantidad deseada, por ejemplo, al menos 20 %, al menos un 30 %, al menos un 40 %, al menos un 50 %, al menos un 60 %, al menos un 70 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 %, al menos un 100 % o más, en comparación con un control adecuado.

Se entiende que para obtener una respuesta terapéutica a la enfermedad o afección puede necesitarse múltiples administraciones del agente. Por lo tanto, una cantidad terapéuticamente eficaz abarca una dosis fraccionaria que contribuye en combinación con administraciones anteriores o posteriores para lograr un resultado terapéutico en el paciente. Por ejemplo, se puede administrar una cantidad terapéuticamente eficaz de un agente en una dosis única o en varias dosis, por ejemplo diariamente, durante un ciclo de tratamiento. Sin embargo, la cantidad terapéuticamente eficaz puede depender del sujeto a tratar, la gravedad y el tipo de la afección a tratar, y la forma de administración. Se

puede empaquetar una forma de dosificación unitaria del agente en una cantidad terapéutica, o en múltiplos de la cantidad terapéutica, por ejemplo, en un vial (por ejemplo, con una tapa perforable) o una jeringa que tiene componentes estériles.

- 5 **Vector:** Un vector es una molécula de ácido nucleico que permite la inserción de ácido nucleico extraño sin alterar la capacidad del vector para replicarse y/o integrarse en una célula hospedadora. Un vector puede incluir secuencias de un ácido nucleico que le permiten replicarse en una célula hospedadora, tal como un origen de replicación. Un vector también puede incluir uno o más genes marcadores seleccionables y otros elementos genéticos. Un vector de expresión es un vector que contiene las secuencias reguladoras necesarias para permitir la transcripción y traducción del gen o genes insertados. En algunas realizaciones en el presente documento, el vector es un vector de virus adenoasociado (AAV). En algunas realizaciones, el vector es un vector gamma retrovítico, un vector lentivítico o un vector adenovítico.

15 A menos que se indique lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que el que entiende comúnmente una persona normalmente experta en la técnica a la cual pertenece la presente divulgación. Los términos en singular "un", "uno", "una", y "el/la" incluyen las referencias en plural a menos que el contexto dicte claramente lo contrario. "Que comprende A o B" significa que incluye A, o B, o A y B. Se entiende además que todos los tamaños de bases o los tamaños de aminoácidos, y todos los valores de pesos moleculares o masas moleculares, proporcionados para los ácidos nucleicos o polipéptidos son aproximados, y se proporcionan para la descripción. Aunque los métodos y materiales similares o equivalentes a los descritos en el presente documento pueden usarse en la práctica o prueba de la presente divulgación, a continuación se describen métodos y materiales adecuados. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluidas las explicaciones de términos. Además, los materiales, métodos y ejemplos son solamente ilustrativos y no pretenden ser limitantes.

25

II. Promotores Optimizados para la Transcripción Dirigida al Hígado

En el presente documento, se proporcionan nuevos promotores para promover la transcripción en tejido y/o células del hígado. Como se analiza en el Ejemplo 2, los nuevos promotores se diseñaron usando un enfoque iterativo que finalmente identificó varias secuencias promotoras que proporcionan niveles de transcripción inesperadamente altos (como se analiza midiendo la actividad de la proteína expresada), y son sustancialmente más cortas que las secuencias promotoras anteriores, como la secuencia promotora HLP.

35 En la invención, se proporciona una molécula de ácido nucleico recombinante que comprende un promotor que comprende un primer elemento de respuesta que comprende un conjunto de sitios de unión a factores de transcripción (TF), que incluyen: un sitio de unión al TF HNF1a, un sitio de unión al TF HNF1-1, un sitio de unión al TF HNF4, un sitio de unión al TF HNF3a, un sitio de unión al TF HNF1-2, un sitio de unión al TF HNF3-2, un sitio de unión al TF HP1, una caja TATA; y un Sitio de Inicio de la Transcripción. Estos son los sitios de unión al factor de transcripción incluidos en el promotor HCB.

40

En la invención, el primer elemento de respuesta puede comprender una secuencia de nucleótidos que no tiene más de 160 nucleótidos de longitud (tal como no más de 150 nucleótidos de longitud, tal como de 146 nucleótidos de longitud).

45 En la invención, el sitio de unión al TF HNF1a comprende o consiste en los nucleótidos 1-12 de la SEQ ID NO: 4; el sitio de unión al TF HNF1-1 comprende o consiste en los nucleótidos 16-23 de la SEQ ID NO: 4; el sitio de unión al TF HNF4 comprende o consiste en los nucleótidos 26-36 de la SEQ ID NO: 4; el sitio de unión al TF HNF3a comprende o consiste en los nucleótidos 39-45 de la SEQ ID NO: 4; el sitio de unión al TF HNF1-2 comprende o consiste en los nucleótidos 48-62 de la SEQ ID NO: 4; el sitio de unión al TF HNF3-2 comprende o consiste en los nucleótidos 65-71 de la SEQ ID NO: 4; el sitio de unión al TF HP1 comprende o consiste en los nucleótidos 75-87 de la SEQ ID NO: 4; la caja TATA comprende o consiste en los nucleótidos 108-114 de la SEQ ID NO: 4; y el Sitio de Inicio de la Transcripción (TSS) comprende o consiste en los nucleótidos 116-146 de la SEQ ID NO: 4.

50

55 En algunas realizaciones, el primer elemento de respuesta comprende, de 5' a 3', el sitio de unión al TF HNF1a, el sitio de unión al TF HNF1-1, el sitio de unión al TF HNF-4, el sitio de unión al TF HNF3a, el sitio de unión al TF HNF1-2, el sitio de unión al TF HNF3-2, el sitio de unión al TF HP1, la caja TATA y el Sitio de Inicio de la Transcripción (TSS, por sus siglas en inglés).

60 En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 4 (HCB), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

65 En algunas realizaciones, la molécula de ácido nucleico recombinante puede comprender un promotor que comprende el primer elemento de respuesta como se analiza anteriormente, y puede comprender además un segundo elemento de respuesta.

En algunas realizaciones, el segundo elemento de respuesta puede comprender un elemento de respuesta HSh. Por ejemplo, un elemento de respuesta HSh que comprende o que consiste en la secuencia de nucleótidos establecida como la SEQ ID NO: 111, o una secuencia de al menos un 90 % (tal como al menos 91%, al menos 92%, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunas realizaciones, el segundo elemento de respuesta puede comprender un elemento de respuesta 5'HS. Por ejemplo, un elemento de respuesta 5'HS que comprende o que consiste en la secuencia de nucleótidos establecida como los nucleótidos 6-32 de SEQ ID NO: 111, o una secuencia al menos un 90 % (tal como al menos 91%, al menos 92%, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunas realizaciones, el segundo elemento de respuesta puede comprender un elemento de respuesta 3'HS. Por ejemplo, un elemento de respuesta 3'HS que comprende o que consiste en la secuencia de nucleótidos establecida como los nucleótidos 44-68 de SEQ ID NO: 111, o una secuencia al menos un 90 % (tal como al menos 91%, al menos 92%, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 102 (HSh-HCB), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 104 (5'HSh-HCB), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunas realizaciones, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 103 (3'HSh-HCB),

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 7 (ABP-HP1-God-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 105 (HSh-SynO-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 106 (sHS-SynO-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 107 (Agro), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 108 (HS-SynO-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 112 (HNF1-ShortABPEXact-SynO-TSS-Int), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 5 (shortABP-HP1-God-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

En algunos aspectos de la divulgación, la molécula de ácido nucleico recombinante comprende un promotor que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 7 (ABP-HP1-God-TSS), o una secuencia al menos un 90 % (tal como, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 98 % o al menos un 99 %) idéntica a la misma.

Los promotores divulgados pueden utilizarse en cualquier situación en la que se desee la transcripción específica del hígado. En varias realizaciones, cualquiera de los promotores de la invención puede incluirse en un vector (tal como un vector AAV) para métodos de terapia génica en los que se desea la expresión específica de hígado de un transgén, tal como la expresión específica de hígado de un factor de coagulación como se divulga en el presente documento.

III. Moléculas de Ácido Nucleico Recombinante que Codifican Factores de Coagulación

El sistema de coagulación de la sangre es una cascada proteolítica. Los factores de coagulación sanguínea están presentes en el plasma como un zimógeno, en otras palabras, en forma inactiva, que al activarse sufre una escisión proteolítica para liberar el factor activo de la molécula precursora. El objetivo final es producir trombina. La trombina convierte el fibrinógeno en fibrina, que forma un coágulo.

El factor X es la primera molécula de la vía común y se activa mediante un complejo de moléculas que contienen fIX activado, fVIII, calcio y fosfolípidos que se encuentran sobre la superficie de las plaquetas. El FVIII es activado por la trombina, y facilita la activación del factor X por fIXa. FVIII, contiene múltiples dominios (A1-A2-B-ap-A3-C1-C2) y circula en la sangre en forma inactiva unido al factor von Willebrand (VWF, por sus siglas en inglés). El dominio C2 está implicado en la unión de fVIII a VWF. La trombina escinde el fVIII causando disociación con el VWF, conduciendo, finalmente a la formación de fibrina a través del fIX. La hemofilia A congénita está asociada con mutaciones genéticas en el gen fVIII y da como resultado una coagulación alterada debido a niveles de fVIII circulantes inferiores a lo normal. La hemofilia B se asocia de manera similar con mutaciones genéticas en el gen fIX.

Los límites del dominio FVIII se refieren a la numeración de la secuencia de aminoácidos de fVIII humano como sigue; restos 1-19 (Secuencia Señal), 20-391 (A1), 392-759 (A2), 760-1667 (B), 1668-1708 (ap), 1709-2038 (A3), 2039-2191 (C1) y 2192-2351 (C2) (véase Gitschier et al., Nature, 1984, 312, 326-330) de la SEQ ID NO: 1:

```

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPVRVPSFPPFNTSVVYKKTFLVEFTDHLF
NIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHITYV
WQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNS
LMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITF
LTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNPSFSFIQI
RSVAKKHPKTWVHYIAAEEDWDYAPLVLPDDRYSKQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGIL
GPLLYGEVGDTLIIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPR
CLTRYSSFVNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLED
PEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMYVEDTLTLFPFSGETVFMMSME
NPGWLILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDI SAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI
PENDIEKTDPFWAHRTMPKIQNVSSDILLMLLRQYDPTPHGLSLSDLQEAKEYTFSDDPSPGAIDSNNSLSEMTFRP
QLHSGDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSSTSNLISTIPSDNLAAGTDNTSSLGPPSPMPVHYDSQL
DTTLFGKKSSPLTESGGPLSLSEENNDKLLSGLMNSQESSWGKNVSSTESGRLFKGKRAHPALLTKDNALFKVSI
SLLKTNKTSNNSATNRKTHIDGPSLLIENSPSVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRLNHMSNKTSSKN
MEMVQQKKEGPIPPDAQNPDMSSFFKMLFLPESARWIQRTHGKNSLNSGQGPSKQLVSLGPEKSVEGQNFLESEKNKV
VGKGEFTKDVGLKEMVFPSSRNLFNLNLDNLHENNTHNQEKKIQEEIEKKETLIQENVVLPQIHTVTGTKNFMKNLFL
LSTRQNVESGYDGAAPVLQDFRSLNDSTNRKTHAHFSKKGEEENLEGLNQTKQIVEKYACTRISPTSQQNFV
TQRSKRALKQFRLPLEETELEKRIIVDDTSTQWSKNMHLTPSTLTQIDYNEKEKGAIQSPLSDCCLTRSHSIPQANR
SPLPIAKVSSFPISIRPIYLTRVLFDQDNSSHLPAASYRKKDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLG
TSATNSVTYKKVENTVLKPKDLPKTSKGVLELLPKVHIYQKDLFPETETSNGPSGHLDLVEGSLGQTEGAIKWNEANRP
GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSQEKSPKTAFFKKKDTILSLNACESNHAIAAINEGQNK
PEIEVTWAKQGRTERLCSQNPVLRKHQREITRTTLQSDQEEIDYDDTISVENMKKEDFDIYDEDENQSPRSFQKKTRH
YFIAAVERLWDYGMSSSPHVLNRNAQSGSVPPQFKKVVFQEFDTGDSFTQPLYRGELNEHLGLLPYIRAEVEDNIMVTF
RNQASRPYSFYSSLSIYEEDQRQGAEPKRNFKVPNETKTYFWKVQHHMPTKDEFDCAWAYFSDVDLEKDVHSGLIG
PLLVCHTNTLNPAHGRQVTQVEFALFFTIFDETSKSWYFTEMNERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPG
LVMAQDQIRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMLALNYLPGVFETVEMLPKAGIWRVECLIGELHAG
MSTLFLVYSNKCQTPGLMASGHIRDQITASGQYQGWAPKYLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQ
GARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNI FNPPIIARYIRLHPHTHYSIRSTLRME
LMGCDLNSCMLPGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTVG
TTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM
EVLGCEAQDLY

```

Como se analiza en el Ejemplo 1, las secuencias de nucleótidos de ADNc que codifican las variantes de fVIII ET3 y HSQ se mejoraron implementando una preferencia de uso de codones específicos para la célula hepática humana en

comparación con la secuencia de nucleótidos de origen natural que codifica la secuencia sin codones optimizados correspondiente para un ser humano. También se realizaron cambios adicionales para mejorar la eficacia de la traducción, tal como la optimización del contenido en GC, estructura secundaria de ARNm, sitios de PoliA prematuros, motivo de inestabilidad de ARN, energía libre estable de ARNm, sitios chi internos, motivos de inestabilidad de ARN, sitios de unión a ribosomas, sitios de corte y empalme crípticos, islas CpG negativas, secuencia SD, cajas TATA y señales terminales crípticas.

Además, se eliminaron los motivos de ADN de CpG porque pueden conducir a la metilación y silenciamiento de genes. Véase Bird, DNA methylation and the frequency of CpG in animal DNA, 1980, Nucleic Acids Res, 8: 1499-1504. Los codones se sustituyeron con la alternativa humana/hepática más utilizada que no dio como resultado la formación de un dinucleótido 5'-CG-3' en la secuencia. La eliminación de CpG también puede reducir cualquier respuesta inmunitaria a un vector, incluido el transgén modificado, potenciando la seguridad y la eficacia del vector. Véase J Clin Invest. 2013, 123(7):2994-3001, titulado "CpG-depleted adeno-associated virus vectors evade immune detection".

ET3 es un fVIII híbrido con el dominio B eliminado (BDD) que contiene dominios humanos y porcinos, es decir, la secuencia (A1 y A3 porcinos, véanse las FIG. 1A y 1B) con un enlazador en el dominio B eliminado. ET3 utiliza una secuencia enlazadora OL derivada de una secuencia porcina de 24 aminoácidos, es decir, la secuencia SFAQNSRPPSASAPKPPVLRHQR (SEQ ID NO: 23) derivada de la porcina. HSQ es una variante de fVIII humano en donde la proteína fVIII humana BDD se sustituye con un enlazador SQ, SFSQNPPVLRHQR (SEQ ID NO: 22) derivado humano de 14 aminoácidos. La secuencia de aminoácidos de HSQ se proporciona como la SEQ ID NO: 3. Tanto HSQ como ET3 contienen la secuencia de reconocimiento RHQR (SEQ ID NO: 24) para la secuencia de procesamiento PACE/furina para el dominio B.

Como se analiza en el Ejemplo 1, la secuencia de nucleótidos que codifica ET3 tenía codones optimizados para la expresión en hígado humano. Se proporciona una secuencia de ET3 con codones de hígado optimizados ejemplar como la SEQ ID NO: 12. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende la secuencia de nucleótidos establecida como la SEQ ID NO: 12, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma. Adicionalmente, se eliminaron los motivos CpG dentro de la secuencia de ET3 con codones optimizados, para proporcionar la secuencia de ET3 con codones de hígado optimizados con CpG eliminados establecida como la SEQ ID NO: 11. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende la secuencia de nucleótidos establecida como la SEQ ID NO: 11, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

Como se analiza en el Ejemplo 1, la secuencia de nucleótidos que codifica HSQ tenía codones optimizados para la expresión en hígado humano. Adicionalmente, se eliminaron los motivos CpG dentro de la secuencia de HSQ con codones optimizados, para proporcionar la secuencia de HSQ con codones de hígado optimizados con CpG eliminados establecida como la SEQ ID NO: 2. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende la secuencia de nucleótidos establecida como la SEQ ID NO: 2, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

Adicionalmente, las secuencias de nucleótidos que codifican ET3 y HSQ se optimizaron para la expresión en células mieloides. Se proporciona una secuencia de ET3 con codones mieloides optimizados con CpG eliminados ejemplar como la SEQ ID NO: 125. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende la secuencia de nucleótidos establecida como la SEQ ID NO: 125, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma. Se proporciona una secuencia de HSQ con codones mieloides optimizados con CpG eliminados ejemplar como la SEQ ID NO: 126. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende la secuencia de nucleótidos establecida como la SEQ ID NO: 126, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

En aspectos adicionales de la presente divulgación, se proporcionan variantes de secuencias codificantes de fIX que están diseñadas para altos niveles de expresión cuando el transgén se expresa por el hígado, que es el tejido diana de muchas estrategias de terapia génica dirigidas a fIX. Para crear esta secuencia codificante, se utiliza una estrategia de optimización de codones dirigida al hígado.

Como se analiza en el Ejemplo 1, la secuencia de nucleótidos que codifica para fIX se optimizó implementando una preferencia de uso de codones específicos para la célula hepática humana en comparación con la secuencia de nucleótidos de origen natural que codifica la secuencia correspondiente sin optimización de condones para un ser humano. También se realizaron cambios adicionales para mejorar la eficacia de la Traducción, tal como la optimización del contenido en GC, estructura secundaria de ARNm, sitios de PoliA prematuros, motivo de inestabilidad de ARN, energía libre estable de ARNm, sitios chi internos, motivos de inestabilidad de ARN, sitios de unión a ribosomas, sitios de corte y empalme crípticos, islas CpG negativas, secuencia SD, cajas TATA y señales terminales crípticas.

Además de ajustar preferencia de uso de codones, las secuencias resultantes se modifican adicionalmente para eliminar los motivos CpG que pueden inhibir la expresión eficaz del transgén. Adicionalmente, en algunos aspectos, la molécula de ácido nucleico fIX recombinante puede codificar fIX con la mutación K5A (Darrel Stafford collagen binding mutation, Gui et al. Blood. 2002, 100(1): 153-8). En determinados aspectos, la molécula de ácido nucleico de

fIX recombinante puede codificar fIX con la mutación R338L (mutación de Padua), que es una ganancia de origen natural de mutación de la función que se ha demostrado que mejora la actividad específica de fIX en 8 veces. Las variantes de secuencia se crearon adicionalmente para reflejar dos polimorfismos principales de fIX en el resto 148, que incluyen alanina o treonina. En algunos aspectos, estas secuencias de fIX pueden injertarse en AAV dirigidos al hígado como diseños transgénicos monocatenarios o bicatenarios autocomplementarios.

En el Ejemplo 1, se analizan secuencias del ácido nucleico recombinante ejemplares que codifican proteínas fVIII o fIX, o variantes de las mismas, que se modifican para la expresión específica de tejido.

10 La SEQ ID NO: 12 proporciona una secuencia de ET3 con codones de hígado optimizados ejemplar. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 12, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

15 La SEQ ID NO: 11 proporciona una secuencia de ET3 con codones de hígado optimizados con CpG eliminados ejemplar. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 11, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

20 La SEQ ID NO: 2 proporciona una secuencia de HSQ con codones de hígado optimizados con CpG eliminados ejemplar. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 2, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

25 La SEQ ID NO: 125 proporciona una secuencia de ET3 con codones mieloides optimizados con CpG eliminados ejemplar. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 125, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

30 La SEQ ID NO: 126 proporciona una secuencia de HSQ con codones mieloides optimizados con CpG eliminados ejemplar. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 125, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

35 En el Ejemplo 1, se analizan secuencias del ácido nucleico recombinante ejemplares que codifican proteínas fIX, o variantes de las mismas, que se modifican para la expresión específica de tejido.

La SEQ ID NO: 124 proporciona una secuencia de fIX con codones de hígado optimizados ejemplar con mutaciones Padua/Malmö y sin CpG. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 124, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

La SEQ ID NO: 8 proporciona una secuencia de fIX con codones de hígado optimizados ejemplar sin CpG y que codifica modificaciones A582. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 8, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

La SEQ ID NO: 9 proporciona una secuencia de fIX con codones de hígado optimizados ejemplar sin CpG y que incluye modificaciones Padua y A582. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 9, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

La SEQ ID NO: 10 proporciona una secuencia de fIX con codones de hígado optimizados ejemplar con mutaciones Padua/Malmö y sin CpG. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 10, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

La SEQ ID NO: 127 proporciona una secuencia de fIX con codones humanos optimizados ejemplar con mutaciones Padua/Malmö y sin CpG. En algunos aspectos de la divulgación, se proporciona una molécula de ácido nucleico recombinante que comprende o que consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 127, o una secuencia al menos un 90 % (tal como al menos un 95 %) idéntica a la misma.

Cualquiera de las moléculas de ácidos nucleicos recombinantes analizadas anteriormente que codifican una proteína fVIII o fIX, o una variante de las mismas, puede incluirse en un vector (tal como un vector AAV) como se describe en el presente documento para realizaciones donde es de interés la expresión de una proteína fVIII o fIX o variante de las mismas.

- En algunos aspectos de la divulgación, se proporciona una proteína aislada que comprende una secuencia de aminoácidos codificada por una de las SEQ ID NO: 8, 9 o 10, tal como las secuencias de aminoácidos establecidas como las SEQ ID NO: 17-18 a continuación. En algunos aspectos de la divulgación, se proporciona una proteína aislada que comprende una secuencia de aminoácidos establecida como la SEQ ID NO: 17, o una secuencia de aminoácidos al menos un 90 % (tal como al menos un 95 %) idéntica a la misma que tiene actividad fIX. En algunos aspectos de la divulgación, se proporciona una proteína aislada que comprende una secuencia de aminoácidos establecida como la SEQ ID NO: 18, o una secuencia de aminoácidos al menos un 90 % (tal como al menos un 95 %) idéntica a la misma que tiene actividad fIX. En algunos aspectos de la divulgación, se proporciona una proteína aislada que comprende una secuencia de aminoácidos establecida como la SEQ ID NO: 19, o una secuencia de aminoácidos al menos un 90 % (tal como al menos un 95 %) idéntica a la misma que tiene actividad fIX.

La SEQ ID NO: 8 codifica la secuencia de aminoácidos establecida como la SEQ ID NO: 17:

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEEKCSFEEAREVF
ENTERTEFWKQYVDGDQCESNPCLNGGCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCCKNSADNKVVCS
CTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLKTRAEAVFPDQVYVNSTEAEITLDNITQSTQSFNDFTRVVGEDAK
PGQFPWQVVLNGKVDACGGSIVNEKWIIVTAAHCVEITGVKITVAGEHNIEETEHEQKRNIRIIPHHNYNAAINKY
NHDIALLELDEPLVLSYVTPICIAKEYTNIIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLRSTKFTI
YNNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKETKLT

La SEQ ID NO: 9 codifica la secuencia de aminoácidos establecida como la SEQ ID NO: 18:

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEEKCSFEEAREVF
ENTERTEFWKQYVDGDQCESNPCLNGGCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCCKNSADNKVVCS
CTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLKTRAEAVFPDQVYVNSTEAEITLDNITQSTQSFNDFTRVVGEDAK
PGQFPWQVVLNGKVDACGGSIVNEKWIIVTAAHCVEITGVKITVAGEHNIEETEHEQKRNIRIIPHHNYNAAINKY
NHDIALLELDEPLVLSYVTPICIAKEYTNIIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLRSTKFTI
YNNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKETKLT

La SEQ ID NO: 10 codifica la secuencia de aminoácidos establecida como la SEQ ID NO: 19:

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQGNLERECMEEEKCSFEEAREVF
ENTERTEFWKQYVDGDQCESNPCLNGGCKDDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCCKNSADNKVVCS
CTEGYRLAENQKSCEPAVPFPCGRVSVSQTSLKTRAEAVFPDQVYVNSTEAEITLDNITQSTQSFNDFTRVVGEDAK
PGQFPWQVVLNGKVDACGGSIVNEKWIIVTAAHCVEITGVKITVAGEHNIEETEHEQKRNIRIIPHHNYNAAINKY
NHDIALLELDEPLVLSYVTPICIAKEYTNIIFLKFGSGYVSGWGRVFKGRSALVLQYLRVPLVDRATCLRSTKFTI
YNNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKVSRYVNWIKETKLT

- Los ácidos nucleicos ejemplares pueden prepararse mediante técnicas de clonación, o pueden generarse sintéticamente. Se conocen ejemplos de técnicas apropiadas de clonación y secuenciación, e instrucciones suficientes para dirigir a los expertos a través de muchos ejercicios de clonación (véase, p. ej., Sambrook et al. (Molecular Cloning: A Laboratory Manual, 4ª Ed, Cold Spring Harbor, Nueva York, 2012) y Ausubel et al. (In Current Protocols in Molecular Biology, John Wiley and Sons, Nueva York, a través del suplemento 104, 2013). La información del producto de los fabricantes de reactivos biológicos y equipos experimentales también proporciona información útil. Dichos fabricantes incluyen SIGMA Chemical Company (Saint Louis, MO), R&D Systems (Minneapolis, MN), Pharmacia Amersham (Piscataway, NJ), CLONTECH Laboratories, Inc. (Palo Alto, CA), Chem Genes Corp., Aldrich Chemical Company (Milwaukee, WI), Glen Research, Inc., GIBCO BRL Life Technologies, Inc. (Gaithersburg, MD), Fluka Chemical-Biochemika Analytika (Fluka Chemie AG, Buchs, Switzerland), Invitrogen (Carlsbad, CA), y Applied Biosystems (Foster City, CA), así como muchas otras fuentes comerciales conocidas por un experto.

- Los ácidos nucleicos también se pueden preparar por métodos de amplificación. Los métodos de amplificación incluyen la reacción en cadena de la polimerasa (PCR), la reacción en cadena de la ligasa (LCR), el sistema de amplificación basado en la transcripción (TAS), el sistema de replicación de secuencias autosostenida (3SR). Los expertos en la materia conocen bien una amplia variedad de métodos de clonación, células hospedadoras y metodologías de amplificación *in vitro*.

IV. Vectores Recombinantes y Aplicaciones de Terapia Génica

- Las secuencias de los ácidos nucleicos y promotores divulgadas en el presente documento son útiles en la producción de vectores (tales como vectores rAAV), y también son útiles como vectores de suministro antisentido, vectores de terapia génica o vectores de vacuna. En determinadas realizaciones, la invención proporciona vectores de suministro de genes que contienen las secuencias de los ácidos nucleicos de la invención. En algunas realizaciones, el vector seleccionado puede suministrarse a un sujeto por cualquier método adecuado, incluyendo inyección intravenosa,

transducción ex vivo, transfección, electroporación, suministro en liposomas, técnicas de fusión de membranas, microgránulos recubiertos con ADN de alta velocidad, infección vírica o fusión de protoplastos, para introducir un transgén en el sujeto.

- 5 En determinados aspectos, la divulgación se refiere a partículas de virus, por ejemplo, cápsides, que contienen las secuencias de los ácidos nucleicos que codifican los promotores y las proteínas divulgados en el presente documento. Las partículas víricas, las cápsides y los vectores recombinantes son útiles en el suministro de un gen heterólogo u otras secuencias de los ácidos nucleicos a una célula diana. Los ácidos nucleicos se pueden utilizar fácilmente en una variedad de sistemas de vectores, cápsides y células hospedadoras. En determinadas realizaciones, los ácidos nucleicos están en vectores contenidos dentro de una cápside que comprende proteínas de protección terminal, incluidas las proteínas de la cápside de AAV vp1, vp2, vp3 y las regiones hipervariables.

- 10 En determinadas realizaciones, los ácidos nucleicos de la invención pueden ser parte de cualquier elemento genético (vector) que se pueda suministrar a una célula hospedadora, por ejemplo, ADN desnudo, un plásmido, fago, transposón, cósmido, episoma, una proteína en un vehículo de suministro no vírico (por ejemplo, un transportador basado en lípidos), virus, etc. que transfieren las secuencias portadas sobre los mismos.

- 15 En determinadas realizaciones, un vector puede ser un vector basado en lentivirus (que contiene genes o secuencias lentivíricas), por ejemplo, que tiene secuencias de los ácidos nucleicos derivadas de pseudotipos VSVG o GP64 o de ambos. En determinadas realizaciones, las secuencias de los ácidos nucleicos derivadas de los pseudotipos VSVG o GP64 pueden ser al menos uno o dos o más genes o fragmentos de genes de más de 1000, 500, 400, 300, 200, 100, 50 o 25 nucleótidos continuos o secuencias de nucleótidos con más de 50, 60, 70, 80, 90, 95 o 99 % de identidad con el gen o fragmento.

- 20 La invención también proporciona un vector o composición de la invención para el uso de un método para inducir la coagulación de la sangre en un sujeto que lo necesite. El método comprende administrar al sujeto una cantidad terapéuticamente eficaz del vector (tal como un vector AAV) que codifica un factor de coagulación como se describe en el presente documento. En algunas realizaciones, el sujeto es un sujeto con un trastorno de la coagulación, tal como hemofilia A o hemofilia B. En algunas realizaciones, el trastorno de la coagulación es hemofilia A y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una proteína con actividad de fVIII. En otras realizaciones, el trastorno de la coagulación es hemofilia B y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una proteína con actividad de fIX.

- 25 En algunas realizaciones, las secuencias de los ácidos nucleicos y promotores de la invención son útiles en la producción de vectores AAV. El AAV pertenece a la familia *Parvoviridae* y al género *Dependovirus*. El AAV es un virus pequeño sin envoltura que empaqueta un genoma de ADN lineal de cadena sencilla. Las cadenas de ADN de AAV tanto en sentido como antisentido se empaquetan en cápsides de AAV con la misma frecuencia.

- 30 El genoma de AAV se caracteriza por dos repeticiones terminales invertidas (ITR) que flanquean dos marcos de lectura abiertos (ORF). En el genoma de AAV2, por ejemplo, los primeros 125 nucleótidos de la ITR son un palíndromo, que se pliega sobre sí mismo para maximizar el emparejamiento de bases y forma una estructura de horquilla en forma de T. Las otras 20 bases de la ITR, llamadas secuencia D, permanecen sin emparejar. Las ITR son secuencias de acción en *cis* importantes para la replicación del ADN del AAV; la ITR es el origen de la replicación y sirve como cebador para la síntesis de la segunda cadena por la ADN polimerasa. El ADN bicatenario formado durante esta síntesis, que se llama monómero de forma replicante, se usa para una segunda ronda de replicación autocebadora y forma un dímero de forma replicante. Estos intermedios de doble cadena se procesan mediante un mecanismo de desplazamiento de cadena, lo que da como resultado un ADN de cadena sencilla que se usa para el empaquetamiento y un ADN de bicatenario que se usa para la transcripción. Ubicados dentro de la ITR se encuentran los elementos de unión a Rep y un sitio de resolución terminal (TRS). Estas características son utilizadas por la proteína reguladora vírica Rep durante la replicación de AAV para procesar los intermedios bicatenarios. Además de su papel en la replicación del AAV, la ITR también es esencial para el empaquetamiento del genoma del AAV, transcripción, regulación negativa en condiciones no permisivas y la integración específica de sitio (Daya y Berns, Clin Microbiol Rev 21 (4): 583-593, 2008).

- 35 El ORF izquierdo del AAV contiene el gen Rep, que codifica cuatro proteínas: Rep78, Rep 68, Rep52 y Rep40. El ORF derecho contiene el gen Cap, que produce tres proteínas de la cápside vírica (VP1, VP2 y VP3). La cápside del AAV contiene 60 proteínas de la cápside vírica dispuestas en una simetría icosaédrica. VP1, VP2 y VP3 están presentes en una relación molar 1:1:10 (Daya y Berns, Clin Microbiol Rev 21 (4): 583-593, 2008).

- 40 Los vectores AAV generalmente contienen un casete de expresión transgénica entre las ITR que reemplaza los genes rep y cap. Las partículas del vector se producen por la cotransfección de células con un plásmido que contiene el genoma del vector y una construcción de empaquetamiento/auxiliar que expresa las proteínas rep y cap en trans. Durante la infección, los genomas de los vectores AAV ingresan al núcleo celular y pueden persistir en múltiples estados moleculares. Un resultado común es la conversión del genoma de AAV a un episoma circular de doble cadena mediante síntesis de la segunda cadena o apareamiento con la cadena complementaria.

- 45 En el contexto de los vectores AAV, los vectores divulgados tienen generalmente un genoma recombinante que

comprende la siguiente estructura:

(5' ITR de AAV) - (promotor) - (transgén) - (3' ITR de AAV)

Como se analiza anteriormente, estos vectores AAV recombinantes contienen un casete de expresión transgénica entre las ITR que reemplaza los genes rep y cap. Las partículas de vector se producen, por ejemplo, mediante la cotransfección de células con un plásmido que contiene el genoma del vector recombinante y una construcción de empaquetamiento/auxiliar que expresa las proteínas rep y cap en trans. Por ejemplo, en algunos aspectos, el vector AAV recombinante puede tener un genoma con una estructura establecida como una de:

- 10 (5' ITR de AAV) - (HCB) - (transgén) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (fVIII) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (ET3) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (ET3, Seq_12) - (3' ITR de AAV)
- 15 (5' ITR de AAV) - (HCB) - (ET3, Seq_11) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (HSQ) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (HSQ, Seq_2) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (flX) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (flX, Seq_124) - (3' ITR de AAV)
- 20 (5' ITR de AAV) - (HCB) - (flX, Seq_8) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (flX, Seq_9) - (3' ITR de AAV)
(5' ITR de AAV) - (HCB) - (flX, Seq_10) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (transgén) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (fVIII) - (3' ITR de AAV)
- 25 (5' ITR de AAV) - (HSh-HCB) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (ET3) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (ET3, Seq_12) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (ET3, Seq_11) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (HSQ) - (3' ITR de AAV)
- 30 (5' ITR de AAV) - (HSh-HCB) - (HSQ, Seq_2) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (flX) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (flX, Seq_124) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (flX, Seq_8) - (3' ITR de AAV)
(5' ITR de AAV) - (HSh-HCB) - (flX, Seq_9) - (3' ITR de AAV)
- 35 (5' ITR de AAV) - (HSh-HCB) - (flX, Seq_10) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (transgén) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (ET3) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (ET3, Seq_12) - (3' ITR de AAV)
- 40 (5' ITR de AAV) - (5'HSh-HCB) - (ET3, Seq_11) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (HSQ) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (HSQ, Seq_2) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (flX) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (flX, Seq_124) - (3' ITR de AAV)
- 45 (5' ITR de AAV) - (5'HSh-HCB) - (flX, Seq_8) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (flX, Seq_9) - (3' ITR de AAV)
(5' ITR de AAV) - (5'HSh-HCB) - (flX, Seq_10) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (transgén) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (fVIII) - (3' ITR de AAV)
- 50 (5' ITR de AAV) - (3'HSh-HCB) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (ET3) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (ET3, Seq_12) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (ET3, Seq_11) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (HSQ) - (3' ITR de AAV)
- 55 (5' ITR de AAV) - (3'HSh-HCB) - (HSQ, Seq_2) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (flX) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (flX, Seq_124) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (flX, Seq_8) - (3' ITR de AAV)
(5' ITR de AAV) - (3'HSh-HCB) - (flX, Seq_9) - (3' ITR de AAV)
- 60 (5' ITR de AAV) - (3'HSh-HCB) - (flX, Seq_10) - (3' ITR de AAV)
(5' ITR de AAV) - (ABP-HP1-God-TSS) - (transgén) - (3' ITR de AAV)
(5' ITR de AAV) - (ABP-HP1-God-TSS) - (fVIII) - (3' ITR de AAV)
(5' ITR de AAV) - (ABP-HP1-God-TSS) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
(5' ITR de AAV) - (ABP-HP1-God-TSS) - (ET3) - (3' ITR de AAV)
- 65 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (ET3, Seq_12) - (3' ITR de AAV)
(5' ITR de AAV) - (ABP-HP1-God-TSS) - (ET3, Seq_11) - (3' ITR de AAV)

- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (HSQ) - (3' ITR de AAV)
 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (HSQ, Seq_2) - (3' ITR de AAV)
 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (flX) - (3' ITR de AAV)
 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (flX, Seq_124) - (3' ITR de AAV)
 5 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (flX, Seq_8) - (3' ITR de AAV)
 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (flX, Seq_9) - (3' ITR de AAV)
 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (flX, Seq_10) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (transgén) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (fVIII) - (3' ITR de AAV)
 10 (5' ITR de AAV) - (HSh-SynO-TSS) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (ET3) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (ET3, Seq_12) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (ET3, Seq_11) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (HSQ) - (3' ITR de AAV)
 15 (5' ITR de AAV) - (HSh-SynO-TSS) - (HSQ, Seq_2) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (flX) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (flX, Seq_124) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (flX, Seq_8) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-SynO-TSS) - (flX, Seq_9) - (3' ITR de AAV)
 20 (5' ITR de AAV) - (HSh-SynO-TSS) - (flX, Seq_10) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (transgén) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (fVIII) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (fVIII-dominio B eliminado) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (ET3) - (3' ITR de AAV)
 25 (5' ITR de AAV) - (sHS-SynO-TSS) - (ET3, Seq_12) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (ET3, Seq_11) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (HSQ) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (HSQ, Seq_2) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (flX) - (3' ITR de AAV)
 30 (5' ITR de AAV) - (sHS-SynO-TSS) - (flX, Seq_124) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (flX, Seq_8) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (flX, Seq_9) - (3' ITR de AAV)
 (5' ITR de AAV) - (sHS-SynO-TSS) - (flX, Seq_10) - (3' ITR de AAV)
- 35 El transgén puede estar flanqueado por secuencias reguladoras tales como una secuencia Kozak 5' y/o una señal de poliadenilación 3'. Por ejemplo, en algunos casos, el vector AAV recombinante puede tener un genoma con una estructura establecida como una de:
- (5' ITR de AAV) - (HCB) - (Kozak) - (transgén) - (señal de poliA) - (señal de poliA) - (3' ITR de AAV)
 40 (5' ITR de AAV) - (HCB) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
 45 (5' ITR de AAV) - (HCB) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
 50 (5' ITR de AAV) - (HCB) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HCB) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
 55 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
 60 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (HSh-HCB) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
 65 (5' ITR de AAV) - (5'HS-HCB) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
 (5' ITR de AAV) - (5'HS-HCB) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)

- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
- 5 (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
- 10 (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (5'HSsh-HCB) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
- 15 (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
- 20 (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (3'HSsh-HCB) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
- 25 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
- 30 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
- 35 (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (ABP-HP1-God-TSS) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
- 40 (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
- 45 (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
- 50 (5' ITR de AAV) - (HSh-SynO-TSS) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (transgén) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (fVIII) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (fVIII-dominio B eliminado) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (ET3) - (señal de poliA) - (3' ITR de AAV)
- 55 (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (ET3, Seq_12) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (ET3, Seq_11) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (HSQ) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (HSQ, Seq_2) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (flX) - (señal de poliA) - (3' ITR de AAV)
- 60 (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (flX, Seq_124) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (flX, Seq_8) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (flX, Seq_9) - (señal de poliA) - (3' ITR de AAV)
- (5' ITR de AAV) - (sHS-SynO-TSS) - (Kozak) - (flX, Seq_10) - (señal de poliA) - (3' ITR de AAV)

- 65 Las ITR de AAV, y otros componentes de AAV seleccionados descritos en el presente documento, pueden seleccionarse fácilmente entre cualquier serotipo de AAV, incluyendo, sin limitación, AAV1, AAV2, AAV3, AAV4, AAV5,

AAV6, AAV7, AAV8, AAV9 y variantes de función de los mismos. Estas ITR u otros componentes de AAV pueden aislarse fácilmente utilizando técnicas disponibles para los expertos en la materia a partir de un serotipo de AAV. Dicho AAV puede aislarse u obtenerse de fuentes académicas, comerciales o públicas (por ejemplo, la American Type Culture Collection, Manassas, Va.). De manera alternativa, las secuencias de AAV pueden obtenerse a través de medios sintéticos u otros medios adecuados por referencia a secuencias publicadas como las que están disponibles en la bibliografía o en bases de datos tales como, por ejemplo, GenBank, PubMed o similares.

En algunas realizaciones, el vector puede ser un vector AAV recombinante que comprende un genoma que comprende una molécula de ácido nucleico que codifica cualquiera de los promotores específicos de hígado de la invención (tal como el promotor HCB, SEQ ID NO: 4) operativamente unido a una molécula de ácido nucleico heterólogo que codifica una variante de fVIII, en donde la molécula de ácido nucleico heterólogo comprende o consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125, o SEQ ID NO: 126, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125 o SEQ ID NO: 126. En varias de tales realizaciones, el genoma de AAV recombinante (de 5' a 3' ITR) no tiene más de 5,1, 5,0, 4,9, 4,8, 4,7, 4,6 o 4,5 kb de longitud.

En algunas realizaciones, el vector puede ser un vector AAV recombinante que comprende un genoma que comprende una molécula de ácido nucleico que codifica cualquiera de los promotores específicos del hígado de la invención (tal como el promotor HCB, SEQ ID NO: 4) operativamente unido a una molécula de ácido nucleico heterólogo que codifica una variante de fIX, en donde la molécula de ácido nucleico heterólogo comprende o consiste en la secuencia del ácido nucleico establecida como la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124, o SEQ ID NO: 127, o una secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124 o SEQ ID NO: 127. En varias de tales realizaciones, el genoma de AAV recombinante (de 5' a 3' ITR) no tiene más de 5,1, 5,0, 4,9, 4,8, 4,7, 4,6 o 4,5 kb de longitud.

El AAV es actualmente uno de los virus más utilizados para terapia génica. Aunque el AAV infecta a los seres humanos y a algunas otras especies de primates, no se sabe que cause enfermedad y provoca una respuesta inmunitaria muy leve. Los vectores de terapia génica que utilizan AAV pueden infectar tanto las células en división como las quiescentes y persisten en un estado extracromosómico sin integrarse en el genoma de la célula hospedadora. Debido a las características ventajosas del AAV, la presente divulgación contempla el uso del AAV para las moléculas de ácidos nucleicos recombinantes y los métodos divulgados en el presente documento.

El AAV posee varias características deseables para un vector de terapia génica, incluida la capacidad de unirse y entrar en las células diana, ingresar al núcleo, la capacidad de expresarse en el núcleo durante un período prolongado de tiempo y baja toxicidad. Sin embargo, el pequeño tamaño del genoma del AAV limita el tamaño del ADN heterólogo que se puede incorporar. Para minimizar este problema, se han construido vectores AAV que no codifican Rep ni el elemento de eficacia de integración (IEE, por sus siglas en inglés). Las ITR se conservan ya que son señales en *cis* requeridas para el empaquetamiento (Daya y Berns, Clin Microbiol Rev 21 (4): 583-593, 2008).

Se conocen métodos para producir rAAV adecuados para terapia génica (véase, por ejemplo, las Solicitudes de Patente de EE. UU. N.º 2012/0100606; 2012/0135515; 2011/0229971; y 2013/0072548; y Ghosh et al., Gene Ther 13(4):321-329, 2006), y pueden utilizarse con las moléculas de ácidos nucleicos recombinantes y los métodos divulgados en el presente documento.

En algunas realizaciones, los ácidos nucleicos de la invención son parte de un casete de expresión o transgén. Véase por ejemplo, la publicación de solicitud de Patente de Estados Unidos 20150139953. El casete de expresión se compone de un transgén y secuencias reguladoras, por ejemplo, por ejemplo un promotor y repeticiones terminales invertidas (ITR) del AAV 5' y 3'. En una realización deseable, se utilizan las ITR del serotipo 2 u 8 del AAV. Sin embargo, se pueden seleccionar ITR de otros serotipos adecuados. Un casete de expresión se empaqueta generalmente en una proteína de la cápside y se suministra a una célula hospedadora seleccionada.

En algunos aspectos, la divulgación proporciona un método para generar un virus adenoasociado (AAV) recombinante que tiene una cápside de serotipo del AAV, o una porción de la misma. Tal método implica el cultivo de una célula hospedadora que contiene una secuencia del ácido nucleico que codifica una proteína de la cápside del serotipo de virus adenoasociados (AAV); un gen rep funcional; un casete de expresión compuesto por repeticiones terminales invertidas (ITR) de AAV y un transgén; y suficientes funciones auxiliares para permitir el empaquetamiento del casete de expresión en la proteína de la cápside del AAV. Véase por ejemplo, la publicación de solicitud de Patente de Estados Unidos 20150139953.

Los componentes para el cultivo en la célula hospedadora para empaquetar un casete de expresión de AAV en una cápside de AAV pueden proporcionarse a la célula hospedadora en trans. De manera alternativa, uno o más de los componentes (por ejemplo, casete de expresión, secuencias rep, secuencias cap y/o funciones auxiliares) pueden proporcionarse mediante una célula hospedadora estable que se ha modificado por ingeniería para contener uno o más de los componentes requeridos utilizando métodos conocidos por los expertos en la materia. De forma más adecuada, dicha célula hospedadora estable contendrá los componente(s) bajo el control de un promotor inducible. Sin embargo, los componente(s) requeridos pueden estar bajo el control de un promotor constitutivo. En otra alternativa

más, una célula hospedadora estable seleccionada puede contener componente(s) seleccionados bajo el control de un promotor constitutivo y otros componente(s) seleccionados bajo el control de uno o más promotores inducibles. Por ejemplo, se puede generar una célula hospedadora estable que deriva de células 293 (que contienen funciones E1 auxiliares bajo el control de un promotor constitutivo), pero que contiene las proteínas rep y/o cap bajo el control de promotores inducibles. Un experto en la materia puede generar otras células hospedadoras estables adicionales.

En algunas realizaciones, la invención se refiere a vectores recombinantes que comprenden una secuencia del ácido nucleico del promotor específico de hígado de la invención en combinación operativa con el transgén. El transgén es una secuencia de un ácido nucleico, heteróloga a las secuencias del vector que flanquean el transgén, que codifica una proteína u otro producto de interés. La secuencia codificante del ácido nucleico está operativamente unida a componentes reguladores de una manera que permite la transcripción, traducción y/o expresión transgénica en una célula hospedadora.

Un transgén típico es una secuencia que codifica un producto que es útil en biología y medicina, tales como proteínas, péptidos, ARN, enzimas, mutantes negativos dominantes o ARN catalíticos. Las moléculas de ARN deseables incluyen ARNm, ARNt, ARNbc, ARN ribosómico, ARN catalíticos, ARNip, microARN, ARN de horquilla pequeña, ARN de corte y empalme en trans y ARN antisentido. Un ejemplo de una secuencia de ARN útil es una secuencia que inhibe o extingue la expresión de una secuencia del ácido nucleico dirigida en el animal tratado. Generalmente, las secuencias diana adecuadas incluyen dianas oncológicas y enfermedades víricas.

El transgén puede usarse para corregir o mejorar las deficiencias genéticas, que pueden incluir deficiencias en las que los genes normales se expresan a niveles inferiores a los normales o deficiencias en las que el producto génico funcional no se expresa. Un tipo preferido de secuencia transgénica codifica una proteína o polipéptido terapéutico que se expresa en una célula hospedadora. La divulgación además contempla el uso de múltiples transgenes, por ejemplo, para corregir o mejorar un defecto genético causado por una proteína de múltiples subunidades. En determinadas situaciones, se puede usar un transgén diferente para codificar cada subunidad de una proteína, o para codificar diferentes péptidos o proteínas. Esto es deseable cuando el tamaño del ADN que codifica la subunidad de la proteína es grande, por ejemplo, para una inmunoglobulina, el factor de crecimiento derivado de plaquetas o una proteína distrofina. Para que la célula produzca la proteína de múltiples subunidades, se infecta una célula con el virus recombinante que contiene cada una de las diferentes subunidades. De manera alternativa, diferentes subunidades de una proteína pueden estar codificadas por el mismo transgén. En este caso, un solo transgén incluye el ADN que codifica cada una de las subunidades, con el ADN de cada subunidad separado por un sitio interno de entrada al ribosoma (IRES). Esto es deseable cuando el tamaño del ADN que codifica cada una de las subunidades y las regiones de control reguladoras en cis tal como un promotor, intrón, señal de poliA, es pequeño, por ejemplo, el tamaño total del ADN que codifica las subunidades y los IRES y las regiones de control regulador en cis es menor de cinco kilobases. Como alternativa a un IRES, el ADN puede estar separado por secuencias que codifican un péptido 2A, que se autoescinde en un evento postraduccional. Véanse, por ejemplo, M. L. Donnelly, et al., J. Gen. Virol., 78(Pt 1):13-21 (Enero de 1997); Furler, S., et al, Gene Ther., 8(11):864-873 (Junio de 2001); Klump H., et al., Gene Ther., 8(10):811-817 (Mayo de 2001). En determinadas realizaciones, los rAAV que portan los transgenes o subunidades deseados se administran conjuntamente para permitirles concatamerizar *in vivo* para formar un genoma de vector único. En dicha una realización, un primer AAV puede portar un casete de expresión que expresa un solo transgén y un segundo AAV puede portar un casete de expresión que expresa un transgén diferente para la coexpresión en la célula hospedadora. Sin embargo, el transgén seleccionado puede codificar cualquier producto biológicamente activo u otro producto, por ejemplo, un producto deseable para estudio.

El casete de expresión se puede portar en cualquier vector adecuado, por ejemplo, un plásmido, que se suministra a una célula hospedadora. Los plásmidos útiles en la presente divulgación pueden modificarse por ingeniería de modo que sean adecuados para la replicación y, opcionalmente, la integración en células procariotas, células de mamífero o ambas. Estos plásmidos (u otros vectores que portan la 5' ITR de AAV-molécula heteróloga-3' ITR) contienen secuencias que permiten la replicación del casete de expresión en eucariotas y/o procariotas y marcadores de selección para estos sistemas. Preferentemente, La molécula que porta el casete de expresión se transfecta en la célula, donde puede existir de forma transitoria. De manera alternativa, el casete de expresión (que porta la 5' ITR de AAV-molécula heteróloga-3' ITR) puede integrarse de manera estable en el genoma de la célula hospedadora, ya sea cromosómicamente o como un episoma. En determinadas realizaciones, el casete de expresión puede estar presente en múltiples copias, opcionalmente en concatámeros cabeza a cabeza, cabeza a cola o cola a cola. Se conocen técnicas de transfección adecuadas y se pueden utilizar fácilmente para suministrar el casete de expresión a la célula hospedadora.

Por lo general, cuando se suministra el vector que comprende el casete de expresión por transfección, pueden ajustarse el vector y las cantidades relativas de ADN del vector a las células hospedadoras, teniendo en cuenta factores tales como el vector seleccionado, el método de suministro y las células hospedadoras seleccionadas. Además del casete de expresión, la célula hospedadora contiene las secuencias que impulsan la expresión de la proteína de la cápside del AAV en la célula hospedadora y las secuencias rep del mismo serotipo que el serotipo de las ITR del AAV encontrados en el casete de expresión, o de un serotipo de complemento cruzado. Aunque las moléculas que proporcionan rep y cap pueden existir en la célula hospedadora de forma transitoria (es decir, a través de la transfección), se prefiere que una o ambas proteínas rep y cap y el (los) promotor(es) que controlan su expresión

se expresen de manera estable en la célula hospedadora, por ejemplo, como un episoma o por integración en el cromosoma de la célula hospedadora.

La célula hospedadora de empaquetamiento generalmente también contiene funciones auxiliares para empaquetar el rAAV de la divulgación. Opcionalmente, estas funciones pueden suministrarse por un herpesvirus. De manera más deseable, las funciones auxiliares necesarias se proporcionan cada una de una fuente de adenovirus de primates humanos o no humanos, tal como las descritas anteriormente y/o están disponibles en una variedad de fuentes, incluida la American Type Culture Collection (ATCC), Manassas, Va. (EE.UU). Las funciones auxiliares deseadas, se pueden proporcionar utilizando cualquier medio que permita su expresión en una célula.

La introducción en la célula hospedadora del vector se puede lograr por cualquier medio conocido en la técnica o como se divulga anteriormente, incluyendo transfección, infección, electroporación, suministro en liposomas, técnicas de fusión de membranas, microgránulos recubiertos con ADN de alta velocidad, infección vírica o fusión de protoplastos, entre otros. Uno o más de los genes adenovíricos pueden integrarse de manera estable en el genoma de la célula hospedadora, expresarse de manera estable como episomas o expresarse de manera transitoria. Todos los productos génicos pueden expresarse de manera transitoria, en un episoma o integrarse de manera estable, o algunos de los productos génicos pueden expresarse de manera estable mientras que otros se expresan de manera transitoria. Además, los promotores para cada uno de los genes adenovíricos pueden seleccionarse independientemente de un promotor constitutivo, un promotor inducible o un promotor adenovírico natural. Los promotores pueden estar regulados por un estado fisiológico específico del organismo o la célula (es decir, por el estado de diferenciación o en células en replicación o quiescentes) o por factores añadidos exógenamente, por ejemplo.

La introducción de las moléculas (como plásmidos o virus) en la célula hospedadora se puede lograr utilizando técnicas conocidas por el experto en la materia. En una realización preferida, se usan técnicas de transfección convencionales, por ejemplo, transfección o electroporación con CaPO_4 , y/o infección por vectores híbridos de adenovirus/AAV en líneas celulares tales como la línea celular de riñón embrionario humano HEK 293 (una línea celular de riñón humano que contiene genes E1 de adenovirus funcionales que proporcionan proteínas E1 de acción en trans).

Un experto en la materia comprenderá fácilmente que las técnicas de AAV pueden adaptarse para su uso en estos y otros sistemas de vectores víricos para el suministro de genes *in vitro*, *ex vivo* o *in vivo*. En determinadas realizaciones, la divulgación contempla el uso de ácidos nucleicos y vectores divulgados en el presente documento en una variedad de sistemas de vectores rAAV y no rAAV. Tales sistemas de vectores pueden incluir, por ejemplo, lentivirus, retrovirus, poxvirus, virus vaccinia y sistemas adenovíricos, entre otros.

Se contempla que las partículas víricas, los ácidos nucleicos y los vectores divulgados en el presente documento sean útiles para una variedad de propósitos, incluido el suministro de moléculas terapéuticas para la expresión génica de proteínas terapéuticas.

Las proteínas terapéuticas codificadas por los ácidos nucleicos (p. ej., operativamente en combinación con los promotores) informadas en el presente documento incluyen las utilizadas para el tratamiento de la hemofilia, incluida la hemofilia B (incluyendo el fIX) y la hemofilia A (incluyendo el fVIII y sus variantes, tal como la cadena ligera y cadena pesada del heterodímero y con el dominio B eliminado; Patente de Estados Unidos N.º 6.200.560 y Patente de Estados Unidos N.º 6.221.349). El gen del Factor VIII codifica 2351 aminoácidos y la proteína tiene seis dominios, designados desde el extremo amino al carboxilo terminal como A1-A2-B-A3-C1-C2 [Wood et al, Nature, 312:330 (1984); Vehar et al., Nature 312:337 (1984); y Toole et al, Nature, 342:337 (1984)]. El fVIII humano se procesa dentro de la célula para producir un heterodímero que comprende principalmente una cadena pesada que contiene los dominios A1, A2 y B y una cadena ligera que contiene los dominios A3, C1 y C2. Tanto el polipéptido de cadena única como el heterodímero circulan en el plasma como precursores inactivos, hasta que se activan mediante la escisión por trombina entre los dominios A2 y B, que libera el dominio B y da como resultado una cadena pesada que consiste en los dominios A1 y A2. El dominio B se elimina en la forma procoagulante activada de la proteína. Adicionalmente, en la proteína natural, dos cadenas polipeptídicas ("a" y "b"), que flanquean el dominio B, están unidas a un catión de calcio divalente.

