

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 848 703**

51 Int. Cl.:

C07K 14/755 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA MODIFICADA
TRAS OPOSICIÓN

T5

86 Fecha de presentación y número de la solicitud internacional: **14.02.2014 PCT/US2014/016441**

87 Fecha y número de publicación internacional: **21.08.2014 WO14127215**

96 Fecha de presentación y número de la solicitud europea: **14.02.2014 E 14751254 (5)**

97 Fecha y número de publicación de la concesión europea modificada tras oposición: **24.01.2024 EP 2956477**

54 Título: **Gen del factor VIII optimizado**

30 Prioridad:

15.02.2013 US 201361765626 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente modificada:
25.06.2024

73 Titular/es:

**BIOVERATIV THERAPEUTICS INC. (100.0%)
225 Second Avenue
Waltham, MA 02451, US**

72 Inventor/es:

**TAN, SIYUAN y
PETERS, ROBERT, T.**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 848 703 T5

DESCRIPCIÓN

Gen del factor VIII optimizado

ANTECEDENTES DE LA INVENCION

5 La vía de coagulación de la sangre implica, en parte, la formación de un complejo enzimático del factor VIIIa (FVIIIa) y el factor IXa (FIXa) (complejo Xasa) sobre la superficie de las plaquetas. El FIXa es una serina proteasa con actividad catalítica relativamente débil sin su cofactor FVIIIa. El complejo Xasa escinde el factor X (FX) en el factor Xa (FXa), que a su vez interacciona con el factor Va (FVa) para escindir protrombina y generar trombina. La hemofilia A es un trastorno hemorrágico provocado por mutaciones y/o deleciones en el gen FVIII (FVIII) que da como resultado una deficiencia de actividad de FVIII (Peyvandi et al. 2006). En algunos casos, los pacientes tienen niveles de FVIII reducidos debido a la presencia de inhibidores de FVIII, tales como anticuerpos anti-FVIII.

10 La hemofilia A se caracteriza por hemorragia espontánea y excesivo sangrado. Con el tiempo, el sangrado repetido en músculos y articulaciones, que frecuentemente empieza en la segunda infancia, da como resultado artropatía hemofílica y daño articular irreversible. Este daño es progresivo y puede conducir a una movilidad fuertemente limitada de articulaciones, atrofia muscular y dolor crónico (Rodríguez-Merchan, E.C., Semin. Thromb. Hemost. 29:87-96 (2003).

15 La enfermedad se puede tratar por tratamiento sustitutivo que se dirige a la restauración de la actividad de FVIII al 1 a 5 % de los niveles normales para prevenir el sangrado espontáneo (véase, por ejemplo, Mannucci, P.M., et al., N. Engl. J. Med. 344:1773-9 (2001)). Están disponibles productos derivados del plasma y de FVIII recombinante para tratar episodios hemorrágicos a demanda o para prevenir que ocurran episodios hemorrágicos tratando profilácticamente. Basándose en la semivida de estos productos (10-12 h) (White G.C., et al., Thromb. Haemost. 77:660-7 (1997); Morfini, M., Haemophilia 9 (supl 1):94-99; discusión 100 (2003)), las pautas de tratamiento requieren una frecuente administración intravenosa, comúnmente dos a tres veces a la semana para la profilaxis y de una a tres veces al día para el tratamiento a demanda (Manco-Johnson, M.J., et al., N. Engl. J. Med. 357:535-544 (2007)). Dicha administración frecuente es poco práctica y cara.

20 Un impedimento importante en proporcionar una proteína FVIII recombinante de bajo coste a los pacientes es el alto coste de la producción comercial. La proteína FVIII se expresa débilmente en los sistemas de expresión heteróloga, dos a tres órdenes de magnitud menos que las proteínas de tamaño similar (Lynch et al., Hum. Gene. Ther.; 4:259-72 (1993)). La mala expresión de FVIII es debida en parte a la presencia de elementos que actúan en cis en la secuencia codificante de FVIII que inhiben la expresión de FVIII, tal como los elementos silenciadores transcripcionales (Hoeben et al., Blood 85:2447-2454 (1995)), secuencias de tipo unión a la matriz (MAR) (Fallux et al., Mol. Cell. Biol. 16:4264-4272 (1996)) y elementos inhibidores de la elongación transcripcional (Koeberl et al., Hum. Gene. Ther.; 6:469-479 (1995)).

25 Los avances en el entendimiento de los presentes inventores de la biología de la expresión de FVIII han conducido al desarrollo de variantes de FVIII más potentes. Por ejemplo, estudios bioquímicos demostraron que el dominio B de FVIII era dispensable para la actividad del cofactor FVIII. La delección del dominio B produjo un aumento de 17 veces en los niveles de ARNm a lo largo de FVIII natural de longitud completa y un aumento del 30 % en proteína secretada. (Toole et al., Proc Natl Acad Sci USA 83:5939-42 (1986)). Esto condujo al desarrollo de un concentrado de proteína FVIII de dominio B delecionado (BDD), que ahora se usa ampliamente en la clínica. Estudios recientes, sin embargo, indican que FVIIIh de longitud completa y BDD se plegaron erróneamente en la luz del RE, dando como resultado la activación de la respuesta a proteínas desplegadas (UPR) y la apoptosis de hepatocitos murinos. Los documentos de patente WO2011005968 y US2009042283 desvelan factor VIII optimizados por codones con delección del dominio B.

30 Así, existe una necesidad en la técnica de secuencias de FVIII que se expresen eficientemente en sistemas heterólogos.

SUMARIO DE LA INVENCION

35 La presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos al menos 90 % idéntica a SEQ ID NO: 1, en donde la secuencia de nucleótidos codifica un polipéptido con actividad de factor VIII. En otra realización, la invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o al menos 100 % idéntica a SEQ ID NO: 1. En otras realizaciones, la invención proporciona una molécula de ácido nucleico aislada que comprende SEQ ID NO: 1.

40 La presente invención también proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 %, o al menos 100 % idéntica a SEQ ID NO: 2, en donde la secuencia de nucleótidos codifica un polipéptido con actividad de factor VIII. En una realización, la invención proporciona una molécula de ácido nucleico aislada que comprende SEQ ID NO: 2.

55 En algunas realizaciones, la molécula de ácido nucleico aislada de la invención tiene un índice de adaptación de codones humanos que es elevado con respecto a SEQ ID NO: 3. En otras realizaciones, la molécula de ácido nucleico

aislada de la invención tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,75, al menos aproximadamente 0,76, al menos aproximadamente 0,77, al menos aproximadamente 0,78, al menos aproximadamente 0,79, o al menos aproximadamente 0,80. En otras realizaciones más, la molécula de ácido nucleico aislada de la invención tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,80, al menos aproximadamente 0,81, al menos aproximadamente 0,82, al menos aproximadamente 0,83, al menos aproximadamente 0,84, al menos aproximadamente 0,85, al menos aproximadamente 0,86, al menos aproximadamente 0,87, o al menos aproximadamente 0,88.

En ciertas realizaciones, la molécula de ácido nucleico aislada de la invención contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 3. En otras realizaciones, la molécula de ácido nucleico aislada de la invención contiene un porcentaje de nucleótidos G/C que es al menos aproximadamente 45 %, al menos aproximadamente 46 %, al menos aproximadamente 47 %, al menos aproximadamente 48 %, al menos aproximadamente 49 %, o al menos aproximadamente 50 %.

En otras realizaciones más, la molécula de ácido nucleico aislada de la invención contiene menos secuencias MARS/ARS (SEQ ID NO: 5 y 6) con respecto a SEQ ID NO: 3. En aún otras realizaciones, la molécula de ácido nucleico aislada de la invención contiene como máximo una secuencia MARS/ARS. En algunas realizaciones, la molécula de ácido nucleico aislada de la invención no contiene una secuencia MARS/ARS.

En algunas realizaciones, la molécula de ácido nucleico aislada de la invención no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 7).

En ciertas realizaciones, la molécula de ácido nucleico aislada de la invención contiene menos secuencias desestabilizantes (SEQ ID NO: 8 y 9) con respecto a SEQ ID NO: 3. En otras realizaciones, la molécula de ácido nucleico aislada de la invención contiene como máximo 4 secuencias desestabilizantes. En otras realizaciones más, la molécula de ácido nucleico aislada de la invención contiene como máximo 2 secuencias desestabilizantes. En aún otras realizaciones, la molécula de ácido nucleico aislada de la invención no contiene una secuencia desestabilizante.

En otras realizaciones, la molécula de ácido nucleico aislada de la invención no contiene una secuencia de poli-T (SEQ ID NO: 10). En aún otras realizaciones, la molécula de ácido nucleico aislada de la invención no contiene una secuencia de poli-A (SEQ ID NO: 11).

En una realización, la molécula de ácido nucleico aislada de la invención comprende además una secuencia heteróloga de nucleótidos. Por ejemplo, la secuencia heteróloga de nucleótidos puede codificar una secuencia heteróloga de aminoácidos que es un extensor de la semivida. En algunas realizaciones, la secuencia heteróloga de aminoácidos es una región constante de inmunoglobulina o una porción de la misma, transferrina, albúmina, polipéptido de unión a albúmina, una secuencia XTEN, Fc, el péptido del extremo C (CTP) de la subunidad β de gonadotropina coriónica humana, o una secuencia PAS. En otras realizaciones, la secuencia heteróloga de aminoácidos es una región Fc o un componente de unión a FcRn. En otras realizaciones más, la secuencia heteróloga de aminoácidos se une al extremo N o el extremo C de la secuencia de aminoácidos codificada por la secuencia de nucleótidos o se inserta entre dos aminoácidos en la secuencia de aminoácidos codificada por la secuencia de nucleótidos.

En una realización particular, la molécula de ácido nucleico aislada de la invención codifica una molécula híbrida de monómero-dímero que comprende el factor VIII.

En otra realización, la molécula de ácido nucleico aislada de la invención está operativamente unida a al menos un secuencia de control de la transcripción.

La presente invención también proporciona un vector que comprende la molécula de ácido nucleico de la invención.

La presente invención también proporciona una célula hospedadora que comprende la molécula de ácido nucleico de la invención. En algunas realizaciones, la célula hospedadora se selecciona del grupo que consiste en: una célula CHO, una célula HEK293, una célula BHK21, una célula PER.C6, una célula NS0 y una célula CAP.

La presente invención también proporciona un método de producción de un polipéptido con actividad de factor VIII, que comprende: cultivar la célula hospedadora de la invención en condiciones por las cuales se produce un polipéptido con actividad de factor VIII; y, recuperar el polipéptido con actividad de factor VIII. En otras realizaciones del método de producción de un polipéptido con actividad de factor VIII, la expresión del polipéptido con actividad de factor VIII es elevada con respecto a una célula hospedadora cultivada en las mismas condiciones que comprende una secuencia de nucleótidos de referencia que comprende SEQ ID NO: 3. En otras realizaciones del método, la célula hospedadora es una célula CHO. En otras realizaciones del método, la célula hospedadora es una célula HEK293.

La presente invención también proporciona un método de aumento de la expresión de un polipéptido con actividad de factor VIII en un sujeto que comprende administrar la molécula de ácido nucleico aislada de la invención o el vector de la invención a un sujeto en necesidad del mismo, en donde la expresión del polipéptido con actividad de factor VIII es elevada con respecto a una molécula de ácido nucleico de referencia que comprende SEQ ID NO: 3 o el vector que comprende la molécula de ácido nucleico de referencia .

La presente invención también proporciona un método de aumento de la expresión de un polipéptido con actividad de factor VIII que comprende cultivar la célula hospedadora de la invención en condiciones por las cuales un polipéptido con actividad de factor VIII se expresa por la molécula de ácido nucleico, en donde la expresión del polipéptido con actividad de factor VIII es elevada con respecto a una célula hospedadora cultivada en las mismas condiciones que comprende una secuencia de ácidos nucleicos de referencia que comprende SEQ ID NO: 3.

La presente invención también proporciona un método de mejora del rendimiento de un polipéptido con actividad de factor VIII que comprende cultivar la célula hospedadora de la invención en condiciones por las cuales se produce un polipéptido con actividad de factor VIII por la molécula de ácido nucleico, en donde el rendimiento del polipéptido con actividad de factor VIII es elevada con respecto a una célula hospedadora cultivada en las mismas condiciones que comprende una secuencia de ácidos nucleicos de referencia que comprende SEQ ID NO: 3.

La presente invención también proporciona un método de tratamiento de un trastorno hemorrágico que comprende: administrar a un sujeto en necesidad del mismo una molécula de ácido nucleico de la invención, un vector de la invención o un polipéptido de la invención. En algunas realizaciones del método de tratamiento de un trastorno hemorrágico, el trastorno hemorrágico se caracteriza por una deficiencia en el factor VIII. En algunas realizaciones, el trastorno hemorrágico es hemofilia. En algunas realizaciones, el trastorno hemorrágico es hemofilia A.

En algunas realizaciones del método de tratamiento de un trastorno hemorrágico, la actividad de factor VIII en plasma 24 horas después de la administración es elevada con respecto a un sujeto administrado con una molécula de ácido nucleico de referencia que comprende SEQ ID NO: 3, un vector que comprende la molécula de ácido nucleico de referencia, o un polipéptido codificado por la molécula de ácido nucleico de referencia.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La Figura 1 es un diagrama esquemático que muestra la localización de diversos sitios en la secuencia codificante del factor VIII BDD. Estos sitios se retiraron durante el proceso de optimización por codones.

La Figura 2 es la secuencia de nucleótidos del factor VIII BDD (SEQ ID NO: 1), optimizado por codones por un primer método de optimización por codones, descrito en el Ejemplo 1.

La Figura 3 es la secuencia de nucleótidos del factor VIII BDD (SEQ ID NO: 2), optimizado por codones por un segundo método de optimización por codones, descrito en el Ejemplo 2.

Las Figuras 4 A-B muestran el ajuste del sesgo de uso de codones en la secuencia de FVIII BDD optimizado (SEQ ID NO: 1). La Figura 4A muestra la frecuencia de codones relativa en la secuencia de FVIII BDD antes de la optimización por codones. El índice de adaptación de codones (IAC) humanos de la secuencia FVIII BDD de partida es 0,74. La Figura 4B muestra la frecuencia de codones relativa en la secuencia de FVIII BDD optimizado (SEQ ID NO: 1). La IAC humana de la secuencia optimizada resultante es 0,88. El eje X indica la posición relativa de los codones a lo largo de la longitud de la secuencia de nucleótidos FVIII BDD. El eje Y indica la frecuencia relativa del codón en cada posición dentro del genoma humano.

Las Figuras 5 A-B muestran la frecuencia de codones humanos óptimos en la secuencia de FVIII BDD optimizado (SEQ ID NO: 1). La Figura 5A muestra la frecuencia de codones óptimos en la secuencia de FVIII BDD antes de la optimización por codones. La Figura 5B muestra la frecuencia de codones óptimos en la secuencia de FVIII BDD después de la optimización por codones (SEQ ID NO: 1). El eje X indica la frecuencia de codones en el genoma humano. El eje Y indica el porcentaje de codones en la secuencia de FVIII BDD que se clasifican en cada categoría delineada en el eje X.

Las Figuras 6 A-B muestra el contenido de G/C de la secuencia de FVIII BDD optimizado (SEQ ID NO: 1). La Figura 6A muestra el contenido de G/C de la secuencia de FVIII BDD antes de la optimización por codones. El contenido de G/C de la secuencia de FVIII BDD inicial es 46,16 %. La Figura 6B muestra el contenido de G/C de la secuencia de FVIII BDD después de de la optimización por codones (SEQ ID NO: 1). El contenido de G/C de la secuencia FVIII BDD optimizado es 51,56 %. El eje X indica la posición relativa de los codones a lo largo de la longitud de la secuencia de nucleótidos de FVIII BDD. El eje Y indica el porcentaje de contenido de G/C.

La Figura 7 es un histograma que muestra la actividad de FVIII en plasma en ratones HemA 24 horas después de la inyección hidrodinámica con plásmidos que contienen o la secuencia de FVIII BDD inicial (círculos), secuencia de FVIII BDD optimizado (SEQ ID NO: 1) (cuadrados) o secuencia de FVIII BDD optimizado (SEQ ID NO: 2) (triángulos).

DESCRIPCIÓN DETALLADA DE LA INVENCION

Las construcciones a modo de ejemplo de la invención se ilustran en las figuras adjuntas y el listado de secuencias. Para proporcionar un claro entendimiento de la memoria descriptiva y las reivindicaciones, a continuación se proporcionan las siguientes definiciones.

I. Definiciones

Se debe observar que el término "un" o "una" entidad se refiere a uno o más de esa entidad: por ejemplo, "una secuencia de nucleótidos" se entiende que representa una o más secuencias de nucleótidos. Como tales, los términos "un" (o "una"), "uno o más," y "al menos uno" se pueden usar indistintamente en el presente documento.

5 El término "aproximadamente" se usa en el presente documento para significar aproximadamente, alrededor o en las regiones de. Cuando el término "aproximadamente" se usa junto con un intervalo numérico, modifica ese intervalo extendiendo los límites por encima y por debajo de los valores numéricos expuestos. En general, el término "aproximadamente" se usa en el presente documento para modificar un valor numérico por encima y por debajo del valor establecido por una varianza del 10 por ciento, arriba o abajo (más alto o más bajo).

10 El término "aislado" para los fines de la presente invención designa un material biológico (célula, ácido nucleico o proteína) que se ha retirado de su entorno original (el entorno en el que está naturalmente presente). Por ejemplo, un polinucleótido presente en el estado natural en una planta o un animal no está aislado, sin embargo, el mismo polinucleótido separado de los ácidos nucleicos adyacentes en los que está naturalmente presente se considera "aislado".

15 "Ácido nucleico", "molécula de ácido nucleico", "oligonucleótido" y "polinucleótido" se usan indistintamente y se refieren a la forma polimérica de éster de fosfato de los ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualquier análogo de fosfoéster de los mismos, tal como fosforotioatos y tioésteres, en cualquier forma monocatenaria, o una hélice bicatenaria. Son posibles hélices de ADN bicatenario-ADN, ADN-ARN y ARN-ARN. El término molécula de ácido nucleico, y en particular ADN o molécula de ARN, solo se refiere a la estructura primaria y secundaria de la molécula, y no lo limita a ninguna forma terciaria particular. Así, este término incluye ADN bicatenario encontrado, entre otros, en moléculas de ADN lineal o circular (por ejemplo, fragmentos de restricción), plásmidos, ADN superenrollado y cromosomas. En la discusión de la estructura de moléculas de ADN bicatenario particulares, las secuencias se pueden describir en el presente documento según la convención normal de dar solo la secuencia en la dirección 5' a 3' a lo largo de la cadena no transcrita de ADN (es decir, la cadena que tiene un homólogo de secuencia al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha experimentado una manipulación biológica molecular. El ADN incluye, pero no se limita a, ADNc, ADN genómico, ADN de plásmido, ADN sintético y ADN semisintético. Una "composición de ácido nucleico" de la invención comprende uno o más ácidos nucleicos como se describe en el presente documento.

30 Como se usa en el presente documento, una "región codificante" o "secuencia codificante" es una porción de polinucleótido que consiste en codones traducibles en aminoácidos. Aunque un "codón de terminación" (TAG, TGA o TAA) no se traduce normalmente en un aminoácido, se puede considerar que es parte de una región codificante, pero cualquier secuencia flanqueante, por ejemplo promotores, sitios de unión al ribosoma, terminadores transcripcionales, intrones y similares, no es parte de una región codificante. Los límites de una región codificante normalmente se determinan por un codón de iniciación en el extremo 5', que codifica el extremo amino del polipéptido resultante, y un codón de terminación de la traducción en el extremo 3', que codifica el extremo carboxilo del polipéptido resultante. Pueden estar presentes dos o más regiones codificantes en una única construcción de polinucleótido, por ejemplo, en un único vector, o en construcciones de polinucleótido separadas, por ejemplo, en vectores separados (diferentes). De esto resulta, entonces, que un único vector puede contener solo una única región codificante, o comprender dos o más regiones codificantes.

40 Ciertas proteínas secretadas por células de mamífero están asociadas con un péptido señal secretor que se escinde de la proteína madura una vez se ha iniciado la exportación de la cadena de proteína en crecimiento a través del retículo endoplasmático rugoso. Los expertos habituales en la técnica conocen que los péptidos señal, en general, se fusionan con el extremo N del polipéptido, y se escinden del polipéptido completo o de "longitud completa" para producir una forma secretada o "madura" del polipéptido. En ciertas realizaciones, se usa un péptido señal nativo, o un derivado funcional de esa secuencia que retiene la capacidad para dirigir la secreción del polipéptido al que está operativamente asociado. Alternativamente, se puede usar un péptido señal de mamífero heterólogo, por ejemplo, un activador de plasminógeno de tejido humano (TPA) o péptido señal de β -glucuronidasa de ratón, o un derivado funcional del mismo.

50 El término "en la dirección 3'" se refiere a una secuencia de nucleótidos que se localiza 3' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos en la dirección 3' se refieren a secuencias que siguen el punto de inicio de la transcripción. Por ejemplo, el codón de iniciación de la traducción de un gen se localiza en la dirección 3' del sitio de inicio de la transcripción.

55 El término "en la dirección 5'" se refiere a una secuencia de nucleótidos que se localiza 5' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos en la dirección 5' se refieren a secuencias que se localizan en el lado 5' de una región codificante o punto de inicio de la transcripción. Por ejemplo, la mayoría de los promotores se localizan en la dirección 5' del sitio de inicio de la transcripción.

Como se usa en el presente documento, el término "región reguladora" se refiere a secuencias de nucleótidos localizadas en la dirección 5' (secuencias no codificantes 5'), dentro de, o en la dirección 3' (secuencias no codificantes 3') de una región codificante, y que influyen en la transcripción, procesamiento de ARN, estabilidad, o traducción de la

región codificante asociada. Las regiones reguladoras pueden incluir promotores, secuencias conductoras de la traducción, intrones, secuencias de reconocimiento de la poliadenilación, sitios de procesamiento de ARN, sitios de unión a efector y estructuras de tallo-bucle. Si está prevista una región codificante para la expresión en una célula eucariota, una señal de poliadenilación y secuencia de terminación de la transcripción se localizará normalmente 3' con respecto a la secuencia codificante.

Un polinucleótido que codifica un producto génico, por ejemplo, un polipéptido, puede incluir un promotor y/u otros elementos de control de la transcripción o de la traducción operativamente asociados a una o más regiones codificantes. En una asociación operativa, una región codificante de un producto génico, por ejemplo, un polipéptido, se asocia con una o más regiones reguladoras de tal forma que pongan la expresión del producto génico bajo la influencia o control de la(s) región (regiones) reguladora(s). Por ejemplo, una región codificante y un promotor están "operativamente asociados" si la inducción de la función de promotor da como resultado la transcripción de ARNm que codifica el producto génico codificado por la región codificante, y si la naturaleza del enlace entre el promotor y la región codificante no interfiere con la capacidad del promotor para dirigir la expresión del producto génico o interferir con la capacidad del molde de ADN a transcribir. También se pueden asociar operativamente otros elementos de control de la transcripción, además de un promotor, por ejemplo, potenciadores, operadores, represores y señales de terminación de la transcripción, a una región codificante para dirigir la expresión del producto génico.

"Secuencias de control de la transcripción" se refiere a secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una secuencia codificante en una célula hospedadora. Los expertos en la técnica conocen una variedad de regiones de control de la transcripción. Estas incluyen, sin limitación, regiones de control de la transcripción que funcionan en células de vertebrado, tales como, pero no se limitan a, segmentos de promotores y potenciadores del citomegalovirus (el promotor temprano inmediato, conjuntamente con el intrón A), virus 40 simio (el promotor temprano) y retrovirus (tales como el virus del sarcoma de Rous). Otras regiones de control de la transcripción incluyen las derivadas de genes de vertebrados tales como actina, proteína de choque térmico, hormona de crecimiento bovina y β -globina de conejo, así como otras secuencias capaces de controlar la expresión génica en células eucariotas. Las regiones de control de la transcripción adecuadas adicionales incluyen promotores y potenciadores específicos de tejido, así como promotores inducibles por linfocinas (por ejemplo, promotores inducibles por interferones o interleucinas).

Similarmente, los expertos habituales en la técnica conocen una variedad de elementos de control de la traducción. Estos incluyen, pero no se limitan a, sitios de unión al ribosoma, codones de inicio y terminación de la traducción, y elementos derivados de picornavirus (particularmente un sitio interno de entrada al ribosoma, o IRES, también denominada una secuencia CITE).

El término "expresión", como se usa en el presente documento, se refiere a un proceso por el que un polinucleótido produce un producto génico, por ejemplo, un ARN o un polipéptido. Incluye, sin limitación, la transcripción del polinucleótido en ARN mensajero (ARNm), ARN de transferencia (ARNt), ARN de horquilla pequeña (ARNhp), ARN interferente pequeño (ARNip), o cualquier otro producto de ARN, y la traducción de un ARNm en un polipéptido. La expresión produce un "producto génico". Como se usa en el presente documento, un producto génico puede ser o un ácido nucleico, por ejemplo, un ARN mensajero producido por transcripción de un gen, o un polipéptido que se traduce de un transcrito. Los productos génicos descritos en el presente documento incluyen además ácidos nucleicos con modificaciones postranscripcionales, por ejemplo, poliadenilación o corte y empalme, o polipéptidos con modificaciones postraduccionales, por ejemplo, metilación, glucosilación, la adición de lípidos, asociación con otras subunidades de proteína, o escisión proteolítica. El término "rendimiento", como se usa en el presente documento, se refiere a la cantidad de un polipéptido producido por la expresión de un gen.

Un "vector" se refiere a cualquier vehículo para la clonación de y/o transferencia de un ácido nucleico en una célula hospedadora. Un vector puede ser un replicón al que se puede unir otro segmento de ácido nucleico de manera que provoque la replicación del segmento unido. Un "replicón" se refiere a cualquier elemento genético (por ejemplo, plásmido, fago, cósmido, cromosoma, virus) que actúa como una unidad de replicación autónoma *in vivo*, es decir, capaz de replicación bajo su propio control. El término "vector" incluye tanto vehículos virales como no virales para introducir el ácido nucleico en una célula *in vitro*, *ex vivo* o *in vivo*. Se conocen un gran número de vectores y se usan en la técnica incluyendo, por ejemplo, plásmidos, virus eucariotas modificados, o virus bacterianos modificados. La inserción de un polinucleótido en un vector adecuado se puede llevar a cabo uniendo los fragmentos de polinucleótidos apropiados en un vector elegido que tiene extremos cohesivos complementarios.

Los vectores se pueden manipular para codificar marcadores o indicadores de selección que proporcionan la selección o identificación de células que han incorporado el vector. La expresión de marcadores o indicadores de selección permite la identificación y/o selección de células hospedadoras que incorporan y expresan otras regiones codificantes contenidas en el vector. Los ejemplos de genes marcadores de selección conocidos y usados en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomycin, gentamicina, kanamicina, higromicina, el herbicida bialafos, sulfonamida y similares; y genes que se usan como marcadores fenotípicos, es decir, genes reguladores de antocianina, gen isopentanol transferasa y similares. Los ejemplos de indicadores conocidos y usados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), -galactosidasa (LacZ), -glucuronidasa (Gus) y similares. También se puede considerar que los marcadores de selección son indicadores.

El término "marcador de selección" se refiere a un factor identificador, normalmente un gen de resistencia a antibióticos o a productos químicos, que es capaz de ser seleccionado para basarse en el efecto del gen marcador, es decir, resistencia a un antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes y similares, en donde el efecto se usa para hacer el seguimiento de la herencia de un ácido nucleico de interés y/o para identificar una célula u organismo que ha heredado el ácido nucleico de interés. Los ejemplos de genes marcadores de selección conocidos y usados en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomycin, gentamicina, kanamicina, higromicina, herbicida bialafos, sulfonamida y similares; y genes que se usan como marcadores fenotípicos, es decir, genes reguladores de antocianina, gen isopentanol transferasa y similares.

El término "gen indicador" se refiere a un ácido nucleico que codifica un factor identificador que es capaz de ser identificado basándose en el efecto del gen indicador, en donde el efecto se usa para hacer el seguimiento de la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés, y/o para medir la inducción o transcripción de la expresión génica. Los ejemplos de genes indicadores conocidos y usados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), β -galactosidasa (LacZ), β -glucuronidasa (Gus) y similares. Los genes marcadores de selección también se pueden considerar genes indicadores.

"Promotor" y "secuencia promotora" se usan indistintamente y se refieren a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una secuencia codificante se localiza 3' con respecto a una secuencia promotora. Los promotores se pueden obtener en su totalidad de un gen nativo, o estar compuestos de elementos diferentes derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintético. Se entiende por los expertos en la técnica que los diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes estadios de desarrollo, o en respuesta a diferentes condiciones medioambientales o fisiológicas. Los promotores que provocan que un gen se exprese en la mayoría de los tipos de células en la mayoría de los casos se denominan comúnmente "promotores constitutivos". Los promotores que provocan que un gen se exprese en un tipo específico de célula se denominan comúnmente "promotores específicos de célula" o "promotores específicos de tejido". Los promotores que provocan que un gen se exprese en un estadio de desarrollo específico o diferenciación celular se denominan comúnmente "promotores específicos del desarrollo" o "promotores específicos de la diferenciación celular". Los promotores que se inducen y provocan que se exprese un gen después de la exposición o tratamiento de la célula con un agente, molécula biológica, producto químico, ligando, luz o similares que inducen el promotor se denominan comúnmente "promotores inducibles" o "promotores regulables". Se reconoce además que puesto que en la mayoría de los casos los límites exactos de secuencias reguladoras no se han definido completamente, fragmentos de ADN de diferentes longitudes pueden tener actividad de promotor idéntica.

La secuencia promotora normalmente está unida en su extremo 3' por el sitio de iniciación de la transcripción y se extiende aguas arriba (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro de la secuencia promotora se encontrará un sitio de iniciación de la transcripción (convenientemente definido, por ejemplo, por mapeo con nucleasa S1), así como dominios de unión a proteína (secuencias consenso) responsables de la unión de ARN polimerasa.

Los términos "endonucleasa de restricción" y "enzima de restricción" se usan indistintamente y se refieren a una enzima que se une y corta dentro de una secuencia de nucleótidos específica dentro de ADN bicatenario.

El término "plásmido" se refiere a un elemento extracromosómico que frecuentemente lleva un gen que no es parte del metabolismo central de la célula, y normalmente en forma de moléculas de ADN bicatenario circular. Dichos elementos pueden ser secuencias que se replican de forma autónoma, secuencias que se integran en el genoma, secuencias de fagos o nucleótidos, lineales, circulares, o superenrolladas, de un ADN o ARN mono o bicatenario, derivados de cualquier fuente, en las que varias secuencias de nucleótidos se han unido o recombinado en una única construcción que es capaz de introducir un fragmento de promotor y secuencia de ADN para un producto génico seleccionado junto con secuencia no traducida de 3' apropiada dentro de una célula.

Los vectores virales eucariotas que se pueden usar incluyen, pero no se limitan a, vectores de adenovirus, vectores de retrovirus, vectores de virus adeno-asociado, poxvirus, por ejemplo, vectores de virus de la variolovacuna, vectores de baculovirus, o vectores de virus del herpes. Los vectores no virales incluyen plásmidos, liposomas, lípidos eléctricamente cargados (citofectinas), complejos de ADN-proteína y biopolímeros.

Un "vector de clonación" se refiere a un "replicón", que es una unidad de longitud de un ácido nucleico que se replica secuencialmente y que comprende un origen de replicación, tal como un plásmido, fago o cósmido, al que se puede unir otro segmento de ácido nucleico de manera que provoque la replicación del segmento unido. Ciertos vectores de clonación son capaces de replicación en un tipo de célula, por ejemplo, bacterias, y expresión en otro, por ejemplo, células eucariotas. Los vectores de clonación normalmente comprenden una o más secuencias que se pueden usar para la selección de células que comprenden el vector y/o uno o más sitios de clonación múltiple para la inserción de secuencias de ácidos nucleicos de interés.

El término "vector de expresión" se refiere a un vehículo diseñado para permitir la expresión de una secuencia de ácidos nucleicos insertada tras la inserción en una célula hospedadora. La secuencia de ácidos nucleicos insertada se pone en asociación operativa con regiones reguladoras, como se ha descrito anteriormente.

5 Los vectores se introducen en células hospedadoras por métodos bien conocidos en la técnica, por ejemplo, transfección, electroporación, microinyección, transducción, fusión de células, DEAE-dextrano, precipitación con fosfato de calcio, lipofección (fusión de lisosomas), uso de una pistola de genes, o un transportador de vector de ADN.

"Cultivo", "para cultivar" y "cultivar", como se usan en el presente documento, significa incubar células en condiciones *in vitro* que permite el crecimiento o división celular, o mantener las células en un estado vivo. "Células cultivadas", como se usa en el presente documento, significa células que son propagadas *in vitro*.

10 Como se usa en el presente documento, el término "polipéptido" pretende englobar un "polipéptido" singular, así como "polipéptidos" plurales, y se refiere a una molécula compuesta de monómeros (aminoácidos) linealmente unidos por enlaces amida (también conocidos como enlaces peptídicos). El término "polipéptido" se refiere a cualquier cadena o cadenas de dos o más aminoácidos, y no se refiere a una longitud específica de producto. Así, los péptidos, dipéptidos, tripéptidos, oligopéptidos, "proteína", "cadena de aminoácidos", o cualquier otro término usado para referirse a una
 15 "polipéptido" se puede usar en lugar de, o indistintamente, con cualquiera de estos términos. El término "polipéptido" también pretende referirse a los productos de modificaciones posteriores a la expresión del polipéptido, que incluyen, sin limitación, glucosilación, acetilación, fosforilación, amidación, derivatización por grupos protectores/de bloqueo conocidos, escisión proteolítica, o modificación por aminoácidos que no existen de forma natural. Un polipéptido se
 20 puede obtener de una fuente biológica natural o se produce por tecnología recombinante, pero no se traduce necesariamente de una secuencia de ácidos nucleicos designada. Se puede generar de cualquier manera, incluyendo por síntesis química.

El término "aminoácido" incluye alanina (Ala o A); arginina (Arg o R); asparagina (Asn o N); ácido aspártico (Asp o D);
 25 cisteína (Cys o C); glutamina (Gln o Q); ácido glutámico (Glu o E); glicina (Gly o G); histidina (His o H); isoleucina (Ile o I); leucina (Leu o L); lisina (Lys o K); metionina (Met o M); fenilalanina (Phe o F); prolina (Pro o P); serina (Ser o S);
 30 treonina (Thr o T); triptófano (Trp o W); tirosina (Tyr o Y); y valina (Val o V). También están dentro del alcance de la invención los aminoácidos no tradicionales e incluyen norleucina, ornitina, norvalina, homoserina, y otros análogos de restos de aminoácidos tales como los descritos en Ellman et al. Meth. Enzym. 202:301-336 (1991). Para generar dichos restos de aminoácidos que no existen de forma natural, se pueden usar los procedimientos de Noren et al.
 35 Science 244:182 (1989) y Ellman et al., arriba. Brevemente, estos procedimientos implican activar químicamente un ARNt supresor con un resto de aminoácido que no existe de forma natural, seguido por transcripción y traducción *in vitro* del ARN. La introducción del aminoácido no tradicional también se puede lograr usando químicas de péptidos conocidas en la técnica. Como se usa en el presente documento, el término "aminoácido polar" incluye aminoácidos que tienen carga neta cero, pero tienen cargas parciales distintas de cero en diferentes porciones de sus cadenas laterales (por ejemplo M, F, W, S, Y, N, Q, C). Estos aminoácidos pueden participar en interacciones hidrófobas e interacciones electrostáticas. Como se usa en el presente documento, el término "aminoácido cargado" incluye aminoácidos que pueden tener carga neta distinta de cero en sus cadenas laterales (por ejemplo, R, K, H, E, D). Estos aminoácidos pueden participar en interacciones hidrófobas e interacciones electrostáticas.

40 Un polipéptido "aislado" o un fragmento, variante, o derivado del mismo, se refiere a un polipéptido que no está en su ámbito natural. No se requiere nivel de purificación particular. Por ejemplo, un polipéptido aislado se puede retirar simplemente de su entorno nativo o natural. Los polipéptidos y proteínas recombinantemente producidos expresados en células hospedadoras se consideran aislados para el fin de la invención, ya que son polipéptidos nativos o recombinantes que han sido separados, fraccionados, o parcialmente o sustancialmente purificados, por cualquier técnica adecuada.

45 También se incluyen en la presente invención fragmentos o variantes de polipéptidos, y cualquier combinación de los mismos. El término "fragmento" o "variante", cuando se refiere a dominios de unión a polipéptido o moléculas de unión de la presente invención, incluye cualquier polipéptido que retenga al menos algunas de las propiedades (por ejemplo, afinidad de unión de FcRn por un dominio de unión de FcRn o variante de Fc, actividad de coagulación para una variante de FVIII, o la actividad de unión de FVIII para el fragmento de VWF) del polipéptido de referencia. Los
 50 fragmentos de polipéptidos incluyen fragmentos proteolíticos, así como fragmentos de delección, además de los fragmentos específicos de anticuerpo tratados en cualquier parte en el presente documento, pero no incluyen el polipéptido de longitud completa que existe de forma natural (o polipéptido maduro). Las variantes de dominios de unión de polipéptido o moléculas de unión de la presente invención incluyen fragmentos como se ha descrito anteriormente, y también polipéptidos con secuencias de aminoácidos alteradas debido a sustituciones de
 55 aminoácidos, delecciones, o inserciones. Las variantes pueden existir de forma natural o no existir de forma natural. Las variantes que no existen de forma natural se pueden producir usando técnicas de mutagénesis conocidas en la técnica. Los polipéptidos de variante pueden comprender sustituciones conservativas o no conservativas de aminoácidos, delecciones o adiciones.

60 Una "sustitución de aminoácidos conservativa" es una en la que el resto de aminoácido se sustituye por un resto de aminoácido que tiene una cadena lateral similar. Se han definido en la técnica las familias de restos de aminoácidos

que tienen cadenas laterales similares, que incluyen cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales beta-ramificadas (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Así, si un aminoácido en un polipéptido se sustituye por otro aminoácido de la misma familia de cadena lateral, se considera que la sustitución es conservativa. En otra realización, una cadena de aminoácidos se puede sustituir conservativamente con una cadena estructuralmente similar que se diferencia en orden y/o composición de los miembros de familia de la cadena lateral.

El término "porcentaje de identidad", como se conoce en la técnica, es una relación entre dos o más secuencias de polipéptidos o dos o más secuencias de polinucleótidos, como se determina comparando las secuencias. En la técnica, "identidad" también significa el grado de vinculación de secuencias entre secuencias de polipéptidos o polinucleótidos, según lo requiera el caso, como se ha determinado por el emparejamiento entre series de dichas secuencias. La "identidad" se puede calcular fácilmente por métodos conocidos, que incluyen, pero no se limitan a, los descritos en: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Parte I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); y Sequence Analysis Primer (Gribskov, M. y Devereux, J., eds.) Stockton Press, New York (1991). Los métodos preferidos para determinar la identidad se diseñan para dar el mejor emparejamiento entre las secuencias probadas. Los métodos para determinar la identidad están codificados en programas informáticos públicamente disponibles. Se pueden realizar alineamientos de secuencias y cálculos del porcentaje de identidad usando software de análisis de secuencias, tal como el programa Megalign del paquete de cálculo bioinformático LASERGENE (DNASTAR Inc., Madison, WI), el paquete de programas GCG (Wisconsin Package versión 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., J Mol. Biol. 215:403 (1990)) y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE. UU.). Dentro del contexto de la presente solicitud se entenderá que donde se usa el software de análisis de secuencias para el análisis, que los resultados del análisis se basarán en los "valores por defecto" del programa referenciado, a menos que se especifique de otro modo. Como se usa en el presente documento, "valores por defecto" significará cualquier conjunto de valores o parámetros que se cargan originalmente con el software cuando se inicia por primera vez. A efectos de determinar el porcentaje de identidad entre una secuencia de FVIII BDD optimizado de la invención y una secuencia de referencia, solo los nucleótidos en la secuencia de referencia correspondientes a los nucleótidos en la secuencia de FVIII BDD optimizado de la invención se usan para calcular el porcentaje de identidad. Por ejemplo, cuando se compara una secuencia de nucleótidos de FVIII de longitud completa que contiene el dominio B con una secuencia de nucleótidos de FVIII deleciónado en el dominio B (BDD) optimizado de la invención, la porción del alineamiento que incluye el dominio A1, A2, A3, C1 y C2 se usará para calcular el porcentaje de identidad. Los nucleótidos en la porción de la secuencia de FVIII de longitud completa que codifican el dominio B (que dará como resultado un gran "hueco" en el alineamiento) no se contarán como un despareamiento.

Como se usa en el presente documento, "nucleótidos correspondientes a nucleótidos en la secuencia de FVIII de BDD optimizado de la invención" se identifican por alineamiento de la secuencia de FVIII BDD optimizado de la invención para maximizar la identidad con la secuencia de FVIII de referencia. El número usado para identificar un aminoácido equivalente en una secuencia de FVIII de referencia se basa en el número usado para identificar el aminoácido correspondiente en la secuencia de FVIII BDD optimizado de la invención.

Una proteína de "fusión" o "quimérica" comprende una primera secuencia de aminoácidos unida a una segunda secuencia de aminoácidos con la que no se une naturalmente en la naturaleza. Las secuencias de aminoácidos que normalmente existen en proteínas separadas se pueden poner juntas en el polipéptido de fusión, o las secuencias de aminoácidos que normalmente existen en la misma proteína se pueden poner en una nueva disposición en el polipéptido de fusión, por ejemplo, fusión de un dominio de factor VIII de la invención con un dominio Fc de Ig. Una proteína de fusión se crea, por ejemplo, por síntesis química, o por creación y traducción de un polinucleótido en el que las regiones de péptido están codificadas en la relación deseada. Una proteína quimérica puede comprender además una segunda secuencia de aminoácidos asociada a la primera secuencia de aminoácidos por un enlace no peptídico covalente, o un enlace no covalente.

Como se usa en el presente documento, el término "semivida" se refiere a una semivida biológica de un polipéptido particular *in vivo*. La semivida se puede representar por el tiempo requerido para que la mitad de la cantidad administrada a un sujeto sea eliminada de la circulación y/u otros tejidos en el animal. Cuando se construye una curva de eliminación de un polipéptido dado en función del tiempo, la curva es normalmente bifásica con una fase α rápida y una fase β larga. La fase α normalmente representa un equilibrio del polipéptido de Fc administrado entre el espacio intra- y extra-vascular y se determina, en parte, por el tamaño del polipéptido. La fase β normalmente representa el catabolismo del polipéptido en el espacio intravascular. En algunas realizaciones, FVIII y las proteínas quiméricas que comprenden FVIII son monofásicas, y así no tienen una fase alfa, sino solo la fase beta individual. Por tanto, en ciertas realizaciones, el término semivida, como se usa en el presente documento, se refiere a la semivida del polipéptido en la fase β .

El término "unido", como se usa en el presente documento, se refiere a una primera secuencia de aminoácidos o secuencia de nucleótidos covalentemente o no covalentemente unida a una segunda secuencia de aminoácidos o secuencia de nucleótidos, respectivamente. La primera secuencia de aminoácidos o de nucleótidos se puede unir directamente o yuxtaponer con la segunda secuencia de aminoácidos o de nucleótidos, o alternativamente una secuencia intercalada puede unir covalentemente la primera secuencia con la segunda secuencia. El término "unido" significa no solo una fusión de una primera secuencia de aminoácidos con una segunda secuencia de aminoácidos en el extremo C o el extremo N, sino que también incluye la inserción de toda la primera secuencia de aminoácidos (o la segunda secuencia de aminoácidos) en cualesquiera dos aminoácidos en la segunda secuencia de aminoácidos (o la primera secuencia de aminoácidos, respectivamente). En una realización, la primera secuencia de aminoácidos se puede unir a una segunda secuencia de aminoácidos por un enlace peptídico o un conector. La primera secuencia de nucleótidos se puede unir a una segunda secuencia de nucleótidos por un enlace fosfodiéster o un conector. El conector puede ser un péptido o un polipéptido (para cadenas de polipéptidos) o un nucleótido o una cadena de nucleótidos (para cadenas de nucleótidos), o cualquier resto químico (para tanto cadenas de polipéptidos como de polinucleótidos). El término "unido" también se indica por un guión (-).

Como se usa en el presente documento, el término "asociado a" se refiere a un enlace covalente o no covalente formado entre una primera cadena de aminoácidos y una segunda cadena de aminoácidos. En una realización, el término "asociado a" significa un enlace no peptídico covalente, o un enlace no covalente. Esta asociación se puede indicar por dos puntos, es decir, (:). En otra realización, significa un enlace covalente, excepto un enlace peptídico. Por ejemplo, el aminoácido cisteína comprende un grupo tiol que puede formar un enlace disulfuro o puente con un grupo tiol en un segundo resto de cisteína. En la mayoría de las moléculas de IgG que existen de forma natural, las regiones CH1 y CL están asociadas por un enlace disulfuro y las dos cadenas pesadas están asociadas por dos enlaces disulfuro en las posiciones correspondientes a 239 y 242 usando el sistema de numeración de Kabat (posición 226 o 229, sistema de numeración EU). Los ejemplos de enlaces covalentes incluyen, pero no se limitan a, un enlace peptídico, un enlace metálico, un enlace de hidrógeno, un enlace disulfuro, un enlace sigma, un enlace pi, un enlace delta, un enlace glucosídico, un enlace agnóstico, un enlace flexionado, un enlace dipolar, un esqueleto pi, un doble enlace, un triple enlace, un cuádruple enlace, un quíntuple enlace, un séxtuple enlace, conjugación, hiperconjugación, aromaticidad, hapticidad o antienlace. Los ejemplos no limitantes de enlace no covalente incluyen un enlace iónico (por ejemplo, enlace catión-pi o enlace de sal), un enlace metálico, un enlace de hidrógeno (por ejemplo, enlace de dihidrógeno, complejo de dihidrógeno, enlace de hidrógeno de baja barrera, o enlace de hidrógeno simétrico), fuerza de van der Waals, fuerza de dispersión de London, un enlace mecánico, un enlace halógeno, aurofilicidad, intercalación, apilamiento, fuerza entrópica o polaridad química.

El término "híbrido monómero-dímero" usado en el presente documento se refiere a una proteína quimérica que comprende una primera cadena de polipéptidos y una segunda cadena de polipéptidos, que están asociadas entre sí por un enlace disulfuro, en donde la primera cadena comprende un factor de coagulación, por ejemplo, factor VIII, y una primera región Fc y la segunda cadena comprende, consiste esencialmente en, o consiste en una segunda región Fc sin el factor de coagulación. La construcción híbrida monómero-dímero es así un híbrido que comprende un aspecto de monómero que tiene solo un factor de coagulación y un aspecto de dímero que tiene dos regiones Fc.

Hemostasia, como se usa en el presente documento, significa la detención o el ralentizamiento del sangrado o la hemorragia; o la detención o el ralentizamiento de la circulación sanguínea a través de un vaso sanguíneo o parte del cuerpo.

Trastorno hemostático, como se usa en el presente documento, significa una afección genéticamente heredada o adquirida caracterizada por una tendencia a hemorragia, o espontáneamente, o como resultado de traumatismo, debido a una capacidad alterada o incapacidad de formar un coágulo de fibrina. Los ejemplos de dichos trastornos incluyen las hemofilias. Las tres formas principales son hemofilia A (deficiencia de factor VIII), hemofilia B (deficiencia de factor IX o "enfermedad de Christmas") y hemofilia C (deficiencia de factor XI, tendencia a sangrado leve). Otros trastornos hemostáticos incluyen, por ejemplo, enfermedad de von Willebrand, deficiencia de factor XI (deficiencia de PTA), deficiencia de factor XII, deficiencias o anomalías estructurales en fibrinógeno, protrombina, factor V, factor VII, factor X o factor XIII, síndrome de Bernard-Soulier, que es un defecto o deficiencia en GPIb. GPIb, el receptor para vWF, puede ser defectuoso y conducir a una ausencia de formación de coágulos primarios (hemostasia primaria) y elevada tendencia al sangrado), y trombostenia de Glanzman y Naegeli (trombostenia de Glanzmann). En insuficiencia hepática (formas agudas y crónicas), existe producción insuficiente de factores de coagulación por el hígado; esto puede aumentar el riesgo de sangrado.

Las moléculas de ácido nucleico aisladas de la invención se pueden usar profilácticamente. Como se usa en el presente documento, el término "tratamiento profiláctico" se refiere a la administración de una molécula antes de un episodio de sangrado. En una realización, el sujeto en necesidad de un agente hemostático general se somete a, o está a punto de someterse a, cirugía. La proteína quimérica de la invención se puede administrar antes o después de la cirugía como un profiláctico. La proteína quimérica de la invención se puede administrar durante o después de la cirugía para controlar un episodio de sangrado agudo. La cirugía puede incluir, pero no se limita a, trasplante de hígado, resección hepática, procedimientos dentales, o trasplante de células madre.

Las moléculas de ácido nucleico aisladas de la invención también se usan para tratamiento a demanda. El término "tratamiento a demanda" se refiere a la administración de una molécula de ácido nucleico aislada en respuesta a

síntomas de un episodio hemorrágico o antes de una actividad que puede provocar sangrado. En un aspecto, el tratamiento a demanda (episódico) se puede administrar a un sujeto cuando el sangrado empieza, tal como después de una lesión, o cuando se espera sangrado, tal como antes de cirugía. En otro aspecto, el tratamiento a demanda se puede administrar antes de actividades que aumentan el riesgo de sangrado, tales como deportes de contacto.

5 Como se usa en el presente documento, el término "sangrado agudo" se refiere a un episodio de sangrado independientemente de la causa subyacente. Por ejemplo, un sujeto puede tener traumatismo, uremia, un trastorno de sangrado hereditario (por ejemplo, deficiencia de factor VII), un trastorno plaquetario, o resistencia debido al desarrollo de anticuerpos contra factores de coagulación.

10 Tratar, tratamiento, tratando, como se usa en el presente documento, se refiere a, por ejemplo, la reducción en la gravedad de una enfermedad o afección; la reducción en la duración de una evolución de la enfermedad; la mejora de uno o más síntomas asociados a una enfermedad o afección; la provisión de efectos beneficiosos a un sujeto con una enfermedad o afección, sin curar necesariamente la enfermedad o afección, o la profilaxis de uno o más síntomas asociados a una enfermedad o afección. En una realización, el término "tratar" o "tratamiento" significa mantener un nivel valle de FVIII de al menos aproximadamente 1 UI/dL, 2 UI/dL, 3 UI/dL, 4 UI/dL, 5 UI/dL, 6 UI/dL, 7 UI/dL, 8 UI/dL, 15 9 UI/dL, 10 UI/dL, 11 UI/dL, 12 UI/dL, 13 UI/dL, 14 UI/dL, 15 UI/dL, 16 UI/dL, 17 UI/dL, 18 UI/dL, 19 UI/dL o 20 UI/dL en un sujeto administrando una molécula de ácido nucleico aislada de la invención. En otra realización, tratar o tratamiento significa mantener un nivel valle de FVIII entre aproximadamente 1 y aproximadamente 20 UI/dL, aproximadamente 2 y aproximadamente 20 UI/dL, aproximadamente 3 y aproximadamente 20 UI/dL, aproximadamente 4 y aproximadamente 20 UI/dL, aproximadamente 5 y aproximadamente 20 UI/dL, 20 aproximadamente 6 y aproximadamente 20 UI/dL, aproximadamente 7 y aproximadamente 20 UI/dL, aproximadamente 8 y aproximadamente 20 UI/dL, aproximadamente 9 y aproximadamente 20 UI/dL, o aproximadamente 10 y aproximadamente 20 UI/dL. Tratamiento o tratar de una enfermedad o afección también puede incluir mantener la actividad de FVIII en un sujeto a un nivel comparable a al menos aproximadamente 1 %, 2 %, 3 %, 25 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, o 20 % de la actividad de FVIII en un sujeto no hemofílico. El nivel valle mínimo requerido para el tratamiento se puede medir por uno o más métodos conocidos y se puede ajustar (aumentar o reducir) para cada persona.

"Administrar", como se usa en el presente documento, significa dar un polipéptido de factor VIII farmacéuticamente aceptable a un sujeto por una vía farmacéuticamente aceptable. Las vías de administración pueden ser intravenosas, 30 por ejemplo, inyección intravenosa e infusión intravenosa. Las vías de administración adicionales incluyen, por ejemplo, administración subcutánea, intramuscular, oral, nasal y pulmonar. Se pueden administrar polipéptidos quiméricos y proteínas híbridas como parte de una composición farmacéutica que comprende al menos un excipiente.

Como se usa en el presente documento, la expresión "sujeto en necesidad del mismo" incluye sujetos, tales como sujetos mamíferos, que se beneficiarían de la administración de una molécula de ácido nucleico de la invención, por 35 ejemplo, para mejorar la homeostasia. En una realización, los sujetos incluyen, pero no se limitan a, individuos con hemofilia. En otra realización, los sujetos incluyen, pero no se limitan a, los individuos que han desarrollado un inhibidor de FVIII y así están en necesidad de una terapia de derivación. El sujeto puede ser un adulto o un menor de edad (por ejemplo, menos de 12 años).

