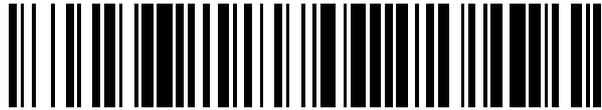


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 437 815**

51 Int. Cl.:

**C12N 15/67** (2006.01)

**C07H 21/00** (2006.01)

**C07K 14/745** (2006.01)

**C12N 15/63** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.10.2001 E 01984006 (5)**

97 Fecha y número de publicación de la concesión europea: **17.07.2013 EP 1325138**

54 Título: **ARN mensajero optimizado**

30 Prioridad:

**11.10.2000 US 686497**

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

**14.01.2014**

73 Titular/es:

**SHIRE HUMAN GENETIC THERAPIES, INC.  
(100.0%)  
300 Shire Way  
Lexington MA 02421, US**

72 Inventor/es:

**SELDON, RICHARD F.;  
MILLER, ALLAN M. y  
TRECO, DOUGLAS S.**

74 Agente/Representante:

**RIZZO, Sergio**

**ES 2 437 815 T3**

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

## DESCRIPCIÓN

ARN mensajero optimizado

**Campo de la invención**

5 [0001] La invención se refiere a métodos para optimizar las propiedades de moléculas de ARNm, moléculas de ARNm optimizado, métodos de utilización de moléculas de ARNm optimizado y composiciones que incluyen moléculas de ARNm optimizado.

**Antecedentes de la invención**

10 [0002] En eucaroytes, la expresión génica se ve afectada, en parte, por la estabilidad y estructura de la molécula de ARN mensajero (ARNm). La estabilidad del ARNm influye en la expresión génica al afectar el nivel en estado estacionario del ARNm. Puede afectar las tasas a las que desaparece el ARNm tras la represión de la transcripción y se acumula después de la inducción de la transcripción. La estructura y la secuencia de nucleótidos de la molécula de ARNm también pueden influir en la eficacia con la que se traducen estas moléculas individuales de ARNm.

15 [0003] La estabilidad intrínseca de una molécula de ARNm determinada está influida por una serie de elementos específicos de la secuencia interna que pueden ejercer un efecto de desestabilización en el ARNm, Estos elementos pueden estar situados en cualquier región de la transcripción y, por ejemplo, se pueden encontrar en la región 5' no traducida (5'UTR), en la región de codificación y en la región 3' no traducida (3' UTR). Está probado que el acortamiento del extremo de poli(a) inicia la descomposición del ARNm (Ross, *Trends in Genetics*, 12:717-175, 1996). La cola de poli(A) influye en la estabilidad del ARNm citoplasmático protegiendo el ARNm de la degradación rápida. Los elementos ricos en adenosina y uridina (AURE) en la 3'UTR también se asocian con ARNm de mamífero inestable. Se ha demostrado que las proteínas que se unen a los AURE, proteínas de unión a AURE (AUPB) pueden afectar a la estabilidad del ARNm. La región de codificación también puede alterar la vida media de muchos ARN. Por ejemplo, la región de codificación puede interactuar con las proteínas que la protegen del ataque endonucleolítico. Además, la eficacia con la que las moléculas de ARNm individuales se traducen tiene una fuerte influencia en la estabilidad de la molécula de ARNm (Herrick *et al.*, *Mol Cell Biol.* 10, 2269-2284, 1990, y Hoekema *et al.*, *Mol Cell Biol.* 7, 2914-2924, 25 1987).

[0004] La naturaleza de cadena sencilla del ARNm le permite adoptar la estructura secundaria y terciaria de un modo dependiente de la secuencia a través del apareamiento de bases complementario. Entre los ejemplos de estas estructuras se incluyen: horquillas de ARN, estructuras tallo-lazo y estructuras más complejas como las bifurcaciones, pseudonudos y triples hélices. Estas estructuras influyen en la estabilidad del ARNm, p.ej., los elementos de la horquilla en la 3'UTR pueden servir como un sitio de escisión de endonucleasa, y afectan a la eficiencia de la traducción.

30

[0005] Además de la estructura del ARNm, el contenido de nucleótidos del ARNm también puede desempeñar un papel en la eficacia con la que se traduce el ARNm. Por ejemplo, el ARNm con un alto contenido de GC en la región 5' no traducida (5'UTR) puede traducirse con poca eficacia y un efecto de traducción reducido puede disminuir la estabilidad del mensaje. Por tanto, la alteración de la secuencia de una molécula de ARNm puede, en última instancia, influir en la estabilidad del transcrito del ARNm, por la influencia en la estabilidad de la traducción del mensaje.

35

[0006] Factor VIII y Factor IX son proteínas plasmáticas importantes que participan en la vía intrínseca de la coagulación sanguínea. Su disfunción o ausencia en los sujetos puede provocar trastornos en la coagulación sanguínea, p.ej, déficit de Factor VIII o Factor IX da lugar a hemofilia A o B, respectivamente. El aislamiento del Factor VIII o Factor IX de la sangre es difícil, p.ej., el aislamiento del Factor VIII se caracteriza por bajo rendimiento y también tiene el peligro asociado de contaminarse con agentes infecciosos como el virus de la hepatitis B, virus de la hepatitis C o VIH. La tecnología de ADN recombinante proporciona un método alternativo para producir Factor VIII o Factor IX biológicamente activos. Aunque estos métodos han cosechado éxito, continúa siendo un desafío mejorar el rendimiento de Factor VIII o Factor IX.

40

[0007] Un enfoque para aumentar la producción de proteína utilizando la tecnología de ADN recombinante es modificar la secuencia de codificación de una proteína de interés, p.ej., Factor VIII o Factor IX, sin alterar la secuencia de aminoácidos del producto génico. Este enfoque implica la alteración de, por ejemplo, la secuencia del gen del Factor VIII o Factor IX nativos, de modo que los codones que no se utilizan tan frecuentemente en células de mamíferos se sustituyen con codones que están sobreexpresados en genes de mamíferos altamente expresados. Seed *et al.*, (WO 98/12207) utilizaron este enfoque con cierto éxito. Descubrieron que la sustitución de codones de mamíferos escasos con los que se utilizan con frecuencia en las células de mamíferos conlleva un aumento de cuatro veces en la producción del Factor VIII a partir de células de mamíferos.

45 50

[0008] D1 (WO98/11206) muestra que un sujeto del que se sospecha que tiene déficit de  $\alpha$ -gal A como la enfermedad de Fabry puede tratarse con (1) células humanas que se han modificado genéticamente para sobreexpresar y secretar

$\alpha$ -gal A humana o (2)  $\alpha$ -gal A humana purificada obtenida a partir de células humanas genéticamente modificadas y cultivadas.

**Resumen de la invención**

5 [0009] La invención proporciona un ácido nucleico sintético de acuerdo con la reivindicación 1. Se describen otras características de la invención en las reivindicaciones dependientes. También se proporciona un vector de acuerdo con la reivindicación 5. Además, se proporciona una célula que contiene el vector según la reivindicación 6. La invención también proporciona un método de producción de alfa-galactosidasa de acuerdo con la reivindicación 7. También se presenta el uso de un ácido nucleico sintético que dirige la síntesis de un mensaje optimizado para la alfa-galactosidasa en la preparación de un medicamento según la reivindicación 10. Además, se muestra un ácido nucleico sintético que dirige la síntesis de un mensaje optimizado para alfa-galactosidasa en la preparación de un medicamento según la reivindicación 11. Otras características de las reivindicaciones 10 y 11 se definen en las reivindicaciones dependientes 12 a 16.

15 [0010] El ácido nucleico sintético puede dirigir la síntesis de un ARNm mensajero optimizado. En un modo de realización preferente, la extensión continua de los codones comunes puede incluir: la secuencia de una pre-proteína; la secuencia de una pro-proteína; la secuencia de una proteína madura; la "pre" secuencia de una pre-proteína; la secuencia "pre-pro" de una pre-proteína; la secuencia "pro" de una pre-pro o una pro-proteína; o una parte de cualquiera de las secuencias anteriormente mencionadas.

[0011] En un modo de realización preferido, la secuencia de ácido nucleico sintético incluye una longitud continua de al menos 90, 95, 100, 125, 150, 200, 250, 300 o más codones, todos ellos codones comunes.

20 [0012] En otro modo de realización preferente, la secuencia de ácido nucleico que codifica una proteína tiene al menos 30, 50, 60, 75, 100, 200 o más codones no comunes o menos comunes sustituidos con un codón común.

[0013] En un modo de realización preferido, el número de codones no comunes o menos comunes reemplazado es inferior a 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1.

25 [0014] En un modo de realización preferido, el número de codones no comunes o menos comunes que quedan es inferior a 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1.

[0015] En modos de realización preferidos, los codones no comunes y menos comunes reemplazados, en conjunto, son igual o inferior al 6%, 5%, 4%, 3%, 2%, 1% de los codones en la secuencia de ácido nucleico sintético.

[0016] En modos de realización preferidos, los codones no comunes y menos comunes restantes, en conjunto, son igual o inferior al 6%, 5%, 4%, 3%, 2%, 1% de los codones en la secuencia de ácido nucleico sintético.

30 [0017] En un modo de realización preferido, todos los codones no comunes o menos comunes de la secuencia de ácido nucleico sintético que codifica una proteína han sido sustituidos con codones comunes.

[0018] En un modo de realización preferido, la secuencia de ácido nucleico sintético codifica una proteína de al menos aproximadamente 90, 95, 100, 120, 130, 150, 200, 500, 700, 1000 o más aminoácidos de longitud.

35 [0019] En varios modos de realización preferidos, al menos el 94%, 95%, 96%, 97%, 98%, 99% de los codones, o todos, en la secuencia de ácido nucleico sintético son codones comunes. De modo preferente, todos los codones en la secuencia de ácido nucleico sintético son codones comunes.

[0020] En modos de realización preferentes, la proteína se expresa en una célula eucariota, p.ej., una célula de mamífero, p.ej., una célula humana y la proteína es una proteína de mamífero, p.ej. una proteína humana.

40 [0021] Además se describe una secuencia de ácido nucleico sintético que puede dirigir la síntesis de un mensaje optimizado que codifica  $\alpha$ -galactosidasa.

[0022] En un modo de realización preferido, se inserta la secuencia de ácido nucleico sintético que codifica  $\alpha$ -galactosidasa, p.ej., mediante transfección, en una célula no transformada, por ejemplo, una célula primaria o secundaria, p. ej., un fibroblasto humano primario.

45 [0023] En otro aspecto, la invención muestra un constructo de ADN o un plásmido, p. ej., un plásmido de expresión o un constructo de ADN, que incluye una secuencia de ácido nucleico sintético descrita en este documento.

[0024] En otro aspecto más, la invención muestra una secuencia de ácido nucleico sintético descrita en el presente documento introducida en el genoma de una célula animal. En un modo de realización preferido, la célula animal es una célula de primate, p. ej., una célula de mamífero, p. ej., una célula humana.

**[0025]** En otro aspecto más, la invención incluye una célula que alberga una secuencia de ácido nucleico sintético descrita en el presente documento, p.ej., una célula de una cepa de célula primaria o secundaria, o una célula de una línea celular continua, p.ej., una célula de melanoma de Bowes (núm. de registro ATCC CRL 9607), una célula Daudi (núm. de depósito ATCC CCL 213), una célula HeLa y un derivado de una célula HeLa (núms. de depósito ATCC CCL 2. CCL 2.1 y CCL 2.2), una célula HL-60 (núm. de depósito ATCC CCL 240), una célula HT-1080 (núm. de depósito ATCC CCL 121), una célula Jurkat (núm. de depósito ATCC TIB 152), una célula de carcinoma KB (núm. de depósito ATCC CCL 17), una célula de leucemia K-562 (núm. de depósito ATCC CCL 243), una célula de cáncer de mama MCF-7 (núm. de depósito ATCC BTH 22), una célula MOLT-4 (núm. de depósito ATCC 1582), una célula Namalwa (núm. de depósito ATCC CRL 1432), una célula Raji (núm. de depósito ATCC CCL 86), una célula RPMI 8226 (núm. de depósito ATCC CCL 155), una célula U-937 (núm. de depósito ATCC CRL 1593), una célula 2R4 de la sublínea WI-38VA13 (núm. de depósito ATCC CLL 75.1), una célula CCRF-CEM (núm. de depósito ATCC CCL 119) y una célula de carcinoma de ovario 2780AD (Van Der Blick *et al.*, *Cancer Res.* 48: 5927-5932, 1988), así como células de heterohibridoma producidas por fusión de células humanas y células de otras especies. En otro modo de realización, la línea celular inmortalizada puede ser una línea celular que no sea una línea celular humana, p.ej., una línea celular CHO o una línea celular COS. En un modo de realización preferido, la célula es una célula no transformada. En un modo de realización preferido, la célula es de una cepa de célula clónica. En diversos modos de realización preferidos, la célula es una célula de mamífero, p.ej., una célula de mamífero primaria o secundaria, p.ej., un fibroblasto, una célula madre hematopoyética, un mioblasto, un queratinocito, una célula epitelial, una célula endotelial, una célula glial, una célula neural, una célula que comprende un elemento formado de la sangre, una célula muscular y precursores de estas células somáticas. En el modo de realización más preferido, la célula es un fibroblasto humano secundario.

**[0026]** En otro aspecto, la invención muestra un método para preparar un ácido nucleico sintético que codifica alfa-galactosidasa que es de al menos 90 codones de longitud. El método incluye la identificación de un codón no común y un codón menos común en una secuencia génica no optimizada que codifica una proteína alfa-galactosidasa; y la sustitución de al menos el 94% de los codones no comunes y menos comunes con un codón común que codifica el mismo aminoácido que el codón sustituido para una expresión de proteína aumentada; en el que por codón común se entiende: Ala (gcc), Arg (cgc), Asn (aac), Asp (gac), Cys (tgc), Gln (cag), Gly (ggc), His (cac), Ile (atc), Leu (ctg), Lys (aag), Pro (ccc), Phe (ttc), Ser (agc), Thr (acc), Tyr (tac), Glu (gag) y Val (gtg); en el que los codones menos comunes son: Gly (ggg), Ile (att), Leu (ctc), Ser (tcc), Val (gtc) y Arg (agg). Preferiblemente, todos los codones distintos de los codones comunes y de los codones menos comunes son codones no comunes.

**[0027]** También se describe en el presente documento que el método incluye la identificación de codones no comunes y menos comunes en el gen no optimizado que codifica la proteína y la sustitución de al menos el 94%, 95%, 96%, 97%, 98%, 99% o más de los codones no comunes y menos comunes con un codón común que codifica el mismo aminoácido que el codón sustituido.

**[0028]** En un modo de realización preferido, la secuencia de ácido nucleico sintético codifica una proteína de al menos aproximadamente 90, 95, 100, 105, 110, 120, 130, 150, 200, 500, 700, 1000 o más codones de longitud.

**[0029]** En modos de realización preferidos, la proteína se expresa en una célula eucariota, p.ej., una célula de mamífero, p.ej., una célula humana y la proteína es una proteína de mamífero, p.ej., una proteína humana.

**[0030]** También se describe en el presente documento un método para preparar una secuencia de ácido nucleico que dirija la síntesis de un mensaje optimizado de una proteína de al menos 90, 100 o 120 aminoácidos de longitud, p.ej., una secuencia de ácido nucleico sintético descrita en este documento. El método incluye: sintetizar al menos dos fragmentos de la secuencia de ácido nucleico, en la que los dos fragmentos codifican partes adyacentes de la proteína y donde ambos fragmentos son ARNm optimizados, p.ej., como se describe en el documento; y unir los dos fragmentos de modo que no se cree un codón no común en un punto de unión, con lo que se obtiene la secuencia de ácido nucleico de ARNm optimizado.

**[0031]** En un modo de realización preferido, los dos fragmentos se unen entre sí de modo que no se duplica un sitio único de endonucleasa de restricción utilizado para crear los dos fragmentos en el punto de unión. En otro modo de realización preferido, los dos fragmentos se unen entre sí de modo que se crea un sitio de restricción único.

**[0032]** En un modo de realización preferido, la secuencia de ácido nucleico sintético codifica una proteína de al menos 90, 95, 100, 105, 110, 120, 130, 150, 200, 500, 700, 1000 o más codones de longitud.

**[0033]** En un modo de realización preferido, se sintetizan al menos 3, 4, 5, 6, 7, 8, 9, 10 o más fragmentos de la secuencia de ácido nucleico.

**[0034]** En un modo de realización preferido, los fragmentos se unen entre sí por una fusión, p.ej., una fusión de extremos romos.

**[0035]** En diversos modos de realización preferidos, al menos el 94%, 95%, 96%, 97%, 98%, 99% de los codones o todos los codones de la secuencia de ácido nucleico sintético son codones comunes. Preferentemente, todos los codones de la secuencia de ácido nucleico sintético son codones comunes.

5 **[0036]** En modos de realización preferidos, el número de codones que no son codones comunes es igual o inferior a 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1.

**[0037]** En modos de realización preferidos, cada fragmento es de al menos 30, 40, 50, 75, 100, 120, 150 o más codones de longitud.

10 **[0038]** En el presente documento se describe un método para proporcionar una proteína a un sujeto, por ejemplo, un humano. Los métodos incluyen: disponer una secuencia de ácido nucleico sintético que puede dirigir la síntesis de un mensaje optimizado para una proteína, p.ej., una secuencia de ácido nucleico sintético descrita en este documento; introducir la secuencia de ácido nucleico sintético que dirige la síntesis de un mensaje optimizado para una proteína en el sujeto; y permitir que el sujeto exprese la proteína, proporcionándole así la proteína al sujeto.

15 **[0039]** En modos de realización preferidos, el método incluye además la inserción de la secuencia de ácido nucleico que dirige la síntesis de un mensaje optimizado en una célula. La célula puede ser una célula autóloga, alogénica o xenogénica, pero se prefiere la autóloga. Una célula preferida es un fibroblasto, una célula madre hematopoyética, un mioblasto, una queratinocita, una célula epitelial, una célula endotelial, una célula glial, una célula neural, una célula que comprende un elemento formado de la sangre, una célula muscular y precursores de estas células somáticas. La secuencia de ácido nucleico sintético optimizado de ARNm puede insertarse en la célula *ex vivo* o *in vivo*. Si se inserta *ex vivo*, la célula puede introducirse en el sujeto.

20 **[0040]** En modos de realización preferidos, al menos el 94%, 95%, 96%, 97%, 98%, 99% de los codones o todos los codones en la secuencia de ácido nucleico sintético son codones comunes. Preferentemente, todos los codones en la secuencia de ácido nucleico sintético son codones comunes.

**[0041]** En modos de realización preferidos, el número de codones que no son codones comunes es igual o inferior a 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1.

25 **[0042]** Además se proporcionan fragmentos de ácido nucleico sintético que codifica una parte de una proteína. Estos fragmentos de ácido nucleico sintético son similares a la secuencia de ácido nucleico sintético de la invención, con la diferencia de que codifican solo una parte de una proteína. Preferentemente, dichos fragmentos de ácido nucleico codifican al menos 50, 60, 70, 80, 100, 110, 120, 130, 150, 200, 300, 400, 500 o más aminoácidos contiguos de la proteína.

30 **[0043]** También se describen en el presente documento células somáticas primarias y secundarias infectadas o transfectadas de origen vertebrado, en concreto, de origen de mamífero, p.ej., de origen humano, de ratón o de conejo, p.ej., células humanas primarias, células humanas secundarias o células de conejo primarias o secundarias. Las células se infectan o transfectan con ácido nucleico sintético exógeno, p.ej., ADN, como se describe en el presente documento. El ácido nucleico sintético puede codificar una proteína, p.ej., una proteína terapéutica, p.ej., una enzima, p.ej.,  $\alpha$ -galactosidasa, una citocina, una hormona, un antígeno, un anticuerpo, un factor de coagulación, p.ej., Factor VIII, Factor IX o una proteína reguladora. También se proporcionan métodos mediante los cuales se transfectan o infectan células primarias y secundarias para incluir ADN sintético exógeno, métodos para producir cepas de células clónicas o cepas de células heterogéneas y métodos de terapia génica en los que se utilizan las células primarias o secundarias transfectadas o infectadas. El ácido nucleico sintético dirige la síntesis de un mensaje optimizado, p.ej., un mensaje optimizado como se describe en el presente documento.

35 **[0044]** Además, se proporcionan aquí células somáticas primarias o secundarias, que han sido transfectadas o infectadas con un ácido nucleico sintético exógeno descrito en el presente documento, que se integran de forma estable en sus genomas o que se expresan en las células de modo episomal. En modos de realización preferidos las células son fibroblastos, queratinocitos, células epiteliales, células endoteliales, células gliales, células neurales, células que comprenden un elemento formado de la sangre, células musculares, otras células somáticas que se pueden cultivar o precursoras de células somáticas. Las células resultantes se denominan, respectivamente, células primarias infectadas o transfectadas o células secundarias infectadas o transfectadas. El ADN sintético exógeno codifica una proteína, o una parte de la misma, p.ej., una proteína terapéutica (p.ej. Factor VIII o Factor IX). En el modo de realización en el que el ADN sintético exógeno codifica una proteína, o una parte de la misma, para que las células receptoras la expresen, la proteína resultante puede ser retenida en la célula, incorporada en la membrana celular o secretada a partir de la célula. En este modo de realización, el ADN sintético exógeno que codifica la proteína se introduce en las células junto con las secuencias de ADN adicionales suficientes para la expresión de ADN sintético exógeno en las células. Las secuencias de ADN adicionales pueden ser de origen viral o no viral. A las células primarias modificadas para expresar ADN sintético exógeno se las denomina en el presente documento células primarias transfectadas o infectadas, que incluyen células extraídas de tejido y colocadas en medio de cultivo por primera vez. Las células secundarias modificadas para expresar

o hacer que el ADN exógeno esté disponible se denominan en este documento células secundarias transfectadas o infectadas.

5 **[0045]** Las células primarias y secundarias transfectadas o infectadas por el método en cuestión, p.ej., cepas de células clonadas, se dividen en tres tipos o categorías 1) células que, como han sido obtenidas, no producen ni contienen la proteína terapéutica, 2) células que producen o contienen la proteína terapéutica, pero en menor cantidad que las normales (en cantidades menores que el nivel inferior fisiológicamente normal) o en forma defectuosa, y 3) células que producen la proteína terapéutica en niveles fisiológicamente normales, pero que han de aumentar o mejorar su contenido o producción. Entre los ejemplos de proteínas que se pueden producir con el presente método se incluyen las citocinas o los factores de coagulación.

10 **[0046]** El ADN sintético exógeno se introduce en la célula primaria o secundaria mediante varias técnicas. Por ejemplo, un constructo de ADN que incluye ADN sintético exógeno que codifica una proteína terapéutica y secuencias de ADN adicionales necesarias para la expresión en células receptoras se puede introducir en células primarias o secundarias mediante electroporación, microinyección u otros medios (p.ej., precipitación con fosfato de calcio, precipitación con fosfato de calcio modificada, precipitación con polibreno, fusión con liposoma, administración de ADN mediada por receptores). Por otro parte, se puede utilizar un vector, como un vector retroviral que incluye el ADN exógeno y, como resultado de la infección con el vector, se pueden modificar genéticamente las células.

15 **[0047]** Además del ADN sintético exógeno, las células primarias y secundarias transfectadas o infectadas pueden contener opcionalmente ADN que codifica a un marcador de selección, que se expresa y confiere a los receptores un fenotipo de selección, como resistencia a antibióticos, resistencia a un agente citotóxico, fototropía nutricional o la expresión de una proteína de superficie. Su presencia hace posible identificar y seleccionar células que contienen el ADN exógeno. Se pueden utilizar una variedad de genes marcadores de selección, como neo, gpt, dhfr, ada, pac, hyg, mdv y hisD.

20 **[0048]** Las células transfectadas o infectadas descritas en la presente invención son útiles, como poblaciones de células primarias o secundarias transfectadas o infectadas, cepas de células clónicas transfectadas o infectadas, cepas de células heterogéneas transfectadas o infectadas y como mezclas de células en las que está presente por lo menos una célula representativa de una de las tres categorías anteriores de células transfectadas o infectadas (p.ej., la mezcla de células contiene básicamente células primarias o secundarias transfectadas o infectadas y puede incluir células primarias o secundarias sin transfectar o sin infectar), como sistema de administración para el tratamiento de un sujeto con una enfermedad anormal o indeseada que responde a la administración de la proteína terapéutica, que es: 1) una proteína terapéutica (p.ej., una proteína que está ausente, infraproducida con relación a las necesidades fisiológicas del sujeto, defectuosa, ineficaz o inapropiadamente utilizada en el sujeto, p.ej., Factor VIII o Factor IX; o 2) una proteína terapéutica con nuevas funciones, como las funciones enzimática o de transporte como la  $\alpha$ -galactosidasa. En el método descrito en el presente documento para la provisión de una proteína terapéutica, se pueden administrar células primarias o secundarias transfectadas o infectadas, cepas de células clónicas o cepas de células heterogéneas a un sujeto que padece la enfermedad anormal o indeseada para ser tratada o prevenida, en cantidad suficiente y por una vía adecuada, para expresar el ADN sintético exógeno a niveles fisiológicos relevantes. Un nivel fisiológicamente relevante es aquel que se aproxima al nivel al que se produce el producto en el cuerpo o que produce la mejoría de la enfermedad anormal o indeseada.

25 **[0049]** Las cepas de células clónicas de células secundarias transfectadas o infectadas (denominadas cepas de células clónicas transfectadas o infectadas) que expresan ADN sintético exógeno (y, opcionalmente, que incluyen un gen marcador de selección) pueden producirse mediante el procedimiento descrito en el presente documento. El método incluye: 1) la provisión de una población de células primarias, obtenidas del sujeto al que se le administrarán las células primarias transfectadas o infectadas o de otro origen, 2) la introducción en las células primarias o en las células secundarias procedentes de células primarias de un constructo de ADN que incluye ADN exógeno como se describe anteriormente y las secuencias de ADN adicionales necesarias descritas anteriormente, que produzcan células primarias o secundarias transfectadas o infectadas; 3) el mantenimiento de estas células primarias o secundarias transfectadas o infectadas en condiciones apropiadas para su propagación; 4) la identificación de una célula primaria o secundaria transfectada o infectada; y 5) la producción de una colonia a partir de la célula primaria o secundaria transfectada o infectada identificada (4) gracias al mantenimiento en condiciones de cultivo apropiadas hasta que se obtiene un número deseado de células. El número deseado de células clónicas es un número suficiente para producir una cantidad de producto terapéuticamente eficaz cuando se administra a un sujeto, p.ej., a un sujeto que padece hemofilia A se le administra una población de células que producen una cantidad de Factor VIII terapéuticamente eficaz, de modo que se trata la enfermedad. El sujeto también puede ser, por ejemplo, un sujeto diagnosticado con hemofilia B o un sujeto con déficit de la  $\alpha$ -galactosidasa como un sujeto que padece la enfermedad de Fabry. El número de células necesarias para una dosis terapéutica determinada depende de muchos factores entre los que se incluyen el nivel de expresión de la proteína, la enfermedad del animal huésped, y las limitaciones relacionadas con el procedimiento de implantación. En general, el número de células necesarias para la implantación está en el intervalo de  $1 \times 10^6$  a  $5 \times 10^9$  y preferentemente de  $1 \times 10^8$  a  $5 \times 10^8$ . En un modo de realización del método, la célula identificada en (4) se somete a aproximadamente 27 duplicaciones (es decir, se somete a 27 ciclos de crecimiento celular y división celular) para producir 100 millones de células clónicas infectadas o transfectadas. En otro modo de realización del método, se

introduce el ADN sintético exógeno en el ADN genómico por recombinación homóloga entre las secuencias de ADN presentes en el constructo de ADN y ADN genómico. En otro modo de realización, el ADN sintético exógeno está presente en forma episomal en una célula transfectada, p.ej., una célula primaria o secundaria.

5 **[0050]** En un modo de realización de producción de una población clónica de células secundarias transfectadas, se combina una suspensión celular que contiene células primarias o secundarias con ADN sintético exógeno que codifica una proteína terapéutica y ADN que codifica un marcador de selección, como el gen neo. Las dos secuencias de ADN están presentes en el mismo constructo de ADN o en dos constructos de ADN separados. La combinación resultante se somete a electroporación, en general a 250-300 voltios con una capacitancia de 960  $\mu$ Faradios y una constante de tiempo apropiada (p.ej., 14 a 20 ms) para que las células recojan el constructo de ADN. En un modo de realización alternativo, se utiliza la microinyección para introducir el constructo de ADN en células primarias o secundarias. En cualquiera de los modos de realización, la introducción del ADN exógeno produce células primarias o secundarias transfectadas. El ADN sintético exógeno introducido en la célula puede integrarse de modo estable en el ADN genómico o está presente en forma episomal en la célula.

10 **[0051]** En el método de producción de cepas de células heterogéneas incluido en este documento, se realizan las mismas etapas que las descritas para la producción de una cepa celular clónica, pero no se aísla una única célula primaria o secundaria transfectada ni se utiliza como célula fundadora. En su lugar, se cultivan dos o más células primarias o secundarias transfectadas para producir una cepa de célula heterogénea. Una cepa de célula heterogénea también puede contener además dos o más células primarias o secundarias transfectadas, células primarias o secundarias no transfectadas.

15 **[0052]** Los métodos descritos en el presente documento tiene una gran aplicabilidad en el tratamiento de enfermedades anormales o no deseadas y se pueden utilizar para producir varias proteínas en una cantidad eficaz para un sujeto. Por ejemplo, pueden utilizarse para producir proteínas secretadas (con efectos predominantemente sistémicos o predominantemente locales, p.ej., Factor VIII y Factor IX), proteínas de membrana (p.ej., para conferir una capacidad de respuesta nueva o mejorada, facilitando la eliminación de un producto tóxico o para marcar o dirigir a una célula) o proteínas intracelulares (p.ej., para afectar a la expresión génica o producir efectos autocrinos).

20 **[0053]** Un método descrito en el presente documento es útil en particular en el tratamiento de las enfermedades anormales o indeseadas en cuanto: 1) es curativo (un tratamiento de terapia génica tiene el potencial de prolongarse durante toda la vida del paciente); 2) permite una dosificación exacta (las células del paciente determinan y suministran continuamente la dosis óptima de la proteína necesaria basándose en las demandas fisiológicas, y las cepas celulares transfectadas o infectadas de forma estable pueden caracterizarse extensamente *in vitro* antes de la implantación, logrando predicciones exactas de la función *in vivo* a largo plazo); 3) es simple de aplicar para el tratamiento de pacientes; 4) elimina problemas en cuanto a la adaptabilidad de los pacientes (tras un tratamiento por terapia génica realizado de una sola vez, ya no son necesarias las inyecciones diarias de proteínas); y 5) reduce los costes del tratamiento (puesto que las propias células del paciente sintetizan la proteína terapéutica, la inversión en una costosa producción y purificación de proteínas se vuelve innecesaria).

25 **[0054]** Como se utiliza en el presente documento, el término "ARN mensajero optimizado" hace referencia a una secuencia de ácido nucleico sintético que codifica una proteína donde al menos un codón no común o un codón menos común en la secuencia que codifica la proteína ha sido sustituido por un codón común.

30 **[0055]** Por "codón común" se entiende el codón más común que representa un aminoácido concreto en una secuencia humana. La frecuencia de los codones en los genes humanos de alta expresión se describe en la Tabla 1 a continuación. Entre los codones comunes se incluyen: Ala (gcc), Arg (cgc), Asn (aac), Asp (gac), Cys (tgc), Gln (cag), Gly (ggc), His (cac), Ile (atc), Leu (ctg), Lys (aag), Pro (ccc), Phe (ttc), Ser (agc), Thr (acc), Tyr (tac), Glu (gag) y Val (gtg) (véase la Tabla 1). Los "codones menos comunes" son codones que se producen con frecuencia en humanos pero que no son codones comunes: Gly (ggg), Ile (att), Leu (etc), Ser (tcc), Val (gtc) y Arg (agg). Todos los codones que no son codones comunes ni codones menos comunes son "codones no comunes".