Una opción de tratamiento para un paciente diagnosticado con hemofilia A es la administración exógena de fVIII recombinante, a veces denominada terapia de reemplazo de fVIII. En algunos pacientes, esta terapia puede conducir al desarrollo de anticuerpos que se unen a la proteína fVIII administrada. Posteriormente, los conjugados unidos fVIII-anticuerpo, generalmente conocidos como inhibidores, interfieren o retrasan la capacidad de fVIII para causar la coagulación sanguínea. Los autoanticuerpos inhibitorios también se producen a veces espontáneamente en un sujeto que no está genéticamente en riesgo de tener hemofilia, denominada hemofilia adquirida. Los ensayos de anticuerpos inhibitorios se realizan generalmente antes del tratamiento exógeno con fVIII para determinar si la terapia anticoagulante será eficaz.

Históricamente, se ha utilizado un "ensayo Bethesda" para cuantificar la dosificación inhibitoria de la concentración de anticuerpos de unión a fVIII. En el ensayo, se preparan diluciones en serie de plasma de un paciente, por ejemplo, antes de someterse a cirugía, y cada dilución se mezcla con un volumen igual de plasma normal como fuente de fVIII. Después de incubar durante un par de horas, se miden las actividades de fVIII en cada una de las mezclas diluidas. Tener concentraciones de inhibidor de anticuerpos que evitan la actividad de coagulación de fVIII después de múltiples

diluciones repetidas indica un mayor riesgo de sangrado incontrolado. Se cree que es poco probable que los pacientes con títulos de inhibidores después de aproximadamente diez diluciones respondan a infusiones exógenas de fVIII para detener el sangrado. Un título Bethesda se define como el valor recíproco de la dilución que da como resultado una inhibición del 50 % de la actividad de FVIII presente en el plasma humano normal. Un título Bethesda superior a 10 se considera el umbral de respuesta a la terapia de reemplazo de FVIII.

En determinadas realizaciones, la invención se refiere a una partícula vírica o cápside que comprende un vector de la invención que comprende un ácido nucleico que codifica un factor de coagulación sanguínea como se divulga en el presente documento para usar en métodos para inducir la coagulación sanguínea, comprendiendo dicho método administrar una cantidad eficaz a un sujeto que lo necesite.

En determinadas realizaciones, el sujeto es diagnosticado con hemofilia A o B o hemofilia adquirida o es poco probable que responda a infusiones exógenas de fVIII.

En algunas realizaciones, la invención se refiere a un vector vírico adenoasociado (AAV) que codifica fIX humano como el vehículo de suministro génico para su uso en métodos de transferencia génica para el tratamiento de la hemofilia B. Si bien varias de estas terapias génicas basadas en AAV para la hemofilia B se han incorporado en ensayos clínicos en seres humanos, se han visto obstaculizadas por la baja expresión de la proteína terapéutica, fIX de coagulación, después de la administración del virus, lo que da como resultado solamente una corrección parcial de la enfermedad. La toxicidad del vector AAV limita la dosis del virus que puede administrarse de manera segura. Para una transición exitosa a una terapia clínicamente viable, un vector debe proporcionar una expresión eficaz de fIX a dosis víricas por debajo del umbral de toxicidad.

El tratamiento de pacientes con inhibidores de fVIII también se ha logrado mediante métodos de inducción de tolerancia inmunitaria (ITI, por sus siglas en inglés) que generalmente implican la infusión diaria de fVIII hasta que disminuyen los niveles de inhibidor/anticuerpo circulante. Sin embargo, un 20-30 % de los pacientes no se vuelven tolerantes después de una terapia de inducción de tolerancia inmunitaria (ITI). La persistencia de los inhibidores de fVIII se asocia con riesgos de morbilidad y mortalidad aumentados. En determinados aspectos, la divulgación se refiere a métodos de inducción de tolerancia inmunitaria que comprenden administrar una cantidad eficaz de un vector o una cápside como se divulga en el presente documento a un sujeto que lo necesite.

En algunas realizaciones, las proteínas terapéuticas codificadas por los ácidos nucleicos (p. ej., operativamente en combinación con promotores) indicadas en el presente documento comprenden los primeros 57 pares de bases de la cadena pesada de fVIII que codifica la secuencia señal de 10 aminoácidos, así como la secuencia de poliadenilación de beta globina humana o la secuencia de poliadenilación de la hormona de crecimiento (hGH). En realizaciones alternativas, los dominios A1 y A2, así como 5 aminoácidos del extremo N del dominio B y/u 85 aminoácidos del extremo C del dominio B, así como los dominios A3, C1 y C2. En otras realizaciones más, los ácidos nucleicos que codifican la cadena pesada y la cadena ligera de fVIII se proporcionan en un único ácido nucleico separado por 42 ácidos nucleicos que codifican 14 aminoácidos del dominio B. Véase la patente de EE.UU. n.º 6.200.560.

Como se usan en el presente documento, una cantidad terapéuticamente eficaz es una cantidad de vector AAV que produce cantidades suficientes de fVIII para disminuir el tiempo que tarda la sangre de un sujeto en coagularse. Por lo general, los hemofílicos graves que tienen menos del 1 % de los niveles normales de fVIII tienen un tiempo de coagulación de la sangre total de más de 60 minutos en comparación con aproximadamente 10 minutos para los no hemofílicos.

La presente divulgación no se limita a ninguna secuencia específica de fVIII o fIX u otra proteína indicada en el presente documento. Se han aislado y generado muchas formas naturales y recombinantes de fVIII. Se pueden encontrar ejemplos de formas de origen natural y recombinantes de fVII en la bibliografía científica y de patentes que incluye, patente de los EE.UU. n.º 5.563.045, patente de los EE.UU. n.º 5.451.521, patente de los EE.UU. n.º 5.422.260, patente de los EE.UU. n.º 5.004.803, patente de los EE.UU. n.º 4.757.006, patente de los EE.UU. n.º 5.661.008, patente de los EE.UU. n.º 5.789.203, patente de los EE.UU. n.º 5.681.746, patente de los EE.UU. n.º 5.595.886, patente de los EE.UU. n.º 5.045.455, patente de los EE.UU. n.º 5.668.108, patente de los EE.UU. n.º 5.633.150, patente de los EE.UU. n.º 5.693.499, patente de los EE.UU. n.º 5.587.310, patente de los EE.UU. n.º 5.171.844, patente de los EE.UU. n.º 5.149.637, patente de los EE.UU. n.º 5.112.950, patente de los EE.UU. n.º 4.886.876, el documento WO 94/11503, el documento WO 87/07144, el documento WO 92/16557, el documento WO 91/09122, el documento WO 97/03195, el documento WO 96/21035, el documento WO 91/07490, el documento EP 0 672 138, el documento EP 0 270 618, el documento EP 0 182 448, el documento EP 0 162 067, el documento EP 0 786 474, el documento EP 0 533 862, el documento EP 0 506 757, el documento EP 0 874 057, el documento EP 0 795 021, el documento EP 0 670 332, el documento EP 0 500 734, el documento EP 0 232 112, el documento EP 0 160 457, Sanberg et al., Int. Congress of the World Fed. Of Hemophilia (1992), y Lind et al., Eur. J. Biochem., 232:19 (1995).

Además, la divulgación no se limita al fVIII humano. De hecho, se pretende que la presente divulgación abarque fVIII de animales que no sean humanos, incluidos, entre otros, animales de compañía (por ejemplo, canino, felinos y equinos), ganado (por ejemplo, bovinos, caprinos y ovinos), animales de laboratorio, mamíferos marinos, grandes felinos, etc.

- Los vectores AAV pueden contener un ácido nucleico que codifica fragmentos de fVIII que en sí mismo no es biológicamente activo, sin embargo, cuando se administra al sujeto mejora o restaura el tiempo de coagulación de la sangre. Por ejemplo, la proteína fVIII comprende dos cadenas polipeptídicas: una cadena pesada y una cadena ligera separadas por un dominio B que se escinde durante el procesamiento. La transducción conjunta de células receptoras con las cadenas pesada y ligera de FVIII conducen a la expresión de fVIII biológicamente activo. La administración de solamente la cadena defectuosa se contempla en pacientes porque la mayoría de los hemofílicos contienen una mutación o delección en una sola de las cadenas (p. ej., cadena pesada o ligera).
- Por lo tanto, en determinadas realizaciones, la divulgación se refiere a vectores de la invención que tienen ácidos nucleicos que codifican una cadena ligera que contiene los dominios A3, C1 y C2 o una cadena pesada que consiste en los dominios A1 y A2.

Descripción adicional de vectores recombinantes y modalidades terapéuticas

- Los vectores recombinantes divulgados en el presente documento (por ejemplo, un vector AAV recombinante) se pueden usar en varias aplicaciones terapéuticas diferentes, dependiendo de la proteína de interés codificada por el vector recombinante.
- En determinados aspectos, los usos son para el tratamiento de la hemocromatosis hereditaria (HH), un trastorno importante de sobrecarga de hierro, enfermedad de Wilson, un trastorno genético de sobrecarga de cobre y deficiencia de alfa1-antitripsina (α 1-AT). En determinados aspectos, la proteína es alfa1-antitripsina humana (α 1-AT, Registro: P01009.3), proteína HFE (Registro NP_000401.1 o Q30201) o proteína hepática ATP7B (registro P35670.4) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de hipercolesterolemia utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para fenilalanina hidroxilasa humana (Registro: P00439.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de tirosinemia tipo 1 usando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para fumarilacetoacetato hidrolasa de humana (Registro: P16930.2) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de tirosinemia Tipo 2 utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para tirosina aminotransferasa humana (Registro: P17735.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de homocistinuria utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para metilentetrahidrofolato reductasa humana (Registro: P42898.3) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de hiperlipidemia e hipercolesterolemia utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para acil-CoA deshidrogenasa de cadena media humana (Registro: P11310.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de Galactosemia utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para galactosa-1-fosfato uridiltransferasa humana (Registro: P07902.3) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento del síndrome de Lesch-Nyhan utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para hipoxantina fosforibosiltransferasa humana (Registro: P00492.2) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de la enfermedad de Gaucher utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para cerebrosidasa humana (Registro: P07602.2, Registro: P04062.3) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de la enfermedad de Tay-Sachs utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para beta-hexosaminidasa humana (Registro: P06865.2) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.
- En determinados aspectos, el uso es para el tratamiento de la enfermedad de Fabry utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para α -galactosidasa humana

(Registro: P06280.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

En determinados aspectos, el uso es para el tratamiento del síndrome de Hunter utilizando un promotor del presente documento en combinación operativa con un ácido nucleico que codifica para iduronato sulfatasa humana (Registro: P22304.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

En determinados aspectos, el uso es para el tratamiento de la enfermedad por almacenamiento de glucógeno la utilizando un promotor en combinación operativa con un ácido nucleico que codifica para glucosa-6-fosfatasa humana (Registro: P35575.2) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

En determinados aspectos, el uso es para el tratamiento del metabolismo del amoniaco utilizando un promotor en combinación operativa con un ácido nucleico que codifica para ornitina transcarbamilasa humana (Registro: P00480.3) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

En determinados aspectos, el uso es para el tratamiento de fenilcetonuria utilizando un promotor en combinación operativa con un ácido nucleico que codifica para el receptor de lipoproteínas de baja densidad humano (Registro: P01130.1) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

En determinados aspectos, el uso es para el tratamiento de acidemia propiónica utilizando un promotor en combinación operativa con un ácido nucleico que codifica para propionil-coenzima A carboxilasa, ya sea PCCA y/o PCCB (Registro: P05166.3 beta, NP_000273.2 alfa, NP_001121164.1 alfa) o variantes con más de 50, 60, 70, 80, 90, 95 o 95 de identidad o similitud de secuencia.

También se contemplan usos en regímenes de vacuna, por ejemplo, para el suministro conjunto de una citocina o para el suministro de un inmunógeno o antígeno.

Las partículas víricas recombinantes, las cápsides o los vectores que comprenden los ácidos nucleicos divulgados en el presente documento pueden suministrarse al hígado a través de la arteria hepática, la vena porta o intravenosamente para producir niveles terapéuticos de proteínas terapéuticas o de factores de coagulación en la sangre. La cápside o vector se suspende preferentemente en un transportador fisiológicamente compatible, pudiendo administrarse a un paciente mamífero humano o no humano. Un experto en la materia puede seleccionar fácilmente transportadores adecuados a la vista de la indicación para la cual se dirige el virus de transferencia. Por ejemplo, un transportador adecuado incluye una solución salina, que puede formularse con una variedad de soluciones tamponantes (por ejemplo, solución salina tamponada con fosfato). Otros transportadores ilustrativos incluyen solución salina estéril, lactosa, sacarosa, fosfato de calcio, gelatina, dextrano, agar, pectina, aceite de sésamo y agua.

Opcionalmente, las composiciones de la divulgación pueden contener otros excipientes farmacéuticamente aceptables, tales como conservantes, o estabilizantes químicos. Conservantes ejemplares adecuados incluyen clorobutanol, sorbato de potasio, ácido sórbico, dióxido de azufre, galato de propilo, los parabenos, etil vainillina, glicerina, fenol y paraclorofenol. Estabilizantes químicos adecuados incluyen gelatina y albúmina.

Las partículas, cápsides o vectores de virus recombinantes se administran en cantidades suficientes para transfectar las células y proporcionar niveles suficientes de transferencia y expresión génica para proporcionar un beneficio terapéutico sin efectos adversos indebidos, o con efectos fisiológicos médicamente aceptables, que se pueden determinar por los expertos en las materias médicas. Las vías de administración convencional y farmacéuticamente aceptables incluyen, aunque sin limitación, suministro directo a un órgano deseado (por ejemplo, el hígado (opcionalmente a través de la arteria hepática) o el pulmón), vía oral, inhalación, intranasal, intratraqueal, intraarterial, intraocular, intravenosa, intramuscular, subcutánea, intradérmica y otras vías de administración parenteral. Las vías de administración pueden combinarse, si se desea.

Las dosis de las partículas, cápsides o vectores de virus recombinantes dependerán principalmente de factores tales como la afección a tratar, la edad, el peso y la salud del paciente, y por lo tanto pueden variar entre los pacientes. Por ejemplo, una dosificación de un ser humano terapéuticamente eficaz del vector vírico está generalmente en el intervalo de aproximadamente 0,1 ml a aproximadamente 100 ml de solución que contiene concentraciones de aproximadamente 1×10^9 a 1×10^{16} genomas del vector del virus.

Otras proteínas terapéuticas útiles codificadas por los ácidos nucleicos (p. ej., operativamente en combinación con promotores) indicadas en el presente documento incluyen hormonas y factores de crecimiento y diferenciación que incluyen, sin limitación, insulina, glucagón, hormona del crecimiento (GH), hormona paratiroidea (PTH), factor liberador de la hormona del crecimiento (GRF), hormona foliculoestimulante (FSH), hormona luteinizante (LH), gonadotropina coriónica humana (hCG), factor de crecimiento endotelial vascular (VEGF), angiopoyetinas, angiostatina, factor estimulante de colonias de granulocitos (G-CSF), eritropoyetina (EPO), factor de crecimiento del tejido conectivo (CTGF), factor de crecimiento fibroblástico básico (bFGF), factor de crecimiento fibroblástico ácido (aFGF), factor de crecimiento epidérmico (EGF), factor de crecimiento derivado de plaquetas (PDGF), factores de crecimiento insulínico I y II (IGF-I e IGF-II), cualquiera de la superfamilia alfa del factor de crecimiento transformante, incluyendo TGFalfa, activinas, inhibinas o cualquiera de las BMP 1-15 de las proteínas morfogenéticas óseas (BMP), cualquiera de la

familia del factor de diferenciación de hereguina/neuregulina/ARIA/neu (NDF) de los factores de crecimiento, factor de crecimiento nervioso (NGF), factor neurotrófico derivado del cerebro (BDNF), neurotrofinas NT-3 y NT-4/5, factor neurotrófico ciliar (CNTF), factor neurotrófico derivado de la línea de células gliales (GDNF), neurturina, agrina, cualquiera de la familia de semaforinas/colapsinas, netrina-1 y netrina-2, factor de crecimiento de hepatocitos (HGF), efrinas, nogina, sonic hedgehog y tirosina hidroxilasa.

Otras proteínas terapéuticas codificadas por los ácidos nucleicos (p. ej., operativamente en combinación con promotores) indicadas en el presente documento incluyen aquellas que regulan el sistema inmunitario, incluyendo, sin limitación, citocinas y linfocinas tales como trombopoyetina (TPO), interleucinas (IL) IL-1 a IL-25 (incluidas IL-2, IL-4, IL-12 e IL-18), proteína quimioatrayente de monocitos, factor inhibidor de la leucemia, factor estimulante de colonias de granulocitos-macrófagos, ligando Fas, factores de necrosis tumoral alfa y beta, interferones alfa, beta y gamma, factor de células madre, ligando flk-2/flt3. Las proteínas producidas por el sistema inmunitario también son útiles. Estas incluyen, sin limitaciones, inmunoglobulinas IgG, IgM, IgA, IgD e IgE, inmunoglobulinas quiméricas, anticuerpos humanizados, anticuerpos monocatenarios, receptores de linfocitos T, receptores de linfocitos T quiméricos, receptores de linfocitos T de cadena sencilla, moléculas MHC de clase I y clase II, así como inmunoglobulinas y moléculas MHC modificadas por ingeniería. Las proteínas útiles también incluyen proteínas reguladoras del complemento tales como proteínas reguladoras del complemento, proteína cofactor de membrana (MCP), factor acelerador de la descomposición (DAF), CR1, CF2 y CD59.

Otras proteínas terapéuticas codificadas por los ácidos nucleicos (p. ej., operativamente en combinación con promotores) indicadas en el presente documento son receptores para las hormonas, factores de crecimiento, citocinas, linfocinas, proteínas reguladoras y proteínas del sistema inmunitario. La divulgación abarca los receptores para la regulación del colesterol y/o la modulación de los lípidos, incluidos el receptor de lipoproteínas de baja densidad (LDL), el receptor de lipoproteínas de alta densidad (HDL), el receptor de lipoproteínas de muy baja densidad (VLDL) y los receptores neutralizantes. La divulgación también abarca proteínas tales como miembros de la superfamilia de receptores de hormonas esteroideas, incluidos los receptores de glucocorticoides y los receptores de estrógenos, los receptores de vitamina D y otros receptores nucleares. Además, proteínas útiles incluyen factores de transcripción como jun, fos, max, mad, factor de respuesta sérica (SRF), AP-1, AP2, myb, MyoD y miogenina, proteínas que contienen caja ETS, TFE3, E2F, ATF1, ATF2, ATF3, ATF4, ZF5, NFAT, CREB, HNF-4, C/EBP, SP1, proteínas de unión a caja CCAAT, factor de regulación de interferón (IRF-1), proteína tumoral de Wilms, proteína de unión a ETS, STAT, proteínas de unión a caja GATA, por ejemplo, GATA-3 y la familia forkhead de proteínas de hélice alada.

Otras proteínas útiles incluyen, carbamoil sintetasa I, ornitina transcarbamilasa, arginosuccinato sintetasa, arginosuccinato liasa, arginasa, fumarilacetoacetato hidrolasa, fenilalanina hidroxilasa, alfa-1 antitripsina, glucosa- 6-fosfatasa, porfobilinógeno desaminasa, cistación beta-sintasa, cetoácido descarboxilasa de cadena ramificada, albúmina, isovaleril-CoA deshidrogenasa, propionil CoA carboxilasa, metil malonil CoA mutasa, glutaril CoA deshidrogenasa, insulina, beta-glucosidasa, piruvato carboxilasa, fosforilasa hepática, fosforilasa cinasa, glicina descarboxilasa, proteína H, proteína T, una secuencia reguladora transmembrana de fibrosis quística (CFTR) y una secuencia de ADNc de distrofina. Aún otras proteínas útiles incluyen enzimas tales como las que pueden ser útiles en la terapia de reemplazo enzimático, que es útil en una variedad de afecciones que son el resultado de una actividad deficiente de la enzima. Por ejemplo, las enzimas que contienen manosa-6-fosfato pueden utilizarse en terapias para enfermedades por almacenamiento lisosómico (por ejemplo, un gen adecuado incluye el que codifica la beta-glucuronidasa (GUSB)).

Otras proteínas útiles incluyen polipéptidos de origen no natural, tales como polipéptidos quiméricos o híbridos que tienen una secuencia de aminoácidos de origen no natural que contiene inserciones, delecciones o sustituciones de aminoácidos. Por ejemplo, las inmunoglobulinas modificadas por ingeniería de cadena sencilla podrían ser útiles en determinados pacientes inmunocomprometidos. Otros tipos de secuencias génicas de origen no natural incluyen moléculas antisentido y ácidos nucleicos catalíticos, tal como ribozimas, que podrían usarse para reducir la sobreexpresión de una diana.

La reducción y/o modulación de la expresión de una proteína es particularmente deseable para el tratamiento de afecciones hiperproliferativas caracterizadas por células en hiperproliferación, como son los cánceres y la psoriasis. Los polipéptidos diana incluyen aquellos polipéptidos que se producen exclusivamente o a niveles más altos en células hiperproliferativas en comparación con las células normales. Los antígenos diana incluyen polipéptidos codificados por oncogenes tales como myb, myc, fyn, y el gen de translocación bcr/abl, ras, src, P53, neu, trk y EGRF. Además de los productos oncogénicos como antígenos diana, los polipéptidos diana para tratamientos y regímenes protectores contra el cáncer incluyen regiones variables de anticuerpos producidos por linfomas de linfocitos B y regiones variables de receptores de linfocitos T de linfomas de linfocitos T que, en algunas realizaciones, también se usan como antígenos diana para una enfermedad autoinmunitaria. Se pueden usar otros polipéptidos asociados a tumores como polipéptidos diana tales como polipéptidos que se encuentran a niveles más altos en células tumorales que incluyen el polipéptido reconocido por el anticuerpo monoclonal 17-1A y los polipéptidos de unión a folato.

Otros polipéptidos y proteínas terapéuticos adecuados incluyen aquellos que pueden ser útiles para tratar individuos que padecen enfermedades y trastornos autoinmunitarios al conferir una respuesta inmunitaria protectora de base amplia contra dianas que están asociadas con la autoinmunidad, incluidos los receptores celulares y las células que

producen anticuerpos "autodirigidos". Las enfermedades autoinmunitarias mediadas por linfocitos T incluyen artritis reumatoide (AR), esclerosis múltiple (EM), síndrome de Sjögren, sarcoidosis, diabetes mellitus dependiente de insulina (I DDM), tiroiditis autoinmunitaria, artritis reactiva, espondilitis anquilosante, esclerodermia, polimiositis, dermatomiositis, psoriasis, vasculitis, granulomatosis de Wegener, enfermedad de Crohn y colitis ulcerosa. Cada una de estas enfermedades se caracteriza por receptores de linfocitos T (TCR) que se unen a antígenos endógenos e inician la cascada inflamatoria asociada con las enfermedades autoinmunitarias.

Los vectores indicados en el presente documento pueden formularse de una manera que permita la expresión de una proteína transportada por los vectores para inducir una respuesta inmunitaria a un antígeno seleccionado. Por ejemplo, para promover una respuesta inmunitaria, el antígeno puede expresarse a partir de un promotor divulgado en el presente documento, el vector puede adyuvarse como se describe en el presente documento, y/o el vector puede introducirse en tejido degenerado.

Ejemplos de antígenos inmunogénicos adecuados incluyen los seleccionados de una variedad de familias víricas. Ejemplos de familias víricas deseables contra las cuales sería deseable una respuesta inmunitaria incluyen, la familia de los picornavirus, que incluye los géneros rinovirus, que son responsables de aproximadamente el 50 % de los casos del resfriado común; los géneros enterovirus, que incluyen los poliovirus, coxsackievirus, ecovirus y enterovirus humanos, tal como el virus de la hepatitis A; y los géneros aphovirus, que son responsables de las enfermedades de la fiebre aftosa, principalmente en animales no humanos. Dentro de la familia de los virus picornavirus, los antígenos diana incluyen VP1, VP2, VP3, VP4 y VPG. Otras familias víricas incluyen los astrovirus y la familia del calcivirus. La familia del calcivirus abarca el grupo de virus Norwalk, que son un importante agente causante de la gastroenteritis epidémica. Aún otra familia vírica deseable para su uso en el direccionamiento de antígenos para inducir respuestas inmunitarias en seres humanos y animales no humanos es la familia de togavirus, que incluye los géneros alfavirus, que incluyen virus Sindbis, virus Ross River y Encefalitis Equina Venezolana, Oriental y Occidental, y rubivirus, incluyendo el virus de la rubéola. La familia flaviviridae incluye dengue, fiebre amarilla, encefalitis japonesa, encefalitis de St. Louis y virus de la encefalitis transmitida por garrapatas. Se pueden generar otros antígenos diana a partir de la hepatitis C o la familia de coronavirus, que incluye una serie de virus no humanos, tales como el virus de la bronquitis infecciosa (aves de corral), el virus gastroentérico transmisible porcino (cerdo), el virus de la encefalomiелitis hemaglutinata porcina (cerdo), virus de peritonitis infecciosa felino (gatos), coronavirus entérico felino (gato), coronavirus canino (perro) y coronavirus respiratorios humanos, que pueden causar el resfriado común y/o hepatitis no A, B o C, y que incluyen la supuesta causa de síndrome respiratorio agudo repentino (SRAS). Dentro de la familia de los coronavirus, los antígenos diana incluyen la glucoproteína E1 (también llamada proteína M o de matriz), E2 (también llamada proteína S o Spike), E3 (también llamada HE o hemaglutina-elterosa) (no presente en todos los coronavirus) o N (nucleocápside). Aún otros antígenos se pueden dirigir contra la familia de arterivirus y la familia de rabdovirus. La familia de los rabdovirus incluye los géneros vesiculovirus (p. ej., El Virus de la Estomatitis Vesicular) y los géneros lisisavirus (p. ej., rabia). Dentro de la familia de los rabdovirus, los antígenos adecuados pueden derivar de la proteína G o la proteína N. La familia filoviridae, que incluye virus de la fiebre hemorrágica los como el virus de Marburg y el Ébola, puede ser una fuente adecuada de antígenos. La familia de paramixovirus incluye el Virus paragripal Tipo 1, el Virus paragripal Tipo 3, el Virus paragripal Tipo 3 bovino, rubulavirus (virus de las paperas, el Virus paragripal Tipo 2, el Virus paragripal Tipo 4, el virus de la enfermedad de Newcastle (pollos), peste bovina, el morbillivirus, que incluye sarampión y moquillo canino y neumovirus, que incluye el virus sincitial respiratorio. El virus de la gripe se clasifica dentro de la familia de ortomixovirus y es una fuente adecuada de antígenos (por ejemplo, la proteína HA, la proteína N1). La familia de los bunyavirus incluye los géneros bunyavirus (encefalitis de California, La Crosse), flebovirus (fiebre del Valle del Rift), hantavirus pumala es el virus de la fiebre de hemahagina), nairovirus (enfermedad de las ovejas de Nairobi) y varios bungavirus no asignados. La familia arenavirus proporciona una fuente de antígenos contra LCM y contra el virus de la fiebre de Lassa. Otra fuente de antígenos es la familia bornavirus. La familia reovirus incluye los géneros reovirus, rotavirus (que causa gastroenteritis aguda en niños), orbivirus, y cultivirus (virus de la garrapata de Colorado, Lebonbo (seres humanos), encefalosis equina, lengua azul). La familia retrovirus incluye la subfamilia oncovirinal que abarca enfermedades humanas y veterinarias tales como el virus de la leucemia felina, HTLV I y HTLV II, lentivirus (que incluye el VIH, el virus de inmunodeficiencia de simios, el virus de inmunodeficiencia felina, el virus de anemia infecciosa equina, y espumavirinal). La familia papovavirus incluye la subfamilia de polioma virus (virus BKU y JCU) y la subfamilia papilomavirus (asociados con cánceres o progresión maligna de papiloma). La familia adenovirus incluye los virus (EX, AD7, ARD, O.B.) que causan enfermedad respiratoria y/o enteritis. El papovirus felino de la familia de los parvovirus (enteritis felina), el panleucovirus felino, el parvovirus canino, y el parvovirus porcino. La familia herpesvirus incluye la subfamilia alfa herpesvirinae, la cual abarca los géneros simplexvirus (HSV I, HSV II), varicelovirus (pseudorabias, varicela zóster) y la subfamilia beta herpesvirinae, la cual incluye los géneros citomegalovirus (HCMV, muromegalovirus) y la subfamilia gamma herpesvirinae, la cual incluye los géneros linfocriptovirus, EBV (linfoma de Burkitts), herpesvirus 6A humanos, 6B y 7, herpesvirus asociados con sarcoma de Kaposi y cercopithecine herpesvirus (virus B), rinotraqueitis infecciosa, virus de la enfermedad de Marek y radinivirus. La familia de los poxvirus incluye la subfamilia cordopoxvirinae, que abarca los géneros ortopoxvirus (Variola major (Viruela) y Vaccinia (viruela bovina), parapoxvirus, avipoxvirus, capripoxvirus, leporipoxvirus, suipoxvirus y la subfamilia entomopoxvirinae. La familia de los hepadnavirus incluye el virus de la hepatitis B. Un virus no clasificado que puede ser una fuente adecuada de antígenos es el virus de la hepatitis delta, el virus de la hepatitis E y los priones. Otro virus que es una fuente de antígenos es el virus Nipán. Aún otras fuentes víricas pueden incluir el virus de la enfermedad infecciosa de la bolsa aviar y el virus del síndrome respiratorio y reproductivo porcino. La familia de alfavirus incluye el virus de la arteritis equina y varios virus de encefalitis.

La presente divulgación también puede abarcar inmunógenos basados en proteínas que son útiles para inmunizar un animal humano o no humano contra otros patógenos, incluyendo bacterias, hongos, microorganismos parásitos o parásitos multicelulares que infectan vertebrados humanos y no humanos, o de una célula cancerosa o célula tumoral.

5 Los ejemplos de patógenos bacterianos incluyen cocos grampositivos patógenos incluyendo neumococos; estafilococos (y las toxinas producidas por ellos, p.ej., enterotoxina B); y estreptococos. Los cocos patógenos gramnegativos incluyen meningococo; gonococo. Los bacilos gramnegativos entéricos patógenos incluyen enterobacterias; pseudomonas, acinetobacterias y eikenella; melioidosis; salmonella; shigella; haemophilus; moraxella; H. ducreyi (que causa chancroide); especies de brucella (brucelosis); Francisella tularensis (que causa tularemia); Yersinia pestis (peste) y otras Yersinia (pasteurella); streptobacillus moniliformis y spirillum; Los bacilos grampositivos incluyen listeria monocytogenes; erysipelotheix rhusiopathiae; Corynebacterium diphtheriae (difteria); cólera; B. anthracis (ántrax); donovanosis (granuloma inguinal); y bartonellosis. Las enfermedades causadas por bacterias anaerobias patógenas incluyen el tétanos; botulismo (Clostridium botulinum y su toxina); Clostridium perfringens y su toxina épsilon; otros clostridios; tuberculosis; lepra; y otras micobacterias. Las enfermedades por

10 enterobacterias patógenas incluyen sífilis; treponematosis; pian, sífilis endémica y pinta; y leptospirosis. Otras infecciones causadas por bacterias patógenas superiores y hongos patógenos incluyen muermo (Burkholderia mallei); actinomicosis; nocardiosis; criptococosis, blastomicosis, histoplasmosis y coccidioidomicosis; candidiasis, aspergilosis y mucormicosis; esporotricosis; paracoccidioidomicosis, petrieliidiosis, torulopsosis, micetoma y cromomicosis; y dermatofitosis. Las infecciones por rickettsia incluyen fiebre tifoidea, fiebre maculosa de las Montañas Rocosas, Fiebre

15 Q (Coxiella burnetti) y viruela rickettsiósica. Ejemplos de micoplasma e infecciones por clamidias incluyen: mycoplasma pneumoniae; linfogranuloma venéreo; psitacosis; e infecciones perinatales por clamidias. Los eucariotas patógenos abarcan protozoos y helmintos patógenos, y las infecciones producidas incluyen: amebiasis; malaria; leishmaniosis; tripanosomiasis; toxoplasmosis; Pneumocystis carinii; Trichans; Toxoplasma gondii; babesiosis; giardiasis; triquinosis; filariasis; esquistosomiasis; nematodos; trematodos o duelas; e infecciones por cestodos (tenia).

25 Muchos de estos organismos y/o las toxinas producidas por los mismos han sido identificados por los Centros para el Control de Enfermedades [(CDC), Departamento de Salud y Servicios Humanos, EE. UU.], como agentes que pueden usarse en ataques biológicos. Por ejemplo, algunos de estos agentes biológicos, incluyen, Bacillus anthracis (ántrax), Clostridium botulinum y su toxina (botulismo), Yersinia pestis (peste), variola mayor (viruela), Francisella tularensis (tularemia) y fiebres hemorrágicas víricas [filovirus (p. ej., Ébola, Marburg) y arenavirus [por ejemplo, Lassa, Machupo]], todos los cuales están actualmente clasificados como agentes de Categoría A; Coxiella burnetti (fiebre Q); especies de Brucella (brucelosis), Burkholderia mallei (muermo), Burkholderia pseudomallei (melioidosis), Ricinus communis y su toxina (toxina ricina), Clostridium perfringens y su toxina (toxina épsilon), Especies de estafilococos y sus toxinas (enterotoxina B), Chlamydia psittaci (psitacosis), amenazas a la seguridad del agua (p. ej., Vibrio cholerae, Crytosporidium parvum), fiebre tifoidea (Rickettsia powazekii) y encefalitis vírica (alfavirus, por ejemplo, encefalitis equina Venezolana; encefalitis equina del este; encefalitis equina del oeste); todos los cuales están actualmente clasificados como agentes de Categoría B; y virus Nipán y hantavirus, que actualmente están clasificados como

30 agentes de Categoría C. Además, otros organismos, que se clasifican así o de manera diferente, pueden identificarse y/o usarse para tal fin en el futuro. Se entenderá fácilmente que los vectores víricos y otras construcciones descritas en el presente documento son útiles para suministrar antígenos de estos organismos, virus, sus toxinas u otros subproductos, que evitarán y/o tratarán infecciones u otras reacciones adversas con estos agentes biológicos.

En determinados aspectos, una proteína para la expresión en un vector de la divulgación es un segmento de una región variable de linfocitos T que desencadena una respuesta inmunitaria, es decir, para eliminar los linfocitos T citotóxicos. En artritis reumatoide (AR), se han caracterizado varias regiones variables específicas de TCR que están implicadas en la enfermedad. Estos TCR incluyen V-3, V-14, V-17 y V-17. Por lo tanto, el suministro de una secuencia del ácido nucleico que codifica al menos uno de estos polipéptidos desencadenará una respuesta inmunitaria que se dirigirá a los linfocitos T implicados en la AR. En esclerosis múltiple (EM), se han caracterizado varias regiones variables específicas de TCR que están implicadas en la enfermedad. Estos TCR incluyen V-7 y V-10. Por lo tanto, el

45 suministro de una secuencia del ácido nucleico que codifica al menos uno de estos polipéptidos desencadenará una respuesta inmunitaria que se dirigirá a los linfocitos T implicados en la EM. En escleroderma, se han caracterizado varias regiones variables específicas de TCR que están implicadas en la enfermedad. Estos TCR incluyen V-6, V-8, V-14 y V-16, V-3C, V-7, V-14, V-15, V-16, V-28 y V-12. Por lo tanto, el suministro de una molécula de ácido nucleico que codifica al menos uno de estos polipéptidos desencadenará una respuesta inmunitaria que se dirigirá a los

50 linfocitos T implicados en escleroderma.

Los vectores víricos recombinantes de la divulgación proporcionan un vehículo de transferencia génica eficaz que puede suministrar proteínas seleccionadas a una célula hospedadora seleccionada *in vivo* o *ex vivo* incluso cuando el organismo tiene anticuerpos neutralizantes de la proteína. En un aspecto, los vectores descritos en el presente

60 documento y las células se mezclan *ex vivo*; las células infectadas se cultivan usando metodologías convencionales; y las células transducidas se vuelven a infundir en el paciente.

Realizaciones adicionales

65 En determinadas realizaciones, la invención se refiere a vectores víricos recombinantes que comprenden una secuencia del ácido nucleico de un promotor específico de hígado de la invención en combinación operativa con una

secuencia del ácido nucleico heterólogo que codifica una proteína. Generalmente, la secuencia del ácido nucleico que codifica la proteína comprende un mayor porcentaje de codones de aminoácidos específicos de células hepáticas en comparación con el uso general de codones humanos. En determinados aspectos, la divulgación se refiere a métodos para tratar a un sujeto diagnosticado con un rasgo genético que da como resultado la expresión de una proteína no funcional mutada o truncada mediante la administración de una cantidad eficaz de un vector divulgado en el presente documento configurado para expresar una proteína funcional del hígado.

En determinadas realizaciones, el vector comprende una secuencia de un ácido nucleico vírico de más de 10, 20, 30, 40, 50, 100 o 200 nucleótidos. En determinadas realizaciones, la secuencia de un ácido nucleico vírico es un segmento del virus adenoasociado humano (hAAV) de los serotipos 1, 2, 3B, 4, 5, 6, 7, 8, 9 o combinaciones o variantes de los mismos, que generalmente comprende una repetición terminal invertida de AAV.

En determinados aspectos, la divulgación se refiere a una partícula vírica, por ejemplo, cápside que comprende un vector divulgado en el presente documento, por ejemplo, el vector empaquetado en una cápside. La cápside puede ser una cápside o partícula recombinante o quimérica, por ejemplo, una cápside que tiene secuencias de aminoácidos que son una combinación de pseudotipos de AAV para VP 1, 2 o 3. Un VP de cápside de AAV puede derivar de un gen humano o gen de AAV animal, o combinaciones con alteraciones genéticamente modificadas, es decir, AAV aislado de células humanas o un primate no humano infectados. Los AAV de animales incluyen los derivados de aves, bovino, porcino, ratones, etc. En determinados aspectos, la cápside puede tener secuencias de aminoácidos que son genéticamente modificados o cápsides sintéticas identificadas mediante métodos como la evolución dirigida o el diseño racional.

En determinadas realizaciones, los vectores divulgados en el presente documento son un ácido nucleico monocatenario o bicatenario o un ácido nucleico autocomplementario de menos de 5,1, 5,0, 4,9, 4,8, 4,7, 4,6 o 4,5 kb de nucleótidos en total. En determinadas realizaciones, el vector es incompetente para la replicación dentro de un hospedador humano, por ejemplo, el vector no codifica una polimerasa vírica.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTA (SEQ ID NO: 21) ATAATCTCAGGACAAACA (SEQ ID NO: 43) y/o TATAAAAGGCCAGCAGCAGCCTGACCACATCT- CATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 21 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 21 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos. En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 4, 5, 6, o 7.

Como se usan en el presente documento, el promotor específico de hígado se refiere a las secuencias en la dirección 5' del sitio de inicio de la transcripción de la proteína a producir. Las secuencias promotoras divulgadas en el presente documento pueden contener combinaciones de otros promotores, potenciadores, otras secuencias y fragmentos de los mismos conocidos. Por ejemplo, la SEQ ID NO: 21 es la secuencia del potenciador ABP acortado. Por sí solo no funciona como un promotor, sirve para potenciar la expresión conferida por el promotor central, SEQ ID NO: 20, SynO, que por sí solo no confiere una expresión génica eficaz. Se contemplan secuencias potenciadoras adicionales que pueden reemplazar o aumentar el potenciador short ABP.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con CGGAGGAGCAAACAGGG (SEQ ID NO: 97) y/o TATAAAAG- GCCAGCAGCAGCCTGACCACATCTCATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 97 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 97 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con CGGAGGAGCAAACAGGGGCTAAGTCCAC (SEQ ID NO: 98) y/o TATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 98 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 98 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con GGCTGCTGGTGAATATTAACCAAGGTC (SEQ ID NO: 99) y/o TATAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 99 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 99 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGT- TATCGGAGGAGCAAACAGGGG CTAAGTCCAC (SEQ ID NO: 100) y/o TATAAAGGCCAGCAGCAGCCTGAC- CACATCTCATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 100 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 100 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGT- TATCGGAGGAGCAAACAGGGA CTAAGTCCAC (SEQ ID NO: 101) y/o TATAAAGGCCAGCAGCAGCCTGAC- CACATCTCATCCTC (SEQ ID NO: 20).

En determinados aspectos, la SEQ ID NO: 20 es 3' o posterior a la SEQ ID NO: 101 entre un enlazador nucleotídico, por ejemplo, conectado con un enlazador como se ilustra: 5'- SEQ ID NO: 101 seguida de un enlazador seguido de SEQ ID NO: 20-3'. El enlazador puede ser de entre 0 a 200 nucleótidos, de 10 a 100 nucleótidos, de 20 a 70 nucleótidos, de 30 a 60 nucleótidos, de 30 a 40 nucleótidos, de 32 a 36 nucleótidos.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 4, 5, 6, 7, 102, 103, 104, 105, 106, 107 o 108.

En determinados aspectos, la secuencia del ácido nucleico del promotor específico de hígado es inferior a 205 o 250 nucleótidos. Las expresiones "menos de 205 o 250 nucleótidos" se refieren a la longitud de una cadena sencilla o la longitud de pares de bases de cadena doble. El promotor funcional es bicatenario después de la conversión intracelular posterior a una infección vírica. No funcionaría si fuera monocatenario.

En determinados aspectos, la divulgación contempla que el primer nucleótido de la secuencia promotora es la 5' "G" en el sitio de unión al TF HFN1a, GTTAAT (SEQ ID NO: 25), por ejemplo, el promotor es 5' - SEQ ID NO: 21, seguido de un enlazador y seguido adicionalmente por un sitio de inicio de la transcripción (TSS), por ejemplo TCATCCTC (SEQ ID NO: 109), en donde el último nucleótido en el sitio de inicio de la transcripción es el final de la secuencia promotora. En determinados aspectos, el enlazador comprende una caja TATAA (SEQ ID NO: 26) y un espaciador rico en GC. En determinadas realizaciones, la divulgación contempla que el primer nucleótido de la secuencia promotora es una 5' "G" en GTTAA (SEQ ID NO: 27), GTTA (SEQ ID NO: 28), GTT (SEQ ID NO: 29). En determinados aspectos, el primer nucleótido de la secuencia promotora es una 5' "T" en TTAAT (SEQ ID NO: 30), TTAA (SEQ ID NO: 31), TTA (SEQ ID NO: 32). En determinados aspectos, el primer nucleótido de la secuencia promotora es una 5' "A" en AAT (SEQ ID NO: 33).

En determinados aspectos, la divulgación contempla un promotor que comprende 5'-SEQ ID NO: 21 o 97 o 98 o 99 o 100 o 101 opcionalmente seguido de un enlazador, seguido de una caja TATAA, seguido de un espaciador rico en GC seguido de un sitio de inicio de la transcripción. En determinados aspectos, el espaciador rico en GC es una secuencia en donde más del 60, 70, 80 o 90 % de los nucleótidos son G o C, por ejemplo, en una ventana de 5 a 35 nucleótidos, de 10 a 30 nucleótidos, o de 15 a 40 nucleótidos, o de 5 a 50 o 60 nucleótidos. En determinados aspectos, el espaciador rico en GC es GGCCAGCAGCAGCCTGACCACATC (SEQ ID NO: 110). En determinadas realizaciones, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 110.

En determinados aspectos, la secuencia promotora específica de hígado comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con una de:

la SEQ ID NO: 34: GTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACA;
la SEQ ID NO: 35: TTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACA;
la SEQ ID NO: 36: TAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACA;
la SEQ ID NO: 37: AATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACA;

- la SEQ ID NO: 38: ATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 39: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 40: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 41: TTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 5 la SEQ ID NO: 42: TTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 44: TGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 45: GTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 46: TGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 47: GCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 10 la SEQ ID NO: 48: CCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 49: CCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 50: CTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAACA;
 la SEQ ID NO: 51: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAAC;
 la SEQ ID NO: 52: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAA;
 15 la SEQ ID NO: 53: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAA;
 la SEQ ID NO: 54: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAA;
 la SEQ ID NO: 55: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACA;
 la SEQ ID NO: 56: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGAC;
 la SEQ ID NO: 57: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGA;
 20 la SEQ ID NO: 58: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGG;
 la SEQ ID NO: 59: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAG;
 la SEQ ID NO: 60: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCA;
 la SEQ ID NO: 61: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTC;
 la SEQ ID NO: 62: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCT;
 25 la SEQ ID NO: 63: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATC;
 la SEQ ID NO: 64: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAAT;
 la SEQ ID NO: 65: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAA;
 la SEQ ID NO: 66: GTTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAAC;
 la SEQ ID NO: 67: TTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAA;
 30 la SEQ ID NO: 68: TAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAA;
 la SEQ ID NO: 69: AATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACA;
 la SEQ ID NO: 70: ATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGAC;
 la SEQ ID NO: 71: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGA;
 la SEQ ID NO: 72: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGG;
 35 la SEQ ID NO: 73: TTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAG;
 la SEQ ID NO: 74: TTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCA;
 la SEQ ID NO: 75: TGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTC;
 la SEQ ID NO: 76: GTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCT;
 la SEQ ID NO: 77: TGGCCCTTGCGATGTTTGCTCTGGTTAATAATC;
 40 la SEQ ID NO: 78: GCCCTTGCGATGTTTGCTCTGGTTAATAAT;
 la SEQ ID NO: 79: CCCTTGCGATGTTTGCTCTGGTTAATAA;
 la SEQ ID NO: 80: CCTTGCGATGTTTGCTCTGGTTAAT A;
 la SEQ ID NO: 81: CTTGCGATGTTT GCTCTGGTTAAT;
 la SEQ ID NO: 82: TTAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAAAC;
 45 la SEQ ID NO: 83: TAATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAA;
 la SEQ ID NO: 84: AATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAA;
 la SEQ ID NO: 85: ATTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAA;
 la SEQ ID NO: 86: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACA;
 la SEQ ID NO: 87: TTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGAC;
 50 la SEQ ID NO: 88: TTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGA;
 la SEQ ID NO: 89: TTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGG;
 la SEQ ID NO: 90: TGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAG;
 la SEQ ID NO: 91: TGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCA;
 la SEQ ID NO: 92: GGCCCTTGCGATGTTTGCTCTGGTTAATAATCTC;
 55 la SEQ ID NO: 93: GCCCTTGCGATGTTTGCTCTGGTTAATAATCT;
 la SEQ ID NO: 94: CCCTTGCGATGTTTGCTCTGGTTAATAATC;
 la SEQ ID NO: 95: CCTTGCGATGTTTGCTCTGGTTAATAAT; o
 la SEQ ID NO: 96: CTTGCGATGTTTGCTCTGGTTAATAA.
- 60 En determinados aspectos, el promotor puede comenzar con el primer nucleótido 5' de la SEQ ID NO: 34-96 y terminar con el último nucleótido del TSS.
- En determinadas realizaciones, la proteína es un fVIII o fIX o una variante de los mismos. En determinadas realizaciones, los esquemas de optimización de codones y promotores divulgados en el presente documento podrían usarse para cualquier terapia génica con AAV dirigida al hígado. Se contemplan otras enfermedades metabólicas causadas por deficiencias de enzimas hepáticas y la expresión de esas proteínas funcionales.
- 65

En determinadas realizaciones, la variante de fVIII comprende un dominio A1, un dominio A2, una secuencia RHQR (SEQ ID NO: 24), un dominio A3, un dominio C1 y un dominio C2. En determinadas realizaciones, la variante de fVIII comprende un dominio B eliminado.

5 En determinadas realizaciones, la variante de fVIII comprende un enlazador de entre dos y cincuenta, o dos y veinticinco, o dos y quince aminoácidos entre el dominio A2 y el dominio A3.

10 En determinadas realizaciones, la variante de fVIII comprende un dominio A1, un dominio A2, un dominio de péptido de activación (ap, por sus siglas en inglés), un dominio A3, un dominio C1 y un dominio C2. En determinadas realizaciones, la variante de fVIII comprende un dominio B eliminado.

15 En determinadas realizaciones, la variante de fVIII comprende un enlazador de entre dos y cincuenta, o dos y veinticinco, o dos y quince aminoácidos entre el dominio A2 y el un dominio de péptido de activación (ap).

En determinadas realizaciones, la variante de fVIII comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 3.

20 En determinadas realizaciones, la variante de fVIII comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 13, 14, 15 o 16.

25 En determinados aspectos, la divulgación se refiere a métodos en los que el uso de codones de un gen se ajusta de acuerdo con el tejido en que se expresará, por ejemplo, tejido hepático. En determinados aspectos, la secuencia del ácido nucleico que codifica una proteína comprende codones que se utilizan o representan de manera diferencial en genes altamente expresados dentro del hígado u otro tejido específico en comparación con el uso de codones de toda la región de codificación del genoma humano y evita los codones que están subrepresentados en el hígado o en otro tejido específico.

30 En determinadas realizaciones, la secuencia del ácido nucleico que codifica la proteína comprende codones para más del 50, 60, 70, 80, 90 o 95 % o 100 % de los aminoácidos que se prefieren como se proporciona en la Figura 2.

35 En determinadas realizaciones, la secuencia del ácido nucleico que codifica una proteína que comprende un mayor porcentaje de codones de aminoácidos específicos de células hepáticas en comparación con el uso general de codones humanos es una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 1.

En determinadas realizaciones, la variante de fIX comprende una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 17, 18 o 19.

40 En determinadas realizaciones, la secuencia del ácido nucleico que codifica una proteína que comprende un mayor porcentaje de codones de aminoácidos específicos de células hepáticas en comparación con el uso general de codones humanos es una secuencia que tiene al menos un 50, 60, 70, 80, 85, 90, 95, 96, 97, 98, 99, o 100 % de identidad de secuencia con la SEQ ID NO: 8, 9 o 10.

45 En determinadas realizaciones, la secuencia del ácido nucleico que codifica la proteína comprende al menos uno de a) a g) en donde,

- a) un codón ATC es en más del 50 % o 52 % para Ile;
- b) un codón ATC es en más del 38 % o 40 % para Thr;
- 50 c) un codón TTC es en más del 57 % o 59 % para Phe;
- d) un codón GAG es en más del 60 % o 62 % para Glu;
- e) un codón CTG es en más del 43 % o 45 % para Leu;
- f) un codón AAG es en más del 60 % o 62 % para Lys; y/o
- 55 g) un codón GAC es en más del 56 % o 58 % para Asp.

En determinadas realizaciones, la secuencia del ácido nucleico que codifica la proteína comprende al menos dos o más, o tres o más, o cuatro o más, cinco o más, cinco o más, seis o más, o todos de a), b), c), d), e), f) y g).

60 En determinadas realizaciones, la secuencia del ácido nucleico que codifica la proteína comprende menos de 100, 50, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 o ningún dinucleótido 5'-CG-3'.

La invención proporciona composiciones farmacéuticas que comprenden un vector de la invención y un excipiente farmacéuticamente aceptable.

65 La invención también proporciona un vector o composición de la invención para su uso en un método para inducir la coagulación de la sangre, comprendiendo dicho método administrar una cantidad eficaz del vector o composición a

un sujeto que lo necesite.

En determinadas realizaciones, el sujeto es diagnosticado con hemofilia A o B o hemofilia adquirida o es poco probable que responda a infusiones exógenas de fVIII.

En determinadas realizaciones, el vector, la partícula vírica o cápside se administra en combinación con un agente inmunosupresor, *por ejemplo*, ciclosporina, tacrolimus, sirolimus, ciclofosfamida, metotrexato, azatioprina, mercaptopurina, fluorouracilo, ácido micofenólico, dactinomicina, fingolimod, anticuerpo o proteína de unión a receptores de linfocitos T, muromonab-CD3, anticuerpo o proteína de unión al receptor de IL-2, basiliximab, daclizumab, IFN-beta recombinante, anticuerpo o proteína de unión a TNF-alfa, infliximab, etanorcept, adalimumab o combinaciones de los mismos.

En determinados aspectos, la divulgación se refiere a sistemas de expresión que comprenden ácidos nucleicos o vectores que comprenden ácidos nucleicos divulgados en el presente documento.

Los siguientes ejemplos se proporcionan para ilustrar determinadas características particulares y/o realizaciones. Estos ejemplos no deben interpretarse como limitantes de la divulgación a las características particulares o realizaciones descritas.

Ejemplos

Ejemplo 1

Optimización de los factores de coagulación y su ADN codificante

Este ejemplo ilustra la optimización de las secuencias codificantes para las proteínas fVIII y fIX para mejorar su utilidad para la expresión *in vivo* y la terapia génica.

La secuencia de nucleótidos del ADNc que codifica para fVIII y fIX se optimizó inicialmente implementando una preferencia de uso de codones específicos para la célula hepática humana en comparación con la secuencia de nucleótidos de origen natural que codifica la secuencia correspondiente sin optimización de codones para un ser humano.

La adaptabilidad de una secuencia de nucleótidos que codifica fVIII al uso de codones de las células hepáticas humanas puede expresarse como índice de adaptación de codones hepáticos (LCAI por sus siglas en inglés). Un índice de adaptación de codones se define como una medida de la adaptabilidad relativa del uso de codones de un gen hacia el uso de codones de genes altamente expresados en el hígado humano. La adaptabilidad relativa de cada codón es la relación entre el uso de cada codón, y el del codón más abundante para el mismo aminoácido. El CAI se define como la media geométrica de estos valores de adaptabilidad relativa. Se excluyen los codones no sinónimos y los codones de terminación (dependientes del código genético). Los valores de CAI varían de 0 a 1, indicando los valores más altos una mayor proporción de los codones más abundantes. Usando las secuencias de 43 genes altamente expresados en el hígado humano, se construyó una tabla personalizada de preferencia de uso de codones específica para el hígado humano (véase la Figura 2A) que difiere sustancialmente de los codones más prevalentes usados en secuencias codificantes humanas totales (Figura. 2B). La secuencia codificante optimizada de las variantes de fVIII ET3 y HSQ se desarrolló con los codones estadísticamente más prevalentes identificados por el análisis de preferencia de uso en el hígado.