Como se usa en el presente documento, el término "factor de coagulación" se refiere a moléculas, o análogos de las mismas, que existen de forma natural o se producen recombinantemente que previenen o disminuyen la duración de 40 un episodio hemorrágico en un sujeto. En otras palabras, significa que las moléculas que tienen actividad pro-coagulante, es decir, son responsables de la conversión de fibrinógeno en una malla de fibrina insoluble que causa que la sangre coagule. Un "factor de coagulación activable" es un factor de coagulación en una forma inactiva (por ejemplo, en su forma de zimógeno) que es capaz de ser convertido en una forma activa.

La actividad de coagulación, como se usa en el presente documento, significa la capacidad de participar en una 45 cascada de reacciones bioquímicas que culmina en la formación de un coágulo de fibrina y/o reduce la intensidad, duración o frecuencia de la hemorragia o episodio hemorrágico.

Como se usa en el presente documento, los términos "heterólogo" o "exógeno" se refieren a dichas moléculas que no se encuentran normalmente en un contexto dado, por ejemplo, en una célula o en un polipéptido. Por ejemplo, se 50 puede introducir una molécula exógena o heteróloga en una célula y solo están presentes después de la manipulación de la célula, por ejemplo, por transfección u otras formas de ingeniería genética o una secuencia heteróloga de aminoácidos pueden estar presentes en una proteína en la que no se encuentra naturalmente.

Como se usa en el presente documento, el término "secuencia de nucleótidos heteróloga" se refiere a una secuencia de nucleótidos que no ocurre naturalmente con una secuencia de polinucleótidos dada. En una realización, la 55 secuencia heteróloga de nucleótidos codifica un polipéptido capaz de prolongar la semivida de FVIII. En otra realización, la secuencia heteróloga de nucleótidos codifica un polipéptido que aumenta el radio hidrodinámico de FVIII. En otras realizaciones, la secuencia heteróloga de nucleótidos codifica un polipéptido que mejora una o más propiedades farmacocinéticas de FVIII sin afectar significativamente su actividad o función biológica (por ejemplo, su actividad procoagulante). En algunas realizaciones, FVIII se une o conecta con el polipéptido codificado por la secuencia heteróloga de nucleótidos por un conector. Los ejemplos no limitantes de restos de polipéptido codificados

por secuencias de nucleótidos heterólogas incluyen una región constante de inmunoglobulina o una porción de la misma, albúmina o un fragmento de la misma, un resto de unión a albúmina, una transferrina, los polipéptidos PAS de la solicitud de patente de EE. UU. N° 20100292130, una secuencia HAP, transferrina o un fragmento de la misma, el péptido del extremo C (CTP) de la subunidad β de gonadotropina coriónica humana, molécula pequeña de unión a albúmina, una secuencia XTEN, restos de unión a FcRn (por ejemplo, regiones Fc completas o porciones de la misma que se unen a FcRn), regiones Fc de una sola cadena (regiones scFc, por ejemplo, como se describe en los documentos de patente US 2008/0260738, WO 2008/012543 o WO 2008/1439545), conectores de poliglicina, conectores de poliserina, péptidos y polipéptidos cortos de 6-40 aminoácidos de dos tipos de aminoácidos seleccionados de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P) con grados variables de estructura secundaria desde menos del 50 % hasta más del 50 %, entre otros, o dos o más combinaciones de los mismos. En algunas realizaciones, el polipéptido codificado por la secuencia heteróloga de nucleótidos se une a un resto no de polipéptido. Los ejemplos no limitantes de los restos no de polipéptido incluyen polietilenglicol (PEG), moléculas pequeñas de unión a albúmina, ácido polisialílico, hidroxietilalmidón (HES), un derivado del mismo, o cualquier combinación de los mismos.

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

En una realización, la "región Fc" se refiere a la porción de una única cadena pesada de Ig que empieza en la región bisagra justo en la dirección 5' del sitio de escisión por papaína (es decir, el resto 216 en IgG, considerando que el primer resto de la región constante de la cadena pesada es 114) y que termina en el extremo C del anticuerpo. Por consiguiente, un dominio Fc completo comprende al menos un dominio bisagra, un dominio CH2 y un dominio CH3.

La región Fc de una región constante de Ig, dependiendo del isotipo de Ig, puede incluir los dominios CH2, CH3 y CH4, así como la región bisagra. Las proteínas quiméricas que comprenden una región Fc de una Ig conceden varias propiedades deseables a una proteína quimérica que incluyen elevada estabilidad, elevada semivida en suero (véase Capon et al., 1989, Nature 337:525), así como unión a receptores de Fc tales como el receptor de Fc neonatal (FcRn) (patentes de EE. UU. N° 6.086.875, 6.485.726, 6.030.613; documentos de patente WO 03/077834; US2003-0235536A1).

Una "secuencia de nucleótidos de referencia", cuando se usa en el presente documento como una comparación con una secuencia de nucleótidos de la invención, es una secuencia de polinucleótidos esencialmente idéntica a la secuencia de nucleótidos de la invención, excepto que las porciones correspondientes a la secuencia de FVIII no están optimizadas. Por ejemplo, la secuencia de nucleótidos de referencia para una molécula de ácido nucleico que consiste en FVIII BDD optimizado por codones de SEQ ID NO: 1 y una secuencia heteróloga de nucleótidos que codifica una región Fc de una sola cadena unida a SEQ ID NO: 1 en su extremo 3' es una molécula de ácido nucleico que consiste en el FVIII BDD original (o "parental") de SEQ ID NO: 3 y la secuencia de nucleótidos heteróloga idéntica que codifica una región Fc de una sola cadena unida a SEQ ID NO: 3 en su extremo 3'.

Un "índice de adaptación de codones", como se usa en el presente documento, se refiere a una medida del sesgo de uso de codones. Un índice de adaptación de codones (IAC) mide la desviación de una secuencia de genes codificantes de proteína dada con respecto a un conjunto de referencia de genes (Sharp PM y Li WH, Nucleic Acids Res. 15(3):1281-95 (1987)). IAC se calcula determinando la media geométrica del peso asociado a cada codón a lo largo de la longitud de la secuencia del gen (medida en codones):

$$IAC = \exp\left(\frac{1}{L} \sum_{i=1}^L \ln(w_i(i))\right), \quad (I)$$

Para cada aminoácido, el peso de cada uno de sus codones, en IAC, se calcula como la relación entre la frecuencia observada del codón (f_i) y la frecuencia del codón sinónimo (f_j) para ese aminoácido:

Fórmula 2:

$$w_i = \frac{f_i}{\max(f_j)} \quad i, j \in [\text{codones sinónimos para el aminoácido}] \quad (II)$$

Como se usa en el presente documento, el término "optimizado", con respecto a secuencias de nucleótidos, se refiere a una secuencia de polinucleótidos que codifica un polipéptido, en donde la secuencia de polinucleótidos se ha mutado para potenciar una propiedad de esa secuencia de polinucleótidos. En algunas realizaciones, la optimización se hace para aumentar los niveles de transcripción, aumentar los niveles de traducción, aumentar los niveles de ARNm en

- estado estacionario, aumentar o disminuir la unión de proteínas reguladoras, tales como factores de transcripción generales, aumentar o disminuir el corte y empalme, o aumentar el rendimiento del polipéptido producido por la secuencia de polinucleótidos. Los ejemplos de cambios que se pueden hacer a una secuencia de polinucleótidos para optimizarla incluyen la optimización por codones, optimización del contenido de G/C, retirada de secuencias de repetición, retirada de elementos ricos en AT, retirada de espacios crípticos de corte y empalme, retirada de elementos que actúan en cis que reprimen la transcripción o traducción, adición o retirada de secuencias de poli-T o poli-A, adición de secuencias alrededor del sitio de inicio de la transcripción que potencian la transcripción, tal como secuencias consenso de Kozak, retirada de secuencias que podrían formar estructuras de tallo-lazo, retirada de secuencias desestabilizantes, y dos o más combinaciones de los mismos.
- 5
- 10 La presente invención se refiere a secuencias del factor VIII optimizado, vectores y células hospedadoras que comprenden secuencias de factor VIII optimizado, y métodos para producir polipéptidos codificados por secuencias de factor VIII optimizado. La presente invención también se refiere a métodos de tratamiento de trastornos hemorrágicos tales como hemofilia que comprenden administrar al sujeto una secuencia de ácidos nucleicos de factor VIII optimizado. La presente invención cumple una necesidad importante en la técnica proporcionando secuencias de
- 15 factor VIII optimizado que demuestran un aumento de la expresión en células hospedadoras, rendimiento mejorado de la proteína factor VIII en métodos para producir factor VIII recombinante, y dan posiblemente como resultado mayor eficacia terapéutica cuando se usa en métodos de terapia génica.
- En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de factor VIII (FVIII), en donde la secuencia de
- 20 nucleótidos es al menos 90 % idéntica a SEQ ID NO: 1. En otras realizaciones, la secuencia de nucleótidos es al menos 91 %, al menos 92 %, al menos 93 %, al menos 94 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 % o al menos 99 % idéntica a SEQ ID NO: 1 y codifica un polipéptido con actividad de FVIII. En otras realizaciones más, la secuencia de nucleótidos comprende SEQ ID NO: 1.
- En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos
- 25 es al menos 95 % idéntica a SEQ ID NO: 2. En otras realizaciones, la secuencia de nucleótidos es al menos 96 %, al menos 97 %, al menos 98 %, o al menos 99 % idéntica a SEQ ID NO: 2 y codifica un polipéptido con actividad de FVIII. En otras realizaciones más, la secuencia de nucleótidos comprende SEQ ID NO: 2.
- SEQ ID NO: 1 y 2 son versiones optimizadas de SEQ ID NO: 3, la secuencia de nucleótidos de FVIII inicial o "parental". SEQ ID NO: 3 codifica un FVIII humano de dominio B deleciónado. Aunque SEQ ID NO: 1 y 2 derivan de una forma
- 30 específica de dominio B deleciónado de FVIII (SEQ ID NO: 3), se debe entender que la presente invención también se refiere a versiones optimizadas de ácidos nucleicos que codifican otras versiones de FVIII. Por ejemplo, otra versión de FVIII puede incluir FVIII de longitud completa, otras deleciones del dominio B de FVIII (descritas a continuación), u otros fragmentos de FVIII que retienen actividad de FVIII.
- "Un polipéptido con actividad de FVIII" como se usa en el presente documento significa un polipéptido FVIII funcional en su función normal en la coagulación, a menos que se especifique de otro modo. El término un polipéptido con actividad de FVIII incluye un fragmento funcional, variante, análogo o derivado del mismo que retiene la función de
- 35 factor VIII natural de longitud completa en la vía de coagulación. "Un polipéptido con actividad de FVIII" se usa indistintamente con proteína FVIII, polipéptido FVIII o FVIII. Los ejemplos de funciones de FVIII incluyen, pero no se limitan a, una capacidad para activar la coagulación, una capacidad para actuar de cofactor para el factor IX, o una capacidad para formar un complejo de tenasa con el factor IX en presencia de Ca²⁺ y fosfolípidos, que luego convierte el factor X en la forma activada Xa. En una realización, un polipéptido que tiene actividad de FVIII comprende dos cadenas de polipéptidos, la primera cadena que tiene la cadena pesada de FVIII y la segunda cadena que tiene la
- 40 cadena ligera de FVIII. En otra realización, el polipéptido que tiene actividad de FVIII es FVIII de una sola cadena. FVIII de una sola cadena puede contener una o más mutaciones o sustituciones en el resto de aminoácido 1645 y/o 1648 correspondientes a la secuencia de FVIII maduro. Véase la solicitud internacional N° PCT/US2012/045784. La proteína FVIII puede ser la proteína FVIII humana, porcina, canina, de rata o murina. Además, las comparaciones entre FVIII de seres humanos y otras especies han identificado restos conservados que es probable que se requieran para la función (Cameron et al., *Thromb. Haemost.* 79:317-22 (1998); documento de patente US 6.251.632).
- El "dominio B" de FVIII, como se usa en el presente documento, es el mismo que el dominio B conocido en la técnica que se define por identidad interna de secuencia de aminoácidos y sitios de escisión proteolítica por trombina, por ejemplo, los restos Ser741-Arg1648 de FVIII humano de longitud completa. Los otros dominios de FVIII humano se definen por los siguientes restos de aminoácidos: A1, restos Ala1-Arg372; A2, restos Ser373-Arg740; A3, restos Ser1690-Ile2032; C1, restos Arg2033-Asn2172; C2, restos Ser2173-Tyr2332. La secuencia de A3-C1-C2 incluye los
- 55 restos Ser1690-Tyr2332. La secuencia restante, los restos Glu1649-Arg1689, se denomina normalmente el péptido de activación de la cadena ligera de FVIII. También se conocen en la técnica las localizaciones de los límites para todos los dominios, que incluyen los dominios B, para FVIII porcino, de ratón y canino. Un ejemplo de un FVIII BDD es el FVIII BDD recombinante REFACTO® (Wyeth Pharmaceuticals, Inc.).
- Un "FVIII de dominio B deleciónado" puede tener las deleciones completas o parciales desveladas en las patentes de
- 60 EE. UU. N° 6.316.226, 6.346.513, 7.041.635, 5.789.203, 6.060.447, 5.595.886, 6.228.620, 5.972.885, 6.048.720,

5.543.502, 5.610.278, 5.171.844, 5.112.950, 4.868.112 y 6.458.563. En algunas realizaciones, una secuencia de FVIII de dominio B deleciónado de la presente invención comprende una cualquiera de las deleciones desveladas en la col. 4, línea 4 a col. 5, línea 28 y los Ejemplos 1-5 de la patente de EE. UU. N° 6.316.226 (también en el documento de patente US 6.346.513). En algunas realizaciones, un FVIII de dominio B deleciónado de la presente invención tiene una deleción desvelada en la col. 2, líneas 26-51 y Ejemplos 5-8 de la patente de EE. UU. N° 5.789.203 (también los documentos de patente US 6.060.447, US 5.595.886 y US 6.228.620). En algunas realizaciones, un FVIII de dominio B deleciónado tiene una deleción descrita en la col. 1, líneas 25 a col. 2, línea 40 de la patente de EE. UU. N° 5.972.885; col. 6, líneas 1-22 y Ejemplo 1 de la patente de EE. UU. N° 6.048.720; col. 2, líneas 17-46 de la patente de EE. UU. N° 5.543.502; col. 4, línea 22 a col. 5, línea 36 de la patente de EE. UU. N° 5.171.844; col. 2, líneas 55-68, Figura 2, y el Ejemplo 1 de la patente de EE. UU. N° 5.112.950; col. 2, línea 2 a col. 19, línea 21 y Tabla 2 de la patente de EE. UU. N° 4.868.112; col. 2, línea 1 a col. 3, línea 19, col. 3, línea 40 a col. 4, línea 67, col. 7, línea 43 a col. 8, línea 26, y col. 11, línea 5 a col. 13, línea 39 de la patente de EE. UU. N° 7.041.635; o col. 4, líneas 25-53, de la patente de EE. UU. N° 6.458.563. En algunas realizaciones, un FVIII de dominio B deleciónado tiene una deleción de la mayor parte del dominio B, pero aún contiene secuencias de extremo amino del dominio B que son esenciales para el procesamiento proteolítico *in vivo* del producto de traducción primaria en dos cadenas de polipéptido, como se desvela en documento de patente WO 91/9122. En algunas realizaciones, un FVIII de dominio B deleciónado se construye con una deleción de aminoácidos 747-1638, es decir, prácticamente una deleción completa del dominio B. Hoeben R.C., et al. J. Biol. Chem. 265 (13): 7318-7323 (1990). Un FVIII de dominio B deleciónado también puede contener una deleción de aminoácidos 771-1666 o aminoácidos 868-1562 de FVIII. Meulien P., et al. Protein Eng. 2(4): 301-6 (1988). Las deleciones del dominio B adicional que son parte de la invención incluyen, por ejemplo: deleción de los aminoácidos 982 a 1562 o 760 a 1639 (Toole et al., Proc. Natl. Acad. Sci. U.S.A. (1986) 83, 5939-5942), 797 a 1562 (Eaton, et al. Biochemistry (1986) 25:8343-8347), 741 a 1646 (Kaufman (solicitud PCT publicada N° WO 87/04187)), 747-1560 (Sarver, et al., DNA (1987) 6:553-564), 741 a 1648 (Pasek (solicitud PCT N° 88/00831)), 816 a 1598 o 741 a 1689 (Lagner (Behring Inst. Mitt. (1988) No 82:16-25, documento de patente EP 295597)). Cada una de las deleciones anteriores se puede hacer en cualquier secuencia de FVIII.

Se desvelan varias moléculas de FVIII funcionales, que incluyen deleciones del dominio B, en las siguientes patentes US 6.316.226 y US 6.346.513, ambas cedidas a Baxter; US 7.041.635 cedidas a In2Gen; US 5.789.203, US 6.060.447, US 5.595.886 y US 6.228.620 cedidas a Chiron; US 5.972.885 y US 6.048.720 cedidas a Biovitrum, US 5.543.502 y US 5.610.278 cedidas a Novo Nordisk; US 5.171.844 cedidas a Immuno Ag; US 5.112.950 cedidas a Transgene S.A.; US 4.868.112 cedidas a Genetics Institute.

Optimización por codones

En una realización, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de ácidos nucleicos se ha optimizado por codones. En otra realización, la secuencia de ácidos nucleicos inicial que codifica un polipéptido con actividad de FVIII y que se somete a la optimización por codones es SEQ ID NO: 3. En algunas realizaciones, la secuencia que codifica un polipéptido con actividad de FVIII se optimiza por codones para la expresión humana. En otras realizaciones, la secuencia que codifica un polipéptido con actividad de FVIII se optimiza por codones para la expresión murina. SEQ ID NO: 1 y 2 son versiones optimizadas por codones de SEQ ID NO: 3, optimizadas para expresión humana.

El término "optimizado por codones", como se refiere a genes o regiones codificantes de moléculas de ácidos nucleicos para la transformación de diversos hospedadores, se refiere a la alteración de codones en el gen o las regiones codificantes de las moléculas de ácidos nucleicos para reflejar el uso típico de codones del organismo hospedador sin alterar el polipéptido codificado por el ADN. Dicha optimización incluye sustituir al menos uno, o más de uno, o un número significativo, de codones con uno o más codones que se usan más frecuentemente en los genes de ese organismo.

Las desviaciones en la secuencia de nucleótidos que comprenden los codones que codifican los aminoácidos de cualquier cadena de polipéptidos permiten variaciones en la secuencia que codifica el gen. Puesto que cada codón consiste en tres nucleótidos, y los nucleótidos que comprenden el ADN están restringidos a cuatro bases específicas, existen 64 combinaciones posibles de nucleótidos, 61 de las cuales codifican aminoácidos (los tres codones restantes codifican señales que terminan la traducción). El "código genético" que muestra qué codones codifican qué aminoácidos se reproduce en el presente documento como la Tabla 1. Como resultado, muchos aminoácidos se designan por más de un codón. Por ejemplo, los aminoácidos alanina y prolina están codificados por cuatro tripletes, serina y arginina por seis, mientras que triptófano y metionina están codificados por solo un triplete. Esta degeneración permite que la composición de bases de ADN varíe en un amplio intervalo sin alterar la secuencia de aminoácidos de las proteínas codificadas por el ADN.

Tabla 1. El código genético estándar

	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC "	TCC "	TAC "	TGC "
	TTA Leu (L)	TCA "	TAA Parada	TGA Parada
	TTG "	TCG "	TAG Parada	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC "	CCC "	CAC "	CGC "
	CTA "	CCA "	CAA Gln (C)	CGA "
	CTG "	CCG "	CAG "	CGG "
A	ATT Ile (I)			
	ATC "	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
		ACC "	AAC "	AGC "
	ATA "	ACA "	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG "	AAG "	AGG "
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC "	GCC "	GAC "	GGC "
	GTA "	GCA "	GAA Glu (E)	GGA "
	GTG "	GCG "	GAG "	GGG "

5 Muchos organismos muestran un sesgo para que el uso de codones particulares codifique la inserción de un aminoácido particular en una cadena de péptido en crecimiento. Se proporciona preferencia de codones, o preferencia codónica, diferencia en el uso de codones entre organismos, por la degeneración del código genético, y está bien documentada entre muchos organismos. La preferencia codónica se correlaciona frecuentemente con la eficiencia de traducción de ARN mensajero (ARNm), que se cree a su vez que es dependiente, entre otros, de las propiedades de los codones que se traducen y la disponibilidad de las moléculas de ARN de transferencia (ARNt) particular. La predominancia de ARNt seleccionados en una célula es, en general, una reflexión de los codones más frecuentemente usados en la síntesis de péptidos. Por consiguiente, los genes se pueden adaptar para la expresión génica óptima en un organismo dado basándose en la optimización por codones.

15 Dado el gran número de secuencias de genes disponible para una amplia variedad de especies animales, vegetales y microbianas, se han calculado las frecuencias relativas de uso de codones. Las tablas de uso de codones están disponibles, por ejemplo, en "Codon Usage Database" disponible en www.kazusa.or.jp/codon/ (visitado el 18 de junio de 2012). Véase Nakamura, Y., et al. Nucl. Acids Res. 28:292 (2000).

Se puede hacer manualmente la asignación aleatoria de codones a una frecuencia optimizada para codificar una secuencia de polipéptidos dada calculando frecuencias de codones para cada aminoácido, y luego asignando los codones a la secuencia de polipéptidos aleatoriamente. Además, se pueden usar diversos algoritmos y programas informáticos de software para calcular una secuencia óptima.

20 En una realización, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 % idéntica a SEQ ID NO: 1, y en donde el índice de adaptación de codones humanos es elevado con respecto a SEQ ID NO: 3. Por ejemplo, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 puede tener un índice de adaptación de codones humanos que es al menos aproximadamente 0,75, al menos aproximadamente 0,76, al menos aproximadamente 0,77, al menos aproximadamente 0,78, al menos aproximadamente 0,79, al menos aproximadamente 0,80, al menos aproximadamente 0,81, al menos aproximadamente 0,82, al menos aproximadamente 0,83, al menos aproximadamente 0,84, al menos aproximadamente 0,85, al menos aproximadamente 0,86, al menos

aproximadamente 0,87, al menos aproximadamente 0,88, al menos aproximadamente 0,89, o al menos aproximadamente 0,90.

5 En otra realización, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de factor VIII (FVIII), en donde la secuencia de nucleótidos es al menos 95 % idéntica a SEQ ID NO: 2, y en donde el índice de adaptación de codones humanos es elevado con respecto a SEQ ID NO: 3. Por ejemplo, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 2 pueden tener un índice de adaptación de codones humanos que es al menos aproximadamente 0,75, al menos aproximadamente 0,76, al menos aproximadamente 0,77, al menos aproximadamente 0,78, al menos aproximadamente 0,79, al menos aproximadamente 0,80, al menos aproximadamente 0,81, al menos aproximadamente 0,82, al menos aproximadamente 0,83, al menos aproximadamente 0,84, al menos aproximadamente 0,85, al menos aproximadamente 0,86, al menos aproximadamente 0,87, al menos aproximadamente 0,88, al menos aproximadamente 0,89, o al menos aproximadamente 0,90.

15 En otras realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de factor VIII, en donde la secuencia de nucleótidos es al menos 85 % idéntica a SEQ ID NO: 1 y tiene una o más de las siguientes características: (1) la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con SEQ ID NO: 3, (2) la secuencia de nucleótidos contiene menos secuencias MARS/ARS en comparación con SEQ ID NO: 3, (3) la secuencia de nucleótidos no contiene el sitio de corte y empalme GGTGAT, (4) la secuencia de nucleótidos contiene menos elementos desestabilizantes, (5) la secuencia de nucleótidos no contiene una secuencia de poli-T, (6) la secuencia de nucleótidos no contiene una secuencia de poli-A, (7) la secuencia de nucleótidos tiene un índice de adaptación de codones que es elevado con respecto a SEQ ID NO: 3, o una combinación de dos o más de dichas características. En una realización particular, la secuencia de nucleótidos contiene todas las características (1) a (6).

25 En otras realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de factor VIII, en donde la secuencia de nucleótidos es al menos 95 % idéntica a SEQ ID NO: 2 y tiene una o más de las siguientes características: (1) la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con SEQ ID NO: 3, (2) la secuencia de nucleótidos contiene menos secuencias MARS/ARS, (3) la secuencia de nucleótidos no contiene el sitio de corte y empalme GGTGAT, (4) la secuencia de nucleótidos contiene menos elementos desestabilizantes, (5) la secuencia de nucleótidos no contiene una secuencia de poli-T, (6) la secuencia de nucleótidos no contiene una secuencia de poli-A, (7) la secuencia de nucleótidos tiene un índice de adaptación de codones que es elevado con respecto a SEQ ID NO: 3, o una combinación de dos o más de dichas características. En una realización particular, la secuencia de nucleótidos contiene todas las características (1) a (6).

Optimización del contenido de G/C

35 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 tiene un contenido de G/C que es al menos aproximadamente 45 %, al menos aproximadamente 46 %, al menos aproximadamente 47 %, al menos aproximadamente 48 %, al menos aproximadamente 49 %, al menos aproximadamente 50 %, al menos aproximadamente 51 %, al menos aproximadamente 52 %, al menos aproximadamente 53 %, al menos aproximadamente 54 %, o al menos aproximadamente 55 %.

45 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 tiene un contenido de G/C que es al menos aproximadamente 45 %, al menos aproximadamente 46 %, al menos aproximadamente 47 %, al menos aproximadamente 48 %, al menos aproximadamente 49 %, al menos aproximadamente 50 %, al menos aproximadamente 51 %, al menos aproximadamente 52 %, al menos aproximadamente 53 %, al menos aproximadamente 54 %, o al menos aproximadamente 55 %.

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

$$\frac{G + C}{A + T + G + C} \times 100 \quad (III)$$

Los genes humanos son altamente heterogéneos en su contenido de G/C, teniendo algunos genes un contenido de G/C de tan solo 20 %, y teniendo otros genes un contenido de G/C de hasta 95 %. En general, los genes ricos en G/C se expresan más altamente. En realidad, se ha demostrado que el aumento del contenido de G/C de un gen puede conducir al aumento de la expresión del gen, debido principalmente a un aumento en la transcripción y mayores niveles de ARNm en estado estacionario. Véase Kudla et al., PLoS Biol., 4(6): e180 (2006).

Secuencias de tipo región de unión a matriz

En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos contiene menos secuencias MARS/ARS con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 contiene como máximo 1 secuencia MARS/ARS. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 no contiene una secuencia MARS/ARS.

En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos contiene menos secuencias MARS/ARS con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 contiene como máximo 1 secuencia MARS/ARS. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 no contiene una secuencia MARS/ARS.

Se han identificado elementos ricos en AT en la secuencia de nucleótidos de FVIII humano que comparten similitud de secuencias con secuencias de replicación autónoma (ARS) de *Saccharomyces cerevisiae* y regiones de unión a la matriz nuclear (MAR). (Fallux et al., Mol. Cell. Biol. 16:4264-4272 (1996). Se ha demostrado que uno de estos elementos se une a factores nucleares *in vitro* y reprime la expresión de un gen indicador de cloranfenicol acetiltransferasa (CAT). Ídem. Se ha supuesto que estas secuencias pueden contribuir a la represión transcripcional del gen FVIII humano. Así, en una realización, todas las secuencias de MAR/ARS se suprimen en el gen FVIII de la presente invención. Existen cuatro secuencias de MAR/ARS ATATTT (SEQ ID NO: 5) y tres secuencias de MAR/ARS AAATAT (SEQ ID NO: 6) en la secuencia de FVIII parental (SEQ ID NO: 3). Todos estos sitios se mutaron para destruir las secuencias de MAR/ARS en las secuencias de FVIII optimizado (SEQ ID NO: 1 y SEQ ID NO: 2). La localización de cada uno de estos elementos, y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 2, a continuación.

ES 2 848 703 T5

Tabla 2: Resumen de cambios a elementos represores

Localización del elemento	Secuencia de FVIII BDD inicial (SEQ ID NO: 3)	Secuencia de FVIII BDD optimizado (SEQ ID NO: 1)	Secuencia de FVIII BDD optimizado (SEQ ID NO: 2)
Secuencias desestabilizantes			
	639 ATTTA	GTTCA	GTTCA
1338	ATTTA	GTTCA	GTTCA
1449	ATTTA	TTTCA	CTTCA
1590	TAAAT	CAAGT	CAAGT
1623	TAAAT	TAAGA	CAAGA
2410	ATTTA	ATCTA	ATCTA
2586	ATTTA	GTTTA	GTTCA
2630	TAAAT	TGAAC	TGAAC
3884	ATTTA	ATCTG	ACCTG
3887	TAAAT	TGAAC	TGAAC
Posibles sitios de unión al promotor			
641	TTATA	TCATT	TCATC
1275	TATAA	TACAA	TACAA
1276	TTATA	CTACA	CTACA
1445	TTATA	TCATT	TCATC
1474	TATAA	TACAA	TACAA
1588	TATAA	TACAA	TACAA
2614	TTATA	CTGTA	CTGTA
2661	TATAA	TATCA	CATTA
3286	TATAA	TACAA	TACAA
3840	TTATA	TTATT	CTACA
Secuencias de tipo unión a la matriz (MARS/ARS)			
1287	ATATTT	GTATCT	GTACCT
1447	ATATTT	ATTTTC	ATCTTC
1577	AAATAT	AAATCT	AGATCT
1585	AAATAT	AAGTAC	AAGTAC
2231	ATATTT	ACATCA	ACATCA
3054	AAATAT	AAACAT	GAACAT
3788	ATATTT	ACATTT	ACATCT
Elementos de secuencia ricos en AU (ARE)			
2468	ATTTTATT	ACTTTATT	ACTTCATT
3790	ATTTTTAA	ATTTTCAA	ATCTTCAA

Localización del elemento	Secuencia de FVIII BDD inicial (SEQ ID NO: 3)	Secuencia de FVIII BDD optimizado (SEQ ID NO: 1)	Secuencia de FVIII BDD optimizado (SEQ ID NO: 2)
Secuencias de poli A/poli T			
3273	AAAAAAA	GAAGAAA	GAAGAAA
4195	TTTTTT	TTCTTT	TTCTTT
Sitios de corte y empalme			
2203	GGTGAT	GGGGAC	GGCGAC

Secuencias desestabilizantes

5 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos contiene menos elementos desestabilizantes con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 elementos desestabilizantes. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elemento desestabilizante. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 % idéntica a SEQ ID NO: 1 no contiene un elemento desestabilizante.

15 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos contiene menos elementos desestabilizantes con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 elementos desestabilizantes. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elemento desestabilizante. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 % idéntica a SEQ ID NO: 2 no contiene un elemento desestabilizante.

25 Existen diez elementos desestabilizantes en la secuencia de FVIII parental (SEQ ID NO: 3); seis secuencias ATTTA (SEQ ID NO: 8) y cuatro secuencias TAAAT (SEQ ID NO: 9). En una realización, las secuencias de estos sitios se mutaron para destruir los elementos desestabilizantes en SEQ ID NO: 1 y SEQ ID NO: 2 de FVIII optimizado. La localización de cada uno de estos elementos, y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas, se muestran en la Tabla 2.

Posibles sitios de unión al promotor

30 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos contiene menos posibles sitios de unión al promotor con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1 contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 posibles sitios de unión al promotor. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1 contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 posible sitio de unión al promotor. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1 no contiene un posible sitio de unión al promotor.

45 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos contiene menos posibles sitios de unión al promotor con respecto a SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII

5 y que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2 contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6, o como máximo 5 posibles sitios de unión al promotor. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2 contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 posible sitio de unión al promotor. En aún otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII y que es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2 no contiene un posible sitio de unión al promotor.

10 Las cajas TATA son secuencias reguladoras encontradas frecuentemente en las regiones promotoras de eucariotas. Sirven de sitio de unión de la proteína de unión a TATA (TBP), un factor de transcripción general. Las cajas TATA comprenden normalmente la secuencia TATAAA (SEQ ID NO: 12) o una variante cercana. Las cajas TATA dentro de una secuencia codificante, sin embargo, pueden inhibir la traducción de proteína de longitud completa. Existen diez posibles secuencias de unión al promotor en la secuencia natural de FVIII BDD (SEQ ID NO: 3); cinco secuencias TATAA (SEQ ID NO: 12) y cinco secuencias TTATA (SEQ ID NO: 13). En una realización, todos los sitios de unión al promotor se suprimen en los genes FVIII de la presente invención. La localización de cada posible sitio de unión al promotor y la secuencia de nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 2.

Otros elementos reguladores negativos que actúan en cis

20 Además de las secuencias de MAR/ARS, elementos desestabilizantes y posibles sitios de promotor descritos anteriormente, se pueden identificar varias secuencias posiblemente inhibitorias adicionales en la secuencia de FVIII BDD natural (SEQ ID NO: 3). Se pueden identificar dos elementos de secuencia ricos en AU (ARE) (SEQ ID NO: 14 y 15), junto con un sitio de poli-A (SEQ ID NO: 11), un sitio de poli-T (SEQ ID NO: 10) y un sitio de corte y empalme (SEQ ID NO: 7) en la secuencia de FVIII BDD natural. Uno o más de estos elementos se pueden retirar de las secuencias de FVIII optimizado. La localización de cada uno de estos sitios y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 2.

25 En ciertas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, en donde la secuencia de nucleótidos no contiene uno o más elementos reguladores negativos que actúan en cis, por ejemplo, un sitio de corte y empalme, una secuencia de poli-T, una secuencia de poli-A, una secuencia ARE, o cualquier combinación de los mismos.

30 En ciertas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, en donde la secuencia de nucleótidos no contiene uno o más elementos reguladores negativos que actúan en cis, por ejemplo, un sitio de corte y empalme, una secuencia de poli-T, una secuencia de poli-A, una secuencia ARE, o cualquier combinación de los mismos.

35 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 7). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos no contiene una secuencia de poli-T (SEQ ID NO: 10). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos no contiene una secuencia de poli-A (SEQ ID NO: 11). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 90 %, al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 1, y en donde la secuencia de nucleótidos no contiene un elemento ARE (SEQ ID NO: 14 o SEQ ID NO: 15).

40 En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 7). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, y

5 en donde la secuencia de nucleótidos no contiene una secuencia de poli-T (SEQ ID NO: 10). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos no contiene una secuencia de poli-A (SEQ ID NO: 11). En algunas realizaciones, la presente invención proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos es al menos 95 %, al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o 100 % idéntica a SEQ ID NO: 2, y en donde la secuencia de nucleótidos no contiene un elemento ARE (SEQ ID NO: 14 o SEQ ID NO: 15).

10 En otras realizaciones, una secuencia de FVIII optimizado de la invención no comprende uno o más de motivos antivirales, estructuras de tallo-lazo y secuencias repetidas.

15 En otras realizaciones más, los nucleótidos que rodean el sitio de iniciación de la transcripción se cambian a una secuencia consenso de Kozak (GCCGCCACCATGC, en donde los nucleótidos subrayados son el codón de iniciación; SEQ ID NO: 16). En otras realizaciones, se pueden añadir o quitar sitios de restricción para facilitar el proceso de clonación.

Secuencias heterólogas de nucleótidos

20 En algunas realizaciones, las moléculas de ácido nucleico aisladas de la invención comprenden además una secuencia heteróloga de nucleótidos. En algunas realizaciones, las moléculas de ácido nucleico aisladas de la invención comprenden además al menos una secuencia heteróloga de nucleótidos. La secuencia heteróloga de nucleótidos se puede unir con las secuencias de nucleótidos de FVIII BDD optimizado de la invención en el extremo 5', en el extremo 3', o se inserta en el centro de la secuencia de nucleótidos de FVIII BDD optimizado. Así, en algunas realizaciones, la secuencia heteróloga de aminoácidos codificada por la secuencia heteróloga de nucleótidos se une al extremo N o al extremo C de la secuencia de aminoácidos de FVIII codificada por la secuencia de nucleótidos o se inserta entre dos aminoácidos en la secuencia de aminoácidos de FVIII. En otras realizaciones, las moléculas de ácido nucleico aisladas de la invención comprenden además dos, tres, cuatro, cinco, seis, siete u ocho secuencias heterólogas de nucleótidos. En algunas realizaciones, todas las secuencias heterólogas de nucleótidos son idénticas. En algunas realizaciones, al menos una secuencia heteróloga de nucleótidos es diferente de las otras secuencias heterólogas de nucleótidos. En algunas realizaciones, la invención puede comprender dos, tres, cuatro, cinco, seis o más de siete secuencias heterólogas de nucleótidos en tándem.

30 En algunas realizaciones, la secuencia heteróloga de nucleótidos codifica una secuencia de aminoácidos. En algunas realizaciones, la secuencia de aminoácidos codificada por la secuencia heteróloga de nucleótidos es un resto heterólogo que puede aumentar la semivida (un "extensor de la semivida") de una molécula de FVIII.

35 En algunas realizaciones, el resto heterólogo es un péptido o un polipéptido con características o sin estructurar o estructuradas que están asociadas con la prolongación de la semivida *in vivo* cuando se incorpora en una proteína de la invención. Los ejemplos no limitantes incluyen albúmina, fragmentos de albúmina, fragmentos Fc de inmunoglobulinas, el péptido del extremo C (CTP) de la subunidad β de gonadotropina coriónica humana, una secuencia HAP, una secuencia XTEN, una transferrina o un fragmento de la misma, un polipéptido PAS, conectores de poliglicina, conectores de poliserina, restos de unión a albúmina, o cualquier fragmento, derivado, variante o combinaciones de estos polipéptidos. En algunos aspectos, un resto heterólogo incluye factor de von Willebrand o un fragmento del mismo. En otros aspectos relacionados, un resto heterólogo puede incluir un sitio de unión (por ejemplo, un aminoácido de cisteína) para un resto no polipeptídico tal como polietilenglicol (PEG), hidroxietilalmidón (HES), ácido polisialílico, o cualquier derivado, variante o combinaciones de estos elementos. En algunos aspectos, un resto heterólogo comprende un aminoácido de cisteína que actúa como un sitio de unión para un resto no polipeptídico tal como polietilenglicol (PEG), hidroxietilalmidón (HES), ácido polisialílico, o cualquier derivado, variante o combinaciones de estos elementos.

45 En una realización específica, un primera secuencia heteróloga de nucleótidos codifica un primer resto heterólogo que es una molécula extensora de la semivida que se conoce en la técnica, y una segunda secuencia heteróloga de nucleótidos codifica un segundo resto heterólogo que también puede ser una molécula extensora de la semivida que se conoce en la técnica. En ciertas realizaciones, el primer resto heterólogo (por ejemplo, un primer resto Fc) y el segundo resto heterólogo (por ejemplo, un segundo resto Fc) se asocian entre sí para formar un dímero. En una realización, el segundo resto heterólogo es un segundo resto Fc, en donde el segundo resto Fc se une a o se asocia con el primer resto heterólogo, por ejemplo, el primer resto Fc. Por ejemplo, el segundo resto heterólogo (por ejemplo, el segundo resto Fc) se pueden unir al primer resto heterólogo (por ejemplo, el primer resto Fc) por un conector o se asocia con el primer resto heterólogo por un enlace covalente o no covalente.

55 En algunas realizaciones, el resto heterólogo es un polipéptido que comprende, que consiste esencialmente en, o que consiste en al menos aproximadamente 10, al menos aproximadamente 100, al menos aproximadamente 200, al menos aproximadamente 300, al menos aproximadamente 400, al menos aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente 700, al menos aproximadamente 800, al menos aproximadamente 900, al menos aproximadamente 1000, al menos aproximadamente 1100, al menos aproximadamente 1200, al menos

aproximadamente 1300, al menos aproximadamente 1400, al menos aproximadamente 1500, al menos aproximadamente 1600, al menos aproximadamente 1700, al menos aproximadamente 1800, al menos aproximadamente 1900, al menos aproximadamente 2000, al menos aproximadamente 2500, al menos aproximadamente 3000, o al menos aproximadamente 4000 aminoácidos. En otras realizaciones, el resto heterólogo es un polipéptido que comprende, que consiste esencialmente en, o que consiste en aproximadamente 100 a aproximadamente 200 aminoácidos, aproximadamente 200 a aproximadamente 300 aminoácidos, aproximadamente 300 a aproximadamente 400 aminoácidos, aproximadamente 400 a aproximadamente 500 aminoácidos, aproximadamente 500 a aproximadamente 600 aminoácidos, aproximadamente 600 a aproximadamente 700 aminoácidos, aproximadamente 700 a aproximadamente 800 aminoácidos, aproximadamente 800 a aproximadamente 900 aminoácidos, o aproximadamente 900 a aproximadamente 1000 aminoácidos.

En ciertas realizaciones, un resto heterólogo mejora una o más propiedades farmacocinéticas de la proteína FVIII sin afectar significativamente su actividad biológica o función.

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

En otras realizaciones, un resto heterólogo aumenta la estabilidad de la proteína FVIII o un fragmento de la misma (por ejemplo, un fragmento que comprende un resto heterólogo después de la escisión proteolítica de la proteína FVIII). Como se usa en el presente documento, el término "estabilidad" se refiere a una medida conocida en la técnica del mantenimiento de una o más propiedades físicas de la proteína FVIII en respuesta a una condición medioambiental (por ejemplo, una temperatura elevada o reducida). En ciertos aspectos, la propiedad física puede ser el mantenimiento de la estructura covalente de la proteína FVIII (por ejemplo, la ausencia de escisión proteolítica, oxidación no deseada o desamidación). En otros aspectos, la propiedad física también puede ser la presencia de la proteína FVIII en un estado apropiadamente plegado (por ejemplo, la ausencia de agregados o precipitados solubles o insolubles). En un aspecto, la estabilidad de la proteína FVIII se mide ensayando un propiedad biofísica de la proteína FVIII, por ejemplo la estabilidad térmica, perfil de desarrollo del pH, retirada estable de la glucosilación, solubilidad, función bioquímica (por ejemplo, capacidad para unirse a una proteína, receptor o ligando), etc., y/o combinaciones de los mismos. En otro aspecto, la función bioquímica se demuestra por la afinidad de unión de la interacción. En un aspecto, una medida de la estabilidad de la proteína es la estabilidad térmica, es decir, la resistencia a la exposición térmica. La estabilidad se puede medir usando métodos conocidos en la técnica, tales como, HPLC (cromatografía líquida de alta resolución), SEC (cromatografía de exclusión por tamaño), DLS (dispersión dinámica de luz), etc. Los métodos para medir la estabilidad térmica incluyen, pero no se limitan a, calorimetría diferencial de barrido (DSC), fluorimetría diferencial de barrido (DSF), dicroísmo circular (CD) y ensayo de exposición térmica.

En ciertos aspectos, una proteína FVIII comprende al menos un extensor de la semivida, es decir, un resto heterólogo que aumenta la semivida *in vivo* de la proteína FVIII con respecto a la semivida *in vivo* de la proteína FVIII correspondiente que carece de dicho resto heterólogo. La semivida *in vivo* de una proteína FVIII se puede determinar por cualquier método conocido por los expertos en la técnica, por ejemplo, ensayos de actividad (ensayo cromogénico o ensayo de aPTT por coagulación de una etapa), ELISA, ROTEM™, etc.

En algunas realizaciones, la presencia de uno o más extensores de la semivida da como resultado que aumente la semivida de la proteína FVIII en comparación con la semivida de la proteína correspondiente que carece de dichos uno o más extensores de la semivida. La semivida de la proteína FVIII que comprende un extensor de la semivida es al menos aproximadamente 1,5 veces, al menos aproximadamente 2 veces, al menos aproximadamente 2,5 veces, al menos aproximadamente 3 veces, al menos aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos aproximadamente 10 veces, al menos aproximadamente 11 veces, o al menos aproximadamente 12 veces más larga que la semivida *in vivo* de la proteína FVIII correspondiente que carece de dicho extensor de la semivida.

En una realización, la semivida de la proteína FVIII que comprende un extensor de la semivida es aproximadamente 1,5 veces a aproximadamente 20 veces, aproximadamente 1,5 veces a aproximadamente 15 veces, o aproximadamente 1,5 veces a aproximadamente 10 veces más larga que la semivida *in vivo* de la proteína correspondiente que carece de dicho extensor de la semivida. En otra realización, la semivida de la proteína FVIII que comprende un extensor de la semivida se extiende aproximadamente 2 veces a aproximadamente 10 veces, aproximadamente 2 veces a aproximadamente 9 veces, aproximadamente 2 veces a aproximadamente 8 veces, aproximadamente 2 veces a aproximadamente 7 veces, aproximadamente 2 veces a aproximadamente 6 veces, aproximadamente 2 veces a aproximadamente 5 veces, aproximadamente 2 veces a aproximadamente 4 veces, aproximadamente 2 veces a aproximadamente 3 veces, aproximadamente 2,5 veces a aproximadamente 10 veces, aproximadamente 2,5 veces a aproximadamente 9 veces, aproximadamente 2,5 veces a aproximadamente 8 veces, aproximadamente 2,5 veces a aproximadamente 7 veces, aproximadamente 2,5 veces a aproximadamente 6 veces, aproximadamente 2,5 veces a aproximadamente 5 veces, aproximadamente 2,5 veces a aproximadamente 4 veces, aproximadamente 2,5 veces a aproximadamente 3 veces, aproximadamente 3 veces a aproximadamente 10 veces,

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

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

En otras realizaciones más, la semivida de la proteína FVIII que comprende un extensor de la semivida es aproximadamente 15 horas a aproximadamente dos semanas, aproximadamente 16 horas a aproximadamente una semana, aproximadamente 17 horas a aproximadamente una semana, aproximadamente 18 horas a aproximadamente una semana, aproximadamente 19 horas a aproximadamente una semana, aproximadamente 20 horas a aproximadamente una semana, aproximadamente 21 horas a aproximadamente una semana, aproximadamente 22 horas a aproximadamente una semana, aproximadamente 23 horas a aproximadamente una semana, aproximadamente 24 horas a aproximadamente una semana, aproximadamente 36 horas a aproximadamente una semana, aproximadamente 48 horas a aproximadamente una semana, aproximadamente 60 horas a aproximadamente una semana, aproximadamente 24 horas a aproximadamente seis días, aproximadamente 24 horas a aproximadamente cinco días, aproximadamente 24 horas a aproximadamente cuatro días, aproximadamente 24 horas a aproximadamente tres días, o aproximadamente 24 horas a aproximadamente dos días.

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

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

En otro aspecto, un resto heterólogo comprende una o más regiones constantes de inmunoglobulina o porciones de la misma (por ejemplo, una región Fc). En una realización, una molécula de ácido nucleico aislada de la invención comprende además una secuencia de ácidos nucleicos heteróloga que codifica una región constante de inmunoglobulina o una porción de la misma. En algunas realizaciones, la región constante de inmunoglobulina o porción de la misma es una región Fc.

Una región constante de inmunoglobulina comprende dominios indicados dominios CH (pesados constantes) (CH1, CH2, etc.). Dependiendo del isotipo (es decir, IgG, IgM, IgA IgD o IgE), la región constante puede comprender tres o cuatro dominios CH. Las regiones constantes de algunos isotipos (por ejemplo, IgG) también contienen una región bisagra. Véase Janeway et al. 2001, Immunobiology, Garland Publishing, N.Y., N.Y.

Se puede obtener de varias fuentes diferentes una región constante de inmunoglobulina o una porción de la misma para producir la proteína FVIII. En una realización, una región constante de inmunoglobulina o una porción de la misma deriva de una inmunoglobulina humana. Se entiende, sin embargo, que la región constante de inmunoglobulina o una porción de la misma puede derivar de una inmunoglobulina de otra especie de mamífero, que incluye, por ejemplo, una especie de roedor (por ejemplo, un ratón, rata, conejo, cobaya) o primate no humano (por ejemplo, chimpancé, macaco). Además, la región constante de inmunoglobulina o una porción de la misma puede derivar de cualquier clase de inmunoglobulina, que incluye IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de inmunoglobulina, que incluye IgG1, IgG2, IgG3 e IgG4. En una realización, se usa el isotipo humano IgG1.

Están disponibles una variedad de las secuencias de genes de la región constante de inmunoglobulina (por ejemplo, secuencias de genes de la región constante humana) en forma de depósitos públicamente accesibles. Se puede

seleccionar la secuencia de dominios de la región constante que tiene una función efectora particular (o que carece de una función efectora particular) o con una modificación particular para reducir la inmunogenicidad. Se han publicado muchas secuencias de anticuerpos y genes que codifican anticuerpos y se pueden derivar secuencias adecuadas de la región constante de Ig (por ejemplo, secuencias de bisagra, CH2, y/o CH3, o porciones de las mismas) de estas secuencias usando técnicas reconocidas en la técnica. Entonces, el material genético obtenido usando cualquiera de los métodos anteriores se puede alterar o sintetizar para obtener polipéptidos tal como se describe en el presente documento. Se apreciará además que el alcance de la presente invención engloba alelos, variantes y mutaciones de secuencias de ADN de la región constante.

Se pueden clonar las secuencias de la región constante de inmunoglobulina o una porción de la misma, por ejemplo, usando la reacción en cadena de la polimerasa y cebadores que se seleccionan para amplificar el dominio de interés. Para clonar una secuencia de la región constante de inmunoglobulina o una porción de la misma de un anticuerpo, se puede aislar ARNm de hibridoma, bazo o células linfáticas, transcribirse de forma inversa en ADN, y amplificarse los genes de anticuerpo por PCR. La amplificación por métodos de PCR se describe en detalle en las patentes de EE. UU. N° 4.683.195; 4.683.202; 4.800.159; 4.965.188; y en, por ejemplo, "PCR Protocols: A Guide to Methods and Applications" Innis et al. eds., Academic Press, San Diego, CA (1990); Ho et al. 1989. *Gene* 77:51; Horton et al. 1993. *Methods Enzymol.* 217:270). Se puede iniciar PCR por cebadores de la región constante consenso o por cebadores más específicos basándose en las secuencias de aminoácidos publicadas de ADN de la cadena pesada y ligera. También se puede usar PCR para aislar clones de ADN que codifican las cadenas ligeras y pesadas del anticuerpo. En este caso, las bibliotecas se pueden cribar por cebadores consenso o sondas homólogas mayores, tales como sondas de región constante de ratón. Se conocen en la técnica numerosos conjuntos de cebadores adecuados para amplificación de genes de anticuerpo (por ejemplo, cebadores de 5' basados en la secuencia del extremo N de anticuerpos purificados (Benhar y Pastan. 1994. *Protein Engineering* 7:1509); amplificación rápida de extremos de ADNc (Ruberti, F. et al. 1994. *J. Immunol. Methods* 173:33); secuencias conductoras de anticuerpos (Larrick et al. 1989 *Biochem. Biophys. Res. Commun.* 160:1250). La clonación de secuencias de anticuerpos se describe además en Newman et al., patente de EE. UU. N° 5.658.570, presentada el 25 de enero de 1995.

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

Como se usa en el presente documento, el término "región Fc" se define como la porción de un polipéptido que corresponde a la región Fc de Ig nativa, es decir, como se forma por la asociación dimérica de los dominios Fc respectivos de sus dos cadenas pesadas. Una región Fc nativa forma un homodímero con otra región Fc. A diferencia, el término "región Fc genéticamente fusionada" o "región Fc de cadena sencilla" (región scFc), como se usa en el presente documento, se refiere a una región Fc dimérica sintética comprendida de dominios Fc genéticamente unidos dentro de una cadena sencilla de polipéptidos (es decir, codificada en una única secuencia genética contigua). Véase la publicación internacional N° WO 201 2/006635.

En una realización, la "región Fc" se refiere a la porción de una única cadena pesada de la Ig que empieza en la región bisagra justo en la dirección 5' del sitio de escisión por papaína (es decir, resto 216 en IgG, que toma el primer resto de la región constante de la cadena pesada para que sea 114) y que termina en el extremo C del anticuerpo. Por consiguiente, una región Fc completa comprende al menos un dominio bisagra, un dominio CH2 y un dominio CH3.

Una región constante de inmunoglobulina o una porción de la misma puede ser un componente de unión de FcRn. FcRn es activo en tejidos epiteliales adultos y se expresa en la luz de los intestinos, vías respiratorias pulmonares, superficies nasales, superficies vaginales, colon y superficies rectales (patente de EE. UU. N° 6.485.726). Un componente de unión de FcRn es una porción de una inmunoglobulina que se une a FcRn.

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

Los componentes de unión de FcRn útiles en la presente invención engloban moléculas que se pueden unir específicamente por el receptor FcRn que incluyen IgG completa, el fragmento Fc de IgG, y otros fragmentos que incluyen la región de unión completa del receptor FcRn. La región de la porción Fc de IgG que se une al receptor FcRn se ha descrito basándose en cristalografía de rayos X (Burmeister et al. 1994, *Nature* 372:379). La principal área de contacto de Fc con el FcRn está cerca del empalme de los dominios CH2 y CH3. Los contactos Fc-FcRn están todos dentro de una cadena pesada sencilla de Ig. Los componentes de unión de FcRn incluyen IgG completa, el fragmento Fc de IgG, y otros fragmentos de IgG que incluyen la región de unión de FcRn completa. Los principales sitios de

contacto incluyen los restos de aminoácidos 248, 250-257, 272, 285, 288, 290-291, 308-311 y 314 del dominio CH2 y los restos de aminoácidos 385-387, 428, y 433-436 del dominio CH3. Las referencias hechas a la numeración de aminoácidos de inmunoglobulinas o fragmentos de inmunoglobulinas, o regiones, se basan todas en Kabat et al. 1991, Sequences of Proteins of Immunological Interest, U.S. Department of Public Health, Bethesda, Md.

