

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 396 915**

51 Int. Cl.:

**A61K 39/21** (2006.01)

**A61K 39/12** (2006.01)

**A61K 39/38** (2006.01)

**C07K 14/16** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.06.2007 E 07798067 (0)**

97 Fecha y número de publicación de la concesión europea: **03.10.2012 EP 2029168**

54 Título: **Secuencias consenso, antígenos y transgenes del VIH-1 del clado A**

30 Prioridad:

**02.06.2006 US 810816 P**

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

**01.03.2013**

73 Titular/es:

**INTERNATIONAL AIDS VACCINE INITIATIVE  
(100.0%)**

**110 WILLIAMS STREET, 27TH FLOOR  
NEW YORK, NY 10038-3901, US**

72 Inventor/es:

**GUPTA, KALPANA y  
JACKSON, NICHOLAS**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 396 915 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

Secuencias consenso, antígenos y transgenes del VIH-1 del clado A

## 5 Campo de la invención

La invención proporciona un polinucleótido que codifica una proteína de fusión del VIH-1 del clado A, que comprende los antígenos Gag, Pol y Nef del VIH-1 del clado A según se define en la reivindicación 1, una composición inmunógena que contiene este polinucleótido así como este polinucleótido para utilizar en un método para generar una respuesta inmunitaria contra el VIH-1.

La presente divulgación se refiere a secuencias consenso de nucleótidos y proteínas para antígenos del VIH-1 del clado A y a secuencias de nucleótidos y proteínas para antígenos del clado A de aislados de campo circulantes del VIH-1 donde las secuencias de los antígenos están estrechamente relacionadas con esas secuencias consenso. En una realización preferida, la presente divulgación se refiere a transgenes del VIH-1 del clado A que derivan de dichas secuencias, y que codifican a Gag, Pol (RT e Int) y Nef (en lo sucesivo, "GRIN") del VIH-1 del clado A. La solicitud también se refiere a vectores que contienen dichos transgenes, que incluyen en una realización preferida, vectores de adenovirus que contienen dichos transgenes. La divulgación también se refiere a composiciones inmunógenas que contienen antígenos del VIH-1 del clado A, secuencias de nucleótidos, vectores o transgenes y a métodos para generar una respuesta inmunitaria contra el VIH-1 en un sujeto mediante administración de una cantidad eficaz de dichas composiciones inmunógenas.

## Antecedentes de la invención

El SIDA, o Síndrome de Inmunodeficiencia Adquirida, es causado por el virus de la inmunodeficiencia humana (VIH) y se caracteriza por varias particularidades clínicas, que incluyen síndromes de consunción, degeneración del sistema nervioso central e inmunosupresión profunda que propicia infecciones oportunistas y neoplasias. El VIH es un integrante de la familia de los lentivirus de los retrovirus animales, que incluyen al virus visna de las ovejas y el virus de inmunodeficiencia de los bovinos, felinos y simios (VIS). Hasta el momento se identificaron dos tipos de VIH, denominados VIH-1 y VIH-2 estrechamente relacionados, de los cuales el VIH-1 es por mucho la causa más común del SIDA. Sin embargo, el VIH-2, que difiere en estructura genómica y antigenicidad, causa un síndrome clínico similar.

Una partícula de VIH infecciosa consta de dos cadenas idénticas de ARN, cada una de aproximadamente 9.2 kb de largo empacadas dentro de un núcleo de proteínas virales. Esta estructura central está rodeada por una envoltura de bicapa de fosfolípidos derivada de la membrana de la célula huésped que también incluye proteínas de membrana codificadas viralmente (Abbas et al., Cellular and Molecular Immunology, 4ª edición, W.B. Saunders Company, 2000, p. 454). El genoma del VIH tiene la organización 5'-LTR-Gag-Pol-Env-LTR-3' característica de la familia de los retrovirus. Repeticiones terminales largas (LTR) en cada extremo del genoma viral sirven como sitios de unión para las proteínas reguladoras transcripcionales del huésped y regulan la integración viral en el genoma del huésped, la expresión de genes virales y la replicación viral.

El genoma del VIH codifica varias proteínas estructurales. El gen Gag codifica proteínas estructurales del núcleo de la nucleocápside y la matriz. El gen Pol codifica las enzimas transcriptasa inversa (RT), integrasa (Int) y proteasa viral necesarias para la replicación viral. El gen tat codifica una proteína que es necesaria para la elongación de los transcritos virales. El gen rev codifica una proteína que promueve la exportación nuclear de los ARN virales incompletamente empalmados o sin empalmar. El producto del gen Vif potencia la infectividad de las partículas virales. El producto del gen vpr promueve la importación nuclear de ADN viral y regula la detención del ciclo celular G2. Los genes vpu y nef codifican proteínas que regulan por disminución la expresión de CD4 de la célula huésped y potencian la liberación del virus desde las células infectadas. El gen Env codifica la glucoproteína de la envoltura viral que es traducida como un precursor de 160 kilodalton (kDa) (gp160) y escindida por una proteasa celular para producir la glucoproteína externa de 120-kDa de la envoltura (gp120) y la glucoproteína transmembrana de 41 kDa de la envoltura (gp41), que son necesarias para la infección de las células (Abbas, pp. 454-456). Gp140 es una forma modificada de la glucoproteína env que contiene la porción de glucoproteína externa de 120 kDa de la envoltura y una parte de la porción gp41 de env y tiene características tanto de gp120 como de gp41. El gen Nef se conserva entre los lentivirus de primates y es uno de los primeros genes virales que es transcrito luego de la infección. Se han descrito varias funciones in vitro, que incluyen el descenso regulado de la expresión superficial de CD4 y MHC de clase I, la señalización y activación de los linfocitos T alterados y una mayor infectividad viral.

La infección por el VIH se inicia con gp120 en la partícula viral que se une a CD4 y moléculas del receptor de quimiocinas (por ejemplo, CXCR4, CCR5) en la membrana celular de las células diana como los linfocitos T CD4+, macrófagos y células dendríticas. El virus unido se fusiona con la célula diana y realiza la transcripción inversa del genoma de ARN. El ADN viral resultante, se integra en el genoma celular, donde dirige la producción de un nuevo ARN viral y consecuentemente de proteínas virales y nuevos viriones. Esos viriones brotan de la membrana de la

célula infectada y establecen infecciones productivas en otras células. Este proceso también mata a la célula infectada originalmente. El VIH también puede matar a las células indirectamente porque el receptor CD4 de los linfocitos T no infectados tiene una fuerte afinidad por gp120 expresada en la superficie de las células infectadas. En este caso, las células no infectadas se unen, a través de la interacción receptor CD4-gp120, a las células infectadas y se fusionan para formar un sincicio, que no puede sobrevivir. La destrucción de linfocitos T CD4+, que son esenciales para la defensa inmunitaria, es una de las principales causas de la disfunción inmunitaria progresiva que es el sello característico del avance de la enfermedad del SIDA. La pérdida de linfocitos T CD4+ perjudica seriamente la capacidad del organismo para luchar contra la mayor parte de los invasores, pero tiene un impacto especialmente grave en las defensas contra virus, hongos, parásitos y algunas bacterias, incluidas las micobacterias.

Los diferentes aislados del VIH-1 se clasificaron en tres grupos: M (principal), O (extraño) y N (no M, no O). El Grupo M del VIH-1 domina la pandemia mundial del VIH (Gaschen et al., (2002) *Science* 296:2354-2360). Desde que el grupo M del VIH-1 inició su expansión en los humanos hace aproximadamente 70 años (Korber et al., *Retroviral Immunology*, Pantaleo et al., eds., Humana Press, Totowa, NJ, 2001, pp. 1-31), se diversificó rápidamente (Jung et al., (2002) *Nature* 418:144). El Grupo M del VIH-1 consiste en un número de diferentes clados (también conocidos como subtipos) así como variantes resultantes de la combinación de dos o más clados, conocidas como formas recombinantes circulantes (CRF). Los subtipos se definen como si tuvieran genomas que son al menos 25% únicos (AIDS epidemic update, diciembre de 2002). Se identificaron once clados y una letra designa cada subtipo. Cuando los clados se combinan entre sí y se establecen exitosamente en el ambiente, como puede ocurrir cuando una persona está infectada con dos subtipos diferentes de VIH, el virus resultante se conoce como una CRF. Hasta ahora, se identificaron aproximadamente 13 CRF. Los clados del VIH-1 también muestran preferencia geográfica. Por ejemplo, el clado A, el segundo clado más frecuente, prevalece en África oriental, mientras que el clado B es común en Europa, las Américas y Australia. El clado C, el subtipo más común, está difundido en el sur de África, India y Etiopía (AIDS epidemic update, diciembre de 2002). Incluso entre clados existe variabilidad del virus entre diferentes cepas y aislados virales.

Esta variabilidad genética del VIH crea un reto científico para el desarrollo de vacunas. Se ha sugerido un enfoque que consiste en desarrollar secuencias consenso basadas en las secuencias de múltiples cepas diferentes del VIH y en desarrollar vacunas basadas en esas secuencias consenso. El fundamento de esos enfoques es que las secuencias consenso codificarán antígenos que están conservados entre las diferentes cepas del VIH y que dichos antígenos por lo tanto, es probable que sean útiles en la generación de respuestas inmunitarias contra múltiples cepas diferentes del VIH. Las secuencias consenso del VIH-1 del clado A fueron generadas por otros. Véase por ejemplo, Nkolola et al. (2004) *Gene Ther.* 2004. Jul. 11 (13): 1068-80, y Korber B (eds) et al. *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos National Laboratory: Los Alamos, Nuevo México, Estados Unidos, (1997) que involucran los transgenes RENTA e HIVA derivados de las secuencias consenso del clado A. Sin embargo, las secuencias consenso descritas en esos artículos parecen haber sido derivadas de la secuencia consenso del VIH-1 del clado A obtenida del laboratorio de Los Alamos, y no se generaron de la misma forma que las secuencias consenso de la presente invención. Además, esas referencias no instruyen acerca del uso de secuencias de las cepas de VIH existentes recientemente circulantes que se corresponden estrechamente con la secuencia consenso. En su lugar involucran el uso de las secuencias consenso en sí mismas.

Malm, M. et al., *Viral Immunol.* 18:678-688, 1. enero de 2005 dan a conocer la protección cruzada de clados inducida por inmunógenos de ADN del VIH-1 que expresan secuencias consenso de múltiples genes y epítomos de los subtipos A, B, C y FGH. Los autores construyeron una plataforma de vacuna con un inmunógeno de ADN del VIH-1 subtipo B que expresa las secuencias consenso enteras rev, nef, tat y gag del VIH-1 con grupos de epítomos celulares adicionales de las regiones env y pol, donde esa plataforma se extendió a tres plásmidos adicionales que expresan los mismos inmunógenos pero provenientes de secuencias ancestrales de los subtipos A o C o FGH. La inmunogenia en ratones Balb/c puso de manifiesto una fuerte producción de IFN- $\gamma$  en respuesta a la reestimulación in vitro con un péptido gag restringido a H-2<sup>d</sup> o incluso más fuerte frente a un epítomo env, donde se detectó inmunidad humoral débil. La vacunación con pistola de genes con un cóctel de los cuatro plásmidos indujo inmunidad antes de la provocación y posteriormente una frecuencia robusta de protección después de la provocación experimental con VIH-1 subtipo A o B/virus de la leucemia murina. Por lo tanto, la protección cruzada de clados observada en este experimento de provocación demostró que esos inmunógenos multigen/multiépítomo de ADN del VIH es probable que sean potentes inmunógenos también contra la infección por el VIH.

WO 2005/047483 A2 proporciona fusiones artificiales diseñadas para provocar una respuesta inmunitaria anti-VIH, así como moléculas de ácido nucleico y vectores de expresión que codifican esas proteínas, donde las fusiones artificiales comprenden dominios de diversas proteínas del VIH, incluidas las proteínas RT, Env, Nef y Tat, así como al menos un epítomo CTL del VIH asociado con no evolución hacia el SIDA de largo plazo. RENTA es una fusión artificial en la que los dominios del VIH son de una secuencia consenso del VIH del clado A y contiene dominios adicionales, útiles por ejemplo, en la vigilancia de los niveles de expresión o la respuesta inmunitaria en animales de laboratorio, donde dichos dominios se incluyen opcionalmente en las fusiones artificiales.

WO 01/47955 A2 da a conocer un inmunógeno en forma estéril adecuado para la administración a un sujeto humano, donde el inmunógeno comprende: al menos una porción de la proteína gag del VIH, donde dicha proteína gag es de un clado de VIH o tiene una secuencia consenso para uno o más clados del VIH y comprende al menos partes de p17 y p24; y un polipéptido sintético compuesto por una pluralidad de secuencias de aminoácidos, donde cada secuencia comprende un epítipo CTL humano de una proteína del VIH, y donde una pluralidad de proteínas del VIH están representadas en el polipéptido sintético, donde dichos epítopos CTL se seleccionan para estimular una respuesta inmunitaria a uno o más clados del VIH de interés.

10 Resumen de la invención

La presente divulgación proporciona secuencias consenso nuevas y mejoradas para antígenos del VIH-1 del clado A. Las secuencias consenso son especialmente ventajosas porque se basan en las secuencias de antígenos de un gran número de cepas del VIH-1 del clado A diferentes, y también porque se basan en las secuencias de antígenos de cepas recientemente aislada del VIH-1 del clado A. En consecuencia, las secuencias consenso de la presente divulgación tienen mayor importancia biológica en comparación con las secuencias consenso del VIH-1 del clado A generadas previamente.

Otra ventaja importante de la presente divulgación es que proporciona antígenos del VIH-1 del clado A y estrategias para producir dichos antígenos, que se derivan de cepas de origen natural del VIH-1 del clado A. Esos antígenos se seleccionan de modo que estén estrechamente relacionados con, o tengan una pequeña "distancia de proteínas" respecto a, las secuencias consenso de la presente divulgación. Una ventaja de utilizar esas secuencias de origen natural con la máxima correspondencia con las secuencias consenso, por contraposición a las secuencias consenso generadas artificialmente, es que se necesitan menos manipulaciones genéticas para generar esas secuencias y se asegura la importancia biológica.

En un primer aspecto la presente invención apunta a un polinucleótido que codifica una proteína de fusión de un clado del VIH-1 donde la proteína de fusión comprende los antígenos Gag, Pol y Nef del VIH-1 del clado A y donde la proteína de fusión comprende la secuencia de aminoácidos de la figura 16. Particularmente, se refiere a secuencias consenso de aminoácidos para los antígenos del VIH-1 del clado A Gag, Pol (compuesto por RT e Int), Nef y Env. En realizaciones preferidas, la invención se refiere a la secuencia consenso de aminoácidos de Gag de la FIG. 1, la secuencia consenso de aminoácidos de Pol de la FIG. 3 o la secuencia consenso de aminoácidos de Nef de la FIG. 7.

La solicitud apunta a un método de identificación de una secuencia consenso de aminoácidos para un antígeno del VIH-1 del clado A de interés, que comprende determinar la secuencia de aminoácidos del antígeno de interés en varias cepas o aislados de campo circulantes del VIH-1, alinear dichas secuencias y determinar la secuencia consenso para ese antígeno.

Además, la solicitud se refiere a un método para identificar un antígeno del VIH-1 del clado A de una cepa circulante o un aislado de campo del VIH-1 del clado A que tenga una secuencia de aminoácidos que sea similar a la secuencia consenso de aminoácidos para ese antígeno del VIH-1 del clado A. En una realización preferida el antígeno del VIH-1 del clado A se selecciona basándose en el grado de similitud con la secuencia consenso, prefiriéndose las secuencias que tienen el mayor grado de similitud con, o la "distancia de proteínas" más pequeña respecto a, la secuencia consenso. En otra realización preferida el antígeno del VIH-1 del clado A se selecciona de una cepa o un aislado de campo recientemente circulante del VIH-1 del clado A. En otra realización la invención se refiere a los antígenos del VIH-1 del clado A identificados con dichos métodos.

Además, la solicitud se refiere a un método para identificar un antígeno del VIH-1 del clado A de una cepa o un aislado de campo circulante del VIH-1 del clado A que tiene una secuencia de aminoácidos similar a la secuencia consenso de aminoácidos para ese antígeno del VIH-1 del clado A y luego realizar mutaciones en la secuencia para anular las funciones biológicas de las secuencias. Es preferible utilizar un enfoque minimalista, es decir, que el número de mutaciones se mantenga en el mínimo, para que se realicen sólo las mutaciones necesarias para anular la función y facilitar la obtención de la aprobación de las autoridades de registro sanitario y se evite la alteración innecesaria de las secuencias de genes del VIH-1 originales. Por ejemplo, en una realización el componente Nef de GRIN no se altera sino más bien la fusión del extremo N-terminal de Nef con el extremo C-terminal de Int anula la función de nef conservando todas las secuencias de nucleótidos originales de Nef.