ET3 es un fVIII híbrido con el dominio B eliminado (BDD) que contiene dominios humanos y porcinos, es decir, secuencia (A1 y A3 porcino, véase la Figura 1A y 1B) con un enlazador en el dominio B eliminado. ET3 utiliza una secuencia enlazadora OL derivada de una secuencia porcina de 24 aminoácidos, es decir, la secuencia SFAQNSRPPSASAPKPPVLRHQ (SEQ ID NO: 23) derivada de la porcina. La secuencia de aminoácidos de ET3 es la SEQ ID NO: 123:

MQLELSTCVFLCLLPLGFSAIRRYLGAVELSWDYRQSELLRELHVDTRFPATAPGALPLGPSVLYKKTVFVEFTDQL
FSVARPRPFWMLLGLPTIQAEVYDVTVVTLKNMASHPVSLHAVGVSWKSSSEGAIEYEDHTSQREKEDDKVLPKGSQTY
VWQVLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLTRERTQNLHEFVLLFAVFDEGKSWHSARND
SWTRAMPAPARAQAPAMHTVNGYVNRSLPGLIGCHKKSVMYHVGIMGTSPEVHSIFLEGHTFLVRHHRQASLEISPLT
FLTAQTFLMDLGQFLFLFCHISSHHHGMEAHVRVESCAEFPQLRRKADEEEDYDDNLYDSMDVVRDLGDDVSPFIQI
RSVAKKHPKTWVHYIAAEEDWDYAPLV LAPDDRYSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGIL

GPLYLYGEVGDITLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPR
CLTRYYSFVNMERDLASGLIGPLLYCYKESVDQGRNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLED
PEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMYEDTTLTLPFSGETVFMMSME
NPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDI SAYLLSKNNAIEPRSFQNSRPPSASAPKPPVLR
HQRDISLPTFQPEEDKMDYDDIFSTETKGEDFDIYGEDENQDPRSFOQRTRHYFIAAVEQLWDYGMSESPRALNRQA
NGEVRPFKKVVFREFADGSFTQPSYRGELNKLHLLGPYIRAEVEDNIMVTFTKNQASRPYSFYSSLI SYPDDQEQGAE
PRHNFVQPNETRYTFWKVQHMAPTEDEFDCKAWAYFSDVDLEKDVHSGLIGPLLYCRANTLNAAHGRQVTVOEFALF
FTIFDETKSWYFTENVERNCRAPCHLQMEDPTLKENYRFHAINGYVMDTLPGLVMAQNQRIRWYLLSMGSNENIHSIH
FSGHVFSVRKKEEYKMAVYNLYPGVFETVEMLPKSGIWRIECLIGEHLQAGMSTTLFVYSKKCQTPGLMASGHIRDF
QITASGQYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQ
TYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQIT
ASSYFTNMFAWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLTSMYVKEFLISSSQDGH
QWTLFFQNGKVKVFQGNQDSFTPVVNSLDPLLLTRYLRHPQSWVHQIALRMEVLGCEAQDLY

HSQ es una variante de fVIII humano en donde la proteína fVIII humana BDD se sustituye con un enlazador SQ, SFSQNPPVLKRHR (SEQ ID NO: 22) derivado humano de 14 aminoácidos. La secuencia de aminoácidos de HSQ es la SEQ ID NO: 3:

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPFPRVPKSFPPNTSVVYKKTFLVEFTVHLF
NIAKPRPFWMLLGLPTIQAEVYDVTVVITLKNMASHPVSLHAVGVSYWKASEGAIEYDDQTSQREKEDDKVFPGGSHTYV
WQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNS
LMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYHVGIMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITF
LTAQTLMLDLGQFLFLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRDNDNSPSFIQI
RSVAKKHPKTWVHYIAAEEDWDYAPLV LAPDDRYSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGIL
GPLYLYGEVGDITLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPR
CLTRYYSFVNMERDLASGLIGPLLYCYKESVDQGRNQIMSDKRNVLFSVFDENRSWYLTENIQRFLPNPAGVQLED
PEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMYEDTTLTLPFSGETVFMMSME
NPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDI SAYLLSKNNAIEPRSFQNSPVPVLKRHRQREITRTTL
QSDQEEIDYDDTISVEMKKEDFDIYEDENQSPRSFOQKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVPOFKKV
VFQEFDTGSFTQPLYRGELNEHLGGLPYIRAEVEDNIMVTFRNQASRPYSFYSSLI SYEEDQEQGAEPRKNFVKPNE
TKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLYVCHTNTLNPAHGRQVTVOEFALFFTIIFDETKSW
YFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIHSIHFSGHVFTVRK
KEEYKMAVYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTTLFVYSNKCQTPGLMASGHIRDFQITASGQYQ
WAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTL
MVFFGNVDSSGIKHNIFNPPIIARYIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFA
TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLTSMYVKEFLISSSQDGHQWTLFFQNGK
VKVFQGNQDSFTPVVNSLDPLLLTRYLRHPQSWVHQIALRMEVLGCEAQDLY

Tanto HSQ como ET3 contienen la secuencia de reconocimiento RHQR (SEQ ID NO: 24) para la secuencia de procesamiento PACE/furina para el dominio B.

La secuencia de ET3 con codones de hígado optimizados es (SEQ ID NO: 12)

ATGCAGCTGGAACGTGTCTACCTGTGTGTTTCTGTGTCTGCTGCCTCTGGGGTTTTCTGCTATCCGCCGCTACTATCTG
GGAGCCGTGGAGCTGTCTGGGACTACAGGCAGAGCGAGCTGCTGAGAGAACTGCACGTGGATACCAGATTCCCAGCT
ACCGCTCCAGGAGCTGTGCTCTGGGCCATCCGTGCTGTACAGAAACCCTCTTCGTGGAGTTTACCAGACAGCTG
TTCAGCGTGGCCAGGCCAAGACCACCTTGGATGGGACTGCTGGGACCAACCATCCAGGCTGAGGTGTACGATACCGTG
GTCGTGACCCTGAAAAACATGGCCTCCCATCCCGTGAGCCTGCACGCTGTGCGGGTGTCTTCTGGAAGTCCAGCGAG
GGAGCCGAGTACGAAGACCATACTCCAGCGCGAGAAAGAACGATAGGTGCTGCCTGGCAAAAGCCAGACCTAT
GTCTGGCAGGTGCTGAAGGAGAACGGACCAACCGCTAGCGACCCACCATGCTGACCTACTCTTATCTGTCCACGTC
GATCTGGTGAAGGACCTGAATTCCGGACTGATCGGAGCTCTGCTGGTGTGTAGAGAGGGAAGCCTGACCAGAGAAAGA
ACCCAGAACCTGCATGAGTTTCGTCTGCTGTTTCGCCGTGTTTGACGAAGGGAAGAGCTGGCACTCTGCCCGCAATGAC
TCCTGGACCAGAGCTATGGATCCAGCTCCTGCTAGAGCTCAGCCTGCTATGCACACCGTCAACGGCTACGTGAATCGG
TCTCTGCCAGGACTGATCGGCTGCCATAAGAAAAGCGTCTATTGGCACGTGATCGGAATGGGCACCAGCCCCGAGGTG
CATTCTATCTTCTGGAAGGCCACCTTTCTGGTCAGGCACCATAGACAGGCTCTCTGGAGATCTCCCTCTGACC
TTCCTGACCGCTCAGACCTTTCTGATGGACCTGGGGCAGTTCTGCTGTTTTGCCATATCTCTTCCACCATCACGGA
GGAATGGAGGCTCAGTTCAGGTGGAATCCTGTGCTGAGGAACACAGCTGAGAAGAAAGGCTGATGAGGAAGAGGAC
TACGACGATAACCTGTATGACAGCGATATGGACGTCGTGCGCCTGGACGGCGACGATGTCAGCCCTTTTCCAGATC
CGGTCTGTGGCCAAGAAACATCCAAAGACCTGGGTCCACTACATCGCCGCTGAAGAGGAAGATTGGGACTATGCCCC
CTGGTGTGGCTCCTGACGATAGATCCTACAAAAGCCAGTATCTGAACAATGGGCCCCAGCGCATCGGACGGAAGTAC
AAGAAAGTGAGGTTTATGGCCTATACCGACGAGACCTTTAAGACCAGAGAGGCTATCCAGCACGAATCCGGGATCCTG
GGACCTCTGCTGTACGGCGAAGTGGGGGATACCTGCTGATCATCTTCAAGAACCAGGCCCTCCAGGCCATACAATATC

TATCCCCATGGCATCACCGACGTGAGACCACTGTACAGCAGGAGACTGCCCAAGGGGGTCAAACACCTGAAGGATTTT
CCCATCCTGCCTGGAGAGATCTTTAAGTATAAATGGACCGTCACCGTGGAAAGACGGGCCCTACCAAGTCCGATCCACGC
TGCTGACCCGGTACTATAGCTCTTTTCGTGAACATGGAGAGAGACCTGGCTAGCGGACTGATCGGACCCCTGCTGATC
TGTTACAAAGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGATAAGAGAAATGTCATCCTGTTCTCCGTGTTT
GACGAGAACCAGCTGGTACCTGACCGAGAACATCCAGCGGTTCTGCCAAATCCAGCTGGAGTGCAGCTGGAGGAC
CCAGAATTTCAGGCTTCCAACATCATGCATAGCATCAATGGCTACGTGTTTCGATAGCCTGCAGCTGTCTGTCTGCCTG
CACGAGGTGGCTACTGTTATATCCTGTCCATCGCGCTCAGACCCACTTCTGTCCGTGTTCTTTAGCGGGTACACC
TTTAAGCATAAAATGGTGTATGAGGATACCTGACCTGTTTCCCCTTTTTCTGGCGAGACCGTGTTCATGTCCATGGAA
AACCTTGGCCTGTGGATCCTGGGGTGCCACAACAGCGACTTCAGGAATAGAGGAATGACCGCCCTGCTGAAAAGTGTCC
AGCTGTGATAAGAATACCGGCGATTACTATGAGGACTCTTACGAAGATATCTCCGCTTATCTGCTGAGCAAGAACAAT
GCCATCGAGCCAGGTCTTTCTGCTCAGAACTCCAGACTCCAAGCGCTTCTGCTCCTAAGCCACCTGTGCTGAGAAGA
CATCAGAGGGACATCTCCCTGCCCTACCTTCCAGCCAGAGGAAGATAAAATGGACTACGACGATATCTTACGACCCGAG
ACCAAGGGGGAAGATTTTGACATCTATGGAGAGGACGAAAACCAGGATCCAAGATCCTTCCAGAAGAGAACCAGACAC
TACTTTATCGCCGCTGTGGAGCAGCTGTGGGACTATGGGATGTCCGAAAGCCCACGGGCCCTGAGGAACAGAGCTCAG
AATGGAGAGGTGCCCCGTTCAAGAAAGTCTGTTTCCGGAGTTTGGCGACGCGAGCTTTACCCAGCCATCTTACAGG
GGGAGCTGAACAAGCATCTGGGGCTGCTGGGACCTATATCAGAGCCGAGGTGCGAAGATAACATCATGTTGACCTTC
AAGAATCAGGCTTCTCGCCCCCTACTCTTTTATCTTCCCTGATCTCTTACCTGACGATCAGGAGCAGGGCGCCGAA
CCTAGGCACAACCTTCGTGCAGCCAAATGAGACCAGAACCTACTTTTGAAGGTGCAGCATCACATGGCTCCCACCGAG
GATGAATTCGACTGCAAAGCTTGGGCCATTTTTTCCGATGTGACCTGGAGAAGGACGTGCATAGCGGCTGATCGGG
CCTCTGCTGATCTGTGCGGCCAACACCTGAATGCTGCTCACGGAAGACAGGTACCGTGCAGGAGTTCGCTCTGTTT
TTTACCATCTTTGACGAAACCAAGAGCTGGTACTTCACCGAGAACGTGGAAGGAATTGCAGAGCCCCCTGTCTATCTG
CAGATGGAGGACCTTACCTGAAGGAAAACATCAGGTTCCACGCGCATCAATGGATATGTCTATGGATACCATGACATGCGC
CTGGTCTATGGCTCAGAACACGCGCATCCGGTGGTACCTGCTGTCTATGGGATCCAACGAGAAATCCATAGCATGCGC
TTCTCTGGCCATGCTTTTTTCCGTGAGGAAGAAAGAGGAATACAAAATGGCCGTGTACAATCTGTATCTGGGGTCTTC
GAGACCGTGGAAATGCTGCCAAGCAAAGTGGGAATCTGGAGAATCGAGTGCCTGATCGGCGAACACCTGCAGGCGGGG
ATGAGCACCACTTCTGGTGTACTCTAAGAAATGTCAGACCCACTGGGGATGGCCTCCGGACATATCCGCGACTTC
CAGATACCCGCTAGCGGACAGTACGGACAGTGGGCTCAAAGCTGGCTAGACTGCACTATTCTGGCTCCATCAACGCC
TGGTCTACCAAGAGCCATTCTCCTGGATCAAGGTGGACCTGCTGGCCCCATGATCATCCACGGAATCAAACCCAG
GGCGTAGGCAGAAGTTCAGCTCTCTGTACATCTCCAGTTTATCATCATGTATAGCCTGGACGGGAAGAAATGGCAG
ACCTACAGAGGCAATTCACCGGGACCTGATGGTCTTCTTTGGAACGTGGATTCAGCGGCATCAAGCACAAACATC
TTCAATCCACCCATCATCGCCGCTACATCCGGCTGCATCTTACCCACTATAGCATCAGGTCTACCTGAGAATGGAG
CTGATGGGATGCGACCTGAACAGCTGTTCTATGCCATGGGACATGGAGTCCAAGGCTATCAGCGATGCCAGATCACC
GCTTCTTCTACTTACCAATATGTTTGTACTGTTCCCAAGCAAGGCTAGACTGCACCTGCAGGGAAGATCCAAC
GCTTGGAGACCCAGGTGAACAATCCTAAGGAGTGGCTGCAGGTGCACTTCCAGAAAACCATGAAGGTACCGGGGTG
ACCACCGAGGAGTGAATCTCTGCTGACCTCCATGTACGTCAAGGAGTTCCTGATCAGCTCTTCCAGGACGGCCAC
CAGTGGACCTGTTCTTTTCAACGGCAAGGTCAAAGTGTTCAGGGGAATCAGGACTCTTTTACCCCGCTGCTGAAC
TCCCTGGATCCTCCACTGCTGACCAGGTACCTGAGAATCCATCCTCAGAGCTGGGTGCACCAGATCGCTCTGAGAATG
GAGGTCTGGGATGCGAAGCTCAGGACCTGTATTGA

Además, los motivos de ADN CpG en la secuencia codificante con codones de hígado optimizados para ET3 y HSQ se eliminaron porque pueden conducir a la metilación y silenciamiento de genes. Véase Bird, DNA methylation and the frequency of CpG in animal DNA, 1980, Nucleic Acids Res, 8: 1499-1504. Los codones se sustituyeron con la alternativa humana/hepática más utilizada (basada en el análisis de preferencia de uso en el hígado analizada

5 anteriormente) que no dio como resultado la formación de un dinucleótido 5'-CG-3 'en la secuencia. Estas modificaciones eliminaron 174 y 175 CpG de las secuencias de ET3 y HSQ con codones de hígado optimizados, respectivamente. La eliminación de CpG también ayuda al vector a evadir la detección inmunitaria, potenciando la seguridad y la eficacia del vector. Véase J Clin Invest. 2013, 123(7):2994-3001, titulado "CpG-depleted adeno-associated virus vectors evade immune detection".

La secuencia de ET3 con codones de hígado optimizados con CpG eliminados es la SEQ ID NO: 11:

```
ATGCAGCTGGAAGTGTCTACCTGTGTGTTTCTGTGTCTGCTGCCTCTGGGGTTTTCTGCTATCAGGAGATACTATCTG
GGAGCTGTGGAGCTGTCTGGGACTACAGGCAGTCTGAGCTGCTGAGAGAACTGCATGTGGATACCAGATTCCCAGCT
ACAGCTCCAGGAGCTCTGCCTCTGGGCCCATCTGTGCTGTACAAGAAAACAGTCTTTGTGGAGTTTACAGACCAGCTG
TTCTCTGTGGCCAGGCCAAGACCACCTTGGATGGGACTGCTGGGACCAACCATCCAGGCTGAGGTGTATGATACAGTG
GTGGTGACCCTGAAAAACATGGCCTCCCATCCTGTGAGCCTGCATGCTGTGGGGGTGTCTTCTGGAAGTCCTCTGAG
GGAGCTGAGTATGAAGACCATACCTCCCAGAGGGAGAAAGAAGATGATAAGGTGCTGCCTGGCAAAAGCCAGACCTAT
GTCTGGCAGGTGCTGAAGGAGAATGGACCAACTGCTTCTGACCCACCATGCCTGACCTACTCTTATCTGTCCCATGTG
```

GATCTGGTGAAGGACCTGAATTCTGGACTGATTGGAGCTCTGCTGGTGTGTAGAGAGGGAAGCCTGACCAGAGAAAGA
 ACCCAGAACCTGCATGAGTTTGTCTGCTGTTTGTCTGTTTGTATGAAGGGAAGAGCTGGCACTCTGCCAGGAATGAC
 TCCTGGACCAGAGCTATGGATCCAGCTCCTGCTAGAGCTCAGCCTGCTATGCACACAGTCAATGGCTATGTGAATAGG
 TCTCTGCCAGGACTGATTGGCTGCCATAAGAAATCTGTCTATTGGCATGTGATTGGAATGGGCACCAGCCCTGAGGTG
 CATTCTATCTTCTTGGGAAGGCCACACCTTTCTGGTCAGGCACCATAGACAGGCCTCTCTGGAGATCTCCCCTCTGACC
 TTCTTGACAGCTCAGACCTTTCTGATGGACCTGGGGCAGTTTCTGCTGTTTGGCCATATCTCTCCCACCATCATGGA
 GGAATGGAGGCTCATGTGAGGGTGAATCCTGTGCTGAGGAACCACAGCTGAGAAGAAAGGCTGATGAGGAAGAGGAC
 TATGATGATAACCTGTATGACTCTGATATGGATGTGGTGGAGGCTGGATGGGGATGATGTGAGCCCTTTTCATCCAGATC
 AGGTCTGTGGCCAAGAAACATCCAAAGACCTGGGTCCACTACATTGCTGCTGAAGAGGAAGATTGGGACTATGCCCC
 CTGGTGTGGCTCCTGATGATAGATCCTACAAAAGCCAGTATCTGAACAATGGGCCCCAGAGGATTGGAAGGAAGTAC
 AAGAAAGTGAGGTTCTATGGCCTATACAGATGAGACCTTTAAGACCAGAGAGGCTATCCAGCATGAATCTGGGATCCTG
 GGACCTCTGCTGTATGGAGAAGTGGGGGATACCCTGCTGATCATCTTCAAGAACCAGGCCCTCCAGGCCATACAATATC
 TATCCCCATGGCATCACAGATGTGAGACCACTGTACAGCAGGAGACTGCCCAAGGGGGTCAAACACCTGAAGGATTTT
 CCCATCCTGCCTGGAGAGATCTTTAAGTATAAATGGACAGTCACAGTGAAGATGGGCCCTACCAAGTCTGATCCAAGG
 TGCCTGACCAGATACTATAGCTCTTTTGTGAACATGGAGAGAGACCTGGCTTCTGGACTGATGGACCCCTGCTGATC
 TGTTACAAAGAGTCTGTGGACCAGAGGGGCAACCAGATCATGTCTGATAAGAGAAATGTCTCTCTGTTCTGTGTTT
 GATGAGAACAGGAGCTGGTACCTGACAGAGAACATCCAGAGGTTCTGCCAAATCCAGCTGGAGTGCAGCTGGAGGAC
 CCAGAATTTTCAGGCTTCCAACATCATGCATAGCATCAATGGCTATGTGTTTGATAGCCTGCAGCTGTCTGTCTGCCG
 CATGAGGTGGCCCTACTGGTATATCCTGTCCATTGGAGCTCAGACAGACTTCTGTCTGTGTTCTTTAGTGGGTACACC
 TTTAAGCATAAAATGGTGTATGAGGATACCCTGACCCTGTTCCCCCTTTCTGGGGAGACAGTGTTCATGTCCATGGAA
 AACCTTGGCCTGTGGATCCTGGGGTGCCCAACTCTGACTTCAGGAATAGAGGAATGACAGCCCTGCTGAAAGTGTCC
 AGCTGTGATAAGAATACAGGGGATTACTATGAGGACTCTTATGAAGATATCTCTGCTTATCTGCTGAGCAAGAACAAT
 GCCATTGAGCCCAGGTCTTTTGTCTCAGAACTCCAGACCTCCATCTGCTTCTGCTCCTAAGCCACCTGTGCTGAGAAGA
 CATCAGAGGGACATCTCCCTGCCTACCTTCCAGCCAGAGGAAGATAAAATGGACTATGATGATATCTTCAGCACAGAG
 ACCAAGGGGGAAGATTTTGACATCTATGGAGAGGATGAAAACAGGATCCAAGATCCTTCCAGAAGAGAACCAGACAC
 TACTTTATTGCTGCTGTGGAGCAGCTGTGGGACTATGGGATGTCTGAAAGCCCAAGGGCCCTGAGGAACAGAGCTCAG
 AATGGAGAGGTGCCAGATTCAAGAAAGTGGTGTTCAGAGAGTTTGTCTGATGGCAGCTTTACCCAGCCATCTTACAGG
 GGGGAGCTGAACAAGCATCTGGGGCTGCTGGGACCTATATCAGAGCTGAGGTGGAAGATAACATCATGGTGACCTTC
 AAGAATCAGGCTTCTAGGCCCTACTCCTTTTATCTTCCCTGATCTCCTACCCTGATGATCAGGAGCAGGGAGCTGAA
 CCTAGGCACAACCTTTGTGAGCCAAATGAGACCAGAACCTACTTTTGAAGGTGCAGCATCATGAGCTCCACAGAG
 GATGAATTTGACTGCAAAGCTTGGGCCCTATTTTCTGATGTGGACCTGGAGAAGGATGTGCATTCTGGCCTGATTGGG
 CCTCTGCTGATCTGTAGGGCCAACACCTGAATGCTGCTCATGGAAGACAGGTACAGTGCAGGAGTTTGTCTGTCTG
 TTTACCATCTTTGATGAAACCAAGAGCTGGTACTTCACAGAGAATGTGGAAGGAATTCAGAGCCCCCTGTCTATCTG
 CAGATGGAGGACCCCTACCCTGAAGGAAAACCTACAGGTTCCATGCCATCAATGGATATGTCTGATATACCTGCCTGGC
 CTGGTCTATGGCTCAGAACCCAGAGGATCAGATGGTACCTGCTGTCTATGGGATCCAATGAGAATATCCATAGCATCCAC
 TTCTCTGGCCATGTCTTTTCTGTGAGGAAGAAAGAGGAATACAAATGGCTGTGTACAATCTGTATCCTGGGGTCTTT
 GAGACAGTGGAAATGCTGCCAAGCAAAGTGGGAATCTGGAGAATTGAGTGCCTGATTGGGGAACACCTGCAGGCTGGG
 ATGAGCACCACCTTCTGGTGTACTCTAAGAAATGTGAGACCCCACTGGGGATGGCCTCTGGACATATCAGGGACTTC
 CAGATCACAGCTTCTGGACAGTATGGACAGTGGGCTCCAAAGCTGGCTAGACTGCACTATTCTGGCTCCATCAATGCC
 TGGTCTACCAAGAGCCATTCTCCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCATGGAATCAAACCCAG
 GGAGCTAGGAGAGAAATTCAGCTCTCTGTACATCTCCAGTTTATCATCATGTATAGCCTGGAGGAAGAAATGGCAG
 ACCTACAGAGGCAATTCAGCTGGGACCCCTGATGGTCTTCTTTGGAAATGTGGATTCTCTGGCATCAAGCACAAATC
 TTCAATCCACCCATCATTGCCAGGTACATCAGGCTGCATCCTACCCACTATAGCATCAGGTCTACCCTGAGAATGGAG
 CTGATGGGATGTGACCTGAACAGCTGTTCTATGCCACTGGGCATGGAGTCCAAGGCTATCTCTGATGCCAGATCACA
 GCTTCTTCTTACTTACCAATATGTTTGTACCTGGTCCCCAAGCAAGGCTAGACTGCACCTGCAGGGAAGATCCAAT
 GCTTGGAGACCCAGGTGAACAATCCTAAGGAGTGGCTGCAGGTGGACTTCCAGAAAACCATGAAGGTACAGGGGTG
 ACCACCCAGGGAGTGAATCTCTGCTGACCTCCATGTATGTCAAGGAGTTCTCTGATCAGCTCTTCCAGGATGGCCAC
 CAGTGGACCCCTGTTCTTTCAGAATGGCAAGGTCAAAGTGTCCAGGGGAATCAGGACTCTTTTACCCAGTGGTGAAC
 TCCCTGGATCCTCAGCTGCTGACCAGGTACCTGAGAATCCATCCTCAGAGCTGGGTGCACCAGATTGCTCTGAGAATG
 GAGGTCTGGGATGTGAAGCTCAGGACCTGTATTGA.

La secuencia de HSQ con codones de hígado optimizados con CpG eliminados es la SEQ ID NO: 2:

ATGCAGATTGAACTGTCTACCTGTTTCTTTCTGTGCCTGCTGAGGTTTTGTTTTTCTGCTACCAGAAGATACTACCTG
 GGAGCTGTGGAAGTGAAGCTGGGATTACATGCAGCTGACCTGGGAGAGAGCTGCCTGTGGATGCTAGATTCCACCTAGA
 GTCCCTAAGTCTTTCCCTTCAACACCTCTGTGGTCTACAAGAAAACCTGTTTGTGGAGTTTACAGACCACCTGTTT
 AACATTGCTAAGCCTAGACCACCATGGATGGGACTGCTGGGACCAACCATCCAGGCAGAGGTGTATGACACAGTGGTC
 ATCACCTGAAAAACATGGCTTCTCACCTGTGTCCCTGCATGCTGTGGGAGTCTCCTACTGGAAGGCCTCTGAAGGG
 GCTGAGTATGATGATCAGACCAGCCAGAGGAAAAAGAGGATGATAAGGTGTTCCCTGGAGGGTCCCATACCTATGTG
 TGGCAGGTCTGAAGGAGAATGGACCAATGGCTTCTGACCTCTGTGCCTGACCTACTCTATCTGTCCCATGTGGAC
 CTGGTCAAGGATCTGAACTCTGGCCTGATTGGGGCTCTGCTGGTGTGTAGGGAAGGGTCCCTGGCCAAGGAGAAAAAC
 CAGACCCTGCATAAGTTCATCCTGCTGTTTGTGTGTTTGTATGAAGGAAAAAGCTGGCACTCTGAGACCAAGAACTCT
 CTGATGCAGGACAGGATGCTGCTTCTGCCAGAGCTTGGCCCAAGATGCACACAGTGAATGGCTATGTCAATAGGAGC

CTGCCCTGGACTGATTGGCTGCCACAGAAAGTCTGTGTATTGGCATGTTCATTGGAATGGGCACCACCCCTGAAGTGCAC
AGCATCTTCTCTGGAGGGGCATACCTTTCTGGTCAGGAACACAGGCAGGCTAGCCTGGAGATCTCTCCAATCACCTTC
CTGACAGCCCAGACCCTGCTGATGGACCTGGGACAGTTCTGTCTGTTTTGCCACATCTCCAGCCACCAGCATGATGGC
ATGGAGGCTTATGTGAAAGTGGACTCCTGTCTGAGGAACCTCAGCTGAGGATGAAGAACAATGAGGAAGCTGAAGAC
TATGATGATGACCTGACAGACTCTGAGATGGATGTGGTCAGGTTTGATGATGATAACTCTCCCTCCTTTATCCAGATC
AGGTCTGTGGCCAAGAAACACCCTAAGACCTGGGTCCATTACATTGCTGCTGAGGAAGAGGACTGGGATTATGCTCCA
CTGGTGTGGCCCTGATGATAGATCCTACAAAAGCCAGTATCTGAACAATGGACCCAGAGGATTGGCAGAAAGTAC
AAGAAAGTGAAGTTTCATGGCTTATACAGATGAGACCTTTAAGACCAGAGAAGCCATCCAGCATGAGTCTGGGATCCTG
GGACCTCTGCTGTATGGGGAAGTGGGGGACACCCTGCTGATCATCTTCAAGAACAGGCCAGCAGGCCTTACAATATC
TATCCACATGGCATCACAGATGTGAGACCTCTGTACTCCAGGAGGCTGCCAAAGGGGGTGAACACCTGAAGGACTTC
CCAATCTGCCTGGGGAATCTTTAAGTATAAATGGACAGTCACAGTGGAGGATGGGCCCACCAAGTCTGACCCTAGG
TGCCTGACCAGATACTATTCTTCTCTTTGTGAATATGGAGAGAGACCTGGCTTCTGGACTGATTGGACCCCTGCTGATC
TGTTACAAAGAGTCTGTGGATCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAATGTGATCCTGTCTCTGTCTTT
GATGAAACAGGTCTTGGTACCTGACAGAGAATCCAGAGGTTCTGCCTAATCCAGCTGGAGTGCAGCTGGAAGAT
CCTGAGTTCAGGCCTCTAACATCATGCATTCATCAATGGCTATGTGTTTGACTCCCTGCAGCTGTCTGTGTGCCTG
CATGAGGTGGCTTACTGGTATATCCTGAGCATTGGAGCCAGACAGATTTCCTGTCTGTGTTCTTTCTGGCTACACC
TTTAAGCATAAAATGGTGTATGAGGACACCCTGACCCCTGTTCCATTTTCTGGAGAACTGTGTTTCATGAGCATGGAG
AATCCTGGGCTGTGGATCCTGGGATGCCACAACCTCTGATTTTCAGGAATAGAGGGATGACAGCCCTGCTGAAAGTGAAGC
TCTTGTGACAAGAACACAGGAGACTACTATGAAGATAGCTATGAGGACATCTCTGCTTATCTGCTGTCCAAAAACAAT
GCCATTGAGCCCAGGAGCTTCTCTCAGAACCCTCCAGTCTGAAGAGGCACCAGAGGGAGATCACCAGAACCACCCCTG
CAGTCTGATCAGGAAGAGATTGACTATGATGATACCATCTCTGTGGAATGAAGAAAGAGGACTTTGATATCTATGAT
GAAGATGAGAACCAGTCTCCAGGTCTTCCAGAAGAAAACCAGACATTACTTTATTGCTGCTGTGGAGAGGCTGTGG
GACTATGGCATGTCCAGCTCTCCTCATGTGCTGAGAAATAGAGCTCAGTCTGGATCTGTCCACAGTTCAAGAAAGT
GTCTTCCAGGAGTTTACAGATGGAAGCTTTACCCAGCCACTGTACAGGGGAGAACTGAATGAGCACCTGGGGCTGCTG
GGACCCATATCAGGGCTGAAGTGGAGGATAACATCATGGTACCTTCAGGAATCAGGCCAGCAGACCCTACTCTTTT
TATTCCAGCCTGATCTCCTATGAAGAGGACCAGAGACAGGAGCTGAACCAAGAAAAAATTTGTGAAGCCTAATGAG
ACCAAAACCTACTTTTGGAGGTGCAGCACCATATGGCCCTACCAAGATGAGTTTGATTGCAAGGCCTGGGCTTAT
TTTTCTGATGTGGATCTGGAGAGGATGTCCACTCTGGCCTGATTGGGCCACTGCTGGTGTGTCTATCCAACACCCTG
AATCCAGCTCATGGAAGGCAGGTGACAGTCCAGGAATTTGCCCTGTTCTTTACCATCTTTGATGAGACCAAGAGCTGG
TACTTCACAGAAAACATGGAGAGGAATTGCAGAGCCCCATGTAACATCCAGATGGAAGACCCACCTTCAAGGAGAAC
TACAGATTTTCATGCTATCAATGGGTATATCATGGATACCCCTGCCAGGACTGGTTCATGGCTCAGGACCAGAGGATCAGA
TGGTACCTGCTTGGCATGGGGTCTAATGAGAATATCCACTCCATCCATTTCTCTGGACATGTGTTTACAGTAAGGAAG
AAAGAAGAGTACAAGATGGCCCTGTACAACCTGTATCCTGGGGTGTGTTGAAACAGTGGAGATGCTGCCTTCCAAGGCT
GGGATCTGGAGGGTGAATGCCTGATTGGGGAGCACCTGCATGCTGGAATGTCTACCCTGTTCTGGTGTACTCCAAT
AAGTGTGACACCCCTGGGGATGGCTTCTGGACATATCAGGGACTTCCAGATCACAGCTTCTGGACAGTATGGACAG
TGGGCTCCTAAGCTGGCTAGACTGCACATATTCTGGCTCCATCAATGCTTGGTCTACCAAGAGCCCTTCTCCTGGATC
AAGGTGGACCTGCTGGCTCCAATGATCATCCATGGCATCAAAACCCAGGGGGCCAGGCAGAAAGTCTCTTCCCTGTAC
ATCAGCCAGTTTATCATCATGTATTCTCTGGATGGGAAGAAATGGCAGACCTACAGAGGCAATTCACAGGGACCCCTG
ATGGTGTCTTGGCATGGGGTCTGGGATCAGGCACAACATCTCAATCCCCCTATGCTTTCAGGTACATC
AGACTGCACCCAACCCATTATTCCATCAGGAGCACCCCTGAGAATGGAGCTGATGGGGTGTGATCTGAACAGCTGTTCT
ATGCCCCCTGGGAATGGAGTCTAAGGCCATCTCTGATGCTCAGATCACAGCCTCCAGCTACTTACCAATATGTTTGCT
ACCTGGTCCCCAAGCAAGGCTAGACTGCATCTGCAGGGAAGAAGCAATGCTTGGAGACCACAGGTGAACAATCCCAAG
GAGTGGCTGCAGGTGGACTTCCAGAAAACCATGAAGGTGACAGGAGTACCACCCAGGGAGTGAAGAGCCTGCTGACC
TCTATGTATGTCAAGGAGTTCTGATCTCTTCCAGCCAGGATGGGCACCAGTGGACCCCTGTTCTTTTCAGAATGGAAG
GTGAAAGTCTTCCAGGGCAATCAGGATTCTTTACCCCTGTGGTCAACAGCCTGGACCCACCCCTGCTGACCAGGTAC
CTGAGAATCCACCCACAGTCTGGGTGCATCAGATTGCTCTGAGGATGGAAGTCTGGGCTGTGAGGCCAGGACCTG
TATTGA

- 5 Este enfoque aumentó el LCAI de 0,62 a 0,86 para fVIII humano (HSQ) con dominio B eliminado. La expresión *in vitro* de las secuencias optimizadas de fVIII se evaluó en células HepG2 transfectadas transitoriamente con los plásmidos de expresión de fVIII correspondientes. La optimización de codones en la secuencia del ácido nucleico que codifica para HSQ aumentó la expresión de HSQ 7,4 veces, y se expresó de manera tan eficaz como ET3 sin codones optimizados (Figura 3A). La optimización de codones en la secuencia del ácido nucleico que codifica para HSQ aumentó la expresión de ET3 5,6 veces. Cuando se probó *in vivo* mediante inyección hidrodinámica de plásmido que
- 10 portaba las variantes de fVIII con codones optimizados en ratones, la actividad de HSQ aumentó a niveles que podrían detectarse mediante el ensayo que eran casi equivalentes a los niveles de fVIII observados en la ET3 sin codones optimizados (Figura 3B). Adicionalmente, se observó un aumento de 17 veces en la actividad para la ET3 con codones optimizados (Figura 3B).
- 15 La eficacia de la traducción a veces también puede mejorarse mediante la optimización del contenido de GC, estructura secundaria de ARNm, sitios de PoliA prematuros, motivo de inestabilidad de ARN, energía libre estable de ARNm, sitios chi internos, motivos de inestabilidad de ARN, sitios de unión a ribosomas, sitios de corte y empalme crípticos, islas CpG negativas, secuencia SD, cajas TATA y señales terminales crípticas, etc. La FIG. 4 ilustra los cambios que se realizaron en un casete AAV que codifica la secuencia de ET3 con codones optimizados y CpG eliminados para

aumentar su producción. Varios elementos de acción en cis dentro de la secuencia de HSQ se modificaron adicionalmente para aumentar la expresión de fVIII, tal como se resume en la siguiente tabla.

Elementos de Acción en CIS	Optimizado	Original
Corte y empalme(GGTAAG)	0	2
Corte y empalme (GGTGAT)	0	1
PoliA(AATAAA) (4789 - AATAAA)	1	1
PoliA(ATTAATA)	0	0
Desestabilizante(ATTTA)	0	6
PoliT(TTTTTT) (4827 - TTTTTT)	1	2
PoliA(AAAAAA)	0	1

- 5 Además de las secuencias con codones de hígado optimizados proporcionadas anteriormente, la secuencia codificante para las proteínas ET3 y HSQ fVIII se optimizó con codones para la expresión en células mieloides. Usando las secuencias de genes altamente expresados en las células mieloides humanas, se construyó una tabla personalizada de preferencia de uso de codones específica para el hígado humano (véase la Figura 2C) que difiere sustancialmente de los codones más prevalentes usados en secuencias codificantes humanas totales (Figura. 2B). La
- 10 secuencia codificante optimizada de las variantes de fVIII ET3 y HSQ se desarrolló con los codones estadísticamente más prevalentes identificados por el análisis de preferencia de uso en el hígado. Además, los motivos de ADN CpG en la secuencia codificante con codones mieloides optimizados para ET3 y HSQ se eliminaron porque pueden conducir a la metilación y silenciamiento de genes.
- 15 La secuencia de ET3 con codones mieloides optimizados con CpG eliminados es la SEQ ID NO: 125:

ATGCAGCTGGAGCTCTCAACCTGTGTGTTCTCTGCCTGCTCCCCCTGGGATTTTCAGCTATCAGGAGATAC
TATCTGGGAGCAGTGGAACTGTCTGGGACTACAGGCAGTCAGAGCTGCTCAGAGAACTGCATGTGGATACTAGGTTT
CCTGCAACAGCTCCTGGAGCACTGCCACTGGGACCTTCAGTGCTGTACAAGAAAAGTCTTTGTGGAGTTTACAGAC
CAGCTGTTTCAGTGTGGCCAGGCCCCAGGCCCCCTGGATGGGGCTGCTGGGACCCACCATCCAGGCTGAAGTGATGAT
ACTGTGGTGGTGACCTGAAAAACATGGCCTCTCATCCAGTCAGCCTGCATGCTGTGGGAGTGAGCTTCTGGAAGAGC
AGTGAGGGAGCTGAGTATGAAGACCATACCTCAGAGGGAGAAAGATGATAAGGTGCTGCCAGGAAAAAGCCAG
ACCTATGTGTGGCAGGTGCTGAAGGAGAATGGCCCTACAGCTTCAGATCCTCCCTGCCTCACATACTCTTATCTGAGC
CATGTGGATCTGGTGAAGGACCTCAATAGTGGCCTGATTGGGGCACTGCTGGTGTGCAGAGAGGGGTCCCTCACAAAGG
GAAAGAACTCAGAACCTGCATGAGTTTGTCTGCTCTTTGCTGTGTTTGATGAGGGAAAGTCTGGCACTCAGCAAGG
AATGACAGCTGGACCAGGCTATGGACCCAGCAGCCAGAGCTCAGCCAGCTATGCACACTGTCAATGGCTATGTG
AATAGGTCCCTGCCTGGACTCATTGGCTGCCATAAGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCAGCCCCA
GAGGTGCATTCCATCTTCTGGAAGGCCACACATTTCTGGTCAGGCACCATAGACAGGCCAGCCTGGAGATCAGCCCCA
CTGACTTTCTCCTCACAGCACAGACATTTCTGATGGACCTGGGGCAGTTCTTGCTCTTTTGCCATATCTCAAGTACCAT
CATGGAGGGATGGAGGCTCATGTGAGGGTGGAAAGCTGTGCAGAGGAACCTCAGCTGAGGAGGAAGGCAGATGAGGAA
GAGGACTATGATGATAACCTGTATGACTCAGATATGGATGTGGTGGAGGCTGGATGGAGATGATGTGAGCCCATTCATC
CAGATCAGGTGAGTGGCTAAGAAACACCTTAAGACCTGGGTCCACTACATTGCAGCTGAAGAGGAAGATTGGGACTAT
GCACCCCTGGTGCTGGCCCCAGATGATAGAAGTTACAAATCTCAGTATCTGAACAATGGGCCCCAGAGGATTGGAAGG
AAGTACAAGAAAGTGAAGTTCATGGCTTATACTGATGAGACCTTTAAGACAAGAGAGGCAATCCAGCATGAAAGTGGC
ATCCTGGGACCACTGCTCTATGGAGAAGTGGGGGATACCCTGCTCATCATCTTCAAGAACCAGGCCTCAAGGCCTTAC
AATATCTATCCCCATGGCATCACAGATGTGAGGCCTCTCTACAGCAGGAGACTGCCCAAGGGAGTCAAACACCTCAAG
GATTTCCCCATCCTGCCAGGGGAAATCTTCAAGTATAAATGGACAGTCACTGTGGAAGATGGGCCAACTAAGTCAGAT
CCTAGGTGCCTGACCAGGTACTATTCTAGCTTTGTGAACATGGAGAGGGACCTGGCTTCAGGACTGATTGGACCTCTG
CTCATCTGCTACAAAGAATCAGTGGACCAGAGGGGCAACCAGATCATGAGTGATAAGAGAAATGTCATCCTGTTCTCA
GTGTTTGTGATGAGAATAGGAGTTGGTATCTGACAGAAAACATCCAGAGGTTCTGCTAATCCTGCAGGAGTGCAGCTG
GAGGACCCAGAAATTCAGGCTTCAAACATCATGCATAGTATCAATGGCTATGTGTTTGATAGTCTGCAGCTCTCTGTC
TGCTGTCATGAGGTGGCCTACTGGTATATCCTCAGCATTGGAGCTCAGACTGACTTCTGAGTGTGTTCTTTTCAGGC

TACACATTCAAGCATAAGATGGTCTATGAAGATACCCTGACACTCTTCCCCTTTTCTGGGGAGACTGTGTTTATGAGC
 ATGGAAAAACCCAGGCCCTGTGGATTCTGGGGTGCCACAACAGTGACTTCAGGAATAGAGGGATGACTGCTCTGCTCAAA
 GTGTCTCATGTGATAAGAATACTGGAGATTACTATGAGGACTCTTATGAAGATATCAGTGCATATCTGCTCTCCAAA
 AACAAATGCCATTGAGCCCAGGTCAATTTGCTCAGAACAGTAGACCACCTTCTGCAAGTGCACCAAAGCCTCCAGTGCTG
 AGGAGACACCAGAGGGACATCAGCCTGCCAACCTTCCAGCCTGAGGAAGATAAAATGGACTATGATGATATCTTCTCC
 ACTGAGACCAAGGGGAAGATTTTGACATCTATGGAGAGGATGAAAACAGGACCCCAGGTCCCTTCCAGAAGAGGACC
 AGACACTACTTTATTGCAGCTGTGGAGCAGCTGTGGGACTATGGCATGTCTGAATCACCTAGAGCTCTGAGGAACAGA
 GCACAGAATGGGGAGGTGCCAGGTTCAAGAAAGTGGTGTTCAGAGAATTTGCAGATGGCTCTTTTACCCAGCCTAGC
 TACAGGGGGGAGCTCAACAAGCATCTGGGGCTGCTGGGACCCCTATATCAGAGCAGAGGTGGAAGATAACATCATGGTG
 ACATTCAAGAATCAGGCCCTCAAGACCCTACAGTTTTTATAGTTCTCTGATCAGCTACCCAGATGATCAGGAGCAGGGG
 GCTGAACCAAGGCACAACCTTTGTGCAGCCTAATGAGACAAGAACCTTACTTTTGGAAGGTCCAGCATCACATGGCTCCC
 ACAGAGGATGAGTTTGACTGCAAGGCCTGGGCATATTTTCTGATGTGGACCTGGAGAAGGATGTGCATAGTGGCCTC
 ATTTGGCCACTGCTCATCTGCAGGGCAAACACACTGAATGCTGCACATGGCAGGCAGGTCACTGTGCAGGAGTTTGCC
 CTGTTCTTTACAATCTTTGATGAAACTAAGTCTGGTACTTCACAGAGAATGTGGAAGGAATTGCAGAGCCCCCTGC
 CATCTCCAGATGGAGGACCAACTCTGAAGGAAAACCTACAGGTTCCATGCTATCAATGGATATGTCAATGGATACACTG
 CCAGGCCTGGTGATGGGCACAGAACCCAGAGGATCAGGTGGTATCTGCTCAGCATGGGGTCCAATGAGAATCCATTTCT
 ATCCACTTCTCAGGACATGTCTTTTTCAGTGAGGAAGAAAGAGGAATATAAAATGGCTGTGTACAATCTGTATCCAGGG
 GTCTTTGAGACAGTGGAATGCTGCCTAGCAAAGTGGGGATCTGGAGAATTGAGTGCCTCATTGGAGAACACCTGCAG
 GCAGGGATGTCCACCACATTTCTGGTGTACTCAAAGAAATGCCAGACTCCCCTGGGGATGGCAAGTGGACATATCAGG
 GACTTCAGATCACTGCATCAGGACAGTATGGACAGTGGGCACCAAAGCTGGCTAGGCTCCACTATAGTGGCTCTATC
 AATGCTTGGAGTACCAAAGAGCCTTTCTCTTGGATCAAGGTGGATCTGCTGGCCCCCATGATCATCCATGGAATCAAA
 ACACAGGGAGCTAGACAGAAGTTCAGCTCCCTGTACATCAGTCAGTTTATCATCATGTATTCTCTGGATGGGAAGAAA
 TGGCAGACCTACAGGGGCAATAGCACTGGGACACTGATGGTCTTCTTTGGAATGTGGATTCAAGTGGCATCAAGCAC
 AACATCTTCAATCTCCCATCATTGCCAGGTACATCAGATGCATGCCACACACTATTCAATCAGGAGTACTCTCAGA
 ATGGAGCTGATGGGGTGTGACCTCAACAGCTGCTCCATGCCACTGGGAATGGAATCCAAGGCAATCTCAGATGCCCAG
 ATCACTGCTTCTAGCTACTTCAACCAATATGTTTGCAACATGGTCACCCAGTAAAGCAAGGCTGCACCTCCAGGGAAGG
 TCCAATGCTTGGAGACCCAGGTGAACAATCCAAAGGAGTGGCTGCAGGTGGACTTTTCAAGAAACCATGAAGGTACA
 GGGGTGACTACCCAGGGAGTGAAAAGTCTGCTCACTCTATGTATGTCAAGGAGTTCCTGATCTCCTCAAGTCAGGAT
 GGCCACCAGTGGACACTGTTCTTTTCAAGATGGCAAGGTCAAAGTGTTCAGGGGAATCAGGACAGCTTTACACCAGTG
 GTGAACAGCCTGGACCCCCCTCTGCTCACTAGATATCTGAGAATCCATCCACAGAGCTGGGTGCACCAGATTGCACTC
 AGAATGGAGGTCTGGGCTGTGAAGCCCAGGACCTGTATTGA

La secuencia de HSQ con codones mieloides optimizados con CpG eliminados es la SEQ ID NO: 126:

ATGCAGATTGAGCTCAGCACCTGCTTCTTTCTGTGCCTGCTCAGGTTTCTGCTTTTTCAGCCACAAGGAGATACTATCTG
 GGAGCTGTGGAAGTGTCTATGGGATTACATGCAGAGTGACCTGGGAGAGCTCCCTGTGGATGCTAGGTTCCCCCAAGG
 GTCCCAAAGTCTTTCCCTTTTAAATACAGTGTGGTCTATAAGAAAACACTCTTTGTGGAATTTACTGATCACCTGTTC
 AACATTGCAAAGCCCAAGGCCTCCCTGGATGGGACTGCTGGGACCTACCATCCAGGCTGAGGTGTATGACACTGTGGTC
 ATCACACTGAAAAACATGGCATCTCACCCCTGTCAGCCTGCATGCAGTGGGAGTCAGCTACTGGAAGGCTTCAGAAGGG
 GCAGAGTATGATGATCAGACAAGCCAGAGAGAAAAAGAGGATGATAAGGTGTTCCAGGAGGGAGCCATACTTATGTG
 TGGCAGGTCTGAAGGAGAATGGCCCAATGGCCAGTGACCCACTGTGCCTCACCTACTCATATCTGAGTCATGTGGAC
 CTGGTCAAGGATCTCAACTCAGGCCTGATTGGGGCACTGCTGGTGTGCAGGGAAGGCTCACTGGCCAAGGAGAAAACC
 CAGACACTGCATAAGTTTATCTCTGCTCTTTGCTGTGTTGATGAAGGGAAATCTTGGCACAGTGAGACCAAGAACAGT
 CTGATGCAGGACAGGGATGCTGCTTCTGCCAGAGCTTGGCCCAAGATGCACACAGTGAATGGATATGTCAATAGGTCC
 CTGCCAGGACTCATTGGCTGCCACAGAAAGTCACTGTATTGGCATGTCAATTGGAATGGGGACCAACACAGAAAGTGCAC
 AGCATCTTCTGGAGGGGCATACCTTTCTGGTCAAGAACACAGGCAGGCCAGCCTGGAGATCAGCCCAATCACCTTTC
 CTGACAGCCCAGACTCTGCTCATGGATCTGGGGCAGTCTCTGCTCTTTTGCCACATCAGCTCCCACCAGCATGATGGA
 ATGGAGGCATATGTGAAAGTGGACTCCTGCCCAGAGGAACACAGCTGAGGATGAAGAACAATGAGGAAGCTGAAGAC
 TATGATGATGACCTGACAGACTCAGAGATGGATGTGGTCAGGTTTATGATGATAACAGCCCTCCTTTTATCCAGATC
 AGAAGTGTGGCCAAGAAACACCCAAAGACATGGGTCCATTACATTGCAGCTGAGGAAGAGGACTGGGATTATGCACCT
 CTGGTGTGGCCCCAGATGATAGATCCTACAAATCACAGTATCTGAACAATGGACCCAGAGGATGGCAGAAAGTAC
 AAGAAAGTGAGGTTTATGGCCTATCTGATGAAACATTTAAGACTAGAGAAGCTATCCAGCATGAGTCAGGCATCCTTG
 GGACCACTGCTCTATGGAGAAGTGGGGACACCCCTGCTCATCATCTTCAAGAACCAGGCTTCCAGGCCATACAAATATC
 TATCCTCATGGCATCACAGATGTGAGACCACTTACTCAAGGAGACTGCCTAAGGGAGTCAAACACCTCAAGGACTTC
 CCTATCCTGCCAGGGGAAATCTTTAAGTATAAATGGACTGTGACAGTGGAGGATGGGCCCCTAAGAGTGACCAAGG
 TGCCTGACCAGATACTATTCAAGTTTGTGAATATGGAAGGGATCTGGCATCAGGACTGATGGACCTCTGCTCATC
 TGCTACAAAGAGAGTGTGGATCAGAGGGGCAACCAGATCATGTGCAGACAAGAGGAATGTGATCCTGTTCAAGTGTCTTT
 GATGAAAAACAGTCTTGGTATCTGCAGAGAACATCCAGAGATTCTGCCAAATCCTGCAGGGGTGCAGCTGGAAGAT
 CCAGAGTTTTCAGGCCTCAAACATCATGATATCAATGGATATGTGTTGACAGTCTGCAGCTCTCTGTGTGCTGCTG
 CATGAAGTGGCCTACTGGTATATCCTGTCCATTGGAGCTCAGACAGATTTCCTGAGTGTGTCTTTTTCAGGCTACACT
 TTTAAGCATAAAATGGTCTATGAGGACACACTGACTCTCTTCCCTTTTGTGGGAAACAGTGTATGAGCATGGAG
 AATCCAGGGCTGTGGATTCTGGGATGCCACAACAGTGATTTCAAGGAATAGAGGCATGACTGCTCTGCTCAAAGTGTCT
 AGCTGTGACAAGAACACAGGGGACTACTATGAAGATTCTTATGAGGACATCAGTGCTTATCTGCTCTCAAAAAACAAT

GCAATTGAACCCAGATCATTCAGTCAGAATCCACCTGTGCTGAAGAGGCACCAGAGAGAGATCACTAGGACTACCCTG
 CAGTCAGATCAGGAAGAGATTGACTATGATGATACCATCTCAGTGGAAATGAAGAAAAGAGGACTTTGATATCTATGAT
 GAAGATGAGAACCAGAGTCCAAGGTCTTTCCAGAAGAAAACCAGACATTACTTTATTGCTGCAGTGGAGAGGCTGTGG
 GATTATGGAATGTCTCAAGTCCACATGTGCTGAGGAATAGGGCAGAGTCTGGCAGTGTCCCTCAGTTCAAGAAAGTG
 GTCTTCCAGGAGTTTACAGATGGCAGCTTCACTCAGCCTCTGTACAGGGGAGAACTCAATGAGCACCTGGGGCTGCTG
 GGACCTATATCAGAGCTGAAGTGGAGGATAACATCATGGTCACCTTCAGGAATCAGGCTTCAAGACCCTACAGTTTT
 TATTCTAGCCTGATCAGCTATGAAGAGGACCAGAGGCAGGGAGCTGAACCTAGGAAAACTTTGTGAAGCCAAATGAG
 ACCAAAACATACTTTTGGAGGTCCAGCACCATGGCACCAACCAAGATGAGTTTGATTGCAAGGCATGGGCTAT
 TTTTCAGATGTGGATCTGGAGAAGGATGTCCACAGTGGCCTCATTGGGCTCTGCTGGTGTGCCATACTAACACCTG
 AATCCAGCTCATGGCAGGCAGGTGACAGTCCAGGAGTTTGACTGTTCTTTACCATCTTTGATGAGACAAAGTCTCTGG
 TACTTCACTGAAAACATGGAGAGGAATTGCAGAGCTCCTTGCAACATCCAGATGGAAGACCCACCTTCAAGGAGAAC
 TACAGATTTTCATGCAATCAATGGGTATATCATGGATACACTGCCAGGACTGGTGATGGCCAGGACAGAGGATCAGA
 TGGTATCTGCTCAGCATGGGTCCAATGAGAATATCCACTCTATCCATTTCAAGTGGACATGTGTTTACAGTCAGAAAAG
 AAAGAAGAGTATAAAATGGCCCTGTACAACCTCTATCCAGGAGTGTGTTGAAACAGTGGAGATGCTGCCAAGCAAGGCT
 GGGATCTGGAGGGTGGAAATGCCTCATTGGGGAGCACCTGCATGCAGGAATGTCAACCTGTTTCTGGTCTACAGTAAT
 AAGTGCCAGACACCTCTGGGAATGGCAAGTGGACATATCAGGGATTTCAGATCACTGCTAGTGGACAGTATGGACAG
 TGGGCACCAAGCTGGCTAGACTCCACTATTAGGCTCAATCAATGCTTGGTCCACCAAGAGCCATTCTCATGGATC
 AAGGTGGACCTGCTGGCTCCTATGATCATCCATGGCATCAAAACACAGGGGGCAAGGCAGAAAGTTCTCCTCACTGTAC
 ATCTCTCAGTTTATCATCATGTATAGCCTGGATGCCAAGAAATGCCACACCTACAGGGCAATAGCACAGGGACTCTG
 ATGGTGTCTTTTGGCAATGTGGACAGCAGTGGGATCAAGCACAACATCTTCAATCCCCCAATCATTGCAAGGTACATC
 AGACTGCACCCACCCATTATTCAATCAGGAGTACACTCAGGATGGAAGTATGGGGTGTGATCTCAACAGTTGCTCT
 ATGCCACTGGGAATGGAGTCCAAGGCAATCTCAGATGCCAGATCACTGCTAGCTCCTACTTCACTAATATGTTTGCT
 ACCTGGAGCCCCCTCCAAAGCAAGGCTGCACCTCCAGGGAAGGAGCAATGCATGGAGGCCTCAGGTGAACAATCCCAAG
 GAATGGCTGCAGGTGGATTTCAGAAAACATGAAGGTGACTGGAGTCACAACTCAGGGAGTGAAAAGTCTGCTCACT
 TCTATGTATGTCAAGGAGTTCTGATCTCAAGTTCTCAGGATGGCCACCAGTGGACCTGTTCTTTTCAAGATGAAAAG
 GTGAAAGTCTTCCAGGGCAATCAGGATTCCTTTACACCAGTGGTCAACTCACTGGACCTCCCTGCTCACTAGATAT
 CTGAGAATCCACCCTCAGAGCTGGGTGCATCAGATTGCTCTCAGAATGGAAGTCTGGGCTGTGAGGCACAGGACCTG
 TATTGA

La expresión *in vitro* de las secuencias de fVIII no optimizadas, optimizadas para hígado y optimizadas para mieloides se evaluó en células HepG2 transfectadas transitoriamente con los plásmidos de expresión de fVIII correspondientes (FIG.5). La optimización específica del tejido conduce a una mayor actividad de FVIII en las células HepG2 (hepáticas) que expresan el fVIII optimizado para el hígado en comparación con las formas no optimizadas u optimizadas para mieloides de ET3 o HSQ. Estos resultados muestran que la optimización para el hígado beneficia específicamente la expresión en células derivadas de hepatocitos y que la expresión de FVIII optimizado para mieloides no beneficia la expresión en células HepG2.