5 Las regiones Fc o componentes de unión de FcRn unidos a FcRn se pueden llevar eficazmente a través de barreras epiteliales por FcRn, proporcionando así un medio no invasivo para administrar por vía sistémica una molécula terapéutica deseada. Además, las proteínas de fusión que comprenden una región Fc o un componente de unión de FcRn son endocitadas por células que expresan el FcRn. Pero en lugar de ser marcadas para degradación, estas proteínas de fusión se recirculan fuera en la circulación otra vez, aumentando así la semivida *in vivo* de estas proteínas.
10 En ciertas realizaciones, las porciones de regiones constantes de inmunoglobulina son una región Fc o un componente de unión de FcRn que normalmente se asocia, mediante enlaces disulfuro y otras interacciones no específicas, con otra región Fc u otro componente de unión de FcRn para formar dímeros y multímeros de orden superior.

Se pueden unir dos receptores de FcRn a una única molécula de Fc. Los datos cristalográficos sugieren que cada molécula de FcRn se une a un único polipéptido del homodímero de Fc. En una realización, la unión del componente de unión de FcRn, por ejemplo, un fragmento Fc de una IgG, a una molécula biológicamente activa proporciona un medio de administración de la molécula biológicamente activa por vía oral, por vía bucal, por vía sublingual, por vía rectal, por vía vaginal, como un aerosol administrado nasalmente o por una vía pulmonar, o por una vía ocular. En otra realización, la proteína FVIII se puede administrar invasivamente, por ejemplo, por vía subcutánea, por vía intravenosa.

Un componente de unión de la región de FcRn es una molécula o porción de la misma que se puede unir específicamente por el receptor de FcRn con el consecuente transporte activo por el receptor de FcRn de la región Fc. Se une específicamente se refiere a dos moléculas que forman un complejo que es relativamente estable en condiciones fisiológicas. La unión específica se caracteriza por una alta afinidad y una capacidad de baja a moderada como se distingue de la unión no específica que normalmente tiene una baja afinidad con una capacidad de moderada a alta. Normalmente, la unión se considera específica cuando la constante de afinidad K_A es superior a $10^6 M^{-1}$, o superior a $10^8 M^{-1}$. Si fuera necesario, se puede reducir la unión no específica sin afectar sustancialmente la unión específica variando las condiciones de unión. Un experto puede optimizar las condiciones de unión apropiadas, tales como concentración de las moléculas, fuerza iónica de la disolución, temperatura, tiempo permitido para la unión, concentración de un agente de bloqueo (por ejemplo, albúmina de suero, caseína de la leche), etc., usando técnicas rutinarias.

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

40 La región Fc se puede obtener de varias fuentes diferentes. En una realización, una región Fc del polipéptido deriva de una inmunoglobulina humana. Se entiende, sin embargo, que un resto Fc puede derivar de una inmunoglobulina de otra especie de mamífero, que incluye, por ejemplo, una especie de roedor (por ejemplo, un ratón, rata, conejo, cobaya) o de primate no humano (por ejemplo, chimpancé, macaco). Además, el polipéptido de los dominios Fc o porciones de los mismos puede derivar de cualquier clase de inmunoglobulina, que incluye IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de inmunoglobulina, que incluye IgG1, IgG2, IgG3 e IgG4. En otra realización, se usa el isotipo humano IgG1.

En ciertas realizaciones, la variante Fc confiere un cambio en al menos una función efectora conferida por un resto Fc que comprende dicho dominio Fc natural (por ejemplo, una mejora o reducción en la capacidad de la región Fc para unirse a receptores de Fc (por ejemplo, FcγRI, FcγRII o FcγRIII) o proteínas del complemento (por ejemplo, C1q), o para desencadenar citotoxicidad dependiente de anticuerpo (ADCC), fagocitosis, o citotoxicidad dependiente del complemento (CDC)). En otras realizaciones, la variante de Fc proporciona un resto de cisteína manipulado.

La región Fc de la invención puede emplear variantes de Fc reconocidas en la técnica que se conocen por conferir un cambio (por ejemplo, un potenciamiento o reducción) en la función efectora y/o unión de FcR o FcRn. Específicamente, una región Fc de la invención puede incluir, por ejemplo, un cambio (por ejemplo, una sustitución) en una o más de las posiciones de aminoácidos desveladas en las publicaciones PCT internacionales WO88/07089A1, WO96/14339A1, WO98/05787A1, WO98/23289A1, WO99/51642A1, WO99/58572A1, WO00/09560A2, WO00/32767A1, WO00/42072A2, WO02/44215A2, WO02/060919A2, WO03/074569A2, WO04/016750A2, WO04/029207A2, WO04/035752A2, WO04/063351A2, WO04/074455A2, WO04/099249A2, WO05/040217A2, WO04/044859, WO05/070963A1, WO05/077981A2, WO05/092925A2, WO05/123780A2, WO06/019447A1, WO06/047350A2, y WO06/085967A2; las publicaciones de patente de EE. UU. N^o US2007/0231329, US2007/0231329, US2007/0237765, US2007/0237766, US2007/0237767, US2007/0243188, US20070248603,

US20070286859, US20080057056; o las patentes de EE. UU. 5.648.260; 5.739.277; 5.834.250; 5.869.046; 6.096.871; 6.121.022; 6.194.551; 6.242.195; 6.277.375; 6.528.624; 6.538.124; 6.737.056; 6.821.505; 6.998.253; 7.083.784; 7.404.956 y 7.317.091. En una realización, se puede hacer el cambio específico (por ejemplo, la sustitución específica de uno o más aminoácidos desvelados en la materia) en una o más de las posiciones de aminoácidos desveladas. En otra realización, se puede hacer un cambio diferente en una o más de las posiciones de aminoácidos desveladas (por ejemplo, la sustitución diferente de una o más posiciones de aminoácidos desvelada en la técnica).

La región Fc o componente de unión de FcRn de IgG se puede modificar según procedimientos bien reconocidos tales como mutagénesis dirigida al sitio y similares para dar IgG modificada o fragmentos Fc, o porciones de los mismos, que se unirán por FcRn. Dichas modificaciones incluyen modificaciones remotas de los sitios de contacto de FcRn, así como modificaciones dentro de los sitios de contacto que preservan o incluso potencian la unión al FcRn. Por ejemplo, se pueden sustituir los siguientes restos de aminoácidos individuales en Fc de IgG1 humana (Fc γ 1) sin pérdida significativa de la afinidad de unión de Fc por FcRn: P238A, S239A, K246A, K248A, D249A, M252A, T256A, E258A, T260A, D265A, S267A, H268A, E269A, D270A, E272A, L274A, N276A, Y278A, D280A, V282A, E283A, H285A, N286A, T289A, K290A, R292A, E293A, E294A, Q295A, Y296F, N297A, S298A, Y300F, R301A, V303A, V305A, T307A, L309A, Q311A, D312A, N315A, K317A, E318A, K320A, K322A, S324A, K326A, A327Q, P329A, A330Q, P331A, E333A, K334A, T335A, S337A, K338A, K340A, Q342A, R344A, E345A, Q347A, R355A, E356A, M358A, T359A, K360A, N361A, Q362A, Y373A, S375A, D376A, A378Q, E380A, E382A, S383A, N384A, Q386A, E388A, N389A, N390A, Y391F, K392A, L398A, S400A, D401A, D413A, K414A, R416A, Q418A, Q419A, N421A, V422A, S424A, E430A, N434A, T437A, Q438A, K439A, S440A, S444A y K447A, donde, por ejemplo, P238A representa prolina natural sustituida con alanina en la posición número 238. Como un ejemplo, una realización específica incorpora la mutación N297A, retirando un sitio de N-glucosilación altamente conservado. Además de alanina, se pueden sustituir otros aminoácidos por los aminoácidos naturales en las posiciones especificadas anteriormente. Las mutaciones se pueden introducir individualmente en Fc, dando lugar a más de cien regiones Fc distintas de Fc nativa.

Ciertas de las mutaciones anteriores pueden conferir nueva funcionalidad a la región Fc o componente de unión de FcRn. Por ejemplo, una realización incorpora N297A, retirando un sitio de N-glucosilación altamente conservado. El efecto de esta mutación es reducir la inmunogenicidad, potenciando así la semivida en circulación de la región Fc, y convirtiendo la región Fc en incapaz de unirse a Fc γ RI, Fc γ RIIA, Fc γ RIIB y Fc γ RIIIA, sin comprometer la afinidad por FcRn (Routledge et al. 1995, *Transplantation* 60:847; Friend et al. 1999, *Transplantation* 68:1632; Shields et al. 1995, *J. Biol. Chem.* 276:6591). Como ejemplo adicional de la nueva funcionalidad que surge de las mutaciones descritas anteriormente, la afinidad por FcRn se puede aumentar más allá de la del natural en algunos casos. Esta elevada afinidad puede afectar una elevada velocidad de "asociación", una reducida velocidad de "disociación", o ambas, una elevada velocidad de "asociación" y una reducida velocidad de "disociación". Los ejemplos de mutaciones que se cree que confieren una elevada afinidad por FcRn incluyen, pero no se limitan a, T256A, T307A, E380A y N434A (Shields et al. 2001, *J. Biol. Chem.* 276:6591).

Además, al menos tres receptores gamma de Fc humanos parecen reconocer un sitio de unión sobre IgG dentro de la región bisagra inferior, en general, los aminoácidos 234-237. Por tanto, otro ejemplo de nueva funcionalidad y posible inmunogenicidad reducida puede surgir de mutaciones de esta región, como, por ejemplo, reemplazando los aminoácidos 233-236 de la IgG1 humana "ELLG" (SEQ ID NO: 29) con la secuencia correspondiente de IgG2 "PVA" (con delección de un aminoácido). Se ha mostrado que Fc γ RI, Fc γ RII y Fc γ RIII, que median en diversas funciones efectoras, no se unirán a IgG1 cuando se hayan introducido dichas mutaciones. Ward y Ghetie 1995, *Therapeutic Immunology* 2:77 y Armour et al. 1999, *Eur. J. Immunol.* 29:2613.

En otra realización, la región constante de inmunoglobulina o una porción de la misma comprende una secuencia de aminoácidos en la región bisagra o una porción de la misma que forma uno o más enlaces disulfuro con una segunda región constante de inmunoglobulina o una porción de la misma. La segunda región constante de inmunoglobulina o una porción de la misma se puede unir a un segundo polipéptido, uniendo la proteína FVIII y el segundo polipéptido. En algunas realizaciones, el segundo polipéptido es un resto potenciador. Como se usa en el presente documento, el término "resto potenciador" se refiere a una molécula, fragmento de la misma o un componente de un polipéptido que es capaz de potenciar la actividad procoagulante de FVIII. El resto potenciador puede ser un cofactor, tal como factor tisular soluble (FTs), o un péptido procoagulante. Así, después de la activación de FVIII, el resto potenciador está disponible para potenciar la actividad de FVIII.

En ciertas realizaciones, una proteína FVIII de la invención comprende una sustitución de aminoácidos en una región constante de inmunoglobulina o una porción de la misma (por ejemplo, variantes Fc), que altera las funciones efectoras independientes de antígeno de la región constante de Ig, en particular la semivida circulante de la proteína.

2. Regiones scFc

En otro aspecto, un resto heterólogo comprende una región scFc (Fc de una sola cadena). En una realización, una molécula de ácido nucleico aislada de la invención comprende además una secuencia de ácidos nucleicos heteróloga que codifica una región scFc. La región scFc comprende al menos dos regiones constantes de inmunoglobulina o porciones de la misma (por ejemplo, restos o dominios Fc (por ejemplo, 2, 3, 4, 5, 6, o más restos o dominios Fc)) dentro de la misma cadena lineal de polipéptidos que son capaces de plegamiento (por ejemplo, plegamiento

intramolecularmente o intermolecularmente) para formar una región scFc funcional que se une por un conector peptídico Fc. Por ejemplo, en una realización, un polipéptido tal como se describe en el presente documento es capaz de unirse, por su región scFc, a al menos un receptor de Fc (por ejemplo un receptor FcRn, FcyR (por ejemplo, FcyRIII), o una proteína del complemento (por ejemplo, C1q)) para mejorar la semivida o desencadenar una función efectora inmunitaria (por ejemplo, citotoxicidad dependiente del anticuerpo (ADCC), fagocitosis o citotoxicidad dependiente del complemento (CDCC) y/o para mejorar la capacidad de fabricación).

3. CTP

En otro aspecto, un resto heterólogo comprende un péptido del extremo C (CTP) de la subunidad β de gonadotropina coriónica humana o fragmento, variante, o derivado de la misma. Se sabe que uno o más péptidos CTP insertados en una proteína recombinante aumentan la semivida *in vivo* de esa proteína. Véase, por ejemplo, la patente de EE. UU. N° 5.712.1:2.

Los péptidos CTP a modo de ejemplo incluyen DPRFQDSSSSKAPPPSLPSPSRLPGSDTPIL (SEQ ID NO: 17) o SSSSKAPPPSLPSPSRLPGSDTPILPQ. (SEQ ID NO: 18). Véase, por ejemplo, la publicación de solicitud de patente de EE. UU. N° US 2009/0087411 A1.

4. Secuencia XTEN

En algunas realizaciones, un resto heterólogo comprende una o más secuencias XTEN, fragmentos, variantes, o derivados de los mismos. Como se usa aquí, "secuencia XTEN" se refiere a polipéptidos de longitud extendida con secuencias sustancialmente no repetitivas que no existen de forma natural, que están compuestas principalmente de aminoácidos hidrófilos pequeños, teniendo la secuencia un bajo grado o ninguna estructura secundaria o terciaria en condiciones fisiológicas. Como resto heterólogo, las XTEN pueden servir de resto de extensión de la semivida. Además, XTEN puede proporcionar propiedades deseables que incluyen, pero no se limitan a, parámetros farmacocinéticos y características de solubilidad potenciados.

La incorporación de un resto heterólogo que comprende una secuencia XTEN en una proteína de la invención puede conferir a la proteína una o más de las siguientes propiedades ventajosas: flexibilidad conformacional, solubilidad acuosa potenciada, alto grado de resistencia a proteasas, baja inmunogenicidad, baja unión a receptores de mamífero, o elevados radios hidrodinámicos (o de Stokes).

En ciertos aspectos, una secuencia XTEN puede aumentar las propiedades farmacocinéticas, tales como mayor semivida *in vivo* o elevada área bajo la curva (ABC), de manera que una proteína de la invención permanezca *in vivo* y tenga actividad procoagulante durante un elevado periodo de tiempo en comparación con una proteína con el mismo resto, pero sin el resto heterólogo XTEN.

Los ejemplos de secuencias XTEN que se pueden usar como restos heterólogos en proteínas quiméricas de la invención se desvelan, por ejemplo, en las publicaciones de patente de EE. UU. N° 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1, o 2011/0172146 A1, o las publicaciones de patente internacional N° WO 2010091122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1 o WO 2011028344 A2.

Las secuencias XTEN a modo de ejemplo que se pueden usar como restos heterólogos en la proteína quimérica de la invención incluyen XTEN AE42-4 (SEQ ID NO: 30, codificada por SEQ ID NO: 31), XTEN 144-2A (SEQ ID NO: 32, codificada por SEQ ID NO: 33), XTEN A144-3B (SEQ ID NO: 34, codificada por SEQ ID NO: 35), XTEN AE144-4A (SEQ ID NO: 36, codificada por SEQ ID NO: 37), XTEN AE144-5A (SEQ ID NO: 38, codificada por SEQ ID NO: 39), XTEN AE144-6B (SEQ ID NO: 40, codificada por SEQ ID NO: 41), XTEN AG144-1 (SEQ ID NO: 42, codificada por SEQ ID NO: 43), XTEN AG144-A (SEQ ID NO: 44, codificada por SEQ ID NO: 45), XTEN AG144-B (SEQ ID NO: 46, codificada por SEQ ID NO: 47), XTEN AG144-C (SEQ ID NO: 48, codificada por SEQ ID NO: 49) y XTEN AG144-F (SEQ ID NO: 50, codificada por SEQ ID NO: 51).

5. Albúmina o fragmento, derivado o variante de la misma

En algunas realizaciones, un resto heterólogo comprende albúmina o un fragmento funcional de la misma. La albúmina de suero humano (HSA, o HA), una proteína de 609 aminoácidos en su forma de longitud completa, es responsable de una significativa proporción de la presión osmótica del suero y también funciona como vehículo de ligandos endógenos y exógenos. El término "albúmina", como se usa en el presente documento, incluye albúmina de longitud completa o un fragmento funcional, variante, derivado o análogo de la misma. Los ejemplos de albúmina o los fragmentos o variantes de la misma se desvelan en las publicaciones de patente de EE. UU. N° 2008/0194481A1, 2008/0004206 A1, 2008/0161243 A1, 2008/0261877 A1 o 2008/0153751 A1, o las publicaciones de solicitud PCT N° 2008/033413 A2, 2009/058322 A1 o 2007/021494 A2.

En una realización, la proteína FVIII de la invención comprende albúmina, un fragmento o una variante de la misma que se une además a un segundo resto heterólogo seleccionado del grupo que consiste en una región constante de inmunoglobulina o porción de la misma (por ejemplo, una región Fc), una secuencia PAS, HES y PEG.

6. Resto de unión a albúmina

En ciertas realizaciones, el resto heterólogo es un resto de unión a albúmina, que comprende un péptido de unión a albúmina, un dominio de unión a albúmina bacteriana, un fragmento de anticuerpo de unión a albúmina, o cualquier combinación de los mismos.

- 5 Por ejemplo, la proteína de unión de albúmina puede ser una proteína de unión de albúmina bacteriana, un anticuerpo o un fragmento de anticuerpo que incluye anticuerpos de dominio (véase la patente de EE. UU. N° 6.696.245). Una proteína de unión de albúmina, por ejemplo, puede ser un dominio de unión de albúmina bacteriana, tal como la de la proteína G estafilocócica (Konig, T. y Skerra, A. (1998) J. Immunol. Methods 218, 73-83). Otros ejemplos de péptidos de unión de albúmina que se pueden usar como componente de conjugación son, por ejemplo, los que tienen una
- 10 secuencia consenso Cys-Xaa₁-Xaa₂-Xaa₃-Xaa₄-Cys, en donde Xaa₁ es Asp, Asn, Ser, Thr, o Trp; Xaa₂ es Asn, Gln, His, Ile, Leu o Lys; Xaa₃ es Ala, Asp, Phe, Trp o Tyr; y Xaa₄ es Asp, Gly, Leu, Phe, Ser o Thr como se describen en la solicitud de patente de EE. UU. 2003/0069395 o Dennis et al. (Dennis et al. (2002) J. Biol. Chem. 277, 35035-35043).

- El dominio 3 de la proteína G estreptocócica, como se desvela por Kraulis et al., FEBS Lett. 378:190-194 (1996) y Linhult et al., Protein Sci. 11:206-213 (2002), es un ejemplo de un dominio de unión a albúmina bacteriana. Los
- 15 ejemplos de péptidos de unión a albúmina incluyen una serie de péptidos que tienen la secuencia central DICLPRWGCLW (SEQ ID NO: 19). Véase, por ejemplo, Dennis et al., J. Biol. Chem. 2002, 277: 35035-35043 (2002). Los ejemplos de fragmentos de unión a albúmina de anticuerpos se desvelan en Muller y Kontermann, Curr. Opin. Mol. Ther. 9:319-326 (2007); Roover et al., Cancer Immunol. Immunother. 56:303-317 (2007), y Holt et al., Prot. Eng. Design Sci., 21:283-288 (2008). Un ejemplo de dicho resto de unión a albúmina es 2-(3-maleimidopropanamido)-6-(4-
- 20 (4-yodofenil)butanamido)hexanoato (marca "Albu") como se desvela por Trussetlet al., Bioconjugate Chem. 20:2286-2292 (2009).

- Se pueden usar ácidos grasos, en particular ácidos grasos de cadena larga (LCFA) y compuestos de unión a albúmina de tipo ácido graso de cadena larga para prolongar la semivida *in vivo* de las proteínas FVIII de la invención. Un
- 25 ejemplo de un compuesto de unión a albúmina de tipo LCFA es ácido 16-(1-(3-(9-(((2,5-dioxopirrolidin-1-ilo)carbonilo)metil)-7-sulfo-9H-fluoren-2-ilamino)-3-oxopropil)-2,5-dioxopirrolidin-3-iltio)hexadecanoico (véase, por ejemplo, el documento de patente WO 2010/140148).

7. Secuencia PAS

- En otras realizaciones, el resto heterólogo es una secuencia de PAS. Una secuencia de PAS, como se usa en el
- 30 presente documento, significa una secuencia de aminoácidos que comprende principalmente restos de alanina y serina, o que comprende principalmente restos de alanina, serina y prolina, formando la secuencia de aminoácidos la conformación de espiral al azar en condiciones fisiológicas. Por consiguiente, la secuencia de PAS es un elemento estructural, un polímero de aminoácido, o un casete de secuencia que comprende, consiste esencialmente en, o que
- 35 consiste en alanina, serina y prolina que se pueden usar como parte del resto heterólogo en la proteína quimérica. Aún, el experto conoce que un polímero de aminoácido también puede formar la conformación de espiral al azar cuando restos distintos de alanina, serina y prolina se añaden como constituyente minoritario en la secuencia de PAS. El término "constituyente minoritario", como se usa en el presente documento, significa que aminoácidos distintos de
- 40 alanina, serina y prolina se pueden añadir a la secuencia de PAS hasta un cierto grado, por ejemplo, hasta aproximadamente 12 %, es decir, aproximadamente 12 de 100 aminoácidos de la secuencia de PAS, hasta aproximadamente 10 %, es decir, aproximadamente 10 de 100 aminoácidos de la secuencia de PAS, hasta aproximadamente 9 %, es decir, aproximadamente 9 de 100 aminoácidos, hasta aproximadamente 8 %, es decir, aproximadamente 8 de 100 aminoácidos, aproximadamente 6 %, es decir, aproximadamente 6 de 100 aminoácidos, aproximadamente 5 %, es decir, aproximadamente 5 de 100 aminoácidos, aproximadamente 4 %, es decir, aproximadamente 4 de 100 aminoácidos, aproximadamente 3 %, es decir, aproximadamente 3 de 100 aminoácidos,
- 45 aproximadamente 2 %, es decir, aproximadamente 2 de 100 aminoácidos, aproximadamente 1 %, es decir, aproximadamente 1 de 100 de los aminoácidos. Los aminoácidos diferentes de alanina, serina y prolina se pueden seleccionar del grupo que consiste en Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr y Val.

- En condiciones fisiológicas, la extensión de secuencia de PAS forma una conformación de espirales al azar y así
- 50 puede mediar en una elevada estabilidad *in vivo* y/o *in vitro* para la proteína FVIII. Puesto que el dominio de espiral al azar no adopta una estructura estable o función por sí mismo, se conserva esencialmente la actividad biológica mediada por la proteína FVIII. En otras realizaciones, las secuencias PAS que forman el dominio de espiral al azar son biológicamente inertes, especialmente con respecto a la proteólisis en plasma sanguíneo, inmunogenicidad, punto isoelectrónico/comportamiento electrostático, unión a receptores de la superficie celular, o internalización, pero son todavía biodegradables, que proporciona claras ventajas con respecto a los polímeros sintéticos tales como PEG.

- Los ejemplos no limitantes de las secuencias PAS que forman la conformación de espiral al azar comprenden una
- 55 secuencia de aminoácidos seleccionada del grupo que consiste en ASPAAPAPASPAAPAPSAPA (SEQ ID NO: 20), AAPASPAPAAPSAPAPAAPS (SEQ ID NO: 21), APSSPSPASPSPPSPASPS (SEQ ID NO: 22),

APSSPSPSAPSSPSPASPS (SEQ ID NO: 23), SPSAPSPSSPASPSPSSPA (SEQ ID NO: 24), AASPAAPSAPAAASPAAPSAPPA (SEQ ID NO: 25) y ASAAAPAAASAAASAPSAAA (SEQ ID NO: 26), o cualquier combinación de las mismas. Los ejemplos adicionales de secuencias PAS se conocen de, por ejemplo, la publicación de patente de EE. UU. N° 2010/0292130 A1 y la publicación de solicitud PCT N° WO 2008/155134 A1.

5 8. Secuencia HAP

En ciertas realizaciones, el resto heterólogo es un polímero de aminoácidos rico en glicina (HAP). La secuencia HAP puede comprender una secuencia repetitiva de glicina, que tiene al menos 50 aminoácidos, al menos 100 aminoácidos, 120 aminoácidos, 140 aminoácidos, 160 aminoácidos, 180 aminoácidos, 200 aminoácidos, 250 aminoácidos, 300 aminoácidos, 350 aminoácidos, 400 aminoácidos, 450 aminoácidos, o 500 aminoácidos de longitud. En una realización, la secuencia de HAP es capaz de prolongar la semivida de un resto fusionado con o unido a la secuencia de HAP. Los ejemplos no limitantes de la secuencia HAP incluyen, pero no se limitan a, $(\text{Gly})_n$, $(\text{Gly}_4\text{Ser})_n$ o $\text{S}(\text{Gly}_4\text{Ser})_n$, en donde n es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20. En una realización, n es 20, 21, 22, 23, 24, 25, 26, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40. En otra realización, n es 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 o 200.

15 9. Transferrina o fragmento de la misma

En ciertas realizaciones, el resto heterólogo es transferrina o un fragmento de la misma. Se puede usar cualquier transferrina para preparar las proteínas FVIII de la invención. Como un ejemplo, la TF humana natural (TF) es una proteína de 679 aminoácidos, de aproximadamente 75 kDa (que no supone glucosilación), con dos dominios principales, N (aproximadamente 330 aminoácidos) y C (aproximadamente 340 aminoácidos), que parece que se originan a partir de una duplicación génica. Véanse los números de acceso de GenBank NM001063, XM002793, M12530, XM039845, XM039847 y S95936 (www.ncbi.nlm.nih.gov). La transferrina comprende dos dominios, dominio N y dominio C. El dominio N comprende dos subdominios, dominio N1 y dominio N2, y el dominio C comprende dos subdominios, dominio C1 y dominio C2.

En una realización, el resto heterólogo de transferrina incluye una variante de corte y empalme de transferrina. En un ejemplo, una variante de corte y empalme de transferrina puede ser una variante de corte y empalme de transferrina humana, por ejemplo, acceso de Genbank AAA61140. En otra realización, la porción de transferrina de la proteína quimérica incluye uno o más dominios de la secuencia de transferrina, por ejemplo, dominio N, dominio C, dominio N1, dominio N2, dominio C1, dominio C2, o cualquier combinación de los mismos.

10. Receptores de la eliminación

En ciertas realizaciones, el resto heterólogo es un receptor de la eliminación, fragmento, variante o derivado del mismo. LRP1 es una proteína de la membrana integral de 600 kDa que participa en la eliminación mediada por receptor de una variedad de proteínas, tales como factor X. Véase, por ejemplo, Narita et al., Blood 91:555-560 (1998).

11. Factor de von Willebrand o fragmentos del mismo

En ciertas realizaciones, el resto heterólogo es factor de von Willebrand (VWF) o fragmentos del mismo.

VWF (también conocido como F8VWF) es una gran proteína multimérica presente en el plasma sanguíneo y producida constitutivamente en endotelio (en los cuerpos de Weibel-Palade), megacariocitos (gránulos α de plaquetas) y tejido conjuntivo subendoteliano. El monómero de VWF básico es una proteína de 2813 aminoácidos. Cada monómero contiene varios dominios específicos con una función específica, los dominios D' y D3 (que juntos se unen al factor VIII), el dominio A1 (que se une al receptor GPIIb de plaquetas, heparina y/o posiblemente colágeno), el dominio A3 (que se une al colágeno), el dominio C1 (en el que el dominio RGD se une a la integrina de plaquetas $\alpha\text{IIb}\beta_3$ cuando esta se activa) y el dominio de "nudo de cisteína" en el extremo C de la proteína (cuyo VWF comparte con el factor de crecimiento derivado de plaquetas (PDGF), factor de crecimiento transformante β (TGF β) y gonadotropina coriónica humana β (βHCG)).

La secuencia de aminoácidos del monómero 2813 para VWF humano se informa como el número de acceso NP000543.2 en Genbank. La secuencia de nucleótidos que codifica el VWF humano se informa como el número de acceso NM000552.3 en Genbank. La secuencia de nucleótidos de VWF humano se designa SEQ ID NO: 27. SEQ ID NO: 28 es la secuencia de aminoácidos codificada por SEQ ID NO: 27. El dominio D' incluye los aminoácidos 764 a 866 de SEQ ID NO: 28. El dominio D3 incluye los aminoácidos 867 a 1240 de SEQ ID NO: 28.

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

En ciertas realizaciones, el resto heterólogo es factor de von Willebrand de longitud completa. En otras realizaciones, el resto heterólogo es un fragmento de factor de von Willebrand Factor. Como se usa en el presente documento, el

término "fragmento de VWF" o "fragmentos de VWF" usado en el presente documento significa cualquier fragmento de VWF que interacciona con FVIII y retiene al menos una o más propiedades que normalmente se proporcionan a FVIII por VWF de longitud completa, por ejemplo, prevención de la activación prematura de FVIIIa, prevención de la proteólisis prematura, prevención de la asociación con membranas de fosfolípido que podrían conducir a eliminación prematura, prevención de la unión a receptores de la eliminación de FVIII que se pueden unir a FVIII desnudo, pero no a FVIII unido a VWF, y/o estabilización de las interacciones de la cadena pesada y la cadena ligera de FVIII. En una realización específica, el resto heterólogo es un fragmento (de VWF) que comprende un dominio D' y un dominio D3 de VWF. El fragmento de VWF que comprende el dominio D' y el dominio D3 puede comprender además un dominio VWF seleccionado del grupo que consiste en un dominio A1, un dominio A2, un dominio A3, un dominio D1, un dominio D2, un dominio D4, un dominio B1, un dominio B2, un dominio B3, un dominio C1, un dominio C2, un dominio CK, uno o más fragmentos de los mismos, y cualquier combinación de los mismos. Los ejemplos adicionales del polipéptido que tiene actividad de FVIII fusionado con el fragmento VWF se desvelan en la solicitud de patente provisional de EE. UU. N° 61/667.901, presentada el 3 de julio de 2012.

12. Restos conectores

En ciertas realizaciones, el resto heterólogo es un conector peptídico.

Como se usa en el presente documento, los términos "conectores peptídicos" o "restos conectores" se refieren a un péptido o secuencia de polipéptidos (por ejemplo, un péptido sintético o secuencia de polipéptidos) que conecta dos dominios en una secuencia de aminoácidos lineal de una cadena de polipéptidos.

En algunas realizaciones, secuencias heterólogas de nucleótidos que codifican conectores peptídicos se pueden insertar entre las secuencias de polinucleótidos de FVIII optimizado de la invención y una secuencia heteróloga de nucleótidos que codifica, por ejemplo, uno de los restos heterólogos descritos anteriormente, tal como albúmina. Los conectores peptídicos pueden proporcionar flexibilidad a la molécula de polipéptido quimérico. Los conectores no se escinden normalmente, sin embargo, dicha escisión puede ser conveniente. En una realización, estos conectores no se retiran durante el procesamiento.

Un tipo de conector que puede estar presentes en una proteína quimérica de la invención es un conector escindible por proteasa que comprende un sitio de escisión (es decir, un sustrato de sitio de escisión por proteasa, por ejemplo, un factor XIa, Xa, o sitio de escisión por trombina) y que puede incluir conectores adicionales en cualquiera del extremo N de o ambos lados del sitio de escisión. Estos conectores escindibles cuando se incorporan en una construcción de la invención dan como resultado una molécula quimérica que tiene un sitio de escisión heterólogo.

En una realización, un polipéptido FVIII comprende dos o más dominios Fc o restos unidos por un conector cscFc para formar una región Fc comprendida en una única cadena de polipéptidos. El conector cscFc está flanqueado por al menos un sitio de procesamiento intracelular, es decir, un sitio escindido por una enzima intracelular. La escisión del polipéptido en el al menos un sitio de procesamiento intracelular da como resultado un polipéptido que comprende al menos dos cadenas de polipéptidos.

Otros conectores peptídicos se pueden usar opcionalmente en una construcción de la invención, por ejemplo, para conectar una proteína FVIII con una región Fc. Algunos conectores a modo de ejemplo que se pueden usar a propósito de la invención incluyen, por ejemplo, polipéptidos que comprenden los aminoácidos GlySer descritos más abajo en más detalle.

En una realización, el conector peptídico es sintético, es decir, no existen de forma natural. En una realización, un conector peptídico incluye péptidos (o polipéptidos) (que pueden o pueden no existir de forma natural) que comprenden una secuencia de aminoácidos que conecta o fusiona genéticamente una primera secuencia de aminoácidos lineal con una segunda secuencia de aminoácidos lineal con la que no está naturalmente unida o genéticamente fusionada en la naturaleza. Por ejemplo, en una realización el conector peptídico puede comprender polipéptidos que no existen de forma natural que son formas modificadas de polipéptidos que existen de forma natural (por ejemplo, que comprende una mutación tal como una adición, sustitución o delección). En otra realización, el conector peptídico puede comprender aminoácidos que no existen de forma natural. En otra realización, el conector peptídico puede comprender aminoácidos que existen de forma natural que ocurren en una secuencia lineal que no ocurre en la naturaleza. En otra realización adicional, el conector peptídico puede comprender una secuencia de polipéptidos que existen de forma natural.

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

En otra realización, un conector peptídico comprende o consiste en un conector de gly-ser. En una realización, un conector de scFc o cscFc comprende al menos una porción de una bisagra de inmunoglobulina y un conector de gly-ser. Como se usa en el presente documento, el término "conector de gly-ser" se refiere a un péptido que consiste en restos de glicina y serina. En ciertas realizaciones, dicho conector de gly-ser se puede insertar entre otras dos secuencias del conector peptídico. En otras realizaciones, un conector de gly-ser se une en uno o ambos extremos de

otra secuencia del conector peptídico. En aún otras realizaciones, dos o más conectores de gly-ser se incorporan en serie en un conector peptídico. En una realización, un conector peptídico de la invención comprende al menos una porción de una región bisagra superior (por ejemplo, derivada de una molécula de IgG1, IgG2, IgG3 o IgG4), al menos una porción de una región bisagra central (por ejemplo, derivada de una molécula de IgG1, IgG2, IgG3 o IgG4) y una serie de restos de aminoácidos gly/ser.

Los conectores peptídicos de la invención tienen al menos un aminoácido de longitud y pueden ser de longitudes variables. En una realización, un conector peptídico de la invención tiene desde aproximadamente 1 hasta aproximadamente 50 aminoácidos de longitud. Como se usa en este contexto, el término "aproximadamente" indica +/- dos restos de aminoácidos. Puesto que la longitud del conector debe ser un entero positivo, la longitud de desde aproximadamente 1 hasta aproximadamente 50 aminoácidos de longitud significa una longitud de desde 1-3 hasta 48-52 aminoácidos de longitud. En otra realización, un conector peptídico de la invención tiene desde aproximadamente 10 hasta aproximadamente 20 aminoácidos de longitud. En otra realización, un conector peptídico de la invención tiene desde aproximadamente 15 hasta aproximadamente 50 aminoácidos de longitud. En otra realización, un conector peptídico de la invención tiene desde aproximadamente 20 hasta aproximadamente 45 aminoácidos de longitud. En otra realización, un conector peptídico de la invención tiene desde aproximadamente 15 hasta aproximadamente 35 o aproximadamente 20 a aproximadamente 30 aminoácidos de longitud. En otra realización, un conector peptídico de la invención tiene desde aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000 o 2000 aminoácidos de longitud. En una realización, un conector peptídico de la invención tiene 20 o 30 aminoácidos de longitud.

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

Los conectores peptídicos se pueden introducir en las secuencias de polipéptidos usando técnicas conocidas en la técnica. Las modificaciones se pueden confirmar por análisis de secuencias de ADN. Se puede usar ADN de plásmido para transformar células hospedadoras para la producción estable de los polipéptidos producidos.

Híbridos de monómero-dímero

En algunas realizaciones, las moléculas de ácido nucleico aisladas de la invención que comprenden además una secuencia heteróloga de nucleótidos codifican una molécula híbrida de monómero-dímero que comprende FVIII.

El término "híbrido de monómero-dímero" usado en el presente documento se refiere a una proteína quimérica que comprende una primera cadena de polipéptidos y una segunda cadena de polipéptidos, que están asociadas entre sí por un enlace disulfuro, en donde la primera cadena comprende el factor VIII y una primera región Fc y la segunda cadena comprende, consiste esencialmente en o consiste en una segunda región Fc sin el FVIII. La construcción híbrida de monómero-dímero es así un híbrido que comprende un aspecto de monómero que tiene solo un factor de coagulación y un aspecto de dímero que tiene dos regiones Fc.

Secuencias de control de la transcripción

En algunas realizaciones, las moléculas de ácido nucleico aisladas de la invención se unen operativamente a al menos una secuencia de control de la transcripción. Una secuencia de control de la transcripción como se usa en el presente documento es cualquier secuencia de nucleótidos reguladora, tal como una secuencia promotora o combinación de promotor-potenciador, que facilita la eficiente transcripción y traducción del ácido nucleico codificante al que se une operativamente. La secuencia de control de la expresión génica puede ser, por ejemplo, un promotor de mamífero o viral, tal como un promotor constitutivo o inducible. Los promotores constitutivos de mamífero incluyen, pero no se limitan a, los promotores para los siguientes genes: promotor de hipoxantina fosforibosil transferasa (HPRT), adenosina desaminasa, piruvato cinasa, beta-actina, y otros promotores constitutivos. Los promotores virales a modo de ejemplo que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores del citomegalovirus (CMV), virus simio (por ejemplo, SV40), virus del papiloma, adenovirus, virus de la inmunodeficiencia humana (VIH), virus del sarcoma de Rous, citomegalovirus, las repeticiones terminales largas (LTR) del virus de la leucemia de Moloney, y otros retrovirus, y el promotor de timidina cinasa del virus del herpes simple. Los expertos habituales en la técnica conocen otros promotores constitutivos. Los promotores útiles como secuencias de expresión génica de la invención también incluyen promotores inducibles. Los promotores inducibles se expresan en presencia de un agente inductor. Por ejemplo, se induce el promotor de metalotioneína para promover la transcripción y traducción en presencia de ciertos iones metálicos. Los expertos habituales en la técnica conocen otros promotores inducibles.

En general, las secuencias de control de la transcripción deben incluir, según sea necesario, secuencias no transcriptoras de 5' y no traductoras de 5' implicadas en el inicio de la transcripción y traducción, respectivamente, tal como una caja TATA, secuencia de terminación, secuencia CAAT, y similares. Especialmente, dichas secuencias no transcriptoras de 5' incluirán una región promotora que incluye una secuencia promotora para el control de la transcripción del ácido nucleico codificante operativamente unido. Las secuencias de expresión génica incluyen opcionalmente secuencias potenciadoras o secuencias activadoras en la dirección 5', según se desee.

Vectores

La invención también proporciona vectores que comprenden las moléculas de ácido nucleico aisladas de la invención. Los vectores adecuados incluyen vectores de expresión, vectores virales y vectores plasmídicos.

Como se usa en el presente documento, un vector de expresión se refiere a cualquier construcción de ácidos nucleicos que contenga los elementos necesarios para la transcripción y la traducción de una secuencia codificante insertada, o en el caso de un vector viral de ARN, los elementos necesarios para la replicación y la traducción, cuando se introducen en una célula hospedadora apropiada. Los vectores de expresión pueden incluir plásmidos, fagémidos, virus y derivados de los mismos.

Los vectores de expresión de la invención incluirán polinucleótidos optimizados que codifican la proteína FVIII BDD descrita en el presente documento. En una realización, las secuencias codificantes optimizadas para la proteína FVIII BDD están operativamente unidas a una secuencia de control de la expresión. Como se usa en el presente documento, dos secuencias de ácidos nucleicos se unen operativamente cuando se unen covalentemente de tal forma que se permita que cada secuencia de ácidos nucleicos componente retenga su funcionalidad. Se dice que una secuencia codificante y una secuencia de control de la expresión génica se unen operativamente cuando se unen covalentemente de tal forma que se ponga la expresión o transcripción y/o traducción de la secuencia codificante bajo la influencia o el control de la secuencia de control de la expresión génica. Se dice que dos secuencias de ADN se unen operativamente si la inducción de un promotor en la secuencia de expresión del gen 5' da como resultado la transcripción de la secuencia codificante y si la naturaleza del enlace entre las dos secuencias de ADN no (1) da como resultado la introducción de una mutación con desplazamiento de marco, (2) interfiere con la capacidad de la región promotora para dirigir la transcripción de la secuencia codificante, o (3) interfiere con la capacidad del transcrito de ARN correspondiente que se va a traducir en una proteína. Así, una secuencia de expresión génica se unirá operativamente a una secuencia de ácidos nucleicos codificante si la secuencia de expresión génica fuera capaz de efectuar la transcripción de esa secuencia de ácidos nucleicos codificante de forma que el transcrito resultante se tradujera en la proteína deseada o polipéptido.

Los vectores virales incluyen, pero no se limitan a, secuencias de ácidos nucleicos de los siguientes virus: retrovirus, tales como el virus de la leucemia murina de Moloney, virus del sarcoma murino de Harvey, virus del tumor mamario murino y virus del sarcoma de Rous; adenovirus, virus adeno-asociado; virus tipo SV40; virus del poliovirus; virus de Epstein-Barr; virus del papiloma; virus del herpes; virus de la variolovacuna; virus de la poliomieltis; y virus de ARN tales como un retrovirus. Se pueden emplear fácilmente otros vectores muy conocidos en la técnica. Ciertos vectores virales se basan en virus eucariotas no citopáticos en los que genes no esenciales se han sustituido con el gen de interés. Los virus no citopáticos incluyen retrovirus, cuyo ciclo vital implica transcripción inversa de ARN viral genómico en ADN con posterior integración proviral en ADN de la célula hospedadora. Los retrovirus han sido autorizados para ensayos de terapia de genes humanos. Los más útiles son los retrovirus que son deficientes en la replicación (es decir, capaces de dirigir la síntesis de las proteínas deseadas, pero incapaces de fabricar una partícula infecciosa). Dichos vectores de expresión retroviral genéticamente alterados tienen utilidad general para la transducción de alta eficiencia de genes *in vivo*. Se proporcionan protocolos convencionales para producir retrovirus deficientes en la replicación (incluyendo las etapas de incorporación de material genético exógeno dentro de un plásmido, transfección de una línea celular de encapsidación con plásmido, producción de retrovirus recombinantes por la línea celular de encapsidación, recogida de partículas virales de medios de cultivo de tejido, e infección de las células diana con partículas virales) en Kriegler, M., Gene Transfer and Expression, A Laboratory Manual, W.H. Freeman Co., New York (1990) y Murry, E. J., Methods in Molecular Biology, Vol. 7, Humana Press, Inc., Clifton, N.J. (1991).

En una realización, el virus es un virus adeno-asociado, un virus de ADN bicatenario. El virus adeno-asociado se puede manipular para ser deficiente en la replicación y es capaz de infectar una amplia variedad de tipos de células y especies. Tiene además ventajas tales como estabilidad al calor y disolventes lipídicos; altas frecuencias de transducción en células de diversos linajes, que incluyen células hematopoyéticas; y carecen de inhibición de la superinfección, permitiendo así múltiples series de transducciones. Supuestamente, el virus adeno-asociado se puede integrar dentro del ADN celular humano de un modo específico de sitio, minimizando así la posibilidad de mutagénesis insercional y variabilidad de la expresión génica insertada característica de la infección retroviral. Además, se han seguido las infecciones por virus adeno-asociados naturales en cultivo de tejido durante más de 100 pases en ausencia de presión selectiva, que implica que la integración genómica del virus adeno-asociado es un acontecimiento relativamente estable. El virus adeno-asociado también puede funcionar de un modo extracromosómico.

Otros vectores incluyen vectores plasmídicos. Los vectores plasmídicos se han descrito ampliamente en la técnica y son bien conocidos para los expertos en la técnica. Véase, por ejemplo, Sambrook et al., Molecular Cloning: A Laboratory Manual, Segunda Edición, Cold Spring Harbor Laboratory Press, 1989. En los últimos años, se ha

encontrado que los vectores plasmídicos son particularmente ventajosos para administrar genes a células *in vivo* debido a su incapacidad para replicarse dentro de e integrarse en un genoma hospedador. Estos plásmidos, sin embargo, que tienen un promotor compatible con la célula hospedadora, pueden expresar un péptido de un gen operativamente codificado dentro del plásmido. Algunos plásmidos comúnmente usados disponibles de proveedores comerciales incluyen pBR322, pUC18, pUC19, diversos plásmidos de pcDNA, pRC/CMV, diversos plásmidos de pCMV, pSV40 y pBlueScript. Los ejemplos adicionales de plásmidos específicos incluyen pcDNA3.1, número de catálogo V79020; pcDNA3.1/hygro, número de catálogo V87020; pcDNA4/myc-His, número de catálogo V86320; y pBudCE4.1, número de catálogo V53220, todos de Invitrogen (Carlsbad, CA.). Otros plásmidos son bien conocidos por los expertos habituales en la técnica. Además, los plásmidos pueden ser diseñados a medida usando técnicas convencionales de biología molecular para retirar y/o añadir fragmentos específicos de ADN.

Células hospedadoras

La invención también proporciona células hospedadoras que comprenden las moléculas de ácido nucleico aisladas de la invención. Como se usa en el presente documento, el término "transformación" se debe usar en un amplio sentido para referirse a la introducción de ADN en una célula hospedadora receptora que cambia el genotipo y, por consiguiente, da como resultado un cambio en la célula receptora.

"Células hospedadoras" se refiere a células que han sido transformadas con vectores construidos usando técnicas de ADN recombinante y que codifican al menos un gen heterólogo. Las células hospedadoras de la presente invención son preferentemente de origen mamífero; lo más preferentemente de origen humano o de ratón. A los expertos en la técnica se les atribuye la capacidad de determinar preferentemente líneas de células hospedadoras particulares que son las más aptadas para su fin. Las líneas de células hospedadoras a modo de ejemplo incluyen, pero no se limitan a, CHO, DG44 y DUXB11 (líneas de ovario de hámster chino, DHFR menos), HELA (carcinoma cervical humano), CVI (línea de riñón de mono), COS (un derivado de CVI con antígeno T del SV40), R1610 (fibroblasto de hámster chino), BALBC/3T3 (fibroblasto de ratón), HAK (línea de riñón de hámster), SP2/O (mieloma de ratón), P3.times.63-Ag3.653 (mieloma de ratón), BFA-1c1BPT (células endoteliales bovinas), RAJI (linfocito humano), PER.C6®, NS0, CAP, BHK21 y HEK 293 (riñón humano). Las líneas de células hospedadoras normalmente están disponibles de servicios comerciales, la Colección Americana de Cultivos de tejido, o de la bibliografía publicada.

La introducción de las moléculas de ácido nucleico aisladas de la invención en la célula hospedadora se puede llevar a cabo por diversas técnicas bien conocidas por los expertos en la técnica. Estas incluyen, pero no se limitan a, transfección (incluyendo electroforesis y electroporación), fusión de protoplastos, precipitación con fosfato de calcio, fusión de células con ADN envuelto, microinyección e infección con virus intacto. Véase Ridgway, A. A. G. "Mammalian Expression Vectors" Capítulo 24.2, pp. 470-472 Vectors, Rodriguez and Denhardt, Eds. (Butterworths, Boston, Mass. 1988). Lo más preferentemente, la introducción de plásmidos en el hospedador es por electroporación. Las células transformadas se cultivan en condiciones apropiadas para la producción de las cadenas ligeras y cadenas pesadas, y se ensayan para la síntesis de proteínas de la cadena pesada y/o ligera. Las técnicas de ensayo a modo de ejemplo incluyen enzimoanálisis de adsorción (ELISA), radioinmunoensayo (RIA) o citometría de flujo activada por fluorescencia (FACS), inmunohistoquímica y similares.

Las células hospedadoras que comprenden las moléculas de ácido nucleico aisladas de la invención se cultivan en un medio de crecimiento apropiado. Como se usa en el presente documento, el término "medio de crecimiento apropiado" significa un medio que contiene los nutrientes requeridos para el crecimiento de las células. Los nutrientes requeridos para el crecimiento celular pueden incluir una fuente de carbono, una fuente de nitrógeno, aminoácidos esenciales, vitaminas, minerales y factores de crecimiento. Opcionalmente, los medios pueden contener uno o más factores de selección. Opcionalmente, los medios pueden contener suero de ternero bovino o suero de ternero fetal (FCS). En una realización, los medios no contienen sustancialmente IgG. El medio de crecimiento se seleccionará, en general, para células que contienen la construcción de ADN por, por ejemplo, selección de fármacos o deficiencia en un nutriente esencial que se complementa por el marcador de selección sobre la construcción de ADN o se co-transfecta con la construcción de ADN. Las células de mamífero cultivadas, en general, se cultivan en medio que contiene suero o sin suero comercialmente disponible (por ejemplo, MEM, DMEM, DMEM/F12). En una realización, el medio es CDoptiCHO (Invitrogen, Carlsbad, CA.). En otra realización, el medio es CD17 (Invitrogen, Carlsbad, CA.). La selección de un medio apropiado para la línea celular particular usada está dentro del nivel de los expertos habituales en la técnica.

Preparación de polipéptidos

Un polipéptido puede ser codificado por las moléculas de ácido nucleico aisladas de la invención. En algunas realizaciones, el polipéptido está codificado por un vector que comprende las moléculas nucleicas aisladas de la invención. En otras realizaciones, el polipéptido se produce por una célula hospedadora que comprende las moléculas nucleicas aisladas de la invención.

En una realización, la invención proporciona un método de producción de un polipéptido con actividad de FVIII, que comprende cultivar una célula hospedadora de la invención en condiciones por las cuales se produce un polipéptido con actividad de FVIII, y recuperar el polipéptido con actividad de FVIII. En algunas realizaciones, la expresión del polipéptido con actividad de FVIII es elevada con respecto a una célula hospedadora cultivada en las mismas

condiciones, pero que contiene una secuencia de nucleótidos de referencia que comprende SEQ ID NO: 3, la secuencia del gen FVIII parental.

5 En otras realizaciones, la invención proporciona un método de aumento de la expresión de un polipéptido con actividad de FVIII que comprende cultivar una célula hospedadora de la invención en condiciones por las cuales un polipéptido con actividad de FVIII se expresa por la molécula de ácido nucleico, en donde la expresión del polipéptido con actividad de FVIII es elevada con respecto a una célula hospedadora cultivada en las mismas condiciones que comprende una molécula de ácido nucleico de referencia que comprende SEQ ID NO: 3.

10 En otras realizaciones, la invención proporciona un método de mejora del rendimiento de un polipéptido con actividad de factor VIII que comprende cultivar una célula hospedadora en condiciones por las cuales se produce un polipéptido con actividad de factor VIII por la molécula de ácido nucleico, en donde el rendimiento del polipéptido con actividad de factor VIII es elevado con respecto a una célula hospedadora cultivada en las mismas condiciones que comprende una secuencia de ácidos nucleicos de referencia que comprende SEQ ID NO: 3.

15 Está disponible una variedad de métodos para producir recombinantemente una proteína FVIII a partir de la molécula de ácido nucleico optimizada de la invención. Se puede producir un polinucleótido de la secuencia deseada por síntesis de ADN en fase sólida *de novo* o por mutagénesis por PCR de un polinucleótido preparado antes. La mutagénesis mediada por oligonucleótidos es un método de preparación de una sustitución, inserción, delección o alteración (por ejemplo, codón alterado) en una secuencia de nucleótidos. Por ejemplo, el ADN inicial se altera hibridando un oligonucleótido que codifica la mutación deseada con un molde de ADN monocatenario. Después de la hibridación, se usa una ADN polimerasa para sintetizar una segunda hebra complementaria completa del molde que incorpora el cebador de oligonucleótidos. En una realización, es suficiente ingeniería genética, por ejemplo, mutagénesis por PCR basada en cebadores, para incorporar una alteración, como se define en el presente documento, para producir un polinucleótido de la invención.

20 Para la producción de proteínas recombinantes, se inserta una secuencia de polinucleótidos optimizada de la invención que codifica la proteína FVIII en un vehículo de expresión apropiado, es decir, un vector que contiene los elementos necesarios para la transcripción y la traducción de la secuencia codificante insertada, o en el caso de un vector viral de ARN, los elementos necesarios para la replicación y la traducción.