Además, la solicitud se refiere a un método para mejorar la estabilidad genética del transgén del VIH-1 del clado A para tecnologías de inserción en vectores virales. El componente PR (proteasa) se quita de Gag-Pol entero-Nef (Pol entero contiene PR, e Int y RT) para que queden sólo las porciones Int y RT de Pol. Esto tiene la ventaja de mayor estabilidad genética y mejores propiedades de clonación y rescate del virus, particularmente usando Ad35 y Ad11. Eliminar a PR de esta manera es un enfoque minimalista puesto que se elimina sólo la subunidad funcional más pequeña de POL, conservando las subunidades funcionales más grandes IN y RT. La solicitud también se refiere a

los antígenos del VIH-1 del clado A seleccionados y producidos utilizando estos métodos.

En una realización el antígeno es un antígeno Gag de una de las cepas indicadas en la tabla 1 y la figura 2. Preferentemente el antígeno Gag se selecciona de una cepa en la cual la "distancia de proteínas" respecto a la secuencia consenso de Gag es menor de 0.07%, o más preferentemente menor de 0.06%, o aún más preferentemente menor de 0.05%. En una realización preferida el antígeno Gag es de la cepa TZA173, la cepa 97TZ02, la cepa KNH1144 o la cepa SE7535UG del VIH-1 del clado A.

En otra realización el antígeno es un antígeno Pol de una de las cepas indicadas en la tabla 2 y la figura 4. Preferentemente el antígeno Pol se selecciona de una cepa en la cual la "distancia de proteínas" respecto a la secuencia consenso de Pol es menor de 0.03%, o más preferentemente menor de 0.025%. En una realización preferida el antígeno Pol es de la cepa MSA4070, la cepa SE7245SO o la cepa SE8538 del VIH-1 del clado A.

Otro antígeno es un antígeno Env de una de las cepas indicadas en la tabla 3 y la figura 6. Preferentemente el antígeno Env se selecciona de una cepa en la cual la "distancia de proteínas" de la secuencia consenso de Gag es menor de 0.1, o más preferentemente menor de 0.08%, o aún más preferentemente menor de 0.07% o todavía más preferentemente menor de 0.065%. En una realización preferida el antígeno Env es de la cepa KEQ23, la cepa TZA341 o la cepa KNH1088 del VIH-1 del clado A.

En otra realización el antígeno es un antígeno Nef de una de las cepas indicadas en la tabla 4 y la figura 8. Preferentemente el antígeno Nef se selecciona de una cepa en la cual la "distancia de proteínas" respecto a la secuencia consenso de Gag es menor de 0.1%, o más preferentemente menor de 0.08% o más preferentemente menor de 0.07%, o más preferentemente menor de 0.06, o aún más preferentemente, menor de 0.05%. En una realización preferida el antígeno Nef es de la cepa MSA4070, o la cepa KNH1211, o la cepa 97TZ03, o la cepa 99UGA070, o la cepa SE8891 UG del VIH-1 del clado A.

La presente invención apunta a las secuencias de nucleótidos que codifican los antígenos del VIH-1 del clado A según se define en la reivindicación 1. La solicitud también se refiere a los vectores que comprenden esas secuencias de nucleótidos. Las secuencias de nucleótidos de la invención y los vectores que las conforman y también los antígenos codificados por las secuencias de nucleótidos de la invención, son útiles para generar una respuesta inmunitaria contra antígenos del VIH del clado A in vivo y son útiles en la producción de vacunas contra cepas del VIH-1 del clado A. Las secuencias de nucleótidos de la invención también pueden ser útiles para expresar y producir los antígenos del VIH-1 del clado A que ellos codifican en las células o in vitro, por ejemplo, para que los antígenos se puedan producir, aislar y/o purificar.

Los nucleótidos de la invención pueden estar alterados en comparación con las secuencias consenso de nucleótidos, o en comparación con las secuencias de aislados circulantes del VIH-1 que están estrechamente relacionadas con dichas secuencias consenso. Por ejemplo, en una realización las secuencias de nucleótidos se pueden mutar para anular la actividad de las proteínas codificadas in vivo. En otra realización a las secuencias de nucleótidos se le puede hacer optimización de codones, por ejemplo, los codones se pueden optimizar para uso humano. Preferentemente, las secuencias de nucleótidos de la invención se mutan para anular la función normal in vivo de las proteínas codificadas y se les optimizan los codones para uso humano. Por ejemplo, cada una de las secuencias de Gag, Pol, Env, Nef, RT e Int se puede modificar de esas maneras.

En una realización preferida, una única secuencia de nucleótidos codifica una proteína de fusión que comprende los antígenos Gag, RT (parte de Pol) y Nef. En este documento las abreviaturas "GRN" y "GRtN" se usan indistintamente para referirse a las proteínas de fusión del VIH-1 del clado A que comprenden los antígenos Gag, RT y Nef y para referirse a las secuencias de nucleótidos que codifican esas proteínas de fusión. En otra realización aún más preferida la secuencia de nucleótidos que codifica a GRN se inserta en un vector adecuado para permitir la expresión de la proteína de fusión GRN. Preferentemente el vector es un vector de adenovirus seleccionado del grupo que consiste en Ad5, Ad35, Ad11, C6 y C7.

En otra realización preferida, una única secuencia de nucleótidos codifica una proteína de fusión que comprende los antígenos Gag, Pol (que incluye RT e INT) y Nef. En este documento las abreviaturas "GRN" y "GRtN" se usan indistintamente para referirse a las proteínas de fusión del VIH-1 del clado A que comprenden los antígenos Gag, Pol y Nef y para referirse a las secuencias de nucleótidos que codifican esas proteínas de fusión. En realizaciones aún más preferidas GRIN tiene la secuencia de aminoácidos que se ilustra en las figuras 16A-16J y es codificada por la secuencia de nucleótidos que se ilustra en las figuras 16A-16J. En una realización aún más preferida la secuencia de nucleótidos que codifica a GRIN se inserta en un vector adecuado para permitir la expresión de la proteína de fusión GRIN. Preferentemente el vector es un vector de adenovirus seleccionado del grupo que consiste en Ad5, Ad35, Ad11, C6 y C7.

En la divulgación una única secuencia de nucleótidos de la invención codifica un antígeno Env del VIH-1 del clado A. El antígeno Env tiene la secuencia de aminoácidos que se ilustra en las Figuras 17A-17 D y es codificado por la

secuencia de nucleótidos que se ilustra en las Figuras 17A-17 D. La secuencia de nucleótidos que codifica a Env se inserta en un vector adecuado para permitir la expresión de la proteína Env. Preferentemente el vector es un vector de adenovirus, más preferentemente un vector de adenovirus seleccionado del grupo que consiste en Ad5, Ad35, Ad11, C6 y C7.

5 La divulgación proporciona métodos para generar una respuesta inmunitaria contra antígenos del VIH-1 del clado A que comprenden administrar a un sujeto una secuencia de nucleótidos o un antígeno de acuerdo con la invención. En realizaciones preferidas el método para generar una respuesta inmunitaria contra el VIH-1 del clado A comprende administrar una secuencia de nucleótidos que codifica a GRIN o a GRN donde la secuencia de  
10 nucleótidos está contenida en un vector de adenovirus seleccionado del grupo que consiste en Ad5, Ad35, Ad11, C6 y C7. En otras realizaciones preferidas, los vectores que comprenden a GRIN o GRN se administran conjuntamente con un vector que comprende una secuencia de nucleótidos que codifica un antígeno Env de la invención.

15 En otra realización, la presente divulgación proporciona composiciones inmunógenas o composiciones para vacuna que comprenden las secuencias de nucleótidos de la divulgación. La invención proporciona una composición inmunógena que comprende el polinucleótido de la reivindicación 1, así como el polinucleótido de la reivindicación 1 para usar en un método para generar una respuesta inmunitaria contra el VIH-1. Cabe señalar que en esta divulgación y particularmente en las reivindicaciones o los párrafos, términos como "comprende", "contiene",  
20 "compuesto de" y análogos pueden tener el significado que se les atribuye en la ley de patentes de los Estados Unidos; por ejemplo, pueden significar "incluye", "incluido", "inclusive" y análogos; y que términos como "que consiste esencialmente en" y "se compone esencialmente de" tienen el significado que se les atribuye en la ley de patentes de los Estados Unidos, por ejemplo, dejan un margen para elementos no explícitamente enumerados, pero excluye elementos que se encuentran en el estado anterior de la técnica o que afectan una característica básica o nueva de la invención.

25 Estas y otras realizaciones se dan a conocer o son obvias a partir de, y abarcadas por, la descripción detallada siguiente.

30 Breve descripción de las figuras

La descripción detallada siguiente, que se proporciona a modo de ejemplo, pero no pretende limitar la invención a las realizaciones específicas descritas, se puede comprender mejor conjuntamente con las figuras adjuntas.

La figura 1 es una secuencia consenso de aminoácidos de la proteína Gag del VIH-1 del clado A.

35 La figura 2 es un gráfico que ilustra la "distancia" de las secuencias de la proteína Gag de cepas circulantes del VIH del clado A respecto a la de la secuencia consenso de la proteína Gag del VIH-1 del clado A.

La figura 3 es una secuencia consenso de aminoácidos de la proteína Pol del VIH-1 del clado A.

La figura 4 es un gráfico que ilustra la "distancia" de las secuencias de la proteína Pol de cepas circulantes del VIH del clado A respecto a la de la secuencia consenso de la proteína Pol del VIH-1 del clado A.

La figura 5 es una secuencia consenso de aminoácidos de la proteína Env del VIH-1 del clado A.

40 La figura 6 es un gráfico que ilustra la "distancia" de las secuencias de la proteína Env de cepas circulantes del VIH del clado A respecto a la de la secuencia consenso de la proteína Env del VIH-1 del clado A.

La figura 7 es una secuencia consenso de aminoácidos de la proteína Nef del VIH-1 del clado A.

La figura 8 es un gráfico que ilustra la "distancia" de las secuencias de la proteína Nef de cepas circulantes del VIH del clado A respecto a la de la secuencia consenso de la proteína Nef del VIH-1 del clado A.

45 La figura 9 es una representación esquemática de los transgenes GRIN y GRN.

La figura 10 ilustra la secuencia de aminoácidos de la proteína Gag de la cepa TZA173 del VIH-1 del clado A que tiene el número de registro de Genbank AY253305.

La figura 11 ilustra la secuencia de aminoácidos de la proteína Pol de la cepa MSA4070 del VIH-1 del clado A que tiene el número de registro de Genbank AF457081.

50 La figura 12 ilustra la secuencia de aminoácidos de la proteína Nef de la cepa MSA4070 del VIH-1 del clado A que tiene el número de registro de Genbank AF457081.

La figura 13 ilustra la secuencia de aminoácidos de la proteína Env de la cepa TZA341 del VIH-1 del clado A que tiene el número de registro de Genbank AY253314.

Las figuras 14A-14 C proporcionan una secuencia de GRIN como la insertada en el vector Ad35.

55 Las figuras 15A-15B proporcionan una secuencia de Env como la insertada en el vector Ad35.

Las figuras 16A-16J proporcionan secuencias de nucleótidos y aminoácidos con del transgén GRIN con codones optimizados.

Las figuras 17A-17 D proporcionan secuencias de nucleótidos y aminoácidos del transgén Env con codones optimizados.

60 La figura 18 ilustra gráficamente la inmunogenia de Ad5-GRIN y Ad5-GRN en ratones, medida por el ensayo de IFN-gamma ELIspot.

La figura 19 ilustra gráficamente la inmunogenia de C7-GRIN y C7-GRN en ratones, medida por el ensayo de IFN-gamma ELIspot.

La figura 20 ilustra gráficamente la inmunogenia de C7-GRIN y C7-GRN en ratones, medida por el ensayo de IL-2

ELISpot.

La figura 21 ilustra gráficamente la inmunogenia de C6-GRIN y C6-GRN en ratones, medida por el ensayo de IFN-gamma ELISpot.

La figura 22 ilustra gráficamente la inmunogenia de C6-GRIN y C6-GRN en ratones, medida por el ensayo de IL-2 ELISpot.

La figura 23A ilustra la inmunogenia por IFN- $\gamma$  ELISpot de Ad35-GRIN/ENV a la dosis de  $10^{10}$  vp luego de un programa de vacunación del mes 0 al 6 en macacus rhesus (monos de laboratorio). Definición de respuesta positiva: para una mezcla de un solo péptido de una sola muestra: respuesta = (media del recuento de péptido - media del recuento de no péptido). Para ser positiva, una respuesta de un sólo péptido debe ser: 1. Media del recuento de péptido  $>4x$  media del recuento de no péptido de la misma placa; 2. Coeficiente de variación entre recuentos repetidos  $\leq 70\%$  3. Respuesta  $> 55$  SFC/ $10^6$ . La media geométrica de las respuestas para células formadoras de manchas (SFC) por millón de PBMC (células mononucleares de sangre periférica) para cada componente antigénico (Gag, RT, IN y ENV) aparece en el eje Y y los tiempos de sangrado en semanas en el eje X.

La figura 23B ilustra la inmunogenia por IFN- $\gamma$  ELISpot de Ad35-GRIN/ENV a la dosis de  $10^{11}$  vp luego de un programa de vacunación del mes 0 al 6 en macacus rhesus. Definición de respuesta positiva: para una mezcla de un solo péptido de una sola muestra: respuesta = (media del recuento de péptido - media del recuento de no péptido). Para ser positiva, una respuesta de un sólo péptido debe ser: 1. Media del recuento de péptido  $> 4x$  media del recuento de no péptido de la misma placa; 2. Coeficiente de variación entre recuentos repetidos  $\leq 70\%$  3. Respuesta  $> 55$  SFC/ $10^6$ . La media geométrica de las respuestas para células formadoras de manchas (SFC) por millón de PBMC para cada componente antigénico (Gag, RT, IN y ENV) aparece en el eje Y y los tiempos de sangrado en semanas en el eje X.

### Descripción detallada de la invención

La presente divulgación se refiere a secuencias consenso de nucleótidos y proteínas para antígenos del VIH del clado A y a aislados de campo circulantes del VIH-1 que se corresponden estrechamente con esas secuencias consenso. La divulgación también se refiere a la versión alterada de esas secuencias, que puede ser alterada de modo que se anule la función de los productos génicos in vivo, a constructos y vectores que comprenden las secuencias de la invención, y a inmunógenos, composiciones inmunógenas y vacunas preparadas usando las secuencias de la invención. La divulgación también se refiere a métodos para generar una respuesta inmunitaria contra antígenos del VIH-1 del clado A en un sujeto y a métodos para inducir inmunidad protectora contra la provocación con el VIH-1. Las diversas realizaciones de la invención se resumen precedentemente en la sección titulada "Resumen de la invención". Se proporcionan más detalles de la invención en la Descripción detallada y los Ejemplos que siguen y también en las Figuras.

Como se describe en el "Resumen de la invención" anterior y los "Ejemplos" siguientes, la presente divulgación proporciona antígenos consenso del VIH-1 del clado A y también antígenos de cepas circulantes del VIH-1 del clado A que están estrechamente relacionadas con esas secuencias consenso. La divulgación también proporciona transgenes y antígenos del VIH-1 codificados por esos transgenes. Esos transgenes comprenden secuencias que codifican los antígenos del VIH-1 del clado A de la divulgación. Por ejemplo, en una realización preferida la presente divulgación proporciona un transgén GRIN (también denominado GRtIN) que comprende los antígenos Gag, Pol (ambos RT e Int) y Nef de la divulgación. La divulgación también proporciona un transgén GRN (también denominado GRtN) que comprende los antígenos Gag, RT y Nef de la divulgación.

La invención proporciona un polinucleótido que codifica una proteína de fusión del VIH-1 del clado A que comprende los antígenos Gag, Pol y Nef del VIH-1 del clado A, donde la proteína de fusión comprende la secuencia de aminoácidos de la figura 16, una composición inmunógena que contiene ese polinucleótido, así como ese polinucleótido para utilizar en un método para generar una respuesta inmunitaria contra el VIH-1.