TABLA 1: Frecuencia de los codones en los genes humanos altamente expresados

	% de incidencia			% de incidencia	
Ala GC	C	53	Cys TG	C	68
	T	17		T	32
	A	13	Gln CA	A	12
	G	17		G	88
Arg CG	C	37	Glu		
	T	7			

## ES 2 437 815 T3

	A	6	GA	A	25
	G	21		G	75
AG	A	10			
	G	18	Gly		
			GG	C	50
Asn				T	12
AA	C	78		A	14
	T	25		G	24
Leu			His		
CT	C	26	CA	C	79
	T	5		T	21
	A	3			
	G	58	Ile		
TT	A	2	AT	C	77
	G	6		T	18
				A	5
Lys			Ser		
AA	A	18	TC	C	28
	G	82		T	13
				A	5
Pro				G	9
CC	C	48	AG	C	34
	T	19		T	10
	A	16			
	G	17	Thr		
Phe			AC	C	57
TT	C	80		T	14
	T	20		A	14
				G	15
			Tyr		
			TA	C	74
				T	26
			Val		
			GT	C	25
				T	7
				A	5
				G	64

**[0056]** La frecuencia de codones de la Tabla 1 se calculó usando el programa GCG establecido por University of Wisconsin Genetics Computer Group. Los números representan el porcentaje de casos en los que se utiliza el codón concreto.

5 **[0057]** El término "célula primaria" incluye las células presentes en una suspensión de células aisladas procedentes de un tejido de vertebrado (antes de ser puestas en placas, es decir, unidas a un sustrato de cultivo de tejido tal como una cápsula o un matraz), las células presentes en un explante de tejido, de ambos tipos anteriores de células colocadas en placas por primera vez y las suspensiones celulares procedentes de estas células colocadas en placas. El término célula secundaria o cepa celular se refiere a las células en todas las etapas posteriores en el cultivo. Es decir, la primera vez que una célula primaria en placa se extrae del sustrato de cultivo y se vuelve a colocar en la placa (transferida), se denomina célula secundaria en este documento, como están todas las células en las transferencias posteriores. Las células secundarias son cepas celulares que constan de células secundarias que se han transferido una o más veces. Una cepa celular consta de células secundarias que: 1) han sido transferidas una o más veces; 2) presentan un número finito de duplicaciones de población media en cultivo; 3) presentan las propiedades de inhibición por contacto, crecimiento dependiente de anclaje (la dependencia de anclaje no se aplica a las células que se propagan en el cultivo en suspensión); y 4) no se immortalizan. Una "cepa de células clónicas" se define como una cepa celular que procede de una única célula fundadora. Una "cepa de células heterogéneas" se define como una cepa de células que procede de dos o más células fundadoras.

20 **[0058]** El término "célula transfectada" se refiere a una célula en la que se introduce una secuencia de ácido nucleico sintético exógeno, p.ej., una secuencia que codifica una proteína. Una vez en la célula, la secuencia de ácido nucleico sintético puede incorporar ADN cromosómico en las células receptoras o puede existir en forma episomal. Los métodos de transfección estándares pueden utilizarse para introducir la secuencia de ácido nucleico sintético en una célula, p.ej., la transfección mediada por liposoma, polibreno, transfección mediada por DEAE-dextrano, electroporación, precipitación con fosfato de calcio o microinyección. El término "transfección" no incluye la administración de ADN o

ARN en una célula por un virus. El término “célula infectada” hace referencia a una célula en la que se introduce una secuencia de ácido nucleico sintético exógeno, p.ej., una secuencia que codifica una proteína, mediante un virus. Entre los virus conocidos por ser útiles para la transferencia génica, se incluyen: un adenovirus, un virus adeno-asociado, un virus del herpes, un virus de las paperas, un poliovirus, un retrovirus, un virus Sindbis, un lentivirus y un virus vaccinia como un virus de la viruela del canario. Otras características y ventajas de la invención serán claras según la siguiente descripción detallada y las reivindicaciones.

### **Descripción detallada de la invención**

**[0059]** Los dibujos se describen primero brevemente.

10 La *Figura 1* es una representación esquemática de estructuras del Factor VIII humano (hFVIII) con el dominio B eliminado (BDD) y de longitud completa.

La *Figura 2* es una representación esquemática del hFVIII de longitud completa.

La *Figura 3* es una representación esquemática del plásmido de expresión de BDD hFVIII de 5R, pXF8.186.

La *Figura 4* es una representación esquemática del plásmido de expresión de BDD hFVIII de LE, pXF8.61.

15 La *Figura 5* es una representación esquemática de los catorce fragmentos (Fragmentos A- Fragmentos N) ensamblados al constructo pXF8.61.

La *Figura 6* es una representación esquemática del ensamblaje de pXF8.61.

La *Figura 7* representa la secuencia de nucleótidos y la secuencia de aminoácidos correspondiente del inserto de Factor VIII (FVIII) con el dominio B suprimido de LE contenido en pAM1-1 (SEQ ID N°:1).

La *Figura 8* es una representación esquemática de los fragmentos ensamblados al constructo pXF8.186.

20 La *Figura 9* representa la secuencia de nucleótidos y la secuencia de aminoácidos correspondiente del inserto de FVIII con el dominio B suprimido de 5Arg (SEQ ID N°:2).

25 La *Figura 10* es una representación esquemática del plásmido de expresión de Factor VIII, pXF8.36. El promotor I temprano inmediato del citomegalovirus (CMV) se representa con un recuadro ligeramente sombreado. Debajo de este recuadro sombreado, se indican las posiciones del sitio donador de empalme (DS) y el aceptor de empalme (SA). La secuencia de ADNc del Factor VIII se representa con un recuadro oscuro. La región 3'UTS de hGH se representa como un recuadro claro. El nuevo casete de expresión se representa con un recuadro sombreado con una flecha que se corresponde con la dirección de transcripción. La delgada línea oscura representa las secuencias de la cadena principal del plásmido. La posición y la dirección de transcripción del gen de la  $\beta$ -lactamasa (*amp*) se indica con la flecha en forma de recuadro oscuro.

30 La *Figura 11* es una representación esquemática del plásmido de expresión de Factor VIII, pXF8.38. El promotor I temprano inmediato del citomegalovirus (CMV) se representa con un recuadro ligeramente sombreado. Debajo de este recuadro sombreado, se indican las posiciones del sitio donador de empalme (SD) y el aceptor de empalme (SA). La secuencia de ADNc del Factor VIII se representa con un recuadro oscuro. La región 3'UTS de hGH se representa como un recuadro claro. El casete de expresión *neo* se representa con un recuadro sombreado con una flecha que se corresponde con la dirección de transcripción. La delgada línea oscura representa las secuencias de la cadena principal del plásmido. La posición y la dirección de transcripción del gen de la  $\beta$ -lactamasa (*amp*) se indica con la flecha en forma de recuadro oscuro.

35 La *Figura 12* es una representación esquemática del plásmido de expresión de Factor VIII, pXF8.269. El promotor de colágeno  $\alpha$ -2 (I) se representa con un recuadro rayado. La región que representa las secuencias no traducidas 5' derivadas de aldolasa se representa con un recuadro ligeramente sombreado. Debajo de este recuadro sombreado, se indican las posiciones del sitio donador de empalme (SD) y el aceptor de empalme (SA). La secuencia de ADNc del Factor VIII se representa con un recuadro sombreado oscuro. La región 3'UTS de hGH se representa como un recuadro claro. El casete de expresión de *neo* se representa con un recuadro sombreado con una flecha que se corresponde con la dirección de transcripción. La delgada línea oscura representa las secuencias de la cadena principal del plásmido. La posición y la dirección de transcripción del gen de la  $\beta$ -lactamasa (*amp*) se indica con la flecha en forma de recuadro oscuro.

40 La *Figura 13* es una representación esquemática del plásmido de expresión de Factor VIII, pXF8.224. El promotor de colágeno  $\alpha$ -2 (I) se representa con un recuadro rayado. La región que representa las secuencias no traducidas 5' derivadas de aldolasa se representan con un recuadro ligeramente sombreado. Debajo de este recuadro sombreado, se indican las posiciones del sitio donador de empalme (SD) y el aceptor de empalme (SA). La

50

secuencia de ADNc del Factor VIII se representa con un recuadro oscuro. La región 3'UTS de hGH se representa como un recuadro claro. El casete de expresión de *neo* se representa con un recuadro sombreado con una flecha que se corresponde con la dirección de transcripción. La delgada línea oscura representa las secuencias de la cadena principal del plásmido. La posición y la dirección de transcripción del gen de la  $\beta$ -lactamasa (*amp*) se indica con la flecha en forma de recuadro oscuro.

La *Figura 14* es una representación esquemática de los fragmentos ensamblados al constructo pFIXABCD. Los sitios de restricción que están cortados aparecen en negrita y las uniones de la última etapa están subrayadas. La dirección de transcripción de la secuencia de FIXABCD se indica con una flecha oscura.

La *Figura 15* representa la secuencia de nucleótidos del inserto FIXABCD (SEQ ID N°:105).

La *Figura 16* es una representación esquemática de los plásmidos de expresión del Factor IX, pXIX76 y pXIX170. Las flechas del interior del círculo indican marcos de lectura abiertos. Las flechas sobre el círculo indican secuencias promotoras; una flecha de doble punta indica un potenciador. Las líneas finas indican intrones o secuencias de vector bacteriano y los recuadros gruesos trazan la secuencia traducida. Las líneas dobles indican secuencias genómicas no transcritas, mientras que las líneas de grosor intermedio indican partes no traducidas del ARNm. El plásmido pXIX170 tiene una secuencia de ADNc del Factor IX que se optimiza, pero pXIX76 no.

La *Figura 17* ilustra la secuencia de nucleótidos del inserto de  $\alpha$ -galactosidasa (SEQ ID N°:106).

La *Figura 18* es una representación esquemática de los plásmidos de expresión de  $\alpha$ -galactosidasa, pXAG94 y pXAG95. Las flechas del interior del círculo indican marcos de lectura abiertos. Las flechas sobre el círculo indican secuencias promotoras; una flecha de doble punta indica un potenciador. Las líneas finas indican intrones o secuencias de vector bacteriano y los recuadros gruesos trazan la secuencia traducida. Las líneas dobles indican secuencias genómicas no transcritas, mientras que las líneas de grosor intermedio indican partes no traducidas del ARNm. El plásmido pXAG95 tiene una secuencia de ADNc de  $\alpha$ -galactosidasa que se optimiza, pero pXAG94 no.

La *Figura 19* es una representación esquemática de los plásmidos de expresión de  $\alpha$ -galactosidasa pXAG73 y pXAG74. Las flechas del interior del círculo indican marcos de lectura abiertos. Las flechas sobre el círculo indican secuencias promotoras; una flecha de doble punta indica un potenciador. Las líneas finas indican intrones o secuencias de vector bacteriano y los recuadros gruesos trazan la secuencia traducida. Las líneas dobles indican secuencias genómicas no transcritas, mientras que las líneas de grosor intermedio indican partes no traducidas del ARNm. El plásmido pXAG74 tiene una secuencia de ADNc de  $\alpha$ -galactosidasa que se optimiza, pero pXAG73 no.

#### Optimización del mensaje

**[0060]** Los métodos descritos en el presente documento se dirigen a mensajes optimizados y secuencias de ácidos nucleicos sintéticos que dirigen la producción de ARNm optimizados. Un ARNm optimizado puede dirigir la síntesis de una proteína de interés, p.ej., una proteína humana, p.ej., un Factor VIII humano, un Factor IX humano o una  $\alpha$ -galactosidasa humana. Puede optimizarse un mensaje para una proteína de interés, p.ej., un Factor VIII humano, un Factor IX humano o una  $\alpha$ -galactosidasa humana, como se describe en este documento, p.ej., reemplazando al menos el 94%, 95%, 96%, 97%, 98%, 99% de los codones no comunes o menos comunes, o preferentemente todos ellos, con un codón común que codifica el mismo aminoácido como se enumera en la Tabla 1.

**[0061]** La región de codificación de una secuencia de ácido nucleico sintético puede incluir la secuencia "cg" sin discriminación alguna, si la secuencia se encuentra en el codón común para ese aminoácido. Como alternativa, puede limitarse la secuencia "cg" en diversas regiones, p.ej., puede diseñarse el primer 20% de la secuencia de codificación para que tenga baja incidencia de la secuencia "cg".

**[0062]** La optimización de un mensaje (y de su secuencia de ADN sintético) puede afectar positiva o negativamente a la expresión génica o a la producción de proteínas. Por ejemplo, reemplazar un codón menos común con un codón más común puede afectar a la vida media del ARNm o alterar su estructura mediante la introducción de una estructura secundaria que interfiere en la traducción del mensaje.

**[0063]** Se puede optimizar todo o parte de un mensaje (o de su gen). En algunos casos, se logra la modulación deseada de la expresión optimizando básicamente el mensaje completo. En otros casos, se conseguirá la modulación deseada optimizando una parte del mensaje o el gen, pero no su totalidad.

**[0064]** El uso del codón de cualquier secuencia de codificación puede modificarse para lograr una propiedad deseada, por ejemplo, niveles elevados de expresión en un tipo de célula específico. El punto de inicio para esta optimización puede ser una secuencia de codificación con el 100% de codones comunes, o una secuencia de codificación que contiene una mezcla de codones comunes y no comunes.

**[0065]** Se generan y analizan dos o más secuencias candidatas que difieren en su uso de codón para determinar si poseen la propiedad deseada. En un principio, pueden evaluarse las secuencias candidatas empleando un ordenador para buscar la presencia de elementos reguladores, como silenciadores o potenciadores, y para buscar la presencia de regiones de la secuencia de codificación que podrían convertirse en dichos elementos reguladores mediante una alteración en el uso de los codones. Los criterios adicionales pueden incluir el enriquecimiento para nucleótidos concretos, p.ej., A, C, G o U, la preferencia codónica para un aminoácido concreto o la presencia o ausencia de estructura secundaria o terciaria de ARNm concreta. El ajuste a la secuencia candidata puede llevarse a cabo basándose en una serie de esos criterios.

**[0066]** Las secuencias candidatas prometedoras se construyen y después se evalúan de modo experimental. Pueden evaluarse múltiples candidatos de forma independiente el uno del otro, o el proceso puede ser iterativo, ya sea utilizando el candidato más prometedor como nuevo punto de inicio o combinando regiones de dos o más candidatos para producir un nuevo híbrido. Pueden incluirse otras fases de modificación y evaluación.

**[0067]** La modificación del uso de codones de una secuencia candidata puede conllevar la creación o destrucción ya sea de un elemento positivo o negativo. En general, un elemento positivo hace referencia a cualquier elemento cuya alteración o eliminación de la secuencia candidata implicaría una disminución en la expresión de la proteína terapéutica, o cuya creación supondría una disminución en la expresión de una proteína terapéutica. Por ejemplo, un elemento positivo puede incluir un potenciador, un promotor, un elemento promotor corriente abajo, un sitio de unión a ADN para un regulador positivo (p.ej., un activador de la transcripción), o una secuencia responsable de conferir o eliminar estructura secundaria o terciaria de ARNm. Un elemento negativo se refiere a cualquier elemento cuya alteración o eliminación de la secuencia candidata conllevaría un aumento en la expresión de la proteína terapéutica, o cuya creación implicaría una disminución en la expresión de la proteína terapéutica. Un elemento negativo incluye un silenciador, un sitio de unión a ADN para un regulador negativo (p.ej., un represor de la transcripción), un sitio de pausa de la transcripción o una secuencia que es responsable de conferir o eliminar estructura secundaria o terciaria del ARNm. En general, aparece con más frecuencia un elemento negativo que uno positivo. Por tanto, es más probable que cualquier cambio de uso de codones que conlleve un aumento en la expresión de proteínas se deba a la destrucción de un elemento negativo que a la creación de un elemento positivo. Además, es más probable que la alteración de la secuencia candidata destruya un elemento positivo que cree un elemento positivo. En un modo de realización, se elige y modifica una secuencia candidata para que aumente la producción de una proteína terapéutica. La secuencia candidata puede modificarse, p.ej., alterando los codones de modo secuencial o alterando los codones al azar en la secuencia candidata. A continuación, se evalúa una secuencia candidata modificada determinando el nivel de expresión de la proteína terapéutica resultante o evaluando otro parámetro, p.ej., un parámetro relacionado con el nivel de expresión. Se selecciona una secuencia candidata que produce un nivel aumentado de una proteína terapéutica en comparación con una secuencia candidata no alterada.

**[0068]** En otro enfoque, se puede modificar y analizar un codón o grupo de codones, p.ej., sin hacer referencia a la estructura del mensaje o de proteína. Como alternativa, se pueden elegir uno o más codones en una propiedad a nivel de mensaje, p.ej., la posición en una región, p.ej., de alto o bajo contenido de GC o AU predeterminado, la posición en una región que tiene una estructura como un potenciador o silenciador, la posición en una región que puede modificarse para introducir una estructura como un potenciador o silenciador, la posición en una región que tiene, o que se prevé que tenga, estructura secundaria o terciaria, p.ej., apareamiento intracadena, apareamiento intercadena, la posición en una región que carece de, o que se prevé que carezca de, estructura secundaria o terciaria, p.ej., apareamiento intracadena o intercadena. Se selecciona una región modificada concreta si produce el resultado deseado.

**[0069]** Los métodos que generan sistemáticamente secuencias candidatas son útiles. Por ejemplo, se puede reemplazar uno o un grupo, p.ej., un bloque contiguo de codones, en varias posiciones de una secuencia de ácido nucleico sintético con codones comunes (o con codones no comunes, si, por ejemplo, la secuencia de inicio ha sido optimizada) y evaluar la secuencia resultante. Se pueden generar candidatos a través de la optimización (o suboptimización) de una "ventana" determinada de codones en la secuencia para generar un primer candidato y, a continuación moviendo la ventana a una nueva posición en la secuencia y optimizando (o suboptimizando) los codones en la nueva posición en la ventana para disponer un segundo candidato. Se pueden evaluar los candidatos mediante la determinación del nivel de expresión que ofrecen o mediante la evaluación de otro parámetro, p.ej., un parámetro relacionado con el nivel de expresión. Se pueden evaluar algunos parámetros con la inspección o con la ayuda de un ordenador, p.ej., la posesión o falta de alto o bajo contenido de AU o GC; un elemento de secuencia como un potenciador o silenciador; estructura secundaria o terciaria, p.ej., apareamiento intracadena o intercadena

**[0070]** Por tanto, los mensajes híbridos, es decir, los mensajes que tienen una región que se optimiza y una región que no se optimiza, pueden evaluarse para determinar si tienen una propiedad deseada. Puede efectuarse la evaluación mediante, p.ej., síntesis del mensaje o mensajes candidatos y determinación de una propiedad como su nivel de expresión. Esta determinación puede llevarse a cabo en un sistema libre de células o en un sistema basado en células. La generación y ensayo de uno o más candidatos también puede realizarse con métodos computacionales, p.ej., en un ordenador. Por ejemplo, se puede utilizar un programa informático para generar una serie de mensajes candidatos y estos mensajes analizados por un programa informático que predice la existencia de elementos de estructura primaria o estructura secundaria o terciaria.

**[0071]** Se puede generar un mensaje candidato mediante la división de una región en subregiones y la optimización de cada subregión. A continuación se combina una subregión optimizada con una subregión no optimizada para producir un candidato. Por ejemplo, se divide una región en tres subregiones, a, b y c, a continuación se optimizan las tres para proporcionar subregiones optimizadas a', b' y c'. A continuación, las subregiones a', b' o c' pueden combinarse con una o más subregiones no optimizadas, p.ej., a, b y c. Por ejemplo, se podría formar y analizar ab'c. Pueden generarse diferentes combinaciones de subregiones optimizadas y no optimizadas. A través de la evaluación de una serie de dichas secuencias candidatas híbridas, es posible analizar el efecto de la modificación de diferentes subregiones y, p.ej., definir la versión concreta de cada subregión que más contribuye a la propiedad deseada. Un candidato preferido puede incluir las versiones de cada subregión que tuvieron más éxito en esta serie de experimentos.

10 Un algoritmo para crear una secuencia candidata optimizada es el siguiente

1. Proporcionar una secuencia de mensaje (un mensaje completo o una parte del mismo). Ir al paso 2.
2. Generar una nueva secuencia candidata mediante la modificación del uso de codones de una secuencia candidata empleando la secuencia candidata más prometedora previamente identificada, o mediante la combinación de regiones de dos o más candidatas previamente identificadas para producir un nuevo híbrido. Ir al paso 3.
3. Evaluar la secuencia candidata y determinar si tiene una propiedad predeterminada. Si la candidata tiene la propiedad predeterminada, entonces ir al paso 4, de lo contrario volver al paso 2.
4. Utilizar la secuencia candidata como un mensaje optimizado.

**[0072]** Los métodos pueden incluir primero la optimización de una secuencia de ácido nucleico sintético de mamífero que codifica una proteína de interés o una parte de la misma, p.ej., Factor VIII humano, Factor IX humano,  $\alpha$ -galactosidasa humana, etc. La secuencia de ácido nucleico sintético puede optimizarse de modo que el 94%, 95%, 96%, 97%, 98%, 99% de los codones, o todos, del ADN sintético se sustituyen con codones comunes. El siguiente paso consiste en determinar la cantidad de proteína producida como resultado de la optimización del mensaje en comparación con la cantidad de proteína producida utilizando la secuencia de tipo salvaje. En los casos en los que la cantidad de proteína producida no es del nivel deseado o esperado, puede ser útil reemplazar uno o más codones comunes de la región que codifica la proteína con un codón menos común o un codón no común. Un mensaje optimizado de mamífero que se rediseña para que los codones comunes se reemplacen por codones de mamíferos menos comunes o no comunes o para que los codones comunes de otras especies eucariotas produzcan la sustitución de al menos el 1%, 5%, 10%, 20% o más de los codones comunes. El rediseño del mensaje optimizado puede llevarse a cabo, por ejemplo, de modo sistemático reemplazando un codón común único con un codón menos común o no común. Como alternativa, un bloque de 2, 4, 6, 10, 20, 40 o más codones puede sustituirse con un codón menos común o no común. El nivel de proteína producido por estos mensajes "optimizados rediseñados" determina qué mensaje optimizado rediseñado se selecciona.

**[0073]** Otro enfoque para la optimización de un mensaje para la expresión de proteína aumentada incluye la alteración del contenido de nucleótido específico de una secuencia de ácido nucleico sintético optimizado. La secuencia de ácido nucleico sintético puede alterarse mediante el aumento o disminución del contenido específico de nucleótido(s), p.ej., el contenido de G, C, A, T, GC o AT de la secuencia. El aumento o la disminución del contenido de nucleótidos específico de una secuencia de nucleótidos sintéticos pueden realizarse sustituyendo el nucleótido de interés por otro nucleótido. Por ejemplo, una secuencia que tiene un gran número de codones que tienen un alto contenido de GC, p.ej., glicina (GGC), puede sustituirse con codones que tiene menor contenido rico en GC, p.ej., glicina (GGT) o un codón rico en AT. Del mismo modo, se puede sustituir una secuencia que tiene un gran número de codones que tienen un alto contenido de AT con codones que tienen un contenido rico en AT menor, p.ej., un codón rico en GC. Puede alterarse cualquier región, o la totalidad, de una secuencia de ácido nucleico sintético de este modo, p.ej., la 5'UTR (p.ej., la región de codificación próxima al promotor), la región de codificación, la secuencia del intrón o la 3'UTR. Preferiblemente, las sustituciones de nucleótidos en la región de codificación no producen una alteración de la secuencia de aminoácidos del producto expresado. Preferentemente, el contenido de nucleótidos, p.ej., el contenido de GC o AT, de una secuencia se aumenta o reduce en un 10%, 20%, 30%, 40% o en un porcentaje mayor.

**[0074]** La secuencia de ácido nucleico sintético puede codificar una proteína de mamífero, p.ej., una humana. La proteína puede ser, p.ej., una que es endógenamente una proteína humana o una proteína modificada genéticamente. Las proteínas modificadas genéticamente incluyen proteínas que difieren de la proteína nativa en uno o más residuos de aminoácidos. Entre los ejemplos de dichas proteínas se incluyen, p.ej., fragmentos o truncamientos internos, deleciones, proteínas de fusión y proteínas que tiene una o más sustituciones de aminoácidos.

**[0075]** Una secuencia que codifica la proteína que tiene uno o más intrones. La secuencia de ácido nucleico sintético puede incluir intrones, tal y como se encuentran en la secuencia no optimizada o puede incluir intrones de un gen no relacionado. En otros modos de realización, pueden modificarse las secuencias de intrones. Por ejemplo, pueden eliminarse todos o partes de uno o más intrones presentes en el gen o pueden añadirse intrones que no se encuentran

en la secuencia. En modos de realización preferidos, uno o más intrones completos presentes en el gen no aparecen en el ácido nucleico sintético. En otro modo de realización, se sustituye todo o parte del intrón presente en un gen es sustituido por otra secuencia, p.ej., una secuencia de intrones procedente de otra proteína.

5 **[0076]** La secuencia de ácido nucleico sintético puede codificar: cualquier proteína que incluya un factor sanguíneo, p.ej., factor V de coagulación sanguínea, factor VII de coagulación sanguínea, factor VIII de coagulación sanguínea, factor IX de coagulación sanguínea, factor X de coagulación sanguínea o factor XIII de coagulación sanguínea, una interleucina, p.ej., interleucina 1, interleucina 2, interleucina 3, interleucina 6, interleucina 11 o interleucina 12; eritropoyetina; calcitonina; hormona del crecimiento; insulina; insulínotropina; factores de crecimiento de tipo insulina; hormona paratiroidea;  $\beta$ -interferón;  $\gamma$ -interferón; factores de crecimiento nervioso; Fish $\beta$ ; factor de necrosis tumoral; glucagón; factor-2 de crecimiento de los huesos, factor-7 de crecimiento de los huesos TSH- $\beta$ ; factores estimulantes de colonias(CSF) de granulocito; CSF de macrófago; CSF de granulocito/macrófago; inmunoglobulinas; anticuerpos catalíticos; proteína-cinasa C; glucocerebrosidasa; superóxido dismutasa; activador tisular del plasminógeno; urocinasa; antitrombina III; ADNasa;  $\alpha$ -galactosidasa; tirosina hidroxilasa; apolipoproteína E; apolipoproteína A-I; globinas; receptor de lipoproteínas de baja densidad; receptor de IL-2; antagonistas de IL-2; alfa-1-antitripsina; modificadores de la respuesta inmune; CD4 soluble; una proteína expresada en condiciones de la enfermedad; y proteínas codificadas por virus (incluido un retrovirus) que se expresan de células de mamífero después de la infección.

10 **[0077]** En modos de realización preferentes, la secuencia de ácido nucleico sintético puede expresar su proteína, p.ej., una eucariota, p.ej., una proteína de mamífero, a un nivel que es al menos el 110%, 150%, 200%, 500%, 1000%, 5000% o incluso 10.000% del nivel expresado por la secuencia de ácido nucleico que no ha sido optimizada. Esta comparación puede hacerse, p.ej., en un sistema de cultivo de células de mamífero *in vitro* en el que las secuencias no optimizadas y las secuencias optimizadas se expresan en las mismas condiciones (por ejemplo, el mismo tipo de célula, las mismas condiciones de cultivo, el mismo vector de expresión).

15 **[0078]** Se conocen en la técnica sistemas de cultivo celular adecuados para medir la expresión de la secuencia de ácido nucleico sintético y las correspondientes secuencias de ácido nucleico no optimizados (p.ej., los vectores pBS phagemic, Stratagene, La Jolla, CA) y se describen en, por ejemplo, los libros de referencia de biología molecular. Los vectores adecuados para la expresión de secuencias de ácido nucleico sintético y no optimizado que codifican la proteína de interés se describen a continuación y en los libros de referencia estándar descritos más abajo. La expresión puede medirse empleando un anticuerpo específico para la proteína de interés (p.ej., ELISA). Los expertos en la materia conocen estos anticuerpos y técnicas de medición

20 **[0079]** En un modo de realización preferido, la proteína es una proteína humana. En modos de realización más preferidos, la proteína es Factor VIII humano y la proteína es un Factor VIII humano con el dominio B suprimido (en adelante, BDD hFVIII del inglés). En otro modo de realización preferido, la proteína es Factor VIII humano con el dominio B suprimido con una secuencia que incluye un sitio de reconocimiento para proteasa intracelular de la clase PACE/furina, como sitio X-ARG-X-X-ARG, una unión a péptido corto, p.ej., una unión a dos péptidos, p.ej., una unión a péptido de ácido glutámico-leucina (LE) o una unión a tres o cuatro péptidos, insertados en la unión de la cadena pesada y ligera (véase la Figura 1).

25 **[0080]** Una gran parte de los codones en los mensajes humanos que codifican Factor VIII y Factor IX son codones no comunes o menos comunes. La sustitución de al menos el 98% de estos codones con codones comunes producirá secuencias de ácido nucleico capaces de una expresión a un nivel más alto en un cultivo celular. Preferentemente, todos los codones se sustituyen con codones comunes y esta sustitución produce al menos un aumento de 2 a 5 veces, más preferentemente de 10 veces y lo más preferentemente de 20 veces en la expresión cuando se compara con una expresión de la secuencia nativa correspondiente en el mismo sistema de expresión.

30 **[0081]** Las secuencias de ácido nucleico sintético de la invención pueden introducirse en las células de un organismo vivo. Las secuencias pueden introducirse directamente, p.ej., a través de recombinación homóloga o a través de un vector. Por ejemplo, los vectores o constructos de ADN pueden utilizarse para introducir una secuencia de ácido nucleico sintético en células de un organismo vivo para terapia génica. Véase, p.ej., la patente de EE. UU. N° 5.460.959 y las solicitudes de EE. UU. pendientes de resolución USSN 08/334.797; USSN 08/231.439; USSN 08/334.455 y USSN 08/928.881.

#### Células infectadas o transfectadas

35 **[0082]** Las células primarias y secundarias para ser transfectadas o infectadas pueden obtenerse a partir de una variedad de tejidos e incluyen tipos de células que pueden mantenerse y propagarse en cultivo. Por ejemplo, las células primarias y secundarias que se pueden transfectar o infectar por el presente procedimiento comprenden fibroblastos, queratinocitos, células epiteliales (p.ej., células epiteliales mamarias, células epiteliales intestinales), células endoteliales, células gliales, células neurales, una célula que comprende un elemento formado de la sangre (p.ej., linfocitos, células de la médula ósea), células del músculo, y precursores de estos tipos de células somáticas. Las células primarias se obtienen preferentemente de sujetos a los que se administran células primarias o secundarias

transfectadas o infectadas. No obstante, las células primarias se pueden obtener de un donante (distinto del receptor) de la misma especie o de otra especie (p.ej., ratón, rata, conejo, gato, perro, vaca, pájaro, oveja, cabra o caballo).

**[0083]** Las células primarias o secundarias de origen vertebrado, en concreto de un mamífero, pueden transfectarse o infectar con ADN sintético exógeno que codifica una proteína terapéutica y producen una proteína terapéutica codificada de modo estable y reproductivo, tanto *in vitro* como *in vivo*, durante prolongados periodos de tiempo. Además, las células primarias y secundarias transfectadas o infectadas pueden expresar el producto codificado *in vivo* en niveles fisiológicamente relevantes, las células pueden recuperarse después de la implantación, una vez vuelta a plantar, para cultivar y mostrar sus propiedades antes de la implantación.

**[0084]** Las células primarias o secundarias transfectadas o infectadas también pueden incluir ADN que codifica un marcador de selección que confiere un fenotipo de selección sobre ellos, facilitando su identificación y aislamiento. En el presente documento se describen métodos para producir células primarias y secundarias transfectadas que expresan de modo estable ADN sintético exógeno, cepas de células clónicas, y cepas de células heterogéneas de estas células transfectadas, métodos para producir las cepas de células clónicas y heterogéneas, y métodos para tratar o prevenir una enfermedad anormal o no deseable a través del uso de poblaciones de células primarias o secundarias transfectadas. Las células primarias y secundarias que se pueden transfectar o infectar por el presente procedimiento comprenden fibroblastos, queratinocitos, células epiteliales (p.ej., células epiteliales mamarias, células epiteliales intestinales), células endoteliales, células gliales, células neurales, una célula que comprende un elemento formado de la sangre (p.ej., un linfocito, una célula de la médula ósea), células del músculo, y precursores de estos tipos de células somáticas. Las células primarias se obtienen preferentemente de sujetos a los que se administran células primarias o secundarias transfectadas o infectadas. No obstante, las células primarias se pueden obtener de un donante (distinto del destinatario) de la misma especie o de otra especie (p.ej., ratón, rata, conejo, gato, perro, vaca, pájaro, oveja, cabra o caballo). También pueden utilizarse células transformadas o inmortalizadas, p.ej., una célula de melanoma de Bowes (núm. de registro ATCC CRL 9607), una célula Daudi (núm. de depósito ATCC CCL 213), una célula HeLa y un derivado de una célula HeLa (núms. de depósito ATCC CCL 2. CCL 2.1 y CCL 2.2), una célula HL-60 (núm. de depósito ATCC CCL 240), una célula HT-1080 (núm. de depósito ATCC CCL 121), una célula Jurkat (núm. de depósito ATCC TIB 152), una célula de carcinoma KB (núm. de depósito ATCC CCL 17), una célula de leucemia K-562 (núm. de depósito ATCC CCL 243), una célula de cáncer de mama MCF-7 (núm. de depósito ATCC BTH 22), una célula MOLT-4 (núm. de depósito ATCC 1582), una célula Namalwa (núm. de depósito ATCC CRL 1432), una célula Raji (núm. de depósito ATCC CCL 86), una célula RPMI 8226 (núm. de depósito ATCC CCL 155), una célula U-937 (núm. de depósito ATCC CRL 1593), una célula 2R4 de la sublínea WI-38VA13 (núm. de depósito ATCC CLL 75.1), una célula CCRF-CEM (núm. de depósito ATCC CCL 119) y una célula de carcinoma de ovario 2780AD (Van Der Blick *et al.*, *Cancer Res.* 48: 5927-5932, 1988), así como células de heterohibridoma producidas por fusión de células humanas y células de otras especies. En otro modo de realización, la línea celular inmortalizada puede ser una línea celular que no sea una línea celular humana, p.ej., una línea celular CHO o una línea celular COS. En un modo de realización preferido, la célula es una célula no transformada. En diversos modos de realización preferidos, la célula es una célula de mamífero, p.ej., una célula de mamífero primaria o secundaria, p.ej., un fibroblasto, una célula madre hematopoyética, un mioblasto, un queratinocito, una célula epitelial, una célula endotelial, una célula glial, una célula neural, una célula que comprende un elemento formado de la sangre, una célula muscular y precursores de estas células somáticas. En el modo de realización más preferido, la célula es un fibroblasto humano secundario.