25 La secuencia de polinucleótidos de la invención se inserta en el vector en el marco de lectura apropiado. El vector de expresión se transfecta entonces en una célula diana adecuada que expresará el polipéptido. Las técnicas de transfección conocidas en la técnica incluyen, pero no se limitan a, precipitación con fosfato de calcio (Wigler et al. 1978, Cell 14:725) y electroporación (Neumann et al. 1982, EMBO, J. 1:841). Se puede utilizar una variedad de sistemas de hospedador-vector de expresión para expresar las proteínas FVIII descritas en el presente documento en células eucariotas. En una realización, la célula eucariota es una célula de animal, que incluye células de mamífero (por ejemplo, células HEK293, PER.C6®, células CHO, BHK, Cos, HeLa). Una secuencia de polinucleótidos de la invención también puede codificar una secuencia señal que permitirá secretar la proteína FVIII. Un experto en la técnica entenderá que mientras se traduce la proteína FVIII, la secuencia señal se escinde por la célula para formar la proteína madura. Se conocen en la técnica diversas secuencias señal, por ejemplo, secuencia señal del factor VII nativo, secuencia señal del factor IX nativo y la secuencia señal de la cadena ligera de IgK de ratón. Alternativamente, donde una secuencia señal no está incluida en la proteína FVIII, se puede recuperar lisando las células.

30 La proteína FVIII se puede sintetizar en un animal transgénico no humano, tal como un roedor, cabra, oveja, cerdo o vaca. El término "animales transgénicos" se refiere a animales no humanos que han incorporado un gen extraño en su genoma. Debido a que este gen está presente en tejidos de la línea germinal, se pasa de los padres a la descendencia. Se introducen genes exógenos en embriones unicelulares (Brinster et al. 1985, Proc. Natl. Acad. Sci. USA 82:4438). Se conocen en la técnica los métodos de producción animales transgénicos que incluyen transgénica que produce moléculas de inmunoglobulina (Wagner et al. 1981, Proc. Natl. Acad. Sci. USA 78: 6376; McKnight et al. 1983, Cell 34 : 335; Brinster et al. 1983, Nature 306: 332; Ritchie et al. 1984, Nature 312: 517; Baldassarre et al. 2003, Theriogenology 59 : 831; Robl et al. 2003, Theriogenology 59: 107; Malassagne et al. 2003, Xenotransplantation 10 (3): 267).

35 Los vectores de expresión pueden codificar marcas que permiten la fácil purificación o identificación de la proteína recombinantemente producida. Los ejemplos incluyen, pero no se limitan a, el vector pUR278 (Ruther et al. 1983, EMBO J. 2: 1791) en el que la secuencia codificante descrita en el presente documento de la proteína FVIII se puede unir en el vector en marco con la región codificante Z de lac de manera que se produzca una proteína híbrida; se pueden usar vectores pGEX para expresar proteínas con una marca de glutatión S-transferasa (GST). Estas proteínas son normalmente solubles y se pueden purificar fácilmente de las células por adsorción a perlas de glutatión-agarosa, seguido por elución en presencia de glutatión libre. Los vectores incluyen sitios de escisión (por ejemplo, PreCission Protease (Pharmacia, Peapack, N. J.)) para la fácil retirada de la marca después de la purificación.

40 Para los fines de la presente invención, se pueden emplear numerosos sistemas de vector de expresión. Estos vectores de expresión normalmente son replicables en los organismos hospedadores ya sea como episomas o como una parte integral del ADN cromosómico del hospedador. Los vectores de expresión pueden incluir secuencias de control de la expresión que incluyen, pero no se limitan a, promotores (por ejemplo, promotores naturalmente

asociados o heterólogos), potenciadores, secuencias señal, señales de corte y empalme, elementos potenciadores y secuencias de terminación de la transcripción. Preferentemente, las secuencias de control de la expresión son sistemas de promotores eucariotas en vectores capaces de transformar o transfectar las células hospedadoras eucariotas. Los vectores de expresión también pueden utilizar elementos de ADN que derivan de virus animales, tales como virus del papiloma bovino, virus del polioma, adenovirus, virus de la variolovacuna, baculovirus, retrovirus (RSV, MMTV o MOMLV), citomegalovirus (CMV) o virus SV40. Otros implican el uso de sistemas policistrónicos con sitios internos de unión al ribosoma.

Comúnmente, los vectores de expresión contienen marcadores de selección (por ejemplo, resistencia a ampicilina, resistencia a higromicina, resistencia a tetraciclina o resistencia a neomicina) para permitir la detección de las células transformadas con las secuencias de ADN deseadas (véase, por ejemplo, Itakura et al., patente de EE. UU. 4.704.362). Se pueden seleccionar las células que han integrado el ADN en sus cromosomas introduciendo uno o más marcadores que permiten la selección de las células hospedadoras transfectadas. El marcador puede proporcionar prototrofia a un hospedador auxotrófico, resistencia a biocidas (por ejemplo, antibióticos) o resistencia a metales pesados tales como el cobre. El gen marcador de selección se puede unir o directamente a las secuencias de ADN a expresar, o introducir en la misma célula por cotransformación.

Un ejemplo de un vector útil para expresar una secuencia de FVIII optimizado es NEOSPLA (patente de EE. UU. N° 6.159.730). Este vector contiene el promotor/potenciador del citomegalovirus, el promotor principal de beta-globina de ratón, el origen de replicación SV40, la secuencia de poliadenilación de la hormona de crecimiento bovina, el exón 1 y el exón 2 de neomicina fosfotransferasa, el gen dihidrofolato reductasa y la secuencia conductora. Se ha encontrado que este vector da como resultado una expresión de muy alto nivel de anticuerpos después de la incorporación de genes de las regiones variables y constantes, transfección en células, seguido por selección en medio que contiene G418 y amplificación con metotrexato. También se enseñan sistemas de vector en las patentes de EE. UU. N° 5.736.137 y 5.658.570. Este sistema proporciona altos niveles de expresión, por ejemplo, > 30 pg/célula/día. Otros sistemas de vector a modo de ejemplo se desvelan, por ejemplo, en la patente de EE. UU. N° 6.413.777.

En otras realizaciones, los polipéptidos se pueden expresar usando construcciones policistrónicas. En estos sistemas de expresión, se pueden producir múltiples productos génicos de interés tales como múltiples polipéptidos de proteína de unión a multímero a partir de una única construcción policistrónica. Estos sistemas usan ventajosamente un sitio interno de entrada al ribosoma (IRES) para proporcionar niveles relativamente altos de polipéptidos en células hospedadoras eucariotas. Las secuencias de IRES compatibles se desvelan en la patente de EE. UU. N° 6.193.980, que también se incorpora en el presente documento.

Más en general, una vez se ha preparado el vector o la secuencia de ADN que codifica un polipéptido, el vector de expresión se puede introducir en una célula hospedadora apropiada. Es decir, las células hospedadoras se pueden transformar. La introducción del plásmido en la célula hospedadora se puede llevar a cabo por diversas técnicas bien conocidas por los expertos en la técnica, como se trata anteriormente. Las células transformadas se cultivan en condiciones apropiadas para la producción del polipéptido FVIII, y se ensayan para la síntesis de polipéptidos FVIII. Las técnicas de ensayo a modo de ejemplo incluyen enzimoinmunoanálisis de adsorción (ELISA), radioinmunoensayo (RIA) o citometría de flujo activada por fluorescencia (FACS), inmunohistoquímica y similares.

En descripciones de procesos para el aislamiento de polipéptidos a partir de hospedadores recombinantes, los términos "célula" y "cultivo celular" se usan indistintamente para indicar la fuente de polipéptido, a menos que se especifique evidentemente de otro modo. En otras palabras, la recuperación del polipéptido de las "células" puede significar cualquier de centrifugación de células completas, o del cultivo celular que contiene tanto el medio como las células suspendidas.

La línea de células hospedadoras usada para la expresión de proteínas es preferentemente de origen mamífero; lo más preferentemente de origen humano o de ratón, ya que los ácidos nucleicos aislados de la invención se han optimizado para la expresión en células humanas. Se han descrito anteriormente líneas de células hospedadoras a modo de ejemplo. En una realización del método para producir un polipéptido con actividad de FVIII, la célula hospedadora es una célula HEK293. En otra realización del método para producir un polipéptido con actividad de FVIII, la célula hospedadora es una célula CHO.

Los genes que codifican los polipéptidos también se pueden expresar en células no de mamífero, tales como células de bacteria o de levadura o vegetales. A este respecto se apreciará que también se pueden transformar diversos microorganismos unicelulares no de mamífero tales como bacterias; es decir, los capaces de ser cultivados en cultivos o fermentación. Las bacterias, que son susceptibles a la transformación, incluyen miembros de las enterobacteriaceae, tales como cepas de *Escherichia coli* o *Salmonella*; Bacillaceae, tales como *Bacillus subtilis*; *Pneumococcus*; *Streptococcus* y *Haemophilus influenzae*. Se apreciará además que, cuando se expresan en bacterias, los polipéptidos normalmente llegan a ser parte de cuerpos de inclusión. Los polipéptidos se deben aislar, purificar y luego ensamblar en moléculas funcionales.

Alternativamente, se pueden incorporar secuencias de nucleótidos optimizadas de la invención en transgenes para la introducción en el genoma de un animal transgénico no humano y la posterior expresión en la leche del animal transgénico (véase, por ejemplo, Deboer et al., documento de patente US 5.741.957, Rosen, documento de patente

US 5.304.489 y Meade et al., documento de patente US 5.849.992). Los transgenes adecuados incluyen secuencias codificantes para polipéptidos en enlace operable con un promotor y potenciador de un gen específico de glándulas mamarias, tal como caseína o beta-lactoglobulina.

5 La producción *in vitro* permite el aumento de escala para dar grandes cantidades de los polipéptidos alterados deseados de la invención. Se conocen en la técnica técnicas para el cultivo de células de mamífero en condiciones de cultivo de tejido e incluyen cultivo en suspensión homogénea, por ejemplo en un reactor de columna aireada o en un reactor continuo con agitador, o cultivo celular inmovilizado o atrapado, por ejemplo en fibras huecas, microcápsulas, sobre microperlas de agarosa o cartuchos de cerámica. Si fuera necesario y/o se deseara, las disoluciones de polipéptidos se pueden purificar por los métodos habituales de cromatografía, por ejemplo filtración en gel, 10 cromatografía de intercambio iónico, cromatografía sobre DEAE-celulosa o cromatografía de (inmuno)afinidad, por ejemplo, después de la biosíntesis preferente de un polipéptido sintético de la región bisagra o antes de o después de la etapa de cromatografía HIC descrita en el presente documento. Se puede unir o incluir opcionalmente una secuencia de marca de afinidad (por ejemplo, una marca de His(6)) dentro de la secuencia de polipéptidos para facilitar la purificación en la dirección 3'.

15 Una vez expresada, la proteína FVIII se puede purificar según procedimientos convencionales de la técnica, que incluyen precipitación con sulfato de amonio, cromatografía en columna de afinidad, purificación por HPLC, electroforesis en gel y similares (véase, en general, Scopes, Protein Purification (Springer-Verlag, N.Y., (1982)). Se prefieren proteínas sustancialmente puras de al menos aproximadamente 90 a 95 % de homogeneidad, y lo más preferido 98 a 99 % o más de homogeneidad, para fines farmacéuticos.

20 Composición farmacéutica

Las composiciones que contienen los ácidos nucleicos aislados de la presente invención, pueden contener un vehículo farmacéuticamente aceptable adecuado. Por ejemplo, pueden contener excipientes y/o auxiliares que facilitan el procesamiento de los compuestos activos en preparaciones diseñadas para administración al sitio de acción.

25 La composición farmacéutica se puede formular para administración parenteral (es decir, intravenosa, subcutánea, o intramuscular) por inyección en bolo. Las formulaciones para inyección se pueden presentar en forma farmacéutica unitaria, por ejemplo, en ampollas o en recipientes de múltiples dosis con un conservante añadido. Las composiciones pueden adoptar formas tales como suspensiones, disoluciones o emulsiones en vehículos aceitosos o acuosos, y contienen agentes de formulación tales como agentes de suspensión, estabilizantes y/o dispersantes. Alternativamente, el principio activo puede estar en forma de polvo para constitución con un vehículo adecuado, por 30 ejemplo, agua sin pirógenos.

Las formulaciones adecuadas para administración parenteral también incluyen disoluciones acuosas de los compuestos activos en forma soluble en agua, por ejemplo, sales solubles en agua. Además, se pueden administrar suspensiones de los compuestos activos como suspensiones aceitosas apropiadas para inyección. Los disolventes o vehículos lipófilos adecuados incluyen aceites grasos, por ejemplo, aceite de sésamo, o ésteres de ácidos grasos 35 sintéticos, por ejemplo, oleato de etilo o triglicéridos. Las suspensiones para inyección acuosa pueden contener sustancias, que aumentan la viscosidad de la suspensión, que incluyen, por ejemplo, carboximetilcelulosa de sodio, sorbitol y dextrano. Opcionalmente, la suspensión también puede contener estabilizadores. También se pueden usar liposomas para encapsular las moléculas de la invención para la administración en células o espacios intersticiales. Los vehículos farmacéuticamente aceptables a modo de ejemplo son disolventes fisiológicamente compatibles, medios 40 de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y de retraso de la absorción, agua, solución salina, solución salina tamponada con fosfato, dextrosa, glicerol, etanol y similares. En algunas realizaciones, la composición comprende agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol, o cloruro sódico. En otras realizaciones, las composiciones comprenden sustancias farmacéuticamente aceptables tales como agentes humectantes o cantidades menores de sustancias auxiliares tales como agentes 45 humectantes o emulsionantes, conservantes o tampones, que potencian la estabilidad en almacén o la eficacia de los principios activos.

Las composiciones de la invención pueden estar en una variedad de formas, que incluyen, por ejemplo, líquidos (por ejemplo, disoluciones inyectables e infusibles), dispersiones, suspensiones, formas farmacéuticas semisólidas y sólidas. La forma preferida depende del modo de administración y la aplicación terapéutica.

50 La composición se puede formular como una disolución, microemulsión, dispersión, liposoma, u otra estructura ordenada adecuada para alta concentración de fármaco. Se pueden preparar disoluciones inyectables estériles incorporando el principio activo en la cantidad requerida en un disolvente apropiado con uno o una combinación de los componentes enumerados anteriormente, según se requiera, seguido por esterilización por filtración. En general, las dispersiones se preparan incorporando el principio activo en un vehículo estéril que contiene un medio de 55 dispersión básico y los otros componentes requeridos de los enumerados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los métodos preferidos de preparación son secado a vacío y liofilización, que dan un polvo del principio activo más cualquier componente deseado adicional de una disolución previamente esterilizada por filtración. La fluidez apropiada de una disolución se puede mantener, por ejemplo, usando un recubrimiento tal como lecitina, por el mantenimiento del tamaño de partículas requerido en el caso de dispersión

y usando tensioactivos. La absorción prolongada de las composiciones inyectables se puede provocar incluyendo en la composición un agente que retrasa la absorción, por ejemplo, sales de monoestearato y gelatina.

El principio activo se puede formular con una formulación o dispositivo de liberación controlada. Los ejemplos de dichas formulaciones y dispositivos incluyen implantes, parches transdérmicos y sistemas de administración microencapsulada. Se pueden usar polímeros biodegradables biocompatibles, por ejemplo, etileno-acetato de vinilo, polianhídridos, ácido poliglicólico, colágeno, poliortoésteres y ácido poliláctico. Se conocen en la técnica los métodos para la preparación de dichas formulaciones y dispositivos. Véase, por ejemplo, Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

Se pueden preparar formulaciones de liberación prolongada inyectables formando matrices microencapsuladas del fármaco en polímeros biodegradables tales como polilactida-poliglicolida. Dependiendo de la relación entre el fármaco y el polímero, y la naturaleza del polímero empleado, se puede controlar la velocidad de liberación del fármaco. Otros polímeros biodegradables a modo de ejemplo son poliortoésteres y polianhídridos. También se pueden preparar las formulaciones inyectables de liberación prolongada atrapando el fármaco en liposomas o microemulsiones.

Se pueden incorporar compuestos activos suplementarios en las composiciones. En una realización, la proteína quimérica se formula con otro factor de coagulación, o una variante, fragmento, análogo, o derivado del mismo. Por ejemplo, el factor de coagulación incluye, pero no se limita a, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, factor XII, factor XIII, protrombina, fibrinógeno, factor de von Willebrand o factor tisular soluble recombinante (rsTF), o formas activadas de cualquiera de los precedentes. El factor de coagulación de agente hemostático también puede incluir fármacos antifibrinolíticos, por ejemplo, ácido épsilon-amino-caproico, ácido tranexámico.

Se pueden ajustar las pautas posológicas para proporcionar la respuesta deseada óptima. Por ejemplo, se puede administrar un único bolo, se pueden administrar varias dosis divididas con el tiempo, o se puede reducir o incrementar proporcionalmente la dosis como se indica por las exigencias de la situación terapéutica. Es ventajoso formular las composiciones parentales en forma de dosificación unitaria para facilitar la administración y uniformidad de la dosificación. Véase, por ejemplo, Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980).

Además del compuesto activo, la forma farmacéutica líquida puede contener componentes inertes tales como agua, alcohol etílico, carbonato de etilo, acetato de etilo, alcohol bencílico, benzoato de bencilo, propilenglicol, 1,3-butilenglicol, dimetilformamida, aceites, glicerol, alcohol tetrahidrofurfurílico, polietilenglicoles, y ésteres de ácidos grasos de sorbitano.

También se describen ejemplos no limitantes de vehículos farmacéuticos adecuados en Remington's Pharmaceutical Sciences por E. W. Martin. Algunos ejemplos de excipientes incluyen almidón, glucosa, lactosa, sacarosa, gelatina, malta, arroz, harina, caliza, gel de sílice, estearato de sodio, monoestearato de glicerol, talco, cloruro sódico, leche desnatada en polvo, glicerol, propileno, glicol, agua, etanol, y similares. La composición también puede contener reactivos de tamponamiento del pH, y agentes humectantes o emulsionantes.

Para administración por vía oral, la composición farmacéutica puede tomar la forma de comprimidos o cápsulas preparadas mediante medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (por ejemplo, jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes (lecitina o goma arábiga), vehículos no acuosos (por ejemplo, aceite de almendra, ésteres aceitosos, alcohol etílico, o aceites vegetales fraccionados) y conservantes (por ejemplo, p-hidroxibenzoatos de metilo o propilo o ácido sórbico). Las preparaciones también pueden incluir aromatizantes, colorantes y edulcorantes. Alternativamente, la composición se puede presentar como un producto seco para constitución con agua u otro vehículo adecuado.

Para administración por vía oral, la composición puede tomar la forma de comprimidos o pastillas para chupar según protocolos convencionales.

Para administración por inhalación, los compuestos para su uso según la presente invención se administran convenientemente en forma de un aerosol nebulizado con o sin excipientes, o en forma de un spray de aerosol de un envase presurizado o nebulizador, con opcionalmente un propulsor, por ejemplo, diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad dosificada. Se pueden formular cápsulas y cartuchos de, por ejemplo, gelatina para su uso en un inhalador o insuflador que contiene una mezcla en polvo del compuesto y una base en polvo adecuada tal como lactosa o almidón.

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

En una realización, una composición farmacéutica comprende una proteína FVIII, el polinucleótido optimizado que codifica la proteína FVIII, el vector que comprende el polinucleótido, o la célula hospedadora que comprende el vector, y un vehículo farmacéuticamente aceptable. En algunas realizaciones, la composición se administra por una vía seleccionada del grupo que consiste en administración tópica, administración intraocular, administración parenteral,

administración intratecal, administración subdural y administración por vía oral. La administración parenteral puede ser intravenosa o administración subcutánea,

5 En otras realizaciones, la composición se usa para tratar una enfermedad o afección hemorrágica en un sujeto en necesidad del mismo. La enfermedad o afección hemorrágica se selecciona del grupo que consiste en un trastorno de la coagulación hemorrágico, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en músculos, hemorragia oral, traumatismo, traumatismo de la cabeza, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intrabdominal, hemorragia intratorácica, fractura de hueso, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, sangrado en la vaina del iliopsoas y cualquier combinación de los mismos. En otras realizaciones más, el sujeto es citado para realizarle una cirugía. En 10 aún otras realizaciones, el tratamiento es profiláctico o a demanda.

Métodos de tratamiento

15 La invención proporciona un método de tratamiento de un trastorno hemorrágico que comprende administrar a un sujeto en necesidad del mismo una molécula de ácido nucleico o vector de la invención. En algunas realizaciones, el trastorno hemorrágico se caracteriza por una deficiencia en el factor VIII. En algunas realizaciones, el trastorno hemorrágico es hemofilia. En algunas realizaciones, el trastorno hemorrágico es hemofilia A. En algunas realizaciones del método de tratamiento de un trastorno hemorrágico, la actividad de factor VIII en plasma 24 horas después de la administración es elevada con respecto a un sujeto administrado con una molécula de ácido nucleico de referencia que comprende SEQ ID NO: 3, un vector que comprende la molécula de ácido nucleico de referencia o un polipéptido codificado por la molécula de ácido nucleico de referencia.

20 La invención también se refiere a un método de tratamiento, mejora o prevención de un trastorno hemostático en un sujeto que comprende administrar una cantidad terapéuticamente eficaz de una molécula de ácido nucleico aislada de la invención. El tratamiento, la mejora y la prevención por la molécula de ácido nucleico aislada puede ser una terapia de derivación. El sujeto que recibe la terapia de derivación puede ya haber desarrollado un inhibidor a un factor de coagulación, por ejemplo, factor VIII, o se somete a desarrollo de un inhibidor del factor de coagulación.

25 Las moléculas de ácidos nucleicos o vectores de la invención tratan o previenen un trastorno hemostático promoviendo la formación de un coágulo de fibrina. La proteína FVIII puede activar un miembro de una cascada de coagulación. El factor de coagulación puede ser un participante en la vía extrínseca, la vía intrínseca, o ambas.

30 Las moléculas de ácidos nucleicos o vectores de la invención se pueden usar para tratar trastornos hemostáticos que se sabe que son tratables con FVIII. Los trastornos hemostáticos que se pueden tratar usando los métodos de la invención incluyen, pero no se limitan a, hemofilia A, hemofilia B, enfermedad de von Willebrand, deficiencia de factor XI (deficiencia de PTA), deficiencia de factor XII, así como deficiencias o anomalías estructurales en el fibrinógeno, protrombina, factor V, factor VII, factor X o factor XIII, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, traumatismo, traumatismo de la cabeza, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intrabdominal, hemorragia intratorácica, fractura de hueso, sangrado del sistema 35 nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal y sangrado en la vaina del iliopsoas. Las composiciones para administración a un sujeto incluyen moléculas de ácidos nucleicos que comprenden una secuencia de nucleótidos optimizada de la invención que codifica un factor de coagulación de FVIII (para aplicaciones de terapia génica), así como moléculas de polipéptido FVIII.

40 En algunas realizaciones, el trastorno hemostático es un trastorno heredado. En una realización, el sujeto tiene hemofilia A. En otras realizaciones, el trastorno hemostático es el resultado de una deficiencia en el factor VIII. En otras realizaciones, el trastorno hemostático puede ser el resultado de un factor de coagulación de FVIII defectuoso.

45 En otra realización, el trastorno hemostático puede ser un trastorno adquirido. El trastorno adquirido puede resultar de una enfermedad o afección secundaria subyacente. La afección sin relacionar puede ser, como un ejemplo, pero no como una limitación, cáncer, una enfermedad autoinmunitaria o embarazo. El trastorno adquirido puede resultar de vejez o de medicación para tratar un trastorno secundario subyacente (por ejemplo, quimioterapia contra el cáncer).

50 La invención también se refiere a métodos de tratamiento de un sujeto que no tiene un trastorno hemostático o una enfermedad o afección secundaria que da como resultado la adquisición de un trastorno hemostático. Así, la invención se refiere a un método de tratamiento de un sujeto en necesidad de un agente hemostático general que comprende administrar una cantidad terapéuticamente eficaz de una molécula de ácido nucleico aislada de la invención. Por ejemplo, en una realización, el sujeto en necesidad de un agente hemostático general se va a someter a, o está a punto de someterse, a cirugía. La molécula de ácido nucleico aislada de la invención se puede administrar antes o después de la cirugía como un profiláctico. Una molécula de ácido nucleico aislada de la invención se puede administrar durante o después de la cirugía para controlar un episodio hemorrágico agudo. La cirugía puede incluir, pero no se limita a, trasplante de hígado, resección de hígado o trasplante de células madre.

55 En otra realización, una molécula de ácido nucleico aislada de la invención se puede usar para tratar un sujeto que tiene un episodio hemorrágico agudo que no tiene un trastorno hemostático. El episodio hemorrágico agudo puede resultar de un traumatismo intenso, por ejemplo, cirugía, un accidente de coche, herida, laceración, arma de fuego o cualquier otro acontecimiento traumático que da como resultado una hemorragia incontrolada.

Las moléculas de ácido nucleico aisladas de la invención se pueden usar para tratar profilácticamente un sujeto con un trastorno hemostático. Las moléculas de ácido nucleico aisladas de la invención se pueden usar para tratar un episodio hemorrágico agudo en un sujeto con un trastorno hemostático.

5 Una composición de proteína FVIII puede administrarse en combinación con al menos otro agente que promueve la hemostasia. Dicho otro agente que promueve la hemostasia es un terapéutico con actividad coagulante demostrada. Como un ejemplo, pero no como una limitación, el agente hemostático puede incluir factor V, factor VII, factor IX, factor X, factor XI, factor XII, factor XIII, protrombina o fibrinógeno o formas activadas de cualquiera de los precedentes. El factor coagulante o agente hemostático también puede incluir fármacos antifibrinolíticos, por ejemplo, ácido épsilon-amino-caproico, ácido tranexámico.

10 En una realización de la invención, la composición (por ejemplo, la molécula de ácido nucleico optimizada que codifica el polipéptido FVIII) es una en la que el FVIII está presente en forma activable cuando se administra a un sujeto. Dicha molécula activable se puede activar *in vivo* en el sitio de la coagulación después de la administración a un sujeto.

15 La molécula de ácido nucleico optimizada que codifica el polipéptido FVIII se puede administrar por vía intravenosa, por vía subcutánea, por vía intramuscular o por cualquier superficie mucosa, por ejemplo, por vía oral, por vía sublingual, por vía bucal, por vía sublingual, por vía nasal, por vía rectal, por vía vaginal o por vía pulmonar.

20 Para administración por vía oral, la composición farmacéutica puede tomar la forma de comprimidos o cápsulas preparadas mediante medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (por ejemplo, jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes (lecitina o goma arábiga), vehículos no acuosos (por ejemplo, aceite de almendra, ésteres aceitosos, alcohol etílico o aceites vegetales fraccionados) y conservantes (por ejemplo, p-hidroxibenzoatos de metilo o propilo o ácido sórbico). Las preparaciones también pueden incluir aromatizantes, colorantes y edulcorantes. Alternativamente, la composición se puede presentar como un producto seco para la constitución con agua u otro vehículo adecuado.

25 Para administración bucal y sublingual, la composición puede tomar la forma de comprimidos, pastillas para chupar o películas de rápida disolución según protocolos convencionales.

30 Para administración por inhalación, el polipéptido que tiene actividad de FVIII se administra convenientemente en forma de un spray de aerosol de un envase presurizado o nebulizador (por ejemplo, en PBS), con un propulsor apropiado, por ejemplo, diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosis se puede determinar proporcionando una válvula para administrar una cantidad dosificada. Se pueden formular cápsulas y cartuchos de, por ejemplo, gelatina para su uso en un inhalador o insuflador que contiene una mezcla en polvo del compuesto y una base de polvo adecuada tal como lactosa o almidón.

35 En una realización, la vía de administración de la molécula de ácido nucleico optimizada que codifica el polipéptido FVIII es parenteral. El término parenteral como se usa en el presente documento incluye administración intravenosa, intrarterial, intraperitoneal, intramuscular, subcutánea, rectal o vaginal. Se prefiere la forma intravenosa de administración parenteral. Mientras que todas estas formas de administración son claramente contempladas por estar dentro del alcance de la invención, una forma para administración sería una disolución para inyección, en particular para inyección intravenosa o intrarterial o goteo. Normalmente, una composición farmacéutica adecuada para inyección puede comprender un tampón (por ejemplo, tampón acetato, fosfato o citrato), un tensioactivo (por ejemplo, polisorbato), opcionalmente un estabilizador (por ejemplo, albúmina humana), etc. Sin embargo, en otros métodos compatibles con las enseñanzas en el presente documento, las moléculas de ácidos nucleicos optimizadas que codifican los polipéptidos FVIII se pueden administrar directamente al sitio de la población celular adversa, aumentando así la exposición del tejido enfermo al agente terapéutico.

45 Las preparaciones para administración parenteral incluyen disoluciones, suspensiones y emulsiones estériles acuosas o no acuosas. Los ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales tales como aceite de oliva, y ésteres orgánicos inyectables tales como oleato de etilo. Los vehículos acuosos incluyen agua, disoluciones, emulsiones o suspensiones alcohólicas/acuosas, que incluyen solución salina y medios tamponados. En el objeto de la invención, los vehículos farmacéuticamente aceptables incluyen, pero no se limitan a, 0,01-0,1 M y preferentemente tampón fosfato 0,05 M o solución salina al 0,8 %. Otros vehículos parenterales comunes incluyen disoluciones de fosfato de sodio, dextrosa de Ringer, dextrosa y cloruro sódico, Ringer con lactato, o aceites no volátiles. Los vehículos intravenosos incluyen reforzadores de fluidos y de nutrientes, reforzadores de electrolitos, tales como los basados en dextrosa de Ringer, y similares. También pueden estar presentes conservantes y otros aditivos, tales como, por ejemplo, antimicrobianos, antioxidantes, agentes quelantes y gases inertes, y similares.

55 Más particularmente, las composiciones farmacéuticas adecuadas para uso inyectable incluyen disoluciones o dispersiones acuosas estériles (cuando son solubles en agua) y polvos estériles para la preparación extemporánea de disoluciones o dispersiones inyectables estériles. En tales casos, la composición debe ser estéril y debe ser fluida hasta el punto de que exista fácil inyectabilidad. Debe ser estable en las condiciones de fabricación y almacenamiento y preferentemente se conservará contra la acción contaminante de microorganismos, tales como bacterias y hongos.

El vehículo puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido, y similares), y mezclas adecuadas de los mismos. La fluidez apropiada se puede mantener, por ejemplo, usando un recubrimiento tal como lecitina, por el mantenimiento del tamaño de partículas requerido en el caso de dispersión y usando tensioactivos.

5 La prevención de la acción de microorganismos se puede lograr por diversos antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido ascórbico, timerosal y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares, polialcoholes, tales como manitol, sorbitol o cloruro sódico en la composición. Se puede provocar la absorción prolongada de las composiciones inyectables incluyendo en la composición un agente que retrase la absorción, por ejemplo, monoestearato de aluminio y gelatina.

10 En cualquier caso, las disoluciones inyectables estériles se pueden preparar incorporando un compuesto activo (por ejemplo, un polipéptido por sí mismo o en combinación con otros agentes activos) en la cantidad requerida en un disolvente apropiado con uno o una combinación de componentes enumerados en el presente documento, según se requiera, seguido por esterilización por filtración. En general, se preparan dispersiones incorporando el compuesto activo en un vehículo estéril, que contiene un medio de dispersión básico y los otros componentes requeridos de los enumerados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los métodos preferidos de preparación son secado a vacío y liofilización, que da un polvo de un principio activo más cualquier componente deseado adicional de una disolución previamente esterilizada por filtración de la misma. Las preparaciones para inyecciones se procesan, se envasan en recipientes tales como ampollas, bolsas, botellas, jeringas o viales, y se sellan en condiciones asépticas según métodos conocidos en la técnica. Además, las preparaciones se pueden envasar y comercializar en forma de un kit. Dichos artículos de fabricación tendrán preferentemente etiquetas o prospectos que indican que las composiciones asociadas son útiles para tratar un sujeto que padece, o predispuesto a trastornos de la coagulación.

25 La composición farmacéutica también se puede formular para administración rectal como un supositorio o enema de retención, por ejemplo, que contiene bases de supositorio convencionales tales como manteca de cacao u otros glicéridos.

30 Las dosis eficaces de las composiciones de la presente invención, para el tratamiento de afecciones, varían dependiendo de muchos factores diferentes, que incluyen medios de administración, sitio diana, estado fisiológico del paciente, si el paciente es humano o un animal, otras medicaciones administradas, y si tratamiento es profiláctico o terapéutico. Normalmente, el paciente es un humano, pero también se pueden tratar mamíferos no humanos que incluyen mamíferos transgénicos. Las dosis de tratamiento se pueden valorar usando métodos rutinarios conocidos por los expertos en la técnica para optimizar la seguridad y la eficacia.

35 Las dosis pueden variar de 1000 ug/kg a 0,1 ng/kg de peso corporal. En una realización, el intervalo de dosis es 1 ug/kg a 100 ug/kg. El polipéptido FVIII o la molécula de ácido nucleico optimizada que codifica el polipéptido FVIII se puede administrar continuamente o en intervalos programados específicos. Se pueden emplear ensayos *in vitro* para determinar intervalos de dosis óptima y/o programas para administración. Se conocen en la técnica ensayos *in vitro* que miden actividad del factor de coagulación. Además, se pueden extrapolar dosis eficaces de las curvas de dosis-respuesta obtenidas de los modelos animales, por ejemplo, un perro hemofílico (Mount et al. 2002, Blood 99 (8); 2670).

40 Las dosis intermedias en los intervalos anteriores también pretenden estar dentro del alcance de la invención. A los sujetos se pueden administrar dichas dosis diariamente, en días alternos, semanalmente o según cualquiera otro programa determinado por análisis empírico. Un tratamiento a un modo de ejemplo implica la administración en múltiples dosis durante un periodo prolongado, por ejemplo, de al menos seis meses. En algunos métodos, se pueden administrar simultáneamente dos o más polipéptidos, en cuyo caso la dosis de cada polipéptido administrado entra dentro de los intervalos indicados.

45 Se pueden administrar los polipéptidos FVIII o las moléculas de ácidos nucleicos optimizadas que codifican los polipéptidos FVIII en múltiples ocasiones. Los intervalos entre dosis únicas pueden ser diariamente, semanalmente, mensualmente o anualmente. Los intervalos también pueden ser irregulares como se indica midiendo niveles en sangre de polipéptido modificado o antígeno en el paciente. Alternativamente, los polipéptidos se pueden administrar como una formulación de liberación sostenida, en cuyo caso se requiere administración menos frecuente. La dosis y frecuencia varían dependiendo de la semivida del polipéptido o polinucleótido en el paciente.

55 La dosis y la frecuencia de administración pueden variar dependiendo de si el tratamiento es profiláctico o terapéutico. En aplicaciones profilácticas, las composiciones que contienen el polipéptido FVIII o la molécula de ácido nucleico optimizada que codifica el polipéptido FVIII o una mezcla de los mismos se administran a un paciente que no está todavía en el estado de enfermedad para potenciar la resistencia del paciente o minimizar los efectos de la enfermedad. Se define que dicha cantidad es una "dosis eficaz profiláctica". Se administra una dosis relativamente baja a intervalos relativamente poco frecuentes durante un largo periodo de tiempo. Algunos pacientes continúan recibiendo tratamiento para el resto de su vida.

Los polipéptidos FVIII o las moléculas de ácidos nucleicos optimizadas que codifican los polipéptidos FVIII se pueden administrar opcionalmente en combinación con otros agentes que son eficaces en el tratamiento del trastorno o afección en necesidad de tratamiento (por ejemplo, profiláctico o terapéutico).

5 Como se usa en el presente documento, la administración de polipéptidos FVIII o las moléculas de ácidos nucleicos optimizadas que codifican los polipéptidos FVIII conjuntamente o en combinación con una terapia complementaria significa la administración o aplicación secuencial, simultánea, coextensiva, simultánea, concomitante o contemporánea de la terapia y los polipéptidos desvelados. Los expertos en la técnica apreciarán que la administración o la aplicación de los diversos componentes de la pauta terapéutica combinada puede ser programada para potenciar la eficacia global del tratamiento. Un experto (por ejemplo, un médico) sería fácilmente capaz de discernir pautas terapéuticas combinadas eficaces sin excesiva experimentación basándose en la terapia complementaria seleccionada las enseñanzas de la presente memoria descriptiva.

15 Se apreciará adicionalmente que el polipéptido FVIII o la molécula de ácido nucleico optimizada que codifica el polipéptido FVIII se puede usar conjuntamente o en combinación con un agente o agentes (por ejemplo, para proporcionar una pauta terapéutica combinada). Los agentes a modo de ejemplo con los que se puede combinar un polipéptido o polinucleótido incluyen agentes que representan el actual tratamiento de referencia para un trastorno particular que está tratándose. Dichos agentes pueden ser de naturaleza química o biológica. El término "biológico" o "agente biológico" se refiere a cualquier agente farmacéuticamente activo hecho de organismos vivos y/o sus productos que está previsto para su uso como un terapéutico.

20 La cantidad de agente que se va a usar en combinación con los polinucleótidos o polipéptidos puede variar por sujeto o se pueden administrar según lo que se conoce en la técnica. Véase, por ejemplo, Bruce A Chabner et al., *Antineoplastic Agents*, en GOODMAN & GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287 (Joel G. Hardman et al., eds., 9ª ed. 1996). En otra realización, se administra una cantidad de dicho agente de acuerdo con el tratamiento de referencia.

25 Como se ha tratado previamente, los polinucleótidos y polipéptidos se pueden administrar en una cantidad farmacéuticamente eficaz para el tratamiento *in vivo* de trastornos de la coagulación. A este respecto, se apreciará que los polipéptidos o polinucleótidos se pueden formular para facilitar la administración y promover la estabilidad del agente activo. Preferentemente, las composiciones farmacéuticas según la presente invención comprenden un vehículo estéril farmacéuticamente aceptable, no tóxico, tal como solución salina fisiológica, tampones no tóxicos, conservantes y similares. Por supuesto, las composiciones farmacéuticas de la presente invención se pueden administrar en dosis únicas o dosis múltiples para proporcionar una cantidad farmacéuticamente eficaz del polipéptido.

30 Están disponibles varias pruebas para evaluar la función del sistema de coagulación: prueba del tiempo de tromboplastina parcial activado (TTPa), ensayo cromogénico, ensayo ROTEM®, prueba del tiempo de protrombina (TP) (también usado para determinar INR), prueba de fibrinógeno (frecuentemente por el método de Clauss), número de plaquetas, prueba de la función plaquetaria (frecuentemente por PFA-100), TCT, tiempo de sangrado, prueba de mezcla (si una anomalía corrige si el plasma del paciente se mezcla con plasma normal), ensayos de factor de coagulación, anticuerpos antifosfolípidos, dímero D, pruebas genéticas (por ejemplo, factor V Leiden, mutación de protrombina G20210A), tiempo del veneno de la víbora de Russell diluido (dRVVT), diversas pruebas de la función plaquetaria, tromboelastografía (TEG o Sonoclot), tromboelastometría (TEM®, por ejemplo, ROTEM®) o tiempo de lisis de euglobulinas (ELT).

40 La prueba de TTPa es un indicador del rendimiento que mide la eficacia de tanto las vías de coagulación "intrínsecas" (también denominadas la vía de activación por contacto) como las vías de coagulación comunes. Esta prueba se usa comúnmente para medir la actividad de coagulación de factores de coagulación recombinantes comercialmente disponibles, por ejemplo, FVIII o FIX. Se usa junto con el tiempo de protrombina (TP), que mide la vía extrínseca.

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

Terapia génica

50 La invención proporciona un método de aumento de la expresión de un polipéptido con actividad de factor VIII en un sujeto que comprende administrar la molécula de ácido nucleico aislada de la invención a un sujeto en necesidad del mismo, en donde la expresión del polipéptido es elevada con respecto a una molécula de ácido nucleico de referencia que comprende SEQ ID NO: 3. La invención también proporciona un método de aumento de la expresión de un polipéptido con actividad de factor VIII en un sujeto que comprende administrar un vector de la invención a un sujeto en necesidad del mismo, en donde la expresión del polipéptido es elevada con respecto a un vector que comprende una molécula de ácido nucleico de referencia.

55 Se ha explorado la terapia génica somática como un posible tratamiento para la hemofilia A. La terapia génica es un tratamiento particularmente atractivo para la hemofilia debido a su posibilidad para curar la enfermedad mediante la producción endógena continua de FVIII tras una administración única de vector. La hemofilia A es muy apta para un

enfoque de sustitución génica debido a que sus manifestaciones clínicas son completamente atribuibles a la falta de un único producto génico (FVIII) que circula en cantidades mínimas (200 ng/ml) en el plasma.

5 Sería terapéuticamente beneficiosa una proteína FVIII que se puede producir *in vivo* en un mamífero, por ejemplo, un paciente humano, usando un enfoque de terapia génica para el tratamiento de una enfermedad hemorrágica o trastorno seleccionado del grupo que consiste en un trastorno hemorrágico de la coagulación, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en músculos, hemorragia oral, traumatismo, traumatismo de la cabeza, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intrabdominal, hemorragia intratorácica, fractura de hueso, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal y sangrado en la vaina del iliopsoas. En una realización, la enfermedad o trastorno hemorrágico es hemofilia. En otra realización, la enfermedad o trastorno hemorrágico es hemofilia A. Esto implica la administración de un ácido nucleico que codifica FVIII optimizado operativamente unido a secuencias de control de la expresión adecuadas. En ciertas realización, estas secuencias se incorporan en un vector viral. Los vectores virales adecuados para dicha terapia génica incluyen vectores adenovirales, vectores lentivirales, vectores baculovirales, vectores virales de Epstein Barr, vectores papovavirales, vectores virales de la variolovacuna, vectores virales del herpes simple y vectores de virus adeno-asociados (AAV). El vector viral puede ser un vector viral defectuoso en la replicación. En otras realizaciones, un vector adenoviral tiene una delección en su gen E1 o gen E3. Cuando se usa un vector adenoviral, el mamífero puede no exponerse a un ácido nucleico que codifica un gen marcador de selección. En otras realizaciones, las secuencias se incorporan en un vector no viral conocido por los expertos en la técnica.

20 Todos los diversos aspectos, realizaciones y opciones descritos en el presente documento se pueden combinar en todas y cada una de las variaciones.

Habiendo descrito, en general, la presente invención, se pueden obtener un entendimiento adicional como referencia a los ejemplos proporcionados en el presente documento. Estos ejemplos son para los fines de ilustración solo y no pretenden ser limitantes.

EJEMPLOS

- 25 Se diseñaron secuencias de FVIII BDD optimizado por dos codones con los siguientes objetivos:
1. Retirar todas las secuencias de la región de tipo unión a matriz (MAR) (ATATTT y AAATAT; SEQ ID NO: 5 y 6, respectivamente);
 2. Retirar todas las secuencias desestabilizantes (ATTTA, SEQ ID NO: 8 y TAAAT, SEQ ID NO: 9);
 3. Retirar las secuencias de unión a promotor (TATAA, SEQ ID NO: 12 y TTATA, SEQ ID NO: 13);
 - 30 4. Retirar los elementos de secuencias ricos en AU (ARE): ATTTTATT (nucleótido 2468) y ATTTTTAA (nucleótido 3790) (SEQ ID NOs: 14 y 15, respectivamente);
 5. Añadir secuencia de Kozak (GCCGCCACCATGC, lo subrayado indica el codón de iniciación de la traducción; SEQ ID NO: 16) para aumentar el inicio traduccional;
 6. Ajustar los sitios de restricción para facilitar la clonación;
 - 35 7. Adaptar el uso de codones a la preferencia codónica de genes de *Homo sapiens*;
 8. Ajustar para evitar regiones de contenido de GC muy alto (>70 %) o bajo (<30 %), que pueden aumentar la estabilidad de ARN o prolongar la semivida de ARN.

Ejemplo 1: Optimización por codones por GENSCRIPT OPTIMUMGENE™

40 Se optimizó la secuencia de nucleótidos de FVIII BDD por codones usando la tecnología de optimización por codones GENSCRIPT OPTIMUMGENE™ (GenScript Corp., New Jersey, EE. UU.). La tecnología de optimización por codones GENSCRIPT OPTIMUMGENE™ se describe en Burgess-Brown et al., Protein Expr Purif. 59(1):94-102 (2008).

Se usaron los siguientes datos de uso de codones humanos para la optimización:

CODÓN	AMINOÁCIDO	FRACCIÓN	FRECUENCIA/MIL
TTT	F	0,45	16,9
TCT	S	0,18	14,6
TAT	Y	0,43	12,0
TGT	C	0,45	9,9

ES 2 848 703 T5

CODÓN	AMINOÁCIDO	FRACCIÓN	FRECUENCIA/MIL
TTC	F	0,55	20,4
TCC	S	0,22	17,4
TAC	Y	0,57	15,6
TGC	C	0,55	12,2
TTA	L	0,07	7,2
TCA	S	0,15	11,7
TAA	*	0,28	0,7
TGA	*	0,52	1,3
TTG	L	0,13	12,6
TCG	S	0,06	4,5
TAG	*	0,20	0,5
TGG	W	1,00	12,8
CTT	L	0,13	12,8
CCT	P	0,28	17,3
CAT	H	0,41	10,4
CGT	R	0,08	4,7
CTC	L	0,20	19,4
CCC	P	0,33	20,0
CAC	H	0,59	14,9
CGC	R	0,19	10,9
CTA	L	0,07	6,9
CCA	P	0,27	16,7
CAA	C	0,25	11,8
CGA	R	0,11	6,3
CTG	L	0,41	40,3
CCG	P	0,11	7,0
CAG	C	0,75	34,6
CGG	R	0,21	11,9
ATT	I	0,36	15,7
ACT	T	0,24	12,8
AAT	N	0,46	16,7
AGT	S	0,15	11,9
ATC	I	0,48	21,4
ACC	T	0,36	19,2
AAC	N	0,54	19,5

CODÓN	AMINOÁCIDO	FRACCIÓN	FRECUENCIA/MIL
AGC	S	0,24	19,4
ATA	I	0,16	7,1
ACA	T	0,28	14,8
AAA	K	0,42	24,0
AGA	R	0,20	11,5
ATG	M	1,00	22,3
ACG	T	0,12	6,2
AAG	K	0,58	32,9
AGG	R	0,20	11,4
GTT	V	0,18	10,9
GCT	A	0,26	18,6
GAT	D	0,46	22,3
GGT	G	0,16	10,8
GTC	V	0,24	14,6
GCC	A	0,40	28,5
GAC	D	0,54	26,0
GGC	G	0,34	22,8
GTA	V	0,11	7,0
GCA	A	0,23	16,0
GAA	E	0,42	29,0
GGA	G	0,25	16,3
GTG	V	0,47	28,9
GCG	A	0,11	7,6
GAG	E	0,58	40,8
GGG	G	0,25	16,4

Se ajustó el uso de codones al sesgo humano con el índice de adaptación de codones (IAC) humanos cambiando desde 0,75 (FVIII BDD natural) hasta 0,88 (FVIII BDD optimizado por GenScript). El contenido de G/C aumentó desde 46,16 % hasta 51,56 %. Se retiraron los picos de contenido de G/C en una ventana de 60 pb. La secuencia resultante de FVIII BDD optimizado por GenScript se desvela en el presente documento como SEQ ID NO: 1 y se muestra en la Figura 2.

5

Ejemplo 2: Optimización por codones por GENEART® GNEOPTIMIZER®

La secuencia de nucleótidos de FVIII BDD se optimizó por codones usando el software GENEART® GNEOPTIMIZER® (Invitrogen LIFE TECHNOLOGIES™ Corp., Grand Island, NY). La tecnología de optimización por codones GENEART® GNEOPTIMIZER® se describe en Graf et al., J. Virol. 74(22):10822-10826 (2000).

10 Se ajustó el uso de codones al sesgo humano con el índice de adaptación de codones (IAC) humanos cambiando desde 0,75 (FVIII BDD natural) hasta 0,96 (FVIII BDD optimizado por GeneArt). El contenido de G/C aumentó desde 46,16 % hasta 59 %. La secuencia resultante de FVIII BDD optimizado por GeneArt se desvela en el presente documento como SEQ ID NO: 2 y se muestra en la Figura 3.

Ejemplo 3: Construcciones de expresión

Todas las construcciones se hicieron en el esqueleto del vector Invitrogen pcDNA™4, que contiene un promotor inmediato-temprano del citomegalovirus humano (CMV), un potenciador de la traducción QBI SP163 y un gen ZEOCIN™ de resistencia para la selección.

pSYN-FVIII-066 conduce la expresión de FVIII BDD natural (SEQ ID NO: 3) en el esqueleto de pcDNA4 (Invitrogen).

5 pSYN-FVIII-116 conduce la expresión de FVIII BDD optimizado por codones (SEQ ID NO: 1) en el esqueleto de pcDNA4. La construcción deriva de pSYN-FVIII-066 reemplazando FVIII BDD natural con FVIII BDD optimizado por codones (SEQ ID NO: 1) usando sitios BsiWI y XhoI.

10 pSYN-FVIII-115 conduce la expresión de FVIII BDD optimizado por codones (SEQ ID NO: 2) en el esqueleto de pcDNA4. La construcción deriva de pSYN-FVIII-066 reemplazando FVIII BDD natural con FVIII BDD optimizado por codones (SEQ ID NO: 2) usando sitios BsiWI y XhoI.

Todas las construcciones se confirmaron por secuenciación de ADN.

Ejemplo 4: La optimización por codones mejora la expresión de FVIII en un ratón Hema

15 Para preguntar si cualquiera de las construcciones de FVIII BDD optimizado por codones da como resultado la elevada expresión de proteínas FVIII, se introdujeron plásmidos de expresión pSYN-FVIII116, pSYN-FVIII115 y el control natural pSYN-FVIII-066 en ratones Hema por inyección hidrodinámica. Posteriormente, se monitorizaron los niveles de expresión de FVIII en cada ratón inyectado por ensayos cromogénicos de FVIII de plasma.

La inyección hidrodinámica es un método no viral eficiente y seguro de administración de genes al hígado en animales pequeños, tales como ratones y ratas. La proteína de interés se produce en el hígado y se puede detectar en el plazo de 24 horas después de la inyección.

20 Se inyectaron por inyección intravenosa ratones Hema que pesaban 20-35 gramos en la vena de la cola con cualquiera de pSYN-FVIII116, pSYN-FVIII115 o el control natural pSYN-FVIII-066. Las inyecciones se constituyeron con 10 ug de ADN de plásmido desnudo libre de endotoxina en disolución al 0,9 % de solución salina estéril, dando un volumen total de 2 mL. Las inyecciones se realizaron rápidamente, necesiéndose no más de 4-7 segundos para inyectar la disolución completa de 2 mL de ADN. Los ratones se monitorizaron estrechamente durante dos horas después de la inyección, o hasta que se reanudó la actividad normal. 24 horas después de la inyección, las muestras se recogieron por extracción de sangre retrorbital, se preparó el plasma y se conservó a -80 °C para análisis adicionales.

30 La actividad de FVIII se midió usando el kit de FVIII COATEST SP de DiaPharma (lote N° N089019) y todas las incubaciones se realizaron en un calentador de placas a 37 °C con agitación. Los patrones de FVIIIr variaron desde 100 mUI/mL hasta 0,78 mUI/mL. Se añadieron un control de ensayo de mezclas de plasmas humanos normales y muestras de plasma (diluidas con tampón IX Coatest) en placas Immulon 2HB de 96 pocillos por duplicado (25 µL/pocillo). Se añadieron mezcla de IXa/FX/fosfolípido recién preparada (50 µL), 25 µL de CaCl₂ 25 mM y 50 µL de sustrato de FXa uno detrás de otro a cada pocillo con una incubación de 5 minutos entre cada adición. Después de la incubación con el sustrato, se añadieron 25 µL de ácido acético al 20 % para terminar la reacción de color y se midió la absorbancia a DO405 con un instrumento SpectraMAX plus (Molecular Devices). Los datos se analizaron con el software SoftMax Pro (versión 5.2). El nivel más bajo de cuantificación (LLOQ) es 7,8 mUI/mL. Los resultados se muestran en la Figura 7.

40 Se detectan bajos niveles de actividad de FVIII BDD tras la administración del plásmido de expresión de FVIII pSYN-FVIII-066 (control de FVIII BDD natural). La actividad de FVIII BDD promedio en ratones de control es aproximadamente 4-5 UI/mL (Figura 7, círculos). A diferencia, en promedio, se observa un aumento de aproximadamente tres veces en la actividad de FVIII BDD en el plasma de ratones administrados con pSYN-FVIII115 o pSYN-FVIII116 optimizado por codones (Figura 7, cuadrados y triángulos). Por tanto, la optimización por codones de FVIII BDD por los enfoques descritos anteriormente mejora la expresión de FVIII en un modelo de ratón Hema.