Los términos "proteína", "péptido", "polipéptido" y "secuencia de aminoácidos" se usan indistintamente en este documento para referirse a polímeros de residuos de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede constar de aminoácidos modificados o análogos de aminoácidos, y puede estar interrumpido por grupos funcionales químicos distintos de los aminoácidos. Los términos también abarcan un polímero de aminoácidos que ha sido modificado de forma natural o por intervención; por ejemplo por formación de enlace disulfuro, glucosilación, lipidación, acetilación, fosforilación, o cualquier otra manipulación o modificación, como la conjugación con un componente de marcado o bioactivo.

En este documento, los términos "antígeno" o "inmunógeno" se usan indistintamente para referirse a una sustancia, normalmente una proteína, que es capaz de inducir una respuesta inmunitaria en un sujeto. El término también se refiere a las proteínas que son inmunológicamente activas en el sentido de que una vez administradas a un sujeto (ya sea directamente o mediante la administración al sujeto de una secuencia de nucleótidos o un vector que codifiquen la proteína) es capaz de provocar una respuesta inmunitaria de tipo humoral y/o celular dirigida contra esa proteína.

Se debe entender que las proteínas y los antígenos de la divulgación pueden diferir de las secuencias exactas ilustradas y descritas en este documento. Por consiguiente, la divulgación contempla deleciones, adiciones y sustituciones a las secuencias mostradas, en tanto las secuencias funcionen de conformidad con los métodos de la divulgación. A este respecto, las sustituciones particularmente preferidas serán generalmente de naturaleza conservadora, es decir, las sustituciones que tienen lugar dentro de una familia de aminoácidos. Por ejemplo, los aminoácidos se dividen generalmente en cuatro familias: (1) ácidos — aspartato y glutamato ; (2) básicos — lisina, arginina, histidina; (3) no polares — alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano; y (4) sin carga polar — glicina, asparragina, glutamina, cistina, serina, treonina, tirosina. Fenilalanina, triptófano y tirosina a veces se clasifican como aminoácidos aromáticos. Es razonablemente previsible que un reemplazo aislado de leucina, con isoleucina o valina, o viceversa; un aspartato con un glutamato o viceversa; una treonina con una serina o viceversa; o un reemplazo conservador similar de un aminoácido con un aminoácido estructuralmente relacionado, no tendrá un efecto importante en la actividad biológica. Las proteínas que tienen sustancialmente la misma secuencia de aminoácidos que las secuencias ilustradas y descritas pero poseen sustituciones de aminoácidos menores que no afectan sustancialmente la inmunogenia de la proteína están comprendidas, por consiguiente, por el alcance de la divulgación.

La solicitud describe secuencias "consenso" de aminoácidos para antígenos del VIH-1 del clado A. Se refiere a secuencias consenso de aminoácidos para los antígenos Gag, Pol (que comprende RT e Int), Nef y Env del VIH-1 del clado A . Se refiere a una secuencia consenso de aminoácidos de Gag de la FIG. 1, la secuencia consenso de aminoácidos de Pol de la FIG. 3, a una secuencia consenso de aminoácidos de Env de la FIG. 5, y/o a una secuencia consenso de aminoácidos de Nef de la figura 7. La solicitud apunta a un método de identificación de una secuencia consenso de aminoácidos para un antígeno del VIH-1 del clado A de interés, que comprende obtener la secuencia de aminoácidos del antígeno de interés en varias cepas o aislados de campo circulantes del VIH-1, alinear dichas secuencias y determinar la secuencia consenso para ese antígeno. Por ejemplo, en una realización se genera una base de datos utilizando secuencias disponibles para cepas circulantes no recombinantes del VIH-1 del clado A, y a continuación los genes individuales del VIH-1 (por ejemplo gag, pol, nef y env) de todas las secuencias de la base de datos se alinean, con insertando guiones para mantener el alineamiento en regiones con inserciones o deleciones en la secuencia, y luego se puede derivar una secuencia consenso de 50%.

La divulgación también se refiere a métodos para identificar antígenos de cepas del VIH-1 del clado A de origen natural que tienen una secuencia de aminoácidos con una "distancia de proteínas" pequeña respecto a la secuencia consenso de aminoácidos de ese antígeno. La "distancia de proteínas" es una medida del nivel de similitud o diferencia entre dos secuencias de aminoácidos. Dos secuencias de aminoácidos que son muy similares tienen una distancia de proteínas baja. Dos secuencias de aminoácidos que son muy diferentes tienen una distancia de proteínas alta. Las distancias de proteínas se calculan preferentemente utilizando la matriz de sustitución de Dayhoff PAM250 (M.O. Dayhoff, ed., 1978, Atlas of Protein Sequence and Structure, Vol. 5) que pondera las sustituciones según el grado de similitud bioquímica. Sin embargo, también se pueden usar otros métodos para determinar la distancia de proteínas.

En este documento los términos "secuencias de nucleótidos" y "secuencias de ácido nucleico" se refieren a ácido desoxirribonucleico (ADN) o secuencias de ácido ribonucleico (ARN), incluidos, pero no exclusivamente, ARN mensajero (ARNm), híbridos de ADN/ARN o ácidos nucleicos sintéticos. El ácido nucleico puede ser monocatenario, o parcial o completamente bicatenario (dúplex). Los ácidos nucleicos dúplex pueden ser homodúplex o heterodúplex.

Como se describe en el "Resumen de la invención" antes y en los "Ejemplos" más adelante, la presente divulgación proporciona antígenos consenso del VIH-1 del clado A y las secuencias de nucleótidos que codifican esos antígenos consenso. La divulgación también se refiere a antígenos de cepas circulantes del VIH-1 del clado A que están estrechamente relacionados con esas secuencias consenso, y a las secuencias de nucleótidos que los codifican. La divulgación también proporciona transgenes del VIH-1 del clado A que comprende secuencias que codifican antígenos del VIH-1 del clado A. Según se usa en este documento, el término "transgén" se usa para referirse a secuencias de nucleótidos "recombinantes" que derivan de secuencias consenso de nucleótidos del VIH-1 del clado A, o de secuencias de nucleótidos que codifican los antígenos de cepas recientemente circulantes del VIH-1 del clado A que se ha identificado que se corresponden estrechamente con esas secuencias consenso. El término "recombinante" significa una secuencia de nucleótidos que ha sido manipulada "por el hombre" y que no se encuentra en la naturaleza, o que está ligada a otra secuencia de nucleótidos o se encuentra en una distribución diferente en la naturaleza. Se entiende que manipulada "por el hombre" significa manipulada por algún medio artificial, que incluye el uso de máquinas, la optimización de codones, enzimas de restricción, etc. Por ejemplo, se proporcionan los transgenes GRIN, GRN y Env.

Los nucleótidos pueden estar alterados en comparación con las secuencias consenso de nucleótidos, o en comparación con las secuencias de aislados circulantes del VIH-1 que están estrechamente relacionadas con dichas secuencias consenso. Por ejemplo, en una realización las secuencias de nucleótidos se pueden mutar para anular la actividad de las proteínas codificadas in vivo. Las secuencias de nucleótidos pueden tener los codones optimizados, por ejemplo, los codones se pueden optimizar para uso humano. Preferentemente las secuencias de nucleótidos se



mutan para anular la función normal in vivo de las proteínas codificadas, y se les optimizan los codones para uso humano. Por ejemplo, cada una de las secuencias de Gag, Pol, Env, Nef, RT e Int se puede modificar de esas maneras.

- 5 Los tipos de mutaciones que se pueden realizar para anular la función in vivo de los antígenos incluyen, pero no exclusivamente, los siguientes que también se describen en el ejemplo 7: mutación de Gly2 a Ala en Gag para eliminar un sitio de miristilación y evitar la formación de partículas semejantes a virus (VLP); mutación de Gag para evitar deslizamiento en la secuencia de desplazamiento del marco de lectura natural para dejar la secuencia de aminoácidos conservada (NFLG) intacta y permitir que se traduzca sólo el producto de la proteína GagPol entera;
- 10 mutación de Asp185 de RT a Ala y mutación de Asp186 a Ala para inactivar los residuos enzimáticos activos. Mutación de Asp 64 de Int a Ala, y mutación de Asp116 a Ala y mutación de Glu 152 a Ala para inactivar los residuos enzimáticos activos.

15 Con respecto a la optimización de codones, las moléculas de ácido nucleico tienen una secuencia de nucleótidos que codifica los antígenos de esta divulgación y que puede ser diseñada para emplear codones que se usan en los genes del sujeto en el cual se va a producir el antígeno. Muchos virus, incluidos el VIH y otros lentivirus, utilizan un gran número de codones raros y, alterando esos codones para que correspondan a codones comúnmente utilizados en el sujeto deseado, se puede lograr una mayor expresión de los antígenos. Preferentemente, los codones utilizados son codones "humanizados", es decir, los codones son los que aparecen frecuentemente en genes humanos altamente expresados (Andre et al., J. Virol. 72:1497-1503, 1998) en lugar de los codones que son frecuentemente utilizados por el VIH. Dicho uso de codones favorece la expresión eficaz de las proteínas del VIH transgénicas en células humanas. Se puede utilizar cualquier método de optimización de codones adecuado. Por ejemplo los codones se pueden optimizar para uso humano según se ilustra en el ejemplo 8. No obstante, se puede utilizar cualquier otro método adecuado de optimización de codones. Dichos métodos y la selección de dichos

20 métodos son bien conocidos por los expertos. Además existen varias compañías que optimizarán codones de secuencias como Geneart (geneart.com). Por lo tanto, las secuencias de nucleótidos se pueden optimizar fácilmente.

25 La solicitud abarca además secuencias de nucleótidos que codifican variantes y derivados funcional y/o antigénicamente equivalentes de los antígenos de la divulgación y sus fragmentos funcionalmente equivalentes. Esas variantes, derivados y fragmentos funcionalmente equivalentes, tienen la capacidad de conservar la actividad antigénica. Por ejemplo, los cambios en una secuencia de ADN que no cambian la secuencia de aminoácidos codificada, así como aquellos que dan como resultado sustituciones de residuos de aminoácidos conservadoras, una o unas pocas deleciones o adiciones de aminoácidos y sustitución de residuos de aminoácidos por análogos de aminoácidos son las que no afectarán significativamente las propiedades del polipéptido codificado. Las sustituciones de aminoácidos conservadoras son glicina/alanina; valina/isoleucina/leucina; asparragina/glutamina; ácido aspártico/ácido glutámico; serina/treonina/metionina; lisina/arginina; y fenilalanina/tirosina/triptofano. Las variantes tienen al menos 50%, al menos 55%, al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97 %, al menos 98% o al menos 99% de homología o identidad con el antígeno, epítipo, inmunógeno, péptido o polipéptido de interés.

30 La identidad de secuencia u homología se determina comparando las secuencias cuando se alinean para maximizar la superposición y la identidad minimizando al mismo tiempo los huecos de la secuencia. En particular, la identidad de secuencia se puede determinar utilizando cualquiera de una serie de algoritmos matemáticos. Un ejemplo no limitante de un algoritmo matemático utilizado para comparar dos secuencias, es el algoritmo de Karlin & Altschul, Proc. Natl. Acad. Sci. USA. 1990; 87:2264-2268, modificado como en Karlin & Altschul, Proc. Natl. Acad. Sci. USA 1993; 90:5873-5877.

35 Otro ejemplo de un algoritmo matemático utilizado para comparar secuencias es el algoritmo de Myers & Miller, CABIOS 1988;4: 11-17. Dicho algoritmo se incorpora en el programa ALIGN (versión 2.0) que forma parte del paquete de software de alineamiento de secuencia GCG. Cuando se utiliza el programa ALIGN para comparar secuencias de aminoácidos, se pueden utilizar una tabla de ponderación de residuos PAM120, una penalización de longitud de hueco de 12 y una penalización de hueco de 4. Aún otro algoritmo útil para identificar regiones de similitud de secuencia local y alineamiento es el algoritmo FASTA como se describe en Pcarson & Lipman, Proc. Natl. Acad. Sci. USA 1988; 85:2444-2448.

40 Resulta ventajoso el uso del software WU-BLAST (Washington University BLAST) versión 2.0. Los programas ejecutables WU-BLAST versión 2.0 para varias plataformas UNIX se pueden descargar de ftp ://blast.wustl.edu/blast/executables. Este programa se basa en WU-BLAST versión 1.4, que a su vez se basa en el dominio público NCBI-BLAST versión 1.4 (Altschul y Gish, 1996, Local alignment statistics, Doolittle ed., Methods in Enzymology 266: 460-480; Altschul et al., Journal of Molecular Biology 1990; 215: 403-410; Gish y States, 1993; Nature Genetics 3: 266-272; Karlin y Altschul, 1993; Proc. Natl. Acad. Sci. EE.UU 90: 5873-5877).

Las diversas secuencias de nucleótidos y transgenes recombinantes se realizan usando técnicas estándar de recombinación de ADN y de clonación. Dichas técnicas son bien conocidas por los expertos. Véase por ejemplo, "Molecular Cloning: A Laboratory Manual", segunda edición (Sambrook et al., 1989).

5 Las secuencias de nucleótidos de la presente invención se pueden insertar en "vectores". El término "vector" es ampliamente utilizado y entendido por los expertos, y en este documento el término "vector" se utiliza coherentemente con su significado para los expertos. Por ejemplo, el término "vector" es utilizado comúnmente por los entendidos en la materia para referirse a un vehículo que permite o facilita la transferencia de moléculas de ácido nucleico de un ambiente a otro, o que permite o facilita la manipulación de una molécula de ácido nucleico.

10 Se puede utilizar cualquier vector que permita la expresión de los transgenes del VIH-1 del clado A de conformidad con la presente divulgación. En ciertas realizaciones, se pueden utilizar transgenes del VIH-1 del clado A de la presente divulgación in vitro (como el uso de sistemas de expresión exentos de células) y/o en cultivos celulares in vitro para producir los antígenos del VIH-1 codificados que se pueden utilizar para diversas aplicaciones, como en la producción de vacunas proteínicas. Para este tipo de aplicaciones, se puede utilizar cualquier vector que permita la expresión de los transgenes del VIH-1 del clado A in vitro o en células en cultivo.

15 Para las aplicaciones en las que se desea que los transgenes se expresen in vivo, por ejemplo cuando se utilizan transgenes en vacunas de ADN o que contienen ADN, se puede utilizar cualquier vector que permita la expresión de los transgenes del VIH-1 del clado A de la presente divulgación y sea seguro para usar in vivo. Los vectores utilizados son seguros para usar en seres humanos, otros mamíferos y animales de laboratorio.

20 Para que los transgenes se expresen, la secuencia de codificación de la proteína debe estar "unida operativamente" a las secuencias reguladoras o de control de ácido nucleico que dirigen la transcripción y traducción de la proteína. Según se usa en este documento, se dice que una secuencia de codificación y una secuencia de ácido nucleico de control o promotora están "unidas operativamente" cuando están unidas covalentemente de manera de colocar la expresión o la transcripción y/o la traducción de la secuencia de codificación bajo la influencia o el control de la secuencia de control de ácido nucleico. La "secuencia de control de ácido nucleico" puede ser cualquier elemento de ácido nucleico, como, pero no exclusivamente, promotores, potenciadores, IRES, intrones y otros elementos descritos que dirigen la expresión de una secuencia de ácido nucleico o secuencia de codificación que está unida operativamente a éstos. En este documento se utilizará el término "promotor" para referirse a un grupo de módulos de control transcripcional que se agrupan alrededor del sitio de iniciación de la ARN polimerasa II y que cuando están unidos operativamente a las secuencias de codificación de las proteínas de la divulgación dan lugar a la expresión de la proteína codificada. La expresión de los transgenes puede estar bajo el control de un promotor constitutivo o de un promotor inducible, que inicia la transcripción sólo cuando se expone a algún estímulo externo particular, como, sin limitación, antibióticos como tetraciclina, hormonas como ecdisona, o metales pesados. El promotor también puede ser específico de un determinado tipo de célula, tejido u órgano. Se conocen muchos promotores y potenciadores adecuados en el área, y se puede utilizar cualquier promotor o potenciador adecuado para la expresión de los transgenes de la divulgación. Por ejemplo, los promotores y/o potenciadores adecuados se pueden seleccionar de la base de datos Eukaryotic Promoter Database (EPDB).

40 Normalmente los vectores se deberían elegir de modo que contengan una región reguladora del gen adecuada, como un promotor o un potenciador, para que los transgenes se puedan expresar.