**[0085]** Como alternativa, el ADN puede administrarse en cualquiera de los tipos de células anteriormente comentadas a través de una infección por vector viral. Los virus conocidos por ser útiles para la transferencia génica incluyen adenovirus, un virus adeno-asociado, un virus del herpes, un virus de las paperas, un poliovirus, un retrovirus, un virus Sindbis, y un virus vaccinia como un virus de la viruela del canario. El uso de vectores virales es muy conocido en la técnica, véase, p.ej., Robbins and Ghizzani, *Mol. Med. Today* 1:410-417, 1995. Se le denomina "célula infectada" a una célula que tiene un ADN exógeno introducido en ella a través de un vector viral.

**[0086]** La descripción también incluye la manipulación genética de una célula que normalmente produce una proteína terapéutica. En este caso, la célula se manipula de modo que la secuencia endógena que codifica la proteína terapéutica se sustituye por una secuencia de codificación optimizada, p.ej., por recombinación homóloga.

#### ADN sintético exógeno

**[0087]** El ADN sintético exógeno incorporado en las células primarias o secundarias por el presente procedimiento puede ser un ADN sintético que codifica una proteína, o una parte de la misma, útil para tratar o prevenir una enfermedad existente.

**[0088]** El ADN sintético incorporado en células primarias o secundarias puede ser un gen completo que codifica una proteína deseada completa o una parte de un gen que codifica, por ejemplo, la parte o partes funcionales o activas de la proteína. La proteína puede ser, por ejemplo, una hormona, una citocina, un antígeno, un anticuerpo, una enzima, un factor de coagulación, p.ej., Factor VIII o Factor XI, una proteína transportadora, un receptor, una proteína reguladora, una proteína estructural o una proteína que no ocurre de modo natural. El ADN puede obtenerse empleando técnicas de modificación genética o procedimientos de síntesis. El ADN introducido en las células primarias o secundarias puede

codificar una o más proteínas terapéuticas. Después de la introducción en las células primarias o secundarias, el ADN sintético exógeno se incorpora de modo estable en el genoma de la célula receptora (junto con las secuencias adicionales presentes en el constructo de ADN utilizado), a partir de la que se expresa u otras funciones. Como alternativa, el ADN sintético exógeno puede existir en forma episomal dentro de las células primarias o secundarias.

5 Marcadores de selección

[0089] En las células primarias o secundarias se puede incorporar una serie de marcadores de selección. Por ejemplo, puede utilizarse un marcador de selección que dispone un fenotipo de selección como la resistencia al fármaco, la auxotrofia nutricional, la resistencia a un agente citotóxico o la expresión de una proteína de la superficie. Entre los genes marcadores de selección que pueden utilizarse se incluyen: neo, gpt, dhfr, ada, pac (puromicina), hyg y hisD. El fenotipo de selección proporcionado hace posible identificar y aislar las células receptoras primarias o secundarias.

10 Constructos de ADN

[0090] Se pueden utilizar los constructos de ADN, que comprenden ADN sintético exógeno y, opcionalmente, el ADN que codifica un marcador de selección, junto con secuencias adicionales necesarias para la expresión del ADN sintético exógeno en las células primarias y secundarias del destinatario, con el fin de transferir células primarias o secundarias en las que se ha de producir el producto codificado. Como alternativa, se pueden utilizar vectores infecciosos con este objetivo como vectores retrovíricos, de herpes, de lentivirus, de adenovirus, de adenovirus-asociado, de paperas y de poliovirus.

[0091] Se puede utilizar un constructo de ADN que incluye el ADN sintético y secuencias adicionales, como las secuencias necesarias para la expresión de ADN sintético exógeno. Un constructo de ADN puede comprender ADN que codifica un marcador de selección, junto con secuencias adicionales, como un promotor, un sitio de poliadenilación y las uniones de corte y empalme, puede utilizarse para producir un fenotipo de selección tras la introducción en células primarias o secundarias. Los dos constructos de ADN se introducen en las células primarias o secundarias, empleando los procedimientos descritos en el presente documento. Como alternativa, se puede utilizar un constructo de ADN que comprende el ADN sintético exógeno, un gen marcador de selección y secuencias adicionales (p.ej., las necesarias para la expresión del ADN sintético exógeno y para la expresión del gen marcador de selección).

20 Transfección de células primarias o secundarias y producción de cepas de células clónicas o heterogéneas

[0092] Se obtiene el tejido vertebrado utilizando procedimientos conocidos como biopsia con sacabocados u otros procedimientos quirúrgicos de obtención de un tejido procedente del tipo celular primario en cuestión. Por ejemplo, la biopsia con sacabocados se utiliza para obtener piel como fuente de fibroblastos o queratinocitos. Se obtiene una mezcla de células primarias del tejido, utilizando métodos conocidos, como la digestión enzimática. Si se utiliza la digestión enzimática, se pueden utilizar enzimas como colagenasa, hialuronidasa, dispasa, pronasa, tripsina, elastasa y quimiotripsina.

[0093] La mezcla de células primarias resultantes se puede transfectar directamente o se puede cultivar en primer lugar, se extrae de la placa de cultivo y se vuelve a suspender antes de que se realice la transfección. Las células primarias o las células secundarias se combinan con ADN sintético exógeno para estar integradas establemente en sus genomas y, opcionalmente, con el ADN que codifica un marcador de selección, y se tratan para realizar la transfección. El ADN sintético exógeno y el ADN que codifica el marcador de selección están cada uno en un constructo separado o en un constructo individual y se utiliza una cantidad apropiada de ADN para asegurar que por lo menos se produce una célula transfectada de forma estable que contiene y que expresa correctamente el ADN exógeno. En general, se utilizan de 0,1 a 500 µg de ADN.

[0094] La transfección de las células primarias o secundarias se efectúa mediante electroporación. Ésta se realiza a un voltaje y una capacitancia apropiados (y constante de tiempo) para producir la entrada del constructo o constructos de ADN en las células primarias o secundarias. La electroporación se puede realizar a lo largo de una amplia gama de valores de voltajes (p.ej., 50 a 2000 voltios) y capacitancia (p.ej., 60-300 µFaradios). En general, se utiliza el ADN total, de aproximadamente 0,1 a 500 µg.

[0095] Las células primarias o secundarias pueden transfectarse utilizando la microinyección. Como alternativa, se pueden utilizar métodos conocidos para transfectar células como precipitación con fosfato de calcio, precipitación con fosfato de calcio modificada y precipitación con polibreno, fusión del liposoma y administración del gen mediada por el receptor. Se aísla, cultiva y subcultiva una célula transfectada establemente, en condiciones de cultivo y durante suficiente tiempo, para propagar las células secundarias transfectadas establemente y producir una cepa de células clónica de células secundarias transfectadas. Como alternativa, se cultiva y subcultiva más de una célula transfectada, lo que produce una cepa de célula heterogénea.

[0096] Las células primarias o secundarias transfectadas experimentan un número suficiente de duplicaciones para producir una cepa de células clónicas o una cepa de células heterogéneas de suficiente tamaño para producir la

proteína terapéutica a un sujeto en cantidades eficaces. En general, por ejemplo, se biopsia 0,1 cm<sup>2</sup> de piel y se supone que contienen 100.000 células; se utiliza una célula para producir una cepa de células clónicas y experimenta aproximadamente 27 duplicaciones para producir 100 millones de células secundarias transfectadas. Si se ha de producir una cepa de células heterogénea a partir de una población original de transfección de aproximadamente 100.000 células, se necesitan solamente 10 duplicaciones para producir 100 millones de células transfectadas.

**[0097]** El número de células necesarias en una cepa de células clónicas o heterogéneas transfectadas es variable y depende de una serie de factores, que comprenden pero no se limitan a, la utilización de células transfectadas, el nivel funcional del ADN exógeno en las células transfectadas, el sitio de implantación de las células transfectadas (por ejemplo, el número de células que se pueden utilizar está limitado por el sitio anatómico de implantación), y la edad, el área superficial y el estado clínico del paciente. Al analizar estos factores en perspectiva, para administrar niveles terapéuticos de hormona humana de crecimiento en un paciente de 10 kg sano, pero con déficit aislado de hormona del crecimiento, serían necesarios aproximadamente de uno a quinientos millones de fibroblastos transfectados (el volumen de estas células es aproximadamente el de la extremidad del pulgar del paciente).

#### Expresión episomal del ADN sintético exógeno

**[0098]** Las secuencias de ADN que están presentes dentro de la célula, pero no se integran en el genoma, son denominadas episomas. Los episomas recombinantes pueden ser útiles en al menos tres casos: 1) si un tipo de célula determinado es incapaz de integrar establemente el ADN sintético exógeno; 2) si a un tipo de célula determinado le afecta desfavorablemente la integración de ADN sintético; y 3) si un tipo determinado de célula es capaz de realizar una función terapéutica mejorada con un ADN episomal en vez de un ADN integrado.

**[0099]** Con el uso de la transfección y el cultivo como se describe en el presente documento, puede introducirse ADN sintético exógeno en forma de episomas dentro de células primarias y secundarias de vertebrados. Los plásmidos pueden ser convertidos en dicho episoma mediante la adición de secuencias de ADN para el origen de replicación del virus de Epstein-Barr y antígeno nuclear (Yates, J.L. *Nature* 319:780-7883 (1985)). Como alternativa, se pueden introducir en el constructo secuencias de vertebrados que se repliquen autónomamente (Weidle, U.H. *Gene* 73(2):427-437 (1988)). Estas y otras secuencias derivadas de modo episomal se pueden incluir también en entes artificiales de ADN sin marcadores de selección, como pXGH5 (Selden et al., *Mol Cell Biol.* 6:3173-3179, 1986). El ADN exógeno sintético episomal se introduce a continuación en células primarias o secundarias de un vertebrado, como se describe en la presente solicitud (si se incluye un marcador de selección en el episoma, se utiliza un agente selectivo para tratar las células transfectadas).

#### Implantación de cepas de células clónicas o cepas de células heterogéneas de células secundarias transfectadas

**[0100]** Las células transfectadas o infectadas producidas como se describe anteriormente pueden introducirse en un sujeto al que se le ha de suministrar la proteína terapéutica, empleando métodos conocidos. A continuación, la cepa de células clónicas o cepa de células heterogéneas es introducida en un sujeto, usando métodos conocidos, a través de varias vías de administración y en diversos sitios (p.ej., implantación subcapsular renal, subcutánea, en el sistema nervioso central (inclusive intratecal), intravascular, intrahepática, intraesplácnica, intraperitoneal (inclusive intraomental o intramuscular). En un modo de realización preferido, se introduce la cepa de células clónicas o cepa de células heterogéneas en el omento. El omento es una estructura membranosa que contiene una lámina de grasa. En general, el omento es un pliegue del peritoneo que se extiende del estómago a órganos abdominales adyacentes. El omento mayor está unido al borde inferior del estómago y cuelga enfrente de los intestinos. El otro borde está conectado al colon transversal. El omento menor está unido al borde superior del estómago y se extiende a la capa interior del hígado. Las células se pueden introducir en cualquier parte del omento por implantación quirúrgica, laparoscopia, o inyección directa, p.ej., a través de ultrasonido o aguja guiada por TC: Una vez implantadas en el sujeto, las células producen el producto terapéutico codificado por el ADN sintético exógeno o se ven afectadas por el propio ADN sintético exógeno. Por ejemplo, un sujeto al que se le ha diagnosticado hemofilia A, un trastorno hemorrágico causado por déficit de Factor VIII, una proteína que se encuentra normalmente en la sangre, es un candidato para un tratamiento de terapia génica. En otro ejemplo, un sujeto al que se le ha diagnosticado hemofilia B, un trastorno hemorrágico causado por déficit de Factor IX, una proteína que se encuentra normalmente en la sangre, es un candidato para un tratamiento de terapia génica. Se lleva a cabo una pequeña biopsia de piel al paciente. Este es un procedimiento sencillo que se puede realizar a un paciente no hospitalizado. La muestra de piel, que tiene aproximadamente el tamaño de una cabeza de cerilla, se obtiene, por ejemplo, de por debajo del brazo y se necesita en torno a un minuto para sacarla. Se elabora la muestra, lo que produce el aislamiento de las células del paciente y se modifica genéticamente para producir el Factor IX o el Factor VIII que falta. Basándose en la edad, peso y el estado clínico del paciente, se hace crecer el número de células necesario en un cultivo a gran escala. El proceso completo requiere normalmente 4-6 semanas y, al final de este periodo de tiempo, se introducen en el sujeto, de nuevo un paciente no hospitalizado, aproximadamente 100-500 millones de células genéticamente modificadas (p.ej., inyectándolas de nuevo bajo la piel del paciente). Entonces, el paciente es capaz de producir su propio Factor IX o Factor VIII y deja de padecer hemofilia.

**[0101]** Un enfoque similar se puede utilizar para tratar otros estados o enfermedades. Por ejemplo, se puede tratar la estatura baja con la administración de la hormona humana del crecimiento a un sujeto mediante la implantación de

células primarias o secundarias que expresan la hormona humana del crecimiento; la anemia se puede tratar administrando eritropoyetina (EPO) a un sujeto mediante la implantación de células primarias o secundarias que expresan eritropoyetina (EPO); o la diabetes puede tratarse mediante la administración del péptido similar al glucógeno tipo 1 (GLP-1) a un sujeto mediante la implantación de células primarias o secundarias que expresan GLP-1. Una enfermedad lisosómica (LSD) puede tratarse con este enfoque. LSD representa a un grupo de al menos 41 enfermedades genéticas distintas, cada una de ellas presenta el déficit de una proteína concreta que está implicada en la biogénesis lisosomal. Una LSD concreta puede tratarse administrando una enzima lisosomal a un sujeto mediante la implantación de células primarias o secundarias que expresan la enzima lisosomal, por ejemplo, se puede tratar la enfermedad de Fabry administrando  $\alpha$ -galactosidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\alpha$ -galactosidasa; la enfermedad de Gaucher se puede tratar administrando  $\beta$ -glucoceramidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\beta$ -glucoceramidasa; la MPS (mucopolisacaridosis) tipo 1 (síndrome de Hurler-Scheie) se puede tratar administrando  $\alpha$ -iduronidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\alpha$ -iduronidasa; la MPS tipo II (síndrome de Hunter) se puede tratar administrando  $\alpha$ -L-iduronidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\alpha$ -L-iduronidasa; la MPS tipo III-A (síndrome de Sanfilippo A) se puede tratar administrando glucosamina-N-sulfatasa a un sujeto mediante la implantación de células primarias o secundarias que expresan glucosamina-N-sulfatasa; la MPS tipo III-B (síndrome de Sanfilippo B) se puede tratar administrando alfa-N-acetilglucosaminidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan alfa-N-acetilglucosaminidasa; la MPS tipo III-C (síndrome de Sanfilippo C) se puede tratar administrando acetil-coenzima A: $\alpha$ -glucosmainide-N-acetiltransferasa a un sujeto mediante la implantación de células primarias o secundarias que expresan acetil-coenzima A: $\alpha$ -glucosmainide-N-acetiltransferasa; la MPS tipo III-D (síndrome de Sanfilippo D) se puede tratar administrando N-acetilglucosamina-6-sulfatasa a un sujeto mediante la implantación de células primarias o secundarias que expresan N-acetilglucosamina-6-sulfatasa; la MPS tipo IV-A (síndrome de Morquio A) se puede tratar administrando N-acetilglucosamina-6-sulfatasa a un sujeto mediante la implantación de células primarias o secundarias que expresan N-acetilglucosamina-6-sulfatasa; la MPS tipo IV-B (síndrome de Morquio B) se puede tratar administrando  $\beta$ -galactosidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\beta$ -galactosidasa; la MPS tipo IV (síndrome de Maroteaux-Larry) se puede tratar administrando N-acetilgalactosamina-6-sulfatasa a un sujeto mediante la implantación de células primarias o secundarias que expresan N-acetilgalactosamina-6-sulfatasa; la MPS tipo VII (síndrome de Sly) se puede tratar administrando  $\beta$ -glucuronidasa a un sujeto mediante la implantación de células primarias o secundarias que expresan  $\beta$ -glucuronidasa.

**[0102]** En general, las células utilizadas serán células genéticamente modificadas, específicas del paciente. Sin embargo, es posible obtener células de otro sujeto de la misma especie o de una especie diferente. El uso de esas células puede requerir la administración de un agente inmunosupresor, la alteración de antígenos de histocompatibilidad o el uso de un dispositivo de barrera para evitar el rechazo de las células implantadas. Para muchas enfermedades, será un tratamiento de una sola vez y otras necesitarán múltiples tratamientos de terapia génica.

#### Usos de cepas de células y células primarias y secundarias transfectadas o infectadas

**[0103]** Las cepas de células o células primarias o secundarias transfectadas o infectadas tienen amplia aplicabilidad como vehículo o sistema de administración para proteínas terapéuticas, como enzimas, hormonas, citocinas, antígenos, anticuerpos, factores de coagulación, ARN antisentido, proteínas reguladores, proteínas de transcripción, receptores, proteínas estructurales, nuevas proteínas (no optimizadas) y productos de ácidos nucleicos, así como ADN modificado genéticamente. Por ejemplo, se pueden utilizar células primarias o secundarias transfectadas para suministrar una proteína terapéutica que incluye, pero sin carácter limitativo, Factor VIII, Factor IX, eritropoyetina, alfa-1 antitripsina, calcitonina, glucocerebrosidasa, hormona del crecimiento, lipoproteína de baja densidad (LDL), receptor de IL-2 y sus antagonistas, insulina, globina, inmunoglobulinas, anticuerpos catalíticos, interleucinas, factores de crecimiento similares a la insulina, superóxido-dismutasa, modificadores de respondedores inmunes, interferón y hormona de paratiroides, factores de crecimiento nervioso, activadores del plasminógeno de tejidos y factores estimulantes de colonias. Como alternativa, se pueden utilizar células primarias y secundarias transfectadas para inmunizar a un sujeto (es decir, como una vacuna).

**[0104]** La gran variedad de usos de las cepas de células de la presente invención puede resumirse tal vez de la manera más conveniente como se muestra a continuación. Las cepas de células se pueden utilizar para administrar los siguientes productos terapéuticos.

1. una proteína secretada con efectos predominantemente sistemáticos;
2. una proteína secretada con efectos predominantemente locales;
3. una proteína de membrana que confiere nueva capacidad de respuesta células nueva y mejorada;
4. una proteína de membrana que facilita la eliminación de un producto tóxico;
5. una proteína de membrana que marca o dirige una célula;

6. una proteína intracelular;
7. una proteína intracelular que afecta directamente a la expresión de genes; y
8. una proteína intracelular con efectos autocrinos.

5 [0105] Las células primarias o secundarias transfectadas o infectadas se pueden utilizar para administrar proteínas terapéuticas (p.ej., hormonas, enzimas, factores de coagulación) que actualmente se administran por vía intravenosa, intramuscular o subcutánea, lo que requiere la cooperación del paciente y, con frecuencia, la participación de un equipo médico. Cuando se utilizan células primarias o secundarias transfectadas o infectadas, no es necesaria la amplia purificación del polipéptido antes de que sea administrado a un sujeto, como si es generalmente necesario con un polipéptido aislado. Además, las células primarias o secundarias transfectadas o infectadas de la presente invención producen la proteína terapéutica como se produciría normalmente.

10 [0106] Una ventaja del uso de células primarias y secundarias transfectadas o infectadas consiste en que al controlar el número de células que se introducen en un sujeto, se puede controlar la cantidad de proteína administrada al cuerpo. Además, en algunos casos, es posible eliminar las células transfectadas o infectadas si ya no es necesario el producto. Otra ventaja del tratamiento del uso de células primarias o secundarias transfectadas o infectadas de la presente invención es que la producción del producto terapéutico se puede regular, por ejemplo a través de la administración de zinc, esteroides o un agente que afecte a la transcripción de una proteína, producto o producto de ácido nucleico, o que afecte a la estabilidad de un producto de ácido nucleico.

#### Animales transgénicos

20 [0107] Se han utilizado una serie de métodos para obtener mamíferos no humanos transgénicos. Un mamífero no humano transgénico hace referencia a un mamífero que ha obtenido un gen adicional gracias a la introducción de una secuencia de ácido nucleico sintético exógeno, es decir, un transgén, en sus propias células (p.ej., tanto en las células somáticas como en las germinales) o en una línea germinal de su especie primitiva.

25 [0108] Existen una serie de métodos para introducir ADN exógeno en una línea germinal (p.ej., la introducción en las células somáticas o germinales) de un mamífero. Un método es la microinyección del constructo de un gen en el pronúcleo de un embrión en etapa temprana (p.ej., antes de la etapa de cuatro células) (Wagner *et al.*, *Proc. Natl. Acad. Sci. EE. UU.* 78:5016 (1981); Brinster *et al.*, *Proc Natl Acad Sci EE: UU.* 82:4438 (1985)). Se ha descrito el procedimiento detallado para producir estos ratones transgénicos (véase, p.ej., Hogan *et al.*, *Manipulating the Mouse Embryo*, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY (1986); la patente de EE.UU. Nº. 5.175.383 (1992)). Este procedimiento también ha sido adaptado a otras especies de mamíferos (p.ej., Hammer *et al.*, *Nature* 315:680 (1985); Murray *et al.*, *Reprod. Fert. Devl.* 1:147 (1989); Pursel *et al.*, *Vet. Immunol. Histopath.* 17:303 (1987); Rexroad *et al.*, *J. Reprod. Fert.* 41(supl):119 (1990); Rexroad *et al.*, *Molec. Reprod. Devl.* 1:164 (1989); Simons *et al.*, *BioTechnology* 6:179 (1988); Vize *et al.*, *J. Cell. Sci.* 90:295 (1988); y Wagner, *J. Cell. Biochem.* 13B (supl):164 (1989).

30 [0109] Otro método para producir mamíferos transgénicos de línea germinal es a través del uso de células madre embrionarias o células somáticas (p.ej., embrionarias, fetales o adultas). El constructo del gen puede introducirse en las células madre embrionarias mediante recombinación homóloga (Thomas *et al.*, *Cell* 51:503 (1987); Capecchi, *Science* 244:1288 (1989); Joyner *et al.*, *Nature* 338: 153 (1989)). También puede introducirse un constructo adecuado en las células madre embrionarias a través de transfección mediada de ADN, como electroporación (Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons (1987)). Los procedimientos detallados para cultivar células madre embrionarias (p.ej., ESD-3, ATCC# CCL-1934, ES-E14TG-2a, ATCC# CCL-1821, American Type Culture Collection, Rockville, MD) y los métodos para obtener mamíferos transgénicos a partir de células madre embrionarias se pueden encontrar en *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, ed. E.J. Robertson (IRL Press, 1987). Los métodos para obtener animales transgénicos a partir de células somáticas se presentan en, por ejemplo, WO 97/07669, WO 97/07668 la patente de EE. UU. 5.945.577.

35 [0110] En los métodos mencionados para generar mamíferos transgénicos a través de la línea germinal, se puede introducir el constructo como un constructo lineal, como un plásmido circular o como un vector que puede añadirse y heredarse como un transgén integrado en el genoma huésped. También puede construirse el transgén para permitir que se herede como un plásmido extracromosómico (Gassmann, M. *et al.*, *Proc. Natl. Acad. Sci. EE.UU.* 92:1292 (1995)).

#### Factor VIII humano (EJEMPLO COMPARATIVO)

40 [0111] El hFVIII está codificado por un gen de 186 kilobases (kb), con la región de codificación distribuida entre 26 exones (Gitchier *et al.*, *Nature*, 312:326-330, (1984)). La transcripción del gen y el corte y empalme del transcrito primario resultante produce un ARNm de aproximadamente 9 kb que codifica un producto de traducción primario que contiene 2531 aminoácidos (aa), incluido un péptido señal de 19 aa. Si se excluye el péptido señal, la proteína de 2332 aa tiene una estructura de dominio que puede presentarse del siguiente modo NH<sub>2</sub>-A1-A2-B-A3-C1-C2-COOH, con una masa molecular predicha de 265 kilodaltons (kD). La glicosilación de esta proteína da lugar a un producto con una masa

molecular de aproximadamente 330 kD como determina la SDS-PAGE. En el plasma, hFVIII es una proteína heterodimérica que consiste en una cadena pesada que varía en tamaño de 90 kD a 200 kD en un complejo de ion metálico con una cadena ligera de 80 kD. El complejo heterodimérico se estabiliza además por interacciones con el vWF. La cadena pesada está compuesta por los dominios A1-A2-B y la cadena ligera está compuesta por los dominios A3-C1-C2 (Figura 2). Los sitios de escisión de proteasa en el dominio B representan la variación de tamaño de la cadena pesada, con las especies de 90 kD que no contienen secuencias de dominio B y las especies de 200 kD que contienen un dominio B completo o casi completo. El dominio B no tiene función conocida y está totalmente eliminado después de la activación de hFVIII por la trombina.

**[0112]** Los plásmidos de expresión del Factor VIII humano, plásmidos pXF8.186 (Figura 3), pXF8.61 (Figura 4), pXF8.38 (Figura 11) y pXF8.224 (Figura 13) se describen a continuación. El plásmido del constructo de expresión de hFVIII, pXF8.186, se desarrolló basándose en los estudios de optimización detallados que ofrecieron un alto nivel de expresión de un hFVIII funcional. Dado el tamaño extremadamente grande del gen del hFVIII y la necesidad de transferir la región de codificación completa a las células, los plásmidos de expresión de ADNc se desarrollaron para la producción de cepas de células clónicas transfectadas de manera estable. Se ha demostrado que es difícil lograr expresión de alto nivel de hFVIII empleando el ADNc de 9 kb de tipo salvaje. A continuación, se presentan tres razones potenciales para la expresión pobre. En primer lugar, el ADNc de tipo salvaje codifica los 909 aa, el dominio B glicosilado de modo pesado que está unido de forma transitoria a la cadena pesada y no se le conoce función (Figura 1). La eliminación de la región que codifica el dominio B a partir de los constructos de expresión de hFVIII produce una expresión altamente mejorada de una proteína funcional. El análisis de los derivados de hFVIII que carecen del dominio B han demostrado que la función del hFVIII no se ve afectada negativamente y que dichas moléculas tienen propiedades bioquímicas, inmunológicas y funcionales *in vivo* que son muy similares a la proteína de tipo salvaje. Se han desarrollado dos constructos diferentes de expresión del Factor VIII humano con el dominio B suprimido (BDD hFVIII), que codifican proteínas con diferentes secuencias de aminoácidos que flanquean la eliminación. El plásmido pXF8.186 contiene una eliminación completa del dominio B (aminoácidos 741-1648 de la secuencia de proteínas maduras de tipo salvaje), con la secuencia Arg-Arg-Arg-Arg (RRRR) insertada en la unión de la cadena ligera y la cadena pesada (Figura 1). Esto da como resultado una serie de residuos consecutivos de arginina (RRRRR o 5R) en la unión de la cadena ligera y la cadena pesada, que comprende un sitio de reconocimiento para una proteasa intracelular del tipo PACE/furina, y se predijo que promovería la escisión para producir las cadenas ligera y pesada correctas. El plásmido pXF8.61 también contiene una delección completa del dominio B con un sitio XhoI sintético en la unión. Esta unión provoca la presencia de la secuencia dipéptida Leu-Glu (LE) en la unión de la cadena ligera y la cadena pesada en las dos formas de BDD hFVIII, las proteínas expresadas se denominan en este documento BDD hFVIII de 5R y de LE.

**[0113]** La segunda característica que se ha observado que afecta de modo adverso a la expresión de hFVIII en las células transfectadas hace referencia a la observación en la que se han identificado una o más regiones de la región de codificación que funcionan eficazmente para bloquear la transcripción de la secuencia de ADNc. Los inventores han descubierto ahora que se puede reducir o eliminar la influencia negativa de los elementos de la secuencia mediante la alteración de la secuencia de codificación completa. Con este fin, un ADNc sintético del hFVIII con el dominio B suprimido por completo se preparó como se describe con mayor detalle a continuación. Se llevaron a cabo cambios de bases de silenciamiento en todos los codones que no se correspondían con la secuencia de tripletes más frecuentemente encontrada para ese aminoácido en proteínas humanas altamente expresada, y esos codones se convirtieron a la secuencia de codón más frecuentemente encontrada en humanos para el aminoácido correspondiente. La secuencia de codificación resultante tiene un total de 1094 de 4335 de pares de base que difieren de la secuencia de tipo salvaje, además codifica una proteína con la secuencia de hFVIII de tipo salvaje (con la excepción de la delección del dominio B). Se cambiaron el 25,2% de las bases y el contenido de GC de la secuencia aumentó del 44% al 54%. Esta secuencia alterada en el ADNc de BDD hFVIII se expresa al menos 5,3 veces más eficaz que un constructo de control no alterado.

**[0114]** La tercera característica que se optimizó para mejorar la expresión de hFVIII fue la estructura intrón-extrón del constructo de expresión. El ADNc, por definición, carece de intrones. Si bien esto reduce el tamaño del constructo de expresión, se ha demostrado que los intrones pueden tener fuertes efectos positivos en la expresión génica cuando se añaden a los constructos de expresión de ADNc. La región no traducida 5' del gen humano de beta-actina que contiene un intrón funcional completo se incorporó a los constructos de expresión de BDD hFVIII, pXF8.61 y pXF8.186.

**[0115]** La cuarta característica que puede afectar de modo adverso a la expresión de hFVIII es la estabilidad del ARNm del Factor VII. La estabilidad del mensaje puede afectar al nivel en estado estacionario del ARNm del Factor VIII, e influir a la expresión génica. Las secuencias específicas en el Factor VIII puede alterarse con el fin de aumentar la estabilidad del ARN, p.ej., la eliminación de AURE de la UTR 3' puede resultar en un ARNm del Factor VIII más estable. Los datos presentados a continuación muestran que la nueva modificación genética de la secuencia de codificación tiene utilidad general para la mejora de la expresión de genes de eucariotas de mamíferos y de no mamíferos en células de mamíferos. Los resultados obtenidos en el presente documento con el Factor VIII humano sugieren que la optimización de codones sistemáticas (sin tener en cuenta el contenido CpG) proporciona una estrategia fructífera para mejorar la expresión en las células de mamíferos de una gran variedad de genes de eucariotas.