Adicionalmente, cuando estos experimentos se repitieron utilizando células no humanas (células de riñón de hámster recién nacido), se descubrió que la optimización de la secuencia específica de seres humanos condujo a una disminución de la expresión en las células no humanas (FIG. 6).

Factor IX

Similar a la optimización con codones para el fVIII de coagulación analizada anteriormente, las secuencias de fIX también fueron optimizadas con codón de expresión en hepatocitos usando las mismas tablas de optimización de codones usadas para el factor de coagulación fVIII. Las secuencias de fIX seleccionadas para la optimización incluyen fIX con la mutación protrombogénica "Padua" R338L ("fIX Padua" véase Paolo et al, "X-Linked Thrombophilia with a Mutant Factor IX" N Engl J Med; 361:1671-1675, 2009) y fIX con la variante trombofética "Malmo" 148T (fIX Malmo, "véase Graham et al, "The Malmo Polymorphism of Coagulation Factor IX, An Immunologic Polymorphism Due to Dimorphism of Residue 148 That Is in Linkage Disequilibrium with Two Other FIX Polymorphisms", Am. J. Hum. Genet. 42:573-580, 1988)

Los ADNc de fIX optimizados para hígado y mieloides se diseñaron y optimizaron de acuerdo con las tablas de hígado y mieloides mostradas en la FIG. 2. Todos los casos de "cg" en su secuencia se eliminan mediante la sustitución de codones sinónimos. Ambos ADNc incorporan las sustituciones de Padua y Malmo.

FIX optimizado con codones de hígado con mutaciones Padua/Malmo y sin CpG (Versión 1) (SEQ ID NO: 124)

ATGCAGAGGGTCAATATGATCATGGCTGAATCTCCTGGGCTGATCACCATTGCTGCTGGGATACCTGCTGTCTGCT
GAGTGTACAGTGTTCCTGGACCATGAGAATGCCAATAAGATCCTGAACAGGCCCAAGAGATACAATTCTGGAAAGCTG
GAGGAATTTGTGCAGGGCAACCTGGAGAGGGAATGCATGGAGGAAAAGTGTAGCTTTGAGGAAGCTAGGGAGGTGTTT
GAAAACACAGAGAGGACCACAGAATTCTGGAAGCAGTATGTGGATGGAGATCAGTGTGAGTCCAACCCCTGTCTGAAT
GGAGGGTCTTGCAAGATGATATCAACTCCTATGAGTGTGGTGTCTTTTGGATTGAAGGCCAAAATTTGTGAGCTG
GATGTGACCTGTAACATCAAGAATGGCAGGTGTGAGCAGTTCTGTAAAACTCTGCTGATAATAAGGTGGTCTGCAGC
TGTACAGAAGGCTATAGGCTGGCTGAGAACCAGAAGAGCTGTGAACCAGCTGTGCCCTTCCCTTGTGGGAGGGTGTCT
GTCAGCCAGACCTTAAGCTGACCAGAGCTGAGACTGTGTTCCAGATGTGGATTATGTCAACTCCACAGAGGCTGAA
ACCATCCTGGACAACATCACCCAGTCTACCCAGTCTTCAATGACTTTACCAGAGTGGTGGGAGGAGAGGATGCCAAA
CCAGGCCAGTTCCCCTGGCAGGTGGTCTGAATGGGAAGGTGGATGCTTTTTGTGGGGGATCCATTGTGAATGAGAAA
TGGATTGTACAGCTGCTCACTGTGTGGAGACAGGGGTCAAGATCACTGTGGTGGCTGGAGAGCACAACATTGAGGAA
ACAGAACATACTGAGCAGAAGAGGAATGTGATCAGAATCATCCCTCACCATAACTACAATGCTGCTATCAACAAATAT
AATCATGACATTGCCCTGCTGGAACGGATGAGCCTCTGGTGTCTGAACAGCTATGTACCCCAATCTGCATTGCTGAC
AAAGAGTATACCAATATCTTCTGAAGTTTGGATCTGGATATGTGTCTGGATGGGGAAGGGTCTTCCACAAGGGCAGG
TCTGCCCTGGTGTGTCAGTATCTGAGGGTGCCTCTGGTGGACAGAGCTACCTGCCTGCTGTCTACCAAGTTCACCATC
TACAACAATATGTTCTGTGCTGGATTTCATGAGGGAGGCAGGGACTCCTGTGAGGGGATCTGGAGGCCACATGTG
ACAGAGGTGGAAGGCACCAGCTTCTGACTGGCATCATCTCTTGGGGGGAGGAATGTGCTATGAAGGGGAAATATGGA
ATCTACACCAAAGTGAGCAGGTATGTGAACCTGGATCAAAGAGAAGACCAAACTGACCTGA

FIX optimizado con codones de hígado sin CpG que codifican modificaciones A582 (SEQ ID NO: 8)

ATGCAGAGGGTGAACATGATCATGGCTGAGTCTCCTGGACTGATCACCATCTGCCTGCTGGGCTATCTGCTGTCTGCT
GAGTGTACAGTGTTCCTGGACCATGAAAATGCTAATAAAAATCCTGAACAGGCCAAAGAGGTACAATTCTGGGAAACTG
GAGGAATTTGTGCAGGGAAACCTGGAGAGGGAATGCATGGAGGAAAAGTGTAGCTTTGAGGAAGCCAGGGAGGTGTTT
GAAAATACAGAGAGGACCACAGAGTTCTGGAAACAGTATGTGGATGGGGATCAGTGTGAGTCCAACCCCTGTCTGAAT
GGAGGGTCTTGCAAGGATGATATCAACTCCTATGAGTGTGGTGTCTTTTGGATTGAAGGCCAAGAATTTGTGAGCTG
GATGTGACCTGTAACATCAAAAATGGGAGGTGTGAGCAGTTCTGTAAAGAACTCTGCTGATAATAAAGTGGTCTGCAGC
TGTACAGAAGGCTACAGGCTGGCTGAGAACCAGAAGAGCTGTGAACCAGCTGTGCCCTTCCCTTGTGGGAGGGTGTCT
GTCAGCCAGACCAGCAAGCTGACCAGAGCTGAGGCTGTGTTTCTGATGTGGATTATGTCAACTCTACAGAGGCTGAA
ACCATCCTGGACAACATCACCCAGTCTACCCAGTCTTCAATGACTTTACCAGGTGGTGGGAGGGGAGGATGCTAAG
CCAGGACAGTTCCCCTGGCAGGTGGTCTGAATGGCAAAGTGGATGCTTTTTGTGGGGGCTCCATTGTGAATGAGAAG
TGGATTGTACAGCTGCTCACTGTGTGAAAACCTGGGGTCAAGATCACAGTGGTGGCTGGAGAGCACAACATTGAGGAA
ACTGAACATACAGAGCAGAAAAGGAATGTGATCAGAATCATCCCCCACCATAACTACAATGCTGCTATCAACAAGTAT
AATCATGACATTGCCCTGCTGGAACGGATGAGCCTCTGGTGTCTGAACAGCTATGTACCCCAATCTGCATTGCTGAC
AAGAGTATACCAATATCTTCTGAAATTTGGGTCTGGATATGTGTCTGGGTGGGGAAGGGTCTTCCACAAGGGGAAGG
TCTGCTCTGGTGTGTCAGTATCTGAGGGTGGCCCTGGTGGACAGAGCTACCTGCCTGAGGAGCACAAGTTCACCATC
TACAACAATATGTTCTGTGCTGGATTTCATGAGGGAGGGAGGGACTCCTGTGAGGGAGATTCTGGAGGCCCTCATGTG
ACAGAGGTGGAAGGCACCAGCTTCTGACTGGCATCATCTCTTGGGGGGAGGAATGTGCTATGAAGGGGAAATATGGA
ATCTATACCAAGGTGTCCAGATATGTCAACTGGATCAAAGGAGAAAACCAAGCTGACCTGA

FIX optimizado con codones de hígado sin CpG incluyendo modificaciones de Padua y A582 (SEQ ID NO: 9)

ATGCAGAGGGTGAACATGATCATGGCTGAGTCTCCTGGACTGATCACCATCTGCCTGCTGGGCTATCTGCTGTCTGCT
GAGTGTACAGTGTTCCTGGACCATGAAAATGCTAATAAAAATCCTGAACAGGCCAAAGAGGTACAATTCTGGGAAACTG
GAGGAATTTGTGCAGGGAAACCTGGAGAGGGAATGCATGGAGGAAAAGTGTAGCTTTGAGGAAGCCAGGGAGGTGTTT
GAAAATACAGAGAGGACCACAGAGTTCTGGAAACAGTATGTGGATGGGGATCAGTGTGAGTCCAACCCCTGTCTGAAT
GGAGGGTCTTGCAAGGATGATATCAACTCCTATGAGTGTGGTGTCTTTTGGATTGAAGGCCAAGAATTTGTGAGCTG
GATGTGACCTGTAACATCAAAAATGGGAGGTGTGAGCAGTTCTGTAAAGAACTCTGCTGATAATAAAGTGGTCTGCAGC
TGTACAGAAGGCTACAGGCTGGCTGAGAACCAGAAGAGCTGTGAACCAGCTGTGCCCTTCCCTTGTGGGAGGGTGTCT
GTCAGCCAGACCAGCAAGCTGACCAGAGCTGAGGCTGTGTTTCTGATGTGGATTATGTCAACTCTACAGAGGCTGAA
ACCATCCTGGACAACATCACCCAGTCTACCCAGTCTTCAATGACTTTACCAGGTGGTGGGAGGGGAGGATGCTAAG
CCAGGACAGTTCCCCTGGCAGGTGGTCTGAATGGCAAAGTGGATGCTTTTTGTGGGGGCTCCATTGTGAATGAGAAG
TGGATTGTACAGCTGCTCACTGTGTGAAAACCTGGGGTCAAGATCACAGTGGTGGCTGGAGAGCACAACATTGAGGAA
ACTGAACATACAGAGCAGAAAAGGAATGTGATCAGAATCATCCCCCACCATAACTACAATGCTGCTATCAACAAGTAT
AATCATGACATTGCCCTGCTGGAACGGATGAGCCTCTGGTGTCTGAACAGCTATGTACCCCAATCTGCATTGCTGAC

AAGGAGTATACCAATATCTTCTGAAATTTGGGTCTGGATATGTGTCTGGGTGGGGAAGGGTCTTCCACAAGGGGAAGG
TCTGCTCTGGTGTGTCAGTATCTGAGGGTGGCCCTGGTGGACAGAGCTACCTGCCTGCTGAGCACCAAGTTCACCATC
TACAACAATATGTTCTGTGCTGGATTTCATGAGGGAGGGAGGGACTCCTGTGAGGGAGATTCTGGAGGCCCTCATGTG
ACAGAGGTGGAAGGCACCAGCTTCTGACTGGCATCATCTCTTGGGGGGAGGAATGTGCTATGAAGGGGAAATATGGA
ATCTATACCAAGGTGTCCAGATATGTCAACTGGATCAAAGGAGAAAACCAAGCTGACCTGA

Además, se construyeron otras dos variantes. La primera es una variante optimizada para hígado, Padua, 148T que es muy similar a la secuencia optimizada para hígado anterior, excepto que se sintetizó usando una versión alternativa del algoritmo de optimización de codones.

FIX optimizado con codones de hígado con mutaciones Padua/Malmo y sin CpG (Versión 2) (SEQ ID NO: 10)

5

ATGCAGAGGGTGAACATGATCATGGCTGAGTCTCCTGGACTGATCACCATCTGCCTGCTGGGCTATCTGCTGTCTGCT
GAGTGTACAGTGTTCCTGGACCATGAAAATGCTAATAAAATCCTGAACAGGCCAAAGAGGTACAATTCTGGGAAACTG
GAGGAATTTGTGCAGGGAAACCTGGAGAGGGAATGCATGGAGGAAAAGTGTAGCTTTGAGGAAGCCAGGGAGGTGTTT
GAAAATACAGAGAGGACCACAGAGTTCTGGAAACAGTATGTGGATGGGGATCAGTGTGAGTCCAACCCCTGTCTGAAT
GGAGGGTCTTGCAAGGATGATATCAACTCCTATGAGTGTCTGGTGTCTTTTGGATTGAAGGCAAGAATTGTGAGCTG
GATGTGACCTGTAACATCAAAAATGGGAGGTGTGAGCAGTTCTGTAAGAACTCTGCTGATAATAAAGTGGTCTGCAGC
TGTACAGAAGGCTACAGGCTGGCTGAGAACCAGAAGAGCTGTGAACAGCTGTGCCCTTCCCTTGTGGGAGGGTGTCT
GTCAGCCAGACCAGCAAGCTGACCAGAGCTGAGACAGTGTTCCTGATGTGGATTATGTCAACTCTACAGAGGCTGAA
ACCATCCTGGACAACATCACCCAGTCTACCCAGTCTTCAATGACTTTACCAGGGTGGTGGGAGGGGAGGATGCTAAG
CCAGGACAGTTCCCTTGGCAGGTGGTCTGAAATGGCAAAGTGGATGCTTTTTGTGGGGGCTCCATTGTGAATGAGAAG
TGGATTGTCAGAGTGTCTGCTGTGGAACTGGGGTCAAGATCACAGTGGTGGCTGGAGAGCACAACTTGAGGAA
ACTGAACATACAGAGCAGAAAAGGAATGTGATCAGAATCATCCCCACCATAACTACAATGCTGCTATCAACAAGTAT
AATCATGACATTGCCCTGCTGGAAGTGGATGAGCCTCTGGTGTGTAACAGCTATGTCAACCAATCTGCATTGCTGAC
AAGGAGTATACCAATATCTCCTGAAATTTGGGTCTGGATATGTGTCTGGGTGGGAAGGGTCTTCCACAAGGGAAGG
TCTGCTCTGGTGTCTGAGTATCTGAGGGTGCCTTGGTGGACAGAGCTACCTGCCTGCTGAGCACCAGTTCACCATC
TACAACAATATGTTCTGTGCTGGATTTCATGAGGGAGGGAGGGACTCCTGTGAGGGAGATTCTGGAGGCCCTCATGTG
ACAGAGGTGGAAGGCACCAGCTTCTGACTGGCATCATCTCTTGGGGGAGGAATGTGCTATGAAGGGGAAATATGGA
ATCTATACCAAGGTGTCCAGATATGTCAACTGGATCAAGGAGAAAACCAAGCTGACCTGA

Además, se optimizó una secuencia de fIX con mutaciones Padua/Malmo y sin CpG de acuerdo con la tabla patrón de optimización de codones humanos (véase la FIG. 2B).

10 FIX optimizado con codones humanos con mutaciones Padua/Malmo y sin CpG (SEQ ID NO: 127)

ATGCAGAGGGTGAATATGATTATGGCTGAGTCCCCTGGGCTGATTACCATTTCCTGCTGGGATACCTGCTGTCTGCT
GAGTGTACAGTGTTCCTGGACCATGAGAATGCAAATAAGATCCTGAACAGGCCCAAGATATAATAGTGGAAAGCTG
GAGGAATTTGTGCAGGGCAACCTGGAGAGAGAATGCATGGAGGAAAAGTGTAGCTTTGAGGAAGCCAGGGAGGTGTTT
GAAAATACAGAGAGAACCACAGAATTCTGGAAGCAGTATGTGGATGGAGATCAGTGTGAGAGCAACCCCTGTCTGAAT
GGAGGGAGTTGCAAAGATGATATCAACTCATATGAATGCTGGTGTCTTTTGGATTGAAGGCAAAAATTTGTGAGCTG
GATGTGACCTGTAACATTAAGAATGGGAGGTGTGAGCAGTTTTGTAAAACTCTGCTGATAATAAAGTGGTCTGCAGT
TGTACAGAAGGGTATAGACTGGCTGAGAACCAGAAGTCTGTGAACAGCTGTGCCCTTCCCTTGTGGAAGGGTGTCT
GTCTCCAGACTTCAAACTGACCAGAGCTGAGACTGTGTTTTCTGATGTGGATTATGTCAACAGCACAGAGGCTGAA
ACTATCCTGGACAACATTACTCAGTCTACCCAGAGTTTCAATGACTTTACCAGAGTGGTGGGAGGAGAGGATGCTAAA
CCAGGCCAGTTCCCTTGGCAGGTGGTCTGAAATGGGAAGGTGGATGCATTTTGTGGGGATCTATTGTGAATGAGAAA
TGGATTGTACAGCTGCTCACTGTGTGGAAGTGGGGTCAAGATCACAGTGGTGGCTGGAGAGCACAACATTGAGGAA
ACAGAACATACTGAGCAGAAGAGGAATGTGATCAGAATCATTCCTCACCATAACTACAATGCAGCCATCAACAAATAT
AATCATGACATTGCCCTGCTGGAAGTGGATGAGCCTCTGGTGTGTAACAGCTATGTCAACCAATCTGCATTGCTGAC
AAGGAGTACACTAACATCTTCTGAAATTTGGGTCTGAGATATGTGTCTGGATGGGAAGAGTCTTCCACAAGGGCAGG
TCTGACTGGTGTGCTGAGTATCTGAGAGTGCCTTGGTGGATAGGGCCACTTGTCTGCTGTCTACCAAGTTTACCATC
TACAACAATATGTTCTGTGCTGGATTTCATGAGGGAGGGAGAGACTCCTGTGAGGGAGATTCTGGAGGCCCATGTG
ACAGAGGTGGAAGGCACCAGCTTCTGACAGGCATCATTTCTTGGGGGAGGAATGTGCAATGAAGGGGAAATATGGA
ATCTACACCAAGTGAGCAGGTATGTGAACTGGATCAAGGAAAAGACCAACTGACATGA

La expresión *in vitro* de la secuencia de fIX optimizada para el hígado (SEQ ID NO: 10) y la secuencia de fIX optimizada para humanos (SEQ ID NO: 127) se evaluó en células HepG2 transfectadas de forma transitoria con los correspondientes plásmidos de expresión de fVIII (FIG. 7). La optimización específica del tejido conduce a una mayor actividad de fIX en las células HepG2 (hepáticas) que expresan el fIX optimizado para el hígado en comparación con el fIX optimizado para humanos. Estos resultados muestran que la optimización para hígado beneficia específicamente la expresión en células derivadas de hepatocitos.

Ejemplo 2

Desarrollo de Promotores

Este ejemplo describe el desarrollo iterativo de secuencias promotoras optimizadas para la expresión de proteínas en el tejido y las células del hígado.

Los promotores se sintetizaron *de novo* y se clonaron en un plásmido de expresión que conduce la expresión del fVIII de coagulación. La actividad de FVIII se midió 48 horas después de la transfección mediante un ensayo de coágulo de una etapa. Como comparador, se utilizó el promotor híbrido de hígado (HLP). HLP representa uno de los promotores

dirigidos al hígado más cortos pero más potentes descritos hasta la fecha. El promotor HLP y su uso se describen en McIntosh et al., "Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human fVIII variant", Blood, 25;121(17):3335-44, 2013. La secuencia de HLP se proporciona como la SEQ ID NO: 128:

5

```
TGTTTGCTGCTTGCAATGTTTGCCCATTTAGGGTGGACACAGGACGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGA
TCCCAGCCAGTGGACTTAGCCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCC
CCCGTTGCCCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACT
GACCTGGGACAGTGAATC
```

Diseño Inicial de Promotores (Promotores de 1ª Generación)

10 Para comenzar la construcción de promotores sintéticos dirigidos al hígado, se seleccionaron dos promotores mínimos dirigidos al hígado que fueron plataformas diseñadas para modificaciones adicionales. Estos dos promotores se designan "ABP-SynO" (SEQ ID NO: 131) y "ABP-HPI-God". Estos promotores son fusiones novedosas de elementos de control regulador previamente descritos. ABP "es una región agrupada de sitios de unión a factores de transcripción, " HP1 "es un sitio de unión a factores de transcripción específico, "God" es una región similar a un potenciador que funciona en proximidad directa al sitio de inicio de la transcripción, y "SynO "es un promotor mínimo que contiene el sitio de unión al factor de transcripción HP1 y una caja TATA. Para todas las construcciones, donde no se proporciona dentro del contexto natural, se agregó o se completó una secuencia TATA (TATAAA) inmediatamente 3' de la región promotora.

20 El promotor inicial diseñado se basó en el elemento "ABP", que se describe, por ejemplo, en Rouet et al., "A potent enhancer made of clustered liver-specific elements in the transcription control sequences of human alpha 1-microglobulin/bikunin gene", J Biol Chem., 267(29):20765-73, 1992. ABP comprende la secuencia de nucleótidos establecida como **elemento ABP (SEQ ID NO: 113)**:

25 GTTAATTTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCGAGCATTTACTCTCTCTGTTTGCTCTGGTTAATAAT
CTCAGGAGCACAAACA

Como se ilustra en la Figura 8, ABP comprende los siguientes sitios de unión al TF: TF HNF-1-1 (nucleótidos 16-23 de la SEQ ID NO: 4), HNF-4 (nucleótidos 26-36 de la SEQ ID NO: 4), HNF-3a (nucleótidos 39-45 de la SEQ ID NO: 4), HNF1-2 (nucleótidos 48-62 de la SEQ ID NO: 4) y HNF-3-2 TF (nucleótidos 65-71 de la SEQ ID NO: 4).

30 Varios de los promotores divulgados incluyen un sitio de unión al TF HP1 (GTTAATAATTTTC, nucleótidos 75-87 de la SEQ ID NO: 4). El elemento HP1 se describe, por ejemplo, en Schorpp et al., "Hepatocyte-specific promoter element HP1 of the Xenopus albumin gene interacts with transcriptional factors of mammalian hepatocytes", J Mol Biol., 202(2):307-20, 1988. El sitio de unión al TF HP1 está incluido en el elemento SynO (incluido en varios de los promotores divulgados), que también incluye una caja TATA. La secuencia del elemento SynO se proporciona como **elemento SynO** GAGGTTAATAATTTCCAGATCTCTCTGAGCAATAGTATAAAA (SEQ ID NO: 114)

El elemento SynO se describe, por ejemplo, en Ryffel et al., "Liver cell specific gene transcription *in vitro*: the promoter elements HP1 and TATA box are necessary and sufficient to generate a liver-specific promoter". Nucleic Acids Res., 17(3): 939-953, 1989.

Varios de los promotores divulgados incluyen un "God" que comprende la secuencia establecida como: **elemento God** AGTCATATGTTTGCTCACTGAAGGTTACTAGTTAACAGGCATCCCTTAAACAGGA (SEQ ID NO: 115)

45 El elemento God se describe, por ejemplo, en Godbout et al., "Multiple regulatory elements in the intergenic region between the alpha-fetoprotein and albumin genes", Mol Cell Biol., 6(2):477-87, 1986.

Los elementos ABP, SynO y God se combinaron para formar dos nuevos promotores, "ABP-SynO" y "ABP-HP1-God" como sigue (véase la FIG. 8):

promotor ABP-SynO (SEQ ID NO: 131)

```
GTTAATTTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCGAGCATTTACTCTCTCTGTTTGCTCTGGTTAATAAT
CTCAGGAGCACAAACAGAGGTTAATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAA
```

55

promotor ABP-Hp1-God (SEQ ID NO: 6)

GTAAATTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCAGCATTACTCTCTGTTTGCTCTGGTTAATAAT
CTCAGGAGCACAAACAGAGGTTAATAATTTTCAGTCATATGTTTGCTCACTGAAGGTTACTAGTTAACAGGCATCCCT
TAAACAGGATATAAAA

La expresión *in vitro* de FVIII se ensayó en células HepG2 transfectadas transitoriamente con plásmidos de expresión de FVIII impulsados por el promotor ABP-SynO o ABP-Hp1-God (FIG. 9A). Adicionalmente, la actividad *in vivo* de FVIII en ratones se ensayó mediante plásmido desnudo hidrodinámicamente suministrado que expresaba FVIII impulsado por el promotor ABP-SynO o ABP-Hp1-God (FIG. 9B). A partir de los resultados, se muestra que los promotores tanto ABP-SynO como ABP-Hp1-God impulsan la expresión de FVIII *in vitro* e *in vivo*. Sin embargo, estos diseños iniciales no impulsan la expresión de FVIII de manera tan fuerte como el promotor HLP *in vitro*, y que la expresión *in vivo* disminuye rápidamente en relación con HLP.

Optimización Inicial (Promotores de 2ª Generación)

Para aumentar la fuerza transcripcional de los promotores ABP-SynO y ABP-Hp1-God, se siguieron múltiples estrategias para la optimización. Esto incluye alterar los sitios de unión a factores de transcripción para reflejar la secuencia de unión consenso, eliminar el espacio intermedio entre los sitios de unión a factores de transcripción, agregar sitios de unión del factor de transcripción adicionales, añadir un motivo del sitio de inicio de la transcripción e incluir el intrón de SV40.

Se generó una variante de ABP que contiene sitios de unión a TF consenso, como sigue:

ABP-exact (sitios de unión a factores de transcripción consenso) (SEQ ID NO: 116)

GTAAATCATTAACTTAAAAAGCAGTCAAAAGTCCAAGGTCAAAGGTCAGAGCATTACTCTCTCCAATGTTGACTCT
CGTTAATGATTAAAGGAGCAATTGTTGACTT

Como se ilustra en la Figura 8, ABP-exact comprende los siguientes sitios de unión a TF consenso: HNF-1-1 consenso, HNF-4 consenso, HNF-3a consenso, HNF1-2 consenso y HNF-3-2 consenso. Se generó una versión condensada de ABP-exact que incluye los mismos sitios de unión a TF consenso, pero una secuencia general más corta, denominada Short-ABP-exact, cuya secuencia se establece como:

Short-ABP-exact (SEQ ID NO: 117) GTAAATCATTAACTTAGGTCAAAGGTCAGACAATGTTGACTCTCGTTAATGATTAACCGGAATTGTTGACTT

Las siguientes características se incluyeron adicionalmente en algunos de los promotores divulgados: Un sitio de inicio de la transcripción (TSS), que contiene un 23 que contiene un espaciador rico en GC y se colocó inmediatamente después de una caja TATA en el promotor para un espaciado óptimo con el motivo de inicio de la transcripción inmediatamente después del espaciador (véase FIG. 10). La secuencia de TSS ensayada incluye la secuencia establecida como: GCCAGCAGCAGCCTGACCACATCTCATCCTC (nucleótidos 116-146 de la SEQ ID NO: 4)

Un sitio de unión al factor de transcripción HNF1a. HNF1a es un factor de transcripción dirigido al hígado: GTTAATCATTA (nucleótidos 1-12 de la SEQ ID NO: 4)

Un sitio de unión al factor de transcripción Sp1. Sp1 es un factor de transcripción dirigido al hígado: TGGGCGGAGT (nucleótidos 1-10 de la SEQ ID NO: 121)

Una secuencia de intrón SV40 establecida como:

CTCTAAGGTAAATATAAAATTTTAAAGTGATAATGTGTTAACTACTGATTCTAATTGTTTGTGATTTTAGATTCC
AACCTATGGAAGTGA (nucleótidos 163-225 de la SEQ ID NO: 112)

Estos elementos se combinaron para formar varios promotores nuevos, como sigue (véase la FIG. 11):

ABP-exact-SynO (SEQ ID NO: 118)

GTAAATCATTAACTTAAAAAGCAGTCAAAAGTCCAAGGTCAAAGGTCAGAGCATTACTCTCTCCAATGTTGACTCT
CGTTAATGATTAAAGGAGCAATTGTTGACTTGAAGGTTAATAATTTCCAGATCTCTCTGAGCAATAGTATAAAA

ShortABP-exact-SynO (SEQ ID NO: 119)

GTAAATCATTAACTTAGGTCAAAGGTCAGACAATGTTGACTCTCGTTAATGATTAAACCGGAATTGTTGACTTGAAGGTT
AATAATTTCCAGATCTCTCTGAGCAATAGTATAAAA

ABP-HP1-God-TSS (SEQ ID NO: 7)

GTTAATTTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCGAGCATTACTCTCTCTGTTTGTCTGGTTAATAAT
CTCAGGAGCACAAACAGAGGTTAATAATTTTCAGTCATATGTTTGCTCACTGAAGGTTACTAGTTAACAGGCATCCCT
TAAACAGGATATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

5 HNF1a-ABP-SynO (SEQ ID NO: 120)

GTTAATCATTAAGTCGTTAATTTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCGAGCATTACTCTCTCTGTTT
GCTCTGGTTAATAATCTCAGGAGCACAAACAGAGGTTAATAATTTTCAGATCTCTCTGAGCAATAGTATAAAA

Sp1-ABP-SynO (SEQ ID NO: 121)

10

TGGGCGGAGTGTCGTTAATTTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGCGAGCATTACTCTCTCTGTTTGC
TCTGGTTAATAATCTCAGGAGCACAAACAGAGGTTAATAATTTTCAGATCTCTCTGAGCAATAGTATAAAA

HNF1-ShortABPEXact-SynO-TSS-Int (SEQ ID NO: 112)

GTTAATCATTAAGTCGTTAATCATTAACCTAGGTCAAAGGTCAGACAATGTTGACTCTCGTTAATGATTAACCGGAAT
TGTTGACTTGAGGTTAATAATTTTCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTC
ATCCTCCTCTAAGGTAAATATAAAATTTTTTAAGTGTATAATGTGTAAACTACTGATTCTAATTGTTGTGTATTTTA
GATTCCAACCTATGGAACCTGA

15

Los promotores se sintetizaron *de novo* y se clonaron en un plásmido de expresión que conduce la expresión del FVIII de coagulación. La actividad de FVIII se midió 48 horas después de la transfección mediante un ensayo de coágulo de una etapa. Como comparador, el promotor de hígado híbrido (HLP) se utiliza en este y otros experimentos.

20

La expresión *in vitro* de FVIII se ensayó en células HepG2 transfectadas transitoriamente con plásmidos de expresión de FVIII impulsados por el respectivo promotor (FIG. 12A). Adicionalmente, la actividad *in vivo* de FVIII en ratones se ensayó mediante plásmido desnudo hidrodinámicamente suministrado que expresaba FVIII impulsado por el respectivo promotor (FIG. 12B). A partir de estos datos, se muestra que la adición del motivo de sitio de inicio de la transcripción sintético, nuevo mejora sustancialmente la expresión. Adicionalmente, la adición de sitios de unión a factores de transcripción fuera de su contexto genómico natural tuvo impactos inesperados e impredecibles en la expresión. Cuando se añaden directamente proximal al elemento potenciador ABP, los sitios de unión al factor de transcripción HNF1 y Sp1 anulan completamente la expresión. Cuando se añadió el sitio de unión al factor de transcripción HNF1 directamente proximal al potenciador shortABPEXact, la expresión fue fuerte. Este hallazgo ilustra que las complejas interacciones espacio temporales de los factores de transcripción, ya sea en su contexto genómico natural o no natural, es difícil de modelar o predecir. Además, la alteración de los sitios de unión a factores de transcripción en las secuencias de unión consenso anuló la expresión génica. Sin embargo, la adición del sitio de unión al factor de transcripción HNF1, el sitio de inicio de la transcripción y el intrón SV40 fue suficiente para rescatar la expresión del diseño del promotor shortABPEXact-SynO complementado con HNF1. Como se ilustra en la FIG. 12B, ABP-HP1-God-TSS y HNF1-shortABPEXact-SynO-TSS-Int, mantuvieron ambos una expresión más alta que el promotor HLP durante el transcurso del experimento, lo que demuestra que la adición del sitio de inicio de la transcripción, el sitio de unión al factor de transcripción HNF1, el ajuste de la secuencia ABP a la secuencia consenso y la eliminación del ADN intermedio entre los sitios de unión al factor de transcripción, y/o la adición del intrón SV40 podría mejorar la durabilidad y la fuerza de expresión *in vivo*.

40

Optimización Adicional (Promotores de 3ª Generación)

Si bien el diseño del promotor ABP-Hp1-God-TSS probado excedió con creces la fuerza del promotor HLP, su tamaño (204 pares de bases) permaneció incompatible con algunos empaquetamientos completos de vectores AAV, tal como los que contienen transgenes de FVIII de longitud completa. La reducción adicional en el tamaño del promotor se dirigió por la selección de los elementos más prometedores probados y descritos anteriormente, así como un elemento nuevo, shortABP, que es el potenciador ABP donde se conservan las secuencias de los sitios de unión a factores de transcripción genómico natural, pero se han truncado las secuencias intermedias entre ellos.

45

50 shortABP (nucleótidos 16-71 de la SEQ ID NO: 4): GTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGT-TAATAATCTCAGGACAAACA

Como se ilustra en la Figura 13, shortABP contiene los siguientes sitios de unión a TF: TF HNF-1-1 (nucleótidos 16-23 de la SEQ ID NO: 4), HNF-4 (nucleótidos 26-36 de la SEQ ID NO: 4), HNF-3a (nucleótidos 39-45 de la SEQ ID NO: 4), HNF1-2 (nucleótidos 48-62 de la SEQ ID NO: 4) y HNF-3-2 TF (nucleótidos 65-71 de la SEQ ID NO: 4).

55

Como se ilustra en la Figura 13, el sitio de unión al TF HNF1, el elemento SynO (que contiene un sitio de unión al TF HP1 y una caja TATA), y un sitio de inicio de la transcripción se unieron a shortABP para formar el HNF1-shortABP-SynO-TSS (designado el Paquete Combinatorio Hepático, o HCB), como sigue:

5 **HNF1-shortABP-SynO-TSS (también denominado Paquete Combinatorio Hepático, o HCB) (SEQ ID NO: 4)**

GTTAATCATTAAAGTCGTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACAGAGGTTA
ATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCTC

10 Como se ilustra en la Figura 13, se unieron el sitio de unión al TF HNF1, el elemento God, una caja TATA y un sitio de inicio de la transcripción a shortABP para formar shortABP-HP1-God-TSS, como sigue:
shortABP-HP1-God-TSS (SEQ ID NO: 5)

GTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACATACATTTTCAGTCATATGTTTG
CTCACTGAAGGTTACTAGTTAACAGGCATCCCTTAAACAGGATATAAAAGGCCAGCAGCAGCCTGACCACATCTCATC
CTC

15 Los promotores se sintetizaron *de novo* y se clonaron en un plásmido de expresión que conduce la expresión del fVIII de coagulación. La actividad de FVIII se midió 48 horas después de la transfección mediante un ensayo de coágulo de una etapa. La expresión *in vitro* de FVIII se ensayó en células HepG2 transfectadas transitoriamente con plásmidos de expresión de FVIII impulsados por el respectivo promotor mostrado en la FIG. 14A. Adicionalmente, la actividad *in vivo* de fVIII en ratones se ensayó mediante plásmido desnudo hidrodinámicamente suministrado que expresaba FVIII
20 impulsado por el respectivo promotor (FIG. 14). A partir de estos datos, se muestra que el promotor HNF1-shortABP-SynO-TSS (HCB) promueve una mayor expresión de fVIII en comparación con la de HLP tanto *in vivo* como *in vitro*, mientras mantiene una expresión duradera 4-14 veces mayor que la de HLP *in vivo*, a pesar de que el HNF1-shortABP-SynO-TSS es un 42 % más pequeño que HLP.

25 **Optimización Complementaria (Promotores de 4ª Generación)**

Recientemente se construyó un potente potenciador dirigido al hígado, designado el módulo regulador específico de hepatocitos informático (HSCRM8 o "HS") combinando las secuencias de varias especies en un nuevo potenciador construido por ordenador (véase, por ejemplo, Nair et al., "Computationally designed liver-specific transcriptional modules and hyperactive fIX improve hepatic gene therapy", Blood, 23(20): 3195-3199, 2014). La secuencia del
30 potenciador HS y los sitios de unión a los factores de transcripción correspondientes se proporcionan como sigue:

HS (secuencia potenciadora no humana HSCRM8) (SEQ ID NO: 101)
GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGTTATCGGAGGAGCAAACAGGGACTAAGTCCAC

35 El elemento de respuesta HS incluye los siguientes sitios de unión a TF:

MYOD	GGCTGCTGGTGAATATT, nucleótidos 5-22 de la SEQ ID NO: 101
CEBP	GCTGCTGGTGAA, nucleótidos 7-18 de la SEQ ID NO: 101
Nhf1	GCTGGTGAATATTAACCA, nucleótidos 10-27 de la SEQ ID NO: 101
Lef1/TCF1	TTAACCAAGGT, nucleótidos 21-31 de la SEQ ID NO: 101
CEBP	CGGAGGAGCAAA, nucleótidos 44-55 de la SEQ ID NO: 101
Forkhead	GGAGCAAACAGGG, nucleótidos 48-60 de la SEQ ID NO: 101
Lef1/TCF1	AGGGACTAAG, nucleótidos 57-66 de la SEQ ID NO: 101

El genoma humano se examinó para determinar las secuencias humanas correspondientes a las del elemento HS y se modificó el elemento HS para contener solo secuencias humanas para generar una secuencia potenciadora completamente humana denominada "HSh". La secuencia del potenciador HSh y los sitios de unión a factores de transcripción correspondientes se proporcionan como sigue (véase también la FIG. 15):

40 **HSh** (secuencia genómica humana) (SEQ ID NO: 111)
GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCAGT- TATCGGAGGAGCAAACAGGGGCTAAGTCCAC

45 El elemento de respuesta HSh incluye los siguientes sitios de unión a TF:

MYOD	GGCTGCTGGTGAATATT, nucleótidos 5-22 de la SEQ ID NO: 101
CEBP	GCTGCTGGTGAA, nucleótidos 7-18 de la SEQ ID NO: 101
Nhf1	GCTGGT GAATATTAACCA, nucleótidos 10-27 de la SEQ ID NO: 101

(continuación)

Lef1/TCF1 TTAACCAAGGT, nucleótidos 21-31 de la SEQ ID NO: 101
 CEBP CGGAGGAGCAAA, nucleótidos 44-55 de la SEQ ID NO: 101
 Forkhead GGAGCAAACAGGG, nucleótidos 48-60 de la SEQ ID NO: 101
 Lef1/TCF1 AGGGGCTAAG, nucleótidos 57-66 de la SEQ ID NO: 111

También se utilizaron partes del potenciador HSh, como sigue:

5'HSh (porción 5' de HSCRM8h) (nucleótidos 6-32 de la SEQ ID NO: 111) GGCTGCTGGTGAATATTAACCAAGGTC

- 5 El elemento de respuesta 5'HSh incluye los siguientes sitios de unión a TF:

MYOD GGCTGCTGGTGAATATT, nucleótidos 5-22 de la SEQ ID NO: 101
 CEBP GCTGCTGGTGAA, nucleótidos 7-18 de la SEQ ID NO: 101
 Nhf1 GCTGGTGAATATTAACCA, nucleótidos 10-27 de la SEQ ID NO: 101
 Lef1/TCF1 TTAACCAAGGT, nucleótidos 21-31 de la SEQ ID NO: 101

3'HSh (porción 3' de HSCRM8h) (nucleótidos 44-68 de la SEQ ID NO: 111) CGGAGGAGCAAACAG- GGGCTAAGTC

- 10 El elemento de respuesta 3'HSh incluye los siguientes sitios de unión a TF:

CEBP CGGAGGAGCAAA, nucleótidos 44-55 de la SEQ ID NO: 101
 Forkhead GGAGCAAACAGGG, nucleótidos 48-60 de la SEQ ID NO: 101
 Lef1/TCF1 AGGGGCTAAG nucleótidos 57-66 de la SEQ ID NO: 111

sHS (short HS, se ha eliminado el espacio intermedio entre los grupos 5' y 3' de los sitios de unión a factores de transcripción) (nucleótidos 1-54 de la SEQ ID NO: 106) GGCTGCTGGTGAATATTAACCAAGGTCATCGGAGGAG-CAAACAGGGACTAAGTC

- 15 El elemento de respuesta sHS incluye los siguientes sitios de unión a TF:

MYOD GGCTGCTGGTGAATATT, nucleótidos 5-22 de la SEQ ID NO: 101
 CEBP GCTGCTGGTGAA, nucleótidos 7-18 de la SEQ ID NO: 101
 Nhf1 GCTGGT GAATATTAACCA, nucleótidos 10-27 de la SEQ ID NO: 101
 Lef1/TCF1 TTAACCAAGGT, nucleótidos 21-31 de la SEQ ID NO: 101
 CEBP CGGAGGAGCAAA, nucleótidos 44-55 de la SEQ ID NO: 101
 Forkhead GGAGCAAACAGGG, nucleótidos 48-60 de la SEQ ID NO: 101
 Lef1/TCF1 AGGGACTAAG, nucleótidos 57-66 de la SEQ ID NO: 111

- 20 Adicionalmente, se construyó una forma modificada del promotor shortABP, denominada "supershortABP", eliminando adicionalmente los nucleótidos y reorganizando los sitios de unión a TF. La secuencia del elemento de respuesta supershortABP y los sitios de unión a factores de transcripción correspondientes se proporcionan como sigue:

Super short ABP (elemento más corto basado en ABP) (SEQ ID NO: 122) CCCTTGCTGGTTAATAATCTCAGT-TAATTTGTTTGCACAAACA

- 25 El elemento de respuesta supershortABP incluye los siguientes sitios de unión a TF:

HNF4 CCCTTGC, nucleótidos 1-7 de la SEQ ID NO: 122
 HNF1b TGGTTAATAATCTCA, nucleótidos 8-22 de la SEQ ID NO: 122
 HNF1 GTTAATT, nucleótidos 23-29 de la SEQ ID NO: 122
 HNF3 TGTTTGC, nucleótidos 30-36 de la SEQ ID NO: 122
 HNF3b ACAAACA, nucleótidos 37-43 de la SEQ ID NO: 122

- 30 Como se ilustra en la Figura 16, los elementos anteriores se combinaron para formar varios promotores nuevos, como sigue (véase la FIG. 11):

Agro (SEQ ID NO: 107)

GGCTGCTGGTGAATATTAACCAAGGTCCCTTGCTGGTTAATAATCTCAGTTAATTTGTTTGCACAAACACGGAGGAG
 CAAACAGGGGAGGTTAATAATTTTCTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

- 35 La secuencia Agro incluye 5'HS (nucleótidos 1-27 de la SEQ ID NO: 107), Super short ABP (nucleótidos 28-70 de la SEQ ID NO: 107), 3'HSh (nucleótidos 71-87 de la SEQ ID NO: 107), SynO (nucleótidos 88-110 de la SEQ ID NO: 107),

y TSS (nucleótidos 111-142 de la SEQ ID NO: 107)

HSh-HCB (SEQ ID NO: 102)

GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACTAGGTTA
ATCATTAAAGTCGTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACAGAGGTAAATAA
TTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

5'HSh-HCB (SEQ ID NO: 104)

GGCTGCTGGTGAATATTAACCAAGGTCGTTAATCATTAAAGTCGTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGT
TAATAATCTCAGGACAAACAGAGGTAAATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCT
GACCACATCTCATCCTC

3'HSh-HCB (SEQ ID NO: 103)

CGGAGGAGCAAACAGGGGCTAAGTCGTTAATCATTAAAGTCGTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTA
ATAATCTCAGGACAAACAGAGGTAAATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGA
CCACATCTCATCCTC

HSh-SynO-TSS (SEQ ID NO: 105)

GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCCAGTTATCGGAGGAGCAAACAGGGGCTAAGTCCACGAGGTTA
ATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

HS-SynO-TSS (SEQ ID NO: 108)

GGGGAGGCTGCTGGTGAATATTAACCAAGGTCACCCCAGTTATCGGAGGAGCAAACAGGGACTAAGTCCACGAGGTTA
ATAATTTTCCAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

sHS-SynO-TSS (SEQ ID NO: 106)

GGCTGCTGGTGAATATTAACCAAGGTCATCGGAGGAGCAAACAGGGACTAAGTCGAGGTTAATAATTTTCCAGATCTC
TCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTC

Los promotores se sintetizaron *de novo* y se clonaron en un plásmido de expresión que conduce la expresión del fVIII de coagulación. La actividad de fVIII se midió 48 horas después de la transfección mediante un ensayo de coágulo de una etapa. La expresión *in vitro* de fVIII se ensayó en células HepG2 transfectadas transitoriamente con plásmidos de expresión de fVIII impulsados por el respectivo promotor mostrado en la FIG. 17. Estos resultados demuestran que la fuerza *in vivo* del promotor HCB puede aumentarse mediante la adición de unas secuencias potenciadoras complementarias. El mayor beneficio se observó cuando se añadió el potenciador HSh humano completo al extremo 5' del promotor HCB. Complementar solo las porciones de 5' o 3' del HSh atenuó la expresión mejorada ligeramente en comparación con el módulo HSh completo. Adicionalmente, cuando se eliminó el potenciador shortABP del diseño de HCB y se reemplazó con potenciadores bien HSh o HS, la expresión de fVIII disminuyó ligeramente de los niveles observados en el promotor HSh-HCB, que contiene el potenciador shortABP. Sin embargo, cuando el plásmido sHS-SynO se administró hidrodinámicamente a ratones, no se observó expresión, lo que demuestra que el módulo potenciador shortABP puede ser un potenciador más adecuado para una expresión *in vivo* duradera.

Ejemplo 3

Vector AVV Recombinante para la Expresión de fVIII

Este ejemplo ilustra vectores AAV recombinantes ejemplares que codifican una variante de fVIII que comprende un genoma dimensionado para una expresión óptima de proteínas basada en vectores AAV (es decir, un genoma de 5 kb o menos pb).

La FIG. 4 representa vectores basados en AAV para la expresión de variantes de fVIII que carecen del dominio B. Sin embargo, las construcciones mostradas en la FIG. 4 están por encima del límite de 5,0 kb para la expresión óptima de proteínas de un vector AAV. La eliminación del ADN genómico vírico no esencial y los restos de clonación, y la sustitución con secuencias promotoras más cortas (como el promotor HCB (SEQ ID NO: 4, 146 pb), permitieron el

desarrollo de un genoma de fVIII-AAV de alta expresión de aproximadamente 4,9 kb de longitud.

En la FIG. 18 se representa un vector AAV ejemplar con dicho genoma. Esta figura ilustra un vector AAV que incluye ITR 5' y 3', el promotor HCB (SEQ ID NO: 4), una molécula de ácido nucleico que codifica una proteína fVIII variante que carece del dominio B (tal como cualquiera de las secuencias ET3 o HSQ con CpG eliminados y codones de hígado optimizados proporcionadas en el presente documento) y una secuencia de poli A sintética.

Una secuencia ejemplar de un casete AAV como se muestra en la FIG. 18 se proporciona como la SEQ ID NO: 129, que tiene la siguiente estructura:

(5'AAV2 ITR) - *RE* - (promotor HCB) - Kozak - (región codificante HSQ) - *RE* - (señal de poliadenilación) - *RE* - (3'ITR de AAV2)

Los elementos del casete AAV de la SEQ ID NO: 129 son los siguientes:

Elemento	Inicio (pb)	Final (pb)
5' ITR de AAV2	1	141
Agel (sitio de restricción)	142	147
promotor HCB	148	293
Sitio XhoI destruido	294	299
Secuencia consenso Kozak	300	304
HSQ optimizada para hígado con CpG	305	4678
NotI (sitio de restricción)	4679	4686
Señal de poliA de beta globina de conejo	4687	4735
MunI (sitio de restricción)	4736	4741
3' ITR de AAV2	4742	4882

Los sitios de restricción pueden eliminarse opcionalmente del casete para proporcionar un genoma de AAV recombinante acortado. Adicionalmente, el transgén se puede sustituir según sea necesario. La eliminación de los elementos de los sitios de restricción generaría un vector de 4885 pares de bases (ET3) o 4855 (HSQ).

SEQ ID NO: 129:

```

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCCGC
CGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCAGTGGGGTTCCCTACCGGTGTTAATCAT
TAAGTCGTTAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACAGAGGTTAATAATTTTC
CAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTCGTCGAGCCACCATGCAGAT
CGAACTGTCTACCTGTTTCTTTCTGTGCCTGCTGCGGTTTTGTTTTTCCGCTACCAGAAGATACTACCTGGGAGCCGT
CGAACTGAGCTGGGATTACATGCAGTCTGACCTGGGAGAGCTGCCCGTGGACGCTAGATTCCCACCTAGAGTCCCTAA
GTCTTCCCTTCAACACCAGCGTGGTCTACAAGAAAACCCTGTTTCGTGGAGTTTACCGACCACCTGTTCAACATCGC
TAAGCCTAGACCACCATGGATGGGACTGCTGGGACCAACCATCCAGGCCGAGGTGTACGACACCGTGGTCATCACCCCT
GAAAAACATGGCTTCTCACCCCGTGTCCCTGCATGCTGTGGGCGTCTCCTACTGGAAGGCCAGCGAAGGGGCTGAGTA

```

TGACGATCAGACCAGCCAGCGGGAAAAAGAGGACGATAAGGTGTTCCCTGGCGGGTCCCATACCTACGTGTGGCAGGT
 CCTGAAGGAGAATGGACCAATGGCTTCCGACCCTCTGTGCCTGACCTACTCTTATCTGTCCCACGTGGACCTGGTCAA
 GGATCTGAACAGCGGCTGATCGGGGCTCTGCTGGTGTGTGCGGAAGGGTCCCTGGCCAAGGAGAAAACCCAGACCCCT
 GCATAAGTTTCATCTCTGCTGTTGCGCCGTGTTTGACGAAGGAAAAAGCTGGCACTCTGAGACCAAGAACTCTCTGATGCA
 GGACAGGGATGCCGCTTCCGCCAGAGCTTGGCCCAAGATGCACACCGTGAACGGCTACGTCAATAGGAGCCTGCCTGG
 ACTGATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGCATCGGAATGGGCACCACCCCTGAAGTGCACAGCATCTT
 CCTGGAGGGGCATACCTTTCTGGTCCGCAACCACCGGCAGGCTAGCCTGGAGATCTCTCCAATCACCTTCCTGACCGC
 CCAGACCCTGCTGATGGACCTGGGACAGTTCCTGCTGTTTTGCCACATCTCCAGCCACCAGCATGATGGCATGGAGGC
 TTACGTGAAAGTCGACTCCTGTCCCAGGAACCTCAGCTGAGGATGAAGAACAATGAGGAAGCCGAAGACTATGACGA
 TGACCTGACCAGCAGCGAGATGGATGTGGTCCGCTTCGATGACGATAACTCTCCCTCCTTTATCCAGATCCGGTCCGT
 GGCCAAAGAAACACCTTAAGACCTGGGTCCATTACATCGCCGCTGAGGAAGAGGACTGGGATTATGCTCCACTGGTGCT
 GGCCCCGACGATAGATCTTACAAAAGCCAGTATCTGAACAATGGACCCAGAGGATCGGCAGAAAGTACAAGAAAGT
 GAGGTTTCATGGCTTATACCGATGAGACCTTTAAGACCAGAGAAGCCATCCAGCACGAGTCCGGGATCCTGGGACCTCT
 GCTGTACGGCGAAGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAGCAGGCCCTTACAATATCTATCCACA
 TGGCATCACCGATGTGAGACCTCTGTACTCCCGCCGGCTGCCAAAGGGCGTGAAACACCTGAAGGACTTCCCAATCCT
 GCCCCGGGAAATCTTTAAGTATAAATGGACCGTCACCGTCGAGGATGGGCCCCACCAAGAGCGACCCCTAGGTGCCTGAC
 CAGATACTATTCTTCTTCGTGAATATGGAGAGAGACCTGGCTTCCGGACTGATCGGACCCCTGCTGATCTGTTACAA
 AGAGAGAGATCGACTACGACGATACCATCTCCGTGGAAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGA
 CCGCTCTTGGTACCTGACCGAGAACATCCAGCGGTTCTGCTTAATCCAGCTGGAGTGCAGCTGGAAGATCCCGAGTT
 CCAGGCCTCTAACATCATGCATTCCATCAATGGCTACGTGTTGACTCCCTGCAGCTGAGCGTGTGCCTGCACGAGGT
 CGTTACTTGGTATATCTGAGCATCGGAGCCCAGACCGATTTCTGTCTGTGTTCTTTTCCGGCTACACCTTTAAGCA
 TAAAATGGTGTATGAGGACACCCTGACCCTGTTCCTATTTTCCGGCGAAACCGTGTTTCATGAGCATGGAGAATCCCGG
 GCTGTGGATCCTGGGATGCCACAACCTCCGATTTTCAGGAATAGAGGGATGACCGCCCTGCTGAAAGTGAGCTCTTGTGA
 CAAGAACACCGGAGACTACTATGAAGATAGCTACGAGGACATCTCTGCTTATCTGCTGTCCAAAAACAATGCCATCGA
 GCCAGGAGCTTCTCTCAGAACCCTCCAGTGCTGAAGCGCCACCAGCGGGAGATCACCAGAACCACCCTGCAGAGCGA
 TCAGGAAGAGATCGACTACGACGATACCATCTCCGTGGAAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGA
 GAACCAGTCTCCAGGTCCTTCCAGAAGAAAACCAGACATTACTTTATCGCCGCTGTGGAGCGGCTGTGGGACTATGG
 CATGTCCAGCTCTCCTCACGTGCTGAGAAATAGAGCTCAGTCCGGAAGCGTCCACAGTTCAAGAAAGTGGTCTTCCA
 GGAGTTTACCGACGGAAGCTTTACCCAGCCACTGTACCGCGGCGAACTGAACGAGCACCTGGGGCTGCTGGGACCCTA
 TATCCGGGCTGAAGTGGAGGATAACATCATGCTCACCTTCAGGAATCAGGCCAGCAGACCCCTACTCTTTTTATTCAG
 CCTGATCTCCTACGAAGAGGACCAGAGACAGGAGCTGAACCAAGAAAAAATTCGTGAAGCCTAATGAGACCAAAAC
 CTACTTTTGAAGGTGCAGCACCATATGGCCCCCTACCAAGACGAGTTCGATTGCAAGGCCTGGGCTTATTTTAGCGA
 CGTGGATCTGGAGAAGGACGTCCACTCCGGCTGATCGGGCCACTGCTGGTGTGTATACCAACACCCTGAATCCAGC
 TCACGGAAGGCAGGTGACCGTCCAGGAATTCGCCCTGTTCTTTTACCATCTTTGATGAGACCAAGAGCTGGTACTTCAC
 CGAAAAATGGAGAGGAATTCAGAGCCCCATGTAACATCCAGATGGAAGACCCACCTTCAAGGAGAACTACAGATT
 TCATGCTATCAATGGGTATATCATGGATACCTGCCAGGACTGGTTCATGGCTCAGGACCAGAGGATCAGATGGTACCT
 GCTGAGCATGGGGTCTAACGAGAATATCCACTCCATCCATTTTCAGCGGACACGTGTTTACCGTCCGCAAGAAAGAAGA
 GTACAAGATGGCCCTGTACAACCTGTATCCCGGCGTGTTTCGAAACCGTCGAGATGCTGCCTTCCAAGGCTGGGATCTG
 GCGGGTGGAAATGCCGTGATCGGGGAGCACCTGCATGCCGGAATGTCTACCCTGTTCCCTGGTGTACTCCAATAAGTGTC
 GACCCCCCTGGGGATGGCTAGCGGACATATCCGCGACTTCCAGATCACCGCTTCCGGACAGTACGGACAGTGGGCTCC
 TAAGCTGGCTGAGATGCATCTTCTGGCTCCATCAACGCTTGGTTTACCAAGAGCCCTTCTCCTGGATCAAGTGGA
 CCTGCTGGCTCCAATGATCATCCATGGCATCAAAAACCAGGGGGCCAGGCAGAAAGTTCTCTTCCCTGTACATCAGCCA
 GTTTATCATCATGTATTCTCTGGATGGGAAGAAATGGCAGACCTACAGAGGCAATTCACCGGGACCCCTGATGGTGT
 CTTTGGCAACGTCGACAGCTCTGGGATCAAGCACAACATCTTCAATCCCCCTATCATCGCCGCTACATCCGGCTGCA
 CCAACCCATTATTCCATCCGAGCACCCCTGCGGATGGAGCTGATGGGGTGCATCTGAACAGCTGTTCTATGCCCT
 GGAATGGAGTCTAAGGCCATCTCCGACGCTCAGATCACCGCCTCCAGCTACTTCACCAATATGTTTGCTACCTGGTC
 CCAAGCAAGGCTAGACTGCATCTGCAGGAAGAAGCAACGCTTGAGACCAAGTGAACAATCCCAAGGAGTGGCT
 GCAGGTCGACTTCCAGAAAACCATGAAGGTGACCGGAGTACCACCCAGGGCGTGAAAAGCCTGCTGACCTCTATGTA
 CGTCAAGGAGTTCTCTGATCTCTTCCAGCCAGGACGGGACCAAGTGGACCCCTGTTCTTTTCAAGACGAAAGGTGAAAGT
 CTTCAGGGCAATCAGGATTCTTTTACCCCTGTGGTCAACAGCCTGGACCCACCCCTGCTGACCAGGTACCTGAGAAAT
 CCACCCACAGTCTGGGTGCATCAGATCGCTCTGAGGATGGAAGTCTGGGCTGCGAGGCCAGGACCTGTATTGAGC
 GGCCGCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGCAATTGAGGAACCCCTAGTGATG
 GAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTGCCCCGACGCCCGGGCTTT
 GCCCCGGGCGGCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG

Otro vector AAV ejemplar que incluye un transgén de fVIII impulsado por el promotor HCB se proporciona como la SEQ ID NO: 130, que proporciona un diseño prototípico de un casete de AAV que codifica un transgén terapéutico bajo el control del promotor HCB. Cada elemento está separado por uno o dos sitios de enzimas de restricción (RE, por sus siglas en inglés), que permiten una sustitución fácil de estos elementos. En la SEQ ID NO: 130, el orden de los elementos es como sigue:

(5'ITR de AAV2) - RE - (promotor HCB) - RE - (intrón de MVM) - RE - Kozak - (región codificante de ET3) - RE - (señal de poliadenilación) - RE - (3'ITR de AAV2)

AAV2-HCB-ET3-LCO-NCG-SpA

Elemento	Inicio (pb)	Final (pb)
5' ITR de AAV2	1	141
AgeI (sitio de restricción)	142	147
promotor HCB	148	293
Sall + PacI (sitios de restricción)	294	307
intrón de MVM	308	399
XhoI (sitio de restricción)	400	405
Secuencia consenso Kozak	406	410
ET3 optimizada para hígado sin CpG	411	4814
NotI (sitio de restricción)	4815	4822
Señal de poliA de beta globina de conejo	4823	4871
MunI (sitio de restricción)	4872	4877
3' ITR de AAV2	4878	5018

El intrón y los sitios de restricción pueden eliminarse opcionalmente del casete para proporcionar un genoma de AAV recombinante acortado. Adicionalmente, el transgén se puede sustituir según sea necesario. La eliminación del intrón y de los elementos de los sitios de restricción generaría un vector de 4885 pares de bases (ET3) o 4855 (HSQ).