45 Por ejemplo, cuando el objetivo es expresar los transgenes in vitro, o en cultivos de células, o en cualquier sistema procariota o eucariota, con el fin de producir la(s) proteína(s) codificada(s) por ese transgén, entonces se puede utilizar cualquier vector adecuado dependiendo de la aplicación. Por ejemplo, se pueden utilizar plásmidos, vectores virales, vectores bacterianos, vectores de protozoarios, vectores de insectos, vectores de expresión de baculovirus, vectores de levadura, vectores de células de mamífero y similares. Los vectores adecuados pueden ser seleccionados por los técnicos con experiencia teniendo en cuenta las características del vector y los requisitos para la expresión de los transgenes en las circunstancias identificadas.

50 Cuando el objetivo es expresar los transgenes in vivo en un sujeto, por ejemplo, para generar una respuesta inmunitaria contra un antígeno del HIV-1 y/o inmunidad protectora contra el VIH-1, se deben elegir los vectores de expresión que son adecuados para la expresión en el sujeto, y que son seguros para usar in vivo. Por ejemplo, en algunas realizaciones puede ser deseable expresar los transgenes en un animal de laboratorio, como para las pruebas preclínicas de las composiciones inmunógenas del VIH-1 y las vacunas de la invención. En otras realizaciones, será deseable expresar los transgenes en sujetos humanos, como en los ensayos clínicos y para el uso clínico real de las composiciones inmunógenas y las vacunas de la invención. Puede emplearse cualquier vector adecuado para este tipo de usos, y está dentro de las capacidades de los técnicos con experiencia seleccionar un vector adecuado. En algunas realizaciones puede ser preferible que a los vectores utilizados para esas aplicaciones in vivo se les atenúe la amplificación en el sujeto. Por ejemplo, si se utilizan vectores plasmídicos, preferentemente carecerán de un origen de replicación que funcione en el sujeto con el fin de mejorar la seguridad para el uso in vivo en el sujeto. Si se utilizan vectores virales, preferentemente se atenúan o no se replican defectuosamente en el

sujeto, una vez más, con el fin de mejorar la seguridad para el uso in vivo en el sujeto.

En las realizaciones preferidas se utilizan vectores virales. Los vectores de expresión viral son bien conocidos por los expertos en el área y son, por ejemplo, virus como los adenovirus, virus adeno-asociados (AAV), alfavirus, herpesvirus, retrovirus y poxvirus, incluidos virus avipox, poxvirus atenuados, virus vaccinia y particularmente, el virus vaccinia Ankara modificado (MVA; N° de registro ATCC VR-1566). Dichos virus, cuando se utilizan como vectores de expresión son innatamente no patógenos en los sujetos seleccionados como los seres humanos o fueron modificados para tornarlos no patógenos en los sujetos seleccionados. Por ejemplo, son bien conocidos los alfavirus y adenovirus que tiene replicación defectuosa y se pueden utilizar como vectores para la administración de genes.

En realizaciones particularmente preferidas se utilizan vectores de adenovirus. Se conocen en el área muchos vectores de adenovirus y cualquiera de dichos vectores adecuados puede ser utilizado. En realizaciones preferidas el vector de adenovirus utilizado se selecciona del grupo que consiste en los vectores Ad5, Ad35, Ad11, C6 y C7.

La secuencia del genoma del Adenovirus 5 "(Ad5) fue publicada. (Chroboczek, J., Bieber, F., y Jacrot, B. (1992) The Sequence of the Genome of Adenovirus Type 5 and Its Comparison with the Genome of Adenovirus Type 2, Virology 186, 280-285). Los vectores Ad35 se describen en las patentes de los Estados Unidos N° 6,974,695, 6,913,922 y 6,869,794. Los vectores Ad11 se describen en la patente de los Estados Unidos N° 6,913,922. Los vectores del adenovirus C6 se describen en las patentes de los Estados Unidos N° 6,780,407; 6,537,594; 6,309,647; 6,265,189; 6,156,567; 6,090,393; 5,942,235 y 5,833,975. Los vectores C7 se describen en la patente de los Estados Unidos N° 6,277,558.

También se pueden utilizar vectores de adenovirus con E1 defectuoso o eliminado, E3 defectuoso o eliminado, o E4 defectuoso o eliminado. Ciertos adenovirus con mutaciones en la región E1 tienen un mayor margen de seguridad porque los mutantes de adenovirus con E1 defectuoso tienen una replicación defectuosa en células no permisivas, o, al menos, están muy atenuados. Los adenovirus con mutaciones en la región E3 pueden tener una mayor inmunogenia al interrumpir el mecanismo mediante el cual el adenovirus regula por disminución las moléculas del MHC clase I. Los adenovirus que tienen mutaciones E4 pueden tener el vector de adenovirus con menor inmunogenia debido a la supresión de la expresión génica tardía. Dichos vectores pueden ser particularmente útiles cuando se desea la revacunación repetida utilizando el mismo vector. Se pueden utilizar vectores de adenovirus con eliminación de, o mutación en, E1, E3, E4, E1 y E3 y E1 y E4.

Además, también se pueden emplear vectores de adenovirus "vacíos" (gutless), en los cuales se eliminan todos los genes virales. Dichos vectores requieren un virus cooperador para su replicación y requieren una línea celular 293 humana especial que exprese E1a y Cre, una condición que no existe en el ambiente natural. Dicho vectores "vacíos" son no inmunógenos y por lo tanto los vectores se pueden inocular varias veces para revacunación. Los vectores de adenovirus "vacíos" se pueden utilizar para la inserción de insertos/genes heterólogos como los transgenes e incluso se pueden utilizar para co-administración de un gran número de insertos/genes heterólogos.

La solicitud también abarca un diseño que pone a Env y GRIN en vectores separados para permitir la evaluación de si la inclusión de Env es beneficiosa o perjudicial en cuanto a la inmunidad mediada por células (CMI) y la eficacia protectora. Los beneficios o perjuicios de Env en CMI y la eficacia protectora sigue siendo una incógnita en el campo de la vacuna del VIH.

Las secuencias de nucleótidos y los vectores de la divulgación se pueden administrar a las células, por ejemplo si el objetivo es expresar los antígenos del VIH-1 en las células para producir y aislar las proteínas expresadas, como de las células en cultivo. Para expresar los transgenes en las células se puede utilizar cualquier método adecuado de transfección, transformación o administración de genes. Dichos métodos son bien conocidos por los expertos en el área, y un experto podrá seleccionar fácilmente un método adecuado según la naturaleza de las secuencias de nucleótidos, los vectores y los tipos de células utilizados. Por ejemplo, se podría utilizar transfección, transformación, microinyección, infección, electroporación, lipofección o administración mediada por liposomas. La expresión de los antígenos se puede realizar en cualquier tipo adecuado de célula huésped, como las células bacterianas, las levaduras, las células de insectos y las células de mamíferos. También se pueden expresar los antígenos del VIH-1 del clado A utilizando inclusive sistemas de transcripción y traducción in vitro. Todos esos métodos son bien conocidos por los expertos en el área, y un experto podrá seleccionar fácilmente un método adecuado según la naturaleza de las secuencias de nucleótidos, los vectores y los tipos de células utilizados.

Después de la expresión, los antígenos se pueden aislar y/o purificar o concentrar mediante cualquier técnica adecuada conocida en el área. Por ejemplo, se puede utilizar cromatografía de intercambio aniónico o catiónico, cromatografía en fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía por afinidad, cromatografía de inmunofinidad, cromatografía de hidroxapatita, cromatografía de lectina, cromatografía de tamiz molecular, isoelectroenfoque, electroforesis en gel, o cualquier otro método adecuado o combinación de métodos.

- En realizaciones preferidas, las secuencias de nucleótidos y/o los antígenos se administran in vivo, por ejemplo cuando el objetivo es producir una respuesta inmunógena en un sujeto. Un "sujeto" en el contexto de la presente invención puede ser cualquier animal. Por ejemplo, en algunas realizaciones puede ser deseable expresar los transgenes en un animal de laboratorio, como para las pruebas preclínicas de las composiciones inmunógenas del VIH-1 y las vacunas de la invención. En otras realizaciones, será deseable expresar los transgenes en sujetos humanos, como en los ensayos clínicos y para el uso clínico real de las composiciones inmunógenas y las vacunas de la invención. En realizaciones preferidas, el sujeto es un ser humano, por ejemplo un ser humano que está infectado, o corre riesgo de infección, con el VIH-1.
- Para dichas aplicaciones in vivo las secuencias de nucleótidos y/o los antígenos se administran preferentemente como un componente de una composición inmunógena que comprende las secuencias de nucleótidos y/o los antígenos en mezcla con un vehículo farmacéuticamente aceptable. Las composiciones inmunógenas de la invención son útiles para estimular una respuesta inmunitaria contra el VIH-1 y se pueden usar como uno o más componentes de una vacuna profiláctica o terapéutica contra el VIH-1 para la prevención, la mejora o el tratamiento del SIDA. Los ácidos nucleicos de la invención son particularmente útiles para proporcionar vacunas genéticas, es decir, vacunas para la administración de ácidos nucleicos que codifican los antígenos del VIH-1 del clado A a un sujeto como un ser humano, de modo que los antígenos del VIH-1 del clado A se expresen a continuación en el sujeto para provocar una respuesta inmunitaria.
- Las composiciones de la invención pueden ser suspensiones inyectables, soluciones, aerosoles, polvos liofilizados, jarabes, elixires y similares. Se puede utilizar cualquier forma adecuada de la composición. Para preparar una composición de ese tipo, se mezcla un ácido nucleico o vector que tenga el grado de pureza deseado, con uno o más vehículos y/o excipientes farmacéuticamente aceptables. Los vehículos y los excipientes deben ser "aceptables" en el sentido de ser compatibles con los demás ingredientes de la composición. Los vehículos, excipientes o estabilizantes aceptables no ser tóxicos para los destinatarios en las dosis y las concentraciones empleadas e incluyen, pero no exclusivamente, agua, solución salina, solución salina amortiguada con fosfato, dextrosa, glicerol, etanol, o sus combinaciones, soluciones amortiguadoras, tales como fosfato, citrato y otros ácidos orgánicos; antioxidantes como ácido ascórbico y metionina; conservantes (como cloruro de octadecildimetilbencil amonio; cloruro de hexametonio; cloruro de benzalconio, cloruro de bencetonio; fenol, alcohol butílico o bencílico; alquilparabenos como metil o propilparabeno; catecol; resorcinol; ciclohexanol; 3-pentanol; y m-cresol); polipéptidos de bajo peso molecular (menos de aproximadamente 10 residuos); proteínas como seroalbúmina, gelatina o inmunoglobulinas; polímeros hidrófilos como polivinilpirrolidona; aminoácidos como glicina, glutamina, asparragina, histidina, arginina o lisina; monosacáridos, disacáridos y otros carbohidratos como glucosa, manosa o dextrinas; agentes quelantes como EDTA; azúcares como sacarosa, manitol, trehalosa o sorbitol; contraiones formadores de sales como sodio; complejos metálicos (por ejemplo, complejos de Zn-proteína); y/o tensioactivos no iónicos como TWEEN™, PLURONICS™ o polietilenglicol (PEG).
- Una composición inmunógena o inmunológica también se puede formular en forma de una emulsión de aceite en agua. La emulsión de aceite en agua se puede basar, por ejemplo, en aceite de parafina líquida ligera (tipo Farmacopea Europea); aceite de isoprenoides como escualano, escualeno, EICOSANE™ o tetratetracontano; aceite resultante de la oligomerización de alqueno(s), p. ej., isobuteno o deceno; ésteres de ácidos o de alcoholes que contienen un grupo alquilo lineal, tales como aceites vegetales, oleato de etilo, di(caprilato/caprato)de propilenglicol, tri(caprilato/caprato) de glicerilo o dioleato de propilenglicol; ésteres de ácidos grasos o alcoholes ramificados, por ejemplo, ésteres de ácido isosteárico. El aceite se utiliza ventajosamente en combinación con emulsionantes para formar la emulsión. Los emulsionantes pueden ser tensioactivos no iónicos, como ésteres de sorbitán, mannide (p. ej., oleato de anhidromanitol), glicerol, poliglicerol, propilenglicol y ácido oleico, isosteárico, ricinoleico o hydroxiesteárico, opcionalmente etoxilados, y copolímeros de bloque de polioxipropileno-polioxietileno, como los productos Pluronic®, por ejemplo, L121. El adyuvante puede ser una mezcla de emulsionante(s), agente de formación de micela y aceite como el que se comercializa con el nombre de Provac® (IDEC Pharmaceuticals, San Diego, CA).
- Las composiciones inmunógenas de la invención pueden contener sustancias adicionales, como humectantes o emulsionantes, soluciones amortiguadoras o adyuvantes para mejorar la eficacia de las vacunas (Remington's Pharmaceutical Sciences, 18ª edición, Mack Publishing Company, (ed.) 1980).
- También se pueden incluir adyuvantes. Los adyuvantes incluyen, pero no exclusivamente, sales minerales (por ej.,  $AlK(SO_4)_2$ ,  $AlNa(SO_4)_2$ ,  $AlNH(SO_4)_2$ , sílice, alum,  $Al(OH)_3$ ,  $Ca_3(PO_4)_2$ , caolín, o carbón), polinucleótidos con o sin complejos inmunoestimulantes (ISCOMs) (por ej., oligonucleótidos CpG, como los descritos en Chuang, T.H. et al, (2002) J. Leuk. Biol. 71(3): 538-44; Ahmad-Nejad, P. et al (2002) Eur. J. Immunol. 32(7): 1958-68; ácidos poli IC o poli AU, poliarginina con o sin CpG (también conocida como IC31; véase Schellack, C. et al (2003) Proceedings of the 34th Annual Meeting of the German Society of Immunology; Lingnau, K. et al (2002) Vaccine 20(29-30): 3498-508), JuvaVax™ (patente de los EE.UU. N° 6,693,086), ciertas sustancias naturales (por ej., cera D de Mycobacterium tuberculosis, sustancias encontradas en Corynebacterium parvum, Bordetella pertussis, o miembros del género Brucella), flagelina (Toll-like receptor 5 ligand; véase McSorley, S.J. et al (2002) J. Immunol. 169(7):

3914-9), saponinas como QS21, QS17 y QS7 ( patentes de los EE.UU. N° 5,057,540; 5,650,398; 6,524,584; 6,645,495), monofosforil lípido A, en particular, 3-de-O-acilado monofosforil lípido A (3D-MPL), imiquimod (también conocido como IQM disponible en el comercio como Aldara®; patentes de los EE.UU. N° 4,689,338; 5,238,944; Zuber, A.K. et al (2004) 22(13-14): 1791-8), y el inhibidor CCR5 CMPD167 (véase Veazey, R.S. et al (2003) J. Exp. Med. 198: 1551-1562).

Hidróxido o fosfato de aluminio (alumbre) se utilizan comúnmente en una solución 0.05 a 0.1% en solución salina amortiguada con fosfato. Otros adyuvantes que se pueden utilizar, especialmente con las vacunas de ADN, son la toxina del cólera, especialmente CTA1-DD/ISCOM (véase Mowat, A.M. et al (2001) J. Immunol. 167(6): 3398-405), polifosfazenos (Allcock, H.R. (1998) App. Organometallic Chem. 12(10-11): 659-666; Payne, L.G. et al (1995) Pharm. Biotechnol. 6: 473-93), citocinas como, pero no exclusivamente, IL-2, IL-4, GM-CSF, IL-12, IL-15 IGF-1, IFN- $\alpha$ , IFN- $\beta$  e IFN- $\gamma$  (Boyer et al., (2002) J. Liposome Res. 121:137-142; WO01/095919), proteínas inmunorreguladoras como CD40L (ADX40; véase, por ejemplo, WO03/063899) y el ligando CD1a de linfocitos citotóxicos naturales (también conocido como CRONY o  $\alpha$ -galactosilceramida; véase Green, T.D. et al, (2003) J. Virol. 77(3): 2046-2055), proteínas de fusión inmunoestimulantes como IL-2 fusionada con el fragmento Fc de las inmunoglobulinas (Barouch et al., Science 290:486-492, 2000) y moléculas coestimulantes B7.1 y B7.2 (Boyer), todos los cuales se pueden administrar como proteínas o en forma de ADN, en los mismos vectores de expresión que los que codifican los antígenos de la invención o en vectores de expresión independientes.