Métodos para realizar secuencias de nucleótidos sintéticas

**[0116]** Una secuencia de ácido nucleico sintético que dirige la síntesis de un mensaje optimizado de la invención puede realizarse, p.ej., por cualquiera de los métodos descritos en este documento. Los métodos descritos a continuación son útiles para producir mensajes optimizados por las siguientes razones:

5 1) permiten la producción de una proteína altamente optimizada, p.ej., una proteína que tiene al menos del 94 al 100% de los codones como codones comunes, en especial para proteínas de más de 90 aminoácidos de longitud. El producto final puede estar 100% optimizado, es decir, cada nucleótido es tal y como se elige, sin necesidad de introducir alteraciones no deseadas cada 100-300 bp. Un gen puede sintetizarse con codones 100% óptimos o puede sintetizarse con 100% los codones que se desean. Se pueden introducir o evitar elementos adicionales de la secuencia de ADN sin las limitaciones impuestas por la necesidad de introducir sitios de enzimas de restricción. Estos elementos de secuencia podrían incluir:

- Señales de transcripción, como potenciadores o silenciadores.
- Señales de corte y empalme, por ejemplo, evitando sitios de corte y empalme críticos en un ADNc u optimizando el contexto del sitio de corte y empalme en un gen que contiene intrones. Añadir un intrón a un ADNc puede ayudar a la expresión y permite la introducción de señales de la transcripción en el gen.
- 15 • Señales de inestabilidad – la creación u omisión de secuencias que dirigen la degradación del ARNm.
- Estructura secundaria – la creación u omisión de estructuras secundarias en el ARNm que pueden afectar a la traducción, terminación de la transcripción o estabilidad del ARNm.
- Señales de traducción – Elección del codón. Se puede sintetizar un gen con codones 100% óptimos, o la preferencia codónica por cualquier aminoácido puede alterarse sin restricción para hacer la expresión génica sensible a la concentración de un aminoacil-ARNt, cuya concentración puede variar con el crecimiento o el estado metabólico.

En cada caso, el objetivo será aumentar o disminuir la expresión para lograr la expresión en una forma concreta de regulación.

25 2) mejoran la precisión de la secuencia sintética porque evitan la amplificación por PCR que produce errores en la secuencia amplificada; y

3) reducen el coste de realización de la secuencia sintética de la invención.

**[0117]** La secuencia de ácido nucleico sintético que dirige la síntesis de los mensajes optimizados de la invención puede repararse, p.ej., utilizando la estrategia que se resume con mayor detalle a continuación.

Estrategia para construir una secuencia

30 **[0118]** La etapa inicial es diseñar un protocolo de clonación.

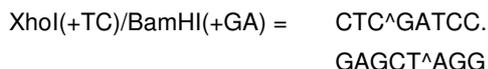
**[0119]** Se genera un archivo de secuencia que contiene 100% la secuencia de ADN deseada. Esta secuencia se analiza para determinar sitios de restricción, incluidos sitios de fusión.

**[0120]** Los sitios de fusión son por orden de preferencia:

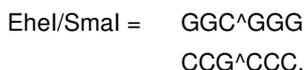
35 A) Las secuencias resultantes de la ligadura de dos salientes complementarios normalmente generados por enzimas de restricción disponibles, p.ej.,

Sall/XhoI =	G <sup>^</sup> TCGAG CAGCT <sup>^</sup> C
o BspDI/BstBI =	AT <sup>^</sup> CGAA TAGC <sup>^</sup> TT
o BstBI/AccI =	TT <sup>^</sup> CGAC AAGC <sup>^</sup> TG.

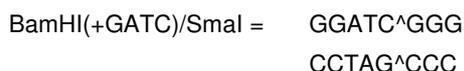
B) Las secuencias resultantes de la ligadura de dos salientes generados llenando parcialmente los salientes de enzimas de restricción disponibles, p.ej.,



C) Las secuencias resultantes de la ligadura roma de dos extremos romos normalmente generados por las enzimas de restricción disponibles, p.ej.,



D) Las secuencias resultante de la ligadura roma de dos extremos romos, donde uno o ambos extremos romos han sido generados por llenado en un saliente, p.ej.,



5 **[0121]** El relleno de un saliente 5' generado por una enzima de restricción se realiza usando una ADN polimerasa, por ejemplo el fragmento de Klenow de la ADN polimerasa I. Si el saliente es para llenarlo por completo, entonces se incluyen en la reacción los cuatro nucleótidos: dATP, dCTP, dGTP y dTTP. Si el saliente es para llenarlo sólo parcialmente, entonces los nucleótidos necesarios se omiten en la reacción. En el punto (B) anterior, el ADN digerido por XhoI se rellenaría con Klenow en presencia de dCTP y dTTP y omitiendo dATP y dGTP. Se determina un orden de las etapas de clonación, lo que permite el uso de sitios de aproximadamente 150-500 bp de distancia. Hay que tener en cuenta que un fragmento debe carecer de la secuencia de reconocimiento para una enzima, sólo si esta enzima se utiliza para clonar el fragmento. Por ejemplo, la estrategia para la construcción de la secuencia de codificación del Factor VIII "deseada" puede utilizar ApaI en una serie de sitios diferentes, debido a que la regla de ensamblaje de los fragmentos - ApaI no se utiliza en los siguientes pasos de clonación.

15 **[0122]** Si existe una región que no tiene sitios útiles disponibles, entonces se puede utilizar una estrategia de secuencia independiente: los fragmentos se clonan en un constructo de ADN que contiene secuencias de reconocimiento para las enzimas de restricción que se escinden fuera de su secuencia de reconocimiento, p.ej. BseRI=

GAGGAGNNNNNNNNNN<sup>^</sup> (SEQ ID N°:5)

CTCCTCNNNNNNNNNN<sup>^</sup>NN (SEQ ID N°:6)

20 Fragmento génico del sitio de clonación del constructo de ADN

**[0123]** La secuencia de reconocimiento de la enzima utilizada para clonar el fragmento se eliminará cuando se libere el fragmento por la digestión con, p.ej., BseRI, dejando un fragmento que comprende el 100% de la secuencia deseada, que luego se puede unir a un fragmento génico adyacente generado de modo similar.

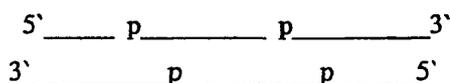
**[0124]** La siguiente etapa es sintetizar los fragmentos de restricción inicial.

25 **[0125]** La síntesis de los fragmentos de restricción iniciales puede lograrse de muchos modos, entre los que se incluyen, sin carácter limitativo:

1. Síntesis química del fragmento completo.

2. Sintetizar dos oligonucleótidos que son complementarios en sus extremos 3', hibridarlos y utilizar el fragmento de Klenow de la ADN polimerasa, o un equivalente, para extender, ofreciendo un fragmento de doble hebra.

3. Sintetizar un número de oligonucleótidos más pequeños, aquellos oligos de cinasa que tienen extremos 5' internos, hibridar todos los oligos y ligarlos, en concreto.



35 **[0126]** Las técnicas 2 y 3 pueden utilizarse en etapas posteriores para unir fragmentos más pequeños entre ellos. La PCR puede utilizarse para aumentar la cantidad de material para la clonación, pero puede conllevar un incremento en el número de mutaciones. Si no se obtiene un fragmento sin errores, entonces se puede utilizar la mutagénesis dirigida al sitio para corregir el mejor aislado. A continuación, le sigue la concatenación de los fragmentos sin errores y la secuenciación de las uniones para confirmar su precisión.

Uso

[0127] Las secuencias de ácido nucleico sintético descritas en el presente documento son útiles para la expresión de una proteína que se expresa normalmente en una célula de mamífero o en un cultivo celular (p.ej., para la producción comercial de proteínas humanas como GH, tPA, GLP-1, EPO,  $\alpha$ -galactosidasa,  $\beta$ -glucoceramidasa,  $\alpha$ -iduronidasa;  $\alpha$ -L-iduronidasa, glucosamina-N-sulfatasa, alfa-N-acetilglucosaminidasa, acetil coenzima A: $\alpha$ -glucosaminidasa-N-acetiltransferasa, N-acetilglucosamina-6-sulfatasa, N-acetilglucosamina-6-sulfatasa,  $\beta$ -galactosidasa, N-acetilgalactosamina-6-sulfatasa,  $\beta$ -glucuronidasa. Factor VIII y Factor IX). Las secuencias de ácido nucleico sintético descritas en el documento son también útiles para la terapia génica. Por ejemplo, se puede introducir directamente una secuencia de ácido nucleico sintético que codifica una proteína seleccionada, p.ej. mediante transfección celular no viral o a través de un vector en una célula, p.ej., una célula transformada o una célula no transformada, que puede expresar la proteína para crear una célula que se puede administrar a un paciente con necesidad de la proteína. Estas técnicas de terapia génica basadas en células se describen con mayor detalle en las solicitudes de EE. UU pendientes de resolución: USSN 08/334.797; USSN 08/231.439; USSN 08/334.455 y USSN 08/928.881.

### Ejemplos

#### I. Constructos de Factor VIII y usos de los mismos (EJEMPLO COMPARATIVO)

##### 15 Construcción de pXF8.61

[0128] Los catorce fragmentos génicos del ADNc optimizado de FVIII con el dominio B suprimido enumerados en la Tabla 2 y mostrados en la Figura 5 (Fragmento A-Fragmento N) se realizaron de la siguiente manera. Se hicieron 92 oligonucleótidos mediante síntesis de oligonucleótidos en un sintetizador ABI 391 (Perkin Elmer). Los 92 oligonucleótidos se enumeran en la Tabla 3. La Figura 5 muestra como estos 92 oligonucleótidos se hibridan para formar los catorce fragmentos génicos de la Tabla 2. Para cada hebra de cada fragmento génico, el primer oligonucleótido (es decir, el más 5') se crearon con un extremo 5'-hidroxilo, y los siguientes oligonucleótidos se fabricaron como 5'-fosforilado para permitir la ligadura de los oligonucleótidos hibridados adyacentes. Para los fragmentos génicos A, B, C, F, G, J, K, L, M y N, seis oligonucleótidos fueron hibridados, ligados y digeridos con EcoRI y HindIII y clonados en pUC18 digerida con EcoRI y HindIII. Para los fragmentos génicos D, E, H e I, ocho oligonucleótidos fueron hibridados, ligados, digeridos con EcoRI y HindIII y clonados en pUC18 digerida con EcoRI y HindIII. Este procedimiento generó catorce plásmidos-pAM1A diferentes a través de pAM1N.

**Tabla 2**

Fragmento	Extremo 5'		Extremo 3'		Nota
A	NheI	1	Apal	279	
B	Apal	279	Pm1I	544	
C	Pm1I	544	Pm1I	829	
D	Pm1I	829	Bg1II/(BamHI)	1172	BamHI sitio 3' a siguientes
E	(Bg1II)/Bam HI	1172	Bg1II	1583	
F	Bg1II	1583	KpnI	1817	
G	KpnI	1817	BamHI	2126	
H	BamHI	2126	Pm1I	2491	
I	Pm1I	2491	KpnI	3170	$\Delta$ BstEII 2661-2955
J	BstEII	2661	BstEII	2955	
K	KpnI	3170	Apal	3482	
L	Apal	3482	SmaI/(EcoRV)	3772	
M	(SmaI)/EcoR V	3772	BstEII	4062	
N	BstEII	4062	SmaI	4348	

30 En la Tabla 2, las posiciones de sitios de restricción están numeradas por la primera base del palíndromo; la numeración empieza en el sitio de NheI.

Tabla 3

Nombre de Oligo	Longitud de Oligo'	Secuencia de oligonucleótidos
AM1Af 1	118	GTAGAATTCGTAGGCTAGCATGCAGATCGAGCTGAGCACCTGCTTCTTCC TGTGCCTGCTGCGCTTCTGCTTCAGCGCCACCCGCGCTACTACCTGGGCGC CGTGGAGCTGAGCTGG (SEQ ID NO:7)
AM1Af 2	104	GACTACATGCAGAGCGACCTGGGCGAGCTGCCCGTGGACGCCCGCTTCCC CCCCCGCGTGCCCAAGAGCTTCCCCTTCAACACCAGCGTGGTGTACAAGAA GAC (SEQ ID NO: 8)
AM1Af 3	88	CCTGTTCGTGGAGTTACCGACCACCTGTTCAACATCGCCAAGCCCCGCC CCCCCTGGATGGGCCTGCTGGGCCCTACAAGCTTTAC (SEQ ID NO: 9)
AM1Ar 1	119	GTAAAGCTTGTAGGGGCCAGCAGGCCCATCCAGGGGGGGCGGGGCTTG GCGATGTTGAACAGGTGGTCCGTTGAAGTCCACGAACAGGGTCTTCTGTAC ACCACGCTGGTGTGAAGG (SEQ ID NO: 10)
AM1Ar 2	107	GGAAGCTCTTGGGCACGCGGGGGGGAAGCGGGCGTCCACGGGCAGCTC GCCAGGTCGCTCTGCATGTAGTCCCAGCTCAGCTCCACGGCGCCAGGTA GTAGCGG (SEQ ID NO: 11)
AM1Ar 3	84	CGGGTGGCGCTGAAGCAGAAGCGCAGCAGGCACAGGAAGAAGCAGGTG CTCAGCTCGATCTGCATGCTAGCCTACGAATTCTAC (SEQ ID NO: 12)
AM1Bf 1	115	GTAGAATTCGTAGGGGCCACCATCCAGGCCGAGGTGTACGACACCGT GGTGTACCCCTGAAGAACATGGCCAGCCACCCCGTGGAGCTGCACGCCGT GGGCGTGGAGCTACTG (SEQ ID NO: 13)
AM1Bf 2	103	GAAGGCCAGCGAGGGCGCCGAGTACGACGACCAGACCAGCCAGCGCGA GAAGGAGGACGACAAGGTGTTCCCGGCGGCAGCCACACCTACGTGTGGC AGGTG (SEQ ID NO: 14)
AM1Bf 3	79	CTGAAGGAGAACGGCCCCATGGCCAGCGACCCCTGTGCCTGACCTACA GCTACCTGAGCCACGTGCTACAAGCTTTAC (SEQ ID NO: 15)
AM1Br 1	107	GTAAAGCTTGTAGCACGTGGCTCAGGTAGCTGTAGGTGAGGCACAGGGG GTCGCTGGCCATGGGGCCGTTCTCCTTCAGCACCTGCCACACGTAGGTGTG GCTGCCG (SEQ ID NO: 16)
AM1Br 2	101	CCGGGGAACACCTTGTCTCCTCCTTCTCGCGCTGGCTGGTCTGGTCTGTCG TACTCGGCGCCCTCGCTGGCCTTCCAGTAGCTCACGCCACGGCGTGCAG (SEQ ID NO: 17)
AM1Br 3	89	GCTCACGGGTGGCTGGCCATGTTCTTCAGGGTGATCACCAGGTGTCTG ACACCTCGGCCTGGATGGTGGGGCCCCCTACGAATTCTAC (SEQ ID NO: 18)
AM1Cf 1	122	GTAGAATTCGTAGCCACGTGGACCTGGTGAAGGACCTGAACAGCGGCCT GATCGGCGCCCTGCTGGTGTGCCGCGAGGGCAGCCTGGCCAAGGAGAAGA CCCAGACCTGCACAAGTTCATC (SEQ ID NO: 19)
AM1Cf 2	110	CTGCTGTTCCCGTGTTCGACGAGGGCAAGAGCTGGCACAGCGAGACCA AGAACAGCCTGATGCAGGACCGCGACGCCGCCAGCGCCCGCGCTGGCCC AAGATGCACAC (SEQ ID NO: 20)
AM1Cf 3	86	CGTGAACGGCTACGTGAACCGCAGCCTGCCCGGCCTGATCGGCTGCCACC GCAAGAGCGTGTACTGGCACGTGCTACAAGCTTTAC (SEQ ID NO: 21)
AM1Cr 1	108	GTAAAGCTTGTAGCACGTGCCAGTACACGCTCTTGGGTTGGCAGCCGATC AGGCCGGCAGGCTGCGGTTACGTAGCCGTTACGGTGTGCATCTTGGGC CAGGCGC (SEQ ID NO: 22)
AM1Cr 2	110	GGGCGTGGCGGCGTCCGCGTCTGCATCAGGCTGTTCTTGGTCTCGCTG TGCCAGCTCTTGCCCTCGTGAACACGGCGAACAGCAGGATGAACTTGTGC AGGGTCTGG (SEQ ID NO: 23)
AM1Cr 3	100	GTCTTCTCCTTGGCCAGGCTGCCCTCGCGGCACACCAGCAGGGCGCCGAT CAGGCCGCTGTTACAGTCTTACCAGGTCCACGTGGTACGAATTCTAC (SEQ ID NO: 24)
AM1Df	99	GTAGAATTCGTAGCACGTGATCGGCATGGGCACCACCCCGAGGTGCAC AGCATCTTCTGGAGGGCCACACCTTCTGGTGCGCAACCACCGCCAGGC (SEQ ID NO: 25)
1		
AM1Df 2	100	CAGCCTGGAGATCAGCCCCATCACCTTCTGACCGCCAGACCCTGCTGA TGGACCTGGGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACCAGCAC (SEQ ID NO: 26)
AM1Df 3	101	GACGGCATGGAGGCTACGTGAAGGTGGACAGCTGCCCGAGGAGCCCC AGCTGCGCATGAAGAACAACGAGGAGGCCGAGGACTACGACGACGACCTG AC (SEQ ID NO: 27)

ES 2 437 815 T3

AM1Df 4	84	CGACAGCGAGATGGACGTGGTGGCCTTCGACGACGACAACAGCCCCAGC TTCATCCAGATCTCTACGGATCCTACAAGCTTTAC (SEQ ID NO: 28)
AM1Dr 1	109	GTAAAGCTTGTAGGATCCGTAGAGATCTGGATGAAGCTGGGGCTGTTGTC GTCGTCGAAGCGCACCCAGTCCATCTCGCTGTCGGTCAGGTCGTCGTCGTA GTCCTCGG (SEQ ID NO: 29)
AM1Dr 2	101	CCTCCTCGTTGTTCTTCATGCGCAGCTGGGGCTCCTCGGGGCAGCTGTCCA CCTTACGTAGGCCTCCATGCCGTCGTGCTGGTGGCTGCTGATGTGGCAG (SEQ ID NO: 30)
AM1Dr 3	102	AACAGCAGGAAGTGGCCAGGTCCATCAGCAGGGTCTGGGCGGTTCAGGA AGGTGATGGGGCTGATCTCCAGGCTGGCCTGGCGGTGGTTGCGCACCCAGG AAG (SEQ ID NO: 31)
AM1Dr 4	72	GTGTGGCCCTCCAGGAAGATGCTGTGCACCTCGGGGGTGGTGCCCATGCC GATCACGTGCTACGAATTCTAC (SEQ ID NO: 32)
AM1Ef 1	122	GTAGAATTCGTAGGGATCCGCAGCGTGGCCAAGAAGCACCCCAAGACCT GGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCTG GTGCTGGCCCCGACGACCCGAG (SEQ ID NO: 33)
AM1Ef 2	120	CTACAAGAGCCAGTACCTGAACAACGGCCCCCAGCGCATCGGCCGCAAG TACAAGAAGGTGCGCTTCATGGCCTACACCGACGAGACCTTCAAGACCCGC GAGGCCATCCAGCAGGAG (SEQ ID NO: 34)
AM1Ef 3	115	CGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCTGCTGA TCATCTTCAAGAACCAGGCCAGCCGCCCTACAACATCTACCCCCACGGCA TCACCGACGTGCGC (SEQ ID NO: 35)
AM1Ef 4	86	CCCCTGTACAGCCGCCGCTGCCAAGGGCGTGAAGCACCTGAAGGACTT CCCCATCCTGCCCCGCGAGATCTCTACAAGCTTTAC (SEQ ID NO: 36)
AM1Er 1	109	GTAAAGCTTGTAGAGATCTCGCCGGCAGGATGGGGAAGTCTTCAGGT GCTTACGCCCTTGGGCAGGCGGCGGTGTACAGGGGGCGCACGTCCGGTG ATGCCGTGGG (SEQ ID NO: 37)
AM1Er 2	114	GGTAGATGTTGTAGGGGCGGCTGGCCTGGTTCTTGAAGATGATCAGCAGG GTGTCGCCACCTCGCCGTACAGCAGGGGGCCAGGATGCCGCTCTCGTGC TGGATGGCCTCGC (SEQ ID NO: 38)
AM1Er 3	121	GGGTCTTGAAGGTCTCGTCCGTGAGGCCATGAAGCGCACCTTCTGTAC TTGCGGCCGATGCGCTGGGGCCGTTGTTTCAGGTAAGGCTCTTGTAGCTG CGGTCTCGGGGGCCAGCAC (SEQ ID NO: 39)
AM1Er 4	99	CAGGGGGGCGTAGTCCCAGTCTCCTCCTCGGCGGCGATGTAGTGCACCC AGGTCTTGGGGTGCTTCTTGGCCACGCTGCGGATCCCTACGAATTCTAC (SEQ ID NO: 40)
AM1Ff1	102	GTAGAATTCGTAGAGATCTTCAAGTACAAGTGGACCGTGACCGTGGAGG ACGGCCCCACCAAGAGCGACCCCCGCTGCCTGACCCGCTACTACAGCAGCT TC (SEQ ID NO: 41)
AM1Ff 2	103	GTGAACATGGAGCGCGACCTGGCCAGCGGCCTGATCGGCCCCCTGCTGAT CTGCTACAAGGAGAGCGTGGACCAGCGCGCAACCAGATCATGAGCGACA AGC (SEQ ID NO: 42)
AM1Ff 3	61	GCAACGTGATCCTGTTACGCGTGTTCGACGAGAACCGCAGCTGGTACCCT ACAAGCTTTAC (SEQ ID NO: 43)
AM1Fr 1	87	GTAAAGCTTGTAGGGTACCAGCTGCGGTTCTCGTGAACACGCTGAACAG GATCACGTTGCGCTTGTGCTCATGATCTGGTTGCCG (SEQ ID NO: 44)
AM1Fr 2	101	CGCTGGTCCACGCTCCTTGTAGCAGATCAGCAGGGGGCCGATCAGGCC GCTGGCCAGTCCGCTCCATGTTACGAAGCTGCTGTAGTAGCGGGTCAG (SEQ ID NO: 45)
AM1Fr 3	78	GCAGCGGGGGTTCGCTCTTGGTGGGGCCGCTCCTCCACGGTCACGGTCCACT TGTACTTGAAGATCTCTACGAATTCTAC (SEQ ID NO: 46)
AM1Gf 1	120	GTAGAATTCGTAGGGTACCAGCTGACCGAGAATCCAGCGCTTCTGCCCCA CCCCGCCGGCGTGCAGCTGGAGGACCCCGAGTCCAGGCCAGCAACATCA TGCACAGCATCAACGGCTAC (SEQ ID NO: 47)
AM1Gf 2	126	GTGTTCCACAGCCTGCAGCTGAGCGTGTGCTGCACGAGGTGGCCTACTG GTACATCCTGAGCATCGGCCGCCAGACCGACTTCTGAGCGTGTCTTTCAG CGGCTACACCTTCAAGCACAAGATG (SEQ ID NO: 48)
AM1Gf 3	95	GTGTACGAGGACACCCTGACCCTGTTCCCTTCAGCGGCGAGACCGTGT CATGAGCATGGAGAACCCCGCCTGTGGATCCCTACAAGCTTTAC (SEQ ID NO: 49)

ES 2 437 815 T3

AM1Gr 1	119	GTAAAGCTTGTAGGGATCCACAGGCCGGGGTCTCCATGCTCATGAACAC GGTCTCGCCGCTGAAGGGGAACAGGGTCAGGGTGTCTCGTACACCATCTT GTGCTTGAAGGTGTAGCC (SEQ ID NO: 50)
AM1Gr 2	124	GCTGAAGAACACGCTCAGGAAGTCGGTCTGGGCGCCGATGCTCAGGATG TACCAGTAGGCCACCTCGTGCAGGCACACGCTCAGCTGCAGGCTGTGCAAC ACGTAGCCGTTGATGCTGTGCATG (SEQ ID NO: 51)
AM1Gr 3	98	ATGTTGCTGGCCTGGAACTCGGGGTCTCCAGCTGCACGCCGGCGGGGTT GGGCAGGAAGCGCTGGATGTTCTCGGTCAGGTACCCTACGAATTCTAC (SEQ ID NO: 52)
AM1Hf 1	111	GTAGAATTCGTAGGGATCCTGGGCTGCCACAACAGCGACTTCCGCAACCG CGGCATGACCGCCTGTGAAGGTGAGCAGCTGCGACAAGAACACCGGCG ACTACTACGAG (SEQ ID NO: 53)
AM1Hf 2	102	GACAGCTACGAGGACATCAGCGCCTACCTGCTGAGCAAGAACAACGCCA TCGAGCCCCGCTGGAGGAGATCACCCGCACCACCCTGCAGAGCGACCAG GAG (SEQ ID NO: 54)
AM1Hf 3	105	GAGATCGACTACGACGACACCATCAGCGTGGAGATGAAGAAGGAGGACT TCGACATCTACGACGAGGACGAGAACCAGAGCCCCCGCAGCTTCCAGAAG AAGACC (SEQ ID NO: 55)
AM1Hf 4	79	CGCCACTACTTCATCGCCCGCTGGAGCGCCTGTGGGACTACGGCATGAG CAGCAGCCCCACGTGCTACAAGCTTTAC (SEQ ID NO: 56)
AM1Hr 1	101	GTAAAGCTTGTAGCACGTGGGGGCTGTGCTCATGCCGTAGTCCCAACAGG CGCTCCACGGCGCGATGAAGTAGTGGCGGGTCTTCTTCTGGAAGCTGCGG (SEQ ID NO: 57)
AM1Hr 2	105	GGGCTCTGGTTCTCGTCTCTCGTAGATGTCGAAGTCTCCTTCTTCATC TCCACGCTGATGGTGTCTCGTAGTCGATCTCCTCCTGGTCGCTCTGCAGGG TG (SEQ ID NO: 58)
AM1Hr 3	108	GTGCGGGTGATCTCCTCCAGGCGGGGCTCGATGGCGTTGTTCTTGCTCAG CAGGTAGGCGCTGATGTCTCGTAGCTGTCCTCGTAGTAGTCGCCGGTGT CTTGTCG (SEQ ID NO: 59)
AM1Hr 4	83	CAGCTGCTACCTTCAGCAGGGCGGTATGCCGCGTTGCGGAAGTCGCTGTTGTGGCAG CCCAGGATCCCTACGAATTCTAC (SEQ ID NO: 60)
AM1lf 1	115	GTAGAATTCGTAGCACGTGCTGCGCAACCGCGCCAGAGCGGCAGCGTG CCCCAGTTCAAGAAGGTGGTGTCCAGGAGTTCACCGACGGCAGCTTCACC CAGCCCCGTGTACCG (SEQ ID NO: 61)
AM1lf 2	111	GGCGAGCTGAACGAGCACCTGGGCCTGCTGGGCCCTACATCCGCGCCG AGGTGGAGGACAACATCATGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCA CCATCTTCGAC (SEQ ID NO: 62)
AM1lf 3	106	GAGACCAAGAGCTGGTACTTCACCGAGAACATGGAGCGCAACTGCCGCG CCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACCGCT TCCACG (SEQ ID NO: 63)
AM1lf 4	85	CCATCAACGGCTACATCATGGACACCCTGCCCGCCTGGTGATGGCCCAG GACCAGCGCATCCGCTGGTACCCTACAAGCTTTAC (SEQ ID NO: 64)
AM1lr 1	115	GTAAAGCTTGTAGGGTACCAGCGGATGCGCTGGTCTGGGCCATCACCAG GCCGGGCAGGGTGTCCATGATGTAGCCGTTGATGGCGTGAAGCGGTAGTT CTCCTTGAAGGTGG (SEQ ID NO: 65)
AM1lr 2	99	GGTCTCCATCTGGATGTTGCAGGGGGCGCGGCAGTTGCGCTCCATGTT TCGGTGAAGTACCAGCTCTTGGTCTCGTGAAGATGGTGAAGAACAGGG (SEQ ID NO: 66)
AM1lr 3	110	CGAACTCCTGCACGGTCACCATGATGTTGTCCTCCACCTCGGCGCGGATG TAGGGGCCAGCAGGCCAGGTGCTCGTTCAGCTCGCCGCGGTACAGGGG CTGGGTGAAG (SEQ ID NO: 67)
AM1lr 4	93	CTGCCGTCGGTGAACCTCCTGGAACACCACCTTCTTGAACCTGGGGCAGCT GCCGCTCTGGGCGCGGTTGCGCAGCACGTGCTACGAATTCTAC (SEQ ID NO: 68)
AM1Jf 1	116	GTAGAATTCGTAGGGTACCTTCCGCAACAGGCCAGCCGCCCTACAGCTTCTACAGCA GCCTGATCAGTACGAGGAGGACAGCGCCAGGGCGCCGAGCCCCGCAAGAATTCT (SEQ ID NO: 69)
AM1Jf 2	120	GTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCACCACA TGGCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCTACTTCAGCG ACGTGGACCTGGAGAAGGAC (SEQ ID NO: 70)
AM1Jf 3	91	GTGCACAGCGCCCTGATCGGCCCTGCTGGTGTGCCACACCAACACCCT GAACCCCGCCACGGCCGCCAGGTGACCCTACAAGCTTTAC (SEQ ID NO: 71)

ES 2 437 815 T3

AM1Jr 1	113	GTAAAGCTTGTAGGGTCACCTGGCGGCCGTGGGCGGGGTTACAGGGTGTG GTGTGGCACACCAGCAGGGGGCCGATCAGGCCGCTGTGCACGTCTCTCC AGGTCCACGTCG (SEQ ID NO: 72)
AM1Jr2	121	CTGAAGTAGGCCAGGCCTTGCAGTCGAACTCGTCCTTGGTGGGGCCAT GTGGTGTGCACCTTCCAGAAGTAGGICTTGGTCTCGTGGGCTTACGAA GTTCTTGCGGGGCTCGGCGC (SEQ ID NO: 73)
AM1Jr 3	93	CCTGGCGCTGGTCTCCTCGTAGCTGATCAGGCTGCTGTAGAAGCTGTAG GGGCGGTGGCCTGGTTGCGGAAGGTCACCCTACGAATTCTAC (SEQ ID NO: 74)
AM1Kf 1	120	GTAGAATTCGTAGGGTACCTGCTGAGCATGGGCAGCAACGAGAACATCC ACAGCATCCACTTCAGCGGCCACGTGTTACCCTGCGCAAGAAGGAGGAG TACAAGATGGCCCTGTACAAC (SEQ ID NO: 75)
AM1Kf 2	122	CTGTACCCCGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGG CATCTGGCGCTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGA GCACCCTGTTCTGGTGTACAG (SEQ ID NO: 76)
AM1Kf 3	102	CAACAAGTGGCAGACCCCTGGGCATGGCCAGCGGCCACATCCGCGAC TTCCAGATCACCGCCAGCGGCCAGTACGGCCAGTGGGCCCCTACAAGCTTT AC (SEQ ID NO: 77)
AM1Kr 1	123	GTAAAGCTTGTAGGGGCCACTGGCCGTAAGTGGCCGCTGGCGGTGATCTG GAAGTCGCGGATGTGGCCGCTGGCCATGCCAGGGGGTCTGGCACTTGT GCTGTACACCAGGAACAGGGTG (SEQ ID NO: 78)
AM1Kr 2	125	CTCATGCCGGCTGTCAGGTGCTCGCCGATCAGGCACTCCACGCGCCAGAT GCCGGCCTTGTGGGACAGCATCTCCACGGTCTCGAACACGCCGGGGTACAG GTTGTACAGGGCCATCTTGTACTC (SEQ ID NO: 79)
AM1Kr 3	96	CTCCTTCTTGCACCGGTGAACACGTGGCCGCTGAAGTGGATGCTGTGGA TGTTCGTGTGCTGCCATGCTCAGCAGGTACCCTACGAATTCTAC (SEQ ID NO: 80)
AM1Lf 1	120	GTAGAATTCGTAGGGGCCCAAGCTGGCCCGCTGCACTACAGCGGCA GCATCAACGCCTGGAGACCAAGGAGCCCTTCAGCTGGATCAAGGTGGAC CTGCTGGCCCCATGATCATC (SEQ ID NO: 81)
AM1Lf 2	116	CACGGCATCAAGACCCAGGGCGCCCGCAGAAAGTTCAGCAGCCTGTACA TCAGCCAGTTCATCATGTACAGCCTGGACGGCAAGAAGTGGCAGACCT ACCGCGGCAACAGCAC (SEQ ID NO: 82)
AM1Lf 3	86	CGGCACCCTGATGGTGTCTTCGGCAACGTGGACAGCAGCGGCATCAAGC ACAACATCTTCAACCCCGGGCTACAAGCTTTAC (SEQ ID NO: 83)
AM1Lr 1	110	GTAAAGCTTGTAGCCCGGGGGTGAAGATGTTGTGCTTGTATGCCGCTG CTGTCCACGTTGCCGAAGAACCATCAGGGTGCCGGTGTGTTGCCGCGG TAGGTCTGC (SEQ ID NO: 84)
AM1Lr 2	113	CACTTCTTGGCGTCCAGGCTGTACATGATGATGAACTGGCTGATGTACAG GCTGCTGAACTTCTGGCGGGCGCCCTGGGTCTTGTATGCCGTGGATGATCAT GGGGGCCAGCAG (SEQ ID NO: 85)
AM1Lr 3	99	GTCCACCTTGATCCAGCTGAAGGGCTCCTTGGTGTCCAGGCGTTGATGC TGCCGCTGTAGTGCAGGCGGGCCAGCTTGGGGGCCCTACGAATTCTAC (SEQ ID NO: 86)
AM1M f1	122	GTAGAATTCGTAGGATATCATCGCCGCTACATCCGCTGCACCCACCC ACTACAGCATCCGCAGCACCTGCGCATGGAGCTGATGGGCTGCGACCTGA ACAGCTGCAGCATGCCCTGG (SEQ ID NO: 87)
AM1M f2	112	GCATGGAGAGCAAGGCCATCAGCGACGCCAGATCACCGCCAGCAGCTA CTTCAACCAACATGTTGCCACCTGGAGCCCCAGCAAGGCCCGCTGCACCT GCAGGGCCGAG (SEQ ID NO: 88)
AM1M f3	89	CAACGCCTGGCGCCCCAGGTGAACAACCCCAAGGATGGCTGCAGGTG GACTTCCAGAAGACCATGAAGGTGACCCTACAAGCTTTAC (SEQ ID NO: 89)
AM1M r1	112	GTAAAGCTTGTAGGGTACCTTACCTGGTCTTCTGGAAGTCCACCTGCAGC CACTCCTTGGGGTGTTCACCTGGGGCGCCAGGCGTTGCTGCGGCCCTGC AGGTGCAGGCG (SEQ ID NO: 90)
AM1M r2	114	GGCCTTGTGGGGCTCCAGGTGGCGAACATGTTGGTGAAGTAGCTGCTGG CGGTGATCTGGGCGTGCCTGATGGCCTTGTCTCCATGCCAGGGGCATGC TGCAGCTGTTTAC (SEQ ID NO: 91)
AM1M r3	97	GTCGCAGCCCATCAGCTCCATGCGCAGGGTGTGCGGATGCTGTAGTGGG TGGGGTGCAGGCGGATGTAGCGGGCGATGATATCTACGAATTCTAC (SEQ ID NO: 92)