- 5 Se generaron partículas de virus AAV2-HCB-ET3-LCO-NCG-SpA y se usaron para transducir ratones. La actividad de fVIII en suero se ensayó en varios puntos temporales después de la transducción (véase la FIG. 19). El ensayo de terapia génica *in vivo* se realizó sustancialmente como se describe en Brown et al. ("Bioengineered Factor FVIII Enables Long-Term Correction of Murine Hemophilia A Following Liver-Directed Adeno-Associated Viral Vector Delivery", Molecular Therapy - Methods and Clinical Development. 1:14036, 2014).
- 10

SEQ ID NO: 130:

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCTGGGCGACCTTTGGTCGCC
CGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTACCGGTGTTAATCAT
TAAGTCGTAAATTTTTGTGGCCCTTGCGATGTTTGCTCTGGTTAATAATCTCAGGACAAACAGAGGTTAATAATTTTC
CAGATCTCTCTGAGCAATAGTATAAAAGGCCAGCAGCAGCCTGACCACATCTCATCCTCGTCGACTTAATTAAGAG
GTAAGGGTTTAAGGGATGGTTGGTTGGTTGGGTTAATGTTTAATTACCTGGAGCACCTGCCTGAAATCACTTTTTT
TCAGGTTGGCTCGAGCCACCATGCAGCTGGAAGTGTCTACCTGTGTGTTTCTGTGTCTGCTGCCTCTGGGGTTTTCTG
CTATCAGGAGATACTATCTGGGAGCTGTGGAGCTGTCTGGGACTACAGGCAGTCTGAGCTGCTGAGAGAAGTGCATG
TGGATACCAGATTCCCAGCTACAGCTCCAGGAGCTCTGCCTCTGGGCCCATCTGTGCTGTACAAGAAAACAGTCTTTG
TGGAGTTTACAGACCAGCTGTTCTCTGTGGCCAGGCCAAGACCACCTTGGATGGGACTGCTGGGACCAACCATCCAGG
CTGAGGTGTATGATACAGTGGTGGTGACCTGAAAAACATGGCCTCCCATCCTGTGAGCCTGCATGCTGTGGGGGTGT
CCTTCTGGAAGTCCTCTGAGGGAGCTGAGTATGAAGACCATACTCCAGAGGGAGAAAGAAGATGATAAGGTGCTGC
CTGGCAAAGCCAGACCTATGTCTGGCAGGTGCTGAAGGAGAATGGACCAACTGCTTCTGACCCACCATGCCTGACCT
ACTCTTATCTGTCCCATGTGGATCTGGTGAAGGACCTGAATTCTGGACTGATTGGAGCTCTGCTGGTGTGTAGAGAGG
GAAGCCTGACCAGAGAAAGAACCCAGAACCTGCATGAGTTTGTCTGCTGTTTGTGTGTTGATGAAGGGAAGAGCT
GGCACTCTGCCAGGAATGACTCCTGGACCAGAGCTATGGATCCAGCTCCTGCTAGAGCTCAGCCTGCTATGCACACAG
TCAATGGCTATGTGAATAGGTCTCTGCCAGGACTGATTGGCTGCCATAAGAAATCTGTCTATTGGCATGTGATTGGAA
TGGGACACCAGCCCTGAGGTGCATTCTATCTTCTGGAAGGCCACACCTTTCTGGTCAGGCACCATAGACAGGCCTCTC

TGGAGATCTCCCCTCTGACCTTCCTGACAGCTCAGACCTTTCTGATGGACCTGGGGCAGTTCTCTGCTGTTTTGCCATA
TCTCTTCCCACCATCATGGAGGAATGGAGGCTCATGTGAGGGTGGAAATCCTGTGCTGAGGAACACAGCTGAGAAGAA
AGGCTGATGAGGAAGAGGACTATGATGATAACCTGTATGACTCTGATATGGATGTGGTGAGGCTGGATGGGGATGATG
TCAGCCCTTTTCATCCAGATCAGGTCTGTGGCCAAGAAACATCCAAAGACCTGGGTCCACTACATTGCTGCTGAAGAGG
AAGATTGGGACTATGCCCCCTGGTGTGGCTCCTGATGATAGATCCTACAAAAGCCAGTATCTGAACAATGGGCCCC
AGAGGATTGGAAGGAAGTACAAGAAAGTGAGGTTTCATGGCCTATACAGATGAGACCTTTAAGACCAGAGAGGCTATCC
AGCATGAATCTGGGATCCTGGGACCTCTGCTGTATGGAGAAGTGGGGGATACCCTGCTGATCATCTTCAAGAACCAGG
CCTCCAGGCCATACAAATATCTATCCCCATGGCATCACAGATGTGAGACCACTGTACAGCAGGAGACTGCCCAAGGGGG
TCAAAACACCTGAAGGATTTCCCCATCCTGCCTGGAGAGATCTTTAAGTATAAATGGACAGTCACAGTGGAAGATGGGC
CTACCAAGTCTGATCCAAGGTGCCTGACCAGATACTATAGCTCTTTTGTGAACATGGAGAGAGACCTGGCTTCTGGAC
TGATTGGACCCCTGTGCTGTTACAAAGAGTCTGTGGACCAGAGGGGCAACCAGATCATGTCTGATAAGAGAAATG
TCATCCTGTTCTCTGTGTTTATGATGAGAACAGGAGCTGGTACCTGACAGAGAACATCCAGAGGTTCTGCCAAATCCAG
CTGGAGTGCAGCTGGAGGACCCAGAATTTTCAGGCTTCCAACATCATGCATAGCATCAATGGCTATGTGTTTGATAGCC
TGCAGCTGTCTGTCTGCCTGCATGAGGTGGCCTACTGGTATATCCTGTCCATTGGAGCTCAGACAGACTTCTGTCTG
TGTCTTTAGTGGGTACACCTTTAAGCATAAAATGGTGTATGAGGATACCCTGACCCTGTCCCCTTTTCTGGGAGA
CAGTGTTCATGTCCATGGAAAACCTGGCCTGTGGATCCTGGGGTGCCACAACCTCTGACTTCAGGAATAGAGGAATGA
CAGCCCTGCTGAAAGTGTCCAGCTGTGATAAGAATACAGGGGATTACTATGAGGACTCTTATGAAGATATCTCTGCTT
ATCTGCTGAGCAAGAACAATGCCATTGAGCCAGGTCTTTTGTCTCAGAACTCCAGACCTCCATCTGCTTCTGCTCCTA
AGCCACCTGTGCTGAGAGACATCAGAGGGACATCTCCCTGCTTACCTTCCAGCCAGAGGAAGATAAAATGGAGATATG
ATGATATCTTCAGCACAGAGACCAAGGGGGAAGATTTTGACATCTATGGAGAGGATGAAAACCAGGATCCAAGATCCT
TCCAGAAGAGAACCAGACACTACTTTATTGCTGCTGTGGAGCAGCTGTGGGACTATGGGATGTCTGAAAGCCCAAGGG
CCCTGAGGAACAGAGCTCAGAATGGAGAGGTGCCAGATTCAAGAAAGTGGTGTTCAGAGAGTTTGCTGATGGCAGCT
TTACCCAGCCATCTTACAGGGGGGAGCTGAACAAGCATCTGGGGCTGCTGGGACCCATATCAGAGCTGAGGTGGAAG
ATAACATCATGGTGACCTTCAAGAATCAGGCTTCTAGGCCCTACTCCTTTTATTCTTCCCTGATCTCCTACCCTGATG
ATCAGGAGCAGGGAGCTGAACCTAGGCACAACCTTTGTGCAGCCAAATGAGACCAGAACCTACTTTTGAAGGTGCAGC
ATCACATGGCTCCCACAGAGGATGAATTTGACTGCAAAAGCTTGGGCCTATTTTCTGATGTGGACCTGGAGAAGGATG
TGCATTCTGGCCTGATTGGGCCTCTGCTGATCTGTAGGGCCAACACCCTGAATGCTGCTCATGGAAGACAGGTCACAG
TGCAGGAGTTTGCTCTGTTCTTTACCATCTTTGATGAAACCAAGAGCTGGTACTTCACAGAGAATGTGGAAAGGAATT
GCAGAGCCCCCTGTCTATCTGCAGATGGAGGACCTACCCTGAAGGAAAACCTACAGGTTCCATGCCATCAATGGATATG
TCATGGATACCCTGCCCTGGCCTGGTCTATGGCTCAGAACCAGAGGATCAGATGGTACCTGCTGTCTATGGGATCCAATG
AGAATATCCATAGCATCCACTTCTCTGGCCATGTCTTTTCTGTGAGGAAGAAAGAGGAATACAAAATGGCTGTGTACA
ATCTGTATCCTGGGGTCTTTGAGACAGTGGAAATGCTGCCAAGCAAAGTGGGAATCTGGAGAATTGAGTGCCTGATTG
GGGAACACCTGCAGGCTGGGATGAGCACCACCTTCTGGTGTACTCTAAGAAATGTCAGACCCCACTGGGGATGGCCT
CTGGACATATCAGGGACTTCCAGATCACAGCTTCTGGACAGTATGGACAGTGGGCTCCAAGCTGGCTAGACTGCACT
ATTCTGGCTCCATCAATGCCTGGTCTACCAAAAGGCCATTCTCCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCA
TCCATGGAATCAAAACCCAGGGAGCTAGGCAGAAGTTTCAGCTCTCTGTACATCTCCAGTTTATCATCATGTATAGCC
TGGATGGGAAGAAATGGCAGACCTACAGAGGCAATTCAGTGGGACCTGATGGTCTTCTTTGGAATGTGGATTCTT
CTGGCATCAAGCACACATCTTCAATCCACCCATCATTGCCAGGTACATCAGGCTGCATCCTACCCACTATAGCATCA
GGTCTACCCTGAGAAATGGAGCTGATGGGATGTGACCTGAACAGCTGTCTATGCCACTGGGCATGGAGTCCAAGGCTA
TCTCTGATGCCAGATCACAGCTTCTTCTACTTCACCAATATGTTTGTACTCTGGTCCCAAGCAAGGCTAGACTGC
ACCTGCAGGGAAGCAATGCTTGGAGACCCAGTCAAGTGAAGTGGCTGCAAGGTGGACTTCCAGAAAA
CCATGAAGGTACAGGGGTGACCACCCAGGGAGTGAAATCTCTGCTGACCTCCATGTATGTCAAGGAGTTCTCTGATCA
GCTCTTCCAGGATGGCCACCAGTGGACCTGTTCTTTTCAAGTGGCAAGGTCAAAGTGTTCAGGGGAATCAGGACT
CTTTTACCCCACTGGTGAACCTCCCTGGATCCTCCACTGCTGACCAGGTACCTGAGAATCCATCCTCAGAGCTGGGTGC
ACCAGATTGCTCTGAGAATGGAGGTCTGGGATGTGAAGCTCAGGACCTGTATTGAGCGCCGCAATAAAATATCTTT
ATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGCAATTGAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCT
GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACAAAGGTGCGCCGACGCGCGGGCTTTGCCCGGGCGGCTCAGTGAG
CGAGCGAGCGCGCAGCTGCCTGCAGG

Ejemplo 4

5 Tratamiento de la hemofilia A humana usando terapia génica basada en AAV

Este ejemplo describe un método ejemplar para el uso clínico de vectores AAV que codifican fVIII para el tratamiento de la hemofilia A.

- 10 Se selecciona un paciente diagnosticado con hemofilia A para el tratamiento. Al paciente se le administra una cantidad terapéuticamente eficaz de un AAV recombinante que codifica la variante de ET3 o HSQ fVIII, tal como AAV-ET3 o AAV-HSQ bajo el control de un promotor HCB como se divulga en el presente documento. El AAV recombinante puede administrarse por vía intravenosa. El médico puede seleccionar una dosis terapéutica adecuada. En algunos casos, la dosis terapéuticamente eficaz está en el intervalo de 1×10^{11} a 1×10^{14} partículas víricas (pv)/kg, tal como
- 15 aproximadamente 1×10^{12} pv/kg. En la mayoría de los casos, al paciente se le administra una dosis única. La salud del sujeto se puede monitorizar con el tiempo para determinar la eficacia del tratamiento.

LISTADO DE SECUENCIAS

<110> Emory University
 CHILDREN'S HEALTHCARE OF ATLANTA, INC.
 Doering, Christopher
 5 Spencer, H Trent
 Brown, Harrison C

 <120> PROMOTORES Y VECTORES RECOMBINANTES PARA LA EXPRESIÓN DE PROTEÍNAS EN EL
 HÍGADO Y USOS DE LOS MISMOS
 10
 <130> 6975-96747-04

 <150> US 62/202.133
 <151> 06/08/2015
 15
 <150> US 62/212.634
 <151> 01/09/2015

 <150> US 62/148.696
 20 <151> 16/04/2015

 <160> 131

 <170> PatentIn versión 3.5
 25
 <210> 1
 <211> 2351
 <212> PRT
 <213> *Homo sapiens*
 30
 <400> 1

ES 2 763 551 T3

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

ES 2 763 551 T3

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
 165 170 175

Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
 195 200 205

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

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
 275 280 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
 290 295 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
 305 310 315 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
 325 330 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
 340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser

ES 2 763 551 T3

370						375						380			
Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr
385					390					395					400
Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro
				405					410					415	
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn
			420					425					430		
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met
		435					440					445			
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu
	450					455					460				
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu
465					470					475					480
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro
				485					490					495	
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys
			500					505					510		
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe
		515					520					525			
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp
	530					535					540				
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg
545					550					555					560
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu
				565					570					575	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val
			580					585					590		
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu
		595					600					605			
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp
	610					615					620				

ES 2 763 551 T3

Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val	625	630	635	640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp	645	650	655	
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe	660	665	670	
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr	675	680	685	
Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro	690	695	700	
Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly	705	710	715	720
Met	Thr	Ala	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp	725	730	735	
Tyr	Tyr	Glu	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys	740	745	750	
Asn	Asn	Ala	Ile	Glu	Pro	Arg	Ser	Phe	Ser	Gln	Asn	Ser	Arg	His	Pro	755	760	765	
Ser	Thr	Arg	Gln	Lys	Gln	Phe	Asn	Ala	Thr	Thr	Ile	Pro	Glu	Asn	Asp	770	775	780	
Ile	Glu	Lys	Thr	Asp	Pro	Trp	Phe	Ala	His	Arg	Thr	Pro	Met	Pro	Lys	785	790	795	800
Ile	Gln	Asn	Val	Ser	Ser	Ser	Asp	Leu	Leu	Met	Leu	Leu	Arg	Gln	Ser	805	810	815	
Pro	Thr	Pro	His	Gly	Leu	Ser	Leu	Ser	Asp	Leu	Gln	Glu	Ala	Lys	Tyr	820	825	830	
Glu	Thr	Phe	Ser	Asp	Asp	Pro	Ser	Pro	Gly	Ala	Ile	Asp	Ser	Asn	Asn	835	840	845	
Ser	Leu	Ser	Glu	Met	Thr	His	Phe	Arg	Pro	Gln	Leu	His	His	Ser	Gly	850	855	860	
Asp	Met	Val	Phe	Thr	Pro	Glu	Ser	Gly	Leu	Gln	Leu	Arg	Leu	Asn	Glu	865	870	875	880

ES 2 763 551 T3

Lys Leu Gly Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys
 885 890 895
 Val Ser Ser Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn
 900 905 910
 Leu Ala Ala Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met
 915 920 925
 Pro Val His Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys
 930 935 940
 Ser Ser Pro Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu
 945 950 955 960
 Asn Asn Asp Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu
 965 970 975
 Ser Ser Trp Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe
 980 985 990
 Lys Gly Lys Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala
 995 1000 1005
 Leu Phe Lys Val Ser Ile Ser Leu Leu Lys Thr Asn Lys Thr Ser
 1010 1015 1020
 Asn Asn Ser Ala Thr Asn Arg Lys Thr His Ile Asp Gly Pro Ser
 1025 1030 1035
 Leu Leu Ile Glu Asn Ser Pro Ser Val Trp Gln Asn Ile Leu Glu
 1040 1045 1050
 Ser Asp Thr Glu Phe Lys Lys Val Thr Pro Leu Ile His Asp Arg
 1055 1060 1065
 Met Leu Met Asp Lys Asn Ala Thr Ala Leu Arg Leu Asn His Met
 1070 1075 1080
 Ser Asn Lys Thr Thr Ser Ser Lys Asn Met Glu Met Val Gln Gln
 1085 1090 1095
 Lys Lys Glu Gly Pro Ile Pro Pro Asp Ala Gln Asn Pro Asp Met
 1100 1105 1110
 Ser Phe Phe Lys Met Leu Phe Leu Pro Glu Ser Ala Arg Trp Ile
 1115 1120 1125

ES 2 763 551 T3

Gln Arg	Thr His	Gly Lys	Asn	Ser Leu	Asn Ser	Gly	Gln Gly	Pro
1130			1135			1140		
Ser Pro	Lys Gln	Leu Val	Ser	Leu Gly	Pro Glu	Lys	Ser Val	Glu
1145			1150			1155		
Gly Gln	Asn Phe	Leu Ser	Glu	Lys Asn	Lys Val	Val	Val Gly	Lys
1160			1165			1170		
Gly Glu	Phe Thr	Lys Asp	Val	Gly Leu	Lys Glu	Met	Val Phe	Pro
1175			1180			1185		
Ser Ser	Arg Asn	Leu Phe	Leu	Thr Asn	Leu Asp	Asn	Leu His	Glu
1190			1195			1200		
Asn Asn	Thr His	Asn Gln	Glu	Lys Lys	Ile Gln	Glu	Glu Ile	Glu
1205			1210			1215		
Lys Lys	Glu Thr	Leu Ile	Gln	Glu Asn	Val Val	Leu	Pro Gln	Ile
1220			1225			1230		
His Thr	Val Thr	Gly Thr	Lys	Asn Phe	Met Lys	Asn	Leu Phe	Leu
1235			1240			1245		
Leu Ser	Thr Arg	Gln Asn	Val	Glu Gly	Ser Tyr	Asp	Gly Ala	Tyr
1250			1255			1260		
Ala Pro	Val Leu	Gln Asp	Phe	Arg Ser	Leu Asn	Asp	Ser Thr	Asn
1265			1270			1275		
Arg Thr	Lys Lys	His Thr	Ala	His Phe	Ser Lys	Lys	Gly Glu	Glu
1280			1285			1290		
Glu Asn	Leu Glu	Gly Leu	Gly	Asn Gln	Thr Lys	Gln	Ile Val	Glu
1295			1300			1305		
Lys Tyr	Ala Cys	Thr Thr	Arg	Ile Ser	Pro Asn	Thr	Ser Gln	Gln
1310			1315			1320		
Asn Phe	Val Thr	Gln Arg	Ser	Lys Arg	Ala Leu	Lys	Gln Phe	Arg
1325			1330			1335		
Leu Pro	Leu Glu	Glu Thr	Glu	Leu Glu	Lys Arg	Ile	Ile Val	Asp
1340			1345			1350		
Asp Thr	Ser Thr	Gln Trp	Ser	Lys Asn	Met Lys	His	Leu Thr	Pro

ES 2 763 551 T3

1355		1360		1365
Ser Thr Leu Thr Gln Ile Asp Tyr Asn Glu Lys Glu Lys Gly Ala				
1370		1375		1380
Ile Thr Gln Ser Pro Leu Ser Asp Cys Leu Thr Arg Ser His Ser				
1385		1390		1395
Ile Pro Gln Ala Asn Arg Ser Pro Leu Pro Ile Ala Lys Val Ser				
1400		1405		1410
Ser Phe Pro Ser Ile Arg Pro Ile Tyr Leu Thr Arg Val Leu Phe				
1415		1420		1425
Gln Asp Asn Ser Ser His Leu Pro Ala Ala Ser Tyr Arg Lys Lys				
1430		1435		1440
Asp Ser Gly Val Gln Glu Ser Ser His Phe Leu Gln Gly Ala Lys				
1445		1450		1455
Lys Asn Asn Leu Ser Leu Ala Ile Leu Thr Leu Glu Met Thr Gly				
1460		1465		1470
Asp Gln Arg Glu Val Gly Ser Leu Gly Thr Ser Ala Thr Asn Ser				
1475		1480		1485
Val Thr Tyr Lys Lys Val Glu Asn Thr Val Leu Pro Lys Pro Asp				
1490		1495		1500
Leu Pro Lys Thr Ser Gly Lys Val Glu Leu Leu Pro Lys Val His				
1505		1510		1515
Ile Tyr Gln Lys Asp Leu Phe Pro Thr Glu Thr Ser Asn Gly Ser				
1520		1525		1530
Pro Gly His Leu Asp Leu Val Glu Gly Ser Leu Leu Gln Gly Thr				
1535		1540		1545
Glu Gly Ala Ile Lys Trp Asn Glu Ala Asn Arg Pro Gly Lys Val				
1550		1555		1560
Pro Phe Leu Arg Val Ala Thr Glu Ser Ser Ala Lys Thr Pro Ser				
1565		1570		1575
Lys Leu Leu Asp Pro Leu Ala Trp Asp Asn His Tyr Gly Thr Gln				
1580		1585		1590

ES 2 763 551 T3

Ile	Pro	Lys	Glu	Glu	Trp	Lys	Ser	Gln	Glu	Lys	Ser	Pro	Glu	Lys
1595						1600					1605			
Thr	Ala	Phe	Lys	Lys	Lys	Asp	Thr	Ile	Leu	Ser	Leu	Asn	Ala	Cys
1610						1615					1620			
Glu	Ser	Asn	His	Ala	Ile	Ala	Ala	Ile	Asn	Glu	Gly	Gln	Asn	Lys
1625						1630					1635			
Pro	Glu	Ile	Glu	Val	Thr	Trp	Ala	Lys	Gln	Gly	Arg	Thr	Glu	Arg
1640						1645					1650			
Leu	Cys	Ser	Gln	Asn	Pro	Pro	Val	Leu	Lys	Arg	His	Gln	Arg	Glu
1655						1660					1665			
Ile	Thr	Arg	Thr	Thr	Leu	Gln	Ser	Asp	Gln	Glu	Glu	Ile	Asp	Tyr
1670						1675					1680			
Asp	Asp	Thr	Ile	Ser	Val	Glu	Met	Lys	Lys	Glu	Asp	Phe	Asp	Ile
1685						1690					1695			
Tyr	Asp	Glu	Asp	Glu	Asn	Gln	Ser	Pro	Arg	Ser	Phe	Gln	Lys	Lys
1700						1705					1710			
Thr	Arg	His	Tyr	Phe	Ile	Ala	Ala	Val	Glu	Arg	Leu	Trp	Asp	Tyr
1715						1720					1725			
Gly	Met	Ser	Ser	Ser	Pro	His	Val	Leu	Arg	Asn	Arg	Ala	Gln	Ser
1730						1735					1740			
Gly	Ser	Val	Pro	Gln	Phe	Lys	Lys	Val	Val	Phe	Gln	Glu	Phe	Thr
1745						1750					1755			
Asp	Gly	Ser	Phe	Thr	Gln	Pro	Leu	Tyr	Arg	Gly	Glu	Leu	Asn	Glu
1760						1765					1770			
His	Leu	Gly	Leu	Leu	Gly	Pro	Tyr	Ile	Arg	Ala	Glu	Val	Glu	Asp
1775						1780					1785			
Asn	Ile	Met	Val	Thr	Phe	Arg	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Ser
1790						1795					1800			
Phe	Tyr	Ser	Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly
1805						1810					1815			
Ala	Glu	Pro	Arg	Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr
1820						1825					1830			

ES 2 763 551 T3

Tyr	Phe	Trp	Lys	Val	Gln	His	His	Met	Ala	Pro	Thr	Lys	Asp	Glu
1835						1840					1845			
Phe	Asp	Cys	Lys	Ala	Trp	Ala	Tyr	Phe	Ser	Asp	Val	Asp	Leu	Glu
1850						1855					1860			
Lys	Asp	Val	His	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Val	Cys	His
1865						1870					1875			
Thr	Asn	Thr	Leu	Asn	Pro	Ala	His	Gly	Arg	Gln	Val	Thr	Val	Gln
1880						1885					1890			
Glu	Phe	Ala	Leu	Phe	Phe	Thr	Ile	Phe	Asp	Glu	Thr	Lys	Ser	Trp
1895						1900					1905			
Tyr	Phe	Thr	Glu	Asn	Met	Glu	Arg	Asn	Cys	Arg	Ala	Pro	Cys	Asn
1910						1915					1920			
Ile	Gln	Met	Glu	Asp	Pro	Thr	Phe	Lys	Glu	Asn	Tyr	Arg	Phe	His
1925						1930					1935			
Ala	Ile	Asn	Gly	Tyr	Ile	Met	Asp	Thr	Leu	Pro	Gly	Leu	Val	Met
1940						1945					1950			
Ala	Gln	Asp	Gln	Arg	Ile	Arg	Trp	Tyr	Leu	Leu	Ser	Met	Gly	Ser
1955						1960					1965			
Asn	Glu	Asn	Ile	His	Ser	Ile	His	Phe	Ser	Gly	His	Val	Phe	Thr
1970						1975					1980			
Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys	Met	Ala	Leu	Tyr	Asn	Leu	Tyr
1985						1990					1995			
Pro	Gly	Val	Phe	Glu	Thr	Val	Glu	Met	Leu	Pro	Ser	Lys	Ala	Gly
2000						2005					2010			
Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile	Gly	Glu	His	Leu	His	Ala	Gly
2015						2020					2025			
Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr	Ser	Asn	Lys	Cys	Gln	Thr	Pro
2030						2035					2040			
Leu	Gly	Met	Ala	Ser	Gly	His	Ile	Arg	Asp	Phe	Gln	Ile	Thr	Ala
2045						2050					2055			
Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro	Lys	Leu	Ala	Arg	Leu	His
2060						2065					2070			

ES 2 763 551 T3

Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser	Thr	Lys	Glu	Pro	Phe	Ser
2075						2080					2085			
Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro	Met	Ile	Ile	His	Gly	Ile
2090						2095					2100			
Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys	Phe	Ser	Ser	Leu	Tyr	Ile	Ser
2105						2110					2115			
Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu	Asp	Gly	Lys	Lys	Trp	Gln	Thr
2120						2125					2130			
Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu	Met	Val	Phe	Phe	Gly	Asn
2135						2140					2145			
Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn	Ile	Phe	Asn	Pro	Pro	Ile
2150						2155					2160			
Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro	Thr	His	Tyr	Ser	Ile	Arg
2165						2170					2175			
Ser	Thr	Leu	Arg	Met	Glu	Leu	Met	Gly	Cys	Asp	Leu	Asn	Ser	Cys
2180						2185					2190			
Ser	Met	Pro	Leu	Gly	Met	Glu	Ser	Lys	Ala	Ile	Ser	Asp	Ala	Gln
2195						2200					2205			
Ile	Thr	Ala	Ser	Ser	Tyr	Phe	Thr	Asn	Met	Phe	Ala	Thr	Trp	Ser
2210						2215					2220			
Pro	Ser	Lys	Ala	Arg	Leu	His	Leu	Gln	Gly	Arg	Ser	Asn	Ala	Trp
2225						2230					2235			
Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys	Glu	Trp	Leu	Gln	Val	Asp	Phe
2240						2245					2250			
Gln	Lys	Thr	Met	Lys	Val	Thr	Gly	Val	Thr	Thr	Gln	Gly	Val	Lys
2255						2260					2265			
Ser	Leu	Leu	Thr	Ser	Met	Tyr	Val	Lys	Glu	Phe	Leu	Ile	Ser	Ser
2270						2275					2280			
Ser	Gln	Asp	Gly	His	Gln	Trp	Thr	Leu	Phe	Phe	Gln	Asn	Gly	Lys
2285						2290					2295			
Val	Lys	Val	Phe	Gln	Gly	Asn	Gln	Asp	Ser	Phe	Thr	Pro	Val	Val

ES 2 763 551 T3

2300

2305

2310

Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His
2315 2320 2325

Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu Val Leu
2330 2335 2340

Gly Cys Glu Ala Gln Asp Leu Tyr
2345 2350

<210> 2

<211> 4374

<212> ADN

<213> Secuencia artificial

<220>

<223> fVIII recombinante

<400> 2

ES 2 763 551 T3

atgcagattg aactgtctac ctgtttcttt ctgtgcctgc tgaggttttg tttttctgct	60
accagaagat actacctggg agctgtggaa ctgagctggg attacatgca gtctgacctg	120
ggagagctgc ctgtggatgc tagattccca cctagagtcc ctaagtcctt ccccttcaac	180
acctctgtgg tctacaagaa aaccctgttt gtggagttta cagaccacct gttcaacatt	240
gctaagccta gaccaccatg gatgggactg ctgggaccaa ccatccaggc agaggtgtat	300
gacacagtgg tcatcaccct gaaaaacatg gcttctcacc ctgtgtccct gcatgctgtg	360
ggagtctcct actggaaggc ctctgaaggg gctgagtatg atgatcagac cagccagagg	420
gaaaaagagg atgataaggt gttccctgga ggggccata cctatgtgtg gcaggtcctg	480
aaggagaatg gaccaatggc ttctgaccct ctgtgcctga cctactctta tctgtcccat	540
gtggacctgg tcaaggatct gaactctggc ctgattgggg ctctgctggg gtgtagggaa	600
gggtccctgg ccaaggagaa aaccagacc ctgcataagt tcatcctgct gtttgctgtg	660
tttgatgaag gaaaaagctg gcactctgag accaagaact ctctgatgca ggacagggat	720
gctgcttctg ccagagcttg gcccaagatg cacacagtga atggctatgt caataggagc	780
ctgcctggac tgattggctg ccacagaaag tctgtgtatt ggcatgtcat tggaatgggc	840
accaccctg aagtgcacag catcttcctg gaggggcata ctttctggt caggaaccac	900
aggcaggcta gcctggagat ctctccaatc accttcctga cagcccagac cctgctgatg	960
gacctgggac agttcctgct gttttgccac atctccagcc accagcatga tggcatggag	1020
gcttatgtga aagtggactc ctgtcctgag gaacctcagc tgaggatgaa gaacaatgag	1080
gaagctgaag actatgatga tgacctgaca gactctgaga tggatgtggg caggtttgat	1140
gatgataact ctccctcctt tatccagatc aggtctgtgg ccaagaaaca ccctaagacc	1200

ES 2 763 551 T3

tgggtccatt	acattgctgc	tgaggaagag	gactgggatt	atgctccact	ggtgctggcc	1260
cctgatgata	gacccataca	aagccagtat	ctgaacaatg	gaccccagag	gattggcaga	1320
aagtacaaga	aagtgaggtt	catggcttat	acagatgaga	cctttaagac	cagagaagcc	1380
atccagcatg	agtctgggat	cctgggacct	ctgctgtatg	gggaagtggg	ggacaccctg	1440
ctgatcatct	tcaagaacca	ggccagcagg	ccttacaata	tctatccaca	tggcatcaca	1500
gatgtgagac	ctctgtactc	caggaggctg	ccaaaggggg	tgaaacacct	gaaggacttc	1560
ccaatcctgc	ctggggaaat	ctttaagtat	aaatggacag	tcacagtgga	ggatgggccc	1620
accaagtctg	accctaggtg	cctgaccaga	tactattctt	cctttgtgaa	tatggagaga	1680
gacctggctt	ctggactgat	tggaccctg	ctgatctgtt	acaaagagtc	tgtggatcag	1740
aggggcaacc	agatcatgtc	tgacaagagg	aatgtgatcc	tgttctctgt	ctttgatgaa	1800
aacaggtctt	ggtacctgac	agagaacatc	cagaggttcc	tgccaatcc	agctggagtg	1860
cagctggaag	atcctgagtt	ccaggcctct	aacatcatgc	attccatcaa	tggctatgtg	1920
tttgactccc	tgcagctgtc	tgtgtgcctg	catgaggtgg	cttactggta	tatcctgagc	1980
attggagccc	agacagattt	cctgtctgtg	ttcttttctg	gctacacctt	taagcataaa	2040
atggtgtatg	aggacaccct	gaccctgttc	ccattttctg	gagaaactgt	gttcatgagc	2100
atggagaatc	ctgggctgtg	gacccctggg	tgccacaact	ctgatttcag	gaatagaggg	2160
atgacagccc	tgctgaaagt	gagctcttgt	gacaagaaca	caggagacta	ctatgaagat	2220
agctatgagg	acatctctgc	ttatctgctg	tccaaaaaca	atgccattga	gcccaggagc	2280
ttctctcaga	accctccagt	gctgaagagg	caccagaggg	agatcaccag	aaccaccctg	2340
cagtctgatc	aggaagagat	tgactatgat	gataccatct	ctgtggaaat	gaagaaagag	2400
gactttgata	tctatgatga	agatgagaac	cagtctccca	ggtccttcca	gaagaaaacc	2460
agacattact	ttattgctgc	tgtggagagg	ctgtgggact	atggcatgtc	cagctctcct	2520
catgtgctga	gaaatagagc	tcagtctgga	tctgtccac	agttcaagaa	agtggctctc	2580
caggagttta	cagatggaag	ctttaccag	ccactgtaca	ggggagaact	gaatgagcac	2640
ctggggctgc	tgggacccta	tatcagggct	gaagtggagg	ataacatcat	ggtcaccttc	2700
aggaatcagg	ccagcagacc	ctactctttt	tattccagcc	tgatctccta	tgaagaggac	2760
cagagacagg	gagctgaacc	aagaaaaaac	tttgtgaagc	ctaagagagc	caaaacctac	2820
ttttggaagg	tgcagcacca	tatggccctt	accaaagatg	agtttgattg	caaggcctgg	2880
gcttattttt	ctgatgtgga	tctggagaag	gatgtccact	ctggcctgat	tgggccactg	2940
ctggtgtgtc	ataccaacac	cctgaatcca	gctcatggaa	ggcaggtgac	agtccaggaa	3000
tttgcctgtg	tctttaccat	ctttgatgag	accaagagct	ggtacttcac	agaaaacatg	3060

ES 2 763 551 T3

```

gagaggaatt gcagagcccc atgtaacatc cagatggaag accccacctt caaggagaac 3120
tacagatttc atgctatcaa tgggtatatc atggataccc tgccaggact ggtcatggct 3180
caggaccaga ggatcagatg gtacctgctg agcatggggt ctaatgagaa tatccactcc 3240
atccatttct ctggacatgt gtttacagta aggaagaaag aagagtacaa gatggccctg 3300
tacaacctgt atcctggggt gtttgaacaa gtggagatgc tgccttccaa ggctgggatc 3360
tggaggggtg aatgcctgat tggggagcac ctgcatgctg gaatgtctac cctgttcctg 3420
gtgtactcca ataagtgtca gacccccctg gggatggctt ctggacatat cagggacttc 3480
cagatcacag cttctggaca gtatggacag tgggctccta agctggctag actgcactat 3540
tctggctcca tcaatgcttg gtctacaaa gagcctttct cctggatcaa ggtggacctg 3600
ctggctccaa tgatcatcca tggcatcaaa acccaggggg ccaggcagaa gttctcttcc 3660
ctgtacatca gccagtttat catcatgtat tctctggatg ggaagaaatg gcagacctac 3720
agaggcaatt ccacagggac cctgatggtg ttctttggca atgtggacag ctctgggatc 3780
aagcacaaca tcttcaatcc ccctatcatt gccaggatca tcagactgca cccaacccat 3840
tattccatca ggagcaccct gagaatggag ctgatggggt gtgatctgaa cagctgttct 3900
atgccctggt gaatggagtc taaggccatc tctgatgctc agatcacagc ctccagctac 3960
ttcaccaata tgtttgctac ctggccccca agcaaggcta gactgcatct gcagggaaga 4020
agcaatgctt ggagaccaca ggtgaacaat occaaggagt ggctgcaggt ggacttccag 4080
aaaaccatga aggtgacagg agtcaccacc caggagtgta aaagcctgct gacctctatg 4140
tatgtcaagg agttcctgat ctcttccagc caggatgggc accagtggac cctgttcttt 4200
cagaatggaa aggtgaaagt cttccagggc aatcaggatt cctttacccc tgtgggtcaac 4260
agcctggacc caccctgct gaccaggtac ctgagaatcc acccacagtc ctgggtgcat 4320
cagattgctc tgaggatgga agtcctgggc tgtgaggccc aggacctgta ttga 4374

```

<210> 3
 <211> 1457
 <212> PRT
 <213> Secuencia artificial

 <220>
 <223> fVIII recombinante

 <400> 3

Met	Gln	Ile	Glu	Leu	Ser	Thr	Cys	Phe	Phe	Leu	Cys	Leu	Leu	Arg	Phe
1				5					10					15	
Cys	Phe	Ser	Ala	Thr	Arg	Arg	Tyr	Tyr	Leu	Gly	Ala	Val	Glu	Leu	Ser
			20					25					30		

ES 2 763 551 T3

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35 40 45
 Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50 55 60
 Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Val His Leu Phe Asn Ile
 65 70 75 80
 Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85 90 95
 Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
 100 105 110
 His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
 115 120 125
 Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140
 Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160
 Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
 165 170 175
 Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190
 Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
 195 200 205
 Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
 210 215 220
 Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240
 Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255
 Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270
 Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
 275 280 285

ES 2 763 551 T3

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
 290 295 300
 Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
 305 310 315 320
 Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
 325 330 335
 Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
 340 345 350
 Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365
 Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
 370 375 380
 Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
 385 390 395 400
 Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
 405 410 415
 Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
 420 425 430
 Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
 435 440 445
 Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
 450 455 460
 Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
 465 470 475 480
 Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
 485 490 495
 His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
 500 505 510
 Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
 515 520 525
 Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
 530 535 540

ES 2 763 551 T3

Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg
545					550					555					560
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu
				565					570					575	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val
			580					585					590		
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu
		595					600					605			
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp
	610					615					620				
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val
625					630					635					640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp
				645					650						655
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe
			660					665						670	
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr
		675					680					685			
Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro
	690					695					700				
Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly
705					710					715					720
Met	Thr	Ala	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp
				725					730					735	
Tyr	Tyr	Glu	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys
			740					745					750		
Asn	Asn	Ala	Ile	Glu	Pro	Arg	Ser	Phe	Ser	Gln	Asn	Pro	Pro	Val	Leu
		755					760					765			
Lys	Arg	His	Gln	Arg	Glu	Ile	Thr	Arg	Thr	Thr	Leu	Gln	Ser	Asp	Gln
	770					775					780				
Glu	Glu	Ile	Asp	Tyr	Asp	Asp	Thr	Ile	Ser	Val	Glu	Met	Lys	Lys	Glu

ES 2 763 551 T3

785		790		795		800
Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe						
		805		810		815
Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp						
		820		825		830
Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln						
		835		840		845
Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr						
		850		855		860
Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His						
		865		870		880
Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile						
		885		890		895
Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser						
		900		905		910
Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg						
		915		920		925
Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val						
		930		935		940
Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp						
		945		950		960
Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu						
		965		970		975
Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His						
		980		985		990
Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe						
		995		1000		1005
Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn						
		1010		1015		1020
Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys						
		1025		1030		1035

ES 2 763 551 T3

Glu	Asn	Tyr	Arg	Phe	His	Ala	Ile	Asn	Gly	Tyr	Ile	Met	Asp	Thr
1040						1045					1050			
Leu	Pro	Gly	Leu	Val	Met	Ala	Gln	Asp	Gln	Arg	Ile	Arg	Trp	Tyr
1055						1060					1065			
Leu	Leu	Ser	Met	Gly	Ser	Asn	Glu	Asn	Ile	His	Ser	Ile	His	Phe
1070						1075					1080			
Ser	Gly	His	Val	Phe	Thr	Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys	Met
1085						1090					1095			
Ala	Leu	Tyr	Asn	Leu	Tyr	Pro	Gly	Val	Phe	Glu	Thr	Val	Glu	Met
1100						1105					1110			
Leu	Pro	Ser	Lys	Ala	Gly	Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile	Gly
1115						1120					1125			
Glu	His	Leu	His	Ala	Gly	Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr	Ser
1130						1135					1140			
Asn	Lys	Cys	Gln	Thr	Pro	Leu	Gly	Met	Ala	Ser	Gly	His	Ile	Arg
1145						1150					1155			
Asp	Phe	Gln	Ile	Thr	Ala	Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro
1160						1165					1170			
Lys	Leu	Ala	Arg	Leu	His	Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser
1175						1180					1185			
Thr	Lys	Glu	Pro	Phe	Ser	Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro
1190						1195					1200			
Met	Ile	Ile	His	Gly	Ile	Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys	Phe
1205						1210					1215			
Ser	Ser	Leu	Tyr	Ile	Ser	Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu	Asp
1220						1225					1230			
Gly	Lys	Lys	Trp	Gln	Thr	Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu
1235						1240					1245			
Met	Val	Phe	Phe	Gly	Asn	Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn
1250						1255					1260			
Ile	Phe	Asn	Pro	Pro	Ile	Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro
1265						1270					1275			

ES 2 763 551 T3

Thr	His	Tyr	Ser	Ile	Arg	Ser	Thr	Leu	Arg	Met	Glu	Leu	Met	Gly
1280						1285					1290			
Cys	Asp	Leu	Asn	Ser	Cys	Ser	Met	Pro	Leu	Gly	Met	Glu	Ser	Lys
1295						1300					1305			
Ala	Ile	Ser	Asp	Ala	Gln	Ile	Thr	Ala	Ser	Ser	Tyr	Phe	Thr	Asn
1310						1315					1320			
Met	Phe	Ala	Thr	Trp	Ser	Pro	Ser	Lys	Ala	Arg	Leu	His	Leu	Gln
1325						1330					1335			
Gly	Arg	Ser	Asn	Ala	Trp	Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys	Glu
1340						1345					1350			
Trp	Leu	Gln	Val	Asp	Phe	Gln	Lys	Thr	Met	Lys	Val	Thr	Gly	Val
1355						1360					1365			
Thr	Thr	Gln	Gly	Val	Lys	Ser	Leu	Leu	Thr	Ser	Met	Tyr	Val	Lys
1370						1375					1380			
Glu	Phe	Leu	Ile	Ser	Ser	Ser	Gln	Asp	Gly	His	Gln	Trp	Thr	Leu
1385						1390					1395			
Phe	Phe	Gln	Asn	Gly	Lys	Val	Lys	Val	Phe	Gln	Gly	Asn	Gln	Asp
1400						1405					1410			
Ser	Phe	Thr	Pro	Val	Val	Asn	Ser	Leu	Asp	Pro	Pro	Leu	Leu	Thr
1415						1420					1425			
Arg	Tyr	Leu	Arg	Ile	His	Pro	Gln	Ser	Trp	Val	His	Gln	Ile	Ala
1430						1435					1440			
Leu	Arg	Met	Glu	Val	Leu	Gly	Cys	Glu	Ala	Gln	Asp	Leu	Tyr	
1445						1450					1455			

<210> 4

<211> 146

5 <212> ADN

<213> Secuencia artificial

<220>

10 <223> secuencia promotora recombinante

<400> 4

gttaatcatt aagtcgttaa tttttgtggc ccttgcatg tttgctctgg ttaataatct	60
caggacaaac agaggttaat aattttccag atctctctga gcaatagtat aaaaggccag	120
cagcagcctg accacatctc atcctc	146

ES 2 763 551 T3

5	<210> 5 <211> 159 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante	
10	<400> 5	
	gttaatTTTT gtggcccttg cgatgtttgc tctggttaat aatctcagga caaacataca	60
	ttttcagtca tatgtttgct cactgaaggt tactagttaa caggcatccc ttaaacagga	120
	tataaaaggc cagcagcagc ctgaccacat ctcaccc	159
15	<210> 6 <211> 172 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante	
20	<400> 6	
	gttaatTTTT aaaaagcagt caaaagtcca agtggccctt gcgagcattt actctctctg	60
	tttgctctgg ttaataatct caggagcaca aacagagggt aataattttc agtcatatgt	120
	ttgctcactg aaggttacta gttaacaggc atcccttaaa caggatataa aa	172
25		
30	<210> 7 <211> 204 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante	
35	<400> 7	
	gttaatTTTT aaaaagcagt caaaagtcca agtggccctt gcgagcattt actctctctg	60
	tttgctctgg ttaataatct caggagcaca aacagagggt aataattttc agtcatatgt	120
	ttgctcactg aaggttacta gttaacaggc atcccttaaa caggatataa aaggccagca	180
	gcagcctgac cacatctcat cctc	204
40	<210> 8 <211> 1386 <212> ADN <213> Secuencia artificial <220> <223> flX recombinante	
45	<400> 8	

ES 2 763 551 T3

```

atgcagaggg tgaacatgat catggctgag tctcctggac tgatcaccat ctgcctgctg      60

ggctatctgc tgtctgctga gtgtacagtg ttcctggacc atgaaaatgc taataaaatc      120
ctgaacaggc caaagaggta caattctggg aaactggagg aatttgtgca gggaaacctg      180
gagaggggaat gcatggagga aaagtgtagc tttgaggaag ccagggaggt gtttgaaaat      240
acagagagga ccacagagtt ctggaaacag tatgtggatg gggatcagtg tgagtccaac      300
ccctgtctga atggaggggc ttgcaaggat gatatcaact cctatgagtg ctggtgtcct      360
tttggatttg aaggcaagaa ttgtgagctg gatgtgacct gtaacatcaa aaatgggagg      420
tgtgagcagt tctgtaagaa ctctgctgat aataaagtgg tctgcagctg tacagaaggc      480
tacaggctgg ctgagaacca gaagagctgt gaaccagctg tgccttccc ttgtgggagg      540
gtgtctgtca gccagaccag caagctgacc agagctgagg ctgtgtttcc tgatgtggat      600
tatgtcaact ctacagaggc tgaaccatc ctggacaaca tcaccagtc taccagctcc      660
ttcaatgact ttaccagggt ggtgggaggg gaggatgcta agccaggaca gttcccctgg      720
caggtggtcc tgaatggcaa agtggatgct ttttgtgggg gctccattgt gaatgagaag      780
tggattgtca cagctgctca ctgtgtggaa actgggggtca agatcacagt ggtggctgga      840
gagcacaaca ttgaggaaac tgaacataca gagcagaaaa ggaatgtgat cagaatcatc      900
ccccaccata actacaatgc tgctatcaac aagtataatc atgacattgc cctgctggaa      960
ctggatgagc ctctggtgct gaacagctat gtcaccccaa tctgcattgc tgacaaggag     1020
tataccaata tcttcctgaa atttgggtct ggatatgtgt ctgggtgggg aagggtcttc     1080
cacaagggaa ggtctgctct ggtgctgcag tatctgaggg tgcccctggt ggacagagct     1140
acctgcctga ggagcaccaa gttcaccatc tacaacaata tgttctgtgc tggatttcat     1200
gagggaggga gggactcctg tcaggagat tctggaggcc ctcatgtgac agaggtggaa     1260
ggcaccagct tcctgactgg catcatctct tggggggagg aatgtgctat gaaggggaaa     1320
tatggaatct ataccaaggt gtccagatat gtcaactgga tcaaggagaa aaccaagctg     1380
acctga                                           1386

```

<210> 9
 <211> 1386
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> flX recombinante

<400> 9

ES 2 763 551 T3

```

atgcagaggg tgaacatgat catggctgag tctcctggac tgatcaccat ctgcctgctg      60
ggctatctgc tgtctgctga gtgtacagtg ttctctggacc atgaaaatgc taataaaatc      120
ctgaacaggc caaagaggta caattctggg aaactggagg aatttgctga gggaaacctg      180
gagagggaat gcatggagga aaagtgtagc tttgaggaag ccagggaggt gtttgaaaat      240

acagagagga ccacagagtt ctggaaacag tatgtggatg gggatcagtg tgagtccaac      300
ccctgtctga atggagggtc ttgcaaggat gatatcaact cctatgagtg ctggtgtcct      360
tttggaattg aaggcaagaa ttgtgagctg gatgtgacct gtaacatcaa aaatgggagg      420
tgtgagcagt tctgtaagaa ctctgctgat aataaagtgg tctgcagctg tacagaaggc      480
tacaggctgg ctgagaacca gaagagctgt gaaccagctg tgcccttccc ttgtgggagg      540
gtgtctgtca gccagaccag caagctgacc agagctgagg ctgtgtttcc tgatgtggat      600
tatgtcaact ctacagaggc tgaaaccatc ctggacaaca tcaccagtc taccagtc      660
ttcaatgact ttaccagggt ggtgggaggg gaggatgcta agccaggaca gttcccctgg      720
caggtggtcc tgaatggcaa agtggatgct ttttgtgggg gctccattgt gaatgagaag      780
tggaattgtca cagctgctca ctgtgtggaa actggggtca agatcacagt ggtggctgga      840
gagcacaaca ttgaggaaac tgaacatata gagcagaaaa ggaatgtgat cagaatcatc      900
ccccaccata actacaatgc tgctatcaac aagtataatc atgacattgc cctgctggaa      960
ctggatgagc ctctggtgct gaacagctat gtcaccccaa tctgcattgc tgacaaggag     1020
tataccaata tcttcctgaa atttgggtct ggatatgtgt ctgggtgggg aagggtcttc     1080
cacaaggga ggtctgctct ggtgctgcag tatctgaggg tgcccctggt ggacagagct     1140
acctgcctgc tgagcaccaa gttcaccatc tacaacaata tgttctgtgc tggatttcat     1200
gagggaggga gggactcctg tcaggagat tctggaggcc ctcatgtgac agagggtgaa     1260
ggcaccagct tcctgactgg catcatctct tggggggagg aatgtgctat gaaggggaaa     1320
tatggaatct ataccaaggt gtccagatat gtcaactgga tcaaggagaa aaccaagctg     1380
acctga                                           1386

```

<210> 10
 <211> 1386
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> flX recombinante

<400> 10

ES 2 763 551 T3

```

atgcagagggg tgaacatgat catggctgag tctcctggac tgatcaccat ctgcctgctg      60
ggctatctgc tgtctgctga gtgtacagtg ttcctggacc atgaaaatgc taataaaatc      120
ctgaacaggc caaagaggta caattctggg aaactggagg aatttgtgca gggaaacctg      180
gagaggggaat gcatggagga aaagtgtagc tttgaggaag ccagggaggt gtttgaaaaat      240
acagagagga ccacagagtt ctggaaacag tatgtggatg gggatcagtg tgagtccaac      300
ccctgtctga atggaggggc ttgcaaggat gatatcaact cctatgagtg ctggtgtcct      360

tttggatttg aaggcaagaa ttgtgagctg gatgtgacct gtaacatcaa aaatgggagg      420
tgtgagcagt tctgtaagaa ctctgctgat aataaagtgg tctgcagctg tacagaaggc      480
tacaggctgg ctgagaacca gaagagctgt gaaccagctg tgcccttccc ttgtgggagg      540
gtgtctgtca gccagaccag caagctgacc agagctgaga cagtgtttcc tgatgtggat      600
tatgtcaact ctacagaggc tgaaaccatc ctggacaaca tcaccagtc taccagtc      660
ttcaatgact ttaccagggt ggtgggaggg gaggatgcta agccaggaca gttcccctgg      720
caggtggtcc tgaatggcaa agtggatgct ttttgtgggg gctccattgt gaatgagaag      780
tggattgtca cagctgctca ctgtgtggaa actgggggtca agatcacagt ggtggctgga      840
gagcacaaca ttgaggaaac tgaacataca gagcagaaaa ggaatgtgat cagaatcatc      900
ccccaccata actacaatgc tgctatcaac aagtataatc atgacattgc cctgctggaa      960
ctggatgagc ctctggtgct gaacagctat gtcaccccaa tctgcattgc tgacaaggag      1020
tataccaata tcttcctgaa atttgggtct ggatatgtgt ctgggtgggg aagggctctc      1080
cacaagggaa ggtctgctct ggtgctgcag tatctgaggg tgcccctggt ggacagagct      1140
acctgcctgc tgagcaccaa gttcaccatc tacaacaata tgttctgtgc tggatttcat      1200
gagggagggg gggactcctg tcagggagat tctggaggcc ctcatgtgac agaggtggaa      1260
ggcaccagct tcctgactgg catcatctct tggggggagg aatgtgctat gaaggggaaa      1320
tatggaatct ataccaagggt gtccagatat gtcaactgga tcaaggagaa aaccaagctg      1380
acctga

```

<210> 11
 <211> 4404
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> fVIII recombinante
 <400> 11