ES 2 848 703 T5

SECUENCIAS

SEQ ID NO: 1 - FVIII BDD optimizado

CGTACGGCCGCCACCATGCAGATTGAGCTGTCTACTTGCTTTTTTCCCTGCGCCTGCTGAGGTTTGCTTTCCGCTACACG
AAGGTATTATCTGGGGCTGTGGAACGTCTTTGGGATTACATGCAGAGTGACCTGGGAGAGCTGCCAGTGGACGCAAGGT
TTCCCCCTAGAGTCCCTAAGTCATTCCCCTTCAACACTAGCGTGGTCTACAAGAAAACACTGTTTCGTGGAGTTTACTGAT
CACCTGTTCAACATCGCAAAGCTAGGCCACCCTGGATGGGACTGCTGGGGCCAACAATCCAGGCCGAGGTGTACGACAC
CGTGGTCATTACACTTAAGAACATGGCCTCACACCCCGTGAGCCTGCATGCTGTGGGCGTCAGCTACTGGAAGGCTTCCG
AAGGAGCAGAGTATGACGATCAGACTTCCCAGAGAGAAAAGAGGACGATAAAGGTGTTTCCCTGGCGGATCTCATACTAC
GTGTGGCAGGTCCTGAAAGAGAATGGCCCTATGGCCTCCGACCCTCTGTGCCTGACCTACTCTTATCTGAGTCACGTGGA
CCTGGTCAAGGATCTGAACAGCGGCCCTGATCGGAGCCCTGCTGGTGTGCAGGGAAGGAAGCCTGGCTAAGGAGAAAACCC
AGACACTGCATAAGTTTATTCTGCTGTTTCGCGGTGTTTGACGAAGGAAATCAAGGGCACAGCGAGACAAAGAATAGTCTG
ATGACAGGACAGGGATGCCGCTTCAGCCAGAGCTTGGCCAAAATGCACACTGTGAACGGCTACGTCGAATCGCTCACTGCC
TGGGCTGATCGGCTGCCACCAGAGCGTGTATTGGCATGTCACTCGGGATGGGCACCACACTGAAGTGCACCTCCATTT
TCCCTGGAGGGACATACCTTTCTGGTCCGCAACCACCGACAGGCTTCCCTGGAGATCTCTCCAATTACCTTCTGACAGCA
CAGACTCTGCTGATGGACCTGGGGCAGTTCCTGCTGTTTTGGCCACATCAGCTCCCACCAGCATGATGGCATGGAGGCTTA
CGTAAAAGTGGACTCTGTCCCGAGGAACCTCAGCTGCGGATGAAGAACAATGAGGAAGCAGAAGACTATGACGATGACC
TGACCGACTCCGAGATGGATGTGGTCCGATTTCGATGACGATAACAGCCCTCCTTTATCCAGATTAGATCTGTGGCCAAG
AAACACCCTAAGACATGGGTCCATTACATCGCAGCCGAGGAAGAGGACTGGGATTAAGCACCCTGGTGTGGCACCAGA
CGATCGCTCCTACAAATCTCAGTATCTGAACAATGGGCCACAGAGGATTGGCAGAAAGTACAAGAAAGTGGGTTTCATGG
CATATACCGATGAGACCTTCAAGACTCGCGAAGCCATCCAGCACGAGAGCGGCATCCTGGGACCCTGCTGTACGGAGAA
GTGGGAGACACCTGCTGATCATTTCGAAGAACCAGGCCAGCCGGCCTTACAATATCTATCCACATGGGATTACAGATGT
GCGCCCTCTGTACAGCAGGAGACTGCCAAAGGGCGTCAAACACCTGAAGGACTTCCCAATCCTGCCCGGAGAAAATCTTCA
AGTACAAGTGGACTGTCAACCTCGAGGATGGCCCCACTAAGAGCGACCCTCGGTGCCTGACCCGCTACTAATCTAGTTTC
GTGAATATGGAAAGAGATCTGGCAAGCGGACTGATCGGACCCTGCTGATTTGTTACAAAGAGAGCGTGGATCAGAGAGG
CAACCAGATCATGTCCGACAAGCGGAATGTGATTCTGTTTCAGTGTCTTTGACGAAAACAGGTCATGGTACCTGACCGAGA
ACATCCAGAGATTCCTGCCTAATCCAGCTGGGGTGCAGCTGGAAGATCCTGAGTTTCAGGCATCTAACATCATGCATAGT
ATTAATGGCTACGTGTTCCAGAGTTTGCAGCTGAGCGTGTGCCTGCACGAGGTCGCTTACTGGTATATCCTGAGCATGG
GGCACAGACAGATTCCTGAGCGTGTCTTTTCCGGCTACACTTTAAGCATAAAATGGTCTATGAGGACACACTGACTC

ES 2 848 703 T5

TGTTCCCTTCAGCGGCGAAACCGTGTTTATGAGCATGGAGAATCCCGGACTGTGGATTCTGGGGTGCCACAACAGCGAT
TTCAGAAATCGCGAATGACTGCCCTGCTGAAAGTGTCAAGCTGTGACAAGAACACCGGGGACTACTATGAAGATTGATA
CGAGGACATCAGCGCATATCTGCTGTCCAAAAACAATGCCATTGAACCCCGGCTCTTTAGTCAGAATCCTCCAGTGCTGA
AGCGGCACCAGCGGAGATCACCCGACTACCTGCAGAGTGATCAGGAAGAGATCGACTACGACGATACAATTTCTGTG
GAAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGAGAACCAGAGTCTCGATCATTCCAGAAGAAAACAGGCA
TTACTTTATTGCCGAGTGGAGCGGCTGTGGGATATGGCATGTCTCTAGTCCCTCACGTGCTGCGAAATAGGGCCAGT
CAGGAAGCGTCCCACAGTTCAAGAAAGTGGTCTTCAGGAGTTTACAGACGGGTCCTTTACTCAGCCACTGTACAGGGC
GAACTGAACGAGCACCTGGGACTGCTGGGGCCCTATATCAGAGCAGAAGTGGAGGATAACATTTATGGTCACCTTCAGAAA
TCAGGCCCTCTCGCCCTTACAGTTTTTATTCAAGCCTGATCTCTTACGAAGAGGACCAGCGACAGGGAGCTGAACCACGAA
AAAACCTCGTGAAGCCTAATGAGACCAAAACATACTTTTGAAGGTGCAGCACCATATGGCCCCAACAAAAGACGAGTTC
GATTGCAAGGCATGGGCCATTTTTCTGACGTGGATCTGGAGAAGGACGTGCACAGTGGCCTGATTGGCCACTGTCTGGT
GTGCCATACTAACCCCTGAATCCAGCCCACGGCCGGCAGGTCACTGTCCAGGAGTTGCTCTGTTCTTTACCATCTTTG
ATGAGACAAAAGAGCTGGTACTTCACCGAAAACATGGAGCGAAATTCAGGGGCTCCATGTAACATTGAGATGGAAAGACCC
ACATTCAGAGGAGAATACCGCTTTTATGCTATCAATGGATACATCATGGATACTCTGCCCGGGCTGGTCATGGCACAGGA
CCAGAGAATCCGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAATATCCACTCAATTCATTTACCGGGCAGCTGTTTA
CTGTCAGGAAGAAAGAAGAGTACAAGATGGCCCTGTACAACCTGTATCCCGGCTGTTGAAACCGTCGAGATGCTGCC
AGCAAGGCCGGAATCTGGAGAGTGGAAATGCTGATTGGAGAGCACCTGCATGCTGGGATGCTTACCCTGTTTCTGGTGTA
CAGTAATAAGTGTGACACCCCTGGGAATGGCATCCGGGCATATCAGGGATTTCAGATTACCGCATCTGGACAGTACG
GACAGTGGGCACCTAAGCTGGCTAGACTGCATATTCGGATCTATCAACGCTTGGTCCACAAAAGAGCCTTTCTCTTGG
ATTAAGGTGGACCTGCTGGCCCAATGATCATTGATGGCATCAAACTCAGGGAGCTCGGCAGAAGTTCTCTCTCTGTGA
CATCTCACAGTTTATCATCATGTACAGCCTGGATGGGAAGAAATGGCAGACATACCGCGCAATAGCACAGGAACCTGTA
TGGTGTCTTTGGCAACGTGGACAGCAGCGGAATCAAGCACAACATTTTCAATCCCCCTATCATTGCTAGATAACATCCGG
CTGCACCCAACCCATTATTCTATTGAAAGTACACTGAGGATGGAAGTGTGGGATGCGATCTGAACAGTGTGTTCAATGCC
CCTGGGATGGAGTCCAAGGCAATCTCTGACGCCAGATFACCCTAGCTCCTACTTCACTAATATGTTTGTACTACCTGGA
GCCCTTCCAAAGCAAGACTGCACCTGCAAGGCCGAGCAACGCATGGCCACCACAGGTGAACAATCCCAAGGAGTGGTTG
CAGGTGATTTTCAGAAAACATGAAGGTGACCGGGGTCAACACTCAGGGCGTGAAGAGTCTGCTGACCTCAATGTACGT
CAAGGAGTTCCTGATCTCTAGTTTACAGGACGGACATCAGTGGACACTGTTCTTTTGAACCGGGAAGGTGAAGTCTTCC
AGGGCAATCAGGATTCCTTTACACCTGTGGTCAACAGTCTAGACCCCTCCACTGCTGACCAGATACCTGAGAATCCACCT
CAGTCTCGGTTGCACCAGATTCGCCCTGAGAATGGAAGTGTGGGATGCGAGGCCAGGATCTGTACTGATAACTCGAGTC
GACC

SEQ ID NO: 2 - FVIII BDD optimizado

GGCGGCCCGTACGGCCGCCACCATGCAGATCGAGCTGTCTACCTGCTTCTTCTGTGCCTGCTGCGGTTCTGCTTCAGC
GCCACCCGGCGGTAACCTGGGCGCGTGGAACTGAGCTGGGACTACATGCAGAGCGACCTGGGGGAGCTGCCCGTGGGA
CGCCAGATTCACCCCAAGAGTGCACCAAGAGCTTCCCTTCAACACCTCCGTGGTGTACAAGAAAACCTGTTCGTCGAGT
TCACCGACCACTGTCAATATCGCCAAAGCCAGACCCCTTGGATGGGCTGCTGGGCCCTACAATCCAGGECGAGGTG
TACGACACCGTGGTATCACCTTAAGAACAAGGCCAGCCACCCCGTGTCCCTGCACGCGGTGGGCGTGTCTACTGGAA
GGCCTCTGAGGGCGCTGAGTACGACGACAGCCAGCCAGCGCGAGAAAGAGGACGACAAAGTCTTCTGCGGCGAGCC
ATACCTACGTGTGGCAGGTCTGAAAGAAAACGGCCCTATGGCCTCCGACCCCTGTGCTGACCTACAGCTACCTGACC
CACGTGGACCTGGTCAAGGACCTGAACAGCGGCTGATTGGCCCTGCTCGTGTGTAGAGAGGGCAGCCTCGCCAAAGA
GAAAACCCAGACCTGCACAAGTTTATCCTGCTGTTGCGCGTCTTCGACAGGGCAAGAGCTGGCACAGCGAGACAAAGA
ACAGCCTGATGCAGGACCGGACCGCCCTCTGCCAGACCTGGCCTAAGATGCACACCGTGAACGGCTACGTGAACAGA
AGCCTGCCCGACTGATCGGCTGCCACCGAAGTCCGTGTACTGGCACGTGATGGCATGGGCACCCACCCCGAGGTGCA

ES 2 848 703 T5

CAGCATCTTTCTGGAAGGCCACACCTTCCTCGTGGGAACCACAGACAGGCCAGCCTGGAAATCAGCCCTATCACCTTCC
TGACCGCCAGACACTGCTGATGGACCTGGGCCAGTTCTGTGTTTTGCCACATCAGCAGCCACCAGCACGACGGCATG
GAAGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAACCCAGCTGCGGATGAAGAACAACGAGGAAGCCGAGGACTACGA
CGACGACCTGACCGACAGCGAGATGGACGTCGTGCGCTTCGACGACGACAACAGCCCCAGCTTCATCCAGATCAGAAGCG
TGGCCAAGAAGCACCCCAAGACCTGGGTGCACTATATCGCCGCCGAGGAAGAGGACTGGGACTACGCCCTCTGGTGCTG
GCCCCGACGACAGAAGCTACAAGAGCCAGTACCTGAACAATGCCCCCAGCGGATCGGCCGGAAGTACAAGAAAGTGG
GTTTCATGGCCTACACCGACGAGACATCAAGACCAGAGAGGCCATCCAGCACGAGAGCGGCATCCTGGCCCCCTGTGT
ATGGCGAAGTGGGCGACACCCCTGCTGATCATCTTCAAGAACCAGGCCAGCCGGCCCTACAACATCTACCCCCACGGCATC
ACCGACGTGCGGCCCTGTACAGCAGACGGCTGCCAAGGGCGTGAAGCACCTGAAGGACTTCCCATCTCTGCCCGGGA
GATCTTCAAGTACAAGTGGACCGTGACCGTGGAAAGTGGCCCCACCAAGAGCGACCCCAAGATGCCTGACCCGGTACTACA
GCAGCTTCTGCAACATGGAACGGGACCTGGCCCTCGGGCTGATCGGCCCTCTGTGATCTGTACAAAGAAGCGTGGAC
CAGCGGGCAACCAGATCATGAGCGACAAGCGGAACCGTGATCCTGTTGAGCGTGTTCGATGAGAATCGTCTCTGGTACCT
GACCGAGAATATCCAGCGGTTCTCTGCCCAACCCCTGCCGGCGTGCAGCTGGAAGATCCCGAGTTCAGGCCAGCAACATCA
TGCCTCCATCAATGGCTACGTGTTGACAGCCCTCCAGCTGAGCGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTG
AGCATCGGCCCCAGACCGACTTCTGAGCGTGTCTTTCAGCGCTACACCTTCAAGCACAAGATGGTGTACGAGGATAC
CCTGACCTGTTCCTTCTCGGCCGAAACCGTGTTCATGAGCATGGAAAACCCGGCCCTGTGGATTCTGGGCTGCCACA
ACAGCGACTTCAAAAACCGGGCATGACCGCCCTGCTGAAGGTGTCCAGCTGCGACAAGAACACCGGCACTACTACGAG
GACAGCTATGAGGACATCAGCGCCTACCTGCTGAGCAAGAACAACGCCATCGAGCCAGATCCTTACGCCAGAACCCCC
CCTGCTGAAGCGGCACCAGAGAGATCACCCGACACCCTGCGAGTCCGACCAGGAAGAGATTGATTACGACGACACCA
TCAGCCTCGAGATCAAGAAAGACCATTTGACATCTACGACGACCCAGACAGCAACACAGCCCCCCTCCTCCAGAAACAA
ACCCGGCACTACTTCAATGCGCCGTGGAAAGACTGTGGGACTACGGCATGAGCAGCAGCCCCACGTGCTGCGGAACAG
AGCCAGAGCGGCAGCGTGCCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTCACCGACGGCAGCTTACCCAGCCCCGT
ATCGGGCGAGCTGAACGAGCACCTGGGACTGCTGGGACCTTACATTAGAGCCGAGGTGAAGATAACATCATGGTCACC
TTCAGAAACCAGGCCCTCCAGACCTACAGCTTCTACAGCAGCCTGATCAGTACGAAGAGGACCAGCGGCAGGGCGCCGA
ACCCCGGAAGAACTTCTGTAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTGCAGCACACATGGCCCCACAAAAG
ACGAGTTCGACTGCAAGSCCTGGGCTACTTCTCCGATGAGGACCTGGAAGGACCTGCACTCTGGCCCTGATTGGACCT
CTGCTCGTCTGCCACACCAACCCCTGAACCCCGCCACGGCCGGCAGGTACAGTGCAGGAATTTGCCCTGTTCTTAC
CATCTTCGATGAGACAAAGAGCTGGTACTTACCGGAGAACATGGAAAGAACTGTAGAGCCCCCTGCAACATCCAGATGG
AAGTCTTACCTTCAAAGAGAACTATCGTTCCAGCCATCAACGGCTACATCATGACACCTGCCCCGGCTGTTGATG
GCCAGGATCAGAGAATCCGGTGGTATCTGCTGAGCATGGGACGCAACGAGAACATCCACAGCATCCACTCAGCGGCCA
CGTGTTCACAGTGGGGAAGAAAGAGAGTACAAGATGGCCCTGTACAACCTGTACCCCGCGCTGTTTCGAGACAGTGGAAA
TGCTGCCAGCAAGGCCGGCATCTGGCGGTGGAATGTCTGATCGCCGAGCATCTGCACGCCGAATGAGCACCTGTTT
CTGGTGTACAGCAACAAGTGCCAGACCCCTCTGGGCATGGCCAGCGGCCACATCCGGGACTTCCAGATCACCGCCTCCGG
CCAGTACGGCCAGTGGGCCCTAAGCTGGCCGGCTCCACTACTCCGGATCTATCAACGCCCTGGTCCACCAAGAGCCCT
TCAGCTGGATCAAGGTGGACCTGCTGGCCCTATGATCATCCACGGAATCAAGACCCAGGGCCGACAGAGAAGTTCAGC
AGCCTGTACATCAGCCAGTTGATCATCATGTACAGCCTGGACGGCAAGAAAGTGGCAGACCTACCCGGGCAACAGCACCGG
CACCTGATGGTGTCTTTCGGCAACGTGGACAGCAGCGGCATCAAGCACAAACATCTTCAACCCCCATCATTTGCCGGT
ACATCCGGCTGCACCCACCCACTACAGCATCCGGTCCACCTGCGGATGGAACGTGATGGGCTGCGACCTGAACCTTTCG
AGCATGCCCTGGGGATGGAAAGCAAGGCCATCAGCGACGCCAGATCACAGCCAGCAGCTACTTCAACCAACATGTTCCG
CACCTGGTCCCCAAGCAAAGCCCGCTGCATCTCAAGGCAGAAGCAATGCCTGGCGGCTCAGGTCAACAACCCCAAG
AATGGCTCCAGTGGACTTTCAGAAAACCATGAAGGTCAACGGCGTGACCAACCCAGGGCGTGAAGGCTGCTGACCTCT
ATGTACGTGAAAGAGTCTCTGATCAGCAGCAGCCAGGACGGCCACAGTGGACCTCTTCTTTCAGAACGGCAAACTGAA
AGTGTTCAGGGCAACACAGGACTCCTTTACCCCGTGTCAACTCTCTAGACCTCCACTGCTGACCAGATACCTGAGAA
TCCACCTCAGTCTGGGTGCACCAGATGCCCCTGAGAATGGAAGTCTGGGATGCGAGGCCAGGATCTGTACTGATAA
CTCGAGTCGACTTAATTAA

SEQ ID NO: 3 – Secuencia de nucleótidos que codifica FVIII BDD "parental"

ATGCAAATAGAGCTCTCCACCTGCTTCTCTCTGTCCTTTTTCGATTCTGCTTTAGTGCCACCAGAACTACTACCTGGG
 TGCAAGTGGAACTGTTCATGGGACTATATGCAAAGTGATCTCGGTGAGCTGCCTGTGGACGCAAGATTTCCCTCTAGAGTGC
 CAAAATCTTTTCCATTCAACACCTCAGTCGTGTACAAAAGACTCTGTTTGTAGAATTCACGGATCACCTTTTCAACATC
 GCTAAGCCAAGGCCACCCTGGATGGGTCTGCTAGETCTACCATCCAGGCTGAGGTTTATGATACAGTGGTCATTACACT
 TAAGAACATGGCTTCCCATCTGTCTAGTCTTCACTGCTGTGGTGTATCCTACTGGAAAGCTTCTGAGGGAGCTGAATATG
 ATGATCAGACCAGTCAAAGGAGAAAGAAGATGATAAAGTCTTCCCTGGTGGAAAGCCATACATATGTCTGGCAGGTCTTG
 AAAGAGAATGGTCCAAATGGCTCTGACCCACTGTGCCTTACCTACTCATATCTTTCTCAATGGACCTGGTAAAAGACTT
 GAATTCAGGCCCTCATTTGGAGCCCTACTAGTATGTAGAGAAGGGAGTCTGGCCAAGGAAAAGACACAGACCTTGCACAAAT
 TTTACTACTTTTGTCTGTATTTGATGAAGGAAAAGTTGGCACTCAGAAAACAAAGAACTCCTTGATGCAGGATAGGGAT
 GCTGCATCTGCTCGGGCCTGGCCTAAAAATGCACACAGTCAATGGTTATGTAAACAGGTCCTCTGCCAGGTCTGATTGGATG
 CCACAGGAAATCACTCTATTGGCATGTGATTGGAAATGGGCACCCTCCTGAAAGTGCACCTCAATATTCTCGAAGGTCACA
 CATTTCTGTGAGGAACCATCGCCAGGCCCTCTGGAAATCTCGCCAATAACTTTCCCTTACTGCTCAAACACTCTTGATG
 GACCTTGGACAGTTTCTACTGTTTGTCTATATCTCTTCCCACCAACATGATGGCATGGAAGCTTATGTCAAAGTAGACAG
 CTGTCCAGAGGAACCCGAATACGAATGAAAAATAATGAAGAAGCGGAAGACTATGATGATGATCTTACTGATTCTGAAA
 TGGATGTGGTCAGGTTTGTGATGACAACTCTCTCTCTTTATCCAAATTCGCTCAGTTGCCAAGAAGCATCTTAAACT
 TGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTCTCGCCCCGATGACAGAAGTTATAA
 AAGTCAATATTTGAACAATGGCCCTCAGCGGATGGTAGGAAGTACAAAAAGTCCGATTTATGGCATAACAGATGAAA
 CCTTTAAGACTCGTGAAGCTATTGAGCATGAATCAGGAACTTGGGACCTTTACTTTTATGGGGAAGTTGGAGACACACTG
 TTGATTATATTTAAGAAATCAAGCAAGCAGACCATAACATCTACCTCACGGAATCACTGATGTCCGCTCTTTGTATTCT
 AAGGAGATTACAAAAGGTGTAAAACATTTGAAGGATTTTCCAAATCTGCGCAGGAGAAATATTCAAATATAAATGGACAG
 TGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGGTGCTGACCCGCTATTACTCTAGTTTCGTTAATATGGAGAGA
 GATCTAGCTTCAGGACTCATTTGGCCCTCTCTCATCTGCTACAAAGAATCTGTAGATCAAAGAGGAAACCCAGATAATGTC
 AGACAAGAGGAATGTATCTGTTTTCTGTATTTGATGAGAACCAGGCTGGTACCTCACAGAGAATATAACACGCTTTC
 TCCCCAATCCAGCTGGAGTGCAGCTTGGAGTCCAGAGTCCCAAGCCTCCAACATCATGCACAGCATCAATGGCTATGTT
 TTTGATAGTTTGCAGTTGTGAGTTGTTTGCATGAGGTTGGCATACTGGTACATTTCTAAGCATTTGGAGCACAGACTGACTT
 CCTTTCTGTCTCTCTCTGGATATACCTTCAAACACAAAAATGGTCTATGAAGACACACTCACCCCTATTCCCATCTCAG
 GAGAAACTGCTTTCATGTGATGGAAAACCCAGGCTATGGATTCTGGGGTGCCACAACCTCAGACTTTCCGGAACAGAGGC
 ATGACCCGCTTACTGAAGGTTTCTAGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGATATTTGAGC
 ATACTTGTGACTAAAACAATGCCATTTGAACCAAGAAGCTTCTCTCAAACCCACCAGTCTTGAACGCCATCAACGGG
 AAATAACTCGTACTACTCTTCACTCAGATCAAGAGGAAATGACTATGATGATACCATATCACTTCAAATCAAGAAGGAA
 GATTTTGACATTTATGATGAGGATGAAAATCAGAGCCCCCGCAGCTTTCAAAGAAAACACGACACTATTTTATGCTGC
 AGTGGAGAGGCTCTGGGATTTATGGGATGAGTAGCTCCCCACATGTTCTAAGAAACAGGGCTCAGAGTGGCAGTGTCCCTC
 AGTTCAAGAAAGTTGTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATACCGTGGAGAACTAAATGAACAT
 TTGGSACTCCTGGGGCCATATATAAGAGCAGAAGTTGAAGATAATATCATGGTAACTTTAGAAATCAGGCCCTCTCCTCC
 CTATTCTTCTATTTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGAGCAGAACCTAGAAAAAATTTGTCAAGC
 CTAATGAAACAAAACCTTACTTTTGGAAAAGTGAACATCATATGGCACCCACTAAAGATGATTTGACTGCAAAGCCTGG
 GCTTATTTCTCTGATGTTGACCTGGAAAAGATGTGCACCTCAGGCCCTGATTTGGACCCCTTCTGGTCTGCCACACTAACAC
 ACTGAACCCCTGCTCATGGGAGACAAGTGCAGTACAGGAATTTGCTCTGTTTTTACCATCTTTGATGAGACCAAAGCT
 GGTACTTCACTGAAAATATGGAAAAGAACTGCAGGGCTCCTGCAATATCCAGATGGAGATCCCCTTTTAAAGAGAAT

ES 2 848 703 T5

TATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGCTTAGTAATGGCTCAGGATCAAAGGATTCGATG
GTATCTGCTCAGCATGGGCAGCAATGAAAACATCCATTTCTATTGATTTTTCAGTGGACATGTGTTCACTGTACGAAAAAAG
AGGAGTATAAAAATGGCACTGTACAATCTCTATCCAGGTGTTTTTGGACAGTGGAAATGTACCATCCAAAGCTGGAATT
TGGCGGTTGGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTTCTGGTGTACAGCAATAAGTGTCA
GACTCCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGACAATATGGACAGTGGCCCCAA
AGCTGGCCAGACTTCATTATCCGGATCAATCAATGCCTGGAGCACCAAGGAGCCCTTTTCTGGATCAAGGTGGATCTG
TTGGCACCATGATTATTCACGGCATCAAGACCCAGGGTCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTAT
CATCATGTATAGTCTTGTATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCTGGAACCTTAATGGTCTTCTTTGGCA
ATGTGGATTTCATCTGGGATAAAAACACAATATTTTAAACCTCCAATTATGCTCGATACATCCGTTTGCACCAACTCAT
TATAGCATTTCGCAGCACTCTTCGCATGGAGTGTATGGGCTGTGATTTAABTAGTTGCAGCATGCCATTTGGGAATGGAGAG
TAAAGCAATATCAGATGCACAGATTACTGCTTCACTACTTTACCAATATGTTTGGCCACTGGTCTCCTTCAAAGGCTC
GACTTCACCTCCAAGGGAGGAGTAATGCCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAG
AAGACAATGAAAGTCACAGGAGTAATACTACTCAGGGAGTAAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCTCAT
CTCCAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTTTCAGAATGGCAAAGTAAAGGTTTTTCAGGCAATCAACACT
CCTTACACCTGTGGTGAAGTCTCTAGACCCACCGTTACTGACTCGCTACCTTCGAATTCACCCCAAGGTTGGGTGCAC
CAGATTGCCCTGAGGATGGAGGTTCTGGGCTGCGAGGCACAGGACCTCTAC

SEQ ID NO: 4 – Secuencia de aminoácidos de FVIII BDD

ATRRYYLGAVELSWDYMOSDLGELPVDARFPPRPVKSPFPNTSVVYKTLFVEFTDHLFNIKPRPPWMLLGPITQAEV
YDVTVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLS
HVDLVKDLMSGLIGALLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNLSMQDRDAASARAWPKMHTVNGYVNR
SLPGLIGCHRKSVYWHVIGMGTTPVHVSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLLMDLGQFLLFCHISSHQHDM
EAYVKVDSCEPEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHEKTVWHYIAAEEEDWDYAPLVL
APDDRSYKSCYLNNGPQRIGRKYKVRFMAYTDETFKTR&A;TQHESGILGPLLYGEVGDLLLIIFKNQASRPYNIYPHGI
TDVREPLYSRRLPKGVKHLKDFEILPGEIFKYKWTVTVEDEGPTKSDPRCLTRYYSSEFVNMERDLASGLIGPLLYCYKESVD
QRGNQIMSDKRNVLFSVFDENRSWYLTEINIQRFLEPNPAGVQLEDEPEFQASNIMHSINGYVFDLQLSVCLHEVAYWYIL
SIGAQTDPLSVFFSCYTFKHKMYEDTLLTFPESGETVFMSENPGLWILGCHNSDFRNRMGNTALLKVVSSCDKNTGDYEE
DSYEDISAYLLSKNNAIEPRSFQNPVVKRHRQREITRITLQSDQEEIDYDDTISVEMKKEDEFDIYDEDENQSPRSFQKK
TRHYFIAAVERLWYGMSSSPHVLNRNAQSGSVPEFKKVVFEFTDGSFTQPLYRGELEHLGLLGPYIRAEVEDNIMVT
FRNQASRPYSFYSSLSIYEEEDQRQGAEPKRNFKVNETKTYFWKVQHMAPTKDEFCKAWAYFSDVDLEKDVHSGLIGP
LLVCHTNTLNPAGHRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFAHNGYIMDTLPGLVM
AQDQRIRWYLLSMGNSNENIHSIH&S;SGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLF
LVYSNKCQTPLGMASGHIRDFQITASGQYQWAPKLARLHYSINAWSTKEPEFWIKVDLLAPMIHGEKTKQARQKFS
SLYISQFIIMYSLDCKKWQTYRGNSTGTLMVFFGNVDSGKHNIFNPPIIARYIRLHPHYSIRSTLRMELMGCDLNSC
SMPLGMESKAISDAQ&T;ASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLLS
MYVKEFLISSQDGHQWTLFFQNGKVKVQGNQDSFTFVNSLDPELLTRYLRIRHPQSWVHQIALRMEVLGCEAQDLY

SEQ ID NO: 5 - Secuencia de nucleótidos MAR/ARS

5 ATATTT

SEQ ID NO: 6 - Secuencia de nucleótidos MAR/ARS

AAATAT

SEQ ID NO: 7 - Posible sitio de corte y empalme

GGTGAT

10 SEQ ID NO: 8 - Secuencia desestabilizante

ATTTA

SEQ ID NO: 9 - Secuencia desestabilizante

TAAAT

SEQ ID NO: 10 - Secuencia de poli-T

TTTTTT

SEQ ID NO: 11 - Secuencia de poli-A

AAAAAAA

SEQ ID NO: 12 - Sitio de unión al promotor

5 TATAA

SEQ ID NO: 13 - Sitio de unión al promotor

TTATA

SEQ ID NO: 14 - Elementos de secuencia rica en AU (ARE)

ATTTTATT

10 SEQ ID NO: 15 - Elementos de secuencia rica en AU (ARE)

ATTTTTAA

SEQ ID NO: 16 - Secuencia consenso de Kozak

GCCGCCACCATGC

SEQ ID NO: 17 - Péptido CTP

15 DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPIL

SEQ ID NO: 18 - Péptido CTP

SSSSKAPPPSLPSPSRLPGPSDTPILPQ

SEQ ID NO: 19 - Secuencia central de péptidos de unión a albúmina

DICLPRWGCLW

20 SEQ ID NO: 20 - Secuencia PAS

ASPAAPAPASPAAPAPSAPA

SEQ ID NO: 21 - Secuencia PAS

AAPASPAPAAPSAPAPAAPS

SEQ ID NO: 22 - Secuencia PAS

25 APSSPSPSAPSSPSPASPSS

SEQ ID NO: 23 - Secuencia PAS

APSSPSPSAPSSPSPASPS

SEQ ID NO: 24 - Secuencia PAS

SSPSAPSPSSPASPSPPA

30 SEQ ID NO: 25 - Secuencia PAS

AASPAAPSAPPAASPAAPSAPPA

SEQ ID NO: 26 - Secuencia PAS

ASAAAPAAAASAAASAPSAAA

SEQ ID NO: 27 - Longitud completa de la secuencia de nucleótidos del factor de von Willebrand

ES 2 848 703 T5

ATGATTCCTGCCAGATTGCGGGGTGCTGCTTGGCTCTGECCTCATTTTGGCCAGGGACCCTTTGTGCAGAAGGAACCTCGGGCAG
GTCAITCCACGGGCCACTAAGGACGGTCTAAACGGCCCCACGACGAACAGACCGGGAGTAAAACGGTCCCTGGGAAACACGTCTT
CCTTGAGCCCGCTCCAGTAGGTGCCGGGGATGCAGCCTTTCCGGAAGTGAATTCGTCAACACCTTTGATGGGAGCATGTACAGCTT
TGCGGGATACTGCAGTTACCTCCTGGCAGGGGGCTGCCAEEACTACGTCCGAAAAGCCTTCACTGAAGCAGTTGTGGAAACTACCC
TCGTACATGTCGAAACGCCCTATGACGTCAATGGAGGACCGTCCCCGACGGTCTTACGCTCCITCTCGATTATTGGGGACTTCCA
GAATGGCAAGAGAGTGAACCTCTCCCTGTATCTTGGGAAATTTTTTGACATCCATTTGTTTGTCAATGGTTGCCGAGGAAGAGCTAA
TAACCCCTGAAGGTCTTACCTTCTCTCACTCGGAGAGGCACATAGAACCCTTAAAAAAGTGTAGGTAAACAAACAGTTACCCAC
CGTGACACAGGGGGACCAAACAGTCTCCATGCCCTATGCCCTCCAAAGGGCTGTATCTAGAAACTGAGGCTGGTACTACAAGCTGT
CCGGTCAGGCCTTGGCACTGTGTCCCTCGTTCCTCAGAGGTACGGGATACGGAGGTTTCCCGACATAGATCTTTGACTCCGACC
CATGATGTCGACAGGCCACTCCGGAATGGCTTTGTGGCCAGGATCGATGGCAGCGGCAACTTCAAGTCTGTGTGACAGAGAT
ACTTCAACAAGACCTGCGGGCTGTGTGGCAACTTTAACATTAACGAAACACCGGTCTTAGCTACCGTCCGCGTGAAGTTGAGGA
CGACAGTCTGTCTATGAAGTTGTTCTGGACGCCGACACACCGTTGAAATTGTACTTTGCTGAAGATGACTTTATGACCCAAAGAG
GGACCTTGACCTCGGACCCCTTATGACTTTGCCAACTCATGGCTCTGAGCAGTGGAGAACAGTGGTGTGAAACGACTTCTACTGAA
ATACTGGGTTCTCCCTCGAAGTGGAGCCTGGGAATACTGAAACCGTTGAGTACCGGAGACTCCTCACCTCTGTCAACACAGAAC
GGGCATCTCCTCCAGCAGCTCATGCAACATCTCCTCTGGGAAATGCAGAAGGGCCGTGGGAGCAGTCCAGCTTCTGAAGAC
ACCTCGGTGCTTGGCCGTAGAGGAGGGTGTGAGTACGTTGTAGAGGAGACCCCTTTACGTCCTCCCGGACACCCCTCGTCCCG
TCGAAGACTTCTCGTGGAGCCACATTTGCCCGCTGCCACCCTGTGGTGSACCCGAGCCTTTTGTGGCCCTGTGTGAGAAGACTTTG
TGTGAGTGTGCTGGGGGCTGAGTGGCCCTGCCCTGCAACGGCGAGGGTGGGAGACCACTGGGGCTCGGAAAACACCGGGACA

CACYCTTCTGAAACACACTCACACGACCCCCGACCTCACGGGACGGGACGCCTCCTGGAGTACGCCCCGACCTGTGCCAGGAG
GGAAATGGTGTGTACGGCTGGACCGACCCACAGCGCGTGCAGCCAGTGTGCCCTGCTGGTATGGAGGGAGGACCTCATGCCGGCCT
GGACACGGGTCTCCCTTACACAGCATGCCAGCTGGCTGGTGTGCGGCACGTGGGTACACAGGGACGACCATACCTCTATAGG
CAGTGTGTGTCCCCTTGGCCAGGACCTGCCAGAGCCTGCACATCAATGAAATGTGTGAGGAGCGATGCCGTGGATGGCTGCAGCTG
CCCTGAGGATATCCGTACACACAGGGGAACCGGGTCTGGACGGTCTCGGACGTGTAGTTACTTTACACAGTCTCTCGCTACGCAC
CTACCGACGTGCAGGGACTCCGACAGCTCCTGGATGAAGGCCTCTCGCTGGAGAGCACCCAGTGTCCCTGCCTGCATTCGGAAA
GGCTACCCCTCCCGACCTCCCTCTCTCGAGACTGCTGTGAGGACCTACTTCCGAGACGCACCTCTCGTGGCTCACAGGGACG
CACGTAAGGCCTTTCCGGATGGGAGGGCCGTGGAGGGAGAGACTGTGACCAACACCTGCATTTGCCGAAACAGCCAGTGGATCTG
CAGCAATGAAGAATGTCCAGGGGAGTGCCTTGTCACTGGTCAATCCACTTCAAGAGCTTTGACGTTGTGGACGTAAACGGCTTTG
TCGCTCACCTAGACGTGCTTACTTCTTACAGGTCCTCTCACGGAACAGTACCAGTTAGGGTGAAGTTCTCGAACTGAACAGATA
CTTACCTTACAGTGGGATCTGCCAGTACCTGTGCCCCGGGATTGCCAGGACCACTCCTTCTCCATTGTCATFGAGACTGTCCAGT
GTGCTGTTGTCTATGAAGTGGAACTACCCTAGACGGTCAATGGACGACCGGGCCCTAACCGTCTGGTGGAGAGAGGTAACAGTA
ACTCTGACAGGTACACAGACATGACCGGACGCTGTGTGCACCCCTCCGTACCGTCCGGCTGCCTGGCCGCACAAACAGCCTTG
TGAACTGAAGCATGGGGCAGGAGTTGCCATGGATACTGGCCCTGCGACACACGTGGCCGAGGGCAGTGGCAGGCCAGCGGACCGGA
CGTGTGTGCGAAACACTTGAAGTCTGCTACCCCTCCTCAACGGTACCTTGGCCAGGACATCCAGCTCCCCCTCCTGAAAGGTGACC
TCCGCATCCAGCATACAGTACCGGCTCCGTGCGCCTCAGTACGGGGAGGACCTGCAGATGACCGGTCTGTAGGTGCGAGGGGGA
GGACTTCCACTGGAGGCGTAGGTCGTATGTCACTGCCGAGGACCGCGGAGTGCATGCCCTCCTGGACGTACGACTGGGATG
GCCGCGGGAGGCTGCTGTGAAGCTGTCCCCCTCTAAGCCGGGAAACCTTGGCGCCGTGTGTGGAAATACAAATGGCAACAGGGC
GACGCTGACCTACCGGGCCCTCCGACGACCACTTGCACAGGGGACAGATACGGCCCTTCTGGACGCCGACACACCCCTTAATGT
TACCGTGTGTCCCGCTGCACTTCCCTTACCCCTCTGGGCTGGCRGAGCCCGGGTGGAGGACTTGGGAACGCTGGAAGCTGCAC
GGGACTGCCAGGACCTGCAGAAGCAGCACAGTGAAGGAATGGGGAGACCCGACCGYCTCGGGCCCACTCCTGAAGCCCTTGC
GGACTTCGACGTGCCCTGACGGTCTGGACGTCTCGTGTGTCCGATCCCTGCCCCCTCAACCCGCGATGACAGGTTCTCC
GAGGAGGCGTCCGCGTCTGACCTCCCCACATTCGAGGCTGCCATCGTCCCGTACGGCTAGGGACCGGGGAGTGGCGCGT
ACTGCTCCAAGAGGCTCCTCCGCACGCGCCAGGACTGCAGGGGTGTAAGCTCCGGACGGTAGCACGGCAGTCCCGCTGCCCTAC
CTGCGGAACTGCCGTACGACGTGTGCTCCTGCTCGGACGGCCGGAGTGCCTGTGCGGGCCCTGGCCAGCTATGCCGGGGCTG
CGGGCAGCGGATGGACGCTTGACGGCGATGCTGCACACGAGGACGAGCCTGCCGGCGCTCACGGACACCGCCGCGGACCGGTCC
ATACGGCGCGGACCGCGGGGAGAGGCGTGCCTGCGCTGGCGGAGCCAGGCCGCTGTGAGGTGAAGTGCAGGAAAGGCCAGGT
CTACCTGCAGTGGGGACCCCTGCAACCTGCCCTCTCCGCACGCGCAGCGCACCGGGCTCGGTCCGGCGACACTCGACTTGACC
GGCTTCCCGTCCACATGGAGCTCACGCCCTGGGGGACGTTGGAGACCAGCCGCTCTCTCTTACCCGGATGAGGAATGCAATGA
GGCTGCTGGAGGCTGCTTCTGCCCCAGGGCTCTACATGGATGAGAGGGGGACCTGGACGGCGAGAGAGAATGGCCCTA
CTCCTTACGTTACTCCGGACGGACCTCCCGACGAAGACGGGGGTCCCGAGATGTACCTACTCTCCCCCTGTGCGTGCACAGGC
CCAGTGGCCCTGTTACTATGACGGTGAATCTTCCAGCCAGAAGACATCTTCTCAGACCATCACACCATGCTACTGTGAGGATG
ACGCACGGGTCCGGGTACGGGGACAATGATACTGCCACTCTAGAAGTCCGTCTTCTGTAGAAGAGTCTGGTAGTGTGGTACAC
GATGACACTCCTACGCTTCACTGCACCTGTACCATGAGTGGAGTCCCGGAAGCTTGCCTGCCCTGACGCTGTCTCAGCACTCCCTGT
CTCATCCAGCAAAAGGAGCCTATCCTCCGAAGTACCTGACATGCTACTCACCTCAGGGCCCTCGAACACGGACTCCGACACGA
GTCTGTCAGGGGACAGAGTAGCGTCTGTTTCTCGGATAGGACTCGGCCCCCATGGTCAAGCTGGTGTGTCCCGCTGACAACCTGC
GGGCTGAAGGCTCGAGTGTACCAAAACCTGCCAGAACTATGACCTGGAGTGCATGACCCGGGGGTACCAGTTCGACCACACAGG
GGACTGTTGGACGCCGACTTCCCGAGCTCACATGGTTTTGACCGGTCTTGATACTGGACCTCACGTACAGCATGGGCTGTGTCT
CTGGCTGCCCTGCCCCCGGGCATGGTCCGGCATGAGAACAGATGTGTGGCCCTGGAAAGGTGTCCTGCTTCCATCAGGGCATC
GTACCCGACACAGAGACCGACGGAGACGGGGGCCCTTACCAGGCGTACTCTTGTCTACACACCGGGACCTTCCACAGGGACGA
AGGTAGTCCCGTAGGAGTATGCCCTGGAGAAACAGTGAAGATGGCTGCAACACTTGTGTCTGTCGGGACCGGAAGTGGAACTGC
ACAGACCATGTGTGTGATGCCACGTGTCTCATACGGGGACCTCTTTGTCACTTCTAACCCGACGTTGTGAACACAGACAGCCCTGG
CCTTACCTTGCAGTGTCTGGTACACACACTACGGTGCACCTCCAGATCGGCATGGCCACTACCTCACCTTGCAGGGGCTCAA
TACCTGTCCCCGGGAGTGCAGTACGTTCTGGTGCAGGATTACTGCCGAGTGAAGTGTAGCCGTACCGGCTATGGAGTGA

AGCTGCCCGAGTTTATGGACAAGGGGCCCTCACG9TCATGCAAGACCACGTCTAATGACGCCGTCAAACCCCTGGGACCTTTCGG
 ATCCTAGTGGGGAATAAGGGATGCAGCCACCCCTCAGTAAATGCAAGAAACGGGTACCATCCTGGTGGAGGGAGGAGATTGG
 GACCCTGAAAGCCTAGGATCACCCCTTATTCCTACGTCGGTGGGGAGTCACTTTACGTTCTTTGCCAGTGGTAGGACCACCTC
 CCTCTCTCTTTGAGCTGTTGACGGGGAGGTGAATGTGAAGAGGCCCATGAAGGATGAGACTCACTTTGAGGTGGTGGAGTCTGG
 CCGGTACATCATTCTGCTGCTGGGAACTCGACAAACTGCCCTCCACTTACACTTCTCCGGGTACTTCTACTCTGAGTGAAACTC
 CACCACCTCAGACCAGCCATGTAGTAAGACGACGACCCCAAAGCCCTCTCCGTGGTCTGGGACCGCCACCTGAGCATCTCCGTGGT
 CCTGAAGCAGACATACCAGGAGAAAGTGTGTGECCTGTGTGGAAATTTTGTATGTTTGGGAGAGGCACCAGACCTGGCGGTGGAC
 TCGTAGAGGCACCAGGACTTCGTCTGTATGGTCTCTTTACACACCCGGACACACCCCTTAAACTAGGCATCCAGAACATGACCT
 CACCAGCAGCAACCTCCAAGTGGAGGAAGACCCCTGTGGACTTTGGGAACTCCTGGAAGTGAAGTCCGAGTGTCTGACACCGTAG
 GTCTTGTACTGGAGTGGTCTGCTGGTGGAGGTTCACTCCCTTCTGGGACACCTGAAACCCCTTGAGGACCTTCTACTCGAGCGTAC
 ACGACTGTCCAGAAAAGTGCCTCTGGACTCATCCCTGCCACCTGCCATAACAACATCATGAAGCAGACGATGGTGGATTCTCTCT
 GTAGAACTCCTTACCAGTACGTTGGTCTTTTACCGAGACCTGAGTASGGGACGGTGGACGGTATTGTTGTAGTACTTCTGCTGTGTA
 CCACCTAAGGAGACATCTTAGGAATGGTCACTGCACTTCCAGGACTGCAACAAGCTGGTGGACCCCGAGCCATATCTGGATGTCT
 GCATTTACGACACCTGCTCCTGTGAGTCCATTGGGGACTGCGCCTGCTTCCGAAGTCCCTGACGTTGTTCCGACACCTGGGGCTCGG
 TATAGACCTACAGACGTAATGCTGTGGACGAGGACACTCAGETAACCCCTGACCGGAGCAAGTCCGACACCAATTGCTGCCTATG
 CCCAGTGTGTCGCCAGCATGGCAAGGTGGTGCATGGAGGACGGCCACATTTGTCGCCCCAGAGCTGGGAGGAGGAAACGCTGTG
 GTAACGACGGATACGGGTGCACACACGGGCTGTAACCGTCCACCACCTGGACCTCCTGCCGGTGAACACGGGGCTCTCGACGCTCC
 TCTCCTATCTCCGGGAGAACGGGTAAGAGTGTGAGTGGCGCTATAACAGCTGTGCACCTGCCTGTCAAGTCACTGTCTCAGACCCCT
 GAGCCACTGGCCTGCCTGTAGAGGCCCTTGTCCATACTCACACTCACCGGATATTTGTGACACGTTGGACGGACAGTTTCACT
 GCACAGTCTGGGACTCCCTGACCGGACCCACAGCACTGTGTGGAGGCTGCCATGCCACTGCCCTCCAGGAAAATCCTGGAT
 GAGCTTTTGCAGACCTGCGTTGACCCCTGAAGACTGTCCAGTGTGTGAGCGTACACACCTCCCGACGGTACGGGTGACGGGAGGTC
 CCTTTTAGACCTACTCGAAAACGCTCTGGACGCAACTGGGACTTCTGACAGGTACACACTCGTGGCTGGCCGGCTTTTGCCICA
 GGAAAGAAAGTACCTTGAATCCAGTCAACCTGAGCACGTCAGATTGCCACTGTGATGTTGCAACCTCACTCACCGACCGG
 CCGCAAAACGGAGTCTTTCTTTTCTAGTGGAACTTAGGGTCACTGGGACTCGTGACGGTCTAAACGGTGACACTACAACAGTTGGAG
 TGGAGTGAAGCCTGCCAGGAGCCGGGAGGCCCTGGTGGTGCCTCCACAGATGCCCGGTGAGCCCCACCCTCTGTATGTGGAGGA
 CATCTCGSAACCGCTTCACTTGGGACGGTCCCTCGGCCCTCCGGACACCACGGGAGGGTGTCTACGGGGCCACTCGGGGTGTTGA
 GACATACACCTCCTGTAGAGCCTTGGGGCAAGCAGATTTCTACTGCAGCAGGCTACTGGACCTGGTCTTCTGCTGGATGCTC
 CTCAGGCTGTCCGAGGCTGAGTTTGAAGTGTGAAG3CCTTTGTGGCTGCTAAAGATGACGTCGTCGGATGACCTGGACAGAAAG
 GACGACCTACCGAGGAGGTCGACAGGCTCCGACTCAAACCTCAGGACTTCCGSAACAGCTGGACATGATGGAGCGGCTCGCAT
 CTCCAGAAAGTGGGTCCGGTGGCGGTGGAGTACCACGACGGCTCCACGCTACATCGGGCTCAAGGACCCACCTGTACTAC
 CTCGCGACCGGTAGAGGGTCTTACCCAGGCGCACCCGACACCTCATGGTGTGTCGGAGGGTGCSEATGTAGCCGAGTTCCT
 GGGGAGCGACCGTCAAGCTCGCGGCATTTGCCAGCCAGGTGAAGTATGCG3GCAGCCAGGTGECCTCCACCAGCGAGTCTTGA
 AATACACACTGTTCCACCTTCCGCTGGCAGTCTCGACGCGCGTAACGGTCCGTTCACTTCATACGCCCTCGGGTCCACCGGAGGTG
 GTCGCTCCAGAACTTTATGTGTGACAAAGTAACTCTCAGCAAGATCGACCGCCCTGAAGCCTCCCGCATCGCCCTGCTCTGATGG
 CCAGCCAGGAGCCCCAACGGATSTCCGGAACTTTGTCCGCTACTTAGAGTCTTCTAGCTGGCGGACTTCCGAGG3CCTAGCG
 GGACACGACTACCGCTCGGCTCTCGGGTGGCTACAGGGCCTTGAACAG3CGATGCTCCAGGCTGAAGAGAA3AAGCTCA
 TTGTGATCCCGTGGGCACTGGGCCCCATGCCAACCTCAAGCAGATCCGCTCATCGAGAAGCAGGCCCTGCAGGTCCCGGACTT
 CTCTTCTTCCAGTAACTAGGGCCACCCGTAACCCGGGTACGGTTGAGTTCGTCTAGGCGGAGTAGCTCTTCTGTCGGGGAC
 AGAACAGGCCCTCTGCTGAGCAGTGTGGATGAGCTGGAGCAGCAAGGCAAGATCGTTA3CTACCTCTGTGACCTTGCCCT
 GAAGCCCTCCTCCTTGTTCGGAAAGCAGACTCGTCACACCTACTCGACCTCGTCTGTTCCCTGCTCTAGCAATCGATGGAGA
 CACTGGAACGGGGACTTCCGGGAGGAGGTACTCTGCCCCCGACATGGCAAACTCACTGTGG3CCCGGGCTCTTGGGGTT*CG
 ACCCTGGGGCCCAAGAGGAACTCCATGGTTCG3ATGTGGCGATGAGACGGGGGCTGTACCGTGTTCAGTACACCCGGGGCCCG
 AGAACCCCAAAGCTGGGACCCCGGGTCTCTCTTGAAGTACCAAGACTACACCGCTTCTGCTGGAAGGATCGGACAAAATGGT
 GAAGCCGACTTCAACAGGAGCAAGGAGTTCATGGAGGAGTATTACGCGGATGGATGTGGCCAGGACAAAGCAGGACCTTCTCA