AM1Nf 1	122	GTAGAATTCGTAGGGTGACCGCGTGACCACCCAGGGCGTGAAGAGCCT GCTGACCAGCATGTACGTGAAGGAGTTCCTGATCAGCAGCAGCCAGGACG GCCACCAGTGGACCCTGTTCTTC (SEQ ID NO: 93)
AM1Nf 2	104	CAGAACGGCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCACCC CCGTGGTGAACAGCCTGGACCCCCCTGCTGACCCGCTACCTGCGCATCC ACCC (SEQ ID NO: 94)
AM1Nf 3	92	CCAGAGCTGGGTGCACCAGATCGCCCTGCGCATGGAGGTGCTGGGCTGC GAGGCCAGGACCTGTACTAGCTGCCGGGCTACAAGCTTTAC (SEQ ID NO: 95)
AM1Nr 1	118	GTAAAGCTTGTAGCCCGGGCAGCTAGTACAGGTCCTGGGCCTCGCAGCCC AGCACCTCCATGCGCAGGGCGATCTGGTGCACCCAGCTCTGGGGGTGGATG CGCAGGTAGCGGGTCAG (SEQ ID NO: 96)
AM1Nr 2	100	CAGGGGGGGGTCCAGGCTGTTACCACGGGGGTGAAGCTGTCCTGGTTGC CCTGGAACACCTTCACCTTGCCGTTCTGGAAGAACAGGGTCCACTGGTGG (SEQ ID NO: 97)
AM1Nr 3	100	CCGTCTGCTGCTGCTGATCAGGAACTCCTTCACGTACATGCTGGTTCAG CAGGCTCTTCACGCCCTGGGTGGTTCACGCCGGTACCCCTACGAATTCTAC (SEQ ID NO: 98)

**[0129]** Como se ha indicado en la Tabla 2 y mostrado en la Figura 5, el fragmento D se construyó con un sitio de restricción BamHI situado entre el sitio BglIII y el sitio HindIII en el extremo 3' del fragmento. El fragmento I se construyó para llevar el ADN de PmlI (2491) a BstEII (2661) seguido inmediatamente del ADN de BstEII (2955) a KpnI (3170), de modo que la inserción del fragmento BstEII a partir de pAMJ en el sitio BstEII de pAM1 en la orientación correcta generará las secuencias deseadas de 2491 a 3170. El plásmido pAM1B se digirió con Apal y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1A digerido con Apal y HindIII, generando el plásmido pAM1AB. El plásmido pAM1D se digirió con PmlI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1AB digerido con PmlI y HindIII, generando el plásmido pAM1ABD. El plásmido pAM1C se digirió con PmlI y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1ABD digerido con PmlI, generando el plásmido pAM1ABCD, la orientación del inserto se confirmó por la aparición de un fragmento de 111bp de diagnóstico cuando se digirió con MscI. El plásmido pAM1F se digirió con BglIII y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1E digerido con BglIII y HindIII, generando el plásmido pAM1EF. El plásmido pAM1G se digirió con KpnI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1EF digerido con KpnI y HindIII, generando el plásmido pAM1EFG. El plásmido pAM1J se digirió con BstEII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1I digerido con BstEII, generando el plásmido pAM1IJ; la orientación se confirmó por la aparición de un fragmento de 465bp de diagnóstico cuando se digirió con EcoRI y EagI. El plásmido pAM1IJ se digirió con PmlI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1H digerido con PmlI y HindIII, generando el plásmido pAM1HIJ. El plásmido pAM1M se digirió con EcoRI y BstEII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1N digerido con EcoRI y BstEII, generando el plásmido pAM1MN. El plásmido pAM1L se digirió con EcoRI y SmaI y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1MN digerido con EcoRI y EcoRV, generando el plásmido pAM1LMN. El plásmido pAM1LMN se digirió con Apal y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1K digerido con Apal y HindIII, generando el plásmido pAM1KLMN. El plásmido pAM1EFG se digirió con BamHI y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1ABCD digerido con BamHI y BglIII, generando el plásmido pAM1ABCDEFG; la orientación se confirmó por la aparición de un fragmento de 552 bp de diagnóstico cuando se digirió con BglIII y HindIII. El plásmido pAM1KLMN se digirió con KpnI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1HIJ digerido con KpnI y HindIII, generando el plásmido pAM1HIJKLMN. El plásmido pAM1HIJKLMN se digirió con BamHI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM1ABCDEFG digerido con BamHI y HindIII, generando el plásmido pAM1-1. Estas etapas de clonación están representadas en la Figura 6. La Figura 7 muestra la secuencia de ADN del inserto contenida en pAM1-1 (SEQ ID N°:1). Se puede clonar este inserto en cualquier vector de expresión adecuado como un fragmento NheI-SmaI para generar un constructo de expresión. pXF8.61 (Fig. 4), pXF8.38 (Fig. 11) y pXF8.224 (Fig. 13) son ejemplos de este tipo de constructo.

#### Construcción de pXF8.186

**[0130]** La versión "LE" del ADNc optimizado del FVIII con el dominio B suprimido contenido en el pAM1-1 se modificó por la sustitución del dipéptido Leu-Glu (2284-2289) en la unión de las cadenas pesada y ligera con cuatro residuos de Arginina, haciendo un total de cinco residuos de Arginina consecutivos (SEQ ID N°:2). Se logró como se describe a continuación. Los seis oligonucleótidos que se muestran en la Tabla 4 se hibridaron, ligaron, digirieron con EcoRI y HinIII y clonaron en pUC18 digerido con EcoRI y HindIII, generando el plásmido pAM8B. La Figura 8 muestra cómo estos oligonucleótidos se hibridan para formar la secuencia de ADN necesaria. pAM8B se digirió con BamHI y BstXI y el inserto de 230 bp se purificó mediante electroforesis y se utilizó para sustituir el fragmento BamHI (2126)-BstXI (2352)

de la versión "LE" (Véase Figura 7). La Figura 9 muestra la secuencia del ADNc resultante (SEQ ID N°:2). Esta versión "5Arg" del ADNc optimizado del FVIII con el dominio B suprimido puede clonarse en un vector de expresión como un fragmento NheI-SmaI para generar un constructo de expresión. pXF8.168 (Figura 3) es un ejemplo de ese tipo de constructo.

5

Tabla 4

NOMBRE DE OLIGO'	LONGITUD DE OLIGO'	SECUENCIA DE OLIGONUCLEÓTIDOS
AM8F1	140	GTAGAATTCGGATCCTGGGCTGCCACAACAGCGACTT CCGCAACCGCGGCATGACCGCCCTGCTGAAGGTGAGC AGCTGCGACAAGAACACCGGCGACTACTACGAGGAC AGCTACGAGGACATCAGCGCCTACCTGCTG (SEQ ID NO:99)
AM8BF2	57	AGCAAGAACAACGCCATCGAGCCCCGCAGGCGCAGG CGCGAGATCACCCGCACCACC (SEQ ID NO:100)
AM8F4	58	CTGCAGAGCGACCAGGAGGAGATCGACTACGACGAC ACCATCAGCGTGAAGCTTTAC (SEQ ID NO:101)
AM8R1	79	GTAAAGCTTCCACGCTGATGGTGTCTCGTAGTCGAT CTCCTCCTGGTCTGCTCTGCAGGGTGGTGCAGGCTGATCT CGCG (SEQ ID NO:102)
AM8BR2	57	CCTGCGCCTGCGGGGCTCGATGGCGTTGTTCTTGCTCA GCAGGTAGGCGCTGATGTC (SEQ ID NO:103)
AM8BR4	119	CTCGTAGCTGTCTCGTAGTAGTCGCCGGTGTCTTGT CGCAGCTGCTCACCTTCAGCAGGGCGGTCATGCCGCG GTTGCGGAAGTCGCTGTTGTGGCAGCCAGGATCCGA ATTCTAC (SEQ ID NO:104)

#### Construcción de pXF8.36

[0131] El constructo para la expresión del Factor VIII humano, pXF8.36 (Fig.10), es un plásmido de ADN circular de 11,1 kilobases que contiene los siguientes elementos: una región flanqueante 5' del gen temprano inmediato del citomegalovirus (CMV) I compuesta de una secuencia promotora, una secuencia no traducida 5' (5'UTS) y la primera secuencia de intrón para la iniciación de la transcripción del ADNc del Factor VIII. La región del CMV está fusionada próxima a una secuencia de ADNc del Factor VIII con el dominio B suprimido de tipo salvaje. La secuencia de ADNc del Factor VIII se fusiona, en el extremo 3', con un fragmento de 0,3 kb de la secuencia no traducida 3' de la hormona humana del crecimiento. Una secuencia no traducida 3' (3'UTS) y señal de terminación de la transcripción del gen de la hormona humana del crecimiento se utiliza para asegurar el procesamiento del mensaje que sigue inmediatamente al codón de terminación. Un gen marcador de selección (el gen (*neo*) bacteriano de neomicina fosfotransferasa) se inserta corriente abajo del ADNc del Factor VIII para permitir la selección para células de mamíferos transfectadas de modo estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por el promotor temprano del virus de simio 40 (SV40). El amplicón basado en pUC 19 portador del pBR322 derivado de una  $\beta$ -lactamasa (*amp*) y origen de replicación (*ori*) permite la captación, selección y propagación del plásmido en las cepas K-12 de *E. coli*. Esta región se derivó a partir del plásmido pBSII SK+.

#### Construcción de pKF8.38

[0132] El constructo para la expresión del Factor VIII humano, pXF8.38 (Fig.11) es un plásmido de ADN circular de 11,1 kilobases que contiene los siguientes elementos: una región flanqueante 5' del gen temprano inmediato del citomegalovirus (CMV) I compuesta de una secuencia promotora, una secuencia no traducida 5' (5'UTS) y la primera secuencia de intrón para la iniciación de la transcripción del ADNc del Factor VIII. La región del CMV está fusionada próxima a una secuencia de ADNc del Factor VIII con el dominio B suprimido sintética configurada de manera óptima. La secuencia de ADNc del Factor VIII se fusiona, en el extremo 3', con un fragmento de 0,3 kb de la secuencia no traducida 3' de la hormona humana del crecimiento. Una secuencia no traducida 3' (3'UTS) y señal de terminación de la transcripción del gen de la hormona humana del crecimiento se utiliza para asegurar el procesamiento del mensaje que sigue inmediatamente al codón de terminación. Un gen marcador de selección (el gen (*neo*) bacteriano de neomicina fosfotransferasa) se inserta corriente abajo del ADNc del Factor VIII para permitir la selección de células de mamíferos transfectadas de modo estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por

el promotor temprano del virus de simio 40 (SV40). El amplicón basado en pUC 19 portador del pBR322 derivado de una  $\beta$ -lactamasa (amp) y origen de replicación (ori) permite la captación, selección y propagación del plásmido en las cepas K-12 de *E. coli*. Esta región se derivó a partir del plásmido pBSII SK+.

Construido de pXF8.269

5 **[0133]** El constructo para la expresión del Factor VIII humano, pXF8.269 (Fig.12) es un plásmido de ADN circular de 14,8 kilobases que contiene los siguientes elementos: un promotor de colágeno  $\alpha$ -2 (I) humano que contiene 0,17 kb de la secuencia no traducida 5' (5'UTS). La secuencia no traducida 5' (5'UTS) del gen de la Aldolasa A y la primera secuencia de intrón para la iniciación de la transcripción del ADNc del Factor VIII. La región de intrón de la aldolasa está fusionada próxima a una secuencia de ADNc del Factor VIII con el dominio B suprimido sintética de tipo salvaje. Una  
10 secuencia no traducida 3' (3'UTS) y señal de terminación de la transcripción del gen de la hormona humana del crecimiento se utiliza para asegurar el procesamiento del mensaje que sigue inmediatamente al codón de terminación. Un gen marcador de selección (el gen (*neo*) bacteriano de neomicina fosfotransferasa (*neo*)) se inserta corriente abajo del ADNc del Factor VIII para permitir la selección de células de mamíferos transfectadas de modo estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por el promotor SV40. El amplicón basado en  
15 pUC 19 portador del pBR322 derivado de una  $\beta$ -lactamasa (amp) y origen de replicación (ori) permite la captación, selección y propagación del plásmido en las cepas K-12 de *E. coli*. Esta región se derivó a partir del plásmido pBSII SK+.

Construido de pXF8.224

20 **[0134]** El constructo para la expresión del Factor VIII humano, pXF8.224 (Fig.13) es un plásmido de ADN circular de 14,8 kilobases que contiene los siguientes elementos: un promotor de colágeno  $\alpha$ -2 (I) humano que contiene 0,17 kb de la secuencia no traducida 5' (5'UTS). La secuencia no traducida 5' (5'UTS) del gen aldolasa A y la primera secuencia de intrón para la iniciación de la transcripción del ADNc del Factor VIII. La región de intrón de aldolasa está fusionada próxima a una secuencia de ADNc del Factor VIII eliminado del dominio B sintética de manera óptima. Una secuencia  
25 no traducida 3' (3'UTS) y señal de terminación de la transcripción del gen de la hormona del crecimiento humana se utiliza para asegurar el procesamiento del mensaje que sigue inmediatamente al codón de terminación. Un gen marcador de selección (el gen bacteriano de neomicina fosfotransferasa) se inserta corriente abajo del ADNc del Factor VIII para permitir la selección de células de mamíferos transfectadas de modo estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por el promotor SV40. El amplicón basado en  
30 pUC 19 portador del pBR322 derivado de una  $\beta$ -lactamasa (amp) y origen de replicación (ori) permite la captación, selección y propagación del plásmido en las cepas K-12 de *E. coli*. Esta región se derivó a partir del plásmido pBSII SK+.

Ensayo de coagulación

35 **[0135]** Un ensayo de coagulación basado en un tiempo parcial de tromboplastina activado (aPTT) (Proctor, *et al.*, *Am. J. Clin. Path.*, 36:212-219, (1961)) se llevó a cabo para analizar la actividad biológica de las moléculas del Factor VIII humano con el dominio B suprimido expresadas por los constructos en los que se optimizó la región que codifica el BDD hFVIII.

Actividad biológica analizada empleando el ensayo de coagulación

40 **[0136]** Los resultados del ensayo de coagulación basado en aPTT se presentan en la Tabla 5, a continuación. La actividad específica de las preparaciones de hFVIII se presentan como unidades de aPTT por miligramo de proteína de hFVIII determinadas por ELISA. Ambas moléculas de BDD hFVIII derivadas de fibroblasto humano (5R y LE) tienen actividad alta específica cuando se mide el ensayo de coagulación de aPTT. Se ha determinado que estas actividades específicas sean de 2 a 3 veces más altas que las determinadas por el FVIII de longitud completa derivado de células CHO (como se muestra en la Tabla 5). Una media de múltiples determinaciones de actividades específicas para  
45 diversas preparaciones parcialmente purificadas de BDD hFVII de LE y de 5R también muestra valores consistentemente más altos para las moléculas de hFVIII BDD (11.622 unidades/mg y 14.561 unidades/mg para BDD hFVIII de LE en comparación con 7097 unidades/mg para FVIII derivado de células CHO de longitud completa). Se ha observado una ratio y/o extensión aumentada de activación de la trombina para diversas moléculas de hFVIII BDD, posiblemente debido a un efecto del dominio B para proteger las cadenas ligera y pesada de la escisión y activación de la trombina (Eaton *et al.*, *Biochemistry*, 25:8343-8347, (1986), Meulien *et al.*, *Protein Engineering*, 2:301-306, (1988)).

Tabla 5. Actividades específicas de diversas proteínas de hFVIII

Producto del hFVIII	Concentración por ELISA (mg/mL)	Actividad en aPTT (aPTT U/mL)	Actividad específica (aPTT U/mg)
BDD hFVIII de 5R	0,050	1306	26.120
BDD hFVIII de LE	0,124	2908	23.452
FVIII (derivado de CHO) de longitud completa	0,158	1454	9202

Ensayo para Factor Humano VIII en sobrenadantes de cultivo de célula transfectada

5 **[0137]** Las muestras de cultivo celular, sobrenadantes que tienen células transfectadas con el tipo salvaje, o el Factor VIII humano con el dominio B suprimido optimizado se ensayaron para determinar el contenido de Factor VIII humano (hFVIII) empleando un análisis de inmunoabsorción enzimática (ELISA). Este ensayo se basa en el uso de dos anticuerpos monoclonales (mAb) de reacción no cruzada en combinación con muestras de medios de cultivo celular recogidos a partir de sobrenadantes de células transfectadas de fibroblasto humano. Los métodos de transfección e identificación de células transfectadas de modo positivo se describen en la Patente de EE. UU. núm. 5.641.670.

Tabla 6

Plásmido	Secuencia no traducida 5'/promotor	Composición del ADNc del Factor VIII	Media (FVIII mU / 10 <sup>6</sup> Células / 24 horas)	Máximo (FVIII mU / 10 <sup>6</sup> Células / 24 horas)	Número de cepas	Aumento de X veces
pXF836	CMV IE1	Tipo salvaje	567	2557	38	-
pXF8.38	CMV IE1	Configuración óptima	5403	17106	24	9.5X
pXF8.269	Intrón de la aldolasa/ de colágeno $\alpha$ -2 (I)	Tipo salvaje	382	1227	18	-
pXF8.224	Intrón de la aldolasa/ de colágeno $\alpha$ -2 (I)	Configuración óptima	2022	11930	218	5.3X

Unidades ELISA basadas en curvas estándar preparadas de mezcla de plasma normal

10

II. Constructos del Factor IX y usos de los mismos (Ejemplo comparativo)Construcción de gen sintético que codifica el Factor IX de coagulación.

**[0138]** Los cuatro fragmentos enumerados en la Tabla 7 y mostrados en la Figura 14 se realizaron mediante síntesis de oligonucleótidos automática y se clonaron en el plásmido pBS para generar cuatro plásmidos, PFIXA a través de pFIXD.

15

Tabla 7

Fragmento	Extremo 5'	Extremo 3'
A	BamHI 1	StuI(/FspI) 379
B	(StuI/)FspI 379	PfIMI 810
C	PfIMI 810	PstI 1115
D	PstI 1115	BamHI 1500

20

**[0139]** Como se muestra en la figura 14, se utilizaron los plásmidos pFIXA a través de pFIXD para construir pFIXABCD, que porta el gen sintético completo. El fragmento A se sintetizó con un sitio PstI 3' al sitio StuI, y se clonó como un fragmento BamHI-PstI. El plásmido pFIXD se digirió con PstI y HindIII, y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pFIXA digerido con PstI y HindIII, generando el plásmido pFIXAD. El plásmido pFIXB se digirió con EcoRI y PflMI, y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pFIXC digerido con EcoRI y PflMI, generando el plásmido pFIXBC. El plásmido pFIXBC se digirió con FspI y PstI, y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pFIXAD digerido con StuI y PstI, generando el plásmido pFIXABCD.

25

**[0140]** La Figura 15 muestra la secuencia de ADN del inserto BamHI contenida en pFIXABCD. Este inserto se puede clonar en un vector de expresión adecuado como un fragmento BamHI para generar un constructo de expresión. Este ejemplo ilustra como un sitio de fusión se puede utilizar en la construcción, incluso cuando existe una secuencia idéntica muy cerca (los fragmentos A, B y D contienen el hexámero "AGGGCA", el producto de ligadura de extremos romos del ADN digerido con StuI-FspI). Esto es posible porque las enzimas de restricción usadas para crearlos no cortan los sitios de fusión resultantes. Este ejemplo también muestra como los fragmentos génicos pueden sintetizarse con sitios de restricción adicionales fuera de la secuencia génica real, y estos sitios pueden utilizarse para facilitar las etapas de clonación intermedias.

30

Expresión del Factor IX humano a partir de ADNc optimizado y no optimizado

[0141] El constructo para la expresión del Factor IX humano (Figura 16), pXIX76, es un plásmido de ADN circular de 8,4 kilobases (kb) que contiene los siguientes elementos: una región flanqueante 5' del gen temprano inmediato del citomegalovirus (CMV) I compuesta de una secuencia promotora, una secuencia no traducida 5' (5'UTS) y una primera secuencia de intrón (equivalente a los nucleótidos 174328 – 172767 de la colección del Genbank X17403). La región del CMV está fusionada próxima a una secuencia de ADNc del Factor IX de tipo salvaje, con un sitio BamHI en la unión. La secuencia de ADNc del Factor IX está fusionada próxima a un fragmento de 1,5 kb de la región 3' del gen del Factor IX que incluye la señal de terminación de la transcripción (equivalente a los nucleótidos 34335 – 35857 de la colección del Genbank K02402). Un gen marcador de selección (el gen (*neo*) bacteriano de neomicina fosfotransferasa) se inserta corriente arriba de las secuencias del CMV para permitir la selección de células de mamíferos transfectadas de modo estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por el promotor de la timidina cinasa del virus del herpes simple. El casete de expresión neo es equivalente a los nucleótidos 452-1596 de la colección del Genbank U43612. El amplicón basado en pUC 19 portador del pBR322 derivado de una β-lactamasa y origen de replicación permite la selección y propagación del plásmido en *E. coli*.

[0142] El plásmido pXIX170 que contiene una región de codificación del Factor IX con una configuración optimizada se puede derivar a partir del pXIX76 mediante la digestión con BamHI y BclI e inserción del fragmento BamHI mostrado en la Figura 15, de este modo se produce un constructo equivalente que dirige la expresión del Factor IX humano a partir de un ADNc optimizado.

[0143] Se ensayaron muestras de sobrenadantes de cultivo celular procedentes de clones de fibroblasto de prepucio humano normal transfectados con constructos de expresión optimizada o de tipo salvaje para determinar la expresión del Factor IX. Como se muestra en la Tabla 8, podría demostrarse un aumento de 2,7 veces en la expresión media del Factor IX cuando el ADNc optimizado fue sustituido para la secuencia de tipo salvaje.

**Tabla 8:** Datos de expresión para las cepas que expresan el Factor IX

Plásmido	Secuencia no traducida 5'/ Promotor	Composición del ADNc	Media	Máximo	Número de cepas de células
			Nanogramos/10 <sup>6</sup> células/24horas		
pXIX76	CMV	De tipo salvaje	418	8384	144
pXIX170	CMV	Configuración óptima	1127	3316	33

III. Constructos de alfa-galactosidasa y usos de los mismos

25 Construcción de un gen sintético que codifica α-galactosidasa

[0144] Los cuatro fragmentos de genes enumerados en la Tabla 9 se realizaron mediante síntesis de oligonucleótidos automática y se clonaron en el vector pUC18 como los fragmentos EcoRI – HindIII (con el extremo N-terminal de cada fragmento génico adyacente al sitio EcoRI) para generar cuatro plásmidos, pAM2A a través de pAM2D.

**Tabla 9**

Fragmento	Extremo 5'			
A	BamHI	1	PstI	364
B	PstI	364	Bg1II(BamHI)	697
C	(Bg1II/BamHI)	697	SmaI(/StuI)	1012
D	(SmaI/)StuI	1012	XhoI	1347

[0145] Los plásmidos pAM2A a través de pAM2D se utilizaron para construir pAM2ABCD, que porta el gen sintético completo. El plásmido pAM2B se digirió con PstI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM2A digerido con PstI y HindIII, generando el plásmido pAM2AB. El plásmido pAM2D se digirió con StuI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM2C digerido con SmaI y HindIII, generando el plásmido pAM2CD. El plásmido pAM2CD se digirió con BamHI y HindIII y el inserto se purificó mediante electroforesis en gel de agarosa y se insertó en el plásmido pAM2AB digerido con Bg1II y HindIII, generando el plásmido pAM2ABCD.

[0146] La Figura 17 muestra la secuencia de ADN del fragmento BamHI-XhoI contenido en pAM2ABCD. Este inserto se puede clonar en un vector de expresión adecuado como un fragmento BamHI - XhoI para generar un constructo de expresión. Este ejemplo ilustra el uso de los sitios de fusión que surgen a partir de la ligadura de dos salientes complementarios (BgIII/BamHI) y a partir de la ligadura de extremos romos (SmaI/StuI).

Expresión de  $\alpha$ -galactosidasa humana a partir de ADNc optimizado y no optimizado

**[0147]** El constructo para la expresión de  $\alpha$ -galactosidasa humana, plásmido pXAG94 (Figura 18), es un plásmido de ADN circular de 8,5 kb que contiene los siguientes elementos. Un gen marcador de selección (el gen (*neo*) bacteriano de neomicina fosfotransferasa) se inserta corriente arriba del casete de expresión de  $\alpha$ -galactosidasa para permitir la selección de células de mamíferos transfectadas de manera estable empleando el análogo de neomicina G418. La expresión del gen *neo* está controlada por el promotor temprano SV40. En concreto, el fragmento PvuII – HindIII de 342 bp equivalente a los nucleótidos 273 – 1/5243 – 5172 de la colección del Genbank J02400 está fusionado a través de una unión XhoI a un fragmento equivalente a los nucleótidos 502 – 561 de la colección del Genbank J02400, que está fusionado a la región de codificación *neo*, equivalente a los nucleótidos 350 – 1322 de la colección del Genbank U13862. Las señales de poliadenilación para este casete de expresión las proporcionan las secuencias 3393 – 3634 de SYNPRSVNEO. Este marcador de selección se fusiona a una secuencia de plásmido corta, equivalente a los nucleótidos 2067 (PvuII) – 2122 de SYNPR322.

**[0148]** La expresión del ADNc de  $\alpha$ -galactosidasa se dirige desde un potenciador de CMV (equivalente a los nucleótidos 174253 – 173848 de la colección del Genbank X17403). Este ADN se fusiona a través de la secuencia de unión TCGACAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAG (SEQ ID N°:107) a secuencias del factor humano de elongación 1 alfa que se extienden de los nucleótidos -207 al +982 en relación con el sitio caperuza. Estas secuencias proporcionan el promotor del EF1 alfa, el sitio CAP y un intrón de 943 nucleótidos presente en las secuencias no traducidas 5' de este gen. El ADN se fusiona próximo a la secuencia de unión GAATTCTCTAGATCGAATTCCTGCAGCCCGGGGATCCACC (SEQ ID N°:108) seguida inmediatamente de 335 nucleótidos del gen de la hormona humana del crecimiento, empezando con el codón iniciador ATG, equivalente a los nucleótidos 5225 – 5559 de la colección del Genbank J03071. Este ADN codifica para el péptido señal del gen hGH, incluyendo el primer intrón.

**[0149]** Este ADN se fusiona a la parte del ADNc de  $\alpha$ -galactosidasa de tipo salvaje que codifica para los aminoácidos 31 a 429. La región de codificación se fusiona próxima a través de la unión AAAAAAAAAAAAACTCGAGCTCTAG (SEQ ID N°:109) a la región no traducida 3' del gen hGH, que se corresponde con los nucleótidos 6699 – 7321 de la colección del Genbank J03071. Finalmente, este ADN se une al amplicón basado en pUC portador del gen de pBR322 derivado de una  $\beta$ -lactamasa y el origen de la replicación que permite la selección y propagación del plásmido en *E. coli*; las secuencias son equivalentes a los nucleótidos 229 -1/2680 – 281 de SYNPU12V.

**[0150]** El plásmido pXAG95 es equivalente al pXAG94, con la secuencia de ADNc de  $\alpha$ -galactosidasa sustituida con la secuencia de configuración optimizada correspondiente (que codifica para los aminoácidos 31 a 429) de la Figura 17.

**[0151]** El plásmido pXAG73 (Figura 19) es un plásmido de 10 kb similar al pXAG94, pero con las siguientes diferencias. La secuencia de unión GCCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAG (SEQ ID N°: 110) y el ADN de EF1 alfa adyacente hasta + 30 más allá del sitio de caperuza han sido sustituidas con el promotor de la metalotioneína de ratón y el sitio caperuza (nucleótidos -1752 al +54 en relación con el sitio caperuza mMTI). También la unión de la UTS de EF1 $\alpha$  a la secuencia de codificación hGH difiere: las secuencias de EF1 $\alpha$  se extienden hasta +973 a partir del sitio de caperuza EF1 $\alpha$ , seguidas por la unión CTAGGATCCACC (SEQ ID N°:111), en lugar de la unión GAATTCTCTAGATCGAATTCCTGCAGCCCGGGGATCCACC (SEQ ID N°:108) descrita anteriormente.

**[0152]** El plásmido pXAG74 es equivalente al pXAG73, con la secuencia de ADNc de  $\alpha$ -galactosidasa de tipo salvaje sustituida por la secuencia de configuración optimizada correspondiente (que codifica los aminoácidos 31 a 429) de la Figura 17.

**[0153]** La construcción de dichos plásmidos, que incluye la creación de fusiones de  $\alpha$ -galactosidasa de hGH, se describe en la Patente de EE.UU n°. 6.083.725.

**[0154]** Se ensayaron muestras de sobrenadantes de cultivo celular de clones de fibroblasto de prepucio humano normal transfectados con constructos de expresión optimizada o de tipo salvaje para determinar la expresión de  $\alpha$ -galactosidasa.

**Tabla 10:** Datos de expresión para las cepas que expresan la alfa-galactosidasa

Plásmido	Secuencia no traducida 5'/ Promotora	Composición del ADNc	Media	Máximo	Número de cepas de células
			Unidades/10 <sup>6</sup> células/24horas		
pXAG-73	CMV/mMT/EF1a	De tipo salvaje	323	752	12
pXAG-74	CMV/mMT/EF1a	Configuración óptima	1845	8586	27
pXAG-94	CMV/EF1a	De tipo salvaje	417	1758	39

pXAG-95	CMV/EF1a	Configuración óptima	842	3751	75
---------	----------	-------------------------	-----	------	----

5 **[0155]** Como se muestra en la Tabla 10, se observaron aumentos de 5,7 veces y uno de 2,0 veces en la expresión de la  $\alpha$ -galactosidasa media cuando se expresaba el ADNc optimizado procedente de los promotores de EF1a (PXAG-95) y mMT1 (PXAG-74), respectivamente, cuando se compara con las secuencias de codificación de tipo salvaje. Además, también se observaron incrementos significativos en la expresión máxima cuando se expresó el ADNc optimizado procedente de cualquier promotor.