ES 2 763 551 T3

atgcagctgg aactgtctac ctgtgtgttt ctgtgtctgc tgcctctggg gttttctgct	60
atcaggagat actatctggg agctgtggag ctgtcctggg actacaggca gtctgagctg	120
ctgagagAAC tgcAtgtgga taccagattc ccagctacag ctccaggagc tctgcctctg	180
ggcccatctg tgctgtacaa gaaaacagtc tttgtggagt ttacagacca gctgttctct	240
gtggccaggc caagaccacc ttggatggga ctgctgggac caaccatcca ggctgaggtg	300
tatgatacag tggTggtgac cctgaaaaac atggcctccc atcctgtgag cctgcatgct	360
gtgggggtgt ctttctggaa gtcctctgag ggagctgagt atgaagacca tacctcccag	420
agggagaaag aagatgataa ggtgctgcct ggcaaaagcc agacctatgt ctggcaggtg	480
ctgaaggaga atggaccaac tgcttctgac ccaccatgcc tgacctactc ttatctgtcc	540

ES 2 763 551 T3

catgtggatc	tggtgaagga	cctgaattct	ggactgattg	gagctctgct	ggtgtgtaga	600
gaggggaagcc	tgaccagaga	aagaaccag	aacctgcatg	agtttgcct	gctgtttgct	660
gtgtttgatg	aagggaagag	ctggcactct	gccaggaatg	actcctggac	cagagctatg	720
gatccagctc	ctgctagagc	tcagcctgct	atgcacacag	tcaatggcta	tgtgaatagg	780
tctctgccag	gactgattgg	ctgccataag	aaatctgtct	attggcatgt	gattggaatg	840
ggcaccagcc	ctgaggtgca	ttctatcttc	ctggaaggcc	acacctttct	ggtcaggcac	900
catagacagg	cctctctgga	gatctcccct	ctgaccttcc	tgacagctca	gacctttctg	960
atggacctgg	ggcagttcct	gctgttttgc	catatctctt	cccaccatca	tggaggaatg	1020
gaggctcatg	tcaggttgga	atcctgtgct	gaggaaccac	agctgagaag	aaaggctgat	1080
gaggaagagg	actatgatga	taacctgtat	gactctgata	tggatgtggt	gaggctggat	1140
ggggatgatg	tcagcccttt	catccagatc	aggtctgtgg	ccaagaaaca	tcaaagacc	1200
tgggtccact	acattgctgc	tgaagaggaa	gattgggact	atgccccct	ggtgctggct	1260
cctgatgata	gacctaaca	aagccagtat	ctgaacaatg	ggccccagag	gattggaagg	1320
aagtacaaga	aagtgaggtt	catggcctat	acagatgaga	cctttaagac	cagagaggct	1380
atccagcatg	aatctgggat	cctgggacct	ctgctgtatg	gagaagtggg	ggataccctg	1440
ctgatcatct	tcaagaacca	ggcctccagg	ccatacaata	tctatcccca	tggcatcaca	1500
gatgtgagac	cactgtacag	caggagactg	ccaaggggg	tcaaacacct	gaaggatttc	1560
cccacctgct	ctggagagat	ctttaagtat	aaatggacag	tcacagtgga	agatgggcct	1620
accaagtctg	atccaaggtg	cctgaccaga	tactatagct	cttttgtgaa	catggagaga	1680
gacctggctt	ctggactgat	tggacctctg	ctgatctgtt	acaaagagtc	tgtggaccag	1740
aggggcaacc	agatcatgtc	tgataagaga	aatgtcatcc	tgttctctgt	gtttgatgag	1800
aacaggagct	ggtacctgac	agagaacatc	cagaggttcc	tgccaaatcc	agctggagtg	1860
cagctggagg	accagaatt	tcaggcttcc	aacatcatgc	atagcatcaa	tggctatgtg	1920
tttgatagcc	tgacagctgc	tgtctgcctg	catgaggtgg	cctactggta	tatcctgtcc	1980
attggagctc	agacagactt	cctgtctgtg	ttcttttagtg	ggtacacctt	taagcataaa	2040
atggtgtatg	aggataccct	gacctgttcc	cccttttctg	gggagacagt	gttcatgtcc	2100
atggaaaacc	ctggcctgtg	gatcctgggg	tgccacaact	ctgacttcag	gaatagagga	2160
atgacagccc	tgctgaaagt	gtccagctgt	gataagaata	caggggatta	ctatgaggac	2220
tcttatgaag	atatctctgc	ttatctgctg	agcaagaaca	atgccattga	gccaggtct	2280
tttgctcaga	actccagacc	tccatctgct	tctgctccta	agccacctgt	gctgagaaga	2340
catcagaggg	acatctccct	gcctaccttc	cagccagagg	aagataaaat	ggactatgat	2400

ES 2 763 551 T3

gatatcttca	gcacagagac	caagggggaa	gattttgaca	tctatggaga	ggatgaaaac	2460
caggatccaa	gaccttcca	gaagagaacc	agacactact	ttattgctgc	tgtggagcag	2520
ctgtgggact	atgggatgtc	tgaaagccca	agggccctga	ggaacagagc	tcagaatgga	2580
gaggtgcca	gattcaagaa	agtgggtgttc	agagagtttg	ctgatggcag	ctttaccag	2640
ccatcttaca	ggggggagct	gaacaagcat	ctggggctgc	tgggacccta	tatcagagct	2700
gaggtggaag	ataacatcat	ggtgaccttc	aagaatcagg	cttctaggcc	ctactccttt	2760
tattcttccc	tgatctccta	ccctgatgat	caggagcagg	gagctgaacc	taggcacaac	2820
tttgtgcagc	caaatgagac	cagaacctac	ttttggaagg	tgcagcatca	catggctccc	2880
acagaggatg	aatttgactg	caaagcttgg	gcctatTTTT	ctgatgtgga	cctggagaag	2940
gatgtgcatt	ctggcctgat	tgggcctctg	ctgatctgta	gggccaacac	cctgaatgct	3000
gctcatggaa	gacaggtcac	agtgcaggag	tttgctctgt	tctttaccat	ctttgatgaa	3060
accaagagct	ggtacttcac	agagaatgtg	gaaaggaatt	gcagagcccc	ctgtcatctg	3120
cagatggagg	accctaccct	gaaggaaaac	tacaggttcc	atgccatcaa	tggatatgtc	3180
atggataccc	tgcttggcct	ggtcatggct	cagaaccaga	ggatcagatg	gtacctgctg	3240
tctatgggat	ccaatgagaa	tatccatagc	atccacttct	ctggccatgt	cttttctgtg	3300
aggaagaaaag	aggaatacaa	aatggctgtg	tacaatctgt	atcctggggg	ctttgagaca	3360
gtggaaatgc	tgccaagcaa	agtgggaatc	tggagaattg	agtgcctgat	tggggaacac	3420
ctgcaggctg	ggatgagcac	caccttctct	gtgtactcta	agaaatgtca	gacccctctg	3480
gggatggcct	ctggacatat	cagggacttc	cagatcacag	cttctggaca	gtatggacag	3540
tgggctccaa	agctggctag	actgcactat	tctggctcca	tcaatgcctg	gtctaccaa	3600
gagccattct	cctggatcaa	ggtggacctg	ctggcccca	tgatcatcca	tggaatcaa	3660
accaggggag	ctaggcagaa	gttcagctct	ctgtacatct	cccagtttat	catcatgtat	3720
agcctggatg	ggaagaaatg	gcagacctac	agaggcaatt	ccactgggac	cctgatggtc	3780
ttctttggaa	atgtggattc	ctctggcatc	aagcacaaca	tcttcaatcc	acctatcatt	3840
gccaggtaca	tcaggctgca	tcctaccac	tatagcatca	ggtctaccct	gagaatggag	3900
ctgatgggat	gtgacctgaa	cagctgttct	atgccactgg	gcatggagtc	caaggctatc	3960
tctgatgccc	agatcacagc	ttcttcctac	ttcaccaata	tgtttgctac	ctggccccca	4020
agcaaggcta	gactgcacct	gcagggaaga	tccaatgctt	ggagaccca	ggtgaacaat	4080
cctaaggagt	ggctgcaggt	ggacttccag	aaaacctga	aggtcacagg	ggtgaccacc	4140
cagggagtga	aatctctgct	gacctccatg	tatgtcaagg	agttcctgat	cagctcttcc	4200
caggatggcc	accagtggac	cctgttcttt	cagaatggca	aggtcaaagt	gttcaggggg	4260
aatcaggact	cttttaccct	agtgggtgaac	tccctggatc	ctccactgct	gaccagggtac	4320

ES 2 763 551 T3

ctgagaatcc atcctcagag ctgggtgcac cagattgctc tgagaatgga ggtcctggga 4380
tgtgaagctc aggacctgta ttga 4404

<210> 12
<211> 4404
<212> ADN
<213> Secuencia artificial

<220>
<223> fVIII recombinante

<400> 12

atgcagctgg aactgtctac ctgtgtgttt ctgtgtctgc tgcctctggg gttttctgct 60
atccgccgct actatctggg agccgtggag ctgtcctggg actacaggca gagcgagctg 120
ctgagagaac tgcacgtgga taccagattc ccagctaccg ctccaggagc tctgcctctg 180
ggcccatccg tgctgtacaa gaaaaccgtc ttctgtggagt ttaccgacca gctgttcagc 240
gtggccaggc caagaccacc ttggatggga ctgctgggac caaccatcca ggctgaggtg 300
tacgataccg tggctctgac cctgaaaaac atggcctccc atcccgtgag cctgcacgct 360
gtcggggtgt ctttctggaa gtccagcgag ggagccgagt acgaagacca tacctccag 420
cgcgagaaag aagacgataa ggtgctgcct ggcaaaagcc agacctatgt ctggcaggtg 480
ctgaaggaga acggaccaac cgctagcgac ccaccatgcc tgacctactc ttatctgtcc 540
cacgtcgatc tgggtgaagga cctgaattcc ggactgatcg gagctctgct ggtgtgtaga 600
gaggaagcc tgaccagaga aagaaccag aacctgcatg agttcgtcct gctgttcgcc 660
gtgtttgacg aaggaagag ctggcactct gcccgcaatg actcctggac cagagctatg 720
gatccagctc ctgctagagc tcagcctgct atgcacaccg tcaacggcta cgtgaatcgg 780
tctctgccag gactgatcgg ctgccataag aaaagcgtct attggcacgt gatcggaatg 840
ggcaccagcc ccgaggtgca ttctatcttc ctggaaggcc acacctttct ggtcaggcac 900
catagacagg cctctctgga gatctcccct ctgaccttcc tgaccgctca gacctttctg 960
atggacctgg ggcagttcct gctgttttgc catatctctt cccaccatca cggaggaatg 1020
gaggctcagc tcagggtgga atcctgtgct gaggaaccac agctgagaag aaaggctgat 1080
gaggaagagg actacgacga taacctgtat gacagcgata tggacgtcgt gcgcctggac 1140
ggcgacgatg tcagcccttt catccagatc cggctctgtg ocaagaaaca tccaaagacc 1200
tggtgccact acatcgccgc tgaagaggaa gattgggact atgccccct ggtgctggct 1260
cctgacgata gatcctacaa aagccagtat ctgaacaatg ggcccagcg catcgacgg 1320
aagtacaaga aagtgaggtt catggcctat accgacgaga cctttaagac cagagaggct 1380
atccagcacg aatccgggat cctgggacct ctgctgtacg gcgaagtggg ggataccctg 1440

ES 2 763 551 T3

ctgatcatct tcaagaacca ggctccagg ccatacaata tctatcccca tggcatcacc	1500
gacgtgagac cactgtacag caggagactg cccaaggggg tcaaacacct gaaggatttc	1560
cccatcctgc ctggagagat ctttaagtat aaatggaccg tcaccgtgga agacgggcct	1620
accaagtccg atccacgctg cctgaccggg tactatagct ctttcgtgaa catggagaga	1680
gacctggcta gcgactgat cggaccctg ctgatctgtt acaaagagag cgtggaccag	1740
aggggcaacc agatcatgtc tgataagaga aatgtcatcc tgttctccgt gtttgacgag	1800
aaccgcagct ggtacctgac cgagaacatc cagcggttcc tgccaaatcc agctggagtg	1860
cagctggagg acccagaatt tcaggcttcc aacatcatgc atagcatcaa tggctacgtg	1920
ttcgatagcc tgcagctgtc tgtctgcctg cacgaggtgg cctactggta tatcctgtcc	1980
atcggcgctc agaccgactt cctgtccgtg ttctttagcg ggtacacctt taagcataaa	2040
atggtgtatg aggataccct gaccctgttc cccttttctg gcgagaccgt gttcatgtcc	2100
atggaaaacc ctggcctgtg gatcctgggg tgccacaaca gcgacttcag gaatagagga	2160
atgaccgccc tgctgaaagt gtccagctgt gataagaata ccggcgatta ctatgaggac	2220
tcttacgaag atatctccgc ttatctgctg agcaagaaca atgccatcga gcccaggtct	2280
ttcgctcaga actccagacc tccaagcgt tctgctccta agccacctgt gctgagaaga	2340
catcagaggg acatctccct gcctaccttc cagccagagg aagataaaat ggactacgac	2400
gatatcttca gcaccgagac caagggggaa gattttgaca tctatggaga ggacgaaaac	2460
caggatccaa gatccttcca gaagagaacc agacactact ttatcgccgc tgtggagcag	2520
ctgtgggact atgggatgtc cgaaagccca cgggccctga ggaacagagc tcagaatgga	2580
gaggtgcccc gcttcaagaa agtcgtgttc cgggagtttg ccgacggcag ctttaccag	2640
ccatcttaca ggggggagct gaacaagcat ctggggctgc tgggacccta tatcagacc	2700
gaggtcgaag ataacatcat ggtgaccttc aagaatcagg cttctcgccc ctactccttt	2760
tattcttccc tgatctccta ccctgacgat caggagcagg gcgccgaacc taggcacaac	2820
ttcgtgcagc caaatgagac cagaacctac ttttggaagg tgcagcatca catggctccc	2880
accgaggatg aattcgactg caaagcttgg gcctattttt ccgatgtcga cctggagaag	2940
gacgtgcata ggggcctgat cgggcctctg ctgatctgtc gcgccaaacac cctgaatgct	3000
gctcacggaa gacaggtcac cgtgcaggag ttcgctctgt tctttaccat ctttgacgaa	3060
accaagagct ggtacttcac cgagaacgtg gaaaggaatt gcagagcccc ctgtcatctg	3120
cagatggagg accctaccct gaaggaaaac tacaggttcc acgccatcaa tggatatgtc	3180
atggataccc tgcccggcct ggtcatggct cagaaccagc gcatccgggt gtacctgtg	3240
tctatgggat ccaacgagaa tatccatagc atccacttct ctggccatgt cttttccgtg	3300
aggaagaaaag aggaatacaa aatggccgtg tacaatctgt atcctgggggt cttcgagacc	3360

ES 2 763 551 T3

```

gtggaaatgc tgccaagcaa agtgggaatc tggagaatcg agtgcctgat cggcgaacac      3420
ctgcaggccg ggatgagcac caccttcctg gtgtactcta agaaatgtca gacccccactg      3480
gggatggcct ccggacatat ccgcgacttc cagatcaccg ctagcggaca gtacggacag      3540
tgggctccaa agctggctag actgcactat tctggctcca tcaacgcctg gtctacaaaa      3600
gagccattct cctggatcaa ggtggacctg ctggccccc tcatcatcca cggaatcaaa      3660
accagggcg ctaggcagaa gttcagctct ctgtacatct cccagtttat catcatgtat      3720
agcctggacg ggaagaaatg gcagacctac agaggcaatt ccaccgggac cctgatggtc      3780
ttctttggaa acgtggattc cagcggcatc aagcacaaca tcttcaatcc acccatcatc      3840
gcccgtaca tccggctgca tcctaccac tatagcatca ggtctaccct gagaatggag      3900
ctgatgggat gcgacctgaa cagctgttct atgccactgg gcatggagtc caaggctatc      3960
agcgatgcc agatcaccgc ttcttcctac ttcaccaata tgtttgctac ctggtcccca      4020
agcaaggcta gactgcacct gcagggaaga tccaacgctt ggagacccca ggtgaacaat      4080
cctaaggagt ggctgcaggt cgacttcag aaaaccatga aggtcaccgg ggtgaccacc      4140
caggagagtga aatctctgct gacctccatg tacgtcaagg agttcctgat cagctcttcc      4200
caggacggcc accagtggac cctgttcttt cagaacggca aggtcaaagt gttccagggg      4260
aatcaggact cttttacccc cgtcgtgaac tccctggatc ctccactgct gaccaggtac      4320
ctgagaatcc atcctcagag ctgggtgcac cagatcgctc tgagaatgga ggtcctggga      4380
tgccaagctc aggacctgta ttga      4404

```

<210> 13

<211> 391

<212> PRT

<213> Secuencia artificial

<220>

<223> secuencia promotora recombinante

<400> 13

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1               5               10               15

```

```

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
                20               25               30

```

```

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
                35               40               45

```

```

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50               55               60

```

ES 2 763 551 T3

Tyr	Lys	Lys	Thr	Leu	Phe	Val	Glu	Phe	Thr	Val	His	Leu	Phe	Asn	Ile	65	70	75	80
Ala	Lys	Pro	Arg	Pro	Pro	Trp	Met	Gly	Leu	Leu	Gly	Pro	Thr	Ile	Gln	85	90	95	
Ala	Glu	Val	Tyr	Asp	Thr	Val	Val	Ile	Thr	Leu	Lys	Asn	Met	Ala	Ser	100	105	110	
His	Pro	Val	Ser	Leu	His	Ala	Val	Gly	Val	Ser	Tyr	Trp	Lys	Ala	Ser	115	120	125	
Glu	Gly	Ala	Glu	Tyr	Asp	Asp	Gln	Thr	Ser	Gln	Arg	Glu	Lys	Glu	Asp	130	135	140	
Asp	Lys	Val	Phe	Pro	Gly	Gly	Ser	His	Thr	Tyr	Val	Trp	Gln	Val	Leu	145	150	155	160
Lys	Glu	Asn	Gly	Pro	Met	Ala	Ser	Asp	Pro	Leu	Cys	Leu	Thr	Tyr	Ser	165	170	175	
Tyr	Leu	Ser	His	Val	Asp	Leu	Val	Lys	Asp	Leu	Asn	Ser	Gly	Leu	Ile	180	185	190	
Gly	Ala	Leu	Leu	Val	Cys	Arg	Glu	Gly	Ser	Leu	Ala	Lys	Glu	Lys	Thr	195	200	205	
Gln	Thr	Leu	His	Lys	Phe	Ile	Leu	Leu	Phe	Ala	Val	Phe	Asp	Glu	Gly	210	215	220	
Lys	Ser	Trp	His	Ser	Glu	Thr	Lys	Asn	Ser	Leu	Met	Gln	Asp	Arg	Asp	225	230	235	240
Ala	Ala	Ser	Ala	Arg	Ala	Trp	Pro	Lys	Met	His	Thr	Val	Asn	Gly	Tyr	245	250	255	
Val	Asn	Arg	Ser	Leu	Pro	Gly	Leu	Ile	Gly	Cys	His	Arg	Lys	Ser	Val	260	265	270	
Tyr	Trp	His	Val	Ile	Gly	Met	Gly	Thr	Thr	Pro	Glu	Val	His	Ser	Ile	275	280	285	
Phe	Leu	Glu	Gly	His	Thr	Phe	Leu	Val	Arg	Asn	His	Arg	Gln	Ala	Ser	290	295	300	
Leu	Glu	Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	Met	305	310	315	320

ES 2 763 551 T3

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
325 330 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
370 375 380

Pro Ser Phe Ile Gln Ile Arg
385 390

<210> 14

<211> 391

<212> PRT

<213> Secuencia artificial

<220>

<223> secuencia promotora recombinante

<400> 14

Met Gln Leu Glu Leu Ser Thr Cys Val Phe Leu Cys Leu Leu Pro Leu
1 5 10 15

Gly Phe Ser Ala Ile Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Arg Gln Ser Glu Leu Leu Arg Glu Leu His Val Asp Thr
35 40 45

Arg Phe Pro Ala Thr Ala Pro Gly Ala Leu Pro Leu Gly Pro Ser Val
50 55 60

Leu Tyr Lys Lys Thr Val Phe Val Glu Phe Thr Asp Gln Leu Phe Ser
65 70 75 80

Val Ala Arg Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile
85 90 95

Gln Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala
100 105 110

Ser His Pro Val Ser Leu His Ala Val Gly Val Ser Phe Trp Lys Ser
115 120 125

ES 2 763 551 T3

Ser Glu Gly Ala Glu Tyr Glu Asp His Thr Ser Gln Arg Glu Lys Glu
 130 135 140
 Asp Asp Lys Val Leu Pro Gly Lys Ser Gln Thr Tyr Val Trp Gln Val
 145 150 155 160
 Leu Lys Glu Asn Gly Pro Thr Ala Ser Asp Pro Pro Cys Leu Thr Tyr
 165 170 175
 Ser Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu
 180 185 190
 Ile Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Thr Arg Glu Arg
 195 200 205
 Thr Gln Asn Leu His Glu Phe Val Leu Leu Phe Ala Val Phe Asp Glu
 210 215 220
 Gly Lys Ser Trp His Ser Ala Arg Asn Asp Ser Trp Thr Arg Ala Met
 225 230 235 240
 Asp Pro Ala Pro Ala Arg Ala Gln Pro Ala Met His Thr Val Asn Gly
 245 250 255
 Tyr Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Lys Lys Ser
 260 265 270
 Val Tyr Trp His Val Ile Gly Met Gly Thr Ser Pro Glu Val His Ser
 275 280 285
 Ile Phe Leu Glu Gly His Thr Phe Leu Val Arg His His Arg Gln Ala
 290 295 300
 Ser Leu Glu Ile Ser Pro Leu Thr Phe Leu Thr Ala Gln Thr Phe Leu
 305 310 315 320
 Met Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His His
 325 330 335
 His Gly Gly Met Glu Ala His Val Arg Val Glu Ser Cys Ala Glu Glu
 340 345 350
 Pro Gln Leu Arg Arg Lys Ala Asp Glu Glu Glu Asp Tyr Asp Asp Asn
 355 360 365
 Leu Tyr Asp Ser Asp Met Asp Val Val Arg Leu Asp Gly Asp Asp Val
 370 375 380

ES 2 763 551 T3

Ser Pro Phe Ile Gln Ile Arg
385 390

<210> 15
<211> 371
<212> PRT
<213> Secuencia artificial

<220>
<223> secuencia promotora recombinante

<400> 15

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

ES 2 763 551 T3

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

Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg Gln Val Thr
210 215 220

Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser
225 230 235 240

Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn
245 250 255

Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His Ala
260 265 270

Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu Val Met Ala Gln
275 280 285

Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn
290 295 300

Ile His Ser Ile His Phe Ser Gly His Val Phe Thr Val Arg Lys Lys
305 310 315 320

Glu Glu Tyr Lys Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu
325 330 335

Thr Val Glu Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys
340 345 350

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

Tyr Ser Asn
370

<210> 16
<211> 371
<212> PRT
<213> Secuencia artificial

<220>
<223> secuencia promotora recombinante

<400> 16

Asp Ile Ser Leu Pro Thr Phe Gln Pro Glu Glu Asp Lys Met Asp Tyr
1 5 10 15

ES 2 763 551 T3

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

ES 2 763 551 T3

Ile Asn Gly Tyr Val Met Asp Thr Leu Pro Gly Leu Val Met Ala Gln
275 280 285

Asn Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn
290 295 300

Ile His Ser Ile His Phe Ser Gly His Val Phe Ser Val Arg Lys Lys
305 310 315 320

Glu Glu Tyr Lys Met Ala Val Tyr Asn Leu Tyr Pro Gly Val Phe Glu
325 330 335

Thr Val Glu Met Leu Pro Ser Lys Val Gly Ile Trp Arg Ile Glu Cys
340 345 350

Leu Ile Gly Glu His Leu Gln Ala Gly Met Ser Thr Thr Phe Leu Val
355 360 365

Tyr Ser Lys
370

<210> 17
<211> 461
<212> PRT
<213> Secuencia artificial

<220>
<223> flX recombinante

<400> 17

Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
35 40 45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
50 55 60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65 70 75 80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
85 90 95

ES 2 763 551 T3

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
 100 105 110
 Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
 115 120 125
 Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
 130 135 140
 Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
 145 150 155 160
 Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
 165 170 175
 Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
 180 185 190
 Glu Ala Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
 195 200 205
 Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
 210 215 220
 Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
 225 230 235 240
 Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
 245 250 255
 Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
 260 265 270
 Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
 275 280 285
 His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
 290 295 300
 Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
 305 310 315 320
 Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
 325 330 335
 Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
 340 345 350

ES 2 763 551 T3

Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
355 360 365

Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Arg
370 375 380

Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
385 390 395 400

Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
405 410 415

Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
420 425 430

Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser
435 440 445

Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
450 455 460

<210> 18

<211> 461

<212> PRT

<213> Secuencia artificial

<220>

<223> flX recombinante

<400> 18

Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
35 40 45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
50 55 60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65 70 75 80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
85 90 95

ES 2 763 551 T3

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
 100 105 110

Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
 115 120 125

Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
 130 135 140

Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
 145 150 155 160

Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
 165 170 175

Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
 180 185 190

Glu Ala Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
 195 200 205

Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
 210 215 220

Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
 225 230 235 240

Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
 245 250 255

Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
 260 265 270

Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
 275 280 285

His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
 290 295 300

Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
 305 310 315 320

Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
 325 330 335

Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
 340 345 350

ES 2 763 551 T3

Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
355 360 365

Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Leu
370 375 380

Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
385 390 395 400

Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
405 410 415

Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
420 425 430

Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser
435 440 445

Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
450 455 460

<210> 19

<211> 461

<212> PRT

<213> Secuencia artificial

<220>

<223> flX recombinante

<400> 19

Met Gln Arg Val Asn Met Ile Met Ala Glu Ser Pro Gly Leu Ile Thr
1 5 10 15

Ile Cys Leu Leu Gly Tyr Leu Leu Ser Ala Glu Cys Thr Val Phe Leu
20 25 30

Asp His Glu Asn Ala Asn Lys Ile Leu Asn Arg Pro Lys Arg Tyr Asn
35 40 45

Ser Gly Lys Leu Glu Glu Phe Val Gln Gly Asn Leu Glu Arg Glu Cys
50 55 60

Met Glu Glu Lys Cys Ser Phe Glu Glu Ala Arg Glu Val Phe Glu Asn
65 70 75 80

Thr Glu Arg Thr Thr Glu Phe Trp Lys Gln Tyr Val Asp Gly Asp Gln
85 90 95

ES 2 763 551 T3

Cys Glu Ser Asn Pro Cys Leu Asn Gly Gly Ser Cys Lys Asp Asp Ile
 100 105 110
 Asn Ser Tyr Glu Cys Trp Cys Pro Phe Gly Phe Glu Gly Lys Asn Cys
 115 120 125
 Glu Leu Asp Val Thr Cys Asn Ile Lys Asn Gly Arg Cys Glu Gln Phe
 130 135 140
 Cys Lys Asn Ser Ala Asp Asn Lys Val Val Cys Ser Cys Thr Glu Gly
 145 150 155 160
 Tyr Arg Leu Ala Glu Asn Gln Lys Ser Cys Glu Pro Ala Val Pro Phe
 165 170 175
 Pro Cys Gly Arg Val Ser Val Ser Gln Thr Ser Lys Leu Thr Arg Ala
 180 185 190
 Glu Thr Val Phe Pro Asp Val Asp Tyr Val Asn Ser Thr Glu Ala Glu
 195 200 205
 Thr Ile Leu Asp Asn Ile Thr Gln Ser Thr Gln Ser Phe Asn Asp Phe
 210 215 220
 Thr Arg Val Val Gly Gly Glu Asp Ala Lys Pro Gly Gln Phe Pro Trp
 225 230 235 240
 Gln Val Val Leu Asn Gly Lys Val Asp Ala Phe Cys Gly Gly Ser Ile
 245 250 255
 Val Asn Glu Lys Trp Ile Val Thr Ala Ala His Cys Val Glu Thr Gly
 260 265 270
 Val Lys Ile Thr Val Val Ala Gly Glu His Asn Ile Glu Glu Thr Glu
 275 280 285
 His Thr Glu Gln Lys Arg Asn Val Ile Arg Ile Ile Pro His His Asn
 290 295 300
 Tyr Asn Ala Ala Ile Asn Lys Tyr Asn His Asp Ile Ala Leu Leu Glu
 305 310 315 320
 Leu Asp Glu Pro Leu Val Leu Asn Ser Tyr Val Thr Pro Ile Cys Ile
 325 330 335
 Ala Asp Lys Glu Tyr Thr Asn Ile Phe Leu Lys Phe Gly Ser Gly Tyr
 340 345 350

ES 2 763 551 T3

Val Ser Gly Trp Gly Arg Val Phe His Lys Gly Arg Ser Ala Leu Val
355 360 365

Leu Gln Tyr Leu Arg Val Pro Leu Val Asp Arg Ala Thr Cys Leu Leu
370 375 380

Ser Thr Lys Phe Thr Ile Tyr Asn Asn Met Phe Cys Ala Gly Phe His
385 390 395 400

Glu Gly Gly Arg Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro His Val
405 410 415

Thr Glu Val Glu Gly Thr Ser Phe Leu Thr Gly Ile Ile Ser Trp Gly
420 425 430

Glu Glu Cys Ala Met Lys Gly Lys Tyr Gly Ile Tyr Thr Lys Val Ser
435 440 445

Arg Tyr Val Asn Trp Ile Lys Glu Lys Thr Lys Leu Thr
450 455 460

<210> 20

<211> 39

5 <212> ADN

<213> Secuencia artificial

<220>

10 <223> secuencia promotora recombinante

<400> 20

tataaaaggc cagcagcagc ctgaccacat ctcacctc 39

<210> 21

15 <211> 38

<212> ADN

<213> Secuencia artificial

<220>

20 <223> secuencia promotora recombinante

<400> 21

gtaatttt gtggcccttg cgaatttgc tctggtta 38

<210> 22

25 <211> 14

<212> PRT

<213> Secuencia artificial

<220>

30 <223> enlazador peptídico recombinante

<400> 22

Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His Gln Arg
1 5 10

35

5 <210> 23
 <211> 24
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> enlazador peptídico recombinante
 <400> 23
 Ser Phe Ala Gln Asn Ser Arg Pro Pro Ser Ala Ser Ala Pro Lys Pro
 1 5 10 15
 Pro Val Leu Arg Arg His Gln Arg
 20
 15 <210> 24
 <211> 4
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> secuencia promotora recombinante
 20 <400> 24
 Arg His Gln Arg
 1
 25 <210> 25
 <211> 6
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> secuencia promotora recombinante
 <400> 25
 gttaat 6
 35 <210> 26
 <211> 5
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> secuencia promotora recombinante
 <400> 26
 tataa 5
 45 <210> 27
 <211> 5
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> secuencia promotora recombinante
 55 <400> 27
 gttaa 5

5	<210> 28 <211> 4 <212> ADN <213> Secuencia artificial	
	<220> <223> secuencia promotora recombinante	
10	<400> 28 gtta	4
15	<210> 29 <211> 3 <212> ADN <213> Secuencia artificial	
	<220> <223> secuencia promotora recombinante	
20	<400> 29 gtt	3
25	<210> 30 <211> 5 <212> ADN <213> Secuencia artificial	
30	<220> <223> secuencia promotora recombinante	
	<400> 30 ttaat	5
35	<210> 31 <211> 4 <212> ADN <213> Secuencia artificial	
40	<220> <223> secuencia promotora recombinante	
	<400> 31 ttaa	4
45	<210> 32 <211> 3 <212> ADN <213> Secuencia artificial	
50	<220> <223> secuencia promotora recombinante	
55	<400> 32 tta	3
60	<210> 33 <211> 3 <212> ADN <213> Secuencia artificial	
	<220> <223> secuencia promotora recombinante	
65	<400> 33 aat	3

5	<210> 34	
	<211> 56	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
10	<223> secuencia promotora recombinante	
	<400> 34	
	gttaattttt gtggcccttg cgatgtttgc tctggttaat aatctcagga caaaca	56
	<210> 35	
	<211> 55	
15	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
	<400> 35	
20	ttaatttttg tgcccttgc gatgtttgct ctggttaata atctcaggac aaaca	55
	<210> 36	
	<211> 54	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 36	
	taatttttgc ggcccttgcg atgtttgctc tggtaataa tctcaggaca aaca	54
	<210> 37	
30	<211> 53	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
35	<400> 37	
	aatttttgcg gcccttgcat tgttgctct ggtaataat ctgaggacaa aca	53
	<210> 38	
	<211> 52	
	<212> ADN	
40	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
	<400> 38	
	atttttggg cccttgcat gttgctctg gtaataatc tcaggacaaa ca	52
45	<210> 39	
	<211> 51	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
50	<223> secuencia promotora recombinante	
	<400> 39	
	ttttgtggc ccttgcatg ttgtctctg ttaataatc caggacaaac a	51
	<210> 39	
	<211> 51	
	<212> ADN	
55	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
	<400> 39	
	ttttgtggc ccttgcatg ttgtctctg ttaataatc caggacaaac a	51

	<210> 40	
	<211> 50	
	<212> ADN	
	<213> Secuencia artificial	
5	<220>	
	<223> secuencia promotora recombinante	
	<400> 40	
10	tttgtggcc ctgcgatgt ttgctctggt taataatctc aggacaaaca	50
	<210> 41	
	<211> 49	
	<212> ADN	
15	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
	<400> 41	
20	tttgtggccc ttgcgatgtt tgctctggtt aataatctca ggacaaaca	49
	<210> 42	
	<211> 48	
25	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
30	<400> 42	
	ttgtggccct tgcatgttt gctctggtta ataatctcag gacaaaca 48 <210> 43	
	<211> 18	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
	<400> 43	
40	ataatctcag gacaaaca	18
	<210> 44	
	<211> 47	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 44	
	tgtggccctt gcgatgttg ctctggtaa taatctcagg acaaaca	47
	<210> 45	
55	<211> 46	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 45	
	gtggcccttg cgatgttgc tctggtaaat aatctcagga caaaca	46
65	<210> 46	
	<211> 45	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 46	
	tgcccttgc gatgttgc ctggttaata atctcaggac aaaca	45
10	<210> 47	
	<211> 43	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 47	
20	gcccttgcga tggttgcct ggtaataat ctcaggacaa aca	43
	<210> 48	
	<211> 42	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 48	
30	cccttgcgat gtttgcctg gtaataatc tcaggacaaa ca	42
	<210> 49	
	<211> 41	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 49	
	ccttgcgatg ttgtctctgg ttaataatc caggacaaac a	41
	<210> 50	
	<211> 40	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 50	
	cttgcgatg ttgtctctgg taataatctc aggacaaaca	40
	<210> 51	
55	<211> 55	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 51	
	gttaatttt gtgcccttg cgatgttgc tctggttaat aatctcagga caaac	55
65	<210> 52	
	<211> 54	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 52	
	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga caaa	54
10	<210> 53	
	<211> 53	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 53	
20	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga caa	53
	<210> 54	
	<211> 53	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 54	
30	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga caa	53
	<210> 55	
	<211> 52	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 55	
	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga ca	52
	<210> 56	
	<211> 51	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 56	
	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga c	51
	<210> 57	
55	<211> 50	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 57	
	gttaatttt gtggcccttg cgatgttgc tctggtaat aatctcagga	50
65	<210> 58	
	<211> 49	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 58	
	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatctcagg	49
10	<210> 59	
	<211> 48	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 59	
20	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatctcag	48
	<210> 60	
	<211> 47	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 60	
30	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatctca	47
	<210> 61	
	<211> 46	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 61	
	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatctc	46
	<210> 62	
	<211> 45	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 62	
	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatct	45
	<210> 63	
55	<211> 44	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 63	
	gttaatTTTT gtggcccttg cgatgtttgc tctgggtaaat aatc	44
65	<210> 64	
	<211> 43	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 64	
	gttaat tttt gtggcccttg cgatgtttgc tctggtaat aat	43
10	<210> 65	
	<211> 42	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 65	
20	gttaat tttt gtggcccttg cgatgtttgc tctggtaat aa	42
	<210> 66	
	<211> 55	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 66	
30	gttaat tttt gtggcccttg cgatgtttgc tctggtaat aatctcagga caaac	55
	<210> 67	
	<211> 53	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 67	
	ttaat ttttg tggcccttgc gatgtttgct ctggttaata atctcaggac aaa	53
	<210> 68	
	<211> 51	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 68	
	taat ttttgt ggcccttgcg atgtttgctc tggtaataa tctcaggaca a	51
	<210> 69	
55	<211> 49	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 69	
	aat ttttgtg gcccttgcca tgtttgctct ggttaataat ctccaggaca	49
65	<210> 70	
	<211> 47	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 70	
	attttgtgg cccttgcatgt gttgctctg gtaataatc tcaggac	47
10	<210> 71	
	<211> 45	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 71	
20	ttttgtggc ccttgcatgt ttgctctgg ttaataatc cagga	45
	<210> 72	
	<211> 43	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 72	
30	ttttgtggcc ctgcatgt ttgctctgt taataatctc agg	43
	<210> 73	
	<211> 41	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 73	
	tttgtggccc ttgcatgtt tgctctggt aataatctca g	41
	<210> 74	
	<211> 39	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 74	
	ttgtggccct tgcatgtt gctctggtta ataactca	39
	<210> 75	
55	<211> 37	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
60	<223> secuencia promotora recombinante	
	<400> 75	
	tgtggccctt gcatgttg ctctggtta taactc	37
65	<210> 76	
	<211> 35	

	<212> ADN	
	<213> Secuencia artificial	
	<220>	
5	<223> secuencia promotora recombinante	
	<400> 76	
	gtggcccttg cgatgttgc tctggtaaat aatct	35
10	<210> 77	
	<211> 33	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> secuencia promotora recombinante	
	<400> 77	
20	tggcccttgc gatgttgc tctggtaata atc	33
	<210> 78	
	<211> 30	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<223> secuencia promotora recombinante	
	<400> 78	
30	gcccttgcca tgttgctct ggtaataat	30
	<210> 79	
	<211> 28	
	<212> ADN	
35	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
40	<400> 79	
	cccttgcat gttgctctg gtaataa 28 <210> 80	
	<211> 26	
	<212> ADN	
	<213> Secuencia artificial	
45	<220>	
	<223> secuencia promotora recombinante	
	<400> 80	
50	ccttgcatg ttgctctgg ttaata	26
	<210> 81	
	<211> 24	
	<212> ADN	
55	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
60	<400> 81	
	cttgcatgt ttgctctgg taat	24
	<210> 82	
	<211> 54	
65	<212> ADN	
	<213> Secuencia artificial	

	<220>	
	<223> secuencia promotora recombinante	
5	<400> 82 ttaatatttg tggcccttgc gatgtttgct ctggttaata atctcaggac aaac	54
	<210> 83	
	<211> 52	
10	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
15	<400> 83 taatttttg tggcccttgcg atgtttgctc tggtaataa tctcaggaca aa	52
	<210> 84	
	<211> 50	
20	<212> ADN	
	<213> Secuencia artificial	
	<220>	
25	<223> secuencia promotora recombinante	
	<400> 84 aatttttg tggcccttgcga tgtttgctct ggttaataat ctcaggacaa	50
30	<210> 85	
	<211> 49	
	<212> ADN	
	<213> Secuencia artificial	
35	<220>	
	<223> secuencia promotora recombinante	
	<400> 85 atttttg tgg cccttgcat gtttgcctg gtaataatc tcaggacaa	49
40	<210> 86	
	<211> 47	
	<212> ADN	
	<213> Secuencia artificial	
45	<220>	
	<223> secuencia promotora recombinante	
	<400> 86 ttttgtggc ccttgcatg ttgtctctg ttaataatc caggaca	47
50	<210> 87	
	<211> 45	
	<212> ADN	
55	<213> Secuencia artificial	
	<220>	
	<223> secuencia promotora recombinante	
60	<400> 87 ttttgtggc cttgcgatg ttgtctctg taataatctc aggac	45
	<210> 88	
	<211> 43	
65	<212> ADN	
	<213> Secuencia artificial	

	<220>	
	<223> secuencia promotora recombinante	
5	<400> 88 ttgtggccc tgcgatgtt tgctctggtt aataatctca gga	43
	<210> 89 <211> 41 <212> ADN <213> Secuencia artificial	
10		
	<220>	
	<223> secuencia promotora recombinante	
15	<400> 89 ttgtggccct tgcgatgtt gctctggtta ataatctcag g	41
	<210> 90 <211> 39 <212> ADN <213> Secuencia artificial	
20		
	<220>	
	<223> secuencia promotora recombinante	
25	<400> 90 tgtggccctt gcgatgttg ctctggttaa taatctcag	39
	<210> 91 <211> 36 <212> ADN <213> Secuencia artificial	
30		
	<220>	
	<223> secuencia promotora recombinante	
35	<400> 91 tggcccttgc gatgttgct ctggttaata atctca	36
40	<210> 92 <211> 34 <212> ADN <213> Secuencia artificial	
45		
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 92 ggcccttgcg atgttgctc tggtaataa tctc	34
	<210> 93 <211> 32 <212> ADN <213> Secuencia artificial	
55		
	<220>	
	<223> secuencia promotora recombinante	
60	<400> 93 gcccttgcga tggttgctct ggttaataat ct	32
	<210> 94 <211> 30 <212> ADN <213> Secuencia artificial	
65		

	<220>	
	<223> secuencia promotora recombinante	
5	<400> 94 cccttgcatg gttgctctg gtaataatc	30
	<210> 95 <211> 28 <212> ADN <213> Secuencia artificial	
10		
	<220>	
	<223> secuencia promotora recombinante	
15	<400> 95 ccttgcatg ttgctctgg ttaataat	28
	<210> 96 <211> 26 <212> ADN <213> Secuencia artificial	
20		
	<220>	
	<223> secuencia promotora recombinante	
25	<400> 96 cttgcatgt ttgctctgg taataa	26
	<210> 97 <211> 17 <212> ADN <213> Secuencia artificial	
30		
	<220>	
	<223> secuencia promotora recombinante	
35	<400> 97 cggaggagca aacaggg	17
40	<210> 98 <211> 28 <212> ADN <213> Secuencia artificial	
45		
	<220>	
	<223> secuencia promotora recombinante	
50	<400> 98 cggaggagca aacaggggct aagtcac	28
	<210> 99 <211> 27 <212> ADN <213> Secuencia artificial	
55		
	<220>	
	<223> secuencia promotora recombinante	
60	<400> 99 ggctgctggg gaatattaac caaggtc	27
	<210> 100 <211> 71 <212> ADN <213> Secuencia artificial	
65		

ES 2 763 551 T3

	<220>		
	<223> secuencia promotora recombinante		
5	<400> 100		
	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg	60	
	gctaagtcca c	71	
10	<210> 101 <211> 71 <212> ADN <213> Secuencia artificial		
15	<220> <223> secuencia promotora recombinante		
	<400> 101		
	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg	60	
	actaagtcca c	71	
20	<210> 102 <211> 220 <212> ADN <213> Secuencia artificial		
25	<220> <223> secuencia promotora recombinante		
	<400> 102		
30	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg	60	
	gctaagtcca ctaggttaat cattaagtcg ttaatttttg tggcccttgc gatgtttgct	120	
	ctggttaata atctcaggac aaacagaggt taataatttt ccagatctct ctgagcaata	180	
	gtataaaaagg ccagcagcag cctgaccaca tctcatcctc	220	
35	<210> 103 <211> 171 <212> ADN <213> Secuencia artificial		
40	<220> <223> secuencia promotora recombinante		
	<400> 103		
	cggaggagca aacaggggct aagtcgttaa tcattaagtc gttaattttt gtggcccttg	60	
	cgatgtttgc tctgggttaat aatctcagga caaacagagg ttaataattt tccagatctc	120	
	tctgagcaat agtataaaag gccagcagca gcctgaccac atctcatcct c	171	
45	<210> 104 <211> 173 <212> ADN <213> Secuencia artificial		

ES 2 763 551 T3

	<220>		
	<223> secuencia promotora recombinante		
5	<400> 104		
	ggctgctggt gaatattaac caaggtcgtt aatcattaag tcgttaattt ttgtggccct	60	
	tgcgatgttt gctctgggta ataatctcag gacaaacaga ggtaataat tttccagatc	120	
	tctctgagca atagtataaa aggccagcag cagcctgacc acatctcatc etc	173	
10	<210> 105 <211> 146 <212> ADN <213> Secuencia artificial		
15	<220> <223> secuencia promotora recombinante <400> 105		
	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg	60	
	gctaagtcca cgagggttaat aattttccag atctctctga gcaatagtat aaaaggccag	120	
	cagcagcctg accacatctc atcctc	146	
20	<210> 106 <211> 129 <212> ADN <213> Secuencia artificial		
25	<220> <223> secuencia promotora recombinante <400> 106		
30			
	ggctgctggt gaatattaac caaggtcatc ggaggagcaa acagggacta agtcgaggtt	60	
	aataattttc cagatctctc tgagcaatag tataaaaggc cagcagcagc ctgaccacat	120	
	ctcatcctc	129	
35	<210> 107 <211> 142 <212> ADN <213> Secuencia artificial		
40	<220> <223> secuencia promotora recombinante <400> 107		
	ggctgctggt gaatattaac caaggtcccc ttgctgggta ataatctcag ttaatttggt	60	
	tgcacaaaca cggaggagca aacaggggag gtaataaatt ttctataaaa ggccagcagc	120	
	agcctgacca catctcatcc tc	142	

ES 2 763 551 T3

5	<210> 108 <211> 146 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante		
10	<400> 108		
	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg		60
	actaagtcca cgagggttaat aattttccag atctctctga gcaatagtat aaaaggccag		120
	cagcagcctg accacatctc atcctc		146
15	<210> 109 <211> 8 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante		
20	<400> 109 tcctctc	8	
25	<210> 110 <211> 24 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante		
30	<400> 110 ggccagcagc agcctgacca catc	24	
35	<210> 111 <211> 71 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante		
40	<400> 111		
	ggggaggctg ctggtgaata ttaaccaagg tcaccccagt tatcggagga gcaaacaggg		60
45	gctaagtcca c		71
50	<210> 112 <211> 255 <212> ADN <213> Secuencia artificial <220> <223> secuencia promotora recombinante		
55	<400> 112		

ES 2 763 551 T3

	gttaatcatt aagtcgttaa tcattaactt aggtcaaagg tcagacaatg ttgactctcg	60
	ttaatgatta accggaattg ttgacttgag gttaataatt ttccagatct ctctgagcaa	120
	tagtataaaa ggccagcagc agcctgacca catctcatcc tcctctaagg taaatataaa	180
	atttttaagt gtataatgtg ttaactact gattctaatt gtttgtgtat ttagattcc	240
	aacctatgga actga	255
5	<210> 113 <211> 94 <212> ADN <213> Secuencia artificial	
10	<220> <223> secuencia promotora recombinante <400> 113	
	gttaatttttt aaaaagcagt caaaagtcca agtggccctt gcgagcattt actctctctg	60
	tttgctctgg ttaataatct caggagcaca aaca	94
15	<210> 114 <211> 43 <212> ADN <213> Secuencia artificial	
20	<220> <223> secuencia promotora recombinante	
25	<400> 114 gaggttaata atttccaga tctctctgag caatagtata aaa43	
30	<210> 115 <211> 55 <212> ADN <213> Secuencia artificial	
	<220> <223> secuencia promotora recombinante	
35	<400> 115 agtcatatgt ttgctcactg aaggttacta gttaacaggc atccctaaa cagga	55
40	<210> 116 <211> 108 <212> ADN <213> Secuencia artificial	
	<220> <223> secuencia promotora recombinante	
45	<400> 116	
	gttaatcatt aacttaaaaa gcagtcaaaa gtccaaagggt caaaggtcag agcatttact	60
	ctctccaatg ttgactctcg ttaatgatta aggagcaatt gttgactt	108
50	<210> 117 <211> 72	

ES 2 763 551 T3

	<212> ADN		
	<213> Secuencia artificial		
	<220>		
5	<223> secuencia promotora recombinante		
	<400> 117		
	gttaatcatt aacttaggtc aaaggtcaga caatgttgac tctcgttaat gattaaccgg	60	
	aattgttgac tt	72	
10	<210> 118		
	<211> 151		
	<212> ADN		
	<213> Secuencia artificial		
15	<220>		
	<223> secuencia promotora recombinante		
	<400> 118		
20	gttaatcatt aacttaaaaa gcagtcaaaa gtccaaaggt caaaggtcag agcatttact	60	
	ctctccaatg ttgactctcg ttaatgatta aggagcaatt gttgacttga ggttaataat	120	
	tttccagatc tctctgagca atagtataaa a	151	
	<210> 119		
	<211> 115		
25	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> secuencia promotora recombinante		
30	<400> 119		
	gttaatcatt aacttaggtc aaaggtcaga caatgttgac tctcgttaat gattaaccgg	60	
	aattgttgac ttgaggttaa taattttcca gatctctctg agcaatagta taaaa	115	
35	<210> 120		
	<211> 152		
	<212> ADN		
	<213> Secuencia artificial		
40	<220>		
	<223> secuencia promotora recombinante		
	<400> 120		
	gttaatcatt aagtcgttaa tttttaaaaa gcagtcaaaa gtccaagtgg cccttgcgag	60	
	catttactct ctctgtttgc tctggttaat aatctcagga gcacaaacag aggttaataa	120	
45	ttttccagat ctctctgagc aatagtataa aa	152	
	<210> 121		
	<211> 150		

ES 2 763 551 T3

<212> ADN
 <213> Secuencia artificial
 <220>
 5 <223> secuencia promotora recombinante
 <400> 121

 tgggcggagt gtcgtaatt tttaaaaagc agtcaaaagt ccaagtggcc cttgcgagca 60
 tttactctct ctgtttgctc tggttaataa tctcaggagc acaaacagag gttaataatt 120

 ttccagatct ctctgagcaa tagtataaaa 150
 10
 <210> 122
 <211> 43
 <212> ADN
 <213> Secuencia artificial
 15
 <220>
 <223> secuencia promotora recombinante
 <400> 122
 20 cccttgctgg ttaataatct cagtaattt gttgcacaa aca 43
 <210> 123
 <211> 1467
 <212> PRT
 25 <213> Secuencia artificial
 <220>
 <223> fVIII recombinante
 30 <400> 123

ES 2 763 551 T3

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

ES 2 763 551 T3

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

ES 2 763 551 T3

Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro	405	410	415
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn	420	425	430
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met	435	440	445
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu	450	455	460
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu	465	470	475
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro	485	490	495
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys	500	505	510
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe	515	520	525
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp	530	535	540
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg	545	550	555
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu	565	570	575
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val	580	585	590
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu	595	600	605
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp	610	615	620
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val	625	630	635
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp	645	650	655

ES 2 763 551 T3

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
 660 665 670
 Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
 675 680 685
 Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
 690 695 700
 Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
 705 710 715 720
 Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
 725 730 735
 Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
 740 745 750
 Asn Asn Ala Ile Glu Pro Arg Ser Phe Ala Gln Asn Ser Arg Pro Pro
 755 760 765
 Ser Ala Ser Ala Pro Lys Pro Pro Val Leu Arg Arg His Gln Arg Asp
 770 775 780
 Ile Ser Leu Pro Thr Phe Gln Pro Glu Glu Asp Lys Met Asp Tyr Asp
 785 790 795 800
 Asp Ile Phe Ser Thr Glu Thr Lys Gly Glu Asp Phe Asp Ile Tyr Gly
 805 810 815
 Glu Asp Glu Asn Gln Asp Pro Arg Ser Phe Gln Lys Arg Thr Arg His
 820 825 830
 Tyr Phe Ile Ala Ala Val Glu Gln Leu Trp Asp Tyr Gly Met Ser Glu
 835 840 845
 Ser Pro Arg Ala Leu Arg Asn Arg Ala Gln Asn Gly Glu Val Pro Arg
 850 855 860
 Phe Lys Lys Val Val Phe Arg Glu Phe Ala Asp Gly Ser Phe Thr Gln
 865 870 875 880
 Pro Ser Tyr Arg Gly Glu Leu Asn Lys His Leu Gly Leu Leu Gly Pro
 885 890 895
 Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr Phe Lys Asn
 900 905 910

ES 2 763 551 T3

Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile Ser Tyr Pro
 915 920 925

Asp Asp Gln Glu Gln Gly Ala Glu Pro Arg His Asn Phe Val Gln Pro
 930 935 940

Asn Glu Thr Arg Thr Tyr Phe Trp Lys Val Gln His His Met Ala Pro
 945 950 955 960

Thr Glu Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr Phe Ser Asp Val
 965 970 975

Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly Pro Leu Leu Ile
 980 985 990

Cys Arg Ala Asn Thr Leu Asn Ala Ala His Gly Arg Gln Val Thr Val
 995 1000 1005

Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser
 1010 1015 1020

Trp Tyr Phe Thr Glu Asn Val Glu Arg Asn Cys Arg Ala Pro Cys
 1025 1030 1035

His Leu Gln Met Glu Asp Pro Thr Leu Lys Glu Asn Tyr Arg Phe
 1040 1045 1050

His Ala Ile Asn Gly Tyr Val Met Asp Thr Leu Pro Gly Leu Val
 1055 1060 1065

Met Ala Gln Asn Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly
 1070 1075 1080

Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val Phe
 1085 1090 1095

Ser Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Val Tyr Asn Leu
 1100 1105 1110

Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys Val
 1115 1120 1125

Gly Ile Trp Arg Ile Glu Cys Leu Ile Gly Glu His Leu Gln Ala
 1130 1135 1140

Gly Met Ser Thr Thr Phe Leu Val Tyr Ser Lys Lys Cys Gln Thr

ES 2 763 551 T3

1145		1150		1155
Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile Thr	1160	1165		1170
Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu	1175	1180		1185
His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe	1190	1195		1200
Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His Gly	1205	1210		1215
Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile	1220	1225		1230
Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln	1235	1240		1245
Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe Gly	1250	1255		1260
Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro Pro	1265	1270		1275
Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser Ile	1280	1285		1290
Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn Ser	1295	1300		1305
Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp Ala	1310	1315		1320
Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp	1325	1330		1335
Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn Ala	1340	1345		1350
Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val Asp	1355	1360		1365
Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly Val	1370	1375		1380

ES 2 763 551 T3

Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile Ser
1385 1390 1395

Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn Gly
1400 1405 1410

Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro Val
1415 1420 1425

Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile
1430 1435 1440

His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu Val
1445 1450 1455

Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1460 1465

<210> 124

<211> 1386

<212> ADN

<213> Secuencia artificial

<220>

<223> secuencia promotora recombinante

<400> 124

ES 2 763 551 T3