Las composiciones inmunógenas se pueden diseñar para introducir los antígenos, ácidos nucleicos o vectores de expresión del VIH-1 del clado A en un sitio de acción deseado y liberarlo a una velocidad apropiada y controlable. En el área se conocen métodos de preparación de formulaciones de liberación controlada. Por ejemplo, se pueden producir preparaciones de liberación controlada mediante el uso de polímeros para complejear o absorber el inmunógeno y/o la composición inmunógena. Se pueden preparar formulaciones de liberación controlada utilizando macromoléculas apropiadas (por ejemplo, poliésteres, ácidos poliamino, polivinilo, pirrolidona, etilenoivinilacetato, metilcelulosa, carboximetilcelulosa o sulfato de protamina) que se sabe que proporcionan las características de liberación controlada o el perfil de liberación deseados. Otro método posible para controlar la duración de la acción por una preparación de liberación controlada es incorporar los principios activos en las partículas de un material polimérico como, por ejemplo, poliésteres, ácidos poliamino, hidrogeles, ácido poliláctico, ácido poliglicólico, copolímeros de esos ácidos o copolímeros de vinilacetato de etileno. Alternativamente, en lugar de incorporar esos ingredientes activos en partículas poliméricas, es posible atrapar esos materiales en microcápsulas preparadas, por ejemplo, mediante técnicas de coacervación o por polimerización interfacial, por ejemplo, microcápsulas de hidroximetilcelulosa o gelatina y poli-(metilmacrilato), respectivamente, en sistemas de administración de fármacos coloidales (por ejemplo, liposomas, microesferas de albúmina, microemulsiones, nanopartículas y nanocápsulas) o en macroemulsiones. Dichas técnicas se dan a conocer en New Trends and Developments in Vaccines, Voller et al. (eds.), University Park Press, Baltimore, Md., 1978 y Remington's Pharmaceutical Sciences, 16ª edición.

Las dosis adecuadas de los antígenos, ácidos nucleicos y vectores de expresión (colectivamente, los inmunógenos) del VIH-1 del clado A en la composición inmunógena, pueden ser determinadas fácilmente por los expertos. Por ejemplo, la dosificación de los inmunógenos puede variar dependiendo de la vía de administración y el tamaño del sujeto. Las dosis adecuadas pueden ser determinadas por los expertos, por ejemplo, midiendo la respuesta inmunitaria de un sujeto, como un animal de laboratorio, utilizando técnicas inmunológicas convencionales, y ajustando las dosis según corresponda. Dichas técnicas para medir la respuesta inmunitaria del sujeto incluyen, pero no exclusivamente, ensayos de liberación de cromo, ensayos de unión al tetrámero, ensayos de IFN- $\gamma$  ELISPOT, ensayos de IL-2 ELISPOT, ensayos de citocinas intracelulares y otros ensayos de detección inmunológicos, por ejemplo, como los detallados en el texto "Antibodies: A Laboratory Manual" por Ed Harlow y David Lane.

Cuando se proporcionan profilácticamente las composiciones inmunógenas de la invención se administran idealmente a un sujeto antes de la infección por el VIH, o de evidencia de infección por el VIH, o antes de cualquier síntoma debido al SIDA, especialmente en sujetos de alto riesgo. La administración profiláctica de las composiciones inmunógenas puede servir para proporcionar inmunidad protectora a un sujeto contra la infección por el VIH-1 o para evitar o atenuar el avance del SIDA en un sujeto ya infectado con el VIH-1. Cuando se proporcionan terapéuticamente, las composiciones inmunógenas pueden servir para mejorar y tratar los síntomas del SIDA y se utilizan ventajosamente tan pronto después de la infección como sea posible, preferentemente antes de la aparición de cualquier síntoma del SIDA pero también se pueden utilizar en (o después de) la aparición de los síntomas de la enfermedad.

Las composiciones inmunógenas se pueden administrar usando cualquier método de administración adecuado, incluidos, pero no exclusivamente, la administración intramuscular, intravenosa, intradérmica, por la mucosa y tópica. Dichas técnicas son bien conocidas por los expertos. Ejemplos más específicos de métodos de administración son la inyección intramuscular, la inyección intradérmica y la inyección subcutánea. Sin embargo, la administración no necesita estar limitada a los métodos de inyección. Además, la administración de ADN al tejido animal se logró mediante liposomas catiónicos (Watanabe et al., (1994) Mol. Reprod. Dev. 38:268-274; y WO 96/20013), inyección directa de ADN desnudo en el tejido muscular del animal (Robinson et al., (1993) Vaccine

11:957-960; Hoffman et al., (1994) Vaccine 12: 1529-1533; Xiang et al., (1994) Virology 199: 132-140; Webster et al., (1994) Vaccine 12: 1495-1498; Davis et al., (1994) Vaccine 12: 1503-1509; y Davis et al., (1993) Hum. Mol. Gen. 2: 1847-1851), o inyección intradérmica de ADN usando tecnología de "pistola de genes" (Johnston et al., (1994) Meth. Cell Biol. 43:353-365). Alternativamente, las vías de administración pueden ser oral, intranasal o cualquier otra vía adecuada. La administración también se puede llevar a cabo a través de una superficie de mucosa como la mucosa oral, vaginal o anal.

Los programas de vacunación (o regímenes) para los animales (incluidos los humanos) son bien conocidos y se pueden determinar fácilmente para el sujeto y la composición inmunógena particulares. Por lo tanto, se pueden administrar los inmunógenos una o más veces al sujeto. Preferentemente, hay un intervalo de tiempo establecido entre administraciones separadas de la composición inmunógena. Si bien este intervalo varía para cada sujeto, normalmente varía desde 10 días hasta varias semanas y a menudo es de 2, 4, 6 u 8 semanas. Para los seres humanos, el intervalo es normalmente de 2 a 6 semanas. Los regímenes de vacunación habitualmente implican de 1 a 6 administraciones de la composición inmunógena, pero pueden ser tan sólo uno o dos o cuatro. Los métodos para inducir una respuesta inmunitaria también pueden incluir la administración de un adyuvante con los inmunógenos. En algunos casos, vacunas de refuerzo anuales, bianuales u a otros intervalos largos (5-10 años) pueden complementar el protocolo de vacunación inicial.

Los métodos de la presente también incluyen diversos regímenes de sensibilización/refuerzo, especialmente regímenes de sensibilización con ADN/refuerzo con adenovirus. En esos métodos, una o más vacunas de sensibilización son seguidas por una o más vacunas de refuerzo. La composición inmunógena real puede ser la misma o diferente en cada vacunación y el tipo de composición inmunógena (por ejemplo, que contiene proteína o vector de expresión), la vía y la formulación de los inmunógenos también pueden variar. Por ejemplo, si se usa un vector de expresión para los pasos de sensibilización y refuerzo, puede ser del mismo tipo o de otro tipo (por ejemplo, ADN o vector de expresión bacteriano o viral). Un régimen de sensibilización/refuerzo útil prevé dos vacunas de sensibilización, separadas cuatro semanas, seguidas de dos vacunas de refuerzo 4 y 8 semanas después de la última vacunación de sensibilización. También debe ser fácilmente evidente para un técnico con experiencia en el área que existen varias permutaciones y combinaciones que se engloban usando los vectores de expresión de ADN bacterianos y virales de la divulgación, para proporcionar los regímenes de sensibilización y refuerzo.

También se dan a conocer métodos para inducir una respuesta inmunitaria contra el VIH en un sujeto mediante administración de una composición inmunógena de la invención, preferentemente que comprende un vector de adenovirus que contiene uno o más de los antígenos del VIH-1 del clado A de la invención, (preferentemente GRIN, GRN, o Env o una combinación de éstos), una o más veces a un sujeto en el que se expresan los antígenos del VIH-1 del clado A a un nivel suficiente para inducir una respuesta inmunitaria específica en el sujeto. Dichas vacunas se pueden repetir varias veces a intervalos de tiempo de al menos 2, 4 o 6 semanas (o más) conforme a un régimen de vacunación deseado.

Las composiciones inmunógenas de la invención se pueden administrar solas, o se pueden coadministrar o administrar secuencialmente con otros inmunógenos del VIH o composiciones inmunógenas del VIH, por ejemplo, con "otras" composiciones inmunológicas, antigénicas o para vacuna o terapéuticas, mediante las cuales se proporcionan composiciones multivalentes o "cocktail" o una combinación de composiciones, y métodos para emplearlas. Una vez más, los ingredientes y la forma (secuencial o coadministración) de administración, así como las dosis, se pueden determinar teniendo en cuenta factores tales como la edad, el género, el peso, la especie y la afección particular del sujeto, y la vía de administración.

Cuando se utilizan en combinación, los otros inmunógenos del VIH se pueden administrar al mismo tiempo o en momentos diferentes como parte de un régimen de vacunación global, por ejemplo, como parte de un régimen de sensibilización-refuerzo u otro protocolo de vacunación. Se conocen muchos otros inmunógenos del VIH en el área, uno de dichos inmunógenos preferidos es HIVA (descrito en WO 01/47955), que se puede administrar como una proteína, en un plásmido (por ejemplo, pThr.HIVA) o en un vector viral (por ejemplo, MVA.HIVA). Otro de dichos inmunógenos del VIH es RENTA (descrito en PCT/US2004/037699), que también se puede administrar como una proteína, en un plásmido (por ejemplo, pThr.RENTA) o en un vector viral (por ejemplo, MVA.RENTA).

Por ejemplo, un método para inducir una respuesta inmunitaria contra el VIH en un sujeto humano comprende administrar al menos una dosis sensibilizante de un inmunógeno del VIH y al menos una dosis de refuerzo de un inmunógeno del VIH, donde el inmunógeno de cada dosis puede ser el mismo o diferente, siempre que al menos uno de los inmunógenos sea un antígeno del VIH-1 del clado A de la divulgación, un ácido nucleico que codifica un antígeno del VIH-1 del clado A de la divulgación o un vector de expresión, preferentemente un vector de adenovirus, que codifica un antígeno del VIH-1 del clado A, y donde los inmunógenos se administran en una cantidad o se expresan a un nivel suficiente para inducir una respuesta inmunitaria específica del VIH en el sujeto. La respuesta inmunitaria específica del VIH puede incluir una respuesta inmunitaria de linfocitos T específicas del VIH o una respuesta inmunitaria de linfocitos B específicos del VIH. Dichas vacunaciones se pueden realizar a intervalos,

preferentemente de al menos 2 a 6 o más semanas.

Los ejemplos no limitantes siguientes se dan con el fin de ilustrar diversas realizaciones de la divulgación.

5 **Ejemplos**

Ejemplo 1: secuencia consenso para Gag del VIH del clado A

Tabla 1			
	Distancia respecto al consenso	País	Año
A_consensu	0		
A_97TZ02_1	0.04081	TZ	1997
A_TZA173_1	0.0425	TZ	2001
A_KNH1144_	0.04259	KE	2000
A_SE7535UG	0.04303	UG	1994
A_KNH1211_	0.04463	KE	2000
A_KSM4024_	0.04684	KE	2000
A_KNH1207_	0.04701	KE	2000
A_SE6594UG	0.04709	UG	1993
A_92UG037_	0.05079	UG	1992
A_TZA195_1	0.05127	TZ	2001
A_MSA4079_	0.05279	KE	2000
A_TZA341_1	0.05523	TZ	2001
A_MSA4072_	0.05583	KE	2000
A_MSA4076_	0.056	KE	2000
A_KNH1199_	0.05687	KE	2000
A_MSA4070_	0.05947	KE	2000
A_98UG5713	0.06038	UG	1998
A_KEQ23-17	0.06072	KE	1994
A_KNH1209_	0.06101	KE	2000
A_NKU3005_	0.06108	KE	2000
A_SE7253SO	0.06113	SO	1994
A_98UG5713	0.06119	UG	1998
A_SE8538TZ	0.06137	TZ	1995
A_KNH1088_	0.06262	KE	1999
A_KER2008_	0.065	KE	2000
A_99UGA070	0.06531	UG	1999
A_KER2012-	0.06654	KE	2000
A_KER2009_	0.0674	KE	2000
A_99UGG033	0.06871	UG	1999
A_KSM4030-	0.07026	KE	2000
A_KSM4021-	0.07145	KE	1999
A_98UG5713	0.07189	UG	1998
A_SE8891UG	0.07197	UG	1995
A_SE8131UG	0.07462	UG	1995
A_97TZ03_1	0.07653	TZ	1997
A_KNH1135_	0.07687	KE	1999

## ES 2 396 915 T3

	Distancia respecto al consenso	País	Año
A_98UG5714	0.0781	UG	1998
A_UGU455_1	0.08349	UG	1985
A_MSA4069_	0.08867	KE	2000

5 Se analizaron las secuencias de aminoácidos de las proteínas Gag de 39 cepas no recombinantes del VIH del clado A. La tabla 1 enumera las 39 cepas utilizadas y se refiere a cada una por su número de registro de Genbank. La tabla 1 también identifica el país y el año del aislamiento de cada una de estas 39 cepas. 20 de las cepas fueron de Kenia, 12 de Uganda, 6 de Tanzania y 1 de Somalia. 20 de las cepas se aislaron entre 2000 y 2002, 10 se aislaron entre 1997 y 1999, 6 se aislaron entre 1994 y 1996 y 3 se aislaron antes de 1993.

10 Las secuencias de la proteína Gag se alinearon agregando espacios para preservar el alineamiento en regiones con inserciones o deleciones. Se derivó una secuencia consenso de 50%. La secuencia consenso de aminoácidos se muestra en la figura 1. En la figura 1 los espacios que se agregaron para conservar el alineamiento en regiones con inserciones o deleciones están representados por guiones, y las posiciones para las cuales no se logró un consenso de 50% están representadas por una "X".

15 Para cada una de las 39 secuencias utilizadas para generar la secuencia consenso, la "distancia" de esa secuencia respecto a la secuencia consenso se calculó utilizando la matriz de sustitución de Dayhoff PAM250, que pondera las sustituciones según el grado de similitud bioquímica. Como se muestra en la tabla 1, la distancia de la secuencia de cada cepa respecto a la secuencia consenso varió entre 4 y 9%.

20 La figura 2 ilustra en forma gráfica la distancia de la secuencia de aminoácidos de cada cepa respecto a la secuencia consenso de aminoácidos e identifica las cuatro cepas que tienen las secuencias más similares a las secuencias consenso. Esas cuatro cepas son la cepa 97TZ02 de un individuo de bajo riesgo en la región de Mbeya del suroeste de Tanzania en 1997 que tiene el número de registro de Genbank AF361872, la cepa TZA173 obtenida de un donante de sangre anónimo en la región de Mbeya al suroeste de Tanzania en 2001 que tiene el número de registro de Genbank AY253305, la cepa KNH1144 obtenida de un donante de sangre anónimo en Kenia meridional en el año 2000 que tiene el número de registro de Genbank AF4587006, y la cepa SE7535 obtenida en 1994 en Suecia de un individuo que se pensaba que había sido infectado en Uganda que tiene el número de registro de Genbank AF069671.

30 Ejemplo 2: secuencia consenso para Pol del VIH del clado A

35 Se analizaron las secuencias de aminoácidos de las proteínas Pol de 36 cepas no recombinantes del VIH del clado A. La tabla 2 enumera las 36 cepas utilizadas y se refiere a cada una por su número de registro de Genbank. La tabla 2 también identifica el país y el año del aislamiento de cada una de esas 36 cepas. 20 de las cepas fueron de Kenia, 9 de Uganda, 6 de Tanzania y 1 de Somalia. 19 de las cepas se aislaron entre 2000 y 2002, 10 se aislaron entre 1997 y 1999, 4 se aislaron entre 1994 y 1996 y 3 se aislaron antes de 1993.

40 Se alinearon las secuencias de las proteínas Pol. No hubo inserciones ni eliminaciones. Se derivó una secuencia consenso de 50%. La secuencia consenso de aminoácidos se muestra en la figura 3. En la figura 3 las posiciones para las cuales no se logró un consenso de 50% están representadas por una "X". Hubo 4 de dichas posiciones en los 947 residuos de aminoácidos. Para cada una de las 36 secuencias utilizadas para generar la secuencia consenso, la "distancia" de esa secuencia respecto a la secuencia consenso se calculó utilizando la matriz de sustitución de Dayhoff PAM250, que pondera las sustituciones según el grado de similitud bioquímica. Como se muestra en la tabla 2, la distancia de la secuencia de cada cepa respecto a la secuencia consenso varió entre 1.5 y 4.8%.