GCCIGTTTTAACCACTTCGGCTGAAGTTGTCTCGTTCCTCAAGTACCTCCTCCACTAAGTCGCTACCTACACCCCGTCTGTGC
ATCCACGTCACGGTGTGCAGTACTCCTACATGGTGACCGTGGAGTACCCTTCAGCGAGGCACAGTCCAAAGGGGACATCCTGCA
GCGGTCGCGAGACGTAGGTGCAGTGCCACGACGTATGAGGATGTACCACTGGCACCTCATGGGGAAGTCGCTCCGTGTGAGGTTT
CCCCGTAGGACGTGCGCCACGCTCTGATCCGCTACCAGGGGCGCAACAGGACCAACACTGGGCTGGCCCTGCGGTACCTCTCTGA
CCACAGCTTCTGGTTCAGCCAGGCTGACCGGGAGCAGGCGCTAGGCGATGGTCCCGCGTGTCTGGTGTGACCCGACCGGGAC
GCCATGGAGAGACTGGTGTGGAAGAACCAGTCCGTCCTCCTGCGCTCGTCCGCCCCAACCTGGTCTACATGGTACCAGGAAATCC
TCCCTCTGATGAGATCAAGAGGCTGCCGAGAGACATCCAGGTGGTGCCTATTGCAGTGGGCCCTAATGGGGTTGGACAGATGTAC
CAGTGGCCTTTAGGACGGAGACTACTCTAGTTCTCCGACGGACCTCTGTAGTCCACCACGGGTAACCTCACCAGGGATTACCCAA
UGTGCAGGAGCTGCAGAGGATGGCTGGCCCAATGCCCTATCCTCATCCAGGACTTTGAGACGCTCCCGGAGAGGGTCTCTGACC
TGGTGTGCAGGTTGCACGCTCCTCGACCTCTCCTAACCGACCGGGTTACGGGGATAGGAGTAGGTCCTGAAACTCTGCGAGGGGGC
TCTCCGAGGACTGGACACGACCTGAGGTGTGTCTCCGAGAGGGGCTGCAGATCCUACCCCTCTCCCTGCACCTGACTGCAGCC
AGCCCTGGACGTGATCCTTCTCCTGGATGGCTCCTCCTCCACGACGAGGCTCTCCCGACGCTCAGGGTGGGAGAGGGGACG
TGGACTGACGTCGGTCCGGGACCTGCCTAGGAAGAGGACCTACCGAGGAGGAGTTTCCAGCTTCCTATTTTGTGAAATGAAGA
GTTTCCCAAGGCTTCATTTCAAAGCCAAATATAGGGCCTCGTCTCACTCAGGTGTGAGTGTCTCAAAGGTCGAGAATAAA
ACTACTTTACTTCTCAAAGCGCTCCGAAAGTAAAGTTTTCSETTATATCCCGGAGCAGAGTGAAGTCCACAGTACGACGAGTATG
GAAGCATACCACCACTGACCTGCCATGGAACTGGTCCCGGAGAAAGCCATTTGCTGAGCCTTGTGGAAGTGCATGCAGCGGGAG
GGAGCCCTCATACCTTCGTAGTGGTGGTAACTGCACGCTACCTTCCACCGAGGCTCTTTCCGGTAAACGACTCGGAACAGCTCC
AGTACGTCGCCCCCTCCCTCCGGGACGCCAAATCCGGGATGCCCTGGGCTTTGCTGTGCGATACTTGACTTCAGAAATGCATGGTCC
AGGCCGGGAGCCTCAAAGGCGGTGGTGCATCCTGGTGTGGTTTAGCCCTACGGAAACCGGAGACACGCTATGAACTGAAGTC
TTTACGTACCACGGTCCGGCCCTCGGAGTTTCCGCCACCAGTAGGACACGAGCAGTCTCTGTGGATTGAGTGGATGCAGGAGCT
GATGCCGCCAGTCCAACAGAGTGACAGTGTCCCTATTGGAATGGAGATCGCTACGATGCAGTGCCTGCAGAGACACCTAAGTC
ACCTACGTCGTCGACTACGGCGTCCAGGTTGTCTCACTGTCACAAAGGATAACCTTAACCTTAGCGGATGCTACGTCCTCCAGCTA
CGGATCTTGGCAGGCCACGAGGCGACTCCAACGTGGTGAAGCTCCAGCGAATCGAAGACCTCCCTACCATGGTCACTTGGGCAA
TTCTTGGGTCGATGCTTAGAACCCTCCGGGCTCGTCCGCTGAGGTTGCACCACTTCGAGGTCCGCTTAGCTTTCGGAGGGATGGTAC
CACTGGAACCCGTTAAGGAACTCCACAACTGTCTCTGCATTTGTTAGGATTTGCATGGATGAGGATGGGAATGACAAGAGGCC
CGGGGACGTCGTGACCTTGCAGACCACTGCCACGGAGGTTTGGACACGAGACCTAAACAATCCTAAACGTACCTACTCTACCC
TTACTCTTCTCCCGCCCCCTGCAGACCTGGAACCGTCTGCTCACGGTCAACCTGACTTCCAGCCAGATGCCACACCTTCTGAA
GAGTCATCGGGTCAACTGTGACCGGGGCTGAGGCTTCTGTGCCCTAACAGCCAGTCCCTGTGGCACTGAACGGTCCGGTCTACCG
GTCTGGAACGACTCTCAGTAGCCAGTTGACACTGGCCCCGACTCCGGAAGCACGGGATTGTCCGTGAGGGGACTTAAAGTGGAA
AGAGACCTGTGGCTGCCGCTGGACCTGCCCTGCTGTGTCACAGGCGACTCCACTCGGCACATCGTACCTTTGATGGGAGAAAT
TCAAATTTACCTTCTCTGGACACCGACGGGACCTGGACGGGGACRCACACGCTCCGTCGAGGTGAGCCGTGTAGCACTGGAA
ACTACCCGCTCTTAAAGTTGCTGACTGGCAGCTGTTCTTATGTCCFATTTCAAACAAGGAGCAGGACCTGGAGGTGATCTCCATA
ATGGTGCCTGCAGCCCTGGAGCAAGGCGAGGCGACTGACCGTGCACAAGAATAAGGATAAAGTTTGTTCCTCGTCTGGACCT
CCACTAAGAGCTATTACCACGGACGTCGGGACCTCCTTCCGTCCCGTCCGATGAAATCCATCGAGGTGAAGCACAGTCCCTCTCCG
TCGAGTGCACAGTACATGGAGGTGACGGTGAATGGGAGACTGGTCTCTGTTCCTTACGACGTACTTTAGGTAGCTCCACTTCGT
GTCACGGGAGAGGCACTCSACGTGTCACGTACCTCCACTGCCACTTACCCTCTGACCAGAGACAAGGAATGCTGGTGGGAACA
TGSAAGTCAACGTTTATGGTGCCATCATGCATGAGETCAGATTCAATCACCTTGGTCAACATCTTACATTCACTCCACAAAACAAT
GAACCCACCCCTTGTACCTTCAGTTGCAAAATACCACGGTAGTACGTACTCCAGTCTAAGTTAGTGGAAACAGTGTAGAAAGTGAAGT
GAGGTGTTTTGTTACTGTCCAACTGCAGCTCAGCCCCAAGACTTTTGTCTCAAAGACGATAGTCTGTGTGGGATCTGTGATGAG
AACGGAGCCAATGACTTCATGCTGAGGGAACAAGGTTGACGTCGAGTCCGGGTTCTGAAAACGAAGTTTCTGCATACCAGACACAC
CCTAGACACTACTCTGCTCGGTTACTGAACTACGACTCCCTAGGCACAGTCAACACAGACTGGAAAACACTTGTTCAGGAATGG
ACTGIGCAGCGGCCAGGGCAGACGTGCCAGCCATCCTGGAGGAGCAGTGTCTGTCCCGTGTGAGTGGTGTCTGACCTTTTGTG
AACAACTCCTTACCTGACACGTCGCCGCTCCGTCCTGCACCGTCCGGTAGGACCTCCTGTCACAGAACAGSCCGACAGCTCCAC
TGCCAGGTCTCTCTTACCACGTGTTGCTGAATGCCACAAGGTCCTGGCTCCAGCCACATTTCTATGCCATCTGCCAGCAGACAG

ES 2 848 703 T5

GSCTGTCGAGGGTGACGGTCCAGGAGGAGAAATGGTGACAAACGACTTACGGTGTTCAGGACCGAGGTGCGTGTAAAGATACGGTAG
ACGGTCGTCTGTCTTGCCACCAGGAGCAAGTGTGTGAGGTGATCGCCTCTTATGCCACCTCTGTGCGACCAACGGGGTCTGCGT
TGACTGGAGGACACCTCATTTCTGTGCTAACGGTGGTCTCGTTCCACACTCCACTAGCGGAGAATACGGGTGGAGACAGCCTGG
TTGCCCCAGACGCAACTGACCTCCTGTGGACTAAAGACACGAATGTCATGCCACCATCTCTGGTCTACAACCACTGTGAGCATGG
CTGTCCCCGGCCTGTGATGGCAACGTGAGCTCCTGTGGGGACCATCCCTCCGAAGTACAGTACGGGTGGTAGAGACCAGATGTTG
GTGACACTCGTACCAGACAGGGGCGGTGACACTACCGTTGCACFCGAGAGACACCCCTGGTAGGGAGGCTTCGCTGTTTCTGCCCTCC
AGATAAAGTCAATGTTGCAAGGCAGCTGTGTCCCTGAAGAGCCCTGCCTCAGTGCATTTGGTGAGGATGAGTCCAGCACCAGTTCCG
ACAAAGACGGGAGGTCTATTTCACTACAACCTTCCGTGACACAGGGACTTCTCCGGACGTGAGTCACTAACCCTCCTACCTCA
GGTCTGGTCAACCTGGAAGCCTGGGTCCCGGACCACCAGCCCTGTGAGATCTGCACATGCCTCAGCGGGCGGAAGGTCAACTGCA
CAACGCAGCCCTGCCCCACGGCCAAAGGACCTTCGGACCCAGGGCCTGGTGGTGGGACAGTCTAGACGTGTACGGAGTCCCGCCG
CTTCCAGTTGACGTGTTGCGTGGGACGGGGTGCCTGTTGCTCCACCGTGTGGCCTGTGTGAAGTAGCCCGCCTCCGCCAGAATG
CAGACCAGTGTGCCCGAGTATGAGTGTGTGTGACCCAGTGAAGTGTGACCCGAGGGTGCACACCCGACACACTTCATCGGGC
GGAGGCGGTCTTACGTCTGCTCAGCAGGGGCTCATACTCACACACACACTGGGTCACTCGACACTGTGTGCCCCAGTGCCTCACT
GTGAACGTGGCCTCCAGCCACACTGACCAACCTGGCGAGTGCAGACCACAACTTCACCTGCGCCTGCAGGAGGAGGAGTACGG
GGGTCAACGGAGTGCACACTGCACCGGAGGTGGGGTGTGACTGGTGGGACCGCTCACGTCTGGGTGAAGTGGACGGCGACGCTCT
TCCTCCTCACCAAAAGAGTGTCCCAACCTCCTGCCCGCCGACCGTTTGGCCACCCCTCGGAAGACCCAGTGTGTGATGAGTAT
GAGTGTCCCTGCAACTGTGTCAACGTTTTCTCACAGGGGTGGGAGGACGGGGGCGTGGCAAACGGGTGGGAAGCCTCTGGGTCA
CGACACTACTCATACTCACACGGACGTTGACACAGTTGTCCACAGTGAAGTGTCCCTTGGGTACTTGGCTCAACCGCCACCAAT
GACTGTGGCTGTACACACACCACTGCTTCCGACAAAGGTGTGTGTCCACCAGTGTCACTCGACAGGGGAACCCATGAACCGGA
GTTGGCGTGGTTACTGACACCGACATGGTGTGCTGACGGAAGGGCTGTTCCACACACAGGTGGGAAGCACCATCTACCTGTG
GGCCAGTTCGGGAGGAGGGTGGGATGTGTGCACCTGCACCGACATGGAGGATGCCGTGATGGGCTCCCGTGGCCACTTCGT
GGTAGATGGGACACCCGTCAGACCCCTCCTCCCGACGCTACACACGTGGACGTGGCTGTACTCCTACGGCACTACCCGGAGGGC
CACCGGGTGTCTCCAGAAGCCCTGTGAGGACAGCTGTGGTGGGCTTCACTTACGTTCTGCATGAAGGCGAGTGTGTGGAAG
GTGCTGCCATCTGCTGTGAGCAGGAGGTCTTCCGGACACTCCTGTGACAGCCAGCCCGAAGTGAATGCAAGACGTAATCTCCG
CTCACGACACTTCCACGGACGGTAGACGGACACTCGTGGTACTGGCTCACCGCGGGGGACTCCAGTCTTCTGGAAGAGTGT
CGGCTCCAGTGGGCTCCCGGAGAACCCTGCTCATCAATGAGTGTGACCACTGACCGAGTGGGCGCCCGCTGAGGGTFCAGA
AGGACCTTCTCACAGCCGAGGTCACCCGGAGGGCTCTTGGGACGGAGTAGTACTCACACTCCGAGTGAAGGAGGAGTCTT
TATACAAACAAAGAACGTCCTCCTGCCCCAGCTGGAGTCCCTGTCTGCCCTCGGCTTTGAGTGTGAGTGTAAAGACAGGCTCAC
TTCCTCCTCCAGAAATATGTTGTTTCTTGCAGAGGACGGGGTGCACCTCCAGGGACAGACGGGAGCCCGAAAGTCGACTCGAC
ATTCCTGCTCAGCGTGTCCCAAGCTGTGCTGTGAGCGCATGGAGCCCTGCATGCTCAATGGCACTGTCTATGGGCCCGGAAGA
CTGTGATGATCGATGTGTGGAGTGCACGACGGTTCCAGACGGACACTCGCGTACCTCCGGACGTACGAGTTACCGTGACAGTA
ACCCGGGCTTCTGACACTACTAGTACACACGACGACCTGCCGCTGCATGGTGCAGGTGGGGTCACTCTCTGATTCAGGCTGG
AGTGCAGGAAGACCACCTGCAACCCCTGCCCTGGGTTACAAGGAAGTGTGGACGGCGACGTACCACGTCCACCCCCAGTAGAG
ACCTAAGTTCGACCTCAGTCTTCTTGGTGGAGCTTGGGACGGGGACCCAATGTTCTTCAAATAACACAGGTGAATGTTGTG
GGAGATGTTTGCTACGGCTTGACCACTTCCAGCTAAGAGGAGGACAGATCATGACACTGAAGCGTGTGAGACGCTTTTTATTTGTG
TCCACTTACAACACCCCTACAACGGATGCCGAACCTGGTAAGTCGATTCCTCCTCCTGTCTAGTACTGTGACTTCGCACACTCT
CGCAGCAGGATGGCTGTGATACTCACTTCTGCAAGGTCAATGAGAGAGGAGTACTTCTGGGAGAAGAGGGTCCAGGCTGCCCA
CCCTTGTATGAACACAAGGGTCTTACCGACACTATGAGTGAAGACGTTCCAGTTACTCTCTCTCTCATGAAGACCCCTCTTCTCCC
AGTGTCCGACGGGTGGGAACTACTTGTGTTCTGTCTTCTGAGGAGGTTAAATATGAAAATCCAGGCACCTGCTGTGACACA
TGTGAGGAGCCTGAGTGAACGACATCACTGCCAGGCTGCAGTATGACAGAACGACTCCCTCCATTTTAATACTTTAAGGTCCGT
GGACGACACTGTGTACTCTCGACTCAGCTTGTGTAGTGCAGGTCGACGTCATACTCAAGTGGGAAGCTGTAAAGTGTGAA
GTAGAGGTGGATATCCACTACTGCCAGGGCAAATGTGCCAGAAAGCCATGTACTCCATTGACATCAACGATGATGTTCCACCCCT
CGACATTCAGACTTACTTCCACCTATAGGTGATGACGGTCCCGTTTACACGGTCTTTCGGTACATGAGGTAAGTGTAGTTGCTA
CAGCAGGACCACTGCTCCTGTGCTCTCCGACACGGACGGAGCCCATGCAGGTGGCCTGCACCTGCACCAATGGCTCTGTTGTGTA
CCATGAGGTTCTCAATCGTCTGCTGCTCAGGAGGACGACAGGCTGTGCTGCTCCCTCGGGTACGTCACCCGGACGTTGAGTGGTTA
CCGAGACAAACATGTTACTCCAAGAGTTAGCCATGGAGTGGAAATGCTCCCGCAGGAGTGCAGCAAGTGA

SEQ ID NO: 28 - Longitud completa de la secuencia de péptidos del factor de von Willebrand (X es cualquier aminoácido natural)

ES 2 848 703 T5

MIPARFAGVLLALALILPGT LCAEGTRGRSSTARCSLFGS DFNNTFDGSMYSFAGYCSYL
 LAGGCQKRFSSTIIGDFQNGK RVSLSVYLGEFFDIHLFVNG TVTQGDQRVSMFYASKGLYL
 ETEAGYYKLSGEAYGFVARI DGSGNFQVLLSDRYFNKTCG LCGNFNIFAEDDFMTQEGTL
 TSDPYDFANSWALSSGEQWC ERASPPSSSSCNISSGEMQKG LWEEQCQLLKSTSVFARCHPL
 VDPEPFVALCEKTLCECAGG LECACPALLEYARTCAQEGM VLYGWTDHSACSPVCPAGME
 YRQCVSPCARTCQSLHINEM CQERCVDGCSCPEGQLLEDEG LCVESTTECPVHSGKRYPPG
 TSLSRDCNTCICRNSQWICS NEECPGECCLVTGQSHFKSFD NRYFTFSGICQYLLARDCCD
 HSFSIVTETVQCADDRDAVC TRSVTVRLPGLHNSLVKLBH GAGVAMDGQDIQLPLLKGD
 RIQHTVTASVRLSYGEDLQM DWDGRGRLLVVKLSPVYAGKT CGLCGNYNGNQGDDEFLTSPG
 LAEPRVEDFGNAWKLHGDCQ DLQKQHSDEPCALNPRMTRFS EEACAVLTSPTFEACHRAVS
 PLFYLRNCRYDVCSGSDGRE CLCGALASYAAACAGRGVRV AWREPGRCELNCPKGQVYLQ
 CGTPCNLTCSRSLSPDEECN EACLEGCFCPPGLYMDERGD CVPKAQCPCYYDGEIFQPED
 IFSDHHTMCYCEDGFMHCTM SGVPGSLLPDAVLSSPLSHR SKRSLSCRPPMVKLVCPADN
 LRAEGLECTKTCQNYDLECM SMGCVSGCLCPPGMVRHENR CVALERCPCFHQKEYAPGE
 TVKIGCNTCVCRDRKWNCTD HVCDATCSTIGMAHYLTFDG LKYLFPGECQYVVLVQDYCGS
 NPGTFRLLVGNKGCSPSVK CKKRVTILVEGGEIELEFDEG VNVKRPMDETHFEVVESEGR
 YIILLGKALSVVWDRHLSI SVVLKQTYQEKVCGLCGNFD GIQNNDLTSSNLQVEEDPVD
 FGNSWKVSSQCADTRKVELD SSPATCHNNIMKQTMVDSSC RILTSDFVQDCNKLVDPEPY
 LDVCIYDTCSCESIGDCACF CDTIAAYAHVCAQHKGKVTW RTATLCPQSCEERNLRENGY
 ECEWRYNSCAPACQVTCQHP EPLACPVQCVEGCHAHCPPG KILDELLQTCVDPEDCPVCE
 VAGRRFASGKKVTLNPSDPE HCQICHCDVNLTCACQEP GGLVVPPTDAPVSPTTLYVE
 DTSEPLHDFYCSRLI,DLVF LLDGSSRLSEAEFEVLKAFV VDMMERLRISQKWVRVAVVE
 YHDGSHAYIGLKDRKRPESEL RRIASQVKYAGSQVASTSEV LKYTFEQIFSKIDRPEASRI
 ALLLMSQEPQRMSRNFVRY VOGLKKKKVIVIPVGIGPHA NLKQIRLIEKQAPENKAFVL
 SSVDELEQQRDEIVSYLCDL APEAPPPTLPPDMAQVTVGP GLLGVSTLGPKRNSMVLDA
 FVLEGSDBKIGEADFNRSKEF MEEVIQRMDVQDSIHVTVL QYSYMTVEYPFSEAQSKGD
 ILQRVREIRYQCGNRTNTGL ALRYLSDHSFLVSQGDREQA PNLVYMTGNPASDEIKRLP
 GDIQVVPIGVGPNANVQELE RIGWPNAPILIQDFETLPRE APDLVLQRCCSGEGLQIPTL

ES 2 848 703 T5

SPAPDCSQPLDVLILLLDGSS SFPASYFDEMKSFAKAFISK ANIGPRLTQVSVLQYGSITT
IDVPWNVVPKHAHLLSLVDV MQREGGFSQIGDALGFAVRY LTSEMHGARPGASKAVVILV
TDVSVDSVDAADAARSNRV TVFPIGIGDRYDAAQLRILA GPAGDSNVVKLQRIEDLPTM
VTLGNSELHKLCSGFVRIEM DEDGNEKRPGDVWTLDPQCH TVTCQPDGQTLKSHRVNCD
RGLRSPSPNSQSEVQVEETC GCRWTCPCVCTGSSTRHIVT FDGQNFKLTGSCSYVLFQNK
EQDLEVILHNGACSPGARQG CMKSIEVKHSALSVEKXSDM EVTVNGRLVSVPPYVGGNMEV
NVYGAIMHEVRFNHLGHIFFTPQNEFQLQLSPKT FASKTYGLCGICDENGANDE
MLRDGTVTTDWKTILVQEWTV QRFQGTCPPILEEQCLVPDS SHCQVLLLPLFAECHKVLAP
ATFYAICQOBSCHQEQVCEV IASYAHLCRTNGVCVDWRTP DFCAMSCPPSLVYNHCEHGC
PRHCDGNVSSCGDHPSEGCF CPPDKVMLEGSCVPEEACTQ CIGEDGVQHQFLEAWVPDHQ
PCQICTCLSGRKVNCTTQPC PTAKAPTCLCEVARLRQNA DQCCPEYECVCDPVSODLPP
VPHCERGLQPTLTNPGECP NFTCACRKEECKRVSPPSCP PHRLPTLRKTQCCDEYECAC
NCVNSTVSCPLGYLASTATN DCGCTTTTCLPDKVCVHRST IYPVGQFWEEGCDVCTCTDM
EDAVMGLRVAQCSQKPCEDS CRSGFTYVLHEGECGRCLP SACEVVTGSPRGDSQSSWKS
VGSQWASPENPCLINECVRV KEEVFIQQRNVSCPQLEVPV CPSGFQLSCKTSACCPSCRC
ERMEACMLNGTVIGPGKTM IDVCTTCRCMVQVGVISGFK LECRKTTCNCPCLGYKEENN
TGECCGRCLPTACTIQLRGG QIMTLKRDETLQDGGCDTHFC KVNERGEYFWEKRVITGCPPE
DEHKCLAEGGKIMKIPGTCC DTCEEPECNDITARLQYVKV GSCKSEVEVDIHYCQGGKAS
KAMYSIDINDVQDQCCSCSP TRTEPMQVALHCTNGSVVYH EVLNAMECKCSPRKCCK

SEQ ID NO: 29 - Aminoácidos 233-236 de IgG1 humana

ELLG

SEQ ID NO: 30 - XTEN AE42-4, secuencia de proteínas

5 GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS

SEQ ID NO: 31 - XTEN AE42-4, secuencia de ADN

GGCGCGCCAGGTTCTCCTGCTGGCTCCCCACCTCAACAGAAGAGGGGACAAGCCAAAGC
GCTACGCCTGAGAGTGGCCCTGGCTCTGAGCCAGCCACCTCCGGCTCTGAAAACCCCTGCC
TCGAGC

SEQ ID NO: 32 - XTEN AE144-2A, secuencia de proteínas

TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG
TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPG
TSESATPESGPGTSESATPESGPG

10 SEQ ID NO: 33 - XTEN AE144-2A, secuencia de ADN

GGCGCGCCAACCAGTACGGAGCCGTCGGAGGGGAGCGCACCAGGAAGCCGGGCTGGGAGC
CCGACTTCTACCGAAGAGGGTACATCTACCGAACCAAGTGAAGGTTACAGCACCAGGCACC
TCAACAGAACCCCTCTGAGGGCTCGGCGCCTGGTIACAAGTGAGTCCGCCACCCAGAAATCC
GGGCCTGGGACAAGCACAGAACCTTCGGAAGGGAGTGCCCCTGGAACATCCGAATCGGCA
ACCCAGAAATCAGGGCCAGGATCTGAGCCCGGACTTCGGGCTCCGAGACGCTGGGACA
TCCACCCAGCCCTCCGAAGGATCAGCCCCAGGCACCAGCACGGAGCCCTCTGAGGGAAGC
GCACCTGGTACCAGCGAAAGCGCAACTCCCGAATCAGGTCCCGGIACGAGCGAGTCGGCG
ACCCCGGAGAGCGGGCCAGGTGCCTCGAGC

SEQ ID NO: 34 - XTEN AE144-3B, secuencia de proteínas

ES 2 848 703 T5

SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG
SPAGSPTSTEEGTSTEPSEGSAPG

SEQ ID NO: 35 - XTEN AE144-3B, secuencia de ADN

GGCGCGCCAAGTCCCGCTGGAAGCCCAACTAGCACCCGAAGAGGGGACCTCAGAGTCCGCC
ACCCCCGAGTCCGGCCCTGGCTCTGAGCCTGCCACTAGCGGCTCCGAGACTCCTGGCACA
TCCGAAAGCGCTACACCCGAGAGTGGACCCGGCACCTCTACCGAGCCCAGTGAGGGCTCC
GCCCCTGGAACAAGCACCCGAGCCAGCGAAGGCAGCGCCCCAGGGACCTCCACAGAGCCC
AGTGAAGGCAGTGCTCCTGGCACCAGCACCCGAACCAAGCGAGGGCTCTGCACCCGGGACC
TCCACCCGAGCCAAGCGAAGGCTCTGCCCTGGCACTTCCACCGAGCCCAGCGAAGGCAGC
GCCCCTGGGAGCCCCGCTGGCTCTCCCACCAAGCACTGAGGAGGGGCACATCTACCGAACCA
AGTGAAGGCTCTGCACCAGGTGCCTCGAGC

SEQ ID NO: 36 - XTEN AE144-4A, secuencia de proteínas

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEG
TSESATPESGPGTSTEPSEGSAPG

5

SEQ ID NO: 37 - XTEN AE144-4A, secuencia de ADN

GGCGCGCCAACGTCCGAAAGTGCTACCCCTGAGTCAGGCCCTGGTAGTGAGCCTGCCACA
AGCGGAAGCGAAACTCCGGGGACCTCAGAGTCTGCCACTCCCGAATCGGGGCCAGGCTCT
GAACCGGCCACTTCAGGGAGCGAAACACCAGGAACATCGGAGAGCGCTACCCCGGAGAGC
GGGCCAGGAAGTAGTACTGAGCCTAGCGAGGGAAGTGCACCTGGTACAAGCGAGTCCGCC
ACACCCGAGTCTGGCCCTGGCTCTCCAGCGGGCTCACCCACGAGCACTGAAGAGGGCTCT
CCCGCTGGCAGCCCAACGTGACAGAAGAAGGATCACCCAGCAGGCTCCCCACATCAACA
GAGGAGGGTACATCAGAATCTGCTACTCCCGAGAGTGGACCCGGTACCTCCACTGAGCCC
AGCGAGGGGAGTGCACCAGGTGCCTCGAGC

SEQ ID NO: 38 - XTEN AE144-5A, secuencia de proteínas

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
TSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
SPAGSPTSTEEGSPAGSPTSTEEG

10

SEQ ID NO: 39 - XTEN AE144-5A, secuencia de ADN

GGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACA
TCTGGGTCCGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGGTCC
GAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAGAGCGCTACACCAGAAAGC
GGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCT
CCTACATCCACGGAGGAGGGCACTTCCGAATCCCGCCACCCCGGAGTCAGGGCCAGGATCT
GAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGT
GGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTACCAGCGGGCTCG
CCCACCAGCACTGAAGAAGGTGCCTCGAGC

SEQ ID NO: 40 - XTEN AE144-6B, secuencia de proteínas

TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG
SEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG
TSESATPESGPGTSTEPSEGSAPG

15

SEQ ID NO: 41 - XTEN AE144-6B, secuencia de ADN

GGCGCGCCAAACATCTACCGAGCCTTCCGAAGGCTCTGCCCTGGGACCTCAGAATCTGCA
ACCCCTGAAAAGCGGCCCTGGAACCTCCGAAAGTGCCACTCCCGAGAGCGGCCAGGGACA
AGCGAGTCAGCAACCCCTGAGTCTGGACCCGGCAGCGAGCCTGCAACCTCTGGCTCAGAG
ACTCCCGGCTCAGAACCCGCTACCTCAGGCTCCGAGACACCCGGCTCTCCTGCTGGGAGT
CCCCTTCCACCGAGGAAGGAACATCCACTGAGCCTAGTGAGGGCTCTGCCCTGGAACC
AGCACAGAGCCAAGTGAGGGCAGTGCAACCAGGATCCGAGCCAGCAACCAGCGGTCCGAG
ACTCCCGGGACCTCTGAGTCTGCCACCCAGAGAGCGGACCCGGCACTTCAACCGAGCCC
TCCGAAGGATCAGCACCAGGTGCCTCGAGC

SEQ ID NO: 42 - XTEN AG144-1, secuencia de proteínas

PGSSPSASTGTGPGSSPSASTGTGPGTSGSTASSSPGSSTPSGATGSPGSSPSASTGTG
PGASPGTSSSTGSPGTSGSTASSSPGSSTPSGATGSPGTSGSTASSSPGASPGTSSSTG
PGASPGTSSSTGSPGTSGSTASS

SEQ ID NO: 43 - XTEN AG144-1, secuencia de ADN

GGCGCGCCACCCGGGTCGTCCCCGTCCGGCTCCACCGGAACAGGGCCAGGGTCATCCCCG
TCAGCGTCGACTGGGACGGGACCCGGGACACCCGGTTCGGGACTGCATCCTCCTCGCT
GGTTCGTCCACCCCGTCAGGAGCCACGGGTTCCGCCGGAAGCAGCCCAAGCGCATCCACT
GGTACAGGGCCTGGGGCTTACCCGGTACTTCATCCACGGGGTACCCGGGAACGCCCGGA
TCGGGGACGGCTTCCICATCACCAGGATCGTCAACACCCTCGGGCGCAACGGGCAGCCCC
GGAACCCCTGGTTCCGGTACGGCGTTCGTCCAGCCCCGGTCCGAGCCCCGGGAACAAGCTCG
ACAGGATCGCCTGGGGCGTCCACCGGCACGTCCAGCACAGGCAGCCCCGGAAACCCCTGGA
TCGGGAACCCGCTCGTCAAGCGCCTCGAGC

5

SEQ ID NO: 44 - XTEN AG144-A, secuencia de proteínas

GASPGTSSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTSGSTASSSPGSSTPSGATGSP
GSSPSASTGTGPGASPGTSSSTGSPGTSGSTASSSPGSSTPSGATGSPGTSGSTASSSP
GASPGTSSSTGSPGASPGTSSSTGSP

SEQ ID NO: 45 - XTEN AG144-A, secuencia de ADN

GGCGCGCCAGGTGCCTCGCCGGAACATCATCAACTGGTTACCCGGGTCATCCCCCTCG
GCCTCAACCGGGACGGGTCCCGGCTCATCCCCAGCGCCAGCACTGGAACAGGTCTGGC
ACTCCTGGTTCGGTACGGCATCGTTCATCCCCGGAAGCTCAACACCCTCCGGAGCGACA
GGATCACCTGGTCTGTCACCTTCGGCGTCAACTGGAACGGGGCCAGGGCCCTCACCCGGA
ACGTCTCGACTGGGTGCCTGCTACGCCGGGATCAGGAACGGCCTCATCCTCGCCTGGG
TCCTCAACGCCCTCGGGTCCGACTGGTTCCGCCGGAACTCCTGGCTCGGGGACGGCCTCG
TCGTCCCTGGGGCATCACCGGGACGAGCTCCACGGGGTCCCCTGGAGCGTCCCGGGG
ACCTCCTCGACAGGTAGCCCGCCTCGAGC

10 SEQ ID NO: 46 - XTEN AG144-B, secuencia de proteínas

GTPGSGTASSSPGSSTPSGATGSPGASPGTSSSTGSPGTSGSTASSSPGSSTPSGATGSP
GSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSSTGSP
GASPGTSSSTGSPGASPGTSSSTGSP

SEQ ID NO: 47 - XTEN AG144-B, secuencia de ADN

GGCGCGCCAGGTACACCGGGCAGCGGCACGGCTTCGTGTCACCCGGCTCGTCCACACCG
TCGGGAGCTACGGGAAGCCCAGGAGCGTACCGGGAACGTCTGTC AACGGGGTACCCGGGT
ACGCCAGGTAGCGGCACGGCCAGCAGCTCGCCAGGTTTCATCGACCCCGTCGGGAGCGACT
GGGTCCGCCGGATCAAGCCCGTCAGCTTCCACTGGAACAGGACCCGGGTCTCGCCGTCA
GCCTCAACGGGGACAGGACCTGGTTTCATCGACCCGTACGGGGCGACAGGCTCGCCCGGA
TCGTCAACACCCTCGGGGGCAACGGGGAGCCCTGGTGCCTCGCCTGGAACCTCATCCACC
GGAAGCCCGGGGGCCTCGCCGGGTACGAGCTCCACGGGATCGCCCGGAGCGTCCCCCGGA
ACTTCAAGCACAGGGAGCCCTGCCTCGAGC

SEQ ID NO: 48 - XTEN AG144-C, secuencia de proteínas

ES 2 848 703 T5

GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP
GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP
GSSTPSGATGSPGASPGTSSTGSP

SEQ ID NO: 49 - XTEN AG144-C, secuencia de ADN

GGCGCGCCAGGTACACCCGGATCGGGTACAGCGTCATCGAGCCCCGGTGGCTCACCTGGT
ACGTCGAGCACCGGGTCGCCAGGGGGCGTCCCCTGGGACGTCCTCAACAGGCTCGCCCCGGT
GCGTCACCCGGCACGTCGTCCACGGGTTCACCTGGTAGCTCCCCTTCGGCGTCCACTGGC
ACCGGGCCCTGGAATCCGGGGAGCGGCACAGCGAGCTCGTCGCCGGGAGCATCGCCTGGG
ACATCGAGCACCGGGTCGCCAGGAGCATCGCCCGGAACATCCAGCACAGGAAGCCCCGGC
GCGTCGCCCCGGGACATCAAGCACAGGTTCCCCTGGGATCGAGCACGCCGTCGGGAGCCACT
GGATCACCCAGGGAGCTCGACACCTTCGGGGCAACGGGATCGCCCCGGAGCCAGCCCCGGT
ACGTCGAAGCACTGGCTCCCCTGCCTCGAGC

SEQ ID NO: 50 - XTEN AG144-F, secuencia de proteínas

GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP
GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTSGSGTASSSPGSSPSASTGTGP
GSSTPSGATGSPGASPGTSSTGSP

5

SEQ ID NO: 51 - XTEN AG144-F, secuencia de ADN

GGCGCGCCAGGCTCCAGCCCCCTCCGCGAGCACGGGAACCGGACCAGGTTTCGTCACCCCTCA
GCATCAACGGGGACGGGACCGGGGGCGTCACCAGGAACGTCCTCCACCGGCTCGCCGGGT
GCATCACCCGGAACGTCATCGACCGGATCGCCAGGGAGCTCGACGCCATCAGGCGCAACA
GGATCACCTGGCTCAAGCCCTAGCGCGTCAACCGGCACGGGTCCGGGTGCCTCCCCTGGC
ACGTCACAGCACCGGATCACCCGGATCGAGCCCATCCGCCTCAACCGGAACCGGACCCGGT
ACACCAGGGTCGGGAACAGCCTCCTCGTCACCAGGCTCCTCAACCCCCCTCGGGAGCCACG
GGTTCGCCCCGGTTCGTCAACGCCTTCGGGAGCAACTGGTAGCCCCGGGAGCATCGCCAGGA
ACTTCGAGCACGGGGTCGCCCCGCCTCGAGC

LISTADO DE SECUENCIAS

<110> Biogen Idec MA Inc.

10

<120> Gen de factor VIII optimizado

<130> 2159,378PC01

<140> Por asignar

15

<141> 14-02-2014

<150> 61/765.626

<151> 15-02-2013

20

<160> 51

<170> PatentIn versión 3.5

<210> 1

25

<211> 4404

ES 2 848 703 T5

<212> ADN

<213> Secuencia artificial

<220>

5

<223> FVIII BDD optimizado

<400> 1

```

cgtaaggccg ccaccatgca gattgagctg tctacttgct ttttcctgtg cctgctgagg      60
ttttgctttt ccgctacacg aaggtattat ctgggggctg tggaaactgtc ttgggattac      120
atgcagagtg acctgggaga gctgccagtg gacgcaaggt ttccccctag agtcccctaag      180
tcattcccct tcaacactag cgtggtctac aagaaaacac tgttcctgga gtttactgat      240
cacctgttca acatcgcaaa gcctaggcca cctgggatgg gactgctggg gccacaacac      300
caggccgagg tgtacgacac cgtggtcatt acacttaaga acatggcctc acaccccggt      360
agcctgcatg ctgtgggctg cagctactgg aaggcttccg aaggagcaga gtatgacgat      420
cagacttccc agagagaaaa agaggacgat aagggtgttc ctggcggatc tcatacctac      480
gtgtggcagg tcctgaaaga gaatggccct atggcctccg accctctgtg cctgacctac      540
tcttatctga gtcacgtgga cctggtcaag gatctgaaca gcggcctgat cggagccctg      600
ctggtgtgca gggaaggaa cctggctaag gagaaaaccc agacactgca taagtccatt      660
ctgctgttcg ccgtgtttga cgaagggaaa tcatggcaca gcgagacaaa gaatagtctg      720
atgcaggaca gggatgccgc ttcagccaga gcttggccca aatgacacac tgtgaacggc      780
tacgtcaatc gctcactgcc tgggctgacg ggctgccacc gaaagagcgt gtattggcat      840
gtcatcggga tgggcaccac acctgaagtg cactccattt tcctggaggg acataccttt      900
ctggtccgca accaccgaca ggcttccttg gagatctctc caattacctt cctgacagca      960
cagactctgc tgatggacct ggggcagttc ctgctgtttt gccacatcag ctcccaccag     1020
catgatggca tggaggctta cgtgaaagtg gactcttgtc ccgaggaacc tcagctgcgg     1080

```

ES 2 848 703 T5

atgaagaaca atgaggaagc agaagactat gacgatgacc tgaccgactc cgagatggat 1140
gtggtccgat tcgatgacga taacagcccc tcctttatcc agattagatc tgtggccaag 1200
aaacacccta agacatgggt ccattacatc gcagccgagg aagaggactg ggattatgca 1260
ccactggtgc tggcaccaga cgatcgctcc taaaaatctc agtatctgaa caatgggcca 1320
cagaggattg gcagaaagta caagaaagtg cggttcatgg catataccga tgagacctc 1380
aagactcgcg aagccatcca gcacgagagc ggcacccctg gaccactgct gtacggagaa 1440
gtgggagaca ccctgctgat cttttcaag aaccaggcca gccggcctta caatatctat 1500
ccacatggga ttacagatgt gcgccctctg tacagcagga gactgccaaa gggcgtcaaa 1560
cacctgaagc acttcccaat cctgcccgga gaaatcttca agtacaagtg gactgtcacc 1620
gtcagaggatg gcccactaa gagcgacct cggcgcctga cccgctacta ttctagtttc 1680
gtgaatatgg aaagagatct ggcaagcggc ctgatcggac cactgctgat ttgttacaaa 1740
gagagcgtgg atcagagagg caaccagatc atgtccgaca agcggaatgt gattctgttc 1800
agtgtctttg acgaaaacag gtcattgtac ctgaccgaga acatccagag attcctgcct 1860
aatccagctg gggcgcagct ggaagatcct gagtttcagg catctaacat catgcatagt 1920
attaatggct acgtgttoga cagtttgcag ctgagcgtgt gcctgcacga ggtcgttac 1980
tggtatatcc tgagcattgg ggcacagaca gatttctga gcgtgttctt ttccggctac 2040
acttttaagc ataaaatggc ctatgaggac aactgactc tgttcccctt cagcggcgaa 2100
accgtgttta tgagcatgga gaatcccgga ctgtggattc tggggtgcca caacagcgt 2160
ttcagaaatc gcggaatgac tgcctgctg aaagtgtcaa gctgtgacaa gaacaccggg 2220
gactactatg aagattcata cgaggacatc agcgcatac tgctgtccaa aaacaatgcc 2280
attgaacccc ggtcttttag tcagaatcct ccagtgtga agcggcacca gcgcgagatc 2340
accgcacta ccctgcagag tgatcaggaa gagatcgact acgacgatac aatttctgtg 2400
gaaatgaaga aagaggactt cgatatctat gacgaagatg agaaccagag tcctcgatca 2460
ttccagaaga aaaccaggca ttactttatt gccgcagtgg agcggctgtg ggattatggc 2520
atgtcctcta gtcctcagc gctgcgaaat agggcccagt caggaagcgt cccacagttc 2580
aagaaagtgg tcttccagga gtttacagac gggtccttta ctacagccact gtacaggggc 2640
gaactgaacg agcacctggg actgctgggg ccctatatca gagcagaagt ggaggataac 2700
attatggtca ccttcagaaa tcaggcctct cggccttaca gtttttatte aagcctgatc 2760
tcttacgaag aggaccagcg acagggagct gaaccacgaa aaaacttcgt gaagccta 2820
gagacaaaa catacttttg gaaggtgcag caccatattg cccaacaaa agacgagttc 2880
gattgcaagg catgggcta tttttctgac gtggatctgg agaaggacgt gcacagtggc 2940
ctgattggcc cactgctggt gtgccatac aacaccctga atccagccca cggccggcag 3000

ES 2 848 703 T5

gtcactgtcc aggagttcgc tctgttcttt accatctttg atgagacaaa gagctggtac 3060
 ttcaccgaaa acatggagcg aaattgcagg gctccatgta acattcagat ggaagacccc 3120
 acattcaagg agaactaccg ctttcatgct atcaatggat acatcatgga tactctgccc 3180
 gggctggtca tggcacagga ccagagaatc cggtggtatc tgctgagcat ggcagcaac 3240
 gagaatatcc actcaattca tttcagcggg cacgtgttta ctgtcaggaa gaaagaagag 3300
 tacaagatgg ccctgtacaa cctgtatccc ggcgtgttcg aaaccgtcga gatgctgcct 3360
 agcaaggccg gaatctggag agtggaatgc ctgattggag agcacctgca tgctgggatg 3420
 tctaccctgt ttctggtgta cagtaataag tgtcagacac ccctgggaat ggcatccggg 3480
 catatcaggg atttccagat taccgcatct ggacagtacg gacagtgggc acctaagctg 3540
 gctagactgc actattccgg atctatcaac gcttgggtcca caaaagagcc tttctcttgg 3600
 attaaggtgg acctgctggc cccaatgatc attcatggca tcaaaactca gggagctcgg 3660
 cagaagttct cctctctgta catctcacag tttatcatca tgtacagcct ggatgggaag 3720
 aaatggcaga cataccgcyg caatagcaca ggaactctga tgggtgttct tggcaacgtg 3780
 gacagcagcg gaatcaagca caacattttc aatcccceta tcattgctag atacatccgg 3840
 ctgcacccaa cccattatc tattcgaagt aactgagga tggaaactgat gggatgcgat 3900
 ctgaacagtt gttcaatgcc cctggggatg gagtccaag caatctctga cgccagatt 3960
 accgctagct cctacttcac taatatgttt gctacctgga gcccttcaa agcaagactg 4020
 cacctgcaag gccgcagcaa cgcattggcg ccacaggtga acaatcccaa ggagtgggtg 4080
 caggtcgatt ttcagaaaac tatgaaggtg accgggggtca caactcaggg cgtgaaaagt 4140
 ctgctgacct caatgtacgt caaggagttc ctgatctcta gttcacagga cggacatcag 4200
 tggacactgt tctttcagaa cgggaaggtg aaagtcttcc agggcaatca ggattccttt 4260
 acacctgtgg tcaacagtct agaccctcca ctgctgacca gatacctgag aatccaccct 4320
 cagtctggg tgcaccagat tgcctgaga atggaagtgc tgggatgca gcccaggat 4380
 ctgtactgat aactcgagtc gacc 4404

<210> 2

<211> 4419

5

<212> ADN

<213> Secuencia artificial

<220>

<223> FVIII BDD optimizado

10

<400> 2

ES 2 848 703 T5

ggcgcgcccg tacggccgcc accatgcaga tcgagctgtc tacctgcttc ttctgtgcc 60
 tgctgcggtt ctgcttcagc gccaccggc ggtactacct gggcgccgtg gaactgagct 120
 gggactacat gcagagcgac ctgggggagc tgcccgtgga cgccagattc cccccaagag 180
 tgccaagag cttccccttc aacacctccg tgggtgataa gaaaaccctg ttctgagagt 240
 tcaccgacca cctgttcaat atcgccaagc ccagaccccc ctggatgggc ctgctgggcc 300
 ctacaatcca ggccgaggtg tacgacaccg tggtcatac ccttaagaac atggccagcc 360
 accccgtgtc cctgcacgcc gtgggctgtt cctactggaa ggctctgag ggcgctgagt 420
 acgacgacca gaccagccag cgcgagaaa aggacgaaa agtctttcct ggcgagcc 480
 atacctactg gtggcaggtc ctgaaagaaa acggccctat ggctccgac ccctgtgcc 540
 tgacctacag ctacctgagc cacgtggacc tggtaagga cctgaacagc ggctgattg 600
 ggcacctgct cgtgtgtaga gagggcagcc tcgccaaaaga gaaaaccag accctgcaca 660
 agttcatcct gctgttcgcc gtgttcgacg agggcaagag ctggcacagc gagacaaaaga 720
 acagcctgat gcaggaccg gacgcccct ctgccagagc ctggcctaag atgcacaccg 780
 tgaacggcta cgtgaacaga agcctgccc gactgatcgg ctgccaccgg aagtccgtgt 840
 actggcacgt gatcgcatg ggcaccacc ccgaggtgca cagcatctt ctggaaggcc 900
 acaccttct cgtgcggaac cacagacagg ccagcctgga aatcagccct atcaccttc 960
 tgaccgcca gacactgctg atggacctg gccagttcct gctgttttgc cacatcagca 1020
 gccaccagca cgacggcatg gaagcctacg tgaagggtgga cagctgcccc gaggaacccc 1080
 agctgcggat gaagaacaac gaggaagccc aggactacga cgacgacctg accgacagcg 1140
 agatggacgt cgtgcgcttc gacgacgaca acagccccag ctcatccag atcagaagcg 1200
 tggccaagaa gcaacccaag acctgggtgc actatatcgc cgccgaggaa gaggactggg 1260
 actacgccc tctggtgctg gccccgacg acagaagcta caagagccag tacctgaaca 1320
 atggccccc ggggatcggc cggaagtaca agaaagtgcg gttcatggcc tacaccgacg 1380
 agacattcaa gaccagagag gccatccagc acgagagcgg catcctgggc ccctgctgt 1440
 atggcgaagt gggcgacacc ctgctgatca tcttaagaa ccaggccagc cggccctaca 1500
 acatctaccc ccacggcatc accgacgtgc ggcccctgta cagcagacgg ctgccaagg 1560
 gcgtgaagca cctgaaggac ttccccatc tgcccggcga gatcttcaag tacaagtgga 1620
 ccgtgaccgt ggaagatggc cccaccaaga gcgaccccag atgctgacc cgtactaca 1680
 gcagcttctg gaacatggaa cgggacctgg cctccgggct gatggccct ctgctgatct 1740
 gctacaaaga aagcgtggac cagcgggca accagatcat gagcgacaag cggaacgtga 1800
 tcctgttcag cgtgttcgat gagaatcggc cctggtacct gaccgagaat atccagcgt 1860
 tcctgccc aaactgccc gtgcagctg aagatcccga gttccaggcc agcaacatca 1920
 tgcactccat caatggctac gtgttcgaca gctccagct gagcgtgtgc ctgcacgag 1980
 tggcctactg gtacatcctg agcatcggc ccagaccga ctctctgagc gtgttcttca 2040

ES 2 848 703 T5

gcggtacac cttcaagcac aagatggtgt acgaggatac cctgaccctg ttccccttct 2100
ccggcgaaac cgtgttcattg agcatggaaa accccggcct gtggattctg ggctgccaca 2160
acagcgactt cagaaaccgg ggcatgaccg ccctgctgaa ggtgtccagc tgcgacaaga 2220
acaccggcga ctactacgag gacagctatg aggacatcag cgcctacctg ctgagcaaga 2280
acaacgccat cgagcccaga tccttcagcc agaaccccc cgtgctgaag cggcaccaga 2340
gagagatcac ccggaccacc ctgcagtccg accaggaaga gattgattac gacgacacca 2400
tcagcgtcga gatgaagaaa gaggatttcg acatctacga cgaggacgag aaccagagcc 2460
cccgtcctt ccagaagaaa acccggcact acttcattgc cgcctggaa agactgtggg 2520
actacggcat gagcagcagc cccacgtgc tgcggaacag agcccagagc ggcagcgtgc 2580
cccagttcaa gaaagtgtg ttccaggagt tcaccgacgg cagcttcacc cagcccctgt 2640
atcggggcga gctgaacgag cacctgggac tgctgggacc ttacattaga gccgaggtgg 2700
aagataacat catggtcacc ttcagaaacc aggcctccag accctacagc ttctacagca 2760
gcctgatcag ctacgaagag gaccagcggc agggcgccga accccggaag aacttcgtga 2820
agcccaacga gactaagacc tacttctgga aggtgcagca ccacatggcc cccacaaagg 2880
acgagttcga ctgcaaggcc tgggcctact tctccgatgt ggacctgaa aaggacgtgc 2940
actctggcct gattggacct ctgctcgtct gccacaccaa cacctgaac cccgcccacg 3000
gccggcaggt cacagtgcag gaattgccc tgttcttcac catcttcgat gagacaaaga 3060
gctggtactt caccgagaac atggaagaa actgtagagc cccctgcaac atccagatgg 3120
aagatcctac cttcaaagag aactatcgtt tccacgccat caacggctac atcatggaca 3180
ccctgcccgg cctggtcattg gcccaggatc agagaatccg gtggtatctg ctgagcatgg 3240
gcagcaacga gaacatccac agcatccact tcagcggcca cgtgttcaca gtgcggaaga 3300
aagaagagta caagatggcc ctgtacaacc tgtaccccgg cgtgttcgag acagtggaaa 3360
tgctgcccag caaggccggc atctggcggg tggaatgtct gatcggcgag catctgcacg 3420
ccggaatgag caccctgttt ctggtgtaca gcaacaagtg ccagaccct ctgggcatgg 3480
ccagcggcca catccgggac ttccagatca ccgcctccgg ccagtacggc cagtgggcc 3540
ctaagctggc ccggctccac tactccggat ctatcaacgc ctggtccacc aaagagccct 3600
tcagctggat caaggtggac ctgctggccc ctatgatcat ccacggaatc aagaccaggg 3660
gcccagaca gaagttcagc agcctgtaca tcagccagtt catcatcatg tacagcctgg 3720
acggcaagaa gtggcagacc taccggggca acagcaccgg caccctgatg gtgttcttcg 3780
gcaacgtgga cagcagcggc atcaagcaca acatcttcaa ccccccatc attgcccgg 3840
acatccggct gcaccccacc cactacagca tccggctcac cctgcggatg gaactgatgg 3900

ES 2 848 703 T5

gctgcgacct gaactcttgc agcatgcccc tggggatgga aagcaaggcc atcagcgacg	3960
cccagatcac agccagcagc tacttcacca acatgttcgc cacctggtcc ccaagcaaag	4020
cccgcctgca tctccaaggc agaagcaatg cctggcggcc tcaggtcaac aaccccaaag	4080
aatggctcca ggtggacttt cagaaaacca tgaaggtcac aggcgtgacc acccagggcg	4140
tgaaaagcct gctgacctct atgtacgtga aagagttcct gatcagcagc agccaggacg	4200
ggcaccagtg gaccctgttc tttcagaacg gcaaagtgaa agtgttcag ggcaaccagg	4260
actcctttac ccccggtggtc aactctctag accctccact gctgaccaga tacctgagaa	4320
tccaccctca gtcctgggtg caccagattg ccctgagaat ggaagtgctg ggatgcgagg	4380
cccaggatct gtactgataa ctcgagtcga cttaattaa	4419

<210> 3

<211> 4371

5

<212> ADN

<213> Secuencia artificial

<220>

<223> FVIII BDD parental

10

<400> 3

ES 2 848 703 T5

atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc	60
accagaagat actacctggg tgcagtggaa ctgtcatggg actatatgca aagtgatctc	120
ggtgagctgc ctgtggacgc aagatctcct cctagagtgc caaaatcttt tccattcaac	180
acctcagctg tgtacaaaa gactctgttt gtagaattca cggatcacct tttcaacatc	240
gctaagccaa ggccaccctg gatgggtctg ctaggtccta ccatccaggc tgaggtttat	300
gatacagtg tcatcact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt	360
ggtgtatcct actggaaagc ttctgagga gctgaatatg atgatcagac cagtcaaagg	420
gagaaagaag atgataaagt cttccctggg ggaagccata catatgtctg gcaggctctg	480
aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat	540
gtggacctgg taaaagactt gaattcaggc ctcatggag ccctactagt atgtagagaa	600
gggagtctgg ccaaggaaa gacacagacc ttgcacaaat ttatactact ttttgctgta	660
tttgatgaag ggaaaagttg gcaactcagaa acaaagaact ccttgatgca ggatagggat	720
gctgcatctg ctcgggcctg gcctaaaatg cacacagtca atggttatgt aaacaggtct	780
ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcatgtgat tggaatgggc	840
accactcctg aagtgcactc aatattcctc gaaggtcaca catttcttgt gaggaacctat	900
cgccaggcgt ccttgaaaat ctgcacaata actttcctta ctgctcaaac actcttgatg	960
gaccttgac agtttctact gttttgtcat atctcttccc accaacatga tggcatggaa	1020
gcttatgtca aagtagacag ctgtccagag gaacccaac tacgaatgaa aaataatgaa	1080