```

atgcagaggg tcaatatgat catggctgaa tctcctgggc tgatcaccat ttgcctgctg      60
ggatacctgc tgtctgctga gtgtacagtg ttcctggacc atgagaatgc caataagatc      120
ctgaacaggc ccaaaagata caattctgga aagctggagg aattttgtgca gggcaacctg      180
gagaggggaat gcatggagga aaagtgtagc tttgaggaag ctagggaggt gtttgaaaac      240
acagagagga ccacagaatt ctggaagcag tatgtggatg gagatcagtg tgagtccaac      300
ccctgtctga atggagggtc ttgcaaagat gatatcaact cctatgagtg ctggtgtcct      360
tttgattttg aaggcaaaaa ttgtgagctg gatgtgacct gtaacatcaa gaatggcagg      420
tgtgagcagt tctgtaaaaa ctctgctgat aataagggtg tctgcagctg tacagaaggc      480
tataggctgg ctgagaacca gaagagctgt gaaccagctg tgcccttccc ttgtgggagg      540
gtgtctgtca gccagacctc taagctgacc agagctgaga ctgtgttccc agatgtggat      600
tatgtcaact ccacagaggc tgaaaccatc ctggacaaca tcacccagtc taccagtc     660
ttcaatgact ttaccagagt ggtgggagga gaggatgcc aaccaggcca gttcccctgg      720
caggtggtcc tgaatgggaa ggtggatgct ttttgtgggg gatccattgt gaatgagaaa      780
tggtattgtca cagctgctca ctgtgtggag acaggggtca agatcactgt ggtggctgga      840
gagcacaaca ttgaggaaac agaacatact gagcagaaga ggaatgtgat cagaatcatc     900

cctcaccata actacaatgc tgctatcaac aaatataatc atgacattgc cctgctggaa      960
ctggatgagc ctctggtgct gaacagctat gtcaccccaa tctgcattgc tgacaaagag     1020
tataccaata tcttcctgaa gtttgatctt ggatatgtgt ctggatgggg aagggtcttc     1080
cacaagggca ggtctgccct ggtgctgcag tatctgaggg tgccctctggt ggacagagct     1140
acctgcctgc tgtctaccaa gttcaccatc tacaacaata tgttctgtgc tggatttcat     1200
gagggaggca gggactcctg tcagggggat tctggaggcc cacatgtgac agagggtgaa     1260
ggcaccagct tcctgactgg catcatctct tggggggagg aatgtgctat gaaggggaaa     1320
tatggaatct acaccaaagt gagcaggtat gtgaactgga tcaaagagaa gaccaaactg     1380
acctga                                           1386

```

<210> 125
 <211> 4404
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> secuencia promotora recombinante
 <400> 125

ES 2 763 551 T3

atgcagctgg agctctcaac ctgtgtgttc ctctgcctgc tccccctggg attttcagct	60
atcaggagat actatctggg agcagtggaa ctgtcctggg actacaggca gtcagagctg	120
ctcagagaac tgcatgtgga tactaggttc cctgcaacag ctccctggagc actgccactg	180
ggaccttcag tgctgtacaa gaaaactgtc tttgtggagt ttacagacca gctgttcagt	240
gtggccaggc ccaggccccc ctggatgggg ctgctgggac ccaccatcca ggctgaagtg	300
tatgatactg tgggtgtgac cctgaaaaac atggcctctc atccagtcag cctgcatgct	360
gtgggagtga gcttctggaa gagcagtgag ggagctgagt atgaagacca tacctcacag	420
aggagaaaag aagatgataa ggtgctgcca ggaaaaagcc agacctatgt gtggcaggtg	480
ctgaaggaga atggccctac agcttcagat cctccctgcc tcacatactc ttatctgagc	540
catgtggatc tgggtgaagga cctcaatagt ggcctgattg gggcactgct ggtgtgcaga	600
gaggggtccc tcacaaggga aagaactcag aacctgcatg agtttgtcct gctctttgct	660
gtgtttgatg agggaaagt cctggcactca gcaaggaatg acagctggac cagggctatg	720
gaccagcac cagccagagc tcagccagct atgcacactg tcaatggcta tgtgaatagg	780
tccctgcctg gactcattgg ctgccataag aaatcagtct attggcatgt gattggaatg	840
ggcaccagcc cagaggtgca ttccatcttc ctggaaggcc acacatttct ggtcaggcac	900
catagacagg ccagcctgga gatcagccca ctgactttcc tcacagcaca gacatttctg	960
atggacctgg ggcagttcct gctcttttgc catatctcaa gtcaccatca tggagggatg	1020
gaggctcatg tcagggtgga aagctgtgca gaggaacctc agctgaggag gaaggcagat	1080

ES 2 763 551 T3

gaggaagagg	actatgatga	taacctgtat	gactcagata	tggatgtggt	gaggctggat	1140
ggagatgatg	tcagcccatt	catccagatc	aggtcagtgg	ctaagaaaca	ccctaagacc	1200
tgggtccact	acattgcagc	tgaagaggaa	gattgggact	atgcaccctt	ggtgctggcc	1260
ccagatgata	gaagttacaa	atctcagtat	ctgaacaatg	ggccccagag	gattggaagg	1320
aagtacaaga	aagtgaggtt	catggcttat	actgatgaga	cctttaagac	aagagaggca	1380
atccagcatg	aaagtggcat	cctgggacca	ctgctctatg	gagaagtggg	ggataccctg	1440
ctcatcatct	tcaagaacca	ggcctcaagg	ccttacaata	tctatcccca	tggcatcaca	1500
gatgtgaggc	ctctctacag	caggagactg	cccaagggag	tcaaacacct	caaggatttc	1560
cccatcctgc	caggggaaat	cttcaagtat	aaatggacag	tactgtgga	agatgggcca	1620
actaagtcat	atcctaggtg	cctgaccagg	tactattcta	gctttgtgaa	catggagagg	1680
gacctggctt	caggactgat	tggacctctg	ctcatctgct	acaagaatc	agtggaccag	1740
aggggcaacc	agatcatgag	tgataagaga	aatgtcatcc	tgttctcagt	gtttgatgag	1800
aataggagtt	ggtatctgac	agaaaacatc	cagaggttcc	tgcctaatac	tgcaggagtg	1860
cagctggagg	accagaatt	tcaggcttca	aacatcatgc	atagtatcaa	tggctatgtg	1920
tttgatagtc	tgcagctctc	tgtctgcctg	catgaggtgg	cctactggta	tatcctcagc	1980
attggagctc	agactgactt	cctgagtgtg	ttcttttcag	gctacacatt	caagcataag	2040
atggtctatg	aagataccct	gacactcttc	cccttttctg	gggagactgt	gtttatgagc	2100
atggaaaacc	caggcctgtg	gattctgggg	tgccacaaca	gtgacttcag	gaatagaggg	2160
atgactgctc	tgctcaaagt	gtcctcatgt	gataagaata	ctggagatta	ctatgaggac	2220
tcttatgaag	atatcagtgc	atatctgctc	tccaaaaaca	atgccattga	gcccaggtca	2280
tttgctcaga	acagtagacc	accttctgca	agtgcaccaa	agcctccagt	gctgaggaga	2340
caccagaggg	acatcagcct	gccaaccttc	cagcctgagg	aagataaaat	ggactatgat	2400
gatatcttct	ccactgagac	caagggggaa	gattttgaca	tctatggaga	ggatgaaaac	2460
caggacccca	ggtccttcca	gaagaggacc	agacactact	ttattgcagc	tgtggagcag	2520
ctgtgggact	atggcatgtc	tgaatcacct	agagctctga	ggaacagagc	acagaatggg	2580
gaggtgccca	ggttcaagaa	agtgtgttcc	agagaatttg	cagatggctc	ttttaccag	2640
cctagctaca	ggggggagct	caacaagcat	ctggggctgc	tgggacccta	tatcagagca	2700
gaggtggaag	ataacatcat	ggtgacattc	aagaatcagg	cctcaagacc	ctacagtttt	2760
tatagttctc	tgatcagcta	cccagatgat	caggagcagg	gggctgaacc	aaggcacaac	2820
tttgtgcagc	ctaatgagac	aagaacttac	ttttggaagg	tccagcatca	catggctccc	2880
acagaggatg	agtttgactg	caaggcctgg	gcataatttt	ctgatgtgga	cctggagaag	2940

ES 2 763 551 T3

gatgtgcata gtggcctcat tgggccactg ctcatctgca gggcaaacac actgaatgct 3000
gcacatggca ggcaggtcac tgtgcaggag tttgccctgt tctttacaat ctttgatgaa 3060
actaagtcct ggtacttcac agagaatgtg gaaaggaatt gcagagcccc ctgccatctc 3120
cagatggagg acccaactct gaaggaaaac tacaggttcc atgctatcaa tggatatgtc 3180
atggatacac tgccaggcct ggtgatggca cagaaccaga ggatcagggtg gtatctgctc 3240
agcatggggt ccaatgagaa tatccattct atccacttct caggacatgt cttttcagtg 3300
aggaagaaag aggaatataa aatggctgtg tacaatctgt atccaggggt ctttgagaca 3360
gtggaaatgc tgcctagcaa agtggggatc tggagaattg agtgcctcat tggagaacac 3420
ctgcaggcag ggatgtccac cacatttctg gtgtactcaa agaaatgcca gactcccctg 3480
gggatggcaa gtggacatat cagggacttc cagatcactg catcaggaca gtatggacag 3540
tgggcaccaa agctggctag gctccactat agtggctcta tcaatgcttg gagtaccaa 3600
gagcctttct cttggatcaa ggtggatctg ctggccccc tgatcatcca tggaatcaaa 3660
acacaggag ctagacagaa gttcagctcc ctgtacatca gtcagtttat catcatgtat 3720
tctctggatg ggaagaaatg gcagacctac aggggcaata gcaactgggac actgatggtc 3780
ttctttggaa atgtggattc aagtggcatc aagcacaaca tcttcaatcc tcccatcatt 3840
gccaggtaca tcagactgca tcccacacac tattcaatca ggagtactct cagaatggag 3900
ctgatggggt gtgacctcaa cagctgctcc atgccactgg gaatggaatc caaggcaatc 3960
tcagatgccc agatcactgc ttctagctac ttcaccaata tgtttgcaac atggtcaccc 4020
agtaaagcaa ggctgcacct ccagggaagg tccaatgctt ggagacccca ggtgaacaat 4080
ccaaaggagt ggctgcagggt ggactttcag aaaaccatga aggtcacagg ggtgactacc 4140
cagggagtga aaagtctgct cacctctatg tatgtcaagg agttcctgat ctccctcaagt 4200
caggatggcc accagtggac actgttcttt cagaatggca aggtcaaagt gttccagggg 4260
aatcaggaca gctttacacc agtgggtgaac agcctggacc cccctctgct cactagatat 4320
ctgagaatcc atccacagag ctgggtgcac cagattgcac tcagaatgga ggtcctgggc 4380
tgtgaagccc aggacctgta ttga 4404

<210> 126

<211> 4374

5 <212> ADN

<213> Secuencia artificial

<220>

10 <223> secuencia promotora recombinante

<400> 126

atgcagattg agctcagcac ctgcttcttt ctgtgcctgc tcaggttctg cttttcagcc 60

acaaggagat actatctggg agctgtggaa ctgtcatggg attacatgca gagtgcacctg 120

ggagagctcc ctgtggatgc taggttcccc ccaaggggtcc caaagtcttt cccttttaat	180
accagtgtgg tctataagaa aacactcttt gtggaattta ctgatcacct gttcaacatt	240
gcaaagccaa ggccctccctg gatgggactg ctgggaccta ccatccaggc tgaggtgtat	300
gacactgtgg tcatcacact gaaaaacatg gcatctcacc ctgtcagcct gcatgcagtg	360
ggagtgcagct actggaaggc ttcagaaggg gcagagtatg atgatcagac aagccagaga	420
gaaaaagagg atgataaggt gttcccagga gggagccata cttatgtgtg gcaggtcctg	480
aaggagaatg gcccaatggc cagtgaacca ctgtgcctca cctactcata tctgagtcatt	540
gtggacctgg tcaaggatct caactcaggc ctgattgggg cactgctggg gtgcagggaa	600
ggctcactgg ccaaggagaa aaccacagaca ctgcataagt tcatcctgct ctttgcctgtg	660
tttgatgaag ggaaatcttg gcacagttag accaagaaca gtctgatgca ggacagggat	720
gctgcttctg ccagagcttg gcccaagatg cacacagtga atggatatgt caataggctc	780
ctgccaggac tcattggctg ccacagaaag tcagtgtatt ggcattgcat tggaatgggc	840
accacaccag aagtgcacag catcttcctg gaggggcata cctttctggg caggaaccac	900
aggcaggcca gcctggagat cagcccaatc accttcctga cagcccagac tctgctcatg	960
gatctggggc agttcctgct cttttgccac atcagctccc accagcatga tggaatggag	1020
gcatatgtga aagtggactc ctgccagag gaaccacagc tgaggatgaa gaacaatgag	1080
gaagctgaag actatgatga tgacctgaca gactcagaga tggatgtggg caggtttgat	1140
gatgataaca gccctcctt tatccagatc agaagtgtgg ccaagaaaca ccaaagaca	1200
tgggtccatt acattgcagc tgaggaagag gactgggatt atgcacctct ggtgctggcc	1260
ccagatgata gatcctacaa atcacagtat ctgaacaatg gaccccagag gattggcaga	1320
aagtacaaga aagtgaggtt catggcctat actgatgaaa catttaagac tagagaagct	1380
atccagcatg agtcaggcat cctgggacca ctgctctatg gagaagtggg ggacaccctg	1440
ctcatcatct tcaagaacca ggcttccagg ccatacaata tctatcctca tggcatcaca	1500
gatgtgagac cactctactc aaggagactg cctaagggag tcaaacacct caaggacttc	1560
cctatcctgc caggggaaat ctttaagtat aaatggactg tgacagtgga ggatgggccc	1620
actaagagtg acccaaggtg cctgaccaga tactattcaa gttttgtgaa tatggaaagg	1680
gatctggcat caggactgat tggacctctg ctcatctgct acaaagagag tgtggatcag	1740
aggggcaacc agatcatgtc agacaagagg aatgtgatcc tgttcagtgt ctttgatgaa	1800
aacaggtctt ggtatctgac agagaacatc cagagattcc tgccaaatcc tgcagggtg	1860
cagctggaag atccagagtt tcaggcctca aacatcatgc atagtatcaa tggatatgtg	1920
tttgacagtc tgcagctctc tgtgtgcctg catgaagtgg cctactggta taccctgtcc	1980

ES 2 763 551 T3

attggagctc	agacagattt	cctgagtgtg	ttcttttcag	gctacacttt	taagcataaa	2040
atggtctatg	aggacacact	gactctcttc	ccttttagtg	gggaaacagt	gtttatgagc	2100
atggagaatc	cagggctgtg	gattctggga	tgccacaaca	gtgatttcag	gaatagaggc	2160
atgactgctc	tgctcaaagt	gtctagctgt	gacaagaaca	caggggacta	ctatgaagat	2220
tcttatgagg	acatcagtgc	ttatctgctc	tccaaaaaca	atgcaattga	accagatca	2280
ttcagtcaga	atccacctgt	gctgaagagg	caccagagag	agatcactag	gactaccctg	2340
cagtcagatc	aggaagagat	tgactatgat	gataccatct	cagtggaaat	gaagaaagag	2400
gactttgata	tctatgatga	agatgagaac	cagagtccaa	ggtctttcca	gaagaaaacc	2460
agacattact	ttattgctgc	agtggagagg	ctgtgggatt	atggaatgtc	ctcaagtcca	2520
catgtgctga	ggaatagggc	acagtctggc	agtgtccctc	agttcaagaa	agtggctctc	2580
caggagttta	cagatggcag	cttcactcag	cctctgtaca	ggggagaact	caatgagcac	2640
ctggggctgc	tgggacccta	tatcagagct	gaagtggagg	ataacatcat	ggtcaccttc	2700
aggaatcagg	cttcaagacc	ctacagtttt	tattctagcc	tgatcagcta	tgaagaggac	2760
cagaggcagg	gagctgaacc	taggaaaaac	tttgtgaagc	caaatgagac	caaaacatac	2820
ttttggaagg	tccagcacca	catggcacca	accaaagatg	agtttgattg	caaggcatgg	2880
gcctattttt	cagatgtgga	tctggagaag	gatgtccaca	gtggcctcat	tgggcctctg	2940
ctggtgtgcc	atactaacac	cctgaatcca	gctcatggca	ggcaggtgac	agtccaggag	3000
tttgactgtt	tctttaccat	ctttgatgag	acaaagtcct	ggtacttcac	tgaaaacatg	3060
gagaggaatt	gcagagctcc	ttgcaacatc	cagatggaag	acccacctt	caaggagaac	3120
tacagatttc	atgcaatcaa	tgggtatatc	atggatacac	tgccaggact	ggtgatggcc	3180
caggaccaga	ggatcagatg	gtatctgctc	agcatggggt	ccaatgagaa	tatccactct	3240
atccatttca	gtggacatgt	gtttacagtc	agaaagaaag	aagagtataa	aatggccctg	3300
tacaacctct	atccaggagt	gtttgaaaca	gtggagatgc	tgccaagcaa	ggctgggac	3360
tggaggggtg	aatgcctcat	tggggagcac	ctgcatgcag	gaatgtcaac	cctgtttctg	3420
gtctacagta	ataagtgcc	gacacctctg	ggaatggcaa	gtggacatat	cagggatttc	3480
cagatcactg	ctagtggaca	gtatggacag	tgggcaccaa	agctggctag	actccactat	3540
tcaggctcaa	tcaatgcttg	gtccaccaa	gagccattct	catggatcaa	ggtggacctg	3600
ctggctccta	tgatcatcca	tggcatcaaa	acacaggggg	caaggcagaa	gttctcctca	3660
ctgtacatct	ctcagtttat	catcatgtat	agcctggatg	gcaagaaatg	gcagacctac	3720
aggggcaata	gcacagggac	tctgatgggt	ttctttggca	atgtggacag	cagtgggatc	3780
aagcacaaca	tcttcaatcc	cccaatcatt	gcaaggtaga	tcagactgca	ccccacccat	3840
tattcaatca	ggagtacact	caggatggaa	ctgatggggt	gtgatctcaa	cagttgctct	3900

ES 2 763 551 T3

atgccactgg gaatggagtc caaggcaatc tcagatgccc agatcactgc tagtcctac	3960
ttcactaata tgtttgctac ctggagcccc tccaaagcaa ggctgcacct ccaggggaagg	4020
agcaatgcat ggaggcctca ggtgaacaat cccaaggaat ggctgcaggt ggatttccag	4080
aaaactatga aggtgactgg agtcacaact cagggagtga aaagtctgct cacttctatg	4140
tatgtcaagg agttcctgat ctcaagttct caggatggcc accagtggac cctgttcttt	4200
cagaatggaa aggtgaaagt cttccagggc aatcaggatt cctttacacc agtgggtcaac	4260
tcactggacc ctccccctgct cactagatat ctgagaatcc accctcagag ctgggtgcat	4320
cagattgctc tcagaatgga agtcctgggc tgtgaggcac aggacctgta ttga	4374

<210> 127

<211> 1386

<212> ADN

<213> Secuencia artificial

<220>

<223> secuencia promotora recombinante

<400> 127

ES 2 763 551 T3

```

atgcagaggg tgaatatgat tatggctgag tcccctgggc tgattacat ttgcctgctg      60
ggatacctgc tgtctgctga gtgtacagtg ttcctggacc atgagaatgc aaataagatc      120
ctgaacaggg ccaaaagata taatagtgga aagctggagg aatttgtgca gggcaacctg      180
gagagagaat gcatggagga aaagtgtagc tttgaggaag ccaggagggt gtttgaaaat      240
acagagagaa ccacagaatt ctggaagcag tatgtggatg gagatcagtg tgagagcaac      300
ccctgtctga atggagggag ttgcaaagat gatatcaact catatgaatg ctggtgtcct      360
tttgattttg aaggcaaaaa ttgtgagctg gatgtgacct gtaacattaa gaatgggagg      420
tgtgagcagt tttgtaaaaa ctctgctgat aataaggtgg tctgcagttg tacagaaggg      480
tatagactgg ctgagaacca gaagtctgtg gaaccagctg tgcccttccc ttgtggaagg      540
gtgtctgtct cccagacttc aaaactgacc agagctgaga ctgtgtttcc tgatgtggat      600
tatgtcaaca gcacagaggc tgaaactatc ctggacaaca ttactcagtc taccagagt      660
ttcaatgact ttaccagagt ggtgggagga gaggatgcta aaccaggcca gttcccctgg      720
caggtggtcc tgaatgggaa ggtggatgca ttttgtggg gatctattgt gaatgagaaa      780
tggtattgtc cagctgctca ctgtgtggaa actggggtca agatcacagt ggtggctgga      840
gagcacaaca ttgaggaaac agaacatact gagcagaaga ggaatgtgat cagaatcatt      900
cctcaccata actacaatgc agccatcaac aaatataatc atgacattgc cctgctggaa      960
ctggatgagc ctctggtgct gaacagctat gtcacaccaa tctgcattgc tgacaaggag     1020
tacactaaca tcttcctgaa gtttgggtca ggatatgtgt ctggatgggg aagagtcttc     1080