45

	Distancia respecto al consenso	País	Año
A_pol.cons	0		
A_MSA4070_	0.01479	KE	2000
A_SE7253SO	0.01582	SO	1994
A_SE8538TZ	0.01898	TZ	1995
A_KER2012-	0.02329	KE	2000
A_97TZ02_3	0.0235	TZ	1997
A_KEQ23-17	0.02445	KE	1994
A_KNH1211_	0.02449	KE	2000



## ES 2 396 915 T3

	Distancia respecto al consenso	País	Año
A_TZA341_3	0.0246	TZ	2001
A_KSM4024_	0.02528	KE	2000
A_97TZ03_3	0.02544	TZ	1997
A_KNH1088_	0.02544	KE	1999
A_MSA4076_	0.02564	KE	2000
A_KNH1207_	0.0265	KE	2000
A_NKU3005_	0.02661	KE	2000
A_TZA173_3	0.02756	TZ	2001
A_MSA4079_	0.02762	KE	2000
A_KER2009_	0.02765	KE	2000
A_TZA195_3	0.02881	TZ	2001
A_KSM4021-	0.02881	KE	1999
A_SE7535UG	0.02883	UG	1994
A_MSA4069_	0.02886	KE	2000
A_SE6594UG	0.02889	UG	1993
A_98UG5713	0.02975	UG	1998
A_KNH1135_	0.0299	KE	1999
A_92UG037_	0.02993	UG	1992
A_KNH1209_	0.03202	KE	2000
A_99UGG033	0.03291	UG	1999
A_KER2008_	0.03294	KE	2000
A_KSM4030-	0.0343	KE	2000
A_KNH1199_	0.03439	KE	2000
A_99UGA070	0.03537	UG	1999
A_MSA4072_	0.03625	KE	2000
A_KNH1144_	0.03863	KE	2000
A_98UG5713.	0.04178	UG	1998
A_UGU455_3	0.04294	UG	1985
A_98UG5713	0.04808	UG	1998

La figura 4 ilustra en forma gráfica la distancia de la secuencia de aminoácidos de cada cepa respecto a la secuencia consenso de aminoácidos e identifica las cuatro cepas que tienen las secuencias más similares a las secuencias consenso. Esas tres cepas son la cepa MSA4070 de un donante de sangre anónimo en Kenia meridional en el año 2000, la cepa SE7235SO que se obtuvo en 1994 de un individuo en Suecia que se pensaba que había sido infectado en Somalia, y la cepa SE8538 que se obtuvo en 1995 de un individuo en Suecia que se pensaba que había sido infectado en Tanzania.

5

Ejemplo 3: secuencia consenso para Env del VIH del clado A

10

	Dist respecto a A.cons	país	año
A.cons	0		
A_KEQ23-17	0.06307	KE	1994
A_TZA341_1	0.06413	TZ	2001
A_KNH1088_	0.06524	KE	1999
A_KNH1209_	0.0699	KE	2000
A_KNH1144_	0.07088	KE	2000

## ES 2 396 915 T3

	Dist respecto a A.cons	país	año
A_99UGA070	0.07365	UG	1999
A_MSA4072_	0.07516	KE	2000
A_KSM4021_	0.0778	KE	1999
A_97TZ02_1	0.07825	TZ	1997
A_KNH1199_	0.07883	KE	2000
A_MSA4079_	0.07944	KE	2000
A_SE7535UG	0.08375	UG	1994
A_SE8538TZ	0.08432	TZ	1995
A_98UG5713	0.08462	UG	1998
A_97TZ03_1	0.08541	TZ	1997
A_MSA4070_	0.0874	KE	2000
A_NKU3005_	0.0884	KE	2000
A_TZA173_1	0.09046	TZ	2001
A_KNH1207_	0.09106	KE	2000
A_TZA195_1	0.09389	TZ	2001
A_MSA4076_	0.09517	KE	2000
A_92UG037_	0.098	UG	1992
A_98UG5714	0.09816	UG	1998
A_SE7253SO	0.09886	SO	1994
A_KER2012-	0.09984	KE	2000
A_98UG5713	0.10139	UG	1998
A_SE6594UG	0.10195	UG	1993
A_SE8891UG	0.10225	UG	1995
A_UGU455_1	0.10314	UG	1985
A_KER2009_	0.10338	KE	2000
A_KNH1211_	0.11319	KE	2000
A_SE8131UG	0.11321	UG	1995
A_MSA4069_	0.11507	KE	2000
A_99UGG033	0.11653;	UG	1999
A_KNH1135	0.117131	KE	1999
A_KER2008	0.12689	KE	2000

5 Se analizaron las secuencias de aminoácidos de las proteínas Env de 36 cepas no recombinantes del VIH del clado A. La tabla 3 enumera las 36 cepas utilizadas y se refiere a cada una por su número de registro de Genbank. La tabla 3 también identifica el país y el año del aislamiento de cada una de esas 36 cepas. 18 de las cepas fueron de Kenia, 11 de Uganda, 6 de Tanzania y 1 de Somalia. 17 de las cepas se aislaron entre 2000 y 2002, 10 se aislaron entre 1997 y 1999, 6 se aislaron entre 1994 y 1996 y 3 se aislaron antes de 1993.

10 Las secuencias de las proteínas Env se alinearon agregando espacios para preservar el alineamiento en regiones con inserciones o deleciones. Hubo muchas regiones con una amplia heterogeneidad en la longitud de las inserciones/deleciones. Se derivó una secuencia consenso de 50%. La secuencia consenso de aminoácidos se muestra en la figura 5. En la figura 5 los espacios que se agregaron para conservar el alineamiento en regiones con inserciones o deleciones están representados por guiones, y las posiciones para las cuales no se logró un consenso de 50% están representadas por una "X". Hubo muchas posiciones de aminoácidos para las cuales no se alcanzó un consenso de 50%.

15 Para cada una de las 36 secuencias utilizadas para generar la secuencia consenso, la "distancia" de esa secuencia respecto a la secuencia consenso se calculó utilizando la matriz de sustitución de Dayhoff PAM250, que pondera las sustituciones según el grado de similitud bioquímica. Como se muestra en la tabla 3, la distancia de la secuencia de

## ES 2 396 915 T3

cada cepa respecto a la secuencia consenso varió entre 6.3 y 12.7%.

La figura 6 ilustra en forma gráfica la distancia de la secuencia de aminoácidos de cada cepa respecto a la secuencia consenso de aminoácidos e identifica las cuatro cepas que tienen las secuencias más similares a las secuencias consenso. Esas tres cepas fueron KEQ23 de una trabajadora social diplomada en Kenia en 1994, TZA341 que fue de un donante de sangre anónimo en Tanzania en 2002 y KNH1088 que fue de un donante de sangre anónimo en Kenia en 1999.

Ejemplo 4: secuencia consenso para Nef del VIH del clado A

Se analizaron las secuencias de aminoácidos de las proteínas Nef de 38 cepas no recombinantes del VIH del clado A. La tabla 4 enumera las 38 cepas utilizadas y se refiere a cada una por su número de registro de Genbank. El país y el año de aislamiento de cada una de esas 38 cepas se describe en las tablas 1 a 3 en los ejemplos anteriores. Más de la mitad de las cepas fueron de Kenia, con una porción sustancial proveniente de Uganda y unas pocas cepas de Tanzania. Aproximadamente la mitad de la cepas se aislaron entre 2000 y 2002.

Tabla 4

	A.cons
A_MSA4070	0.0318
A_KNH1211	0.04807
A_97TZ03_1	0.0535
A_99UGA070	0.05354
A_SE8891UG	0.05383
A_KEQ23-17	0.06476
A_98UG5713	0.07043
A_NKU3005_	0.0709
A_SE7535UG	0.07117
A_98UG5714	0.07613
A_SE6594UG	0.07634
A_TZA341_1	0.0805
A_MSA4069	0.08097
A_KNH1199	0.08213
A_97TZ02_1	0.08276
A_KSM4030-	0.08704
A_KSM4021-	0.08795
A_MSA4076	0.08873
A_KNH1209	0.0899
A_KER2012-	0.09224
A_KNH1144	0.09577
A_KER2008	0.09703
A_MSA4072	0.09892
A_98UG5713	0.09892
A_99UGG033	0.09967
A_KNH1088	0.10303
A_92UG037	0.10654
A_SE8538TZ	0.10996
A_KER2009	0.1102
A_MSA4079	0.11083

	A.cons
A_KSM4024	0.11126
A_SE8131UG	0.11326
A_SE7253SO	0.11453
A_KNH1207	0.11549
A_TZA173_1	0.13766
A_98UG5713	0.1399
A_UGU455_1	0.15688
A_KNH1135	0.16076
A.cons	0

5 Las secuencias de las proteínas Nef se alinearon con espacios para preservar el alineamiento en regiones con inserciones o deleciones. Se derivó una secuencia consenso de 50%. La secuencia consenso de aminoácidos se muestra en la figura 7. En la figura 7 los espacios que se agregaron para conservar el alineamiento en regiones con inserciones o deleciones están representados por guiones, y las posiciones para las cuales no se logró un consenso de 50% están representadas por una "X". Hubo 6 posiciones de aminoácidos para las cuales no se alcanzó un consenso de 50%.

10 Para cada una de las 38 secuencias utilizadas para generar la secuencia consenso la "distancia" de esa secuencia respecto a la secuencia consenso se calculó utilizando la matriz de sustitución de Dayhoff PAM250, que pondera las sustituciones según el grado de similitud bioquímica. Como se muestra en la tabla 4, la distancia de la secuencia de cada cepa respecto a la secuencia consenso varió entre 3.2 y 16.1% con una distancia media de 9.3%.

15 La figura 8 ilustra en forma gráfica la distancia de la secuencia de aminoácidos de cada cepa respecto a la secuencia consenso de aminoácidos e identifica las cinco cepas que tienen las secuencias más similares a las secuencias consenso. Esas cinco cepas fueron MSA4070 y KNH1211, las cuales fueron de donantes anónimos en Kenia meridional y se obtuvieron en el año 2000, 97TZ03 de un individuo de bajo riesgo en la región de Mbeya del suroeste de Tanzania se obtuvo en 1997, y UGA070 y SE8891, ambas eran de individuos en Uganda y se obtuvieron en 1999 y 1995, respectivamente.

20 Ejemplo 5. cepas del VIH del clado A que son las más similares a las secuencias consenso del VIH del clado A.

25 Como se describe en los ejemplos 1 a 4 antes y como se resume en la tabla 5, se identificaron las cepas del VIH del clado A que tienen las secuencias Gag, Pol, Env y Nef más similares a las secuencias consenso de cada una de esas proteínas. Además, se identificaron las cepas que eran más similares globalmente a la secuencia consenso puntuando a cada una de las cepas según su similitud con la secuencia consenso de una proteína particular donde la cepa puntuada con el número 1 fue aquella cuya secuencia para esa proteína era la más similar a la de la secuencia consenso y después sumando las puntuaciones para cada cepa para las 4 proteínas (es decir Gag, Pol, Env y Nef). Las seis cepas que fueron más similares globalmente a la secuencia consenso en las cuatro proteínas estudiadas se indican a continuación en la tabla 6. Se puede observar que la cepa 97TZ02 tiene una secuencia que es globalmente más similar a las secuencias consenso de cada una de los genes de Gag, Pol, Env y Nef.

Tabla 5

Gag	Pol	Env	Nef
97TZ02	MSA4070	KEQ23	MSA4070
TZA173	SE7245SO	TZA341	KNH1211
KNH1144	SE8538	KNH1088	97TZ03
SE7535UG			99UGA070
			SE8891UG

Tabla 6

	gag	pol	env	nef	sum
A_97TZ02_1	1	5	9	15	30
A_KEQ23-17	18	6	1	6	31

35

	gag	pol	env	nef	sum
A_MSA4070_	16	1	16	1	34
A_TZA341_1	12	8	2	12	34
A_SE7535UG	4	20	12	9	45
A_KNH1211_	5	7	31	2	45

Ejemplo 6. Construcción de los transgenes GRIN, GRN y Env.

5 Se hicieron constructos de transgenes con secuencias de proteínas del VIH del clado A derivadas de los aislados de campo circulantes del VIH-1 identificados recientemente que se correspondían mejor con la secuencia consenso del VIH del clado A para cada una de esas proteínas. Esta estrategia se desarrolló para maximizar la importancia biológica de las secuencias del VIH del clado A utilizadas. Se debe comprender que también se pueden usar otras secuencias, es decir secuencias distintas de las secuencias específicas descritas en este ejemplo, de conformidad con esta divulgación. Si se usan otras secuencias es preferible que las secuencias se seleccionen de modo que  
10 deriven de aislados de campo recientes y que tengan secuencias que sean similares a las secuencias consenso del VIH del clado A descritas en este documento, o a las secuencias consenso del VIH del clado A que se puedan generar en el futuro.

15 Se hicieron constructos denominados GRIN y GRN. El constructo GRIN contenía secuencias del VIH del clado A que codifican las proteínas Gag, Pol (RT e integrasa) y Nef. El transgén GRN contenía secuencias que codifican las proteínas Gag, RT y Nef. Los constructos GRIN y GRN se representan esquemáticamente en la figura 9. Los transgenes GRIN y GRN se prepararon usando la secuencia de la proteína Gag de la cepa TZA173 que tiene el número de registro de Genbank AY253305, la secuencia de Pol (que comprende las secuencias RT e Int) de la cepa MSA4070 que tiene el número de registro de Genbank AF457081 y la secuencia de Nef de la cepa MSA4070 que tiene el número de registro de Genbank AF457081. Esas secuencias se seleccionaron porque fueron de los aislados de campos circulantes del VIH-1 más recientemente identificados que tenían la mayor correspondencia con la secuencia consenso para cada una de Gag, Pol y Nef, respectivamente. Esas secuencias se ilustran en las figuras  
20 10, 11 y 12, respectivamente.

25 También se preparó un constructo de Env con la secuencia de codificación de Env del aislado de campo circulante del VIH-1 más recientemente identificado que tenía la mayor correspondencia con la secuencia consenso de Env, es decir, la secuencia de la proteína Env de la cepa TZA341 que tiene el número de registro de Genbank AY253314. Esta secuencia se ilustra en la figura 13.

30 A todas las secuencias se les hizo optimización de codones para expresión en humanos en GeneArt (Alemania). Las secuencias con los genes optimizados permiten un alto nivel de expresión y la expresión estable de las proteínas en humanos u otras células de mamíferos. Otros detalles del proceso de optimización de codones se proporcionan en el ejemplo 8. También se incorporaron por ingeniería genética a los transgenes mutaciones específicas o dispuestas en un orden específico para anular la función normal de los productos génicos in vivo. Los detalles de esas mutaciones, y los efectos biológicos de cada una, se describen en el ejemplo 7 más adelante.  
35

40 Para los transgenes GRIN y GRN, las secuencias de codificación para cada una de las proteínas Gag, Pol (RT e Int) o RT, y Nef se unieron en el marco para que cada uno de los constructos del transgén (es decir de GRIN o de GRN) codifica una única proteína de fusión. Se realizaron búsquedas Blast para asegurarse de que no se formaran neopítopos en las uniones. Aunque no se utiliza en este ejemplo, se debe señalar que también es posible insertar secuencias espaciadoras entre las secuencias de codificación para los componentes individuales de la proteína de fusión final para permitir el plegado óptimo del dominio de la proteína, por ejemplo, se puede agregar una región espaciadora entre Gag y Pol para permitir que los dominios de la proteína se plieguen en una conformación más natural. Asimismo, se agregaron sitios de restricción únicos en los extremos 5' y 3' de cada secuencia con el fin de  
45 facilitar la unión de cada secuencia (por ejemplo, la unión del extremo 5' de Nef con el extremo 3' de Pol, etc.).

50 Para usar in vivo se insertaron los transgenes GRIN, GRN y Env en los vectores de adenovirus Ad5, Ad35, Ad11, C6 o C7. Para facilitar la clonación se agregaron sitios de restricción únicos en esos vectores en los extremos 5' y 3' de los constructos GRIN, GRN o Env. Las figuras 14A-14 C proporcionan la secuencia de GRIN insertada en el vector Ad35 y muestran los sitios de restricción utilizados para clonar la secuencia de GRIN en el vector Ad35 (subrayados y en negrita). La secuencia también incluye la secuencia promotora CMV en dirección 5' de la secuencia de GRIN. Las figuras 15A-15B proporcionan la secuencia de Env insertada en el vector Ad35 y muestran los sitios de restricción utilizados para clonar la secuencia de Env en el vector Ad35 (subrayados y en negrita). La secuencia también incluye la secuencia promotora CMV en dirección 5' de la secuencia de Env.  
55

Se utilizaron técnicas estándar de clonación y recombinación del ADN para generar todos los constructos anteriores.

Dichas técnicas son bien conocidas por los expertos. Véase por ejemplo, "Molecular Cloning: A Laboratory Manual", segunda edición (Sambrook et al., 1989).