**Reivindicaciones**

- 5
1. Un ácido nucleico sintético que codifica alfa-galactosidasa y que difiere en la secuencia de un ácido nucleico de alfa-galactosidasa de tipo salvaje en la medida en que al menos un codón no común o un codón menos común ha sido reemplazado por un codón común, y donde la secuencia de ácido nucleico sintético presenta al menos una o más de las siguientes propiedades:
- (a) presenta una longitud continua de al menos 90 codones, todos ellos codones comunes
- (b) presenta una longitud continua de codones comunes que comprenden al menos el 60% de los codones de la secuencia de ácidos nucleicos sintéticos;
- 10 (c) al menos el 94% o más de los codones en la secuencia de ácido nucleico sintético que codifica la proteína son codones comunes y la secuencia de ácido nucleico sintético codifica una proteína de al menos aproximadamente 90 aminoácidos de longitud;
- (d) presenta una longitud de al menos 80 pares de base,
- 15 en el que por codón común se entiende Ala (gcc), Arg (cgc), Asn (aac), Asp (gac), Cys (tgc), Gln (cag), Gly (ggc), His (cac), Ile (atc), Leu (ctg), Lys (aag), Pro (ccc), Phe (ttc), Ser (agc), Thr (acc), Tyr (tac), Glu (gag) y Val (gtg);
- en el que los codones menos comunes son: Gly (ggg), Ile (att), Leu (ctc), Ser (tcc), Val (gtc) y Arg (agg); y
- en el que todos los codones que no se consideren comunes o menos comunes son codones no comunes.
2. El ácido nucleico sintético de la reivindicación 1 en el que se inserta el ácido nucleico de la alfa-galactosidasa en una célula no transformada.
- 20 3. El ácido nucleico sintético de la reivindicación 1, en el que el número de codones no comunes o menos comunes restantes es menor de 15.
4. El ácido nucleico sintético de la reivindicación 1, en el que los codones no comunes o menos comunes se reemplazan por codones comunes.
5. Un vector que comprende el ácido nucleico sintético de la reivindicación 1, 3 o 4.
- 25 6. Una célula que comprende el vector de la reivindicación 5.
7. Un método para producir alfa-galactosidasa que comprende el cultivo de la célula de la reivindicación 6 en condiciones en las que se expresa el ácido nucleico.
8. Un método para preparar un ácido nucleico que codifica alfa-galactosidasa cuya longitud alcanza al menos 90 codones, que comprende:
- 30 (a) identificar un codón no común y un codón menos común en una secuencia de genes no optimizada que codifica una proteína de alfa-galactosidasa; y
- (b) reemplazar al menos el 94% de los codones no comunes y menos comunes por un codón común que codifica el mismo aminoácido que el codón sustituido;
- 35 en el que por codón común se entiende Ala (gcc), Arg (cgc), Asn (aac), Asp (gac), Cys (tgc), Gln (cag), Gly (ggc), His (cac), Ile (atc), Leu (ctg), Lys (aag), Pro (ccc), Phe (ttc), Ser (agc), Thr (acc), Tyr (tac), Glu (gag) y Val (gtg);
- en el que los codones menos comunes son Gly (ggg), Ile (att), Leu (ctc), Ser (tcc), Val (gtc) y Arg (agg); y
- en el que todos los codones que no se consideren codones comunes ni codones menos comunes son codones no comunes.
- 40 9. El método de la reivindicación 8, en el que al menos el 96% de los codones no comunes y menos comunes son reemplazados por un codón común que codifica el mismo aminoácido que el codón sustituido.
10. El uso de un ácido nucleico sintético de la reivindicación 1 que permite la síntesis de un mensaje optimizado para alfa-galactosidasa en la preparación de un medicamento para el tratamiento de una enfermedad

**caracterizada por** déficit de la alfa-galactosidasa en un sujeto, en el que se introduce el ácido nucleico sintético en el sujeto y el sujeto expresa la alfa-galactosidasa.

- 5
11. Un ácido nucleico sintético de la reivindicación 1 que permite la síntesis de un mensaje optimizado para alfa-galactosidasa para su uso en el tratamiento de una enfermedad **caracterizada por** déficit de la alfa-galactosidasa en un sujeto, en el que se introduce el ácido nucleico sintético en el sujeto y el sujeto expresa la alfa-galactosidasa.
12. El uso de la reivindicación 10 o del ácido nucleico sintético de la reivindicación 11, en el que se introduce el ácido nucleico sintético en una célula.
- 10
13. El uso de la reivindicación 10 o del ácido nucleico sintético de la reivindicación 11, en el que la célula puede ser una célula xenogénica, alogénica o autóloga.
14. El uso de la reivindicación 10 o del ácido nucleico sintético de la reivindicación 11, en el que al menos el 98% o el conjunto de todos los codones de la secuencia de ácido nucleico sintético son codones comunes.
- 15
15. El uso de la reivindicación 10 o del ácido nucleico sintético de la reivindicación 11, en el que el sujeto padece la enfermedad de Fabry.
16. El ácido nucleico sintético de la reivindicación 1, en el que el ácido nucleico sintético comprende la SEQ ID N°: 106.

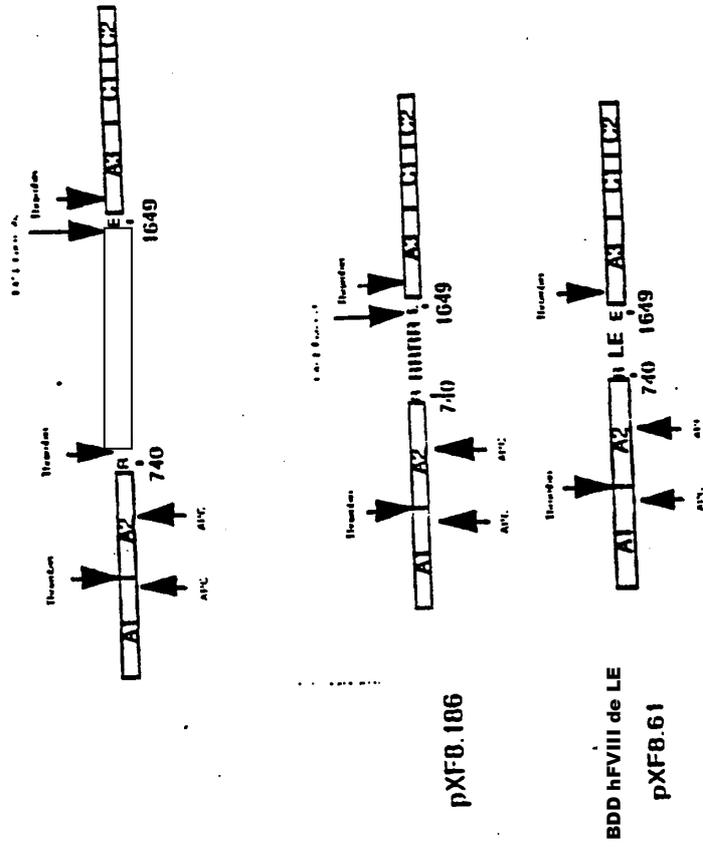
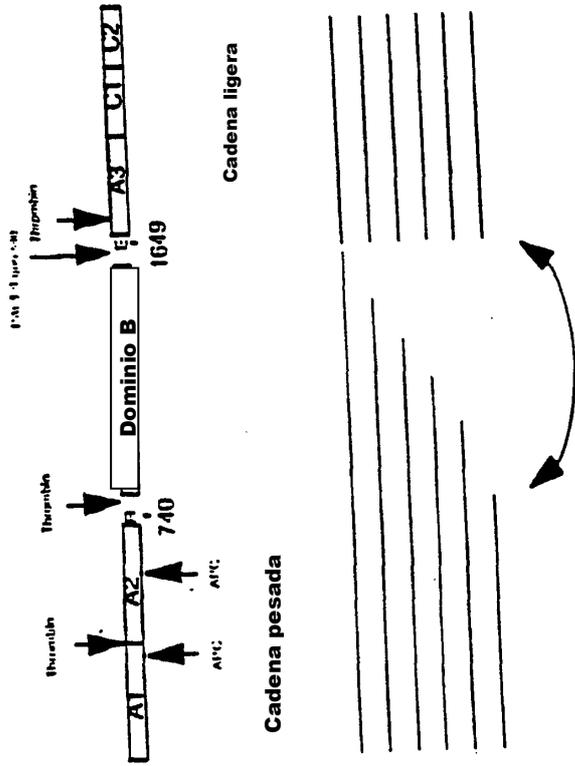


FIG. 1



**Heterogeneidad de hFVIII se debe a la proteólisis con el dominio B**

**FIG. 2**

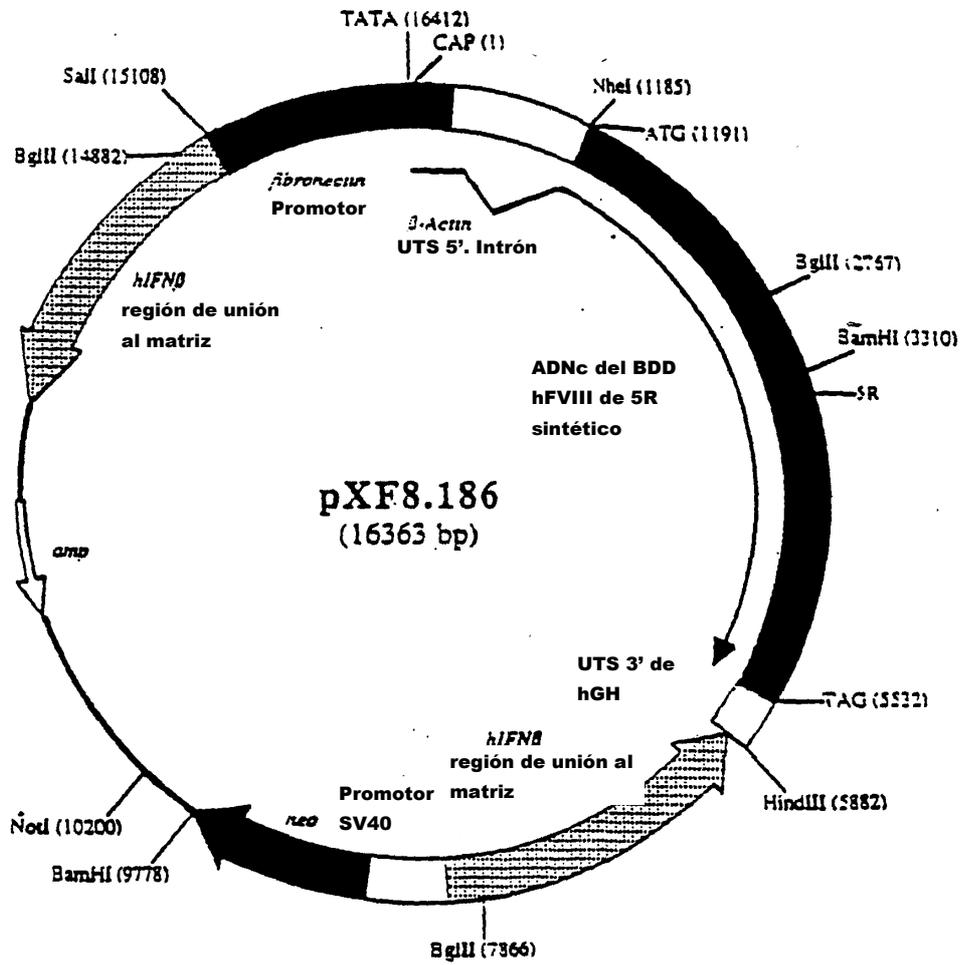


FIG. 3

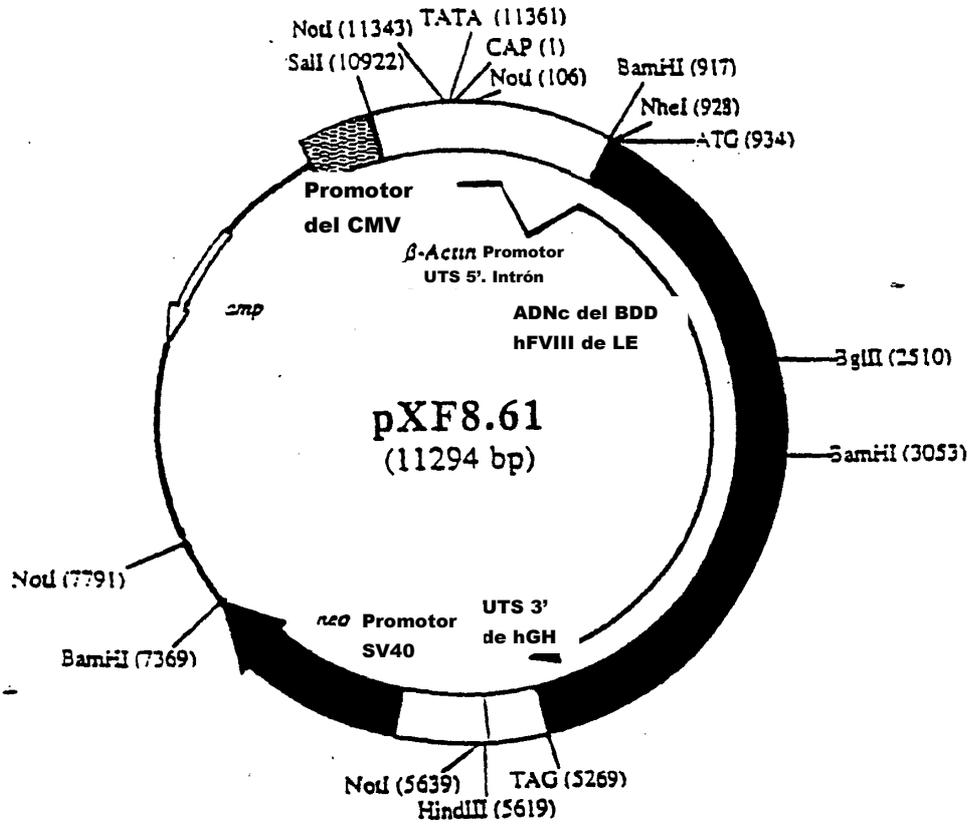


FIG. 4



Frammento B

```

EcoRI|
GTAGAAATTCGTAGGGCCCCACCATTCAGGCGGAGGTGTACGACACCCGTGGTGTATCACCCCTGAAGAACATGGCCAG
CATCTTAAGCATCCCCGGGTGGTAGGATCCGGCTCCACATGCTGTGGCACCCACTAGTGGGACTTCTTGTACCCGGTC
AM1 Br3
3' OH 5' P
CCACCCCGTGAGCCTTGCCACGCCGTGGGGTGAGCTACTG GAAGGCCAGGCTGATGCGCCGAGTACGACGACCCAGA
GGTGGGCACTCG GACGTGCATATTCGCACTCGATGAC CTTCGGGTCCGTCGCCGGCTCATGCTGCTGCTCT
5' P 3' OH
AM1 Br2
3' OH 5' P
CCAGCCAGCGCGGAGAGGAGGACGACAGGTGTTCCCGGTCGGCAGCCACACCTACTGTTGGCAGGTG CTGAAG
GGTCGGTCCGGCTCTTCCCTCTGCTGTTCCACNAGGGGCG GCGGTGGTGTGGATGCCACACCCGTCCAC_GACTTC
5' P 3' OH
AM1 Br1
PmlI
GAGAACGGCCCCCATGGCCAGCGACCCCTGTGCCCTACAGCTACCTAGCCACCGTGTACAAAGCTTTAC
CTCTTGGCCGGGGTACCCTGCTGGGGACACGGACTGGATGGATGGACTCGGTGCCAGGATGTTCCGAATG
AM1 Br1
PmlI HindIII

```

FIG. 5 (2 de 14)

Frammento C

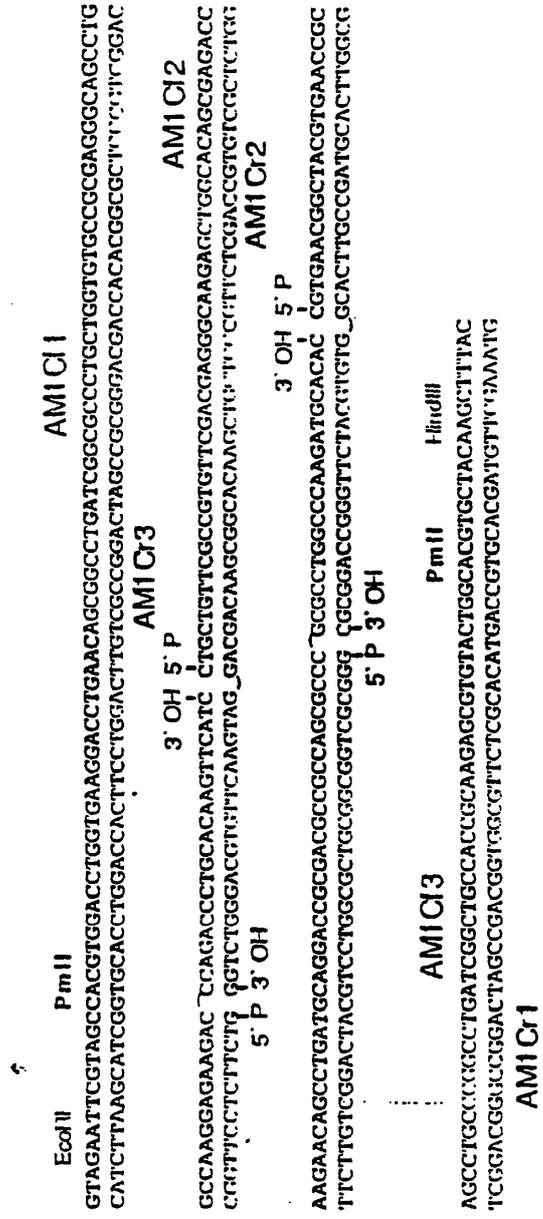


FIG. 5 (3 de 14)

Frammento D

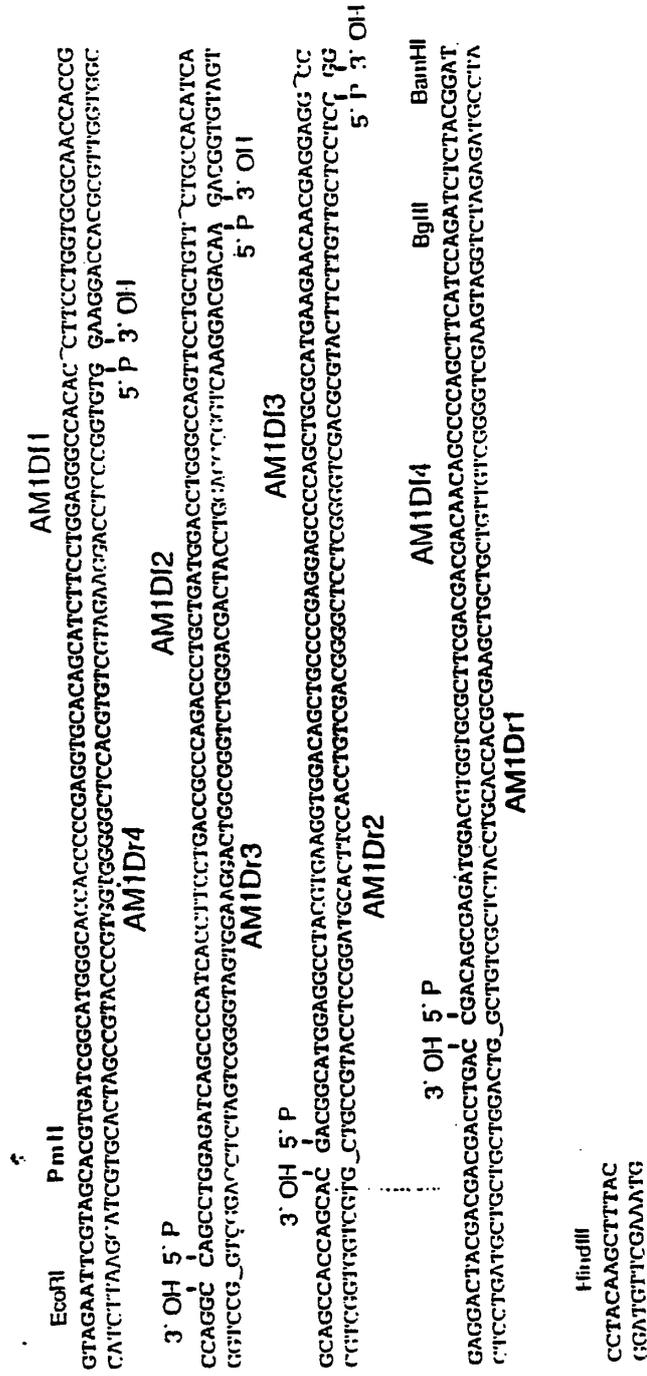


FIG. 5 (4 de 14)



Fraamento F

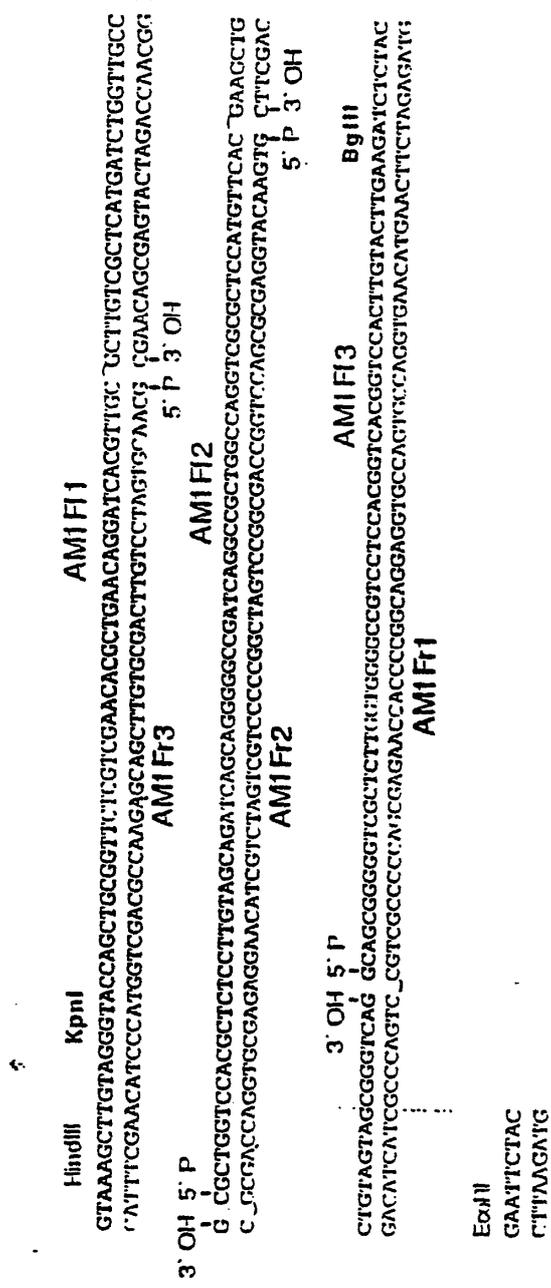


FIG. 5 (6 de 14)

Frammento G

```

EcoRI      KpnI      AM1G11
GTAGAATTCGTAGGGTACCTGACC GGAGAACATCCAGCGCTTCCTGCCCAACCCCGCGGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAG
CATCTTAAAGCATCCCN TGGACTGGCTCTTTGTAATGTCCGCGAAGGACGGGTGGGGCGGCCCAATTCGACCTCCTGGGGCTTAAATTCCTTC
                                     AM1Gr3
CAACATCATCCACAGCATCAACGGCTAC GTGTTGGACAGCCTGCAGCTGAGCGTGTGGCTGCACCGAGGTGGCC TACTGGTACATCCTGAG
GTATGTA GTACGGTGTCTGTAGTTGCCGATG_CACAAAGCTGTCCGACGTGCACACATGTAATTTGTTCCACCCGGATGACCATTGTAGGACTC
5' P 3' OH
                                     AM1Gr2
CATCGGGCCAGACCCGACTTCCAGCGGTGTTCTTACGCTGGTACACCTCAAGCACAAAGTGTGTGTACCGAGGACACCCCTGACCCCTGT
GTAGCCCGGGTCTGGCTGAGGACTCGCACAAAGATCG CCGATGTGGAAGTTCTGTGTTTAC_CACATGCTCCTGTGGGACTGGGACAA
5' P 3' OH
                                     AM1Gr2
                                     BamHI      HindIII
CCCCCTCAGGCTTGGACCGGTTCATGAGCA TGGAGAACCCCGCCCTGTGGATCCCTACAAGCTTTAC
GTGGAGTCCGCCCTCTGGCACAAAGTACTTTTACCCTCTTGGGGCCGGACACCTAGGGATGTTCCTAAATG
AM1Gr1

```

FIG. 5 (7 de 14)



Frammento I

```

EcoRI      PmlI      AM111
GTTAGAAATTCGTAGCACCGTGCCTGGCAACCGCGCCAGAKKCGCAGCGTCCAGGTTCCAGGAGTTACCGACGGCAGCTTTACACCCAG
CAATCTTAAGCATTCTGTCACGACCGCGTTGGCCCGGCTCTGCTCGCACGGGTCAGTTCTTCCACCACAAAGGTCCTCAGTGGCTGCCGTC
5' P 3' OH

AM11r4
3' OH 5' P      ApaI      AM112      BstEII
CCCCGTGACCGC GGGGAGCTGAACGAGCACCTGGGCTTCTGCGCCCTACATCCGGCCGAGGTGGAKKACAAACATCATGTGTGACCCGTGCAGGAGTTGGTC
GTTTACATGGCG_CCCCTCTGACTTCTCTGTTGGACCCGTAATACCCGGGATGTAGGGCGGCTCCACCCTTTTATTAGTACCACTGGCACGTTCTCAAGT
5' P 3' OH

AM11r3
3' OH 5' P      AM113
CTGTCTTCACCATCTTCGAC GAGACCAAGGCTGGTATTTCACTCGAGACATGGGCAACTGCCGCCGCCCTGCAACATCCAGATGGGAGCTCTTACT
CAATAGAAATTTGGTAGAGCTG_CTTCTGGTTCTCGACCATGAAGTGGCTCTTTGTACCCTGCGTTGACGCTTCTGGGACCTTACCTCTCTGGT
5' P 3' OH

AM11r2
3' OH 5' P      AM114      KpnI      HindIII
TTCAAGGAGAACTACCGCTTCACG CCATCAACGGCTACATGGACACCCCTGCCCGCCCTGGTGTATGGCCCCAGGACCCAGCCATCCGCTGGTACCCCTACAA
AAATTTCTTTGTATGCCGGAAGGTGC_GGTAGTTGCCGATTTAGTACCTGTGGACGGGCCGACCCTACTCGGCTCTGGTCCGCTAGGCCACCCATGGGATGTT
5' P 3' OH

GCTTTAC
CGAATG

```

FIG. 5 (9 de 14)









Frammento N

```

EcoRII      BstEII
GTAGAATTCCGTAGGGTGACCGGGCGTGACCCAGCGGTGAAGAGCCCTGCTGACCAGCATGTACCGTGAAGGAGTTCCCTGATCAGCAGCAGCCAGGACGGTCGCA
CATCTTAAGCATCCCACTTGGTCCACTGGTGGTCCCTCCACCTTCTCGGACGACTGGTCGTACATGCAATTTTCTAAGGACTAGTCTGCTCGGTCCCTGACGCTGACG
3' OH 5' P      AM1Nr3      5' P 3' OH
CCAGTGGACCCCTGTCTTC CAGAACGGCAAGGTGAAGGTGTCCAGGGCAACCAAGGACAGCTTCACCCCGTGGTGAACAGCCCTGGACCCCGCCCTGCTTCGAC
GATCACCTGGGACMGMAG_GTCATTGCCGTTCCNCTTCCACAAAGGTCCTCGATGGTCCGACCCACTTGTCCGACCTGGGGGGGATGACATG
AM1Nr2
3' OH 5' P      AM1Nr1      5' P 3' OH
CAGCTTACCTGGGCATCCACC CAGAGCTGGGTGCACCAAGATCGCCCTTCCGCAATGGAGGTGCTGACCTCGGAGCCCGGAGCCCTGTACTAGCTGCCCGGGCTA
GATTAATGACGCGTAGGTGGG_GGTCTCGACCCATTTTCTTAGCGGGACCGCTCCACGACTTATGCTCCGGTCTCTGACATGATCTCATGTTCTGAT
AM1Nr3      AM1Nr1      SmaI
HindIII
CAAGCTTTAC
CTTTTAAATG

```

FIG. 5 (14de 14)

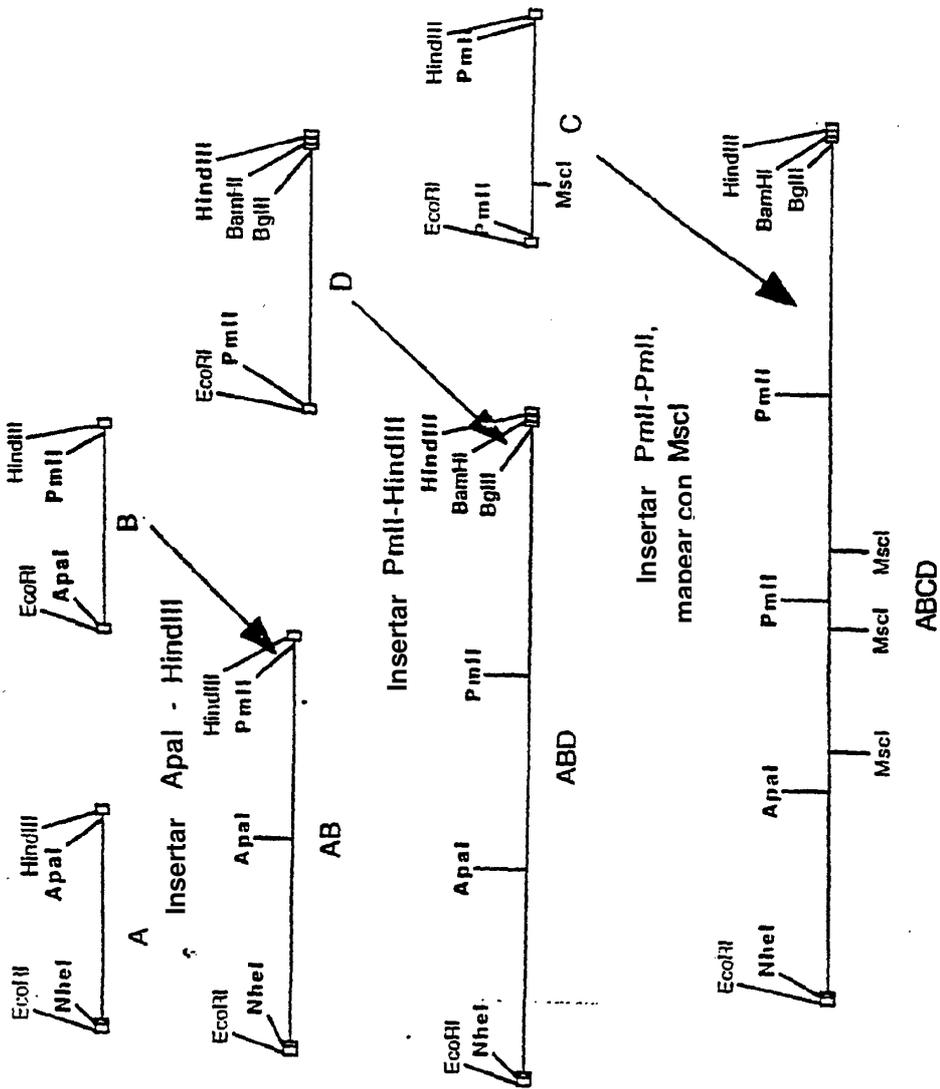


FIG. 6 (Ide 5)