ES 2 848 703 T5

gaagcggag actatgatga tgatcttact gattctgaaa tggatgtggt caggtttgat 1140
 gatgacaact ctcttcctt tatccaaatt cgctcagttg ccaagaagca tcctaaaact 1200
 tgggtacatt acattgctgc tgaagaggag gactgggact atgctccctt agtcctcgcc 1260
 cccgatgaca gaagttataa aagtcaatat ttgaacaatg gccctcagcg gattggtagg 1320
 aagtacaaaa aagtccgatt tatggcatac acagatgaaa cctttaagac tcgtgaagct 1380
 attcagcatg aatcaggaat cttgggacct ttactttatg ggaagttgg agacacactg 1440
 ttgattatat ttaagaatca agcaagcaga ccatataaca tctaccctca cggaatcact 1500
 gatgtccgtc ctttgtattc aaggagatta ccaaaggtg taaaacattt gaaggatttt 1560
 ccaattctgc caggagaaat attcaaatat aaatggacag tgactgtaga agatgggcca 1620
 actaaatcag atcctcgggtg octgaccgcg tattactcta gtttcgtaa tatggagaga 1680
 gatctagctt caggactcat tggccctctc ctcatctgct acaaagaatc tgtagatcaa 1740
 agaggaaaacc agataatgtc agacaagagg aatgtcatcc tgttttctgt atttgatgag 1800
 aaccgaagct ggtacctcac agagaatata caacgcttcc tcccacatcc agctggagtg 1860
 cagcttgagg atccagagtt ccaagcctcc aacatcatgc acagcatcaa tggctatggt 1920
 tttgatagtt tgcagttgtc agtttgtttg catgaggtgg catactggta cattctaagc 1980
 attggagcac agactgactt ctttctgtc ttcttctctg gatatacctt caaacacaaa 2040
 atggctctatg aagacacact caccctattc ccattctcag gagaaactgt cttcatgtcg 2100
 atggaaaacc caggtctatg gattctgggg tgccacaact cagactttcg gaacagaggc 2160
 atgaccgcct tactgaaggt ttctagttgt gacaagaaca ctggtgatta ttacgaggac 2220
 agttatgaag atatttcagc atacttgctg agtaaaaaca atgccattga accaagaagc 2280
 ttctctcaaa acccaccagt cttgaaacgc catcaacggg aaataactcg tactactctt 2340
 cagtcagatc aagaggaaat tgactatgat gataccatat cagttgaaat gaagaaggaa 2400
 gattttgaca ttatgatga ggatgaaaat cagagcccc gcagctttca aaagaaaaca 2460
 cgacactatt ttattgctgc agtggagagg ctctgggatt atgggatgag tagctcccca 2520
 catgttctaa gaaacagggc tcagagtggc agtgtccctc agttcaagaa agttgttttc 2580
 caggaattta ctgatggctc ctttactcag cccttatacc gtggagaact aatgaacat 2640
 ttgggactcc tggggccata tataagagca gaagttgaag ataatatcat ggtaactttc 2700
 agaaatcagg cctctcgtcc ctattccttc tattctagcc ttatttctta tgaggaagat 2760
 cagaggcaag gagcagaacc tagaaaaaac tttgtcaagc ctaatgaaac caaaacttac 2820
 ttttgaaag tgcaacatca tatggcacc actaaagatg agtttgactg caaagcctgg 2880
 gcttatttct ctgatgttga octggaaaaa gatgtgcaact caggcctgat tggaccctt 2940

ES 2 848 703 T5

ctggtctgcc acactaacac actgaaccct gctcatggga gacaagtgac agtacaggaa 3000
 tttgctctgt ttttcacat ctttgatgag accaaaagct ggtacttcac tgaatatg 3060
 gaaagaaact gcagggctcc ctgcaatatc cagatggaag atcccacttt taaagagaat 3120
 tategcttcc atgcaatcaa tggctacata atggatacac tacctggctt agtaatggct 3180
 caggatcaaa ggattcgatg gtatctgctc agcatgggca gcaatgaaa catccattct 3240
 attcatttca gtggacatgt gttcactgta cgaaaaaag aggagtataa aatggcactg 3300
 tacaatctct atccaggtgt ttttgagaca gtggaaatgt taccatcaa agctggaatt 3360
 tggcgggtgg aatgccttat tggcgagcat ctacatgctg ggatgagcac actttttctg 3420
 gtgtacagca ataagtgtca gactcccctg ggaatggctt ctggacacat tagagatttt 3480
 cagattacag cttcaggaca atatggacag tgggccccaa agctggccag acttcattat 3540
 tccggatcaa tcaatgcctg gagcaccaag gagccctttt cttggatcaa ggtggatctg 3600
 ttggcaccaa tgattattca cggcatcaag acccaggtg cccgtcagaa gttctccagc 3660
 ctctacatct ctcagtttat catcatgtat agtcttgatg ggaagaagtg gcagacttat 3720
 cgaggaaatt ccactggaac cttaatggtc ttctttggca atgtggattc atctgggata 3780
 aaacacaata tttttaacc tccaattatt gctcgataca tccgtttgca cccaactcat 3840
 tatagcattc gcagcactct tcgcatggag ttgatgggct gtgatttaa tagttgcagc 3900
 atgccattgg gaatggagag taaagcaata tcagatgcac agattactgc ttcacctac 3960
 tttaccaata tgtttgccac ctggtctcct tcaaaagctc gacttcacct ccaagggagg 4020
 agtaatgcct ggagacctca ggtgaataat ccaaagagt ggctgcaagt ggacttccag 4080
 aagacaatga aagtcacagg agtaactact cagggagtaa aatctctgct taccagcatg 4140
 tatgtgaagg agttcctcat ctccagcagt caagatggcc atcagtggac tctctttttt 4200
 cagaatggca aagtaaaggt ttttcagga aatcaagact ccttcacacc tgtggtgaac 4260
 tctctagacc caccgttact gactcgctac cttcgaattc acccccagag ttgggtgcac 4320
 cagattgcc tgaggatgga ggttctgggc tgcgaggcac aggacctta c 4371

<210> 4

<211> 1438

5 <212> PRT

<213> Secuencia artificial

<220>

<223> FVIII BDD

10

<400> 4

Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
 1 5 10 15

ES 2 848 703 T5

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
 20 25 30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
 35 40 45

Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
 50 55 60

Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
 65 70 75 80

Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
 85 90 95

Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
 100 105 110

Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
 115 120 125

Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
 130 135 140

Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
 145 150 155 160

His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
 165 170 175

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

His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
 195 200 205

His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
 210 215 220

Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
 225 230 235 240

Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
 245 250 255

Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
 260 265 270

ES 2 848 703 T5

Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
 275 280 285

Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
 290 295 300

Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
 305 310 315 320

Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
 325 330 335

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
 340 345 350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
 355 360 365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
 370 375 380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
 385 390 395 400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
 405 410 415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
 420 425 430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
 435 440 445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
 450 455 460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
 465 470 475 480

Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
 485 490 495

His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
 500 505 510

Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
 515 520 525

ES 2 848 703 T5

Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
 530 535 540

Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
 545 550 555 560

Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
 565 570 575

Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
 580 585 590

Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
 595 600 605

Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
 610 615 620

Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
 625 630 635 640

Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
 645 650 655

Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
 660 665 670

Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp
 675 680 685

Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
 690 695 700

Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
 705 710 715 720

Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
 725 730 735

Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His
 740 745 750

Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile
 755 760 765

Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp

ES 2 848 703 T5

Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu
 1025 1030 1035

Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met
 1040 1045 1050

Gly Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val
 1055 1060 1065

Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn
 1070 1075 1080

Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys
 1085 1090 1095

Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His Leu His
 1100 1105 1110

Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln
 1115 1120 1125

Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile
 1130 1135 1140

Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg
 1145 1150 1155

Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro
 1160 1165 1170

Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His
 1175 1180 1185

Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr
 1190 1195 1200

Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp
 1205 1210 1215

Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe
 1220 1225 1230

Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro
 1235 1240 1245

Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser
 1250 1255 1260

ES 2 848 703 T5

Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn
 1265 1270 1275

Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp
 1280 1285 1290

Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr
 1295 1300 1305

Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn
 1310 1315 1320

Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val
 1325 1330 1335

Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly
 1340 1345 1350

Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile
 1355 1360 1365

Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn
 1370 1375 1380

Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro
 1385 1390 1395

Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg
 1400 1405 1410

Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu
 1415 1420 1425

Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
 1430 1435

<210> 5

<211> 6

5 <212> ADN

<213> Secuencia artificial

<220>

<223> MAR/ARS

10

<400> 5

atattt 6

<210> 6

<211> 6

<212> ADN

<213> Secuencia artificial

5

<220>

<223> MAR/ARS

<400> 6

aaatat 6

10

<210> 7

<211> 6

<212> ADN

<213> Secuencia artificial

15

<220>

<223> Posible sitio de corte y empalme

<400> 7

20

ggtgat 6

<210> 8

<211> 5

<212> ADN

25

<213> Secuencia artificial

<220>

<223> Secuencia desestabilizante

30

<400> 8

atta 5

<210> 9

<211> 5

35

<212> ADN

<213> Secuencia artificial

<220>

<223> Secuencia desestabilizante

<400> 9

5 taaat 5

<210> 10

<211> 6

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Secuencia de poli-T

15 <400> 10

ttttt 6

<210> 11

<211> 7

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Secuencia de poli-A

25

<400> 11

aaaaaaa 7

<210> 12

30 <211> 5

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Sitio de unión al promotor

<400> 12

tataa 5

<210> 13
<211> 5
5 <212> ADN
<213> Secuencia artificial

<220>
<223> Secuencia de unión al promotor
10

<400> 13
ttata 5

<210> 14
15 <211> 8
<212> ADN
<213> Secuencia artificial

<220>
20 <223> Elementos de secuencia ricos en AU (ARE)

<400> 14
atattatt 8

25 <210> 15
<211> 8
<212> ADN
<213> Secuencia artificial

30 <220>
<223> Elementos de secuencia ricos en AU (ARE)

<400> 15
atatttaa 8
35

<210> 16
<211> 13

ES 2 848 703 T5

<212> ADN

<213> Secuencia artificial

<220>

5 <223> Secuencia consenso de Kozak

<400> 16

gccgccacca tgc 13

10 <210> 17

<211> 32

<212> PRT

<213> Secuencia artificial

15 <220>

<223> Péptido CTP

<400> 17

Asp Pro Arg Phe Gln Asp Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser
1 5 10 15

Leu Pro Ser Pro Ser Arg Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu
20 25 30

20

<210> 18

<211> 28

<212> PRT

<213> Secuencia artificial

25

<220>

<223> Péptido CTP

<400> 18

Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser Leu Pro Ser Pro Ser Arg
1 5 10 15

Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln
20 25

30

<210> 19

ES 2 848 703 T5

<211> 11

<212> PRT

<213> Secuencia artificial

5

<220>

<223> Secuencia central de péptidos de unión a albúmina

<400> 19

Asp Ile Cys Leu Pro Arg Trp Gly Cys Leu Trp
1 5 10

10

<210> 20

<211> 20

<212> PRT

<213> Secuencia artificial

15

<220>

<223> Secuencia PAS

<400> 20

Ala Ser Pro Ala Ala Pro Ala Pro Ala Ser Pro Ala Ala Pro Ala Pro
1 5 10 15

20

Ser Ala Pro Ala
20

<210> 21

<211> 20

<212> PRT

25

<213> Secuencia artificial

<220>

<223> Secuencia PAS

30

<400> 21

Ala Ala Pro Ala Ser Pro Ala Pro Ala Ala Pro Ser Ala Pro Ala Pro
1 5 10 15

Ala Ala Pro Ser
20

ES 2 848 703 T5

<210> 22

<211> 20

<212> PRT

<213> Secuencia artificial

5

<220>

<223> Secuencia PAS

<400> 22

Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

10

Ser Pro Ser Ser
20

<210> 23

<211> 19

<212> PRT

15

<213> Secuencia artificial

<220>

<223> Secuencia PAS

20

<400> 23

Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

Ser Pro Ser

<210> 24

<211> 20

25

<212> PRT

<213> Secuencia artificial

<220>

<223> Secuencia PAS

30

<400> 24

ES 2 848 703 T5

Ser Ser Pro Ser Ala Pro Ser Pro Ser Ser Pro Ala Ser Pro Ser Pro
1 5 10 15

Ser Ser Pro Ala
20

<210> 25

<211> 24

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Secuencia PAS

10

<400> 25

Ala Ala Ser Pro Ala Ala Pro Ser Ala Pro Pro Ala Ala Ala Ser Pro
1 5 10 15

Ala Ala Pro Ser Ala Pro Pro Ala
20

<210> 26

15 <211> 20

<212> PRT

<213> Secuencia artificial

<220>

20 <223> Secuencia PAS

<400> 26

Ala Ser Ala Ala Ala Pro Ala Ala Ala Ser Ala Ala Ala Ser Ala Pro
1 5 10 15

Ser Ala Ala Ala
20

25 <210> 27

<211> 16842

<212> ADN

<213> Homo sapiens

ES 2 848 703 T5

<400> 27

atgattcctg ccagatttgc cggggtgctg cttgctctgg ccctcatttt gccagggacc 60
 ctttgtgcag aaggaactcg cggcaggta tccacggccc tactaaggac ggtctaacg 120
 gccccacgac gaacgagacc gggagtataa cggtcctctg gaaacacgtc ttccttgagc 180
 gccgtccagt aggtgccggg gatgcagcct tttcgggaagt gacttcgtca acacctttga 240
 tgggagcatg tacagctttg cgggatactg cagttacctc ctggcagggg gctgccagaa 300
 ctacgtcggg aaagccttca ctgaagcagt tgtggaaact accctcgtac atgtcgaaac 360
 gccctatgac gtcaatggag gaccgtcccc cgacggctctt acgctccttc tcgattattg 420
 gggacttcca gaatggcaag agagtgagcc tctccgtgta tcttggggaa ttttttgaca 480
 tccatttgtt tgtcaatggg tgcgaggaag agctaataac ccctgaaggt cttaccgttc 540
 tctcactcgg agaggcacat agaaccctt aaaaaactgt aggtaaacaa acagttacca 600
 accgtgacac agggggacca aagagtctcc atgccctatg cctccaaagg gctgtatcta 660
 gaaactgagg ctgggtacta caagctgtcc ggtgaggcct tggcactgtg tccccctggc 720
 ttctcagagg tacgggatac ggaggtttcc cgacatagat cttgactcc gaccatgat 780
 gttcgacagg cactccgga atggctttgt ggcaggatc gatggcagcg gcaactttca 840
 agtcctgctg tcagacagat acttcaacaa gacctgcggg ctgtgtggca actttaacat 900
 taccgaaaca ccggtcctag ctaccgtcgc cgttgaaagt tcaggacgac agtctgtcta 960
 tgaagtgtt ctggacgccc gacacacogt tgaattgta ctttctgaa gatgacttta 1020
 tgaccaaga agggacctg acctcggacc cttatgactt tgccaactca tgggctctga 1080
 gcagtggaga acagtgggtg gaaacgactt ctactgaaat actgggttct tccctggaac 1140
 tggagcctgg gaatactgaa acggttgagt acccgagact cgtcacctct tgtcaccaca 1200
 gaacgggcat ctctcccag cagctcatgc aacatctcct ctggggaaat gcagaagggc 1260
 ctgtgggagc agtgcagct tctgaagagc acctcgggtg cttgcccgta gaggagggtc 1320
 gtcgagtaag ttgtagagga gacccttta cgtcttcccg gacaccctcg tcacggtcga 1380
 agacttctcg tggagccaca ttgcccgtg ccaccctctg gtggaccccg agccttttgt 1440
 ggccctgtgt gagaagactt tgtgtgagtg tgctgggggg ctggagtgcg cctgccctgc 1500
 aacgggcgac ggtgggagac cacctggggc tcggaaaaca ccgggacaca ctcttctgaa 1560
 acacactcac acgaccccc gacctcacgc ggacgggagc cctcctggag tacgcccgga 1620
 cctgtgcccga ggagggaatg gtgctgtacg gctggaccga ccacagcgcg tgcagcccag 1680
 tgtgccctgc tggatggag ggaggacctc atgcccgcct ggacacgggt cctcccttac 1740
 cagacatgc cgacctggct ggtgtcgcgc acgtcgggtc acacgggagc accatacctc 1800
 tataggcagt gtgtgtcccc ttgcgccagg acctgccaga gcctgcacat caatgaaatg 1860
 tgtcaggagc gatgcgtgga tggctgcagc tgccctgagg atatccgtca cacacagggg 1920
 aacgcggtcc tggacggtct cggacgtgta gttactttac acagtccctg ctacgcacct 1980

ES 2 848 703 T5

accgacgtcg acgggactcc gacagctcct ggatgaaggc ctctgcgtgg agagcaccga 2040
 gtgtccctgc gtgcattccg gaaagcgcta ccctcccggc acctccctct ctcgagactg 2100
 ctgtcgagga cctacttccg gagacgcacc tctcgtggct cacagggacg cacgtaaggc 2160
 ctttcgcatg gggagggccg tggagggaga gagctctgac caacacctgc atttgccgaa 2220
 acagccagtg gatctgcagc aatgaagaat gtccagggga gtgccttgtc actggtcaat 2280
 cccacttcaa gagctttgac gttgtggacg taaacggcct tgtcggtcac ctagacgtcg 2340
 ttacttotta caggtccccc caccggaacag tgaccagtta gggagaagt ctcgaaactg 2400
 aacagatact tcacctcag tgggatctgc cagtacctgc tggcccggga ttgccaggac 2460
 cactccttct ccattgtcat tgagactgtc cagtgtgtcg ttgtctatga agtggaaagt 2520
 accctagacg gtcattggacg accgggcccct aacggtcctg gtgaggaaga ggtaacagta 2580
 actctgacag gtcacacgac atgaccgcga cgctgtgtgc acccgtccg tcaccgtccg 2640
 gctgcctggc ctgcacaaca gccttgtgaa actgaagcat ggggcaggag ttgccatgga 2700
 tactggcgct gcgacacacg tgggagggc agtggcaggc cgacggaccg gacgtgttgt 2760
 cggaacactt tgacttcgta ccccgctcctc aacggtaacct tggccaggac atccagctcc 2820
 ccctcctgaa aggtgacctc cgcattccagc atacagtgac ggccctcctg cgcctcagct 2880
 acggggagga cctgcagatg accggtcctg taggtcgagg gggaggactt tccactggag 2940
 gcgtaggctg tatgtcactg ccggaggcac gcggagtcca tgcacctcct ggacgtctac 3000
 gactgggatg gccgcgggag gctgctggtg aagctgtccc ccgtctatgc cgggaagacc 3060
 tgcggcctgt gtgggaatta caatggcaac cagggcgacg ctgacctac cggcgcctc 3120
 cgacgaccac ttcgacaggg ggcagatacg gcccttctgg acgcggaca cacccttaat 3180
 gttaccgttg gtcccgtgc acttccttac cccctctggg ctggcrgagc cccgggtgga 3240
 ggacttcggg aacgcctgga agctgcacgg ggactgccag gacctgcaga agcagcacag 3300
 tgaaggaatg ggggagacc gaccgctcg gggcccacct cctgaagccc ttgcggacct 3360
 tcgacgtgcc cctgacggtc ctggacgtct tagtcgtgtc cgatccctgc gccctcaacc 3420
 cgcgcatgac caggttctcc gaggaggcgt gcgcggcctc gacgtccccc acattcgagg 3480
 cctgccatcg tgccgtcagc gctagggacg cgggagtgg gcgcgtactg gtccaagagg 3540
 ctcctccgca cgcgccagga ctgcagggg tgtaagctcc ggacggtagc acggcagtcg 3600
 ccgtgccctt acctgcggaa ctgccgtac gacgtgtgct cctgctcggg cggccgcgag 3660
 tgcctgtgcg gcgccctggc cagctatgcc gcggcctgcg ggcgacggga tggacgcctt 3720
 gacggcgtat ctgcacacga ggacgagcct gccggcgctc acggacacgc cgcgggaccg 3780
 gtcgatacgg cgcgggacgc cggggagagg cgtgcgcgtc gcgtggcgcg agccaggccc 3840
 ctgtgagctg aactgcccga aaggccaggt gtacctgcag tgcgggacc cctgcaacct 3900

ES 2 848 703 T5

gccctctcc gcacgcgag cgcaccggc tcggtccgc gacactcgac ttgacgggt 3960
 ttccggcca catggacgtc acgccctggg ggacgttga gacctgccgc tctctctctt 4020
 acccggatga ggaatgcaat gaggcctgcc tggagggctg cttctgcccc ccagggctct 4080
 acatggatga gaggggggac ctggacggcg agagagagaa tgggcctact ccttacgtta 4140
 ctccggacgg acctcccgac gaagacgggg ggtcccgaga tgtacctact ctccccctg 4200
 tgcgtgcca aggccagtg cccctgttac tatgacggtg agatcttcca gccagaagac 4260
 atcttctcag accatcacac catgtgctac tgtgaggatg acgcacgggt tccgggtcac 4320
 ggggacaatg atactgccac tctagaaggt cggctctctg tagaagagtc tggtagtggtg 4380
 gtacacgatg aactcctac gcttcatgca ctgtaccatg agtggagtcc ccggaagctt 4440
 gctgcctgac gctgtcctca gcagtcacct gtctcatcgc agcaaaagga gcctatcctg 4500
 cgaagtacgt gacatggtac tcacctcagg gcccttcgaa cgacggactg cgacaggagt 4560
 cgtcagggga cagagtagcg tcgttttctt cggataggac tcggcccccc atgggtcaagc 4620
 tgggtgtgctc cgctgacaac ctgcgggctg aagggtctga gtgtaccaa acgtgccaga 4680
 actatgacct ggagtgcatt agccgggggg taccagttcg accacacagg gcgactgttg 4740
 gacgcccgac ttcccgagct cacatggttt tgcacggctt tgatactgga cctcacgtac 4800
 agcatgggct gtgtctctgg ctgcctctgc cccccggca tggtcggca tgagaacaga 4860
 tgtgtggccc tggaaaggtg tccctgcttc catcagggca tcgtaccoga cacagagacc 4920
 gacggagacg gggggcccgt accaggccgt actcttctct acacaccggg acctttccac 4980
 agggacgaag gtagtcccgt aggagtatgc ccttgagaa acagtgaaga ttggctgcaa 5040
 cacttgtgtc tgtcgggacc ggaagtggaa ctgcacagac catgtgtgtg atgccacgtg 5100
 tctcatacag gggacctctt tgtcacttct aaccgacgtt gtgaacacag acagccctgg 5160
 ccttcacctt gacgtgtctg gtacacacac tacgggtgac ctccacgatc ggcattggccc 5220
 actacctcac cttcgacggg ctcaaatacc tgttccccgg ggagtgccag tacgttctgg 5280
 tgcaggatta ctgcggcagt gaggtgctag ccgtaccggg tgatggagtg gaagctgccc 5340
 gagtttatgg acaaggggcc cctcacggtc atgcaagacc acgtcctaata gacgccgtca 5400
 aaccctggga cctttcggat cctagtgggg aataagggat gcagccacc ctcagtgaaa 5460
 tgcaagaaac gggtcacat cctggtggag ggaggagaga ttgggaccct ggaaagccta 5520
 ggatcaccct ttattcccta cgtcgggtgg gagtcacttt acgttctttg ccagtggtta 5580
 ggaccacctc cctcctctct ttgagctgtt tgacggggag gtgaatgtga agaggcccat 5640
 gaaggatgag actcactttg aggtggtgga gtctggccgg tacatcattc tgctgctggg 5700
 aactcgaaa actgcccctc cacttacact tctccgggta cttcctactc tgagtgaaac 5760

ES 2 848 703 T5

tccaccacct cagaccggcc atgtagtaag acgacgaccc caaagccctc tccgtggtct 5820
 gggaccgccca cctgagcadc tccgtggtcc tgaagcagac ataccaggag aaagtgtgtg 5880
 gcctgtgtgg gaattttgat gtttcgggag aggcaccaga ccctggcggg ggactcgtag 5940
 aggcaccagg acttcgtctg tatggtcctc tttcacacac cggacacacc cttaaaacta 6000
 ggcacccaga acaatgacct caccagcagc aacctccaag tggaggaaga ccctgtggac 6060
 tttgggaact cctggaaagt gagctcgcag tgtgctgaca ccgtaggtct tgttactgga 6120
 gtggtcgtcg ttggagggtc acctccttct gggacacctg aaacccttga ggacctttca 6180
 ctcgagcgtc acacgactgt ccagaaaagt gcctctggac tcatccctg ccacctgcca 6240
 taacaacatc atgaagcaga cgatggtgga ttctctctgt agaatcctta ccagtgcagt 6300
 ggtcttttca cggagacctg agtaggggac ggtggacggt attgttgtag tacttcgtct 6360
 gctaccacct aaggaggaca tcttaggaat ggtcactgca cttccaggac tgcaacaagc 6420
 tgggtgaccc cgagccatat ctggatgtct gcatttacga cacctgctcc tgtgagtcca 6480
 ttggggactg cgctccttc gaaggtcctg acgttgctcg accacctggg gctcgggata 6540
 gacctacaga cgtaaatgct gtggacgagg acaactcagg aaccctgac gcggacgaag 6600
 tgcgacacca ttgctgccta tgcccacgtg tgtgccacg atggcaaggt ggtgacctgg 6660
 aggacggcca cattgtgcc ccagagctgc gaggagagga acgctgtggt aacgacggat 6720
 acgggtgcac acacgggtcg taccgttcca ccaactggacc tcctgccggt gtaaacacggg 6780
 ggtctcgcag ctccctcct atctccggga gaacgggtat gagtgtgagt ggcgctataa 6840
 cagctgtgca cctgcctgtc aagtcacgtg tcagcacct gagccactgg cctgcctgt 6900
 tagaggccct cttgccata ctcaactca ccgcgatatt gtcgacacgt ggacggacag 6960
 ttcagtgcac agtcgtggga ctcggtgacc ggcgggaca gcagtgtgtg gagggctgcc 7020
 atgccactg ccctccaggg aaaatcctgg atgagctttt gcagacctgc gttgaccctg 7080
 aagactgtcc agtgtgtgag cgtcacacac ctcccacgg tacgggtgac gggaggtccc 7140
 ttttaggacc tactcgaaaa cgtctggacg caactgggac ttctgacagg tcacacactc 7200
 gtggctggcc ggcgttttgc ctcaggaag aaagtcacct tgaatcccag tgaccctgag 7260
 cactgccaga tttgccactg tgatgttgtc aacctcacct caccgaccgg ccgcaaacg 7320
 gagtcccttc tttcagtgga acttagggtc actgggactc gtgacggtct aaacgggtgac 7380
 actacaacag ttggagtgga gtgaagcctg ccaggagccg ggaggcctgg tgggtcctcc 7440
 cacagatgcc ccggtgagcc ccaccactct gtatgtggag gacatctcgg aaccgccgtt 7500
 cacttcggac ggtcctcggc cctccggacc accacggagg gtgtctacgg ggccactcgg 7560
 ggtggtgaga catacacctc ctgtagagcc ttggcggcaa gcacgatttc tactgcagca 7620
 ggctactgga cctggtcttc ctgctggatg gctcctccag gctgtccgag gctgagtttg 7680

ES 2 848 703 T5

aagtgctgaa ggcctttgtg cgtgctaaag atgacgtcgt ccgatgacct ggaccagaag 7740
gacgacctac cgaggaggtc cgacaggctc cgactcaaac ttcacgactt ccggaaacac 7800
gtggacatga tggagcggct gcgcactccc cagaagtggg tccgcgtggc cgtggtggag 7860
taccacgacg gctcccacgc ctacatcggg ctcaaggacc cacctgtact acctcgccga 7920
cgcgtagagg gtcttcaccc aggcgcaccg gcaccacctc atggtgctgc cgagggtgcg 7980
gatgtagccc gagtctctgg ggaagcgacc gtcagagctg cggcgcattg ccagccaggt 8040
gaagtatgcg ggcagccagg tggcctccac cagcgaggtc ttgaaataca cactgttcca 8100
ccttcgctgg cagtctcgac gccgcgtaac ggtcggcca cttcatacgc ccgtcggccc 8160
accggaggtg gtcgctccag aactttatgt gtgacaaggc aatcttcagc aagatcgacc 8220
gccctgaagc ctcccgcac gccctgctcc tgatggccag ccaggagccc caacggatgt 8280
cccggaactt tgtccgctac ttagaagtcg ttctagctgg cgggacttcg gagggcgtag 8340
cgggacgagg actaccggtc ggtcctcggg gttgcctaca gggccttga acaggcgatg 8400
gtccagggcc tgaagaagaa gaaggcatt gtgatcccgg tgggcattgg gcccctgcc 8460
aacctcaagc agatccgcct catcgagaag caggcccctg caggtcccgg acttctctt 8520
cttccagtaa cactagggcc acccgtaacc cggggtacgg ttggagtctg tctaggcggga 8580
gtagctcttc gtccggggac agaacaaggc cttcgtgctg agcagtgtgg atgagctgga 8640
gcagcaaagg gacgagatcg ttagctacct ctgtgacctt gccctgaag cccctcctcc 8700
tcttgctcgg gaagcacgac tcgtcacacc tactcgacct cgtcgtttcc ctgctctagc 8760
aatcgatgga gacactgga oggggacttc ggggaggagg tactctgccc ccgacatgg 8820
cacaagtcac tgtgggcccg gggctcttgg gggtttcgac cctggggccc aagaggaact 8880
ccatggttct ggatgtggcg atgagacggg gggctgtacc gtgttcagtg acaccggggc 8940
cccgagaacc cccaaagctg ggaccccggg ttctccttga ggtaccaaga cctacaccgc 9000
ttcgtcctgg aaggatcgg caaaattggt gaagccgact tcaacaggag caaggagttc 9060
atggaggagg tgattcagcg gatggatgtg gcccaggaca aagcaggacc ttcttagcct 9120
gttttaacca cttcggctga agttgtcctc gttcctcaag tacctcctcc actaagtccg 9180
ctacctacac ccggtcctgt gcatccacgt cacgggtgctg cagtactcct acatggtgac 9240
cgtggagtac cccttcagcg aggcacagtc caaaggggac atcctgcagc ggggtcgaga 9300
cgtagtgca gtgccacgac gtcagtagga tgtaccactg gcacctcatg gggagtcgc 9360
tccgtgtcag gtttcccctg taggacgtcg cccacgctct gatccgctac cagggcggca 9420
acaggaccaa cactgggctg gccctgcggt acctctctga ccacagcttc ttggtcagcc 9480
agggtgaccg ggagcaggcg ctaggcgatg gtcccgccgt tgtcctggtt gtgacccgac 9540

ES 2 848 703 T5

cgggacgcca	tggagagact	ggtgtcgaag	aaccagtcgg	tcccactggc	cctcgtccgc	9600
cccaacctgg	tctacatggt	caccggaat	cctgcctctg	atgagatcaa	gaggctgcct	9660
ggagacatcc	aggtggtgcc	cattggagtg	ggccctaatag	gggttggacc	agatgtacca	9720
gtggccttta	ggacggagac	tactctagtt	ctccgacgga	cctctgtagg	tccaccacgg	9780
gtaacctcac	ccgggattac	ccaacgtgca	ggagctggag	aggattggct	ggcccaatgc	9840
ccctatcctc	atccaggact	ttgagacgct	ccccgagag	gctcctgacc	tgggtgctgca	9900
ggttgcacgt	cctcgacctc	tcctaaccga	ccgggttacg	gggataggag	taggtcctga	9960
aactctgcga	gggggtctctc	cgaggactgg	accacgacgt	gaggtgctgc	tccggagagg	10020
ggctgcagat	ccccacctc	tcccctgcac	ctgactgcag	ccagcccctg	gacgtgatcc	10080
ttctcctgga	tggctcctcc	ctccacgacg	aggcctctcc	ccgacgtcta	ggggtgggag	10140
aggggacgtg	gactgacgtc	ggtcggggac	ctgcactagg	aagaggacct	accgaggagg	10200
agtttcccag	cttcttattt	tgatgaaatg	aagagtctcg	ccaaggcttt	catttcaaaa	10260
gccaatatag	ggcctcgtct	cactcaggtg	tcagtgtctg	tcaaagggtc	gaagaataaa	10320
actactttac	ttctcaaagc	ggttccgaaa	gtaaagtttt	cggttatatc	ccggagcaga	10380
gtgagtccac	agtcacgacg	agtatggaag	catcaccacc	attgacgtgc	catggaacgt	10440
ggtcccggag	aaagcccatt	tgctgagcct	tgtggacgtc	atgcagcggg	agggaggccc	10500
tcataccttc	gtagtgggtg	taactgcacg	gtaccttgca	ccagggcctc	tttcgggtaa	10560
acgactogga	acacctgcag	tacgtcgccc	tccctccggg	cagccaaatc	ggggatgcct	10620
tgggctttgc	tgtgcgatac	ttgacttcag	aaatgcatgg	tgccaggccg	ggagcctcaa	10680
aggcggtggt	catcctggtc	gtcggtttag	cccctacgga	acccgaaacg	acacgctatg	10740
aactgaagtc	tttacgtacc	acggtccggc	cctcggagtt	tccgccacca	gtaggaccag	10800
acggacgtct	ctgtggattc	agtggatgca	gcagctgatg	ccgccaggtc	caacagagtg	10860
acagtgttcc	ctattggaat	tggagatcgc	tacgatgcag	tgctgcaga	gacacctaa	10920
tcacctacgt	cgtcgactac	ggcgggtccag	gttgtctcac	tgtcacaagg	gataacctta	10980
acctctagcg	atgctacgtc	cccagctacg	gatcttggca	ggcccagcag	gcgactccaa	11040
cgtggtgaag	ctccagcga	togaagacct	ccctaccatg	gtcaccttgg	gcaattcctt	11100
gggtcgatgc	ctagaacctg	ccgggtcgtc	cgctgaggtt	gcaccacttc	gaggtcgctt	11160
agcttctgga	gggatggtac	cagtggaacc	cgtaagga	cctccacaaa	ctgtgctctg	11220
gatttggttag	gatttgcatg	gatgaggatg	ggaatgagaa	gaggccggg	gacgtctgga	11280
ccttgccaga	ccagtgccac	ggaggtgttt	gacacgagac	ctaaacaatc	ctaaacgtac	11340
ctactcctac	cottactcct	ctccgggccc	ctgcagacct	ggaacggtct	ggtcacgggtg	11400
accgtgactt	gccagccaga	tggccagacc	ttgctgaaga	gtcatcgggt	caactgtgac	11460

ES 2 848 703 T5

cgggggctga ggccttcgtg ccctaacagc cagtcccctg tggcactgaa cggtcggtct 11520
 accggtctgg aacgacttct cagtagccca gttgacactg gcccccgact cgggaagcac 11580
 gggattgtcg gtcaggggac ttaaagtgga agagacctgt ggctgccgct ggacctgcc 11640
 ctgygtgtgc acaggcagct ccactcggca catcgtgacc tttgatgggc agaatttcaa 11700
 aatttcacct tctctggaca ccgacggcga cctggacggg gacrcacacg tgtccgtcga 11760
 ggtgagccgt gtagcactgg aaactaccg tcttaaagtt gctgactggc agctgttctt 11820
 atgtcctatt tcaaaacaag gagcaggacc tggaggtgat tctccataat ggtgcctgca 11880
 gccctggagc aaggcagggc cgactgaccg tcgacaagaa tacaggataa agttttgttc 11940
 ctcgtcctgg acctccacta agaggtatta ccacggacgt cgggacctcg ttccgtcccg 12000
 tgcataaat ccatcgaggt gaagcacagt gccctctccg tcgagstgca cagtgacatg 12060
 gaggtgacgg tgaatgggag actggtctct gttccttacg acgtacttta ggtagctcca 12120
 cttcgtgtca cgggagaggc agctcsacgt gtcactgtac ctccactgcc acttacctc 12180
 tgaccagaga caaggaatgc tgggtgggaa catggaagtc aacgtttatg gtgccatcat 12240
 gcatgaggtc agattcaatc accttggta catcttcaca ttcactccac aaaacaatga 12300
 acccaccctt gtacctcag ttgcaaatac cacggtagta cgtactccag tctaagttag 12360
 tggaaaccagt gtagaagtgt aagtgaggtg ttttgttact gttccaactg cagctcagcc 12420
 ccaagacttt tgcttcaaag acgtatggtc tgtgtgggat ctgtgatgag aacggagcca 12480
 atgacttcat gctgagggat caaggttgac gtcgagtcgg ggttctgaaa acgaagtttc 12540
 tgcataccag acacacccta gacactactc ttgctcgggt tactgaagta cgactcccta 12600
 ggcacagtca ccacagactg gaaaacactt gttcaggaat ggactgtgca ggggccaggg 12660
 cagacgtgcc agcccatcct ggaggagcag tgtcttgtcc ccgtgtcagt ggtgtctgac 12720
 cttttgtgaa caagtcctta cctgacacgt cgccgggtccc gtctgcacgg tcgggtagga 12780
 cctcctcgtc acagaacagg ccgacagctc ccaactgccag gtcctcctct taccactgtt 12840
 tgctgaatgc cacaaggtcc tggctccagc cacattctat gccatctgcc agcaggacag 12900
 ggctgtcgag ggtgacggtc caggaggaga atggtgacaa acgacttacg gtgttccagg 12960
 accgaggtcg gtgtaagata cggtagacgg tcgtcctgtc ttgccaccag gagcaagtgt 13020
 gtgaggtgat cgctcttat gccacctct gtcggaccaa cggggtctgc gttgactgga 13080
 ggacacctga tttctgtgct aacggtggtc ctggttcaca cactccacta ggggagaata 13140
 cgggtggaga cagcctggtt gcccagacg caactgacct cctgtggact aaagacacga 13200
 atgtcatgcc caccatctct ggtctacaac cactgtgagc atggctgtcc cggcactgt 13260
 gatggcaacg tgagctcctg tggggaccat ccctccgaag tacagtacgg gtggtagaga 13320

ES 2 848 703 T5

ccagatggtg gtgacactcg taccgacagg ggccgtgaca ctaccgttgc actcgaggac 13380
 acccctggta gggaggcttc gctgtttctg ccctccagat aaagtcattg tggaaaggcag 13440
 ctgtgtccct gaagaggcct gcaactcagtg cattgggtgag gatggagtcc agcaccagtt 13500
 cgacaaagac gggaggtcta tttcagtaca accttccgtc gacacaggga cttctccgga 13560
 cgtgagtcac gtaaccactc ctacctcagg tcgtgggtcaa cctggaagcc tgggtcccgg 13620
 accaccagcc ctgtcagatc tgcacatgcc tcagcgggcg gaaggtcaac tgcacaacgc 13680
 agccctgccc caccggccaaa ggaccttcgg acccaggggcc tgggtggtcgg gacagtctag 13740
 acgtgtacgg agtcgcccgc cttccagttg acgtgttgcg tcgggacggg gtgcgggttt 13800
 gctcccacgt gtggcctgtg tgaagttagc cgctccgcc agaatgcaga ccagtgtgc 13860
 cccgagtatg agtgtgtgtg tgaccacagtg agctgtgacc cgagggtgca caccggacac 13920
 acttcatcgg gcggaggcgg tcttacgtct ggtcacgacg gggctcatac tcacacacac 13980
 actgggtcac tgcacactgg tgccccagt gcctcactgt gaacgtggcc tccagcccac 14040
 actgaccaac cctggcgagt gcagacccaa cttcacctgc gcctgcagga aggaggagtg 14100
 acgggggtca cggagtgaca cttgcaccgg aggtcgggtg tgactggttg ggaccgctca 14160
 cgtctgggtt gaagtggacg cggacgtcct tcctcctcac caaaagagtg tccccacct 14220
 cctgcccccc gcaccgtttg cccacccttc ggaagaccga gtgctgtgat gagtatgagt 14280
 gtgcctgcaa ctgtgtcaac gttttctcac aggggtggga ggacgggggg cgtggcaaac 14340
 ggggtgggaag ccttctgggt caccgacacta ctcatactca cacggacgtt gacacagttg 14400
 tccacagtga gctgtccctt tgggtacttg gcctcaaccg ccaccaatga ctgtggctgt 14460
 accacaacca cctgccttcc cgacaagggt tgtgtccacc aggtgtcact cgacagggga 14520
 acccatgaac cggagtggc ggtggttact gacaccgaca tgggtgttggg ggacggaagg 14580
 gctgttccac acacaggtgg gaagcaccat ctaccctgtg ggcagttct gggaggaggg 14640
 ctgcgatgtg tgcacctgca ccgacatgga ggatgccgtg atgggcctcc gcgtggccca 14700
 cttcgtggta gatgggacac ccggtcaaga cctcctccc gacgctacac acgtggacgt 14760
 ggctgtacct cctacggcac taccggagg cgcaccgggt gtgctcccag aagccctgtg 14820
 aggacagctg tcggtcgggc ttcaactacg ttctgcatga aggcgagtgc tgtggaagg 14880
 gcctgccatc tgctgtgag cacgagggtc ttccgggacac tcctgtcgac agccagcccc 14940
 aagtgaatgc aagacgtact tccgctcacg acaccttcca cggacggtag acggacactc 15000
 gtggtgactg gctcaccgcg gggggactcc cagtcttctt ggaagagtgt cggtcccag 15060
 tgggcctccc cggagaacct ctgcctcatc aatgagtgtg caccactgac cgagtggcgc 15120
 cccctgagg gtcagaagga cttctcaca gccgagggtc acccggaggg gcctcttggg 15180
 gacggagttag ttactcacac tccgagtga ggaggaggtc tttatacaac aaaggaacgt 15240

ES 2 848 703 T5

ctcctgcccc cagctggagg tccctgtctg cccctcgggc tttcagctga gctgtaagac 15300
aggctcactt cctcctccag aaatatgttg tttccttgca gaggacgggg gtcgacctcc 15360
agggacagac ggggagcccg aaagtgcact cgacattctg ctgagcgtgc tgcccaagct 15420
gtcgctgtga gcgcatggag gcctgcatgc tcaatggcac tgtcattggg cccggaaga 15480
ctgtgatgat cgatgtgtgc gagtcgcacg acgggttcga cagcgacact cgcgtaacctc 15540
cggacgtacg agttaccgtg acagtaacct gggcccttct gacactacta gctacacacg 15600
acgacctgcc gctgcatggg gcagggtggg gtcactctctg gattcaagct ggagtgcagg 15660
aagaccacct gcaaccctg cccctcgggt tacaaggaag tgctggacgg cgacgtacca 15720
cgtccacccc cagtagagac ctaagttcga cctcacgtcc ttctggtgga cgttggggac 15780
gggggacca atgttccttc aaaataacac aggtgaatgt tgtgggagat gttgcctac 15840
ggcttgacc attcagctaa gaggaggaca gatcatgaca ctgaagcgtg atgagacgct 15900
ttttattgtg tccacttaca acaccctcta caaacggatg ccgaacgtgg taagtcgatt 15960
ctcctcctgt ctagtactgt gacttcgcac tactctgcga ccaggatggc tgtgatactc 16020
acttctgcaa ggtcaatgag agaggagagt acttctggga gaagagggtc acaggctgcc 16080
cacccttga tgaacacaag ggtcctaccg acaactatgag tgaagacggt ccagttactc 16140
tctcctctca tgaagaccct cttctcccag tgtccgacgg gtgggaaact acttgtgttc 16200
tgtcttgctg agggaggtaa aattatgaaa attccaggca cctgctgtga cacatgtgag 16260
gagcctgagt gcaacgacat cactgccagg ctgcagtatg acagaacgac tccctccatt 16320
ttaatacttt taagggtcgt ggacgacact gtgtacactc ctcgactca cgttgctgta 16380
gtgacgggtcc gacgtcatac tcaagggtggg aagctgtaag tctgaagtag aggtggatat 16440
ccactactgc cagggcaaat gtgccagcaa agccatgtac tccattgaca tcaacgatgt 16500
agttccaccc ttcgacatc agacttcac tccacctata ggtgatgacg gtcccgttta 16560
cacggtcggt tcggtacatg aggttaactgt agttgctaca gcaggaccag tgctcctgct 16620
gctctccgac acggacggag cccatgcagg tggccctgca ctgcaccaat ggctctgttg 16680
tgtaccatga ggttctcaat cgtcctggtc acgaggacga cgagaggctg tgctgcctc 16740
gggtacgtcc accgggacgt gacgtggtta ccgagacaac acatggtact ccaagagtta 16800
gccatggagt gcaaatgctc ccccaggaag tgcagcaagt ga 16842

<210> 28

<211> 2813

5 <212> PRT

<213> Homo sapiens

<220>

<221> característica_misc

10 <222> (2016)..(2016)

ES 2 848 703 T5

<223> Xaa puede ser cualquier aminoácido que existe de forma natural

<400> 28

```

Met Ile Pro Ala Arg Phe Ala Gly Val Leu Leu Ala Leu Ala Leu Ile
1           5           10           15

Leu Pro Gly Thr Leu Cys Ala Glu Gly Thr Arg Gly Arg Ser Ser Thr
20           25           30

Ala Arg Cys Ser Leu Phe Gly Ser Asp Phe Val Asn Thr Phe Asp Gly
35           40           45

Ser Met Tyr Ser Phe Ala Gly Tyr Cys Ser Tyr Leu Leu Ala Gly Gly
50           55           60

Cys Gln Lys Arg Ser Phe Ser Ile Ile Gly Asp Phe Gln Asn Gly Lys
65           70           75           80

Arg Val Ser Leu Ser Val Tyr Leu Gly Glu Phe Phe Asp Ile His Leu
85           90           95

Phe Val Asn Gly Thr Val Thr Gln Gly Asp Gln Arg Val Ser Met Pro
100          105          110

Tyr Ala Ser Lys Gly Leu Tyr Leu Glu Thr Glu Ala Gly Tyr Tyr Lys
115          120          125

Leu Ser Gly Glu Ala Tyr Gly Phe Val Ala Arg Ile Asp Gly Ser Gly
130          135          140

Asn Phe Gln Val Leu Leu Ser Asp Arg Tyr Phe Asn Lys Thr Cys Gly
145          150          155          160

Leu Cys Gly Asn Phe Asn Ile Phe Ala Glu Asp Asp Phe Met Thr Gln
165          170          175

Glu Gly Thr Leu Thr Ser Asp Pro Tyr Asp Phe Ala Asn Ser Trp Ala
180          185          190

Leu Ser Ser Gly Glu Gln Trp Cys Glu Arg Ala Ser Pro Pro Ser Ser
195          200          205

Ser Cys Asn Ile Ser Ser Gly Glu Met Gln Lys Gly Leu Trp Glu Gln
210          215          220

Cys Gln Leu Leu Lys Ser Thr Ser Val Phe Ala Arg Cys His Pro Leu

```


ES 2 848 703 T5

Arg Ile Gln His Thr Val Thr Ala Ser Val Arg Leu Ser Tyr Gly Glu
 485 490 495

Asp Leu Gln Met Asp Trp Asp Gly Arg Gly Arg Leu Leu Val Lys Leu
 500 505 510

Ser Pro Val Tyr Ala Gly Lys Thr Cys Gly Leu Cys Gly Asn Tyr Asn
 515 520 525

Gly Asn Gln Gly Asp Asp Phe Leu Thr Pro Ser Gly Leu Ala Glu Pro
 530 535 540

Arg Val Glu Asp Phe Gly Asn Ala Trp Lys Leu His Gly Asp Cys Gln
 545 550 555 560

Asp Leu Gln Lys Gln His Ser Asp Pro Cys Ala Leu Asn Pro Arg Met
 565 570 575

Thr Arg Phe Ser Glu Glu Ala Cys Ala Val Leu Thr Ser Pro Thr Phe
 580 585 590

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

Arg Tyr Asp Val Cys Ser Cys Ser Asp Gly Arg Glu Cys Leu Cys Gly
 610 615 620

Ala Leu Ala Ser Tyr Ala Ala Ala Cys Ala Gly Arg Gly Val Arg Val
 625 630 635 640

Ala Trp Arg Glu Pro Gly Arg Cys Glu Leu Asn Cys Pro Lys Gly Gln
 645 650 655

Val Tyr Leu Gln Cys Gly Thr Pro Cys Asn Leu Thr Cys Arg Ser Leu
 660 665 670

Ser Tyr Pro Asp Glu Glu Cys Asn Glu Ala Cys Leu Glu Gly Cys Phe
 675 680 685

Cys Pro Pro Gly Leu Tyr Met Asp Glu Arg Gly Asp Cys Val Pro Lys
 690 695 700

Ala Gln Cys Pro Cys Tyr Tyr Asp Gly Glu Ile Phe Gln Pro Glu Asp
 705 710 715 720

Ile Phe Ser Asp His His Thr Met Cys Tyr Cys Glu Asp Gly Phe Met
 725 730 735

ES 2 848 703 T5

His Cys Thr Met Ser Gly Val Pro Gly Ser Leu Leu Pro Asp Ala Val
 740 745 750

Leu Ser Ser Pro Leu Ser His Arg Ser Lys Arg Ser Leu Ser Cys Arg
 755 760 765

Pro Pro Met Val Lys Leu Val Cys Pro Ala Asp Asn Leu Arg Ala Glu
 770 775 780

Gly Leu Glu Cys Thr Lys Thr Cys Gln Asn Tyr Asp Leu Glu Cys Met
 785 790 795 800

Ser Met Gly Cys Val Ser Gly Cys Leu Cys Pro Pro Gly Met Val Arg
 805 810 815

His Glu Asn Arg Cys Val Ala Leu Glu Arg Cys Pro Cys Phe His Gln
 820 825 830

Gly Lys Glu Tyr Ala Pro Gly Glu Thr Val Lys Ile Gly Cys Asn Thr
 835 840 845

Cys Val Cys Arg Asp Arg Lys Trp Asn Cys Thr Asp His Val Cys Asp
 850 855 860

Ala Thr Cys Ser Thr Ile Gly Met Ala His Tyr Leu Thr Phe Asp Gly
 865 870 875 880

Leu Lys Tyr Leu Phe Pro Gly Glu Cys Gln Tyr Val Leu Val Gln Asp
 885 890 895

Tyr Cys Gly Ser Asn Pro Gly Thr Phe Arg Ile Leu Val Gly Asn Lys
 900 905 910

Gly Cys Ser His Pro Ser Val Lys Cys Lys Lys Arg Val Thr Ile Leu
 915 920 925

Val Glu Gly Gly Glu Ile Glu Leu Phe Asp Gly Glu Val Asn Val Lys
 930 935 940

Arg Pro Met Lys Asp Glu Thr His Phe Glu Val Val Glu Ser Gly Arg
 945 950 955 960

Tyr Ile Ile Leu Leu Leu Gly Lys Ala Leu Ser Val Val Trp Asp Arg
 965 970 975

His Leu Ser Ile Ser Val Val Leu Lys Gln Thr Tyr Gln Glu Lys Val
 980 985 990

ES 2 848 703 T5

Cys Gly Leu Cys Gly Asn Phe Asp Gly Ile Gln Asn Asn Asp Leu Thr
 995 1000 1005

Ser Ser Asn Leu Gln Val Glu Glu Asp Pro Val Asp Phe Gly Asn
 1010 1015 1020

Ser Trp Lys Val Ser Ser Gln Cys Ala Asp Thr Arg Lys Val Pro
 1025 1030 1035

Leu Asp Ser Ser Pro Ala Thr Cys His Asn Asn Ile Met Lys Gln
 1040 1045 1050

Thr Met Val Asp Ser Ser Cys Arg Ile Leu Thr Ser Asp Val Phe
 1055 1060 1065

Gln Asp Cys Asn Lys Leu Val Asp Pro Glu Pro Tyr Leu Asp Val
 1070 1075 1080

Cys Ile Tyr Asp Thr Cys Ser Cys Glu Ser Ile Gly Asp Cys Ala
 1085 1090 1095

Cys Phe Cys Asp Thr Ile Ala Ala Tyr Ala His Val Cys Ala Gln
 1100 1105 1110

His Gly Lys Val Val Thr Trp Arg Thr Ala Thr Leu Cys Pro Gln
 1115 1120 1125

Ser Cys Glu Glu Arg Asn Leu Arg Glu Asn Gly Tyr Glu Cys Glu
 1130 1135 1140

Trp Arg Tyr Asn Ser Cys Ala Pro Ala Cys Gln Val Thr Cys Gln
 1145 1150 1155

His Pro Glu Pro Leu Ala Cys Pro Val Gln Cys Val Glu Gly Cys
 1160 1165 1170

His Ala His Cys Pro Pro Gly Lys Ile Leu Asp Glu Leu Leu Gln
 1175 1180 1185

Thr Cys Val Asp Pro Glu Asp Cys Pro Val Cys Glu Val Ala Gly
 1190 1195 1200

Arg Arg Phe Ala Ser Gly Lys Lys Val Thr Leu Asn Pro Ser Asp
 1205 1210 1215

Pro Glu His Cys Gln Ile Cys His Cys Asp Val Val Asn Leu Thr

ES 2 848 703 T5

Asp Leu Ala Pro Glu Ala Pro Pro Pro Thr Leu Pro Pro Asp Met
 1460 1465 1470

Ala Gln Val Thr Val Gly Pro Gly Leu Leu Gly Val Ser Thr Leu
 1475 1480 1485

Gly Pro Lys Arg Asn Ser Met Val Leu Asp Val Ala Phe Val Leu
 1490 1495 1500

Glu Gly Ser Asp Lys Ile Gly Glu Ala Asp Phe Asn Arg Ser Lys
 1505 1510 1515

Glu Phe Met Glu Glu Val Ile Gln Arg Met Asp Val Gly Gln Asp
 1520 1525 1530

Ser Ile His Val Thr Val Leu Gln Tyr Ser Tyr Met Val Thr Val
 1535 1540 1545

Glu Tyr Pro Phe Ser Glu Ala Gln Ser Lys Gly Asp Ile Leu Gln
 1550 1555 1560

Arg Val Arg Glu Ile Arg Tyr Gln Gly Gly Asn Arg Thr Asn Thr
 1565 1570 1575

Gly Leu Ala Leu Arg Tyr Leu Ser Asp His Ser Phe Leu Val Ser
 1580 1585 1590

Gln Gly Asp Arg Glu Gln Ala Pro Asn Leu Val Tyr Met Val Thr
 1595 1600 1605

Gly Asn Pro Ala Ser Asp Glu Ile Lys Arg Leu Pro Gly Asp Ile
 1610 1615 1620

Gln Val Val Pro Ile Gly Val Gly Pro Asn Ala Asn Val Gln Glu
 1625 1630 1635

Leu Glu Arg Ile Gly Trp Pro Asn Ala Pro Ile Leu Ile Gln Asp
 1640 1645 1650

Phe Glu Thr Leu Pro Arg Glu Ala Pro Asp Leu Val Leu Gln Arg
 1655 1660 1665

Cys Cys Ser Gly Glu Gly Leu Gln Ile Pro Thr Leu Ser Pro Ala
 1670 1675 1680

Pro Asp Cys Ser Gln Pro Leu Asp Val Ile Leu Leu Leu Asp Gly
 1685 1690 1695

ES 2 848 703 T5

Ser Ser Ser Phe Pro Ala Ser Tyr Phe Asp Glu Met Lys Ser Phe
 1700 1705 1710

Ala Lys Ala Phe Ile Ser Lys Ala Asn Ile Gly Pro Arg Leu Thr
 1715 1720 1725

Gln Val Ser Val Leu Gln Tyr Gly Ser Ile Thr Thr Ile Asp Val
 1730 1735 1740

Pro Trp Asn Val Val Pro Glu Lys Ala His Leu Leu Ser Leu Val
 1745 1750 1755

Asp Val Met Gln Arg Glu Gly Gly Pro Ser Gln Ile Gly Asp Ala
 1760 1765 1770

Leu Gly Phe Ala Val Arg Tyr Leu Thr Ser Glu Met His Gly Ala
 1775 1780 1785

Arg Pro Gly Ala Ser Lys Ala Val Val Ile Leu Val Thr Asp Val
 1790 1795 1800

Ser Val Asp Ser Val Asp Ala Ala Ala Asp Ala Ala Arg Ser Asn
 1805 1810 1815

Arg Val Thr Val Phe Pro Ile Gly Ile Gly Asp Arg Tyr Asp Ala
 1820 1825 1830

Ala Gln Leu Arg Ile Leu Ala Gly Pro Ala Gly Asp Ser Asn Val
 1835 1840 1845

Val Lys Leu Gln Arg Ile Glu Asp Leu Pro Thr Met Val Thr Leu
 1850 1855 1860

Gly Asn Ser Phe Leu His Lys Leu Cys Ser Gly Phe Val Arg Ile
 1865 1870 1875

Cys Met Asp Glu Asp Gly Asn Glu Lys Arg Pro Gly Asp Val Trp
 1880 1885 1890

Thr Leu Pro Asp Gln Cys His Thr Val Thr Cys Gln Pro Asp Gly
 1895 1900 1905

Gln Thr Leu Leu Lys Ser His Arg Val Asn Cys Asp Arg Gly Leu
 1910 1915 1920

Arg Pro Ser Cys Pro Asn Ser Gln Ser Pro Val Lys Val Glu Glu
 1925 1930 1935

ES 2 848 703 T5

Thr Cys Gly Cys Arg Trp Thr Cys Pro Cys Val Cys Thr Gly Ser
 1940 1945 1950

Ser Thr Arg His Ile Val Thr Phe Asp Gly Gln Asn Phe Lys Leu
 1955 1960 1965

Thr Gly Ser Cys Ser Tyr Val Leu Phe Gln Asn Lys Glu Gln Asp
 1970 1975 1980

Leu Glu Val Ile Leu His Asn Gly Ala Cys Ser Pro Gly Ala Arg
 1985 1990 1995

Gln Gly Cys Met Lys Ser Ile Glu Val Lys His Ser Ala Leu Ser
 2000 2005 2010

Val Glu Xaa His Ser Asp Met Glu Val Thr Val Asn Gly Arg Leu
 2015 2020 2025

Val Ser Val Pro Tyr Val Gly Gly Asn Met Glu Val Asn Val Tyr
 2030 2035 2040

Gly Ala Ile Met His Glu Val Arg Phe Asn His Leu Gly His Ile
 2045 2050 2055

Phe Thr Phe Thr Pro Gln Asn Asn Glu Phe Gln Leu Gln Leu Ser
 2060 2065 2070

Pro Lys Thr Phe Ala Ser Lys Thr Tyr Gly Leu Cys Gly Ile Cys
 2075 2080 2085

Asp Glu Asn Gly Ala Asn Asp Phe Met Leu Arg Asp Gly Thr Val
 2090 2095 2100

Thr Thr Asp Trp Lys Thr Leu Val Gln Glu Trp Thr Val Gln Arg
 2105 2110 2115

Pro Gly Gln Thr Cys Gln Pro Ile Leu Glu Glu Gln Cys Leu Val
 2120 2125 2130

Pro Asp Ser Ser His Cys Gln Val Leu Leu Leu Pro Leu Phe Ala
 2135 2140 2145

Glu Cys His Lys Val Leu Ala Pro Ala Thr Phe Tyr Ala Ile Cys
 2150 2155 2160

Gln Gln Asp Ser Cys His Gln Glu Gln Val Cys Glu Val Ile Ala

ES 2 848 703 T5

2165		2170		2175
Ser Tyr	Ala His Leu Cys	Arg Thr Asn Gly Val	Cys Val Asp Trp	
2180		2185	2190	
Arg Thr	Pro Asp Phe Cys	Ala Met Ser Cys Pro	Pro Ser Leu Val	
2195		2200	2205	
Tyr Asn	His Cys Glu His	Gly Cys Pro Arg His	Cys Asp Gly Asn	
2210		2215	2220	
Val Ser	Ser Cys Gly Asp	His Pro Ser Glu Gly	Cys Phe Cys Pro	
2225		2230	2235	
Pro Asp	Lys Val Met Leu	Glu Gly Ser Cys Val	Pro Glu Glu Ala	
2240		2245	2250	
Cys Thr	Gln Cys Ile Gly	Glu Asp Gly Val Gln	His Gln Phe Leu	
2255		2260	2265	
Glu Ala	Trp Val Pro Asp	His Gln Pro Cys Gln	Ile Cys Thr Cys	
2270		2275	2280	
Leu Ser	Gly Arg Lys Val	Asn Cys Thr Thr Gln	Pro Cys Pro Thr	
2285		2290	2295	
Ala Lys	Ala Pro Thr Cys	Gly Leu Cys Glu Val	Ala Arg Leu Arg	
2300		2305	2310	
Gln Asn	Ala Asp Gln Cys	Cys Pro Glu Tyr Glu	Cys Val Cys Asp	
2315		2320	2325	
Pro Val	Ser Cys Asp Leu	Pro Pro Val Pro His	Cys Glu Arg Gly	
2330		2335	2340	
Leu Gln	Pro Thr Leu Thr	Asn Pro Gly Glu Cys	Arg Pro Asn Phe	
2345		2350	2355	
Thr Cys	Ala Cys Arg Lys	Glu Glu Cys Lys Arg	Val Ser Pro Pro	
2360		2365	2370	
Ser Cys	Pro Pro His Arg	Leu Pro Thr Leu Arg	Lys Thr Gln Cys	
2375		2380	2385	
Cys Asp	Glu Tyr Glu Cys	Ala Cys Asn Cys Val	Asn Ser Thr Val	
2390		2395	2400	