Ejemplo 7: Mutaciones para anular la función normal in vivo de las proteínas del VIH del clado A

La tabla 7 resume las mutaciones realizadas por ingeniería genética en las secuencias de GRIN y GRN para anular la función in vivo de sus productos génicos. Esas mutaciones se realizaron usando técnicas estándar de recombinación del ADN. Dichas técnicas son bien conocidas por los expertos.

Tabla 7

Diseño	Gen	Mutación/justificación
Mutación	gag	Gly2 → Ala : elimina el sitio de miristilación evitando la formación de VLP.
Mutación	gag	<i>Para evitar el deslizamiento en la secuencia de cambio de marco natural, la secuencia de ADN se mutó de una manera que deja intacta la secuencia de aminoácidos conservada (NFLG) y permite que sólo se traduzca el producto de la proteína GagPol entera.</i>
Mutación	RT	Asp185 → Ala & Asp186 → Ala: inactiva los residuos enzimáticos activos.
Mutaciones	Integrasa (IN)	Asp 64 → Ala, Asp116 → Ala & Glu 152 → Ala: inactiva los residuos enzimáticos activos.
Sin cambio	Nef	La fusión del extremo N-terminal de nef con el extremo C-terminal de IN evita la miristilación y la acción selectiva sobre la membrana anulando la función de Nef.

La proteína Gag se expresa como una poliproteína precursora de 55 kDa (Pr55<sup>gag</sup>) y es escindida por la proteasa viral del VIH-1. Cuatro proteínas virales principales resultan de la escisión; Matriz (MA), cápside (CA), nucleocápside (NC) y p6; así como dos polipéptidos espaciadores p2 y p1, que representan secuencias entre CA y NC y entre NC y p6, respectivamente.

MA desempeña un papel clave en varios pasos de la replicación del virus, que incluyen la mediación crítica del ensamblado de la partícula viral y el brote de la membrana celular a través de la formación de partículas similares a virus (VLP) (véase Gheysen, D., E. Jacobs, F. de Foresta, D. Thiriart, M. Francotte, D. Thines, y M. De Wilde. (1989). Assembly and release of HIV-1 precursor pr55gag virus-like particles from recombinant baculovirus-infected cells. Cell 59:103-112).

Tanto Pr55gag como MA (p17) se miristilan, es decir, forman enlace amida con el ácido mirístico. Véase Veronese di Marzo, F., Copeland, T. D., Oroszlan, S., Gallo, R. C. y Sarngadharan, M. G. (1988). J. Virol. 62, 795-801. Véase también la sección sobre Nef en el ejemplo 7 por una descripción completa del proceso de miristilación. Diferentes aislados del VIH-1 demuestran que el aceptor de miristilo es el residuo de glicina N-terminal (Gly2). Véase Bryant y Ratner. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc. Natl. Acad. Sci. EE.UU; 87: 523-527.

Bryant y Ratner (1990) demostraron que la sustitución de Gly2 con Ala eliminó la replicación del virus de un clon del HIV-1. El Pr55gag, que carece de la glicina aceptor de miristilo, se acumuló en células Hela infectadas y no se procesó en la cápside del virión maduro. Se concluyó que la miristilación de la Gly2 es necesaria para la asociación estable de la membrana plasmática y el ensamblado posterior de los viriones. Otros grupos demostraron de manera semejante la importancia de la miristilación de Gly2 en la MA. Véase Göttlinger H G, Sodroski J G, Haseltine W A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA; 86:5781-5785, y Paul Spearman, Jaang-Jiun Wang, Nancy Vander Heyden y Lee Ratner. (1994). Identification of Human Immunodeficiency Virus Type 1 Gag Protein Domains Essential to Membrane Binding and Particle Assembly. J. Virol; 68 (5): 3232-3242.

Si se muta la glicina N-terminal aceptor de miristilo (Gly2) en MA, se anula la unión a la membrana y se evita el ensamblado de partículas. Por consiguiente, Gag del clado A se somete a ingeniería para cambiar Gly2 → Ala. Esto provoca la pérdida de la función biológica de Gag.

La transcriptasa inversa (RT) es una enzima viral esencial para la replicación. RT convierte el ARN+ viral entrante en ADNbc, catalizada por las actividades de polimerasa dependiente de ARN y ADN, y de ARNasa H de la enzima. RT es un heterodímero compuesto por las proteínas subunidades p66 y p51. Véase Alfredo Jacobo-Molina et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl. Acad. Sci. USA; 90: 6320-6324. p66 tiene 2 dominios, la polimerasa y la ARNasa H. p51 tiene el mismo dominio polimerasa.

Los residuos Asp-110, Asp-185 y Asp-186 catalíticamente esenciales se encuentran ubicados en el sitio activo

altamente conservado de la ADN polimerasa. Se cree que esos tres residuos denominados "la tríada catalítica" se unen a cationes divalentes necesarios para la función de catálisis. Véase Alfredo Jacobo-Molina et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA*; 90: 6320-6324.

Se demostró que la mutación de los ácidos aspárticos en los residuos 185 y 186 en la asparragina o el glutamato da lugar a proteínas mutantes que fueron catalíticamente inactivas. Véase Lowe DM, Parmar V, Kemp SD, Larder BA. (1991). Mutational analysis of two conserved sequence motifs in HIV-1 reverse transcriptase. *FEBS Lett.*; 6;282(2):231-4.

La mutación de Asp 185 → Ala y Asp 186 → Ala en la RT del clado A inactivará a la enzima polimerasa de RT interrumpiendo la "tríada catalítica". Esto eliminará la función biológica de la RT del clado A.

El ADNc proviral generado por RT se integra en el genoma de la célula huésped a través de la acción de la enzima viral integrasa (Int). Int contiene un dominio de recombinasa de ADN que cataliza dos reacciones endonucleolíticas distintas. La primera reacción, el procesamiento de 3', elimina dinucleótidos de cada extremo de los ADNc produciendo dos extensiones 5' de dos nucleótidos en ambos extremos. En la segunda reacción, Int escinde no específicamente el ADN de la célula huésped y une los grupos libres del extremo 3' del ADNc a los grupos del 5' del ADN escindido de la célula huésped. Las enzimas celulares reparan los huecos dando lugar a un genoma viral totalmente integrado en el ADN de la célula huésped. Véase Coffin JM. *Retroviridae and their Replication*. Capítulo 27. pp. 645-708 & Wong-Staal F. *Human Immunodeficiency Viruses and Their Replication*. Capítulo 28. pp. 709-723. In Fields, BN. & Knipe DM. 2ª edición *Fundamental Virology*. Raven Press. Véase también Engelman A, Mizuuchi K, Craigie R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*; 67(6):1211-1221. Los residuos 50 a 212 del dominio catalítico, contienen una tríada de residuos Asp-64, Asp-116 y Glu-152 (denomina motivo D,D-35-E) que compromete el sitio activo de la enzima. Véase Esposito, D., y R. Craigie. (1999). HIV integrase structure and function. *Adv. Virus Res.* 52:319-333. Véase también Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, y A. M. Skalka. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 19:851-860.

A través de diversas técnicas, algunos grupos han demostrado la anulación de la función de endonucleasa y/o integración de IN a través de la mutación dirigida al sitio de los residuos Asp-64, Asp-116 y Glu-152 en el motivo D, D-35-E. Véase Drelich M, Wilhelm R, Mous J. (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology*;188(2):459-468. Véase también LaFemina RL, Schneider CL, Robbins HL, Callahan PL, LeGrow K, Roth E, Schleif WA, Emini EA. (1992). Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *J Virol*; 66(12):7414-7419. Véase también Leavitt AD, Shiue L, Varmus HE. (1993). Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J Biol Chem*; 268(3):2113-2119.

La mutación de Asp-64→ Ala, Asp-116→ Ala y Glu-152→ Ala en la Int del clado A inactivará a la enzima activa Int interrumpiendo el motivo crítico D,D-35-E. Esto eliminará la función biológica de Int del clado A .

La proteína de factor negativo (Nef) (27-kDa) es la primera proteína viral que se acumula en la célula recién infectada. Véase Haseltine, W. (1991). *Molecular biology of the human immunodeficiency virus type 1*. FASEB. Vol 5. 2349-2360. A través de la miristilación, Nef es capaz de ubicarse del lado del citosol de la membrana celular. Véase Yu G, Felsted RL. (1992). Effect of myristoylation on p27 nef subcellular distribution and suppression of HIV-LTR transcription. *Virology*. 187(1):46-55. Véase también Kaminchik, J., N. Bashan, A. Itach, N. Sarver, M. Gorecki, y A. Panet. (1991). Genetic characterization of human immunodeficiency virus type 1 nef gene products translated in vitro and expressed in mammalian cells. *J. Virol.* 65:583-588. La miristilación de las proteínas es un fenómeno co-traduccional e implica la transferencia de miristato desde la miristil-coenzima A al motivo amino-terminal MGXXX de las proteínas por la enzima N-miristil transferasa (NMT). Véase Towler, D. A., S. P. Adams, S. R. Eubanks, D. S. Towery, E. Jackson-Machelski, L. Glaser y J. I. Gordon (1987). Purification and characterization of yeast myristoyl CoA:protein N- myristoyltransferase. *Proc Natl Acad Sci USA* 84:2708-2712. La metionina líder del polipéptido es escindida por la metionina amino peptidasa durante la traducción y reconoce a NMT el grupo amino terminal de la glicina recién generado del péptido emergente después de que aproximadamente veinte residuos se liberan del ribosoma. NMT transfiere miristato al residuo de glicina (el aceptor de miristilo) y se completa la miristilación. El reemplazo de la penúltima glicina aceptor de miristilo por otro residuo de aminoácido inhibe la miristilación. Véase Towler, D. A., S. R. Eubanks, D. S. Towery, S. P. Adams & L. Glaser (1987). Amino-terminal processing of proteins by N-myristoylation. Substrate specificity of N-myristoyl transferase. *J Biol Chem* 262:1030-1036.

Nef es una proteína multifuncional capaz de modular una serie de moléculas superficiales de la célula infectada, como CD4 (véase Garcia, J. V. y A. D. Miller. (1991). Serine phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* 350:508-511; y Mariani R y Skowronski J. (1993). CD4 down-regulation by nef alleles isolated from human immunodeficiency virus type 1-infected individuals *Proc. Natl. Acad. Sci. USA*. Vol. 90, pp. 5549-5553; y Aiken C, Konner J, Landau NR, Lenburg ME, Trono D (1994). Nef induces CD4 endocytosis:

requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell.* 11;76(5):853-64), CD28 (véase Swigut, T., N. Shohdy y J. Skowronski. (2001). Mechanism for down-regulation of CD28 by Nef. *EMBO J.* 20:1593-1604), MHC-I (véase Schwartz, O., V. Marechal, S. Le Gall, F. Lemonnier y J. M. Heard. (1996). Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat. Med.* 2:338-342), the macrophage-expressed MHC 1b protein HFE (véase Drakesmith H, Chen N, Ledermann H, Sreaton G, Townsend A, Xu XN. (2005). HIV-1 Nef down-regulates the hemochromatosis protein HFE, manipulating cellular iron homeostasis. *Proc Natl Acad Sci USA.* 102(31):11017-22), MHC-II (véase Stumptner-Cuvelette, P., S. Morchoisne, M. Dugast, S. Le Gall, G. Raposo, O. Schwartz y P. Benaroch. (2001). Nef del HIV-1 daña la presentación del antígeno de MHC clase II y la expresión superficial. *Proc. Natl. Acad. Sci. USA* 98:12144-12149), así como interrumpe las vías de transducción de la señal (véase Tolstrup, M., L. Ostergaard, A. L. Laursen, S. F. Pedersen y M. Duch. (2004). VIH/VIS escapan de la vigilancia inmunitaria enfocada en Nef. *Curr. HIV Res.* 2:141-151) a través de la asociación con múltiples cinasas y otras proteínas de la superficie celular de la membrana celular. Los mecanismos de esas acciones y los motivos de nef involucrados aún permanecen sin elucidar completamente.

Específicamente, un mutante de Nef con delección de los 19 aminoácidos N-terminales, incluida la señal de miristilación N-terminal eliminó CD4 y la regulación por disminución de MHC-1, manteniendo simultáneamente la mayor parte de los épitopos de CTL, linfocitos T cooperadores y linfocitos B (véase Peng B, Robert Guroff M (2001). Deletion of N-terminal myristoylation site of HIV Nef abrogates both MHC-1 and CD4 down-regulation. *Immunol Lett.* 78(3):195-200). Otros grupos demostraron que la mutación de la glicina amino terminal de Nef (Gly2) en alanina evita la miristilación (véase Liang, x. et al (2002). Development of HIV-1 Nef vaccine components: immunogenicity study of Nef mutants lacking myristylation and dileucine motif in mice. *Vaccine* 20: 3413-3421, y Kaminchik, J. et al. (1991). Genetic Characterization of Human Immunodeficiency Virus Type 1 nef Gene Products Translated in vitro and Expressed in Mammalian Cells. *J. of Virol.* 65(2): 583-588).

Puesto que el motivo amino terminal MGXXX de Nef del clado A está incluido en la proteína de fusión GRIN, no hay ninguna metionina incipiente para ser escindida por la metionina amino peptidasa durante la traducción. Por consiguiente, no se produce ningún grupo amino-terminal recién generado de glicina y NMT es incapaz de ejecutar la miristilación. En conclusión, la incapacidad de Nef en GRIN para experimentar miristilación anula la función biológica de Nef.

#### Ejemplo 8: Optimización de codones para GRIN (GagPolNef) y Env

El uso de codones para cada uno de GRIN y Env se adaptó al sesgo de codones de los genes humanos. La secuencia de nucleótidos y aminoácidos de la secuencia GRIN con codones optimizados se proporciona en las figuras 16A-16J. La secuencia de nucleótidos y aminoácidos de la secuencia de Env con codones optimizados se proporciona en las Figuras 17A-17 D.

Se evitaron las regiones con muy alto (más del 80%) o muy bajo (menos de 30%) contenido de GC cuando fue posible. Durante el proceso de optimización se evitaron los motivos que actúan en cis siguientes: cajas internas TATA, sitios chi, sitios de entrada ribosómicos, tramos de secuencia ricos en AT o en GC, elementos de secuencia ARE, INS o CRS, secuencias repetidas, estructuras secundarias de ARN, de sitios dadores y aceptores de empalme críptico, puntos de ramificación y sitios de restricción HindIII, NcoI, BglII y BclI excepto como se indica en las secuencias provistas en las figuras 16 y 17. También, se introdujo una secuencia Kozak en dirección al extremo del ATG de inicio para cada una de GRIN y Env para aumentar la iniciación de la traducción y se agregaron dos codones de parada a cada uno de GRIN y Env para asegurar una terminación eficaz. También se agregaron sitios de restricción para facilitar la subclonación, como se indica en las figuras 16 y 17.

#### Ejemplo 9: Estudio de primates no humanos

Se realizó un estudio en primates no humanos (macacos rhesus chinos) con el objetivo principal de evaluar la inmunogenia de GRIN y ENV en un sistema de administración de un vector de adenovirus humano tipo 35 (Ad35). Los animales recibieron una dosis cada vez mayor de Ad35-GRIN/ENV ( $10^9$ ,  $10^{10}$  y  $10^{11}$  partículas de virus [vp]; vía intramuscular) y recibieron dos vacunas en el mes 0 y el mes 6 (con 8 animales por grupo para la primera vacuna y 4 animales por grupo para la segunda vacuna). En diversos momentos (desde la semana 0 hasta la semana 50), los animales se sangraron y se midió la inmunogenia por IFN-gamma ELISpot (véanse las figuras 23A y 23B para las dosis de vp de  $10^{10}$  y  $10^{11}$ , respectivamente).

Se observó una respuesta a la dosis (no se muestran los datos para  $10^9$  vp), por ELISPot, tanto en intensidad como en frecuencia de animales que responden tras la primera vacunación (no se muestran los datos). Se observaron respuestas a todos los componentes antigénicos de la vacuna de GRIN/ENV y las respuestas de IFN $\gamma$  ELISPOT se reforzaron después de la segunda vacunación en el mes 6.