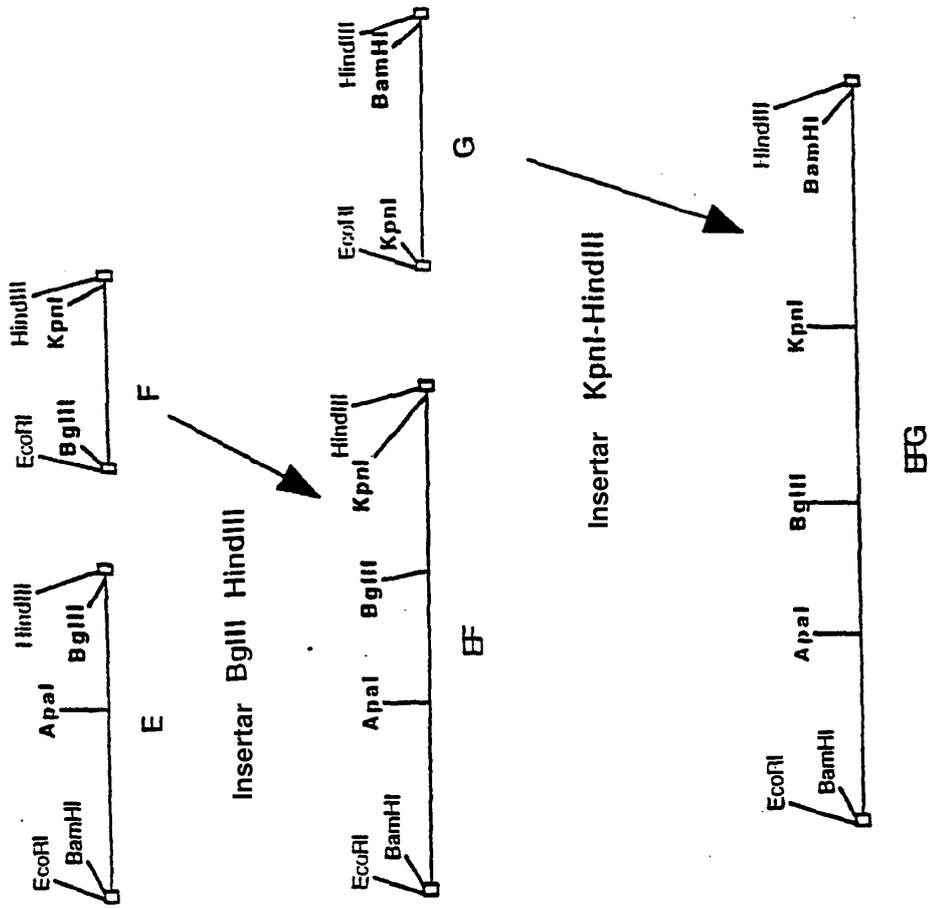
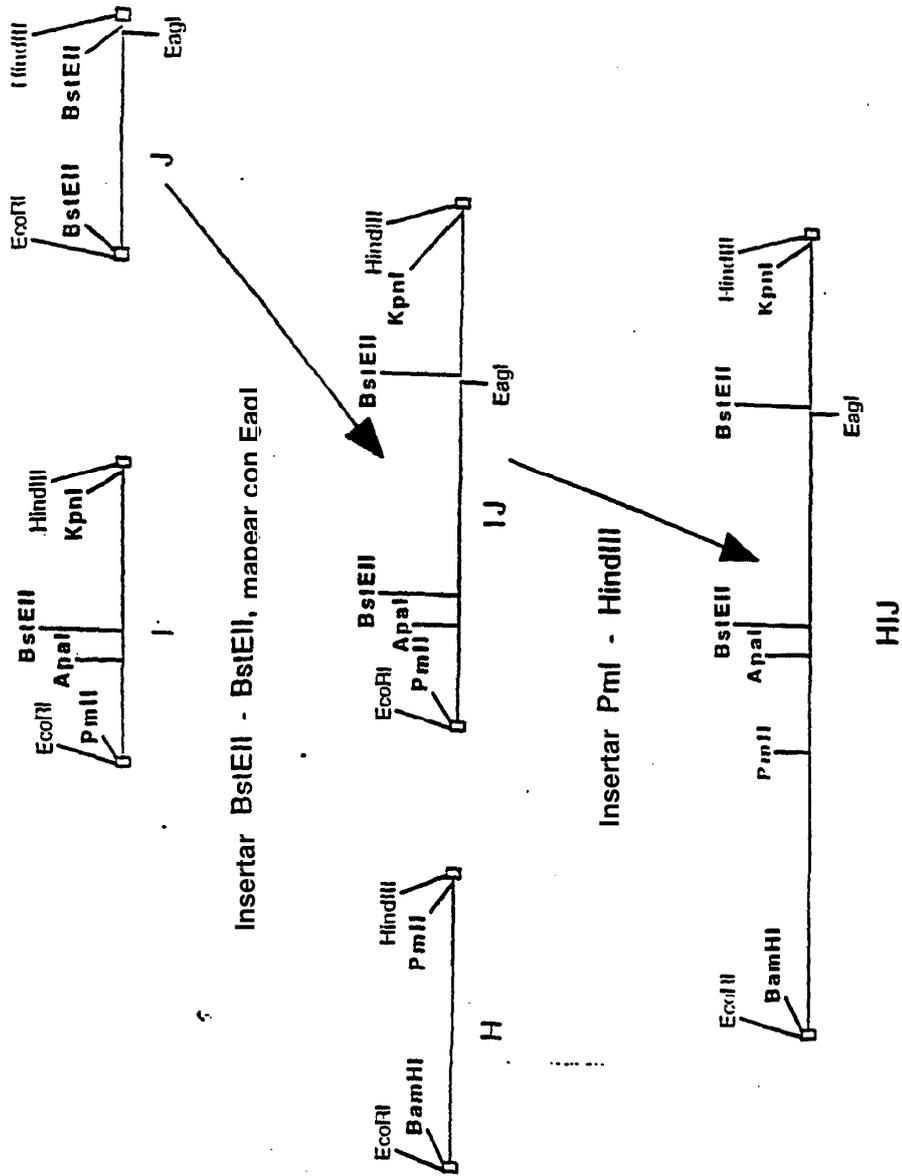


FIG. 6 (2 de 5)



Insertar BstEII - BstEII, mapear con EagI

Insertar PmlI - HindIII

FIG. 6 (3de 5)

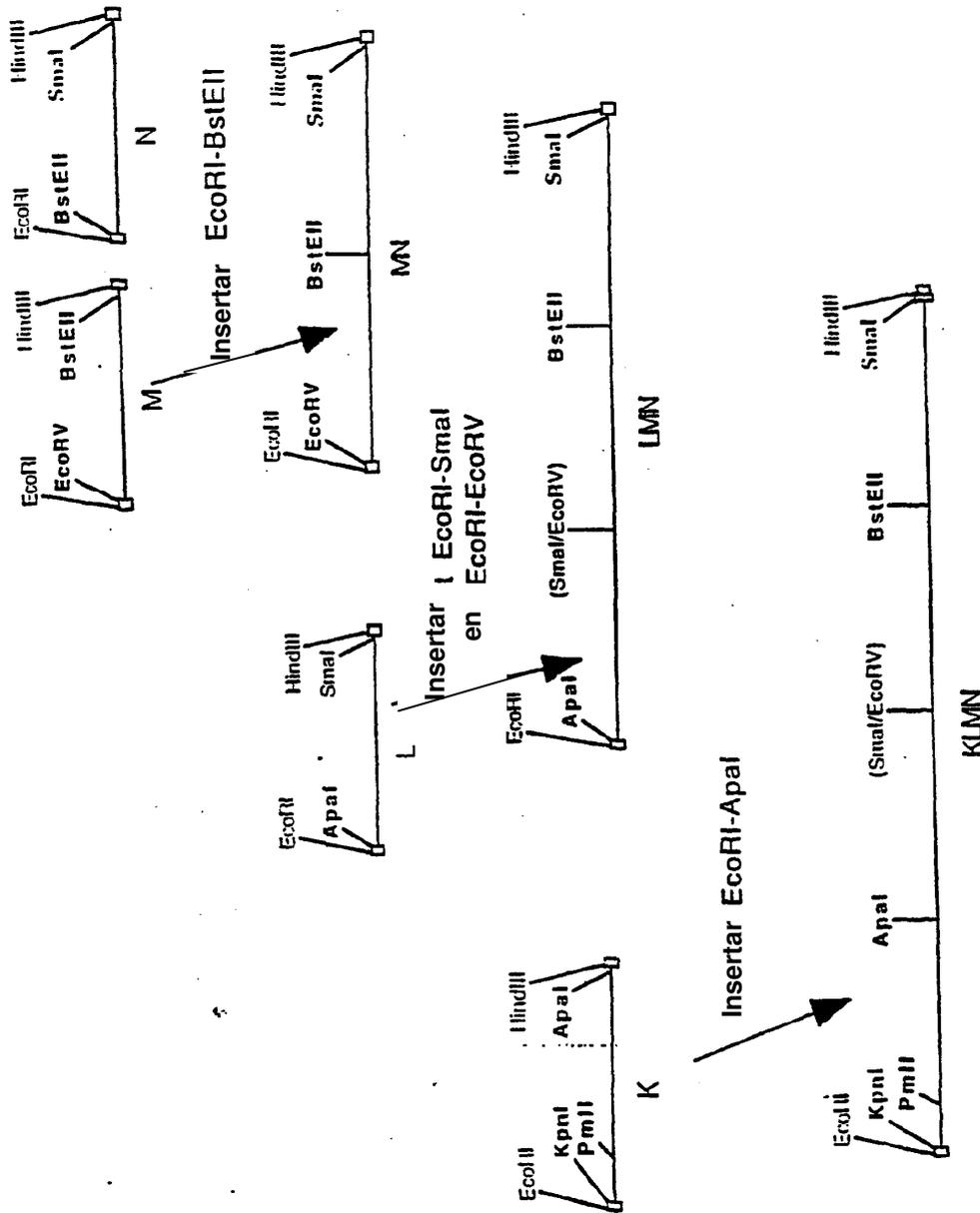


FIG. 6 (4 de 5)

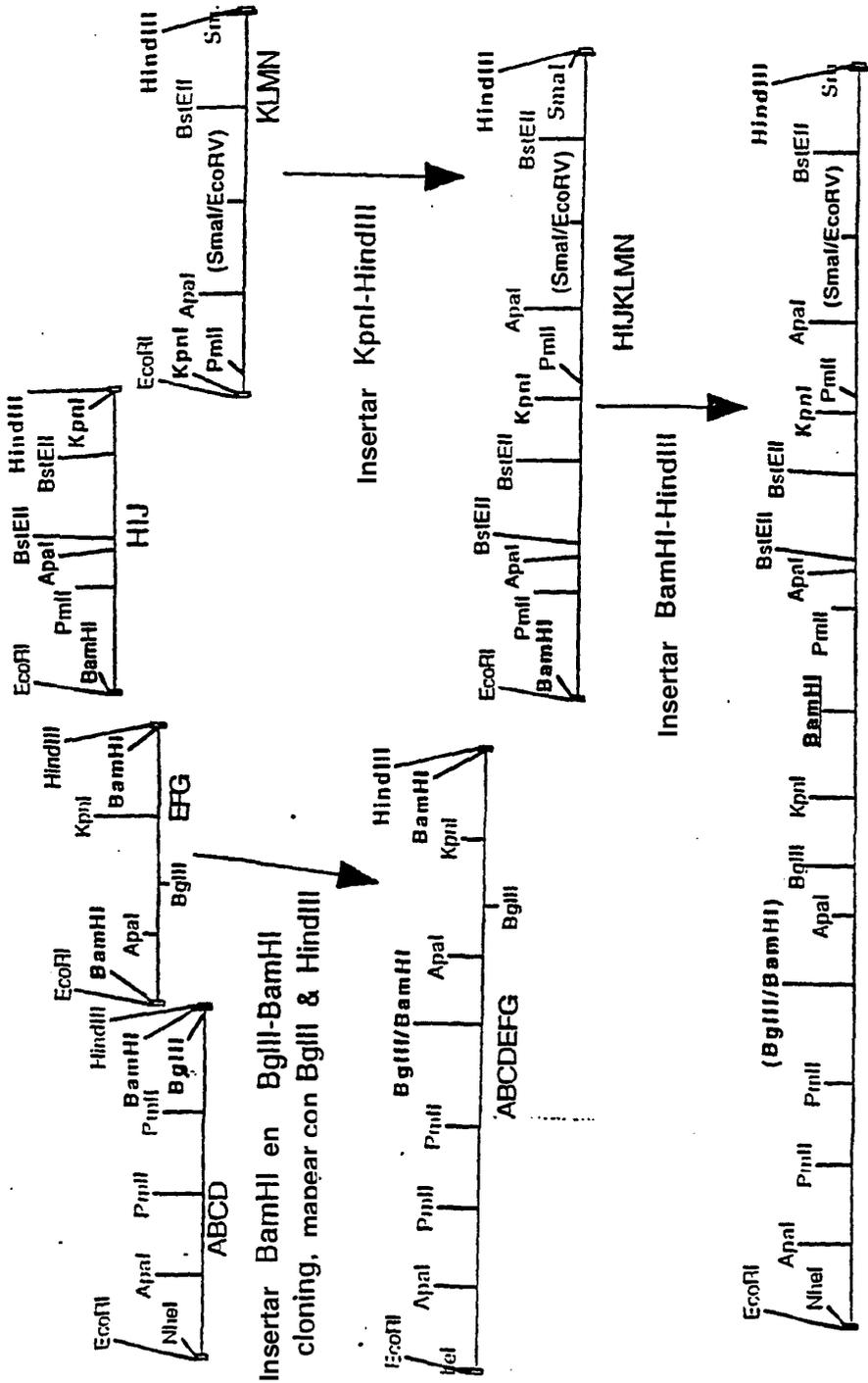


FIG. 6 (5 de 5)

EcoRI NheI  
1 TAGAATTCGGTAGGCTAGCATTCAGATCCAGCTGAGCAGCTGCTTCTTCTCTGCGCTGCTGCGCTTCTGCTTC  
1 MetGlnIleGluLeuSerThrCysPhePheLeuCysLeuLeuArgPheCysPhe  
73 AGCGCCACCCCGCGCTACTACCTGGGCGCCCTGGAGCTGAGCTGGACTACATGCAGAGCGACCTGGCGGAG  
19 SerAlaThrArgArgTyrTyrLeuGlyAlaValGluLeuSerTrpAspTyrMetGlnSerAspLeuGlyGlu  
145 CTGCCCCGTGGACGCCCCGCTTCCCGCCCGCTGCCCCAAGAGCTTCCCTTCACACCAGCGTGGTGTACAAG  
43 LeuProValAspAlaArgPheProProArgValProLysSerPheProPheAsnThrSerValValTyrLys  
217 AAGACCCGTGTTCTGGAGTTCACCGACCACCTGTTCAACATCGCCAAAGCCCCCGCCCCCTGGATGGGCCCTG  
67 LysThrLeuPheValGluPheThrAspHisLeuPheAsnIleAlaLysProArgProProTrpMetGlyLeu  
Apal MscI  
289 CTGGGCCCCACCATCCAGGCGGAGGTGTACGACACCCGTGGTGTATCCCTGAAGAACATGGCCAGCCACCCC  
91 LeuGlyProThrIleGlnAlaGluValTyrAspThrValValIleThrLeuLysAsnMetAlaSerHisPro  
361 GTGAGCCTGCACGCGCTGGGCGTGAGCTACTGGAAGGCCAGCGAGGGCGCGAGTACGACGACCCAGACCAGC  
115 ValSerLeuHisAlaValGlyValSerTyrTrpLysAlaSerGluGlyAlaGluTyrAspAspGlnThrSer  
433 CAGCGCGAGAAGGAGGACGACAAGGTGTTCCCGCGGGCAGCCACCTACGTGTGGCAGGTGCTGAAGGAG  
139 GlnArgGluLysGluAspAspLysValPheProGlyGlySerHisThrTyrValTrpGlnValLeuLysGlu  
MscI PmlI  
505 AACGGCCCCATGGCCAGCGACCCCTGTGCTGACCTACAGCTACCTGAGCCACGTGGACCTGGTGAAGGAC  
163 AsnGlyProMetAlaSerAspProLeuCysLeuThrTyrSerTyrLeuSerHisValAspLeuValLysAsp  
MscI  
577 CTGAACAGCGGCTGATCGGCGCCCTGCTGGTGTGCCCGGAGGGCAGCCTGGCCAAGGAGAAGACCAGACC  
187 LeuAsnSerGlyLeuIleGlyAlaLeuLeuValCysArgGluGlySerLeuAlaLysGluLysThrGlnThr  
649 CTCACCAAGTTCATCCTGCTGTTCCCGCTGTTTCGACGAGGGCAAGAGCTGGCACAGCGAGCAAGAACAGC  
211 LeuHisLysPheIleLeuLeuPheAlaValPheAspGluGlyLysSerTrpHisSerGluThrLysAsnSer  
721 CTGATGCAGGACCGGACGCGCCCGCAGCGCCCGCTGGCCCAAGATGCACACCGTGAACCGCTACGTGAAC  
235 LeuMetGlnAspArgAspAlaAlaSerAlaArgAlaTrpProLysMetHisThrValAsnGlyTyrValAsn  
PmlI  
793 CGCAGCCTGCCCCGCTGATCGGCTGCCACCGCAAGAGCCTGTACTGGCACGTGATCGGCATGGCCACCACC  
259 ArgSerLeuProGlyLeuIleGlyCysHisArgLysSerValTyrTrpHisValIleGlyMetGlyThrThr  
865 CCGGAGGTGCACAGCATCTTCCCTGGAGGGCCACACCTTCCTGGTGGCAACCACCGCCAGGCCAGCTGGAG  
283 ProGluValHisSerIlePheLeuGluGlyHisThrPheLeuValArgAsnHisArgGlnAlaSerLeuGlu  
937 ATCAGCCCCATCACCTTCCCTGACCGCCGAGCCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTCTGCCAC  
307 IleSerProIleThrPheLeuThrAlaGlnThrLeuLeuMetAspLeuGlyGlnPheLeuLeuPheCysHis  
1009 ATCAGCAGCCACCAGCAGCGCCGATGGAGGCTACGTGAAGGTGGACAGCTGCCCGAGGAGCCCCAGCTG  
331 IleSerSerHisGlnHisAspGlyMetGluAlaTyrValLysValAspSerCysProGluGluProGlnLeu  
1081 CGCATGAAGAACAACGAGGAGGCGGAGGACTACGACGACGACCTGACCGACAGCGATGGACGCTGGTGGC  
355 ArgMetLysAsnAsnGluGluAlaGluAspTyrAspAspAspLeuThrAspSerGluMetAspValValArg  
(BglII/BamHI)  
1153 TTCGACGACGACAACAGCCCCAGCTTCATCCAGATCCGACGCTGGCCAAGAAGCACCCCAAGACCTGGGTG  
379 PheAspAspAspAsnSerProSerPheIleGlnIleArgSerValAlaLysLysHisProLysThrTrpVal  
1225 CACTACATCGCCCGGAGGAGGAGGACTGGACTACGCCCCCTGGTGTGGCCCCCGACGACCCGACCTAC  
403 HisTyrIleAlaAlaGluGluAspTrpAspTyrAlaProLeuValLeuAlaProAspAspArgSerTyr  
EagI  
1297 AAGAGCCAGTACCTGAACAACGGCCCCAGCGCATCGGCGCCAAAGTACAAAGGAGGCTTCATGGCCCTAC  
427 LysSerGlnTyrLeuAsnAsnGlyProGlnArgIleGlyArgLysTyrLysLysValArgPheMetAlaTyr  
Apal  
1369 ACCGACGAGACCTTCAAGACCCCGAGGCCATCCAGCAGGAGCGGGCATCCTGGCCCCCTGCTGTACGGC  
451 ThrAspGluThrPheLysThrArgGluAlaIleGlnHisGluSerGlyIleLeuGlyProLeuLeuTyrGly

FIG. 7 (1 de 3)

1441 DAGGTGGGGACACCCCTGCTGATCATCTTCRAGAACCAGGGCCAGCCGCGCTACAACATCTACCCGCGGGC  
 1475▶ GluValGlyAspThrLeuLeuIleIlePheLysAsnGlnAlaSerArgProTyrAsnIleTyrProHisGly  
 1513 ATCACCCGACGTCGCGCCCTGTACAGCCCGCCCTGCCCAAGGGCGTGAAGCACCTGAGGACTTCCCCATC  
 1499▶ IleThrAspValArgProLeuTyrSerArgArgLeuProLysGlyValLysHisLeuLysAspPheProIle  
     BglII  
 1535 CTGCCCCGGGAGATCTTCAAGTACAAGTGGACCTGACCCGTGGAGGACCGCCCCACCAAGAGCCACCCCGG  
 1523▶ LeuProGlyGluIlePheLysTyrLysTrpThrValThrValGluAspGlyProThrLysSerAspProArg  
 1657 TGCCTGACCCGCTACTACAGCAGCTTCCTGAACATGGAGCGCGACCTGGCCAGCGGCTGATCGGCCCCCTG  
 1647▶ CysLeuThrArgTyrTyrSerSerPheValAsnMetGluArgAspLeuAlaSerGlyLeuIleGlyProLeu  
 1729 CTGATCTGCTACAAGGAGCGGTGGACCGCGGCAACCAGATCATGAGCGACAAGCGCAACGTGATCCTG  
 1717▶ LeuIleCysTyrLysGluSerValAspGlnArgGlyAsnGlnIleMetSerAspLysArgAsnValIleLeu  
     KpnI  
 1801 TTCAGCGTGTTCGACGAGAACCAGCTCGTACCTGACCGAGAACATCCAGCGCTTCCTGCCCAACCCCGCC  
 1791▶ PheSerValPheAspGluAsnArgSerTrpTyrLeuThrGluAsnIleGlnArgPheLeuProAsnProAla  
 1873 GGCGTGCAGCTGGAGGACCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGCTACGTGTTCGAC  
 1863▶ GlyValGlnLeuGluAspProGluPheGlnAlaSerAsnIleMetHisSerIleAsnGlyTyrValPheAsp  
 1945 AGCCTGCAGCTGAGCGTGTGCTGCACGAGGTGGCCTACTGGTACATCCTGAGCATCGGGCCCGACCGAC  
 1935▶ SerLeuGlnLeuSerValCysLeuHisGluValAlaTyrTrpTyrIleLeuSerIleGlyAlaGlnThrAsp  
 2017 TTCCTGAGCGTGTCTTCAGCGGTACACCTTCAAGCACAAGATGGTGTACGAGGACACCCCTGACCCCTGTT  
 1997▶ PheLeuSerValPhePheSerGlyTyrThrPheLysHisLysMetValTyrGluAspThrLeuThrLeuPhe  
     BamHI  
 2089 CCCTTCAGCGCGGACGACCGTGTTCATGAGCATGGAGAACCCTGGCCTGTGGATCCTGGGCTGCCACAACAGC  
 2079▶ PropheSerGlyGluThrValPheMetSerMetGluAsnProGlyLeuTrpIleLeuGlyCysHisAsnSer  
 2161 GACTTCCGCAACCCTCGGCATGACCGCCCTGCTGAGGTGAGCAGCTGCGACAAGAACCAGCGGCTACTAC  
 2151▶ AspPheArgAsnArgGlyMetThrAlaLeuLeuLysValSerSerCysAspLysAsnThrGlyAspTyrTyr  
 2233 GAGGACAGCTACGAGGACATCAGCGCCTACCTGCTGAGCAAGAACAACGCCATCGAGCCCCCGCTGGAGGAG  
 2223▶ GluAspSerTyrGluAspIleSerAlaTyrLeuLeuSerLysAsnAsnAlaIleGluProArgLeuGluGlu  
     BstXI  
 2305 ATCACCCGACCAACCCCTGCAGAGCGACAGGAGGAGATCGACTACGACGACACCATCAGCGTGGAGATGAAG  
 2295▶ IleThrArgThrThrLeuGlnSerAspGlnGluGluIleAspTyrAspAspThrIleSerValGluMetLys  
 2377 AAGGAGGACTTCCACATCTACGACGAGGACGAGAACCCAGAGCCCCCGCAGCTTCCAGAGAGACCCCGCAC  
 2367▶ LysGluAspPheAspIleTyrAspGluAspGluAsnGlnSerProArgSerPheGlnLysLysThrArgHis  
     PmlI  
 2449 TACTTCATCGCCCGCGTGGAGCGCCTGTGGGACTACGGCATGAGCAGCAGCCCCACGTTGCTGGCAACCGC  
 2439▶ TyrPheIleAlaAlaValGluArgLeuTrpAspTyrGlyMetSerSerSerProHisValLeuArgAsnArg  
 2521 GCCCAGAGCGGCGAGCGTGCCTCAGTTCAAGAAGGTGGTGTCCAGGAGTTCACCGACGGCAGCTTACCCAG  
 2511▶ AlaGlnSerGlySerValProGlnPheLysLysValValPheGlnGluPheThrAspGlySerPheThrGln  
     ApaI  
 2593 CCCCTGTACCGCGGAGCTGAACGAGCACCTGGGCCCTGCTGGGCCCTACATCCGCGCCGAGGTGGAGGAC  
 2583▶ ProLeuTyrArgGlyGluLeuAsnGluHisLeuGlyLeuLeuGlyProTyrIleArgAlaGluValGluAsp  
     BstEII  
 2665 AACATCATGGTGACCTTCCGCAACCAGGCCAGCCCGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAG  
 2655▶ AsnIleMetValThrPheArgAsnGlnAlaSerArgProTyrSerPheTyrSerSerLeuIleSerTyrGlu  
 2737 GAGGACCGCGCCAGGGCGCCGAGCCCGCAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGG  
 2727▶ GluAspGlnArgGlnGlyAlaGluProArgLysAsnPheValLysProAsnGluThrLysThrTyrPheTrp  
 2809 AAGTGCAGCACCATGCCCCCACCAGGACGAGTTCGACTGCAAGGCCCTGGCCCTACTTCAGCGACGCTG  
 2799▶ LysValGlnHisHisMetAlaProThrLysAspGluPheAspCysLysAlaTrpAlaTyrPheSerAspVal

FIG. 7 (2 de 3)





EcoRI NheI  
1 TAGAATTCCTAGGCTAGCATGCAGATCGAGCTGAGCACCTGCTTCTCCTGTGCTGCTGCGGCTTCTGCTTC  
1 MetGlnIleGluLeuSerThrCysPhePheLeuCysLeuLeuArgPheCysPhe  
73 AGCGCCACCCCGCCCTACTACCTGCGCGCCCTGGAGCTGAGCTGGACTACATGCAGAGCGACCTGGCGGAG  
19 SerAlaThrArgArgTyrTyrLeuGlyAlaValGluLeuSerTrpAspTyrMetGlnSerAspLeuGlyGlu  
145 CTGCCCCGTGGACGCCCGCTTCCCCCCCCGCTGCCCAAGAGCTTCCCCTTCAACACCAGCGTGGTGTACAAG  
43 LeuProValAspAlaArgPheProProArgValProLysSerPheProPheAsnThrSerValValTyrLys  
117 AAGACCCCTGTTCTGGAGTTCACCGACCACCTGTTCAACATCGCCCAAGCCCCGCCCCCTGGATGGGCTG  
67 LysThrLeuPheValGluPheThrAspHisLeuPheAsnIleAlaLysProArgProProTrpMetGlyLeu  
Apal MscI  
289 CTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCCGTGGTGTATCACCCCTGAAGAACATGGCCAGCCACCCC  
91 LeuGlyProThrIleGlnAlaGluValTyrAspThrValValIleThrLeuLysAsnMetAlaSerHisPro  
361 GTGAGCCTGCACGCCGTGGGCGTGGACTACTGGAGGCCAGCGAGGGCGCCGAGTACGACGACCAGACCAGC  
115 ValSerLeuHisAlaValGlyValSerTyrTrpLysAlaSerGluGlyAlaGluTyrAspAspGlnThrSer  
433 CAGCGCCGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGACCCACACCTACGTGTGGCAGGTGCTGAAGGAG  
139 GlnArgGluLysGluAspLysValPheProGlyGlySerHisThrTyrValTrpGlnValLeuLysGlu  
MscI PmlI  
505 AACGGCCCCATGGCCAGCGACCCCCCTGTGCTGACCTACAGCTACCTGAGCCACGTGGACCTGCTGAAGGAC  
163 AsnGlyProMetAlaSerAspProLeuCysLeuThrTyrSerTyrLeuSerHisValAspLeuValLysAsp  
MscI  
577 CTGAACAGCGGCTGATCGGCGCCCTGCTGGTGTGCCCGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACC  
187 LeuAsnSerGlyLeuIleGlyAlaLeuLeuValCysArgGluGlySerLeuAlaLysGluLysThrGlnThr  
649 CTCACAAAGTTCATCCTGCTGTTCCGCGTTCGACGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGC  
211 LeuHisLysPheIleLeuLeuPheAlaValPheAspGluGlyLysSerTrpHisSerGluThrLysAsnSer  
721 CTGATGAGGACCGCGACGCCGCCAGCGCCCCGCTGGCCCCAAGATGCACACCCGTGAACGGCTACGTGAAC  
235 LeuMetGlnAspArgAspAlaAlaSerAlaArgAlaTrpProLysMetHisThrValAsnGlyTyrValAsn  
PmlI  
793 CGCAGCCTGCCCGGCCGATCGGCTGCCACCCGAGAGCGTGTACTGGCAGCTGATCGGCATGGGCACCACC  
259 ArgSerLeuProGlyLeuIleGlyCysHisArgLysSerValTyrTrpHisValIleGlyMetGlyThrThr  
365 CCGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTGTTGCGCAACCACCGCCAGGCCAGCCTGGAG  
383 ProGluValHisSerIlePheLeuGluGlyHisThrPheLeuValArgAsnHisArgGlnAlaSerLeuGlu  
537 ATCAGCCCCATCACCTTCTGACCGCCAGACCTGCTGTGAGCCTGGGCCAGTTCCTGCTTCTCTGCCRC  
307 IleSerProIleThrPheLeuThrAlaGlnThrLeuLeuMetAspLeuGlyGlnPheLeuLeuPheCysHis  
1009 ATCAGCAGCCACCAGCAGCGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCCCAGCTG  
331 IleSerSerHisGlnHisAspGlyMetGluAlaTyrValLysValAspSerCysProGluGluProGlnLeu  
1081 CGCATGAAGAACAACGAGGAGGCCGAGGACTACGACGACGACCTGACCGACAGCGAGATGGACGTGGTGGCC  
355 ArgMetLysAsnAsnGluGluAlaGluAspTyrAspAspAspLeuThrAspSerGluMetAspValValArg  
(BglII/BamHI)  
1153 TTCGACGACGACAACAGCCCCAGCTTCATCCAGATCCCGACGCTGGCCAAGAAGCACCCCAAGACCTGGGTG  
379 PheAspAspAspAsnSerProSerPheIleGlnIleArgSerValAlaLysLysHisProLysThrTrpVal  
1225 CACTACATCGCCCGGAGGAGGACTGGGACTACGCCCCCTGGTGTGGCCCCGACGACCCGAGCTAC  
403 HisTyrIleAlaAlaGluGluGluAspTrpAspTyrAlaProLeuValLeuAlaProAspAspArgSerTyr  
EagI  
1297 AAGAGCCAGTACCTGAACAACGGCCCCCAGCGCATCGGCCCAAGTACAAGAAGGTGCGCTTCATGGCCATC  
427 LysSerGlnTyrLeuAsnAsnGlyProGlnArgIleGlyArgLysTyrLysLysValArgPheMetAlaTyr  
Apal  
1359 ACCGACGAGACCTTCAAGACCCCGGAGGCCATCCAGCACGAGAGCGGCATCCTGGGCCCCCTGCTACGGCC  
451 ThrAspGluThrPheLysThrArgGluAlaIleGlnHisGluSerGlyIleLeuGlyProLeuLeuTyrGly

FIG. 9 (1 de 3)

1441 GAGGTGGGGGACACCCTGCTGATCATCTTCCAGAACCAGGGCCAGCCGCCCTACAACATCTACCCCCACGGC  
 475▶ GluValGlyAspThrLeuLeuIleIlePheLysAsnGlnAlaSerArgProTyrAsnIleTyrProHisGly  
 1513 ATCACCAGCGTGGCCCTCTACAGCCGCCCTGCCAAGGGCTGAAGCAGCTGAAGGACTTCCCCATC  
 499▶ IleThrAspValArgProLeuTyrSerArgArgLeuProLysGlyValLysHisLeuLysAspPheProIle

**BglII**

1585 CTGCCCGGCGAGATCTTCAAGTACAGTGGACCCTGACCCTGGAGGACGGCCCCCAAGAGCGACCCCGC  
 523▶ LeuProGlyGluIlePheLysTyrLysTrpThrValThrValGluAspGlyProThrLysSerAspProArg  
 1657 TGCCTGACCCGCTACTACAGCAGCTTCGTGAACATGGAGCGGACCTGGCCAGCGCCCTGATCGGCCCCCTG  
 547▶ CysLeuThrArgTyrTyrSerSerPheValAsnMetGluArgAspLeuAlaSerGlyLeuIleGlyProLeu  
 1729 CTGATCTGCTACAAGGAGAGCGTGGACCAGCGCGGCAACCCAGATCATGAGCGACAAGCGCAACCTGATCCTG  
 571▶ LeuIleCysTyrLysGluSerValAspGlnArgGlyAsnGlnIleMetSerAspLysArgAsnValIleLeu