cacaagggca ggtctgcact ggtgctgcag tatctgagag tgcctctggt ggatagggcc     1140
acttgtctgc tgtctaccaa gttcaccatc tacaacaata tgttctgtgc tggatttcat     1200
gagggaggga gagactcctg tcaggagatg tctggaggcc cacatgtgac agaggtggaa     1260
ggcaccagct tcctgacagg catcatttcc tggggggagg aatgtgcaat gaaggggaaa     1320
tatggaatct acaccaaagt gagcaggtat gtgaactgga tcaaggaaaa gaccaaactg     1380
acatga                                             1386

```

<210> 128

<211> 252

<212> ADN

<213> Secuencia artificial

<220>

<223> secuencia promotora recombinante

<400> 128

ES 2 763 551 T3

tggtttgctgc ttgcaatggt tgcccatttt aggggtggaca caggacgctg tggtttctga	60
gccaggggggc gactcagatc ccagccagtg gacttagccc ctgtttgctc ctccgataac	120
tgggggtgacc ttggttaata ttcaccagca gcctcccccg ttgcccctct ggatccactg	180
cttaaatacg gacgaggaca gggccctgtc tcctcagctt caggcaccac cactgacctg	240
ggacagtga tc	252

<210> 129
 <211> 4882
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> secuencia de AAV recombinante
 <400> 129

cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcaaag cccgggcgtc	60
ggggcgcctt tggtcgcccgc gcctcagtga gcgagcgagc gcgcagagag ggagtggcca	120
actccatcac taggggttcc taccggtggt aatcattaag tcgttaattt ttgtggcctt	180
tgcgatgttt gctctggtta ataatctcag gacaaacaga ggtaataat tttccagatc	240
tctctgagca atagtataaa aggccagcag cagcctgacc acatctcatc ctgctcgagc	300
caccatgcag atcgaaactgt ctacctgttt ctttctgtgc ctgctgcggt tttgtttttc	360
cgctaccaga agatactacc tgggagccgt cgaactgagc tgggattaca tgcagtctga	420
cctgggagag ctgcccgtgg acgctagatt cccacctaga gtccctaagt ccttcccctt	480
caacaccagc gtggtctaca agaaaaccct gttcgtggag tttaccgacc acctgttcaa	540
catcgctaag cctagaccac catggatggg actgctggga ccaaccatcc aggccgaggt	600
gtacgacacc gtggtcatca ccctgaaaaa catggcttct cccccgtgt ccctgcatgc	660

tgtgggcgtc	tcctactgga	aggccagcga	aggggctgag	tatgacgata	agaccagcca	720
gcgggaaaaa	gaggacgata	aggtgttccc	tggcgggtcc	catacctacg	tgtggcaggt	780
cctgaaggag	aatggaccaa	tggcttccga	ccctctgtgc	ctgacctact	cttatctgtc	840
ccacgtggac	ctggtcaagg	atctgaacag	cggcctgata	ggggctctgc	tgggtgtgtc	900
cgaaggggtc	ctggccaagg	agaaaaccca	gaccctgcat	aagttcatcc	tgctgttcgc	960
cgtgtttgac	gaaggaaaaa	gctggcactc	tgagaccaag	aactctctga	tgcaggacag	1020
ggatgccgct	tccgccagag	cttggtccca	gatgcacacc	gtgaacggct	acgtcaatag	1080
gagcctgcct	ggactgatac	gctgccacag	aaagtccgtg	tattggcatg	tcatcggat	1140
gggcaccacc	cctgaagtgc	acagcatctt	cctggagggg	catacctttc	tgggtccgca	1200
ccaccggcag	gctagcctgg	agatctctcc	aatcaccttc	ctgaccgccc	agaccctgct	1260
gatggacctg	ggacagttcc	tgctgttttg	ccacatctcc	agccaccagc	atgatggcat	1320
ggaggcttac	gtgaaagtgc	actcctgtcc	cgaggaacct	cagctgagga	tgaagaacaa	1380
tgaggaagcc	gaagactatg	acgatgacct	gaccgacagc	gagatggatg	tgggtccgctt	1440
cgatgacgat	aactctccct	cctttatcca	gatccgggtc	gtggccaaga	aacaccctaa	1500
gacctgggtc	cattacatcg	ccgctgagga	agaggactgg	gattatgctc	cactgggtgct	1560
ggcccccgac	gatagatcct	acaaaagcca	gtatctgaac	aatggacccc	agaggatcgg	1620
cagaaagtac	aagaaagtga	ggttcatggc	ttataccgat	gagaccttta	agaccagaga	1680
agccatccag	cacgagtcgg	ggatcctggg	acctctgctg	tacggcgaag	tgggggacac	1740
cctgctgata	atcttcaaga	accaggccag	caggccttac	aatatctata	cacatggcat	1800
caccgatgtg	agacctctgt	actcccggcg	gctgccaaag	ggcgtgaaac	acctgaagga	1860
cttcccaatc	ctgcccgggg	aaatctttaa	gtataaatgg	accgtcaccg	tcgaggatgg	1920
gcccaccaag	agcgacccta	ggtgcctgac	cagatactat	tcttccttcg	tgaatatgga	1980
gagagacctg	gcttccggac	tgatcggacc	cctgctgata	tgttacaaag	agagcgtgga	2040
tcagcgcggc	aaccagatca	tgtctgacaa	gcggaatgtg	atcctgttca	gcgtctttga	2100
cgaaaaccgc	tcttggtagc	tgaccgagaa	catccagcgg	ttcctgccta	atccagctgg	2160
agtgcagctg	gaagatcccg	agttccaggc	ctctaacatc	atgcattcca	tcaatggcta	2220
cgtgttcgac	tccctgcagc	tgagcgtgtg	cctgcacgag	gtcgcttact	ggtatatcct	2280
gagcatcgga	gcccagaccg	atttcctgtc	tgtgttcttt	tccggctaca	cctttaagca	2340
taaaatggtg	tatgaggaca	ccctgaccct	gttcccattt	tccggcgaaa	ccgtgttcat	2400
gagcatggag	aatcccgggc	tgtggatcct	gggatgccac	aactccgatt	tcaggaatag	2460
agggatgacc	gccctgctga	aagtgagctc	ttgtgacaag	aacaccggag	actactatga	2520

agatagctac gaggacatct ctgcttatct gctgtccaaa aacaatgcc a tcgagcccag	2580
gagcttctct cagaaccctc cagtgtctgaa gcgccaccag cgggagatca ccagaaccac	2640
cctgcagagc gatcaggaag agatcgacta cgacgatacc atctccgtgg aaatgaagaa	2700
agaggacttc gatatctatg acgaagatga gaaccagtct cccaggtcct tccagaagaa	2760
aaccagacat tacttttatcg ccgctgtgga ggggtgtgtg gactatggca tgtccagctc	2820
tcctcacgtg ctgagaaata gagctcagtc cggaagcgtc ccacagttca agaaagtggg	2880
cttcaggag tttaccgacg gaagctttac ccagccactg taccgcggcg aactgaacga	2940
gcacctgggg ctgctgggac cctatatccg ggctgaagtg gaggataaca tcatgggtcac	3000
cttcaggaat caggccagca gaccctactc tttttattcc agcctgatct cctacgaaga	3060
ggaccagaga cagggagctg aaccaagaaa aaacttcgtg aagcctaag agaccaaacc	3120
ctacttttgg aaggtgcagc accatatggc ccctaccaa gacgagttcg attgcaaggc	3180
ctgggcttat tttagcgacg tggatctgga gaaggacgtc cactccggcc tgatcggggc	3240
actgctggtg tgtcatacca acaccctgaa tccagctcac ggaaggcagg tgaccgtcca	3300
ggaattcgcc ctgttcttta ccatctttga tgagaccaag agctggtact tcaccgaaaa	3360
catggagagg aattgcagag ccccatgtaa catccagatg gaagaccca cttcaagga	3420
gaactacaga tttcatgcta tcaatgggta tatcatggat accctgccag gactgggtcat	3480
ggctcaggac cagaggatca gatggtacct gctgagcatg gggctaacg agaatatcca	3540
ctccatccat ttcagcggac acgtgtttac cgtccgcaag aaagaagagt acaagatggc	3600
cctgtacaac ctgtatcccg gcgtgttcga aaccgtcgag atgctgcctt ccaaggctgg	3660
gatctggcgg gtggaatgcc tgatcgggga gcacctgcat gccggaatgt ctaccctgtt	3720
cctggtgtac tccaataagt gtcagacccc cctggggatg gctagcggac atatccgcga	3780
cttcagatc accgcttcg gacagtacgg acagtgggt cctaagctgg ctagactgca	3840
ctattctggc tccatcaacg cttggtctac caaagagcct ttctcctgga tcaaggtgga	3900
cctgctggct ccaatgatca tccatggcat caaaaccag ggggccaggc agaagttctc	3960
ttccctgtac atcagccagt ttatcatcat gtattctctg gatgggaaga aatggcagac	4020
ctacagaggc aattccaccg ggaccctgat ggtgttcttt ggcaacgtcg acagctctgg	4080
gatcaagcac aacatcttca atccccctat catcgccgc tacatccggc tgcacccaac	4140
ccattattcc atccgcagca ccctgcggat ggagctgatg gggcgcatc tgaacagctg	4200
ttctatgcc ctgggaatgg agtctaaggc catctccgac gctcagatca ccgcctccag	4260
ctacttcacc aatatgtttg ctacctgtc cccaagcaag gctagactgc atctgcaggg	4320
aagaagcaac gcttgagac cacaggtgaa caatcccaag gagtggctgc aggtcgactt	4380
ccagaaaacc atgaaggtga ccggagtcac caccagggc gtgaaaagcc tgctgacctc	4440

ES 2 763 551 T3

tatgtacgtc aaggagttcc tgatctcttc cagccaggac gggcaccagt ggaccctggt	4500
ctttcagaac ggaaaggtga aagtcttcca gggcaatcag gattccttta cccctgtggt	4560
caacagcctg gaccaccccc tgctgaccag gtacctgaga atccaccac agtcctgggt	4620
gcatcagatc gctctgagga tggaagtcct gggctgogag gcccaggacc tgtattgagc	4680
ggccgcaata aaatatcttt attttcatta catctgtgtg ttggtttttt gtgtgcaatt	4740
gaggaacccc tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag	4800
gccgggagac caaaggctgc ccgacgcccg ggctttgcc gcccggcctc agtgagcgag	4860
cgagcgcgca gctgcctgca gg	4882

<210> 130
 <211> 5018
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> secuencia de AAV recombinante
 <400> 130

ES 2 763 551 T3

cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcaaag cccgggcgctc	60
gggcgacctt tggtcgcccc gcctcagtga gcgagcgagc gcgcagagag ggagtggcca	120
actccatcac taggggttcc taccggtggt aatcattaag tcgttaattt ttgtggccct	180
tgcgatgttt gctctggtta ataatctcag gacaaacaga ggttaataat tttccagatc	240
tctctgagca atagtataaa aggccagcag cagcctgacc acatctcatc ctcgctcgact	300
taattaaaag aggtaagggt ttaagggatg gttggttggg ggggtattaa tgtttaatta	360
cctggagcac ctgcctgaaa tcactttttt tcaggttggc tcgagccacc atgcagctgg	420
aactgtctac ctgtgtgttt ctgtgtctgc tgccctctggg gttttctgct atcaggagat	480
actatctggg agctgtggag ctgtcctggg actacaggca gtctgagctg ctgagagaac	540
tgcattgtga taccagattc ccagctacag ctccaggagc tctgcctctg ggcccatctg	600
tgctgtacaa gaaaacagtc tttgtggagt ttacagacca gctgttctct gtggccaggc	660
caagaccacc ttggatggga ctgctgggac caaccatcca ggctgaggtg tatgatacag	720
tgggtggtgac cctgaaaaac atggcctccc atcctgtgag cctgcatgct gtgggggtgt	780
ccttctggaa gtcctctgag ggagctgagt atgaagacca tacctcccag agggagaaag	840
aagatgataa ggtgctgcct ggcaaaagcc agacctatgt ctggcaggtg ctgaaggaga	900
atggaccaac tgcttctgac ccacatgcc tgacctactc ttatctgtcc catgtggatc	960
tgggtgaagga cctgaattct ggactgattg gagctctgct ggtgtgtaga gagggaagcc	1020
tgaccagaga aagaaccag aacctgcatg agtttgcct gctgtttgct gtgtttgatg	1080

ES 2 763 551 T3

aaggggaagag	ctggcactct	gccaggaatg	actcctggac	cagagctatg	gatccagctc	1140
ctgctagagc	tcagcctgct	atgcacacag	tcaatggcta	tgtgaatagg	tctctgccag	1200
gactgattgg	ctgccataag	aaatctgtct	attggcatgt	gattggaatg	ggcaccagcc	1260
ctgaggtgca	ttctatcttc	ctggaaggcc	acacctttct	ggtcaggcac	catagacagg	1320
cctctctgga	gatctcccct	ctgaccttcc	tgacagctca	gacctttctg	atggacctgg	1380
ggcagttcct	gctgttttgc	catatctctt	cccaccatca	tggaggaatg	gaggctcatg	1440
tcaggggtgga	atcctgtgct	gaggaaccac	agctgagaag	aaaggctgat	gaggaagagg	1500
actatgatga	taacctgtat	gactctgata	tggatgtggt	gaggctggat	ggggatgatg	1560
tcagcccttt	catccagatc	aggtctgtgg	ccaagaaaca	tccaaagacc	tgggtccact	1620
acattgctgc	tgaagaggaa	gattgggact	atgccccct	ggtgctggct	cctgatgata	1680
gacctacaa	aagccagtat	ctgaacaatg	ggccccagag	gattggaagg	aagtacaaga	1740
aagtgaggtt	catggcctat	acagatgaga	cctttaagac	cagagaggct	atccagcatg	1800
aatctgggat	cctgggacct	ctgctgtatg	gagaagtggg	ggataccctg	ctgatcatct	1860
tcaagaacca	ggcctccagg	ccatacaata	tctatcccca	tggcatcaca	gatgtgagac	1920
cactgtacag	caggagactg	cccaaggggg	tcaaacacct	gaaggatttc	cccatcctgc	1980
ctggagagat	ctttaagtat	aaatggacag	tcacagtgga	agatgggcct	accaagtctg	2040
atccaaggtg	cctgaccaga	tactatagct	cttttgtgaa	catggagaga	gacctggctt	2100
ctggactgat	tggaccctg	ctgatctgtt	acaagagtc	tgtggaccag	aggggcaacc	2160
agatcatgtc	tgataagaga	aatgtcatcc	tgttctctgt	gtttgatgag	aacaggagct	2220
ggtacctgac	agagaacatc	cagaggttcc	tgccaaatcc	agctggagtg	cagctggagg	2280
accagaatt	tcaggcttcc	aacatcatgc	atagcatcaa	tggctatgtg	tttgatagcc	2340
tgcagctgtc	tgtctgcctg	catgaggtgg	cctactggta	tatcctgtcc	attggagctc	2400
agacagactt	cctgtctgtg	ttcttttagtg	ggtacacctt	taagcataaa	atgggtgatg	2460
aggataccct	gaccctgttc	cccttttctg	gggagacagt	gttcatgtcc	atggaaaacc	2520
ctggcctgtg	gatcctgggg	tgccacaact	ctgacttcag	gaatagagga	atgacagccc	2580
tgctgaaagt	gtccagctgt	gataagaata	caggggatta	ctatgaggac	tcttatgaag	2640
atatctctgc	ttatctgctg	agcaagaaca	atgccattga	gcccaggtct	tttgctcaga	2700
actccagacc	tccatctgct	tctgctccta	agccacctgt	gctgagaaga	catcagaggg	2760
acatctccct	gcctaccttc	cagccagagg	aagataaaat	ggactatgat	gatatcttca	2820
gcacagagac	caagggggaa	gattttgaca	tctatggaga	ggatgaaaac	caggatccaa	2880
gaccttcca	gaagagaacc	agacactact	ttattgctgc	tgtggagcag	ctgtgggact	2940
atgggatgtc	tgaaagccca	agggccctga	ggaacagagc	tcagaatgga	gaggtgccca	3000

gattcaagaa agtgggtgttc agagagtttg ctgatggcag ctttaccag ccatcttaca	3060
ggggggagct gaacaagcat ctggggctgc tgggacccta tatcagagct gaggtggaag	3120
ataacatcat ggtgaccttc aagaatcagg cttctaggcc ctactccttt tattcttccc	3180
tgatctccta ccctgatgat caggagcagg gagctgaacc taggcacaac tttgtgcagc	3240
caaatgagac cagaacctac ttttggaagg tgcagcatca catggctccc acagaggatg	3300
aatttgactg caaagcttgg gcctatTTTT ctgatgtgga cctggagaag gatgtgcatt	3360
ctggcctgat tgggcctctg ctgatctgta gggccaacac cctgaatgct gctcatggaa	3420
gacaggtcac agtgcaggag tttgctctgt tctttaccat ctttgatgaa accaagagct	3480
ggtacttcac agagaatgtg gaaaggaatt gcagagcccc ctgtcatctg cagatggagg	3540
accctaccct gaaggaaaac tacaggttcc atgccatcaa tggatatgtc atggataccc	3600
tgcctggcct ggtcatggct cagaaccaga ggatcagatg gtacctgtg tctatgggat	3660
ccaatgagaa tatccatagc atccacttct ctggccatgt cttttctgtg aggaagaaag	3720
aggaatacaa aatggctgtg tacaatctgt atcctgggggt ctttgagaca gtggaaatgc	3780
tgccaagcaa agtgggaatc tggagaattg agtgccctgat tggggaacac ctgcaggctg	3840
ggatgagcac caccttctctg gtgtactcta agaaatgtca gacccactg gggatggcct	3900
ctggacatat cagggacttc cagatcacag cttctggaca gtatggacag tgggctccaa	3960
agctggctag actgcactat tctggctcca tcaatgcctg gtctaccaa gagccattct	4020
cctggatcaa ggtggacctg ctggcccca tgatcatcca tggaatcaa acccagggag	4080
ctaggcagaa gttcagctct ctgtacatct ccagtttat catcatgtat agcctggatg	4140
ggaagaaatg gcagacctac agaggcaatt ccaactgggac cctgatggtc ttctttggaa	4200
atgtggattc ctctggcatc aagcacaaca tcttcaatcc acccatcatt gccaggatga	4260
tcaggctgca tcctaccac tatagcatca ggtctaccct gagaatggag ctgatgggat	4320
gtgacctgaa cagctgttct atgccactgg gcatggagtc caaggctatc tctgatgccc	4380
agatcacagc ttcttcctac ttcaccaata tgtttgctac ctggtccca agcaaggcta	4440
gactgcacct gcagggaaga tccaatgctt ggagaccca ggtgaacaat cctaaggagt	4500
ggctgcagggt ggacttcag aaaacatga aggtcacagg ggtgaccacc caggagtgga	4560
aatctctgct gacctcatg tatgtcaagg agttcctgat cagctcttcc caggatggcc	4620
accagtggac cctgttcttt cagaatggca aggtcaaagt gttccagggg aatcaggact	4680
cttttaccac agtgggtgaac tocctggatc ctccactgct gaccaggtag ctgagaatcc	4740
atcctcagag ctgggtgcac cagattgctc tgagaatgga ggtcctggga tgtgaagctc	4800
aggacctgta ttgagcggcc gcaataaaat atctttattt tcattacatc tgtgtgttgg	4860

ES 2 763 551 T3

ttttttgtgt gcaattgagg aaccocctagt gatggagttg gccactccct ctctgcgcgc	4920
tcgctcgtc actgaggccg ggcgaccaa ggtcgccga cgcccggtt ttgcccgggc	4980
ggcctcagt agcgagcgag cgcgcagctg cctgcagg	5018

<210> 131
 <211> 137
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> secuencia promotora recombinante

<400> 131

gttaatTTTT aaaaagcagt caaaagtcca agtggccctt gcgagcattt actctctctg	60
tttgctctgg ttaataatct caggagcaca aacagagggt aataattttc cagatctctc	120
tgagcaatag tataaaa	137

REIVINDICACIONES

1. Una molécula de ácido nucleico recombinante, que comprende:

un promotor específico de hígado que comprende un primer elemento de respuesta de no más de 160 nucleótidos de longitud que comprende:

un sitio de unión al factor de transcripción (TF) HNF1a que comprende o que consiste en los nucleótidos 1-12 de la SEQ ID NO: 4, un sitio de unión al TF HNF1-1 que comprende o que consiste en los nucleótidos 16-23 de la SEQ ID NO: 4;

un sitio de unión al TF HNF4 que comprende o que consiste en los nucleótidos 26-36 de la SEQ ID NO: 4; un sitio de unión al TF HNF3a que comprende o que consiste en los nucleótidos 39-45 de la SEQ ID NO: 4;

un sitio de unión al TF HNF1-2 que comprende o que consiste en los nucleótidos 48-62 de la SEQ ID NO: 4;

un sitio de unión al TF HNF3-2 que comprende o que consiste en los nucleótidos 65-71 de la SEQ ID NO: 4;

un sitio de unión al TF HP1 que comprende o que consiste en los nucleótidos 75-87 de la SEQ ID NO: 4;

una caja TATA que comprende o que consiste en los nucleótidos 108-114 de la SEQ ID NO: 4; y

un Sitio de Inicio de la Transcripción que comprende o que consiste en los nucleótidos 116-146 de la SEQ ID NO: 4.

2. La molécula de ácido nucleico recombinante de la reivindicación 1, en donde:

(a) el primer elemento de respuesta comprende, de 5' a 3', el sitio de unión al TF HNF1a, el sitio de unión al TF HNF1-1, el sitio de unión al TF HNF4, el sitio de unión al TF HNF3a, el sitio de unión al TF HNF1-2, el sitio de unión al TF HNF3-2, el sitio de unión al TF HP1, la caja TATA y el Sitio de Inicio de la Transcripción (TSS); y/o

(b) el primer elemento de respuesta no tiene más de 150 nucleótidos de longitud, en donde, opcionalmente, el primer elemento de respuesta tiene 146 nucleótidos de longitud; y/o

(c) el primer elemento de respuesta comprende o consiste en una secuencia de nucleótidos al menos un 90 % idéntica a la SEQ ID NO: 4, en donde, opcionalmente, el primer elemento de respuesta comprende o consiste en la secuencia de nucleótidos establecida como la SEQ ID NO: 4; y/o

(d) el promotor consiste en el primer elemento de respuesta.

3. La molécula de ácido nucleico recombinante de la reivindicación 1 o 2, que comprende adicionalmente un segundo elemento de respuesta cadena arriba del primer elemento de respuesta, comprendiendo el segundo elemento de respuesta uno de:

un elemento de respuesta HSh que comprende o que consiste en una secuencia de nucleótidos al menos un 90 % idéntica a la SEQ ID NO: 111;

un elemento de respuesta 5'HS que comprende o que consiste en una secuencia de nucleótidos al menos un 90 % idéntica a los nucleótidos 6-32 de la SEQ ID NO: 111; o

un elemento de respuesta 3'HS que comprende o que consiste en una secuencia de nucleótidos al menos un 90 % idéntica a los nucleótidos 44-68 de la SEQ ID NO: 111,

en donde, opcionalmente

el elemento de respuesta HSh comprende o consiste en la secuencia de nucleótidos establecida como la SEQ ID NO: 111;

el elemento de respuesta 5'HS comprende o consiste en los nucleótidos 6-32 de la SEQ ID NO: 111; o

el elemento de respuesta 3'HS comprende o consiste en los nucleótidos 44-68 de la SEQ ID NO: 111.

4. La molécula de ácido nucleico recombinante de la reivindicación 3, en donde el promotor recombinante comprende o consiste en una secuencia de nucleótidos al menos un 90 % idéntica a una de la SEQ ID NO: 102 (HSh-HCB), SEQ ID NO: 104 (5'HSh-HCB) o SEQ ID NO: 103 (3'HSh-HCB), en donde, opcionalmente, el promotor recombinante comprende o consiste en la secuencia de nucleótidos establecida como la SEQ ID NO: 102 (HSh-HCB), SEQ ID NO: 104 (5'HSh-HCB) o SEQ ID NO: 103 (3'HSh-HCB).

5. Un vector que comprende la molécula de ácido nucleico recombinante de una cualquiera de las reivindicaciones anteriores.

6. El vector de la reivindicación 5, en donde el vector es un vector vírico, en donde, opcionalmente, el vector vírico es un vector AAV, un vector gamma retrovírico, un vector lentivírico o un vector adenovírico.

7. El vector de la reivindicación 5 o 6, que comprende una molécula de ácido nucleico que codifica un factor de coagulación que está operativamente unido al promotor.

8. El vector de la reivindicación 7, en donde:

(a) el factor de coagulación es fVIII, o una variante de fVIII ET3 o humana con el dominio B eliminado (HSQ), en donde, opcionalmente, el fVIII comprende la secuencia del ácido nucleico establecida como la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 125, o SEQ ID NO: 126, o una secuencia del ácido nucleico al menos un

90 % idéntica a la SEQ ID NO: 2, SEQ ID NO: 11, SEQ ID NO: 12, la SEQ ID NO: 125 o la SEQ ID NO: 126; o
 (b) el factor de coagulación es fIX, en donde, opcionalmente, el fIX comprende la secuencia del ácido nucleico
 establecida como la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 124, o SEQ ID NO: 127, o una
 secuencia del ácido nucleico al menos un 90 % idéntica a la SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ
 ID NO: 124 o SEQ ID NO: 127.

9. Una composición que comprende el vector de una cualquiera de las reivindicaciones 5-8 en un transportador
 farmacéuticamente aceptable.

10. El vector de la reivindicación 7 u 8, o la composición de la reivindicación 9, para su uso en un método para inducir
 la coagulación de la sangre en un sujeto que lo necesite, comprendiendo dicho método administrar al sujeto una
 cantidad terapéuticamente eficaz del vector o composición.

11. El vector de la reivindicación 7 u 8, o la composición de la reivindicación 9, para su uso en un método de tratamiento
 de un sujeto con un trastorno de la coagulación, comprendiendo dicho método seleccionar un sujeto con el trastorno
 de la coagulación y administrar al sujeto una cantidad terapéuticamente eficaz del vector o la composición.

12. El vector o composición para su uso en el método de la reivindicación 11, en donde el trastorno de la coagulación
 es:

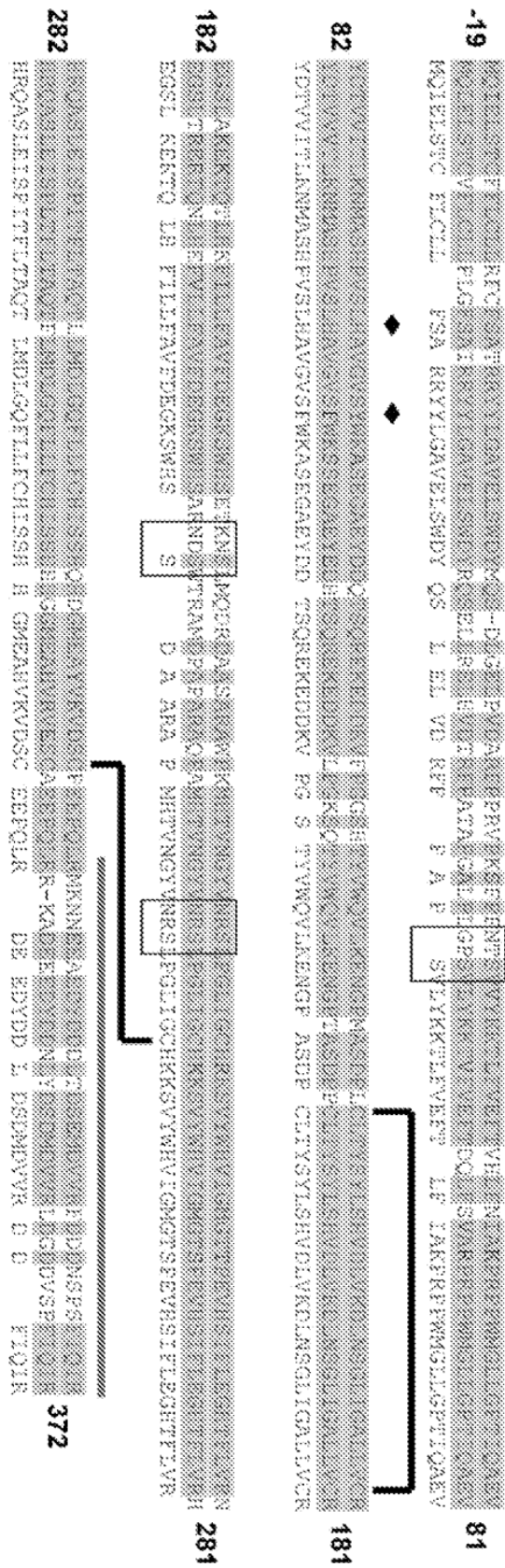
hemofilia A y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una
 proteína con actividad de fVIII; o
 hemofilia B y al sujeto se le administra un vector que comprende una molécula de ácido nucleico que codifica una
 proteína con actividad de fIX.

13. El vector o composición para su uso en el método de una cualquiera de las reivindicaciones 10-12, en donde:

(a) el vector se administra por vía intravenosa; o
 (b) el vector es un vector AAV.

14. El vector o composición para su uso en el método de una cualquiera de las reivindicaciones 10-13, en donde el
 vector se administra en combinación con un agente inmunosupresor, en donde, opcionalmente, el agente
 inmunosupresor es ciclosporina, tacrolimus, sirolimus, ciclofosfamida, metotrexato, azatioprina, mercaptopurina,
 fluorouracilo, ácido micofenólico, dactinomicina, fingolimod, anticuerpo o proteína de unión a receptores de linfocitos
 T, muromonab-CD3, anticuerpo o proteína de unión al receptor de IL-2, basiliximab, daclizumab, IFN-beta
 recombinante, anticuerpo o proteína de unión a TNF-alfa, infliximab, etanercept o adalimumab.

FIG. 1A



Dominio A3

1649
DIS W O D D IDDD S E K EDIDIZ EDENO
1748

1749
KCEIN HIGLIGPITAEVEDNIMVTENQASBPYSYSLSISY DDQ QCAEPK NTV
1848

1849
SCLIGPLIC NTIN AGRQVAVQESALFTIEDIKSWYFTEHRENSORAPC IQMDDPT KENYREHAINQYINDILPOLVMAQ QIRINWYLLSNG
1948

1949
SSENIHSIHSGHVSVRKKEEYKNNALNANZQVEETVEMLERK GIMRIECLIGENT AGNET FLAVS
2019

FIG. 1B

Tabla de optimización de codones humanos específicos de hígado

AA	Codón	Total		Hígado	Hígado:Total	AA	Codón	Total		Hígado	Hígado:Total	AA	Codón	Total		Hígado	Hígado:Total
		Frecuencia	Frecuencia					Frecuencia	Frecuencia					Frecuencia	Frecuencia		
Ala	GCC	0,4	0,44	1,10		Gly	GGT	0,16	0,12	0,75		Pro	CCG	0,11	0,1	0,91	
Ala	GCA	0,23	0,19	0,83		Gly	GGG	0,25	0,28	1,12		Pro	CCT	0,28	0,28	1,00	
Ala	GCT	0,26	0,27	1,04		Gly	GGA	0,25	0,23	0,92		Ser	TCC	0,22	0,28	1,27	
Ala	GCG	0,11	0,1	0,91		His	CAC	0,59	0,59	1,00		Ser	AGT	0,15	0,12	0,80	
Arg	CGC	0,19	0,24	1,26		His	CAI	0,41	0,41	1,00		Ser	TCA	0,15	0,13	0,87	
Arg	AGA	0,2	0,18	0,90		Ile	ATC	0,48	0,61	1,27*		Ser	AGC	0,24	0,25	1,04	
Arg	GGA	0,11	0,1	0,91		Ile	ATT	0,36	0,28	0,78		Ser	TCT	0,18	0,17	0,94	
Arg	CGT	0,08	0,07	0,88		Ile	ATA	0,16	0,11	0,69		Ser	TCG	0,06	0,06	1,00	
Arg	AGG	0,2	0,2	1,00		Leu	CTG	0,41	0,47	1,15		Thr	ACC	0,36	0,47	1,31*	
Arg	CGG	0,21	0,21	1,00		Leu	CTC	0,2	0,23	0,77		Thr	ACA	0,28	0,21	0,75	
Asn	AAC	0,54	0,59	1,09		Leu	CTT	0,13	0,1	0,77		Thr	ACT	0,24	0,19	0,79	
Asn	AAT	0,46	0,41	0,89		Leu	TTA	0,07	0,04	0,57		Thr	ACG	0,12	0,12	1,00	
Asp	GAC	0,54	0,61	1,13*		Leu	TTC	0,13	0,11	0,85		Trp	TGG	1	1	1,00	
Asp	GAT	0,46	0,39	0,85		Leu	CTA	0,07	0,06	0,80		Tyr	TAC	0,57	0,63	1,11	
Cys	TGC	0,55	0,6	1,09		Lys	AAG	0,58	0,65	1,12*		Tyr	TAT	0,43	0,37	0,86	
Cys	TGT	0,45	0,4	0,89		Lys	AAA	0,42	0,35	0,83		Val	GTG	0,47	0,52	1,11	
Gln	CAA	0,25	0,25	1,00		Met	ATG	1	1	1,00		Val	GTT	0,18	0,13	0,72	
Gln	CAG	0,75	0,75	1,00		Phe	TTC	0,55	0,65	1,18*		Val	GTA	0,11	0,08	0,73	
Glu	GAG	0,58	0,65	1,12*		Phe	TTT	0,45	0,35	0,78*		Val	GTC	0,24	0,26	1,08	
Glu	GAA	0,42	0,35	0,83		Pro	CCC	0,33	0,38	1,15							
Gly	GGC	0,34	0,37	1,09		Pro	CCA	0,27	0,24	0,89							

FIG. 2A

Tabla patrón de optimización de codones humanos

Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento
UUU	0,464134	714298	UCU	0,187586	618711	UAU	0,443338	495699	UGU	0,456157	430311
UUC	0,535866	824692	UCC	0,21796	718892	UAC	0,556662	622407	UGC	0,543843	513028
UUA	0,076568	311881	UCA	0,150517	496448	UAA	0	0	UGA	1	10000
UUG	0,129058	525688	UCG	0,054398	179419	UAG	0	0	UGG	1	535595
CUU	0,131716	536515	CCU	0,286731	713233	CAU	0,418515	441711	CGU	0,080108	184609
CUC	0,195577	796638	CCC	0,32347	804620	CAC	0,581485	613713	CGC	0,183777	423516
CUA	0,07138	290751	CCA	0,276603	688038	CAA	0,265017	501911	CGA	0,108812	250760
CUG	0,395702	1611801	CCG	0,113196	281570	CAG	0,734983	1391973	CGG	0,201554	464485
AUU	0,361072	650473	ACU	0,246769	533609	AUA	0,470367	689701	AGU	0,149602	493429
AUC	0,469866	846466	ACC	0,355232	768147	AAC	0,529633	776603	AGC	0,239938	791383
AUA	0,169062	304565	ACA	0,284188	614523	AAA	0,434049	993621	AGA	0,214658	494682
AUG	1	896005	ACG	0,113812	246105	AAG	0,565951	1295568	AGG	0,211091	486463
GUU	0,18177	448607	GCU	0,265922	750096	GAU	0,464542	885429	GGU	0,163087	437126
GUC	0,238306	588138	GCC	0,399781	1127679	GAC	0,535458	1020595	GGC	0,337109	903565
GUA	0,116577	287712	GCA	0,228121	643471	GAA	0,422453	1177632	GGA	0,249922	669873
GUG	0,463346	1143534	GCG	0,106176	299495	GAG	0,577547	1609975	GGG	0,249882	669768

FIG. 2B

Tablas de optimización de codones de Monocitos/Mieloides

Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento	Codón	Frecuencia	Recuento
UUU	0,39	95	UCU	0,15	72	UAU	0,4	70	UGU	0,33	52			
UUC	0,61	150	UCC	0,22	103	UAC	0,6	107	UGC	0,67	106			
UUA	0,05	27	UCA	0,17	77	UAA	0	0	UGA	1	100			
UUG	0,12	67	UCG	0,04	17	UAG	0	0	UGG	1	92			
CUU	0,1	56	CCU	0,22	75	CAU	0,39	54	CGU	0,1	27			
CUC	0,24	139	CCC	0,38	127	CAC	0,61	83	CGC	0,14	37			
CUA	0,07	40	CCA	0,31	104	CAA	0,28	71	CGA	0,09	25			
COG	0,43	251	CCG	0,09	30	CAG	0,72	180	CGG	0,18	48			
AUU	0,26	66	ACU	0,21	78	AUU	0,43	121	AGU	0,14	66			
AUC	0,6	155	ACC	0,42	153	AAC	0,57	161	AGC	0,28	131			
AUA	0,14	37	ACA	0,27	98	AAA	0,37	130	AGA	0,22	58			
AUG	1	137	ACG	0,1	36	AAG	0,63	225	AGG	0,27	71			
GUU	0,16	56	GCU	0,32	114	GAU	0,46	137	GGU	0,13	56			
GUC	0,29	103	GCC	0,35	124	GAC	0,54	161	GGC	0,37	153			
GUA	0,09	32	GCA	0,25	88	GAA	0,44	180	GGA	0,24	98			
GUG	0,47	169	GCG	0,07	25	GAG	0,56	230	GGG	0,26	109			

FIG. 2C

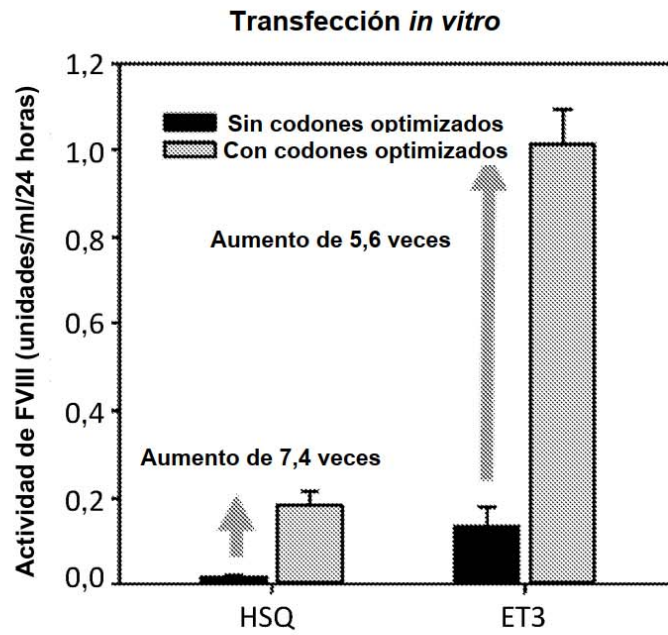


FIG. 3A

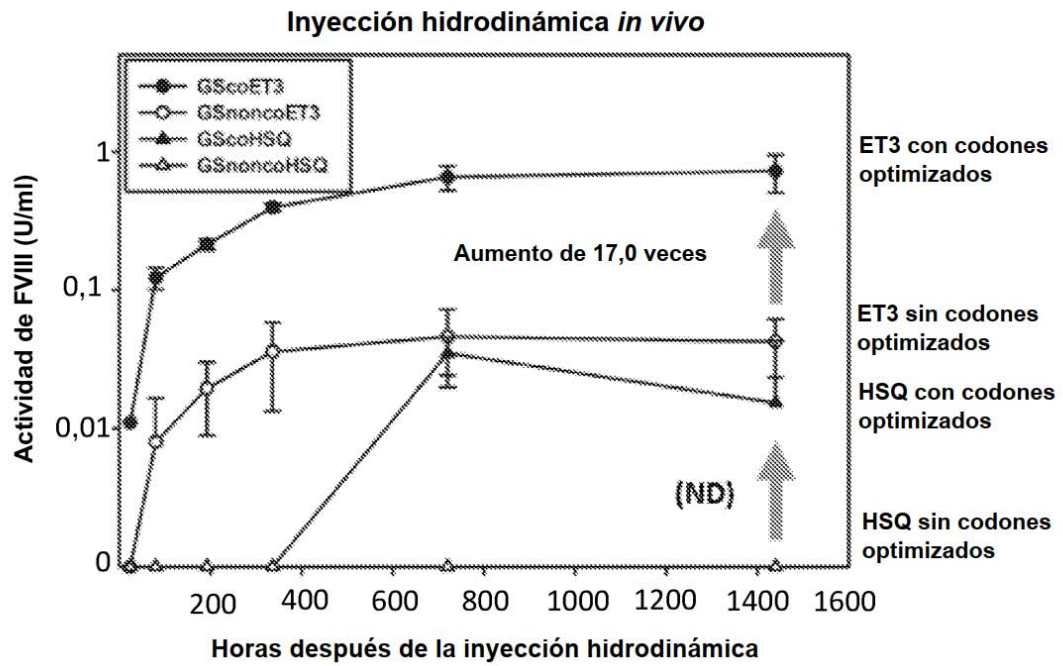
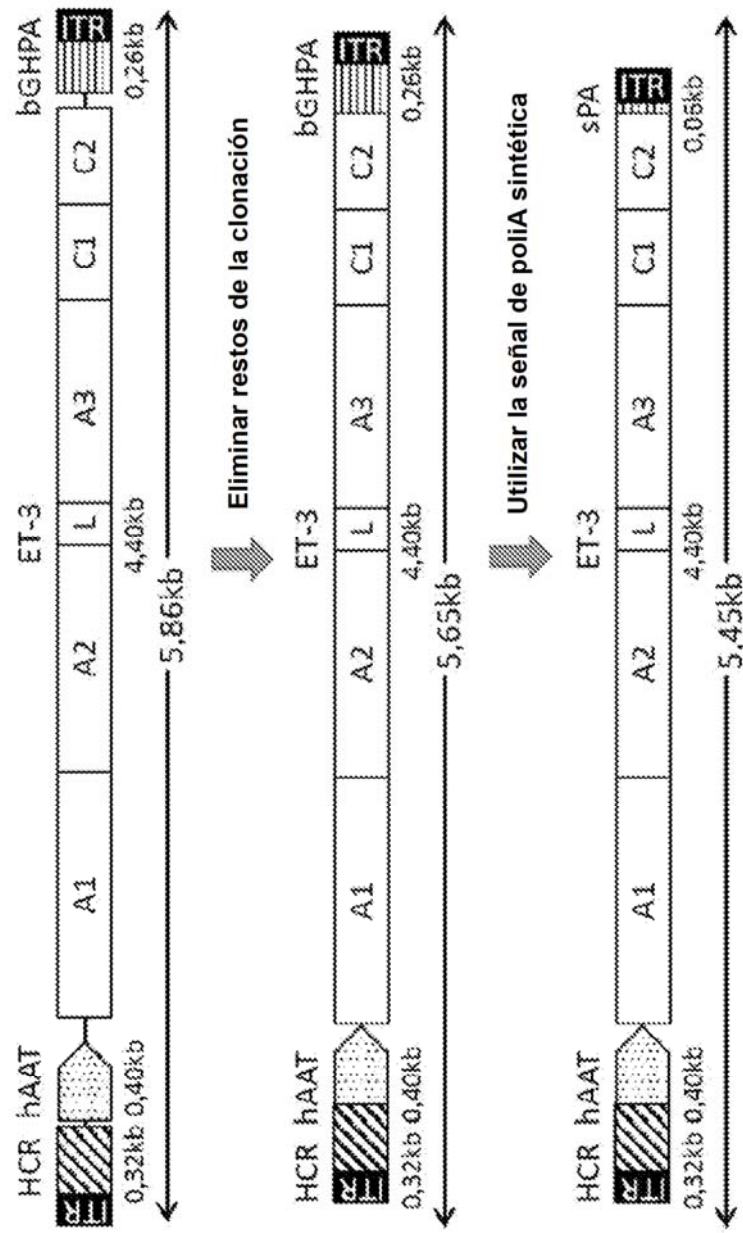


FIG. 3B

FIG. 4



Beneficios de la optimización de codones específicos de tejido
(expresión de células HepG2 por transfección transitoria)

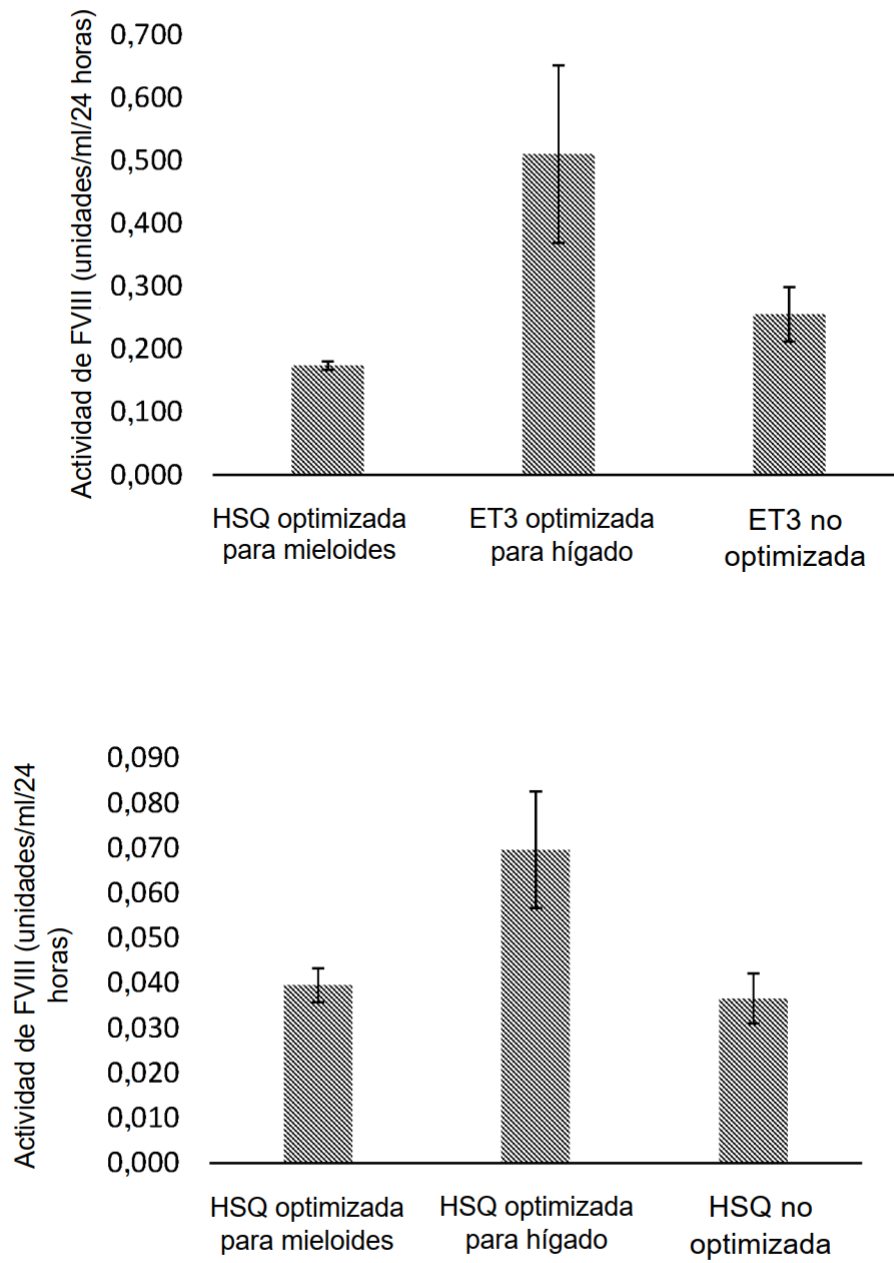


FIG. 5

Expresión de optimización de codones específicos de tejido en células de diferentes especies (expresión en células de riñón de hámster recién nacido por transfección transitoria)

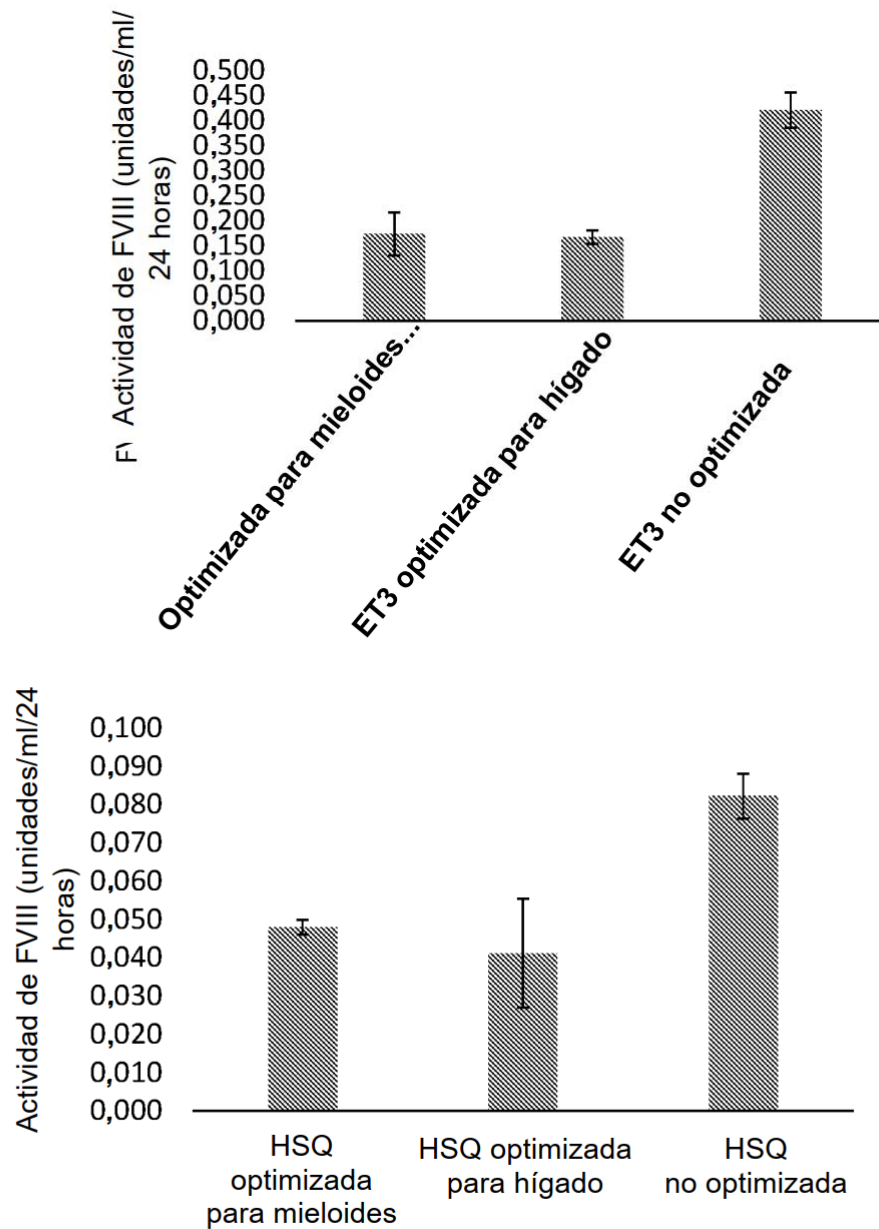


FIG. 6

Optimización en el hígado (factor IX)

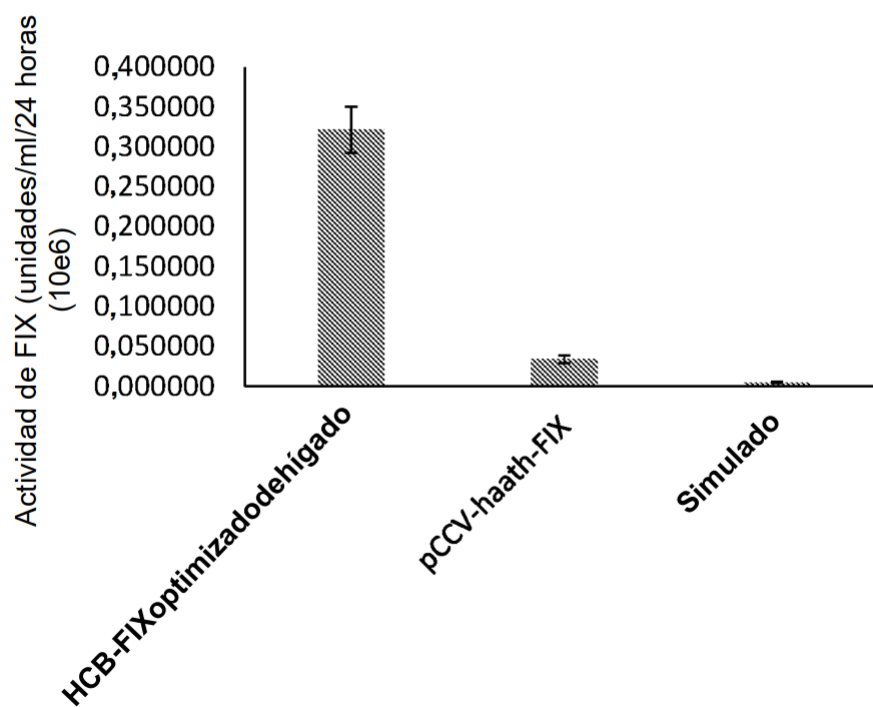
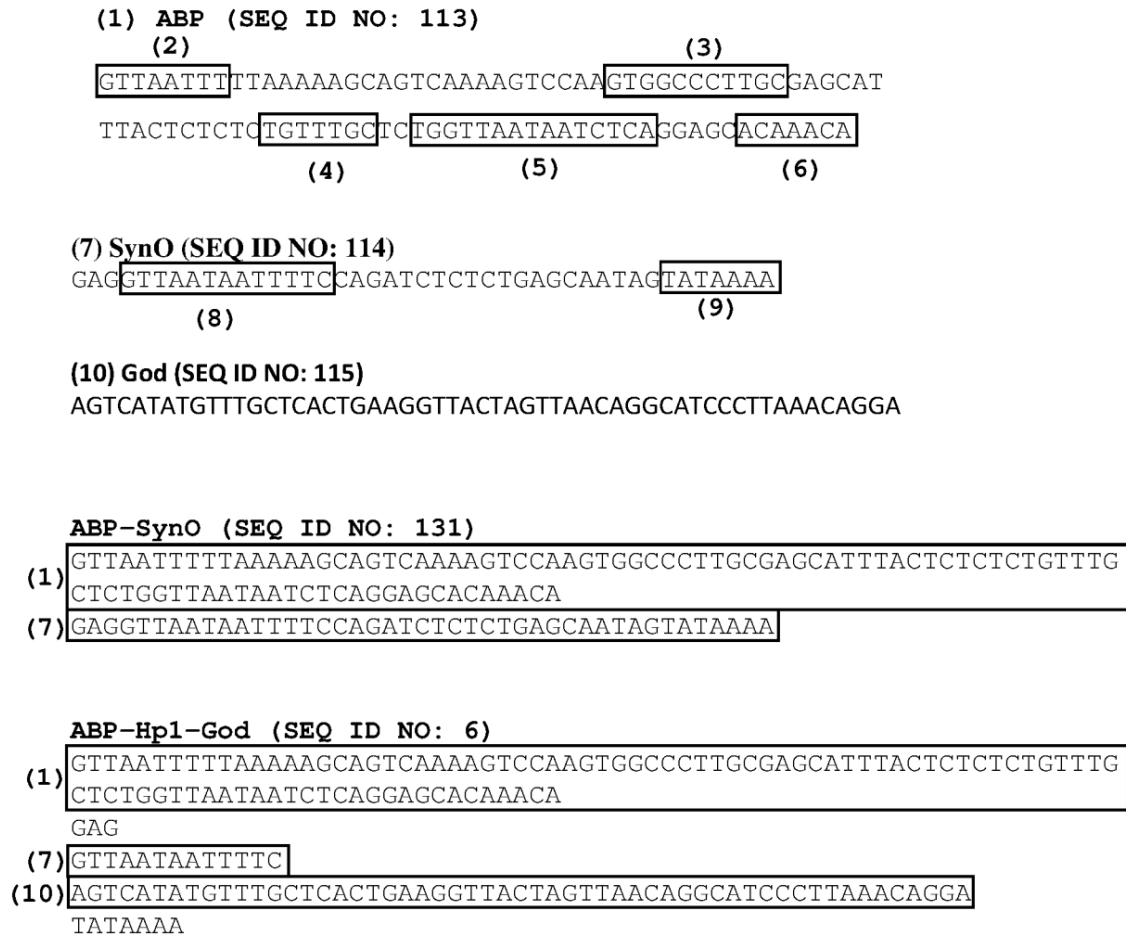


FIG. 7



- (1) Elemento ABP
- (2) Sitio de unión a TF HNF1-1
- (3) Sitio de unión a TF HNF4
- (4) Sitio de unión a TF HNF3a
- (5) Sitio de unión a TF HNF1-2
- (6) Sitio de unión a TF HNF3-2
- (7) Elemento Syno
- (8) Sitio de unión a TF HP1
- (9) Caja TATA
- (10) Elemento God

FIG. 8

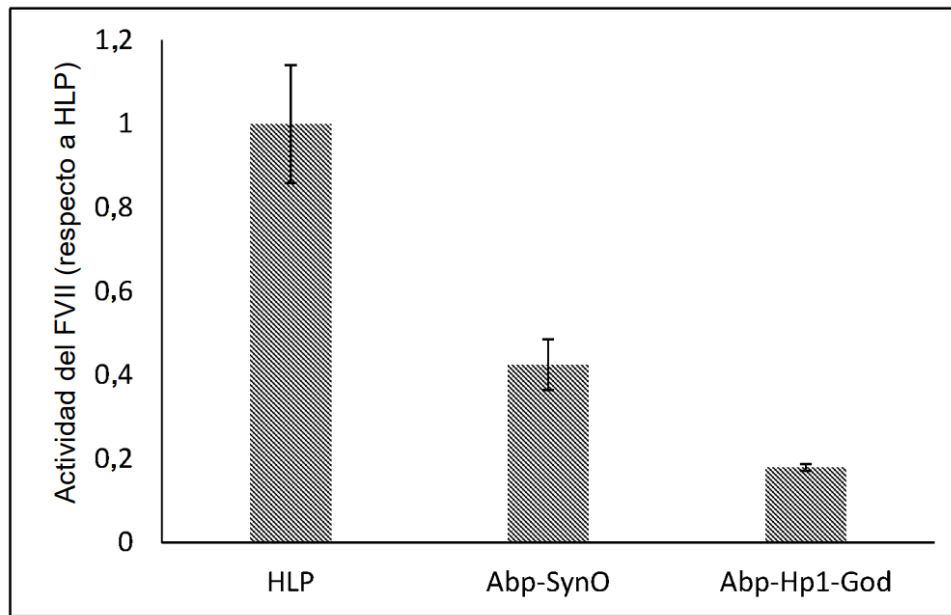


FIG. 9A

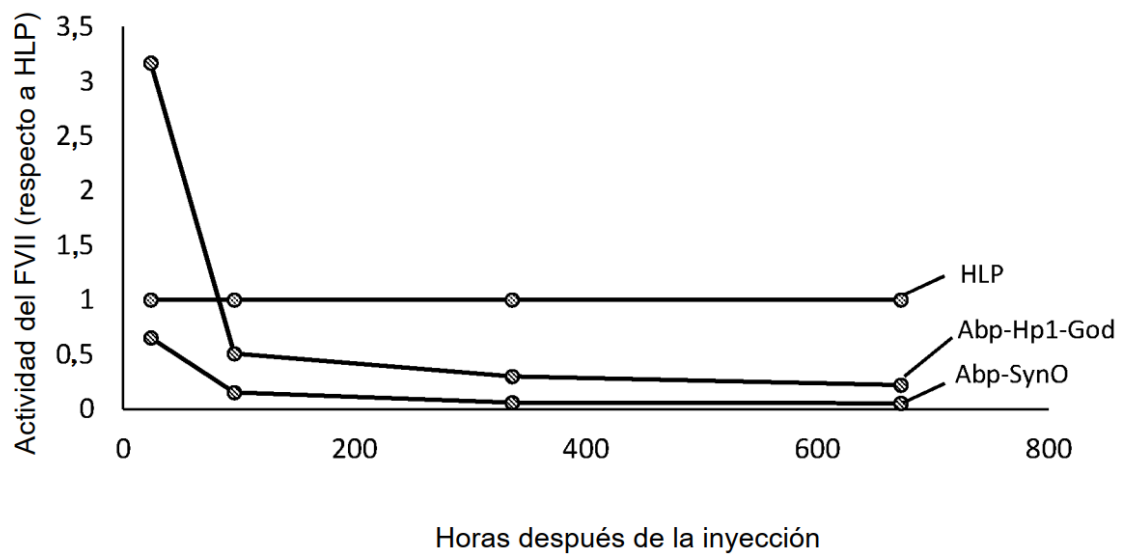


FIG. 9B

(11) ABP-exact (SEQ ID NO: 116)

(12) (13)
 GTTAATCATTAAC TTAAGGTCAGAGCATTTA
 CTCTCTC CAATGTTGACTC TCGTTAATGATTAA GGAGCAATTGTTGACTT
 (14) (15) (16)

(17) Short-ABP-exact (SEQ ID NO: 117)

(12) (13) (14) (15)
 GTTAATCATTAAC TTAAGGTCAGAGCATTTA
 GCAATTGTTGACTT
 (16)

(18) Sitio de inicio de la transcripción, TSS(nuc. 116-146 de la SEQ ID NO: 4)

(19) (20)
 GCCAGCAGCAGCCTGACCACATC TCATCCTC

(21) Sitio de unión a TF HNF1a

GTTAATCATTA

(22) Sitio de unión a TF Sp1

GTTAATCATTA

(23) Intrón SV40

CTCTAAGGTAAATATAAAATTTTAAGTGTATAATGTGTTAACTACTGATTCTAAT
 TGTTTGTGTATTTAGATTCCAACCTATGGAAGTGA

- (11) ABP-exact (sitios de unión al FT consenso)
- (12) Sitio de unión a TF consenso HNF1
- (13) Sitio de unión a TF consenso HNF4
- (14) Sitio de unión a TF consenso HNF3a
- (15) Sitio de unión a TF consenso HNF1
- (16) Sitio de unión a TF consenso HNF3
- (17) Short-ABP-exact
- (18) Sitio de Inicio de la Transcripción (TSS)
- (19) Espaciador rico en GC
- (20) Motivo de Inicio de la transcripción
- (21) Sitio de unión a TF HNF1a (sitio de unión a TF dirigido a hígado)
- (22) Sitio de unión a TF Sp1 (sitio de unión a TF dirigido a hígado)
- (23) Intrón SV40

FIG. 10

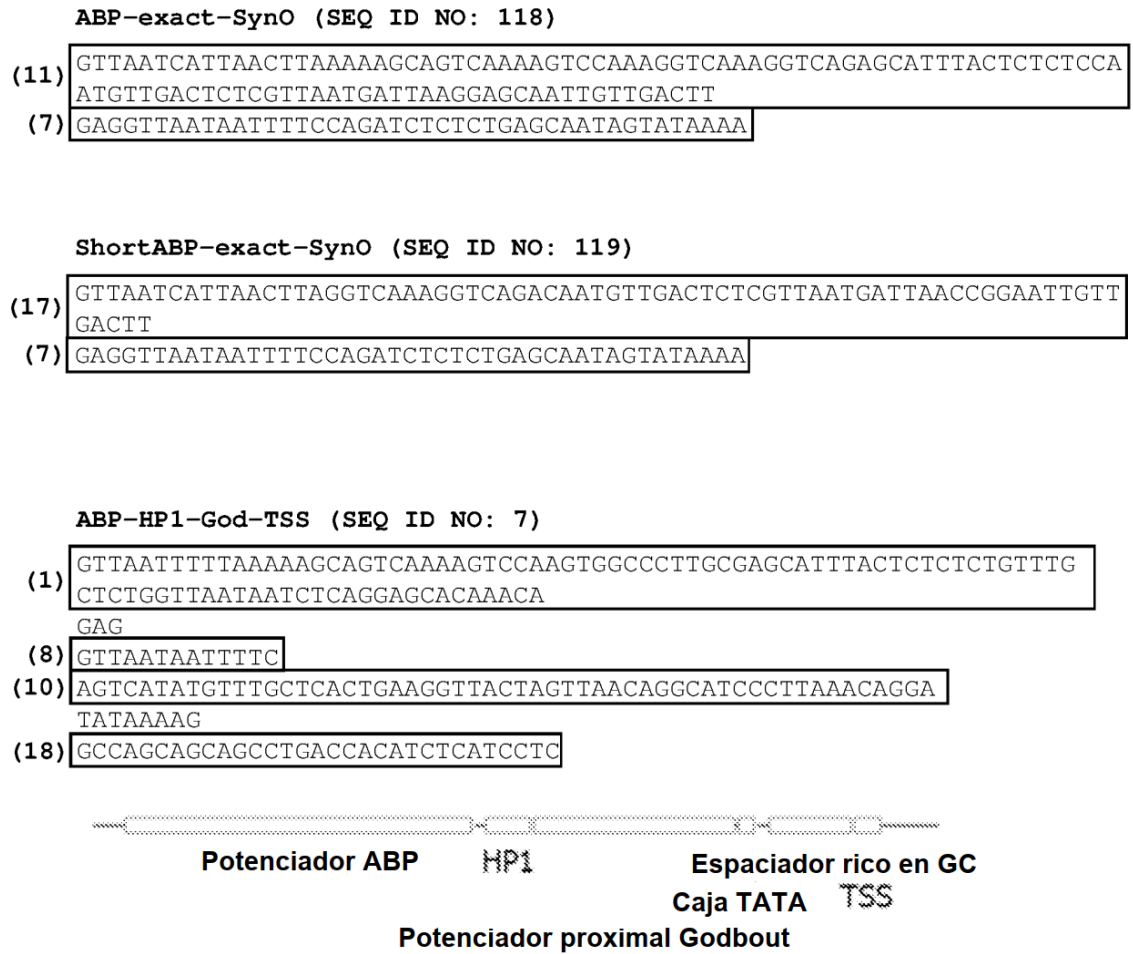


FIG. 11A



FIG. 11B

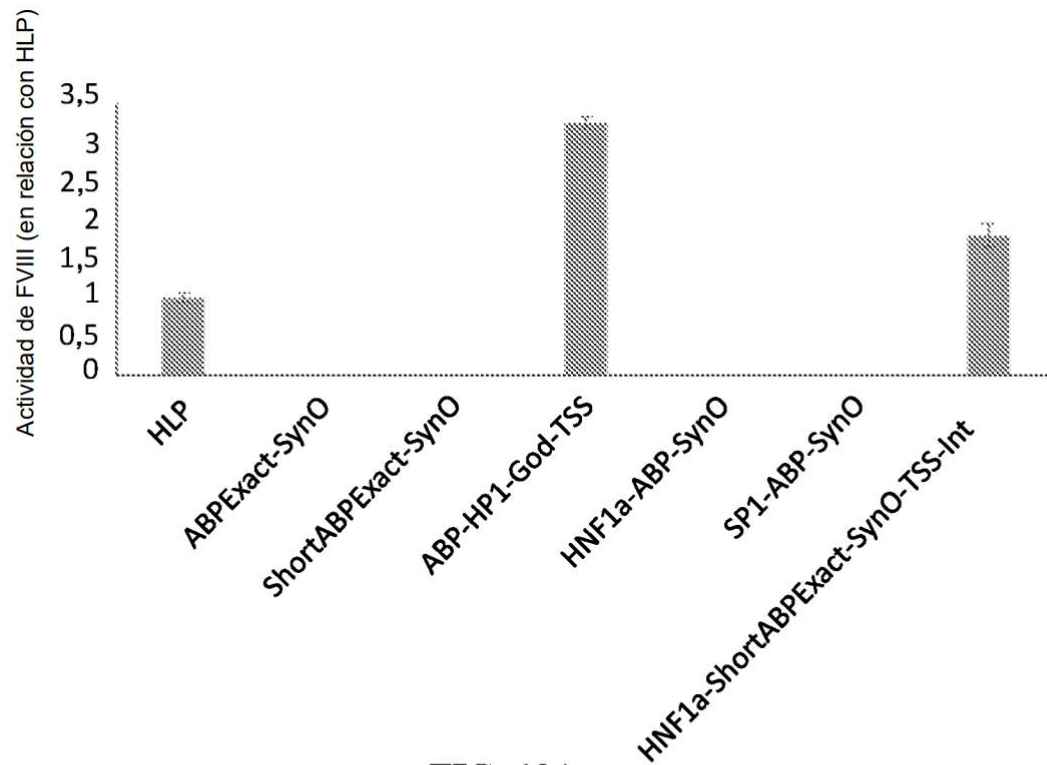


FIG. 12A

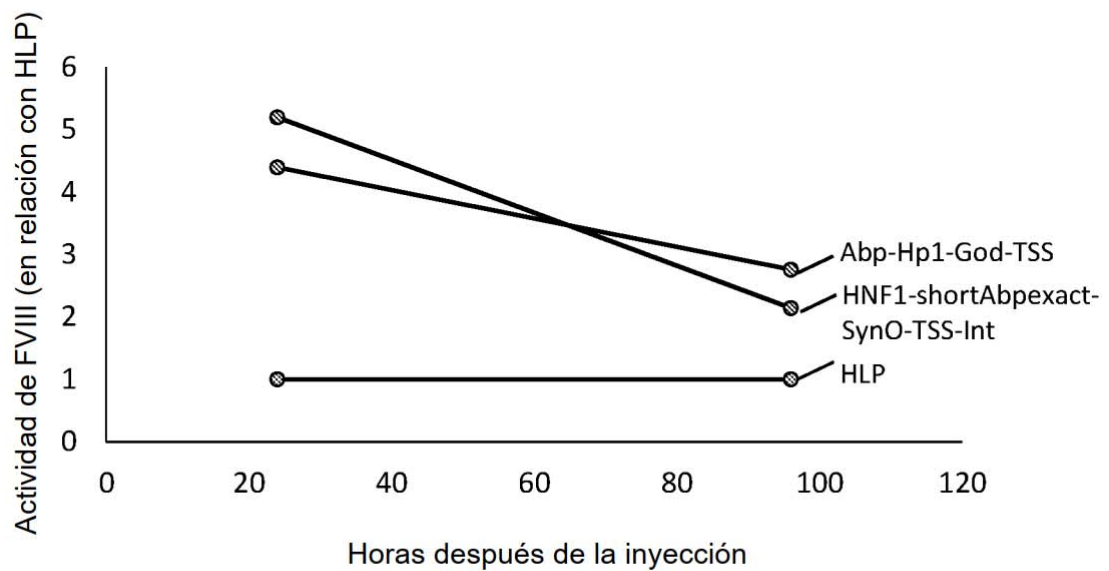


FIG. 12B

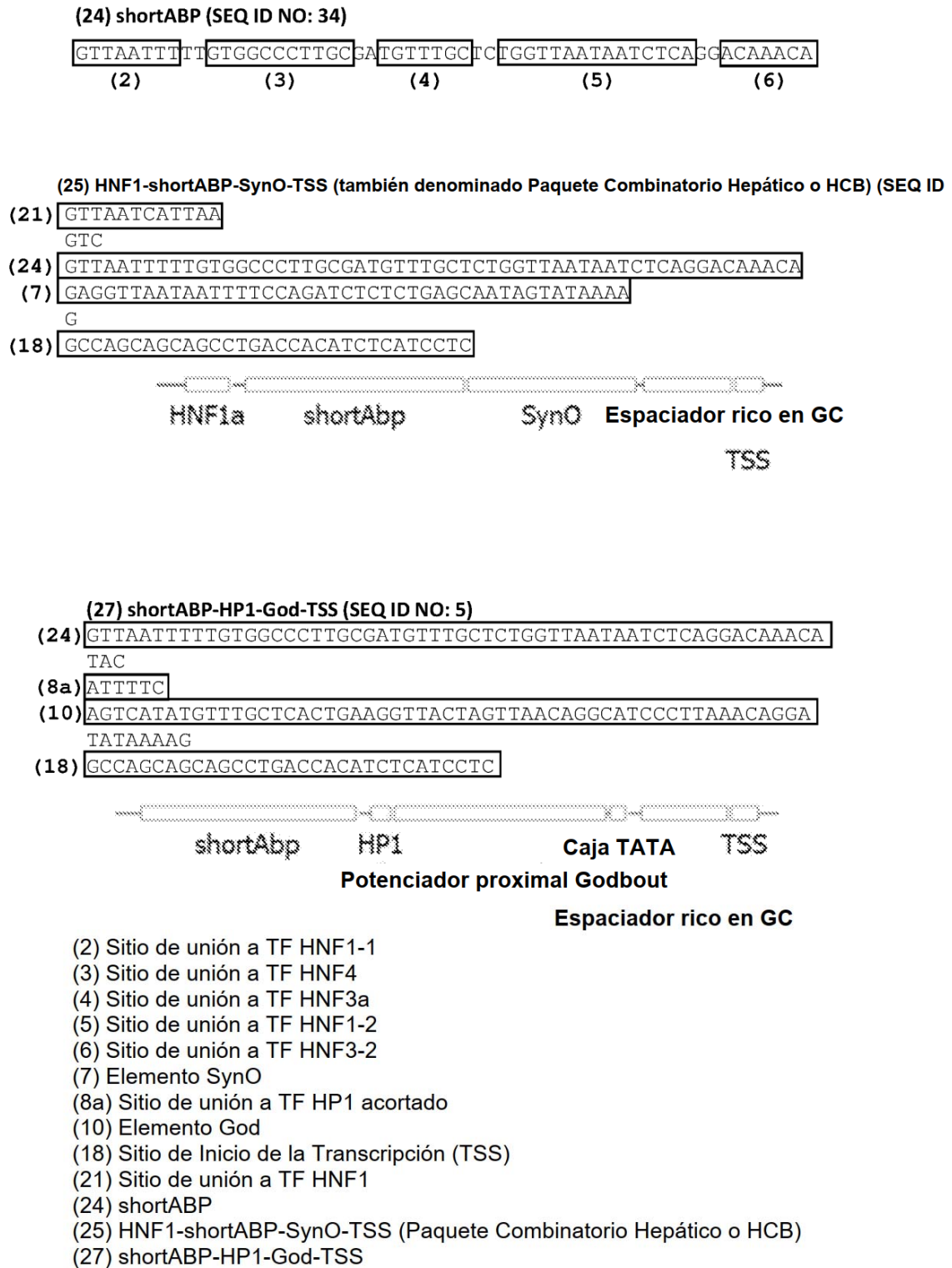


FIG. 13

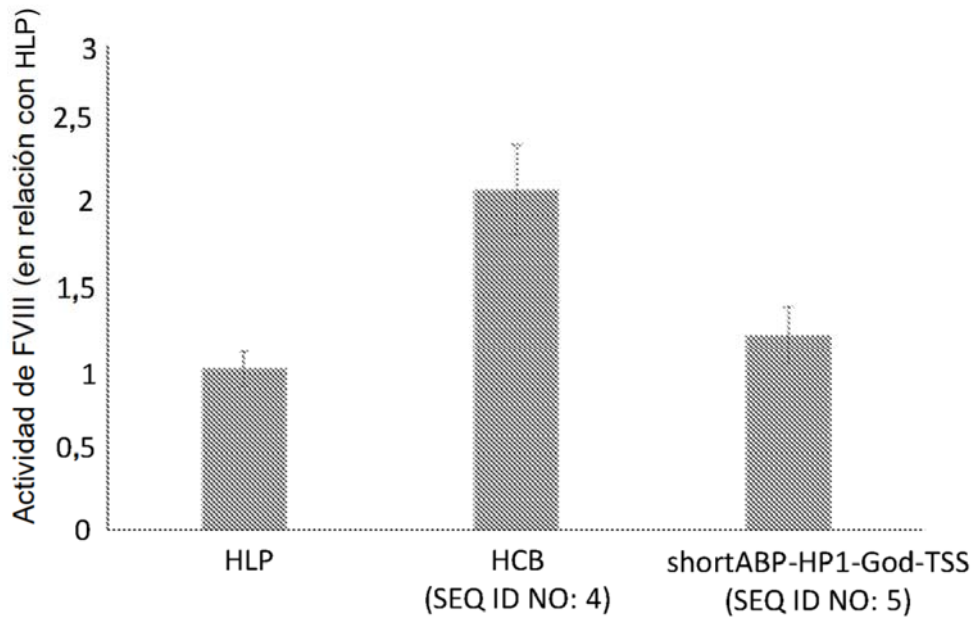


FIG. 14A

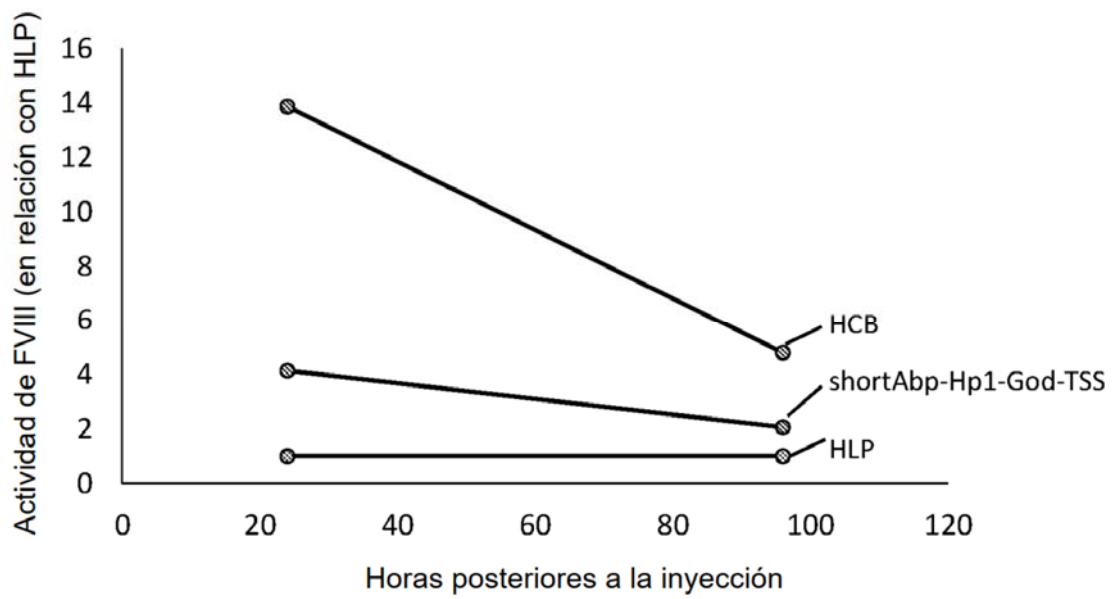


FIG. 14B

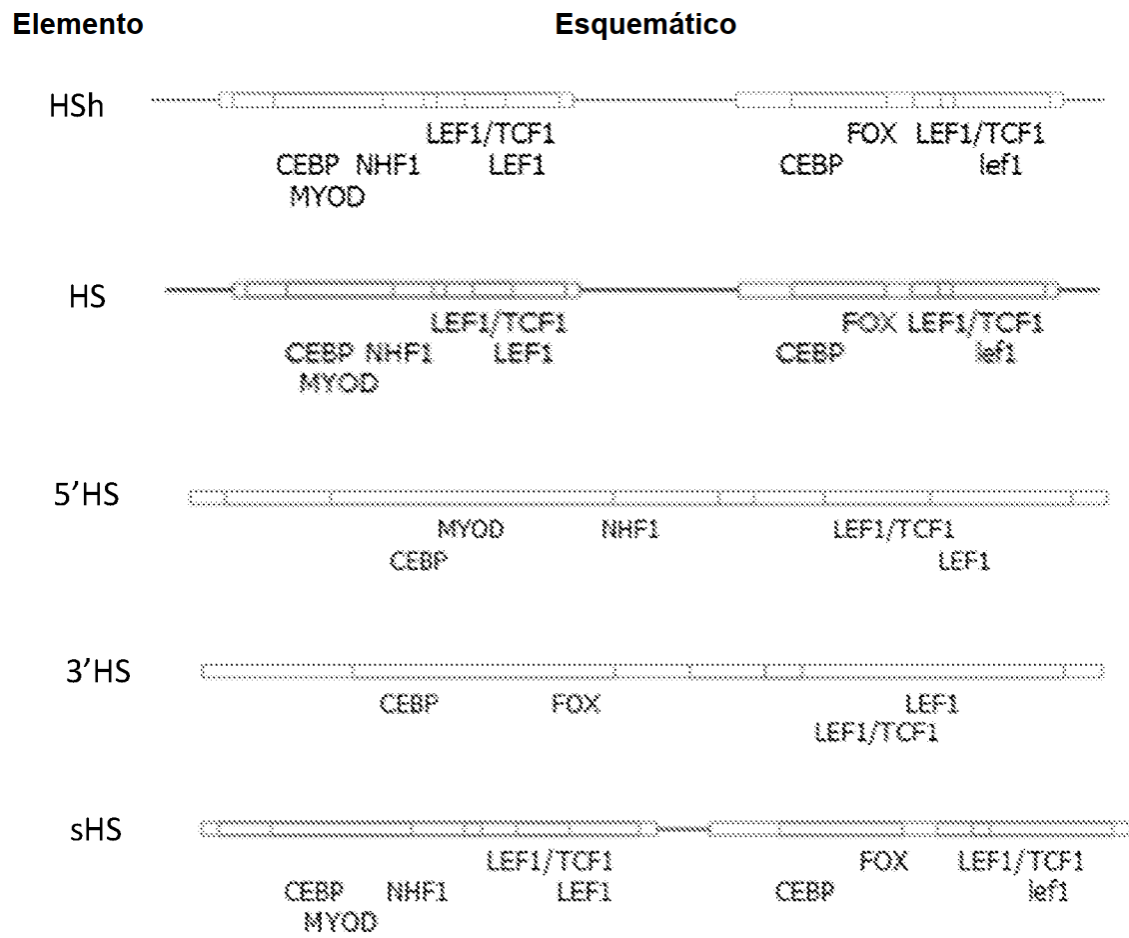


FIG. 15

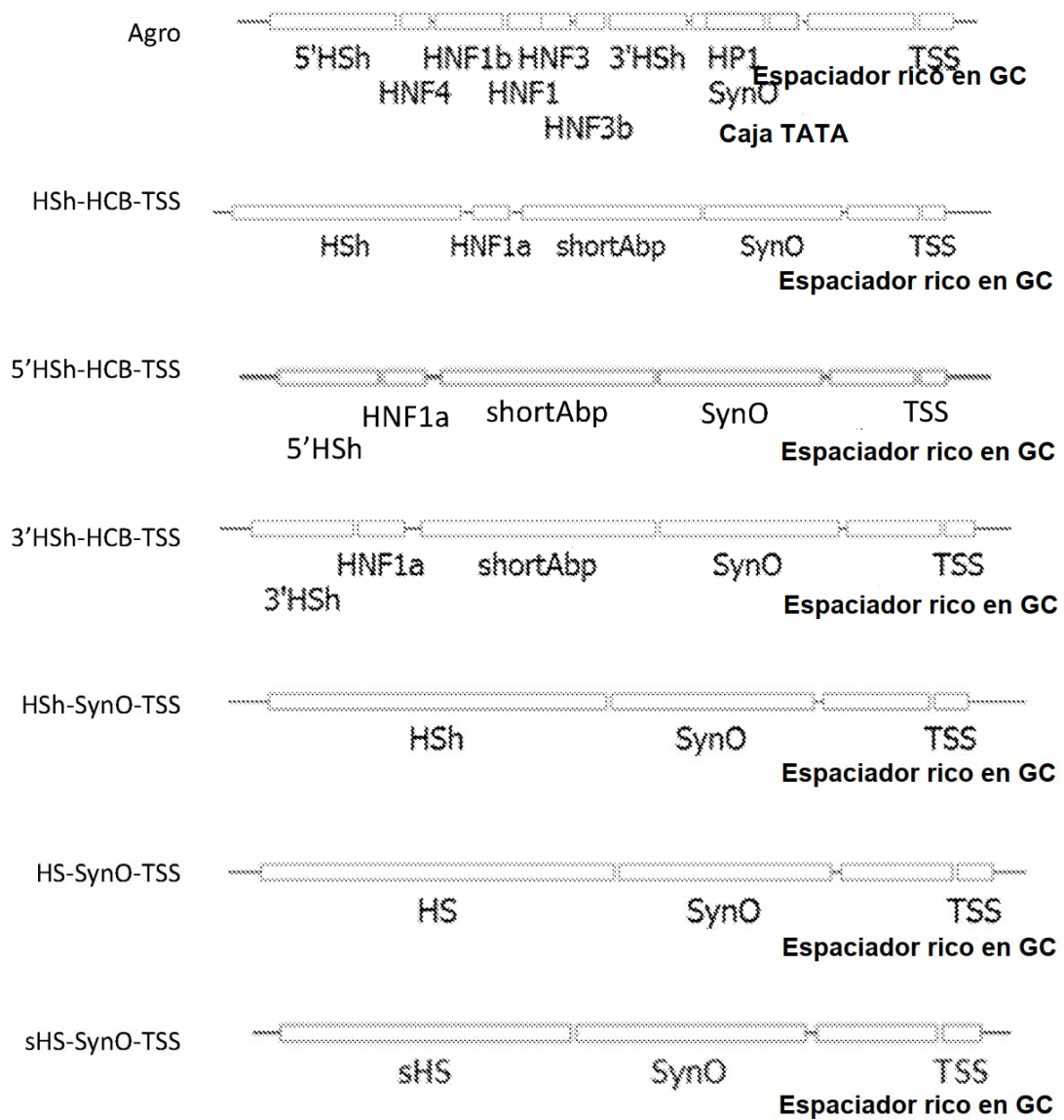


FIG. 16

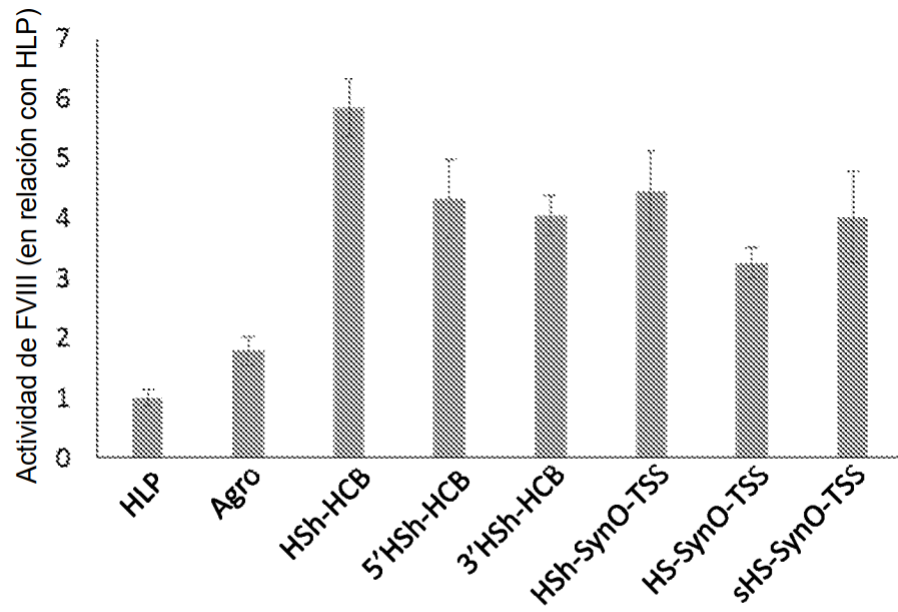


FIG. 17

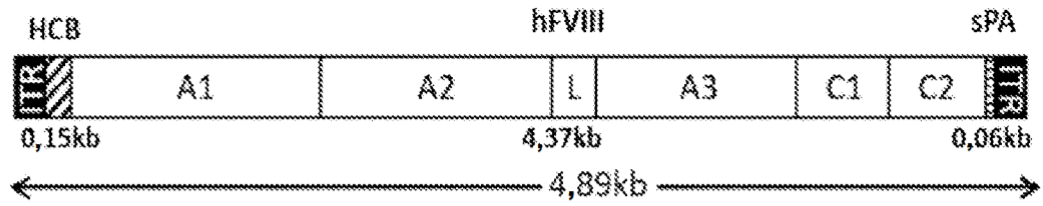


FIG. 18

Ratón Transducido con AAV-HCB-MVM-ET3-LCO-NCG

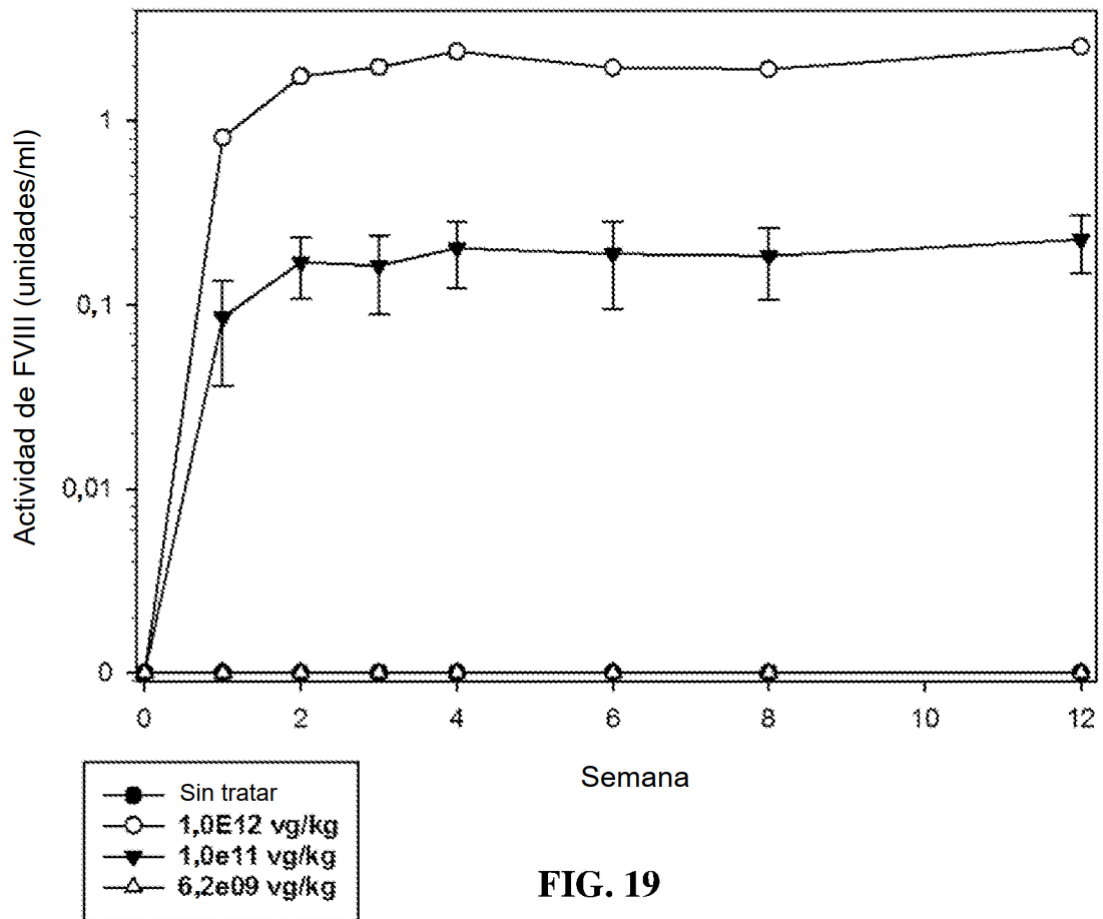


FIG. 19