La divulgación se describe más detalladamente mediante los párrafos numerados siguientes:



1. Una secuencia consenso de nucleótidos para antígenos del VIH-1 del clado A, donde la secuencia comprende secuencias de nucleótidos que codifican a Gag, Pol (RT e Int) y Nef ("GRIN) del VIH-1 del clado A, Gag, RT y Nef ("GRN") del VIH-1 del clado A, o Env del VIH-1 del clado A.
- 5 2. Una secuencia consenso de nucleótidos de acuerdo con el párrafo 1, donde la proteína codificada Gag tiene la secuencia de aminoácidos de la figura 1.
3. Una secuencia consenso de nucleótidos de acuerdo con el párrafo 1, donde la proteína codificada Pol tiene la secuencia de aminoácidos de la figura 3.
4. Una secuencia consenso de nucleótidos de acuerdo con el párrafo 1, donde la proteína codificada Env tiene la secuencia de aminoácidos de la figura 5.
- 10 5. Una secuencia consenso de nucleótidos de acuerdo con el párrafo 1, donde la proteína codificada Nef tiene la secuencia de aminoácidos de la figura 7.
6. Un método para identificar un antígeno del VIH-1 del clado A de una cepa o un aislado de campo circulante del VIH-1 que tiene una secuencia de aminoácidos que es similar a la secuencia consenso de aminoácidos para ese antígeno del VIH-1 del clado A, que comprende comparar las secuencias de aminoácidos de los antígenos de cepas o aislados de campo circulantes del VIH-1 con la secuencia consenso de aminoácidos para esa proteína y seleccionar un antígeno de cepas o aislados de campo circulantes del VIH-1 que tengan una distancia de proteínas pequeña respecto a la secuencia consenso.
- 15 7. Un antígeno del VIH-1 del clado A identificado mediante el método del párrafo 6.
8. Un método para producir un antígeno transgénico del VIH-1 del clado A que comprende seleccionar un antígeno del VIH-1 del clado A utilizando el método del párrafo 6 y mutando la secuencia de nucleótidos que codifica el antígeno donde la mutación anula la función de ese antígeno.
- 20 9. Un método para generar una respuesta inmunitaria contra el VIH-1 que comprende administrar a un sujeto una composición que comprende una secuencia de nucleótidos o antígeno de acuerdo con cualquiera de los párrafos anteriores.

25

**REIVINDICACIONES**

- 5 1. Un polinucleótido que codifica una proteína de fusión del VIH-1 del clado A, donde la proteína de fusión comprende Gag, Pol y Ne del VIH-1 del clado A y donde la proteína de fusión contiene la secuencia de aminoácidos de la figura 16.
2. Una composición inmunógena que comprende el polinucleótido de la reivindicación 1.
- 10 3. El polinucleótido de la reivindicación 1 para usar en un método para generar una respuesta inmunitaria contra el VIH-1.

Consensus Gag  
 MGARASVLSGGKLDAWEKIRLRPGGKKYRLKHLVWASRELERFALNPSLLETAEGCCQIM  
 EQLQPALKTGTEELRSLFNTVATLYCVHQRIDVKDTKEALDKIEEIQNKSKQK---TQQ--  
 AAADTGXSSKVS----  
 QNYPIVQNAQQQMIHQXLSRPTLNAWVKVIEEKAFSPEVPMFSALSEGATPQDLNMMMLNIVG  
 GHQAAMQMLKDTINEEAAEWDRLHPVHAGPIPPGQMPREPRGSDIAGTTSTPQEQGAVMTG  
 NPPIPVGDYIKRWIILGLNKIVRMYSVILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGW  
 MTETLLVQANANPDCKSILRALGXGATLEEMMTACQGVGGPGGHKARVLAEAMSQVQQTN--  
 IMM-QRGNFRGQKR-  
 IKCFNCGKEGHLARNCRAPRKKGCWCKGEGHQMKDCCTERQANFLGKIWPSSKGRPGNFP  
 QSRPEPTAPPAEI-FGMGEEIASPPKQEQK--DREQXXPPLVSLKSLFGNDPLSQ

FIG. 1



Consensus Pol  
 PQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVKQYD  
 QILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIETVPVKKLPGMDGPKV  
 KQWPLTEEKIKALTEICTEMEKESKISKIGPENPYNTPFAIKKKDSTKWRKLVDFRELN  
 KRTQDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDESFRKYTAFITPSTNNEITPG  
 IRYQYNVLPQGWKGSFAIFQSSMTKILEPFRSKNPEIYYQYMDLTVGSDLEIGQHRTK  
 IEELRAHLLSWGFTTPDKKHQKEPFLWMGYELHPDKWTVQPIXLPEKESWTVNDIQKLV  
 GKLNWASQIYAGIKVKQLCKLLRGAKALTDIVTLTEEALELELAENREILKDPVHGVYYDP  
 SKDLIAEIQKQGQDQWTYQIYQEPFKNLKTGKYARKRSAHNTNDVKQLAEVVKVVMESIV  
 IWGKTPKFKLPIQKETWETWMDYWQATWIPEWFEFVNTPLVKLWYQLEKDPXGAETFY  
 VDGAANRETKLGKAGYVTDGRQKVVSLTETTNQKTELHAXLALQDSGSEVNIVTDSQY  
 ALGHIQAQPDSESELVNQIEKLGKDKVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLV  
 LDGIDKAQEEHERYHSNWRXMASDFNLPPIVAKEIVASCDKCOLKGEAMHGQVDCSPGIW  
 QLDCTHLEGKVLVAHVASGYIEAEVIPAETGQETAYFLKLAGRWPVKVVHTDNGSNF  
 TSAAFKAACWWANIQQEFGIPYNPQSQGVVESMKNELKKIIGQVREQAEHLKTAVQMAVF  
 IHNFKRKGIGGYSAGERIIDIIATDIQTKEKQKITKIQNFVYYRDRSRDPIWKGPAKL  
 LWKGEVVIQDNSDIKVVPRRKAIRDYGKQMGADDCVAGRQDED

FIG. 3

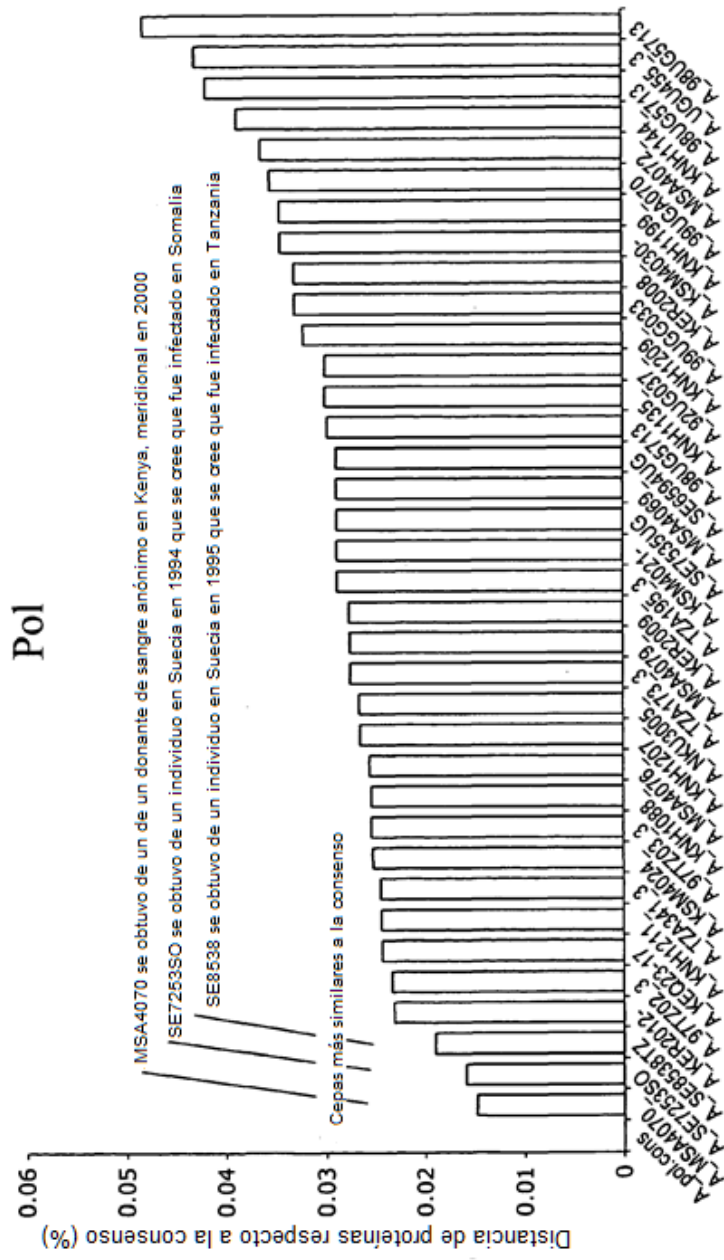


FIG. 4

Consensus Env  
 MRVMGIQRNCQHLLRWG-TMILGMIHCS--XAENLWVTVYYGVVPVWKDAETTLFCASDA  
 KAYXTEHNVWATHACVPTDPNPQEIXLXNVTEEFNMWKNDMVEQMHTDIISLWDQSLKP  
 CVKLTPLCVTLXCXXXXXXXXXXXXXXXXXXXXXXXXX--IKNCSEFNMTTLELRDKKQ  
 KVSLSFYRLDVVQI-----XXXXXXXXXXSYRLINCNTSAITQACPKVSFEPIPIHY  
 CAPAGFALIKCXDXEFNGTGPCKNVSTVQCTHGKIPVSTQLLLNGSLAEXXVX-IRSEN  
 ITNNAKXIIVQLXXPVXINCTRPNNNTRK---SIRIGPGQAKYATGDIIGDIRQAHCNVS  
 RxxWNXTLQXVAXQLRXXXFXNKTIIFXXSSGGDLEITTHSFNCGGEFFYCNTSGLFNST  
 W-----XXXXXXXXXXSNDITILXCRIKQIVNMWQRXGQAMYAPPIQGVIRCES  
 NITGLLITRDGGXXXXXXXXXNETFRPGGGDMRDNWRSELYKYKVKIEPLGVAPTRAKRR  
 VVEREKRAV-GIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSLLRAIEAQQ  
 HLLKLTVWGIKQLQARVLAVERYLRDQQLLGIWGCSGKLICTTNVPWNSSWSNKSXXEIIW  
 DNMTWLQWDKEISNYTQIIYXLIIEESQNQEKNEQDLLALDKWANLWNWFDISNWL YWIK  
 IFIMIVGGGLIGLRIVFAVLSIINRVRQGYSPLSFQTHTPNPRGLDRPGRIIEEGEGEQGRD  
 RSIRLVSGFLALAWDDLRSCLFSYHRLRDFILJAARTVELLGHSSLKGLRLGWEGLKYL  
 WNLLXYWGRELKISAINLXDTIAIAGVWTDTRVIEIGQRIGRAILHIPRRIRQGLERALL

FIG. 5

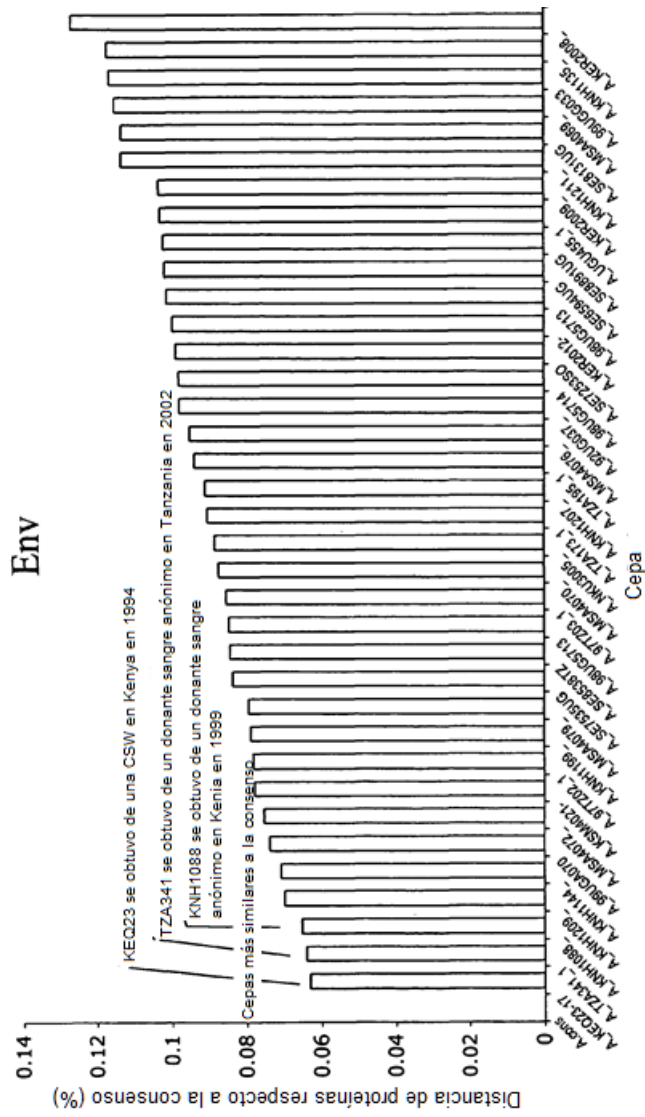


FIG. 6



Consensus Nef  
MGGKWSKSSIVGWPEVRERMRRTPXAAX-----  
GVGAVSQDLKDHGAISSNIN  
H--PSCVWLEAQEEEE--  
VGFPVRPQVPLRPMTYKGAXDLSHFLKEKGGLDGLIYSRK  
RQ  
EILDWVYHTQGYFPDWQNYTPGPGXRYPLTFGWCFKLV  
PVDPDEVEKATEGENSSLHP  
ICQHGMDDEEREVLXWKFDSRLALKHRAXELHPEFYKD

Se muestra la secuencia consenso de 50%. Hay 6 posiciones en las que no se pudo alcanzar un 50% de consenso. La media de la distancia de proteínas en nef fue de 9.3%, rango 3.2% - 16.1%.

*FIG. 7*

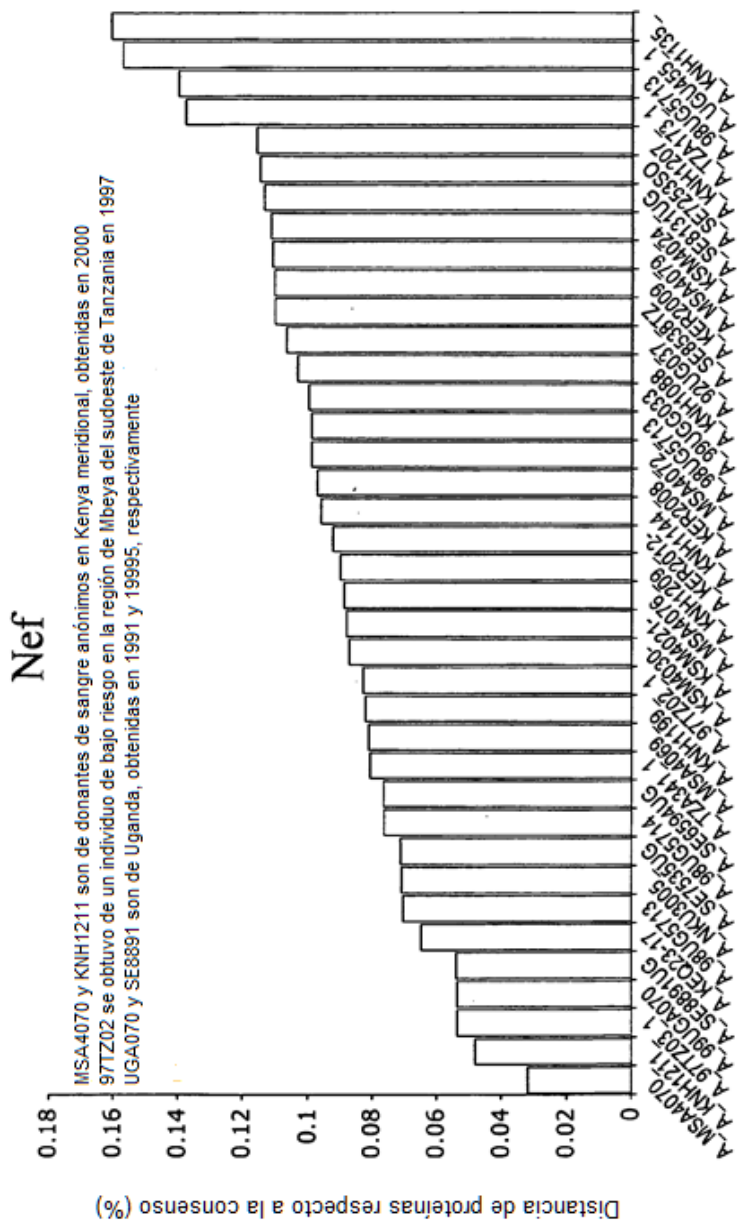