**KpnI**

1801 TTCAGCGTGTTCGACGAGAACCAGCTGGTACCTGACCGAGAACATCCAGCGCTTCTGCCCAACCCCGCC  
 595▶ PheSerValPheAspGluAsnArgSerTrpTyrLeuThrGluAsnIleGlnArgPheLeuProAsnProAla  
 1873 GCGGTGCAGCTGGAGGACCCCSAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGCTACGTGTTCCGAC  
 619▶ GlyValGlnLeuGluAspProGluPheGlnAlaSerAsnIleMetHisSerIleAspGlyTyrValPheAsp  
 1945 AGCCTGCAGCTGAGCGTGTGCCTGCACGAGGTGGCCCTACTGGTACATCCTGAGCATCGGCCCCAGACCGAC  
 643▶ SerLeuGlnLeuSerValCysLeuHisGluValAlaTyrTrpTyrIleLeuSerIleGlyAlaGlnThrAsp  
 2017 TTCCTGAGCGTGTCTTCAGCGCTACACCTTCAAGCACAGATGGTGTACGAGGACACCTGACCCCTGTTT  
 667▶ PheLeuSerValPhePheSerGlyTyrThrPheLysHisLysMetValTyrGluAspThrLeuThrLeuPhe

**BamHI**

2089 CCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCTGGCTGTGGATCCTGGGCTGCCACAACAGC  
 691▶ ProPheSerGlyGluThrValPheMetSerMetGluAsnProGlyLeuTrpIleLeuGlyCysHisAsnSer  
 2161 GACTTCCGCAACCCGGCATGACCCGCCCTGCTGAAGGTGAGCAGCTCGCACAAGAACCAGCGGACTACTAC  
 715▶ AspPheArgAsnArgGlyMetThrAlaLeuLeuLysValSerSerCysAspLysAsnThrGlyAspTyrTyr  
 2233 GAGGACAGCTACGAGGACATCAGCGCTACCTGCTGAGCAAGAACAACGCCATCGAGCCCCGACGGCGCAGG  
 739▶ GluAspSerTyrGluAspIleSerAlaTyrLeuLeuSerLysAsnAsnAlaIleGluProArgArgArgArg

**BstXI**

2305 GCGGAGATCACCCGCACCACCCTGCAGAGCGACCAGGAGGAGATCGACTACGACGACACCATCAGCGTGGAG  
 753▶ ArgGluIleThrArgThrThrLeuGlnSerAspGlnGluGluIleAspTyrAspAspThrIleSerValGlu  
 2377 ATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCGCCGACCTTCCAGAAGAAGACC  
 787▶ MetLysLysGluAspPheAspIleTyrAspGluAspGluAsnGlnSerProArgSerPheGlnLysLysThr

**PmlI**

2449 CGCCACTACTTTCATCGCCGCCGTGGAGCGCCTGTGGGACTACGGCATGAGCAGCAGCCCCACGTGCTGCCG  
 811▶ ArgHisTyrPheIleAlaAlaValGluArgLeuTrpAspTyrGlyMetSerSerSerProHisValLeuArg  
 2521 AACCGCGCCAGAGCGGCGAGCGTGCCTCCAGTTCAAGAAGGTGGTGTCCAGGAGTTCACCGACGGCAGCTTC  
 835▶ AsnArgAlaGlnSerGlySerValProGlnPheLysLysValValPheGlnGluPheThrAspGlySerPhe

**Apal**

2593 ACCCAGCCCTGTACCGCGCGAGCTGAACGAGCACCTGGCCCTGCTGGGCCCTACATCCCGCGCGAGGTG  
 859▶ ThrGlnProLeuTyrArgGlyGluLeuAsnGluHisLeuGlyLeuLeuGlyProTyrIleArgAlaGluVal

**BstEII**

2665 GAGGACAACATCATGGTGACCTTCCGCAACCAGGCCAGCCGCCCTACAGCTTCTACAGCAGCCTGATCAGC  
 883▶ GluAspAsnIleMetValThrPheArgAsnGlnAlaSerArgProTyrSerPheTyrSerSerLeuIleSer  
 2737 TACGAGGAGGACAGCGCCAGGGCGCCGAGCCCGCCAGAACTTCGTGAAGCCCAACGAGACCAAGACCTAC  
 907▶ TyrGluGluAspGlnArgGlnGlyAlaGluProArgLysAsnPheValLysProAsnGluThrLysThrTyr  
 2809 TTCTGGAAGGTGCAGCACCCATGGCCCCCACCAGGACGAGTTCGACTGCAAGGCCCTGGGCCCTACTTCAGC  
 931▶ PheTrpLysValGlnHisHisMetAlaProThrLysAspGluPheAspCysLysAlaTrpAlaTyrPheSer

FIG. 9 (2 de 3)

2681 TACGTGGACCTGGAGAAGGACCTGCACAGCGGCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCCTG  
 955 AspValAspLeuGluLysAspValHisSerGlyLeuIleGlyProLeuLeuValCysHisThrAsnThrLeu  
 EagI BstEII  
 2953 AACCCCGCCACAGGCGCCAGGTGACCTGCAGGAGTTCGCCCTGTTCACCATCTTCGACGAGACCCAG  
 979 AsnProAlaHisGlyArgGlnValThrValGlnGluPheAlaLeuPhePheThrIlePheAspGluThrLys  
 3025 AGCTGGTACTTCACCGAGAACATGGAGCGCAACTGCCCGGCCCTGCAACATCCAGATGGAGGACCCACCC  
 1003 SerTrpTyrPheThrGluAsnMetGluArgAsnCysArgAlaProCysAsnIleGlnMetGluAspProThr  
 3097 TTCAAGGAGAACTACCGCTTCCACGCCATCAACGGCTACATCATGGACACCCCTGCCCGGCTGGTGATGGCC  
 1027 PheLysGluAsnTyrArgPheHisAlaIleAsnGlyTyrIleMetAspThrLeuProGlyLeuValMetAla  
 KpnI  
 3169 CAGGACCAGCGCATCCGCTGGTACCTGCTGAGCATGGGCAGCAACGAGAACATCCACAGCATCCACTTCAGC  
 1051 GlnAspGlnArgIleArgTrpTyrLeuLeuSerMetGlySerAsnGluAsnIleHisSerIleHisPheSer  
 PmlI  
 3241 GCCACCGTGTTCACCGTGGCAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTGTACCCCGGCGTGTTC  
 1075 GlyHisValPheThrValArgLysLysGluGluTyrLysMetAlaLeuTyrAsnLeuTyrProGlyValPhe  
 3313 CAGACCGTGGAGATGCTGCCCGCAAGCGCCGATCTGGCGGTGGAGTGCCTGATCGGCGAGCACCTGCAC  
 1099 GluThrValGluMetLeuProSerLysAlaGlyIleTrpArgValGluCysLeuIleGlyGluHisLeuHis  
 3385 CGCCGCATGAGCACCCCTGTTCCTGGTGTACAGCAACAAGTGCCAGACCCCTGGGGATGGCCAGCGGCCAC  
 1123 AlaGlyMetSerThrLeuPheLeuValTyrSerAsnLysCysGlnThrProLeuGlyMetAlaSerGlyHis  
 ApaI  
 3457 ATCCGCGACTTCCAGATCACCCGACGGCCAGTACGGCCAGTGGGCCCCAAGCTGGCCCCCTGCCTACTAC  
 1147 IleArgAspPheGlnIleThrAlaSerGlyGlnTyrGlyGlnTrpAlaProLysLeuAlaArgLeuHisTyr  
 3529 AGCGGCAGCATCAACGCCCTGGAGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGTGGCCCCATG  
 1171 SerGlySerIleAsnAlaTrpSerThrLysGluProPheSerTrpIleLysValAspLeuLeuAlaProMet  
 3601 ATCATCCACGGCATCAAGACCCAGGGCGCCGCCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATC  
 1195 IleIleHisGlyIleLysThrGlnGlyAlaArgGlnLysPheSerSerLeuTyrIleSerGlnPheIleIle  
 3673 ATGTACAGCCTGGACGGCAAGAAGTGGCAGACCTACCGCGGCAACAGCACCGGCACCCCTGATGGTGTCTTC  
 1219 MetTyrSerLeuAspGlyLysLysTrpGlnThrTyrArgGlyAsnSerThrGlyThrLeuMetValPhePhe  
 (SmaI/EcoRV)  
 3745 GSCAACCTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCTCATCGCCCGCTACATCCGCCTG  
 1243 GlyAsnValAspSerSerGlyIleLysHisAsnIlePheAsnProProIleIleAlaArgTyrIleArgLeu  
 3817 CACCCGACCCACTACAGCATCCGCGACCCCTCCGCATGGAGCTGATGGGCTCGCACCTGACAGCTGCAGC  
 1267 HisProThrHisTyrSerIleArgSerThrLeuArgMetGluLeuMetGlyCysAspLeuAsnSerCysSer  
 3889 ATGCCCTGGGCATGGAGAGCAAGGCCATCAGCGACGCCCAGATCACCGCCAGCAGCTACTTCACCAACATG  
 1291 MetProLeuGlyMetGluSerLysAlaIleSerAspAlaGlnIleThrAlaSerSerTyrPheThrAsnMet  
 3961 TTCGCCACCTGGAGCCCCAGCAAGGCCCTGCACCTGCAGGGCCCGCAGCAACGCCCTGGCGCCCCAGGTG  
 1315 PheAlaThrTrpSerProSerLysAlaArgLeuHisLeuGlnGlyArgSerAsnAlaTrpArgProGlnVal  
 BstEII  
 4033 AACAAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACCGCGTGACCACCCAGGGC  
 1339 AsnAsnProLysGluTrpLeuGlnValAspPheGlnLysThrMetLysValThrGlyValThrThrGlnGly  
 4105 GTGAAGAGCCTGTGACAGCATGTACGTGAAGGAGTTCCTGATCAGCAGCAGCCAGGACGGCCACCCAGTGG  
 1363 ValLysSerLeuLeuThrSerMetTyrValLysGluPheLeuIleSerSerSerGlnAspGlyHisGlnTrp  
 4177 ACCCTGTCTTCCAGAACGCAAGGTGAAGGTGTCCAGGGCAACCAGGACAGCTTACCCCGTGGTGAAC  
 1387 ThrLeuPhePheGlnAsnGlyLysValLysValPheGlnGlyAsnGlnAspSerPheThrProValValAsn  
 4249 AGCCTGGACCCCCCTGCTGACCCGCTACCTGCGCATCCACCCAGAGCTGGGTGCACCAGATCGCCCTG  
 1411 SerLeuAspProProLeuLeuThrArgTyrLeuArgIleHisProGlnSerTrpValHisGlnIleAlaLeu  
 SmaI HindIII  
 4321 CGCATGGAGGTGCTGGGCTGCGAGGCCAGGACCTGTACTAGCTGCCCGGGCTACAAGCTTTAC  
 1435 ArgMetGluValLeuGlyCysGluAlaGlnAspLeuTyr...

FIG. 9 (3 de 3)

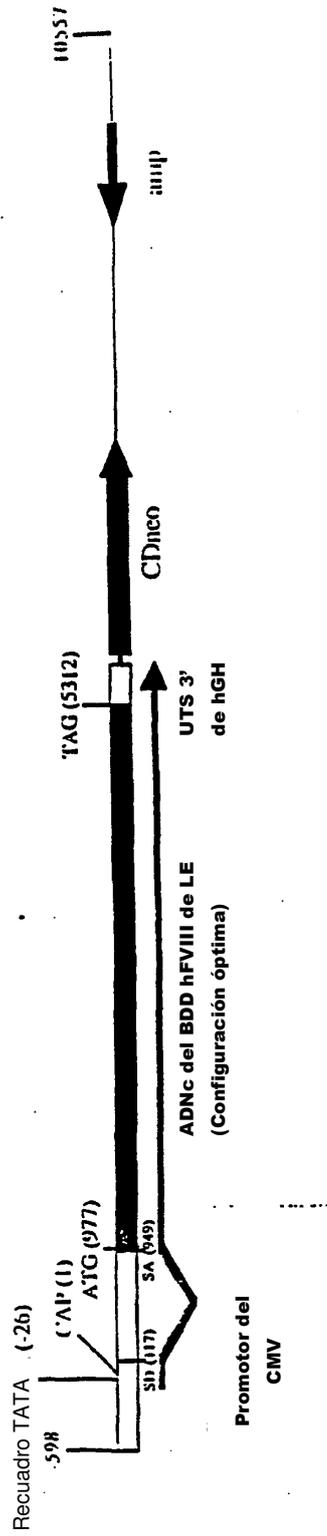


FIG. 10

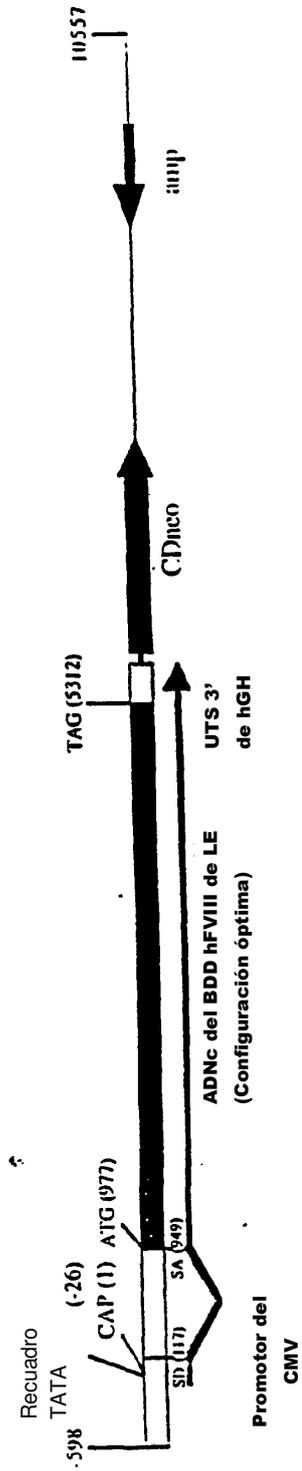


FIG. 11

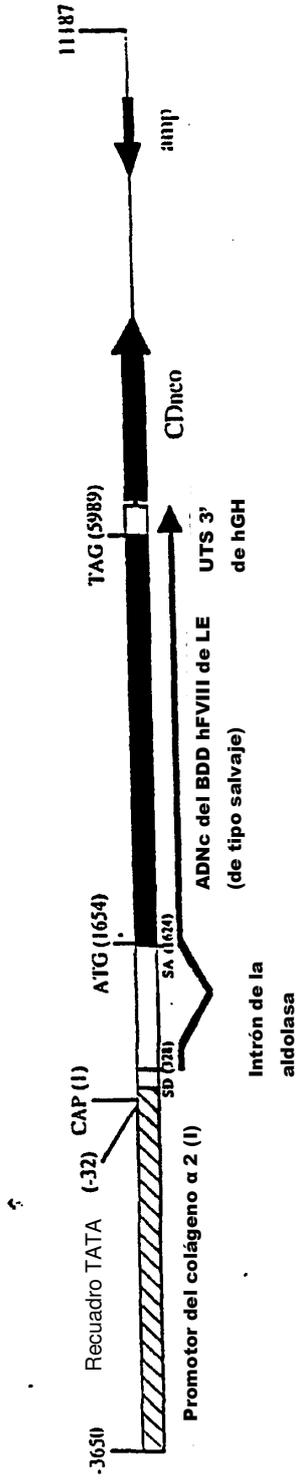


FIG. 12

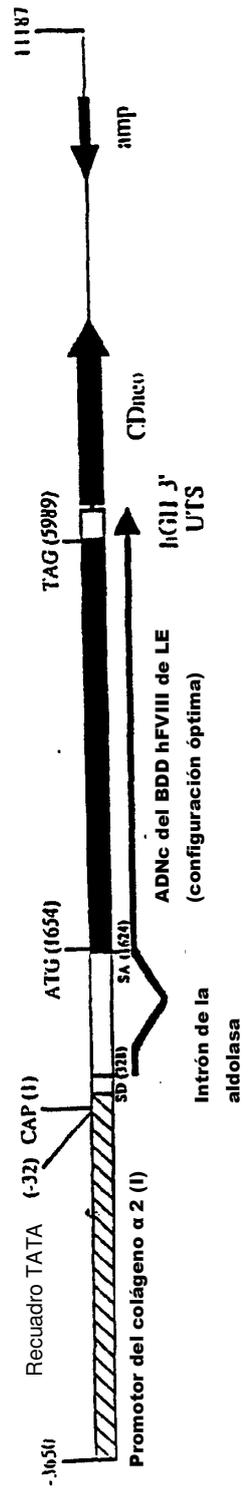


FIG. 13

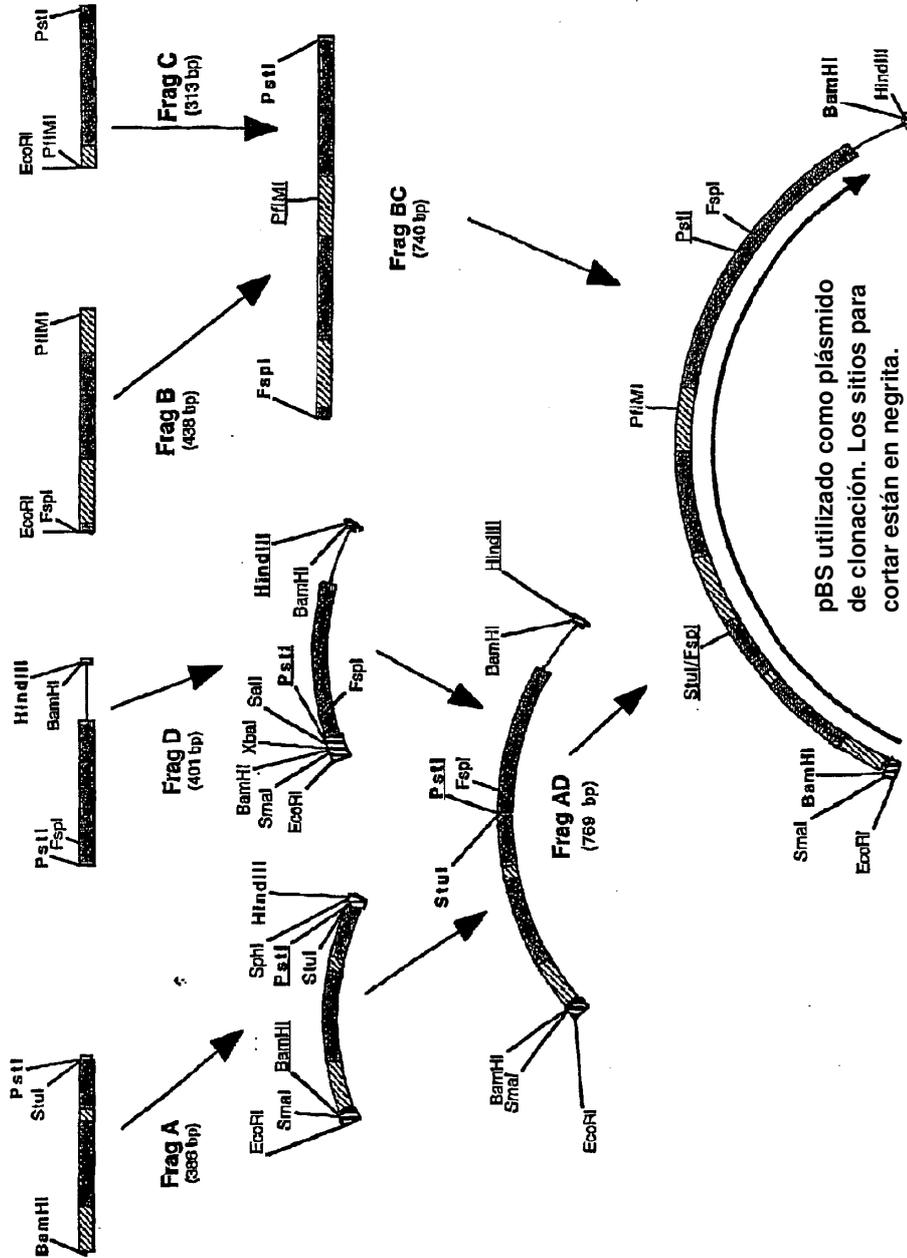
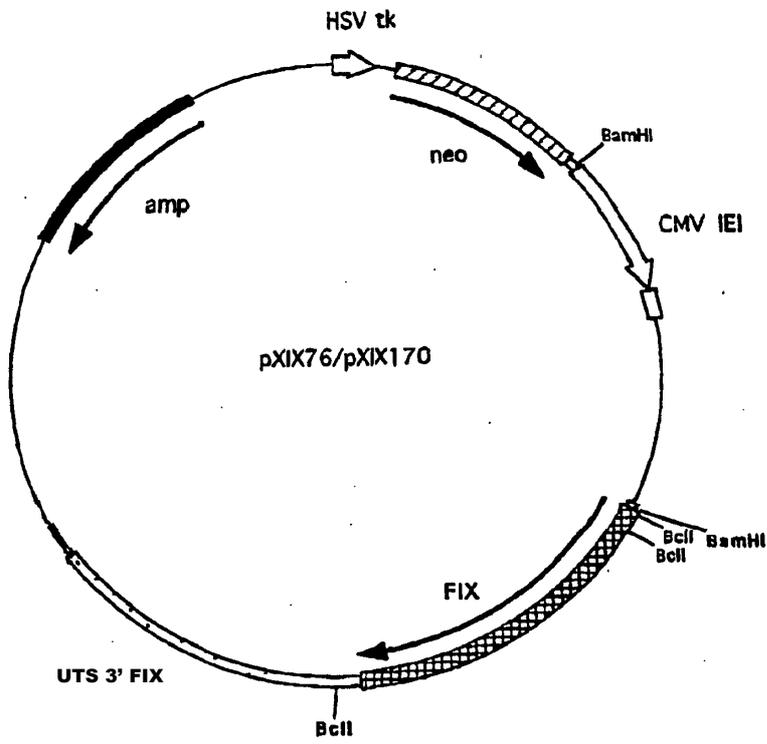


Fig. 14

GGA'TCCATGCAGCGCGTGAACATGATCATGGCCGAGAGCCCCGGCCTGATCACCATCTG  
CCTCTGGGCTACCTGCTGAGCGCCGAGTGCACCGTGTTCCTGGACCACGAGAACGCCA  
ACAAGATCCTGAACCGCCCCAAGCGCTACAACAGCGGCAAGCTGGAGGAGTTCGTGCAG  
GGCAACCTGGAGCGCGAGTGCATGGAGGAGAAGTGCAGCTTCGAGGAGGCCCGCGAGGT  
GTTGAGAACACCGAGCGCACCACCGAGTTCTGGAAGCAGTACGTGGACGGCGACCAGT  
GCGAGAGCAACCCCTGCCTGAACGGCGGCAGCTGCAAGGACGACATCAACAGCTACGAG  
TGC'TGGTGC'CCCTTCGGCTTCGAGGGCAAGAACTGCGAGCTGGACGTGACCTGCAACAT  
CAAGAACGGCCGCTGCGAGCAGTTCTGCAAGAACAGCGCCGACAACAAGGTGGTGTGCA  
GCTGCACCGAGGGCTACCGCCTGGCCGAGAACCAGAAGAGCTGCGAGCCCCGCCGTGCC  
TTCCCTGCGGCCGCGTGAGCGTGAGCCAGACCAGCAAGCTGACCCGCGCCGAGACCGT  
GTTCCCCGACGTGGACTACGTGAACAGCACCGAGGCCGAGACCATCCTGGACAACATCA  
CCCAGAGCACCCAGAGCTTCAACGACTTCACCCGCGTGGTGGGCGGCGAGGACGCCAAG  
CCCCGCCAGTTCCCTGGCAGGTGGTGTGTAACGGCAAGGTGGACGCCTTCTGCGGGG  
CAGCATCGTGAACGAGAAGTGGATCGTGACCGCCGCCACTGCGTGGAGACCGGCGTGA  
AGATCACCGTGGTGGCCGGCGAGCACAACATCGAGGAGACCGAGCACCCGAGCAGAAG  
CGCAACGTGATCCGCATCATCCCCACCACAACACTACAACGCCGCCATCAACAAGTACAA  
CCACGACATCGCCCTGCTGGAGCTGGACGAGCCCCTGGTGTGTAACAGCTACGTGACCC  
CCATCTGCATCGCCGACAAGGAGTACACCAACATCTTCTGAAGTTCGGCAGCGGCTAC  
GTGAGCGGCTGGGGCCGCGTGTTCACAAGGGCCGCGAGCGCCCTGGTGTGCAGTACCT  
GCGCGTGC'CCCTGGTGGACCGCGCCACCTGCCTGCGCAGCACCAAGTTCACCATCTACA  
ACAACATGTTCTGCGCCGGCTTCCACGAGGGCGGGCCGCGACAGCTGCCAGGGCGACAGC  
GGCGGCCCCACGTGACCGAGGTGGAGGGCACCAGCTTCTGACCGGCATCATCAGCTG  
GGGCGAGGAGTGCGCCATGAAGGGCAAGTACGGCATCTACACCAAGGTGAGCCGCTACG  
TGA'ACTGGATCAAGGAGAAGACCAAGCTGACCTAATGAAAAGATGGATTTC'CAAGGTAA  
TTCATTGGAATTGAAAATTAACAGGGCCTCTACTAACTAATCACTTTCCCATCTTTTG  
TTAGATTTGAATATATACATTCTAGGATCC

Fig. 15



*Fig. 16*

GGATCCGCTAGAGCGGAAATTTATGCTGTCCGGTCACCGTGACAATGCAGCTGCGCAAC  
CCCGAGCTGCACCTGGGCTGCGCCCTGGCCCTGCGCTTCTGGCCCTGGTGAAGTGGGA  
CATCCCCGGCGCCCGCGCCCTGGACAACGGCCTGGCCCGCACCCCCACCATGGGCTGGC  
TGCACTGGGAGCGCTTCATGTGCAACCTGGACTGCCAGGAGGAGCCCGACAGCTGCATC  
AGCGAGAAGCTGTTCATGGAGATGGCCGAGCTGATGGTGAGCGAGGGCTGGAAGGACGC  
CGGCTACGAGTACCTGTGCATCGACGACTGCTGGATGGCCCCCAGCGCGACAGCGAGG  
GCCGCTGCAGGCCGACCCCCAGCGCTTCCCCCACGGCATCCGCCAGCTGGCCAACTAC  
GTGCACAGCAAGGGCCTGAAGCTGGGCATCTACGCCGACGTGGGCAACAAGACCTGCGC  
CGGCTTCCCGGGCAGCTTCGGCTACTACGACATCGACGCCCGAGACCTTCGCCGACTGGG  
GCGTGGACCTGCTGAAGTTCCGACGGCTGCTACTGCGACAGCCTGGAGAACCTGGCCGAC  
GGCTACAAGCACATGAGCCTGGCCCTGAACCGCACCGGCCGACGATCGTGTACAGCTG  
CGAGTGGCCCCTGTACATGTGGCCCTTCCAGAAGCCCAACTACACCGAGATCCGCCAGT  
ACTGCAACCACTGGCGCAACTTCGCCGACATCGACGACAGCTGGAAGAGCATCAAGAGC  
ATCCTGGACTGGACCAGCTTCAACCAGGAGCGCATCGTGGACGTGGCCGGCCCCGGCGG  
CTGGAACGACCCCGACATGCTGGTGATCGGCAACTTCGGCCTGAGCTGGAACCAGCAGG  
TGACCCAGATGGCCCTGTGGCCATCATGGCCGCCCCCTGTTTCATGAGCAACGACCTG  
CGCCACATCAGCCCCCAGGCCAAGGCCCTGCTGCAGGACAAGGACGTGATCGCCATCAA  
CCAGGACCCCCCTGGGCAAGCAGGGCTACCAGCTGCGCCAGGGCGACAACCTTCGAGGTGT  
GGGAGCGCCCCCTGAGCGGCCTGGCCTGGGCGGTGGCCATGATCAACCGCCAGGAGATC  
GGCGGCCCCCGCAGCTACACCATCGCCGTGGCCAGCCTGGGCAAGGGCGTGGCCTGCAA  
CCCCGCTGCTTCATCACCCAGCTGCTGCCCGTGAAGCGCAAGCTGGGCTTCTACGAGT  
GGACCAGCCGCTGCGCAGCCACATCAACCCCCACGGCACCGTGTGCTGCAGCTGGAG  
AACACCATGCAGATGAGCCTGAAGGACCTGCTGTAAAAAAAAAAAAAACTCGAG

Fig. 17

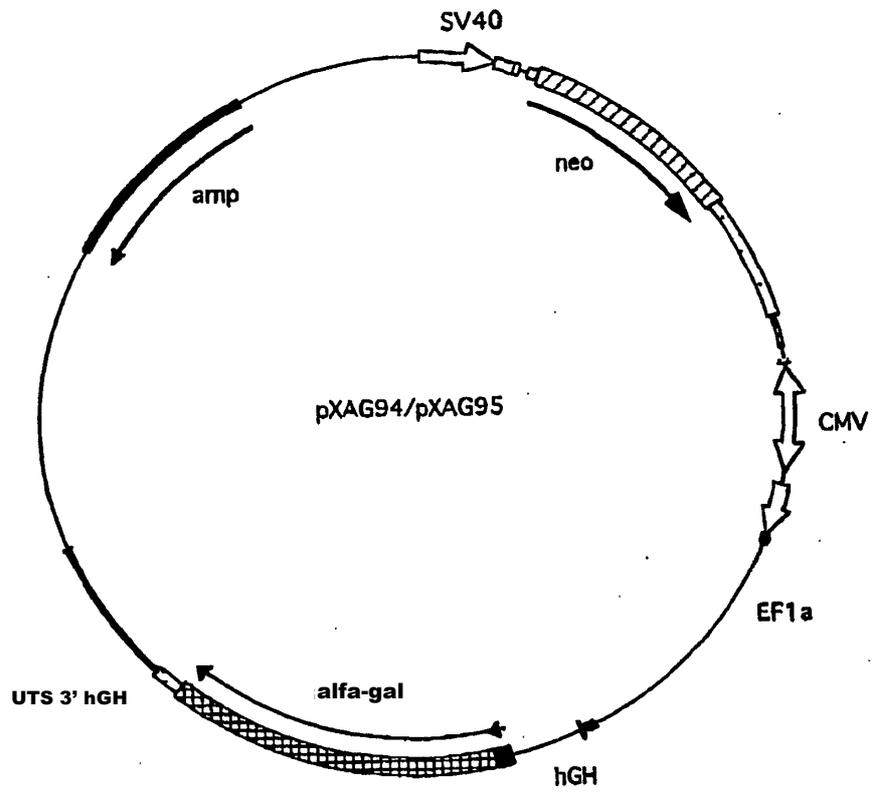


Fig 18

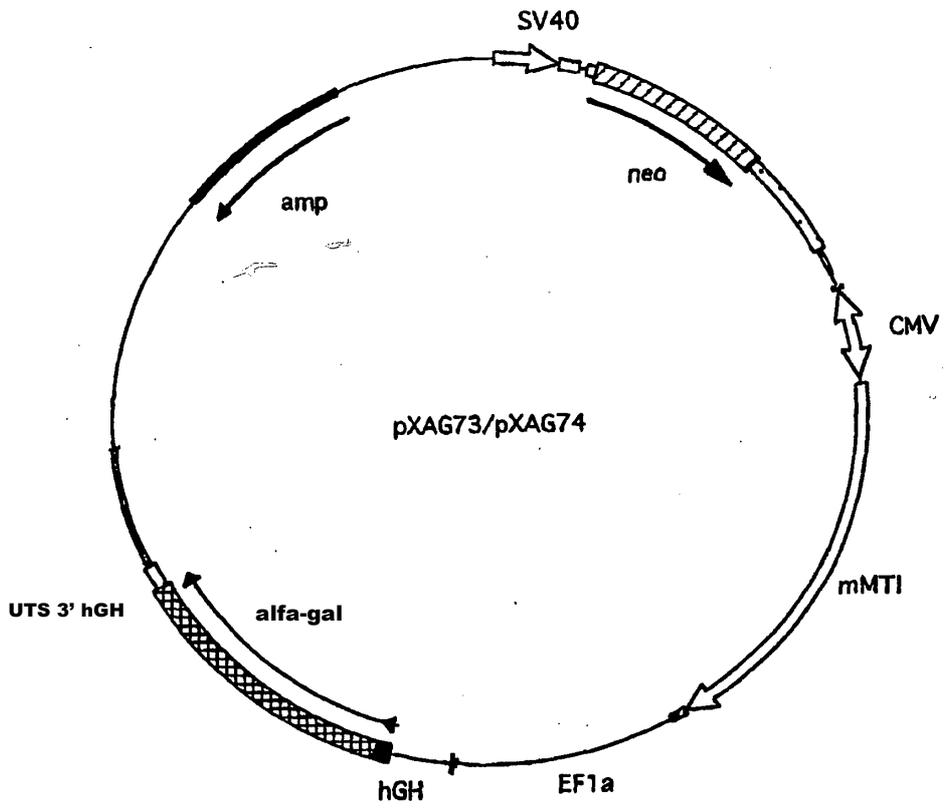


Fig. 19