ES 2 848 703 T5

Ser Cys Pro Leu Gly Tyr Leu Ala Ser Thr Ala Thr Asn Asp Cys
 2405 2410 2415

Gly Cys Thr Thr Thr Thr Cys Leu Pro Asp Lys Val Cys Val His
 2420 2425 2430

Arg Ser Thr Ile Tyr Pro Val Gly Gln Phe Trp Glu Glu Gly Cys
 2435 2440 2445

Asp Val Cys Thr Cys Thr Asp Met Glu Asp Ala Val Met Gly Leu
 2450 2455 2460

Arg Val Ala Gln Cys Ser Gln Lys Pro Cys Glu Asp Ser Cys Arg
 2465 2470 2475

Ser Gly Phe Thr Tyr Val Leu His Glu Gly Glu Cys Cys Gly Arg
 2480 2485 2490

Cys Leu Pro Ser Ala Cys Glu Val Val Thr Gly Ser Pro Arg Gly
 2495 2500 2505

Asp Ser Gln Ser Ser Trp Lys Ser Val Gly Ser Gln Trp Ala Ser
 2510 2515 2520

Pro Glu Asn Pro Cys Leu Ile Asn Glu Cys Val Arg Val Lys Glu
 2525 2530 2535

Glu Val Phe Ile Gln Gln Arg Asn Val Ser Cys Pro Gln Leu Glu
 2540 2545 2550

Val Pro Val Cys Pro Ser Gly Phe Gln Leu Ser Cys Lys Thr Ser
 2555 2560 2565

Ala Cys Cys Pro Ser Cys Arg Cys Glu Arg Met Glu Ala Cys Met
 2570 2575 2580

Leu Asn Gly Thr Val Ile Gly Pro Gly Lys Thr Val Met Ile Asp
 2585 2590 2595

Val Cys Thr Thr Cys Arg Cys Met Val Gln Val Gly Val Ile Ser
 2600 2605 2610

Gly Phe Lys Leu Glu Cys Arg Lys Thr Thr Cys Asn Pro Cys Pro
 2615 2620 2625

Leu Gly Tyr Lys Glu Glu Asn Asn Thr Gly Glu Cys Cys Gly Arg
 2630 2635 2640

ES 2 848 703 T5

Cys Leu Pro Thr Ala Cys Thr Ile Gln Leu Arg Gly Gly Gln Ile
 2645 2650 2655

Met Thr Leu Lys Arg Asp Glu Thr Leu Gln Asp Gly Cys Asp Thr
 2660 2665 2670

His Phe Cys Lys Val Asn Glu Arg Gly Glu Tyr Phe Trp Glu Lys
 2675 2680 2685

Arg Val Thr Gly Cys Pro Pro Phe Asp Glu His Lys Cys Leu Ala
 2690 2695 2700

Glu Gly Gly Lys Ile Met Lys Ile Pro Gly Thr Cys Cys Asp Thr
 2705 2710 2715

Cys Glu Glu Pro Glu Cys Asn Asp Ile Thr Ala Arg Leu Gln Tyr
 2720 2725 2730

Val Lys Val Gly Ser Cys Lys Ser Glu Val Glu Val Asp Ile His
 2735 2740 2745

Tyr Cys Gln Gly Lys Cys Ala Ser Lys Ala Met Tyr Ser Ile Asp
 2750 2755 2760

Ile Asn Asp Val Gln Asp Gln Cys Ser Cys Cys Ser Pro Thr Arg
 2765 2770 2775

Thr Glu Pro Met Gln Val Ala Leu His Cys Thr Asn Gly Ser Val
 2780 2785 2790

Val Tyr His Glu Val Leu Asn Ala Met Glu Cys Lys Cys Ser Pro
 2795 2800 2805

Arg Lys Cys Ser Lys
 2810

<210> 29

<211> 4

5 <212> PRT

<213> Secuencia artificial

<220>

<223> IgG1 humana

10

<400> 29

Glu Leu Leu Gly
 1

ES 2 848 703 T5

<210> 30

<211> 42

<212> PRT

<213> Secuencia artificial

5

<220>

<223> XTEN AE42-4

<400> 30

Gly Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
1 5 10 15

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
20 25 30

Thr Ser Gly Ser Glu Thr Pro Ala Ser Ser
35 40

10

<210> 31

<211> 126

<212> ADN

<213> Secuencia artificial

15

<220>

<223> XTEN AE42-4

20

<400> 31

ggcgcgccag gttctcctgc tggctccccc acctcaacag aagaggggac aagcgaaagc 60
gctacgcctg agagtggccc tggctctgag ccagccacct cggctctga aaccctgcc 120
tcgagc 126

<210> 32

<211> 144

<212> PRT

<213> Secuencia artificial

25

<220>

<223> XTEN AE144-2A

30

ES 2 848 703 T5

<400> 32

```

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
1           5           10           15

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
                20           25           30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
                35           40           45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50           55           60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65           70           75           80

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
                85           90           95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
                100           105           110

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
                115           120           125

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
130           135           140

```

5 <210> 33

<211> 450

<212> ADN

<213> Secuencia artificial

10 <220>

<223> XTEN AE144-2A

<400> 33

```

ggcgcgccaa ccagtacgga gccgtccgag gggagcgcac caggaagccc ggctgggagc      60
ccgacttcta ccgaagaggg tacatctacc gaaccaagtg aaggttcagc accagggacc      120
tcaacagaac cctctgaggg ctcggcgcct ggtacaagtg agtccgccac cccagaatcc      180
gggcctggga caagcacaga accttcggaa gggagtgcc ctggaacatc cgaatcggca      240
accccagaat cagggccagg atctgagccc gcgacttcg gctccgagac gcctgggaca      300
tccaccgagc cctccgaagg atcagcccca ggcaccagca cggagccctc tgaggaagc      360
gcacctggta ccagcgaag cgcaactccc gaatcaggtc ccggtacgag cgagtcggcg      420
accccggaga gcgggcccagg tgcctcgagc

```

ES 2 848 703 T5

<210> 34

<211> 144

<212> PRT

<213> Secuencia artificial

5

<220>

<223> XTEN AE144-3B

<400> 34

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

10

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
20 25 30

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
65 70 75 80

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
100 105 110

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
115 120 125

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 35

<211> 450

15

<212> ADN

<213> Secuencia artificial

<220>

<223> XTEN AE144-3B

20

<400> 35

ES 2 848 703 T5

ggcgcgccaa gtcccgtgg aagcccaact agcaccgaag aggggacctc agagtccgcc 60
 acccccgagt ccggccctgg ctctgagcct gccactagcg gctccgagac tcctggcaca 120
 tccgaaagcg ctacaccga gagtggacct ggcacctta ccgagcccag tgagggctcc 180
 gccctggaa caagcaccga gccagcgaa ggcagcgccc cagggacctc cacagagccc 240
 agtgaaggca gtgctcctgg caccagcacc gaaccaagcg agggctctgc acccgggacc 300
 tccaccgagc caagcgaagg ctctgccctt ggcacttcca ccgagcccag cgaaggcagc 360
 gccctggga gccccgctgg ctctcccacc agcactgagc agggcacatc taccgaacca 420
 agtgaaggct ctgcaccagg tgcctcgagc 450

<210> 36

<211> 144

5

<212> PRT

<213> Secuencia artificial

<220>

<223> XTEN AE144-4A

10

<400> 36

Thr	Ser	Glu	Ser	Ala	Thr	Pro	Glu	Ser	Gly	Pro	Gly	Ser	Glu	Pro	Ala	1	5	10	15
Thr	Ser	Gly	Ser	Glu	Thr	Pro	Gly	Thr	Ser	Glu	Ser	Ala	Thr	Pro	Glu	20	25	30	
Ser	Gly	Pro	Gly	Ser	Glu	Pro	Ala	Thr	Ser	Gly	Ser	Glu	Thr	Pro	Gly	35	40	45	
Thr	Ser	Glu	Ser	Ala	Thr	Pro	Glu	Ser	Gly	Pro	Gly	Thr	Ser	Thr	Glu	50	55	60	
Pro	Ser	Glu	Gly	Ser	Ala	Pro	Gly	Thr	Ser	Glu	Ser	Ala	Thr	Pro	Glu	65	70	75	80
Ser	Gly	Pro	Gly	Ser	Pro	Ala	Gly	Ser	Pro	Thr	Ser	Thr	Glu	Glu	Gly	85	90	95	
Ser	Pro	Ala	Gly	Ser	Pro	Thr	Ser	Thr	Glu	Glu	Gly	Ser	Pro	Ala	Gly	100	105	110	
Ser	Pro	Thr	Ser	Thr	Glu	Glu	Gly	Thr	Ser	Glu	Ser	Ala	Thr	Pro	Glu	115	120	125	
Ser	Gly	Pro	Gly	Thr	Ser	Thr	Glu	Pro	Ser	Glu	Gly	Ser	Ala	Pro	Gly	130	135	140	

ES 2 848 703 T5

<210> 37

<211> 450

<212> ADN

<213> Secuencia artificial

5

<220>

<223> XTEN AE144-4A

<400> 37

```
ggcgcgccaa cgtccgaaag tgctaccctt gagtcaggcc ctggtagtga gcctgccaca      60
agcgggaagcg aaactccggg gacctcagag tctgccactc ccgaatcggg gccaggctct      120
gaaccggcca cttcaggagc cgaaacacca ggaacatcgg agagcgctac cccggagagc      180
gggccaggaa ctagtactga gcctagcgag ggaagtgcac ctggtacaag cgagtcggcc      240
acacccgagt ctggccctgg ctctccagcg ggctcaccca cgagcactga agagggctct      300
cccgtggca  gcccaacgtc gacagaagaa ggatcaccag caggctcccc cacatcaaca      360
gaggagggta catcagaatc tgctactccc gagagtggac ccggtacctc cactgagccc      420
agcgagggga gtgcaccagg tgcctcgagc                                     450
```

10

<210> 38

<211> 144

<212> PRT

<213> Secuencia artificial

15

<220>

<223> XTEN AE144-5A

20

<400> 38

ES 2 848 703 T5

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
 1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
 20 25 30

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
 35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
 50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
 65 70 75 80

Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
 85 90 95

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
 100 105 110

Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
 115 120 125

Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
 130 135 140

<210> 39

<211> 450

5

<212> ADN

<213> Secuencia artificial

<220>

<223> XTEN AE144-5A

10

<400> 39

ggcgcgcaa catcagagag cgccaccct gaaagtggc cgggagcga gccagccaca 60

tctgggtcgg aaacgccagg cacaagtgag tctgcaactc cggagtccgg acctggctcc 120

gagcctgcca ctagcggctc cgagactccg ggaacttccg agagcgctac accagaaagc 180

ggaccgggaa ccagtaccga acctagcgag ggctctgctc cgggcagccc agccggctct 240

cctacatcca cggaggagg cacttccgaa tccgccacc cggagtcagg gccaggatct 300

gaaccgccta cctcaggcag tgagagcga ggaacgagcg agtccgctac accggagagt 360

ggccaggga gccctgctgg atctcctacg tccactgagg aagggtcacc agcgggctcg 420

cccaccagca ctgaagaagg tgcctcgagc 450

ES 2 848 703 T5

<210> 40

<211> 144

<212> PRT

<213> Secuencia artificial

5

<220>

<223> XTEN AE144-6B

<400> 40

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
1 5 10 15
Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30
Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45
Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
50 55 60
Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
65 70 75 80
Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95
Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro Ala
100 105 110
Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125
Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

10

<210> 41

<211> 450

<212> ADN

<213> Secuencia artificial

15

<220>

<223> XTEN AE144-6B

20

<400> 41

ES 2 848 703 T5

```

ggcgcgcca catctaccga gccttccgaa ggctctgccc ctgggacctc agaatctgca      60
accctgaaa gcgccctgg aacctccgaa agtgccactc ccgagagcgg cccagggaca      120
agcgagtcag caaccctga gtctggacctc ggcagcgagc ctgcaacctc tggetcagag      180
actcccggtc cagaaccgc tacctcaggc tccgagacac ccggtctctc tgctgggagt      240
cccacttcca ccgaggaagg aacatccact gagcctagtg agggctctgc ccctggaacc      300
agcacagagc caagtgaggg cagtgcacca ggatccgagc cagcaaccag cgggtccgag      360
actcccgga cctctgagtc tgccacccca gagagcggac ccggcacttc aaccgagccc      420
tccgaaggat cagcaccagg tgcctcgagc                                     450

```

<210> 42

<211> 144

5

<212> PRT

<213> Secuencia artificial

<220>

<223> XTEN AG144-1

10

<400> 42

```

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser
1          5          10          15

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
          20          25          30

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
          35          40          45

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser
50          55          60

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
65          70          75          80

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
          85          90          95

Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser
100         105         110

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
115         120         125

Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser
          130         135         140

```


ES 2 848 703 T5

<210> 43

<211> 450

<212> ADN

<213> Secuencia artificial

5

<220>

<223> XTEN AG144-1

<400> 43

```
ggcgcgccac cggggtcgtc cccgtcggcg tccaccggaa cagggccagg gtcacccccg      60
tcagcgtcga ctgggacggg acccgggaca cccggttcgg ggactgcac ctcctcgcct      120
ggttcgtcca ccccgtcagg agccacgggt tcgccgggaa gcagcccaag cgcacccact      180
ggtacagggc ctggggcttc accgggtact tcatccacgg ggtcaccggg aacgcccgga      240
tcggggacgg cttcctcatc accaggatcg tcaacaccct cgggcgcaac gggcagcccc      300
ggaaccctg gttcgggtac ggcgtcgtcg agccccggtg cgagcccggg aacaagctcg      360
acaggatcgc ctggggcgtc acccggcacg tcgagcacag gcagcccggg aaccctgga      420
tcgggaaccg cgtcgtcaag cgcctcgagc      450
```

10

<210> 44

<211> 144

<212> PRT

<213> Secuencia artificial

15

<220>

<223> XTEN AG144-A

20

<400> 44

ES 2 848 703 T5

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
 1 5 10 15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
 20 25 30

Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
 35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
 50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
 65 70 75 80

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
 85 90 95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
 100 105 110

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
 115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 130 135 140

<210> 45

5 <211> 450

<212> ADN

<213> Secuencia artificial

<220>

10 <223> XTEN AG144-A

<400> 45

ggcgcgccag gtgcctcgcc gggaacatca tcaactgggt caccggggtc atcccctcg 60

gcctcaaccg ggacgggtcc cggctcatcc cccagcgcca gcaactggaac aggtcctggc 120

actcctgggt cgggtacggc atcgatcatcc ccgggaagct caacaccgtc cggagcgaca 180

ggatcacctg gctcgtcacc ttcggcgta actggaacgg ggccaggggc ctcaaccgga 240

acgtcctcga ctgggtcgcc tggtaagccg ggatcaggaa cggcctcatc ctgcctggg 300

tcctcaacgc cctcgggtgc gactggttcg ccgggaactc ctggctcggg gacggcctcg 360

tcgtcgcctg gggcatcacc ggggacgagc tccacgggt cccctggagc gtcaccgggg 420

acctcctcga caggtagccc ggcctcgagc 450

ES 2 848 703 T5

<210> 46

<211> 144

<212> PRT

<213> Secuencia artificial

5

<220>

<223> XTEN AG144-B

<400> 46

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
1 5 10 15

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
20 25 30

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
65 70 75 80

Gly Thr Gly Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
85 90 95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro
100 105 110

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130 135 140

10

<210> 47

<211> 450

15

<212> ADN

<213> Secuencia artificial

<220>

<223> XTEN AG144-B

20

<400> 47

ES 2 848 703 T5

```

ggcgcgccag gtacaccggg cagcggcagc gcttcgtcgt caccgggctc gtccacaccg      60
tcgggagcta cgggaagccc aggagcgtca ccgggaacgt cgtcaacggg gtcaccgggt      120
acgccaggta gcggcacggc cagcagctcg ccaggttcat cgaccccgtc gggagcgact      180
gggtcgcccg gatcaagccc gtcagcttcc actggaacag gacccgggctc gtcgcccgtca      240
gcctcaacgg ggacaggacc tggttcatcg acgccgtcag gggcgacagg ctcgcccgga      300
tcgtcaacac cctcgggggc aacggggagc cctggtgctg cgcctggaac ctcatccacc      360
ggaagcccgg gggcctcgcc gggtaagagc tccacgggat cgcccggagc gtcccccgga      420
acttcaagca cagggagccc tgcctcgagc                                     450

```

<210> 48

<211> 144

5

<212> PRT

<213> Secuencia artificial

<220>

<223> XTEN AG144-C

10

<400> 48

```

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro
1          5          10          15

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
20          25          30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
35          40          45

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
50          55          60

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
65          70          75          80

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
85          90          95

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
100         105         110

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
115         120         125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
130         135         140

```

ES 2 848 703 T5

<210> 49

<211> 450

<212> ADN

<213> Secuencia artificial

5

<220>

<223> XTEN AG144-C

<400> 49

```
ggcgcgccag gtacaccggt atcgggtaca gcgtcatcga gccccggtgc gtcacctggt      60
acgtcgagca cggggtcgcc aggggcgtcc cctgggacgt cctcaacagg ctcgcccggg      120
gcgtcaccgg gcacgtcgtc cacgggttca cctggtagct ccccttcggc gtccactggc      180
accgggcctg gaactccggg gagcggcaca gcgagctcgt cgccgggagc atcgctctgg      240
acatcgagca cggggtcgcc aggagcatcg cccggaacat ccagcacagg aagccccggc      300
gcgtcgcccg ggacatcaag cacaggttcc ccgggatcga gcacgccgtc cggagccact      360
ggatcaccag ggagctcgac accttcgggc gcaacgggat cgcccggagc cagcccgggt      420
acgtcaagca ctggctcccc tgcctcgagc                                     450
```

10

<210> 50

<211> 144

<212> PRT

<213> Secuencia artificial

15

<220>

<223> XTEN AG144-F

20

<400> 50

ES 2 848 703 T5

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
 1 5 10 15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
 20 25 30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 35 40 45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
 50 55 60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
 65 70 75 80

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro
 85 90 95

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
 100 105 110

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
 115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 130 135 140

<210> 51

<211> 450

5

<212> ADN

<213> Secuencia artificial

<220>

<223> XTEN AG144-F

10

<400> 51

ggcgcgccag gctccagccc ctccgcgagc acgggaaccg gaccaggttc gtcaccctca 60

gcatcaacgg ggacgggacc gggggcgtca ccaggaacgt cctccaccgg ctcgccgggt 120

gcatcaccgg gaacgtcatc gaccggatcg ccagggagct cgacgccatc aggcgcaaca 180

ggatcaacctg gctcaagccc tagcgcgtca accggcacgg gtccgggtgc ctcccctggc 240

acgtccagca ccggatcacc cggatcgagc ccatccgcct caaccggaac cggaccgggt 300

acaccagggt cgggaacagc ctctctgtca ccaggctcct caaccctc gggagccagc 360

ggttogcccg gttcgtcaac gccttcggga gcaactgta gccccggagc atcgccagga 420

acttcgagca cggggtcgcc cgcctcgagc 450

REIVINDICACIONES

1. Una molécula de ácido nucleico que comprende una secuencia de nucleótidos al menos 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % idéntica a SEQ ID NO: 1, en donde la secuencia de nucleótidos codifica un polipéptido con actividad de factor VIII.
- 5 2. La molécula de ácido nucleico de la reivindicación 1, en donde la secuencia de nucleótidos codifica FVIII de dominio B deleciónado.
3. La molécula de ácido nucleico de las reivindicaciones 1 o 2, en donde la secuencia de nucleótidos tiene una o más de las siguientes características:
 - 10 (a) la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con SEQ ID NO: 3;
 - (b) la secuencia de nucleótidos contiene menos secuencias MARS/ARS en comparación con SEQ ID NO: 3;
 - (c) la secuencia de nucleótidos no contiene el sitio de corte y empalme GGTGAT (SEQ ID NO: 7);
 - (d) la secuencia de nucleótidos contiene menos elementos desestabilizantes (SEQ ID NO: 8 y SEQ ID NO: 9) con respecto a SEQ ID NO: 3;
 - 15 (e) la secuencia de nucleótidos no contiene una secuencia de poli-T (SEQ ID NO: 10);
 - (f) la secuencia de nucleótidos no contiene una secuencia de poli-A (SEQ ID NO: 11); o
 - (g) el índice de adaptación de codones humanos es elevado con respecto a SEQ ID NO: 3.
4. La molécula de ácido nucleico de la reivindicación 3, en donde
 - 20 (a) el índice de adaptación de codones humanos es al menos aproximadamente 0,75, al menos aproximadamente 0,76, al menos aproximadamente 0,77, al menos aproximadamente 0,78, al menos aproximadamente 0,79, al menos aproximadamente 0,80, al menos aproximadamente 0,81, al menos aproximadamente 0,82, al menos aproximadamente 0,83, al menos aproximadamente 0,84, al menos aproximadamente 0,85, al menos aproximadamente 0,86, al menos aproximadamente 0,87, o al menos aproximadamente 0,88;
 - 25 (b) el porcentaje de nucleótidos G/C es al menos aproximadamente 45 %, al menos aproximadamente 46 %, al menos aproximadamente 47 %, al menos aproximadamente 48 %, al menos aproximadamente 49 % o al menos aproximadamente 50 %;
 - (c) la secuencia de nucleótidos contiene como máximo una secuencia MARS/ARS; o
 - (d) la secuencia de nucleótidos contiene como máximo 4 elementos desestabilizantes.
- 30 5. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1 a 4, que comprende además una secuencia heteróloga de nucleótidos.
6. La molécula de ácido nucleico de la reivindicación 5, en donde la secuencia heteróloga de nucleótidos codifica una secuencia heteróloga de aminoácidos que es un extensor de la semivida.
- 35 7. La molécula de ácido nucleico de la reivindicación 6, en donde la secuencia heteróloga de aminoácidos es una región constante de inmunoglobulina o una porción de la misma, transferrina, albúmina o una secuencia PAS.
8. La molécula de ácido nucleico de la reivindicación 7, en donde la secuencia heteróloga de aminoácidos es una región Fc.
9. La molécula de ácido nucleico de la reivindicación 1, en donde la secuencia de nucleótidos comprende SEQ ID NO: 1.
- 40 10. Un vector que comprende la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1 a 9.
11. El vector de la reivindicación 10, en donde el vector se selecciona de un vector adenoviral, un vector lentiviral, un vector baculoviral, un vector viral de Epstein Barr, un vector papovaviral, un vector de virus variolovacunal, un vector del virus del herpes simple y un vector de virus adenoasociado (AAV).
- 45 12. Una célula hospedadora que comprende la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1 a 9 o el vector de la reivindicación 10 u 11.

13. Un método de producción de un polipéptido con actividad de factor VIII, que comprende: cultivar la célula hospedadora de la reivindicación 12 en condiciones por las cuales se produce un polipéptido con actividad de factor VIII; y, recuperar el polipéptido con actividad de factor VIII.

5 14. La molécula de ácido nucleico de una cualquiera de las reivindicaciones 1 a 9, o el vector de la reivindicación 10 u 11, para su uso en el tratamiento de un trastorno hemorrágico en un sujeto en necesidad del mismo.

Figura 1

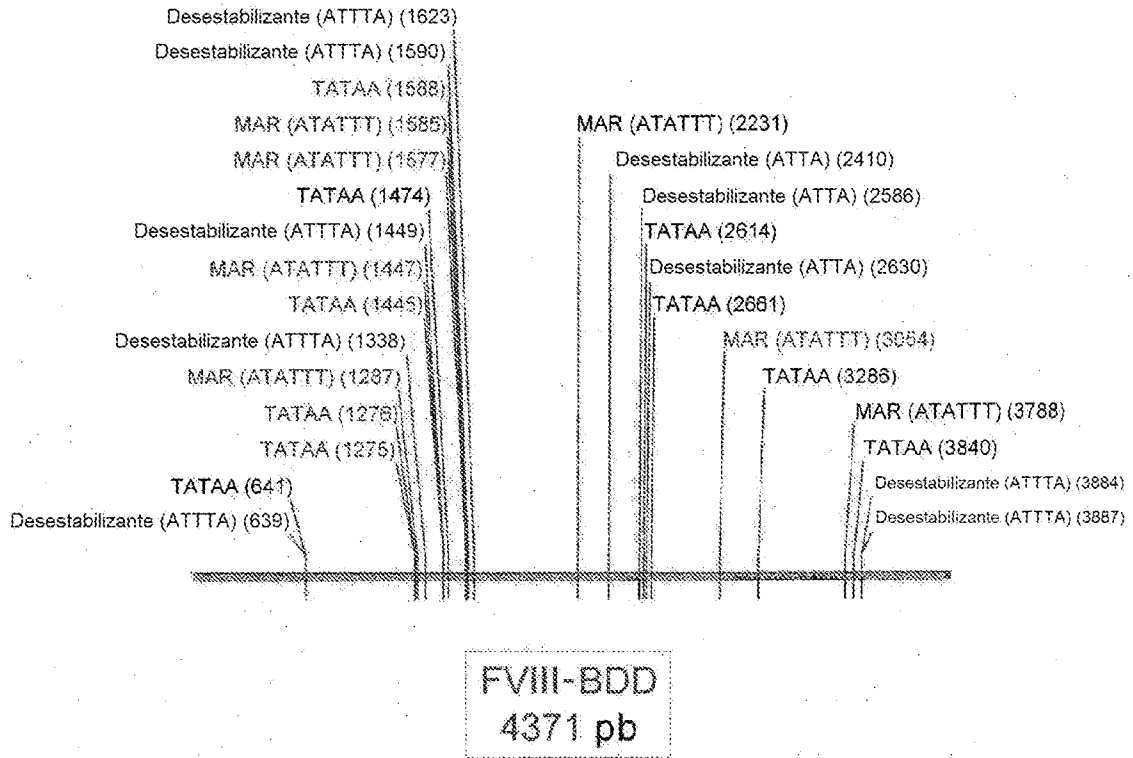


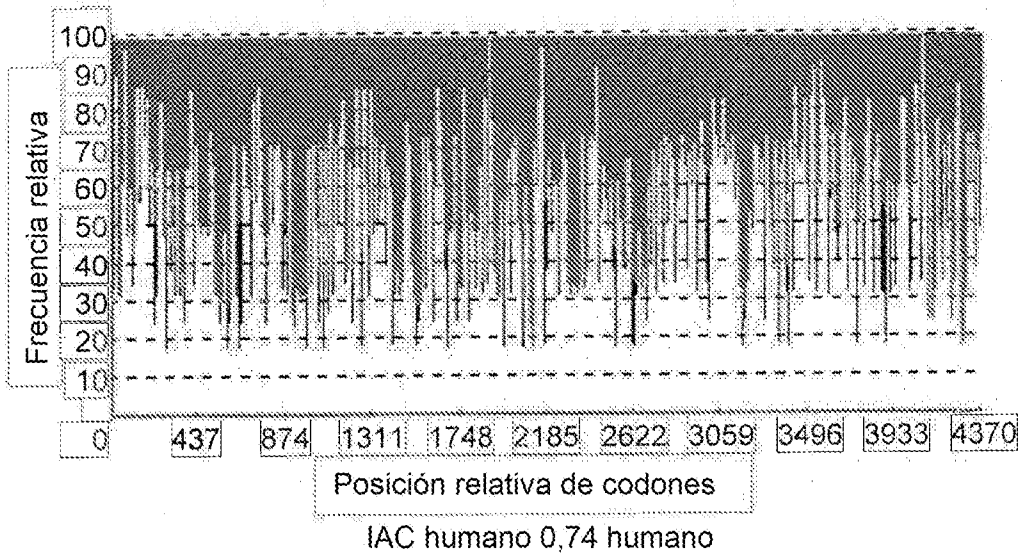
Figura 2

CGTACGGCCGCCACCATTGCAGATTGAGCTGTCTACTTGGCTTTTTCCTGTGCTGCTGAGGTTTTGCTTTTCCGCTACACCG
AAGGTATATCTGGGGCTGTGGAACCTGTCTTGGGATTACATGCAGACTGACCTGGGAGAGCTGCCAGTGGACGCAAGGT
TTCCTCCCTAGAGTCCCTAAGTCATTCCCTTCAACACTIASCCTGTCTACAAGAAAACACTGTTCTGTGGAGTTTACTGAT
CACCTGTTCAACATCSCAAAGCCTAGGCCACCCTGGATGGGACTGCTGGGGCCAAACAATCCAGGCCGAGGTGTACGGCAC
CGTGGTCATTACACTTAAGAACATGGCCTCACACCCCTGAGCCTGCATGCTGTGGGCTCAGCTACTGGAAAGGCTCCG
AAGGAGCAGAGTATGACGATCAGACTTCCAGAGAGAAAAGAGGACGATAAAGGTGTTTTCTGGCGGATCTCATACCTAC
GTGTGGCAGGTCTGTGAAAGAGAAATGGCCCTATGGCCTCCAGACCCTCTGTGCTGACCTACTCTTATCTGAGTCACTGGA
CCTGCTCAAGGATCTGAACAGCGGCTGATCGGAGCCCTGCTGTGTGGAGGGAAGGAAGCCTGGCTAAGGAGAAAACCC
AGACACTGCAATAGTTCATTTCTGTGTTTGGCGTGTGTTGACGAAGGGAAATCATGGCACAGCCGAGACAAGAATAGTCTG
ATGCAGGACAGGGATGCCGCTTCAGCCAGAGCTTGGCCCAAAATGCACACTGTGAACGGCTACGTCATTCGCTCACTGCC
TGGGCTTATCGGCTGCCACCAGAGAGCGTGTATTGGCATGTCTATCGGATGGGCACCACACCTGAAGTGCACCTCAATTT
TCCTGGAGGGACATACCTTTCTGGTCCGCAACCACCGACAGGCTTCCCTGGAGATCTCTCCAATTACCTTCCCTGACAGCA
CAGACTCTGTGATGACCTGGGCGAGTCTCTGCTGTTTTGGCCATCAGCTCCCAACGAGCATGATGGCATGGAGGCTTAC
CGTGAAGTGGACTCTTGTCCGAGGAACCTCAGCTGCGGATGAAGAACAATGAGGAAGCAGAAGACTATGACGATGACC
TGACCGACTCCGAGATGGATGTTGGTCCGATTGATGACGATAACAGCCCTCCTTTATCCAGATTAGATCTGTGGCCAAAG
AAACACCTTAAGACATGGGTCCATTACATCGCAGCCGAGGAAGAGGACTGGGATTTATGCACCCTGGTCTGGCACCAGSA
CGATCCCTCCTACAAATCTCAGTATCTGAACAATGGGCCACAGAGGATGGCCAGAAAAGTACAAGAAAGTGGCGTTCATGG
CATATACCAGTACACTTCAAGACTCGCGAAGCCATCCAGCACGAGAGCGGCATCCTGGGACCCTGCTGTACGGAGAA
GTGGGAGACACCCCTGCTGATCATTTCGAAGAACCCAGGCCAGCCGGCTTACAATATCTATCCAGATGGGATTACAGATGT
GGCCCTCTGTACAGCAGGAGACTGCCAAAGGGCGTCAAACACCTGAAGGACTTCCCAATCTGCCCGGAGAAAATCTTCA
AGTACAAAGTGGACTGTACCCTCGAGGATGGCCCACTAAGAGCGACCCCTCGGTGCTGACCCGCTACTATTCTAGTTC
GTGAATATGAAAGAGATCTGGCAAGCGGACTGATCGGACCCTGCTGATTTGTTACAAGAGAGCGTGGATCAGAGAGG
CAACAGATCATGTCCGACAGCGGAATGTGATTTCTGTTCACTGCTTTGACGAAAACAGGTCATGCTACCTGACCGAGA
ACATCCAGAGATTCTGCTTAATCCAGCTGGGGTGCAGCTGGAAGATCCTGAGTTTCAAGCATCTAACCATCAGATAGT
ATTAATGGCTACGTTCTGACAGTTTGCAGCTGAGCGTGTGCTGACGAGGTCGCTTACTGGTATATCCTGAGCATTTGG
GGCACAGACAGATTTCTGAGCGTGTCTTTTCCGGCTACACTTTTAAGCATAAAATGCTCTAAGAGGACACAGTCACTC
TGTTCCTTFCAGCGGCGAAACCGTGTATTGAGCATGGGAAATCCCGGACTGTGGATTTCTGGGGTCCCAACACACGAT
TTGAGAAATCCCGAATGACTGCCCTGCTGAAAGTGTCAAGCTGTGACAAAGAAACCGGGACTACTATGAAGATTCATA
CCAGACATGAGCCATATCTGTCTCCAAAAGAAATCCCATGAAACCCCGCTTTTAAAGTCAAGTCTCCCATGCTCA
AGGGCACAGCGGGAGATCAACCCCTACCTGCTGAGACTGATCAGGAAAGAGATCCAGTACAGCAATACAAATTTCTGTG
GAAATGAGAAAGGGACTTCCATATCTATGACGAGATGAGAACCCAGAGTCCCTGATCATTCCAGAAAGAAACAGGCT
TACTTTATGCGGCACTGGAGCGGCTGTGGCATATGACATCTGCTTAGTCTGACCTGCTGCGAAATAGGGCCCGAGT
CAGGAAGCGTCCCAAGTTCAAGAAAGTGGTCTTCCAGGATTTACAGACGGGTCCCTTACTCAGCCACTGTACAGGGGC
GAACCTGAACGAGCACTGGGACTGCTGGGCCCTATATCAGAGCAGAAGTGGAGGATAACATTTATGTCACCTTCAGAAA
TCAGGCTCTCGGCTTACAGTTTATTCAGGCTGATCTCTTACGAAGAGGACCAGGACAGGGAGCTGAACCAAGAA
AAAACCTCTGAAAGCTAATGAGACCAAAACATACTTTTGAAGGTGCAGCACCATATGGCCCAACAAAAGACGAGTTC
GATTTCAAGGCATGGGCTATTTTCTGACGTGGATCTGGAGAAAGACCTGCACAGTGGGCTGATTTGGCCACTGCTGTT
GTCCATACTAACACCCCTGAATCCAGCCACCGCCGCGAGCTCACTGTCCAGGAGTTGCTCTGTTCTTTACCATCTTTG
ATGAGACAAAGAGCTGGTACTTCAACGAAACATGGAGCGAAATTTGAGGGCTCCATGTACATTCAGATGGAGACCC
ACATTCAGGAGAACTACCCTTTCTATGCTATCAATGGATACATCATGGATACCTGCCCCGGCTGGTTCATGGCACAGSA
CCAGAGAAATCCGGTGGTATCTGCTGAGCATGGGCGAGCAAGGAGAAATCCCACTCAATTCATTTACAGCGGACAGTGTFTA
CTGTACGGAAGAAAGAGAGTACAAGAATGGCCCTGTACAACCTGTATCCGGGCTGTTCCGAAACGCTCGAGATGCTGCTC
AGCAAGGCCGGAATCTGGAGAGTGGAAATGCTGATTTGGAGAGCACCTGCATGCTGGGATGTCTACCTGTTTCTGTTGTA
CAGTAATAAGTGTGAGACACCCCTGGAAATGGCATCCGGGCATATCAGGATTTCCAGATTACCGCATCTGGACAGTACG
GACAGTGGGACCTAAGTGGCTAGACTGCACTATCCGGATCTATCAAGCCTTGTGTCACAAAAGAGCCTTTCTCTTGG
ATTAAGGTGGACCTGCTGGCCCAATGATCATTCATGGCATCAAACCTCAGGAGCTCGGCAGAAATTTCTCTCTCTGTA
CATCTCACAGTTTATCATCATGTACAGCCTGGATGGGAAGAAATGGCAGACATACCGGGCAATAGCACAGGAACCTGTA
TGGTGTCTTTGGCAACGTTGACAGGAGCGGAATCAAGCACAACTTTCAATCCCTTATCATGCTAGATAGATCCGG
CTGCACCCAAACCAATATTTCTATTCGAAGTACACTGAGGATGGAACTGATGGGATGGATCTGAACAGTTGTTCAATGCC
CCTGGGATGGAGTCCAGGCAATCTCTGACGCCAGATTAACCGCTAGCTCCTACTTCACTAATATGTTTGGTACCTGGA
GCCCTTCAAAGCAAGACTGCACCTGCAAGGCCGAGCAACCGCATGGCCAGCCACAGGTGAACAAATCCCAAGGAGTGGTTC
CAGGTCGATTTTCAAGAAAATATGAAGGTGACCGGGCTCAACAACCTCAGGCGGTGAAAAGTCTGCTGACCTCAATGTACGT
CAAGGAGTCTCTAGTCTCACAGGACGGACATCAGTGGACACTGTTCTTTTCAAGACGGGAAGGTGAAAGTCTTCC
AGGGCAATCAGGATTCCTTTACACCTGTGGTCAACAGTCTAGACCTCCTCACTGCTGACCAGATACCTGAGAAATCCAGCCT
CAGTCTGGGTGCACAGATTGCCCTGGAATGGAAATGCTGGATGCGAGGCCAGGATCTGTACTGATAACTCGAGTC
GACC

SEQ ID NO:1

Figura 4

4A: FVIII BDD antes de la optimización por codones



4B: FVIII BDD después de la optimización por codones con GENESCRIP OPTIMUMGENE™

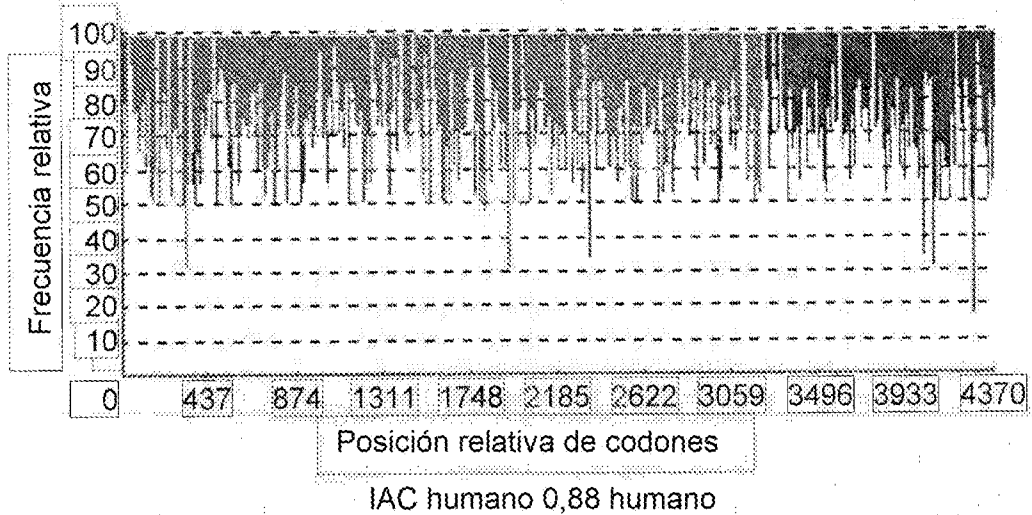
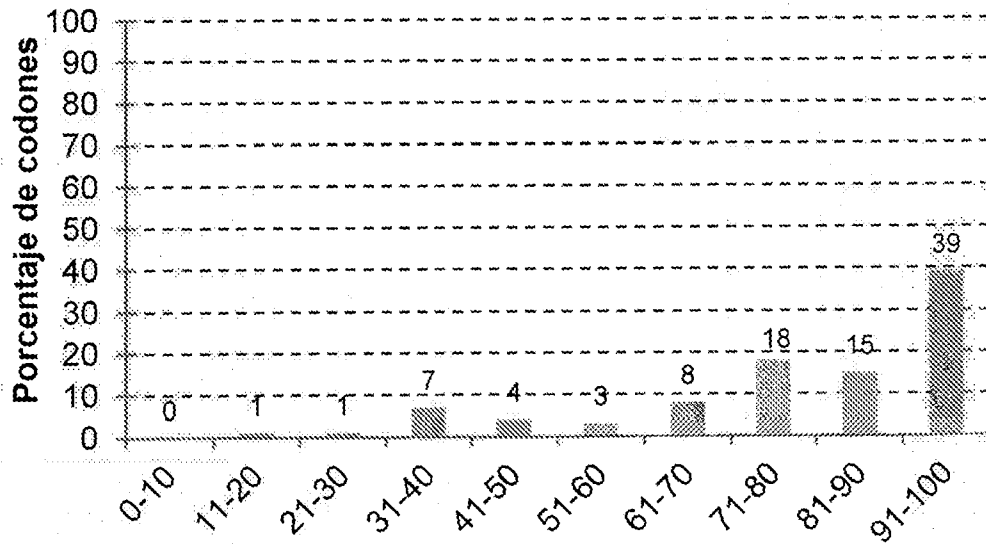


Figura 5

5A: Frecuencia de FVIII BDD de codones óptimos antes de la optimización por codones



5B: Frecuencia de FVIII BDD de codones óptimos después de la optimización por codones con GENSCRIPT OPTIMUMGENE™

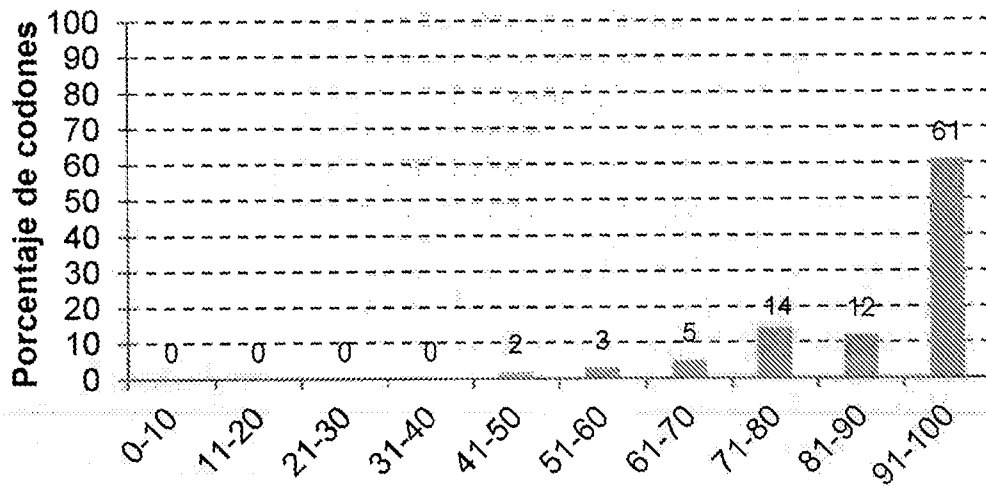
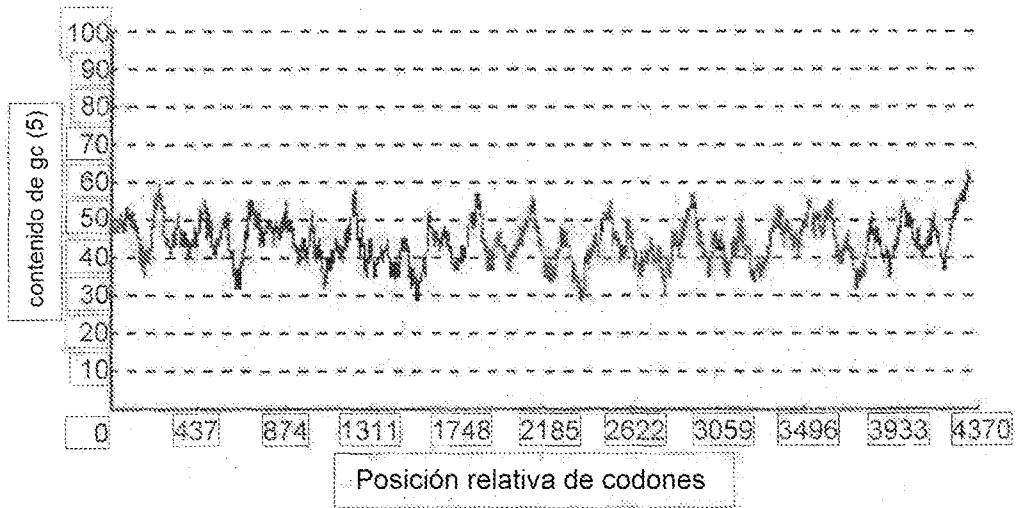


Figura 6

6A: Contenido de G/C de FVIII BDD antes de la optimización por codones



6B: Contenido de G/C de FVIII BDD después de la optimización por codones con GENSCRIPT OPTIMUMGENE™

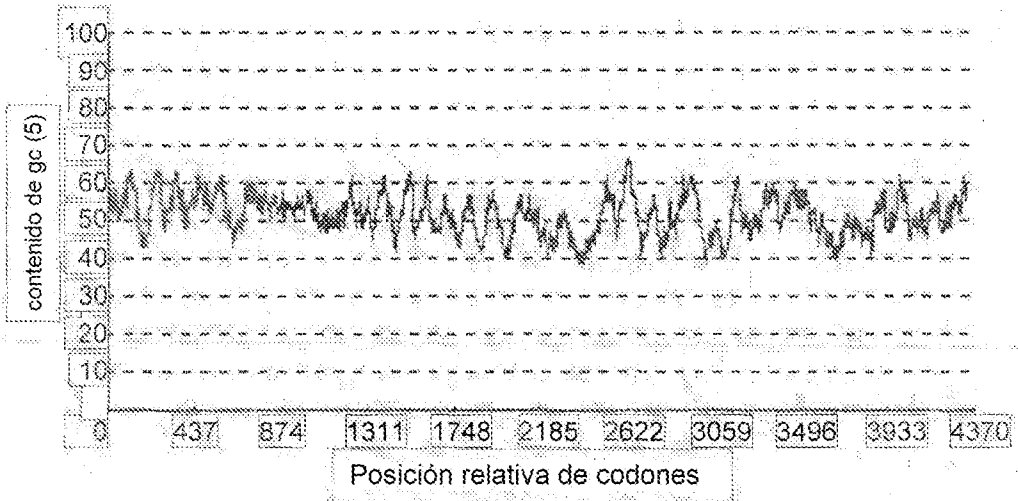


Figura 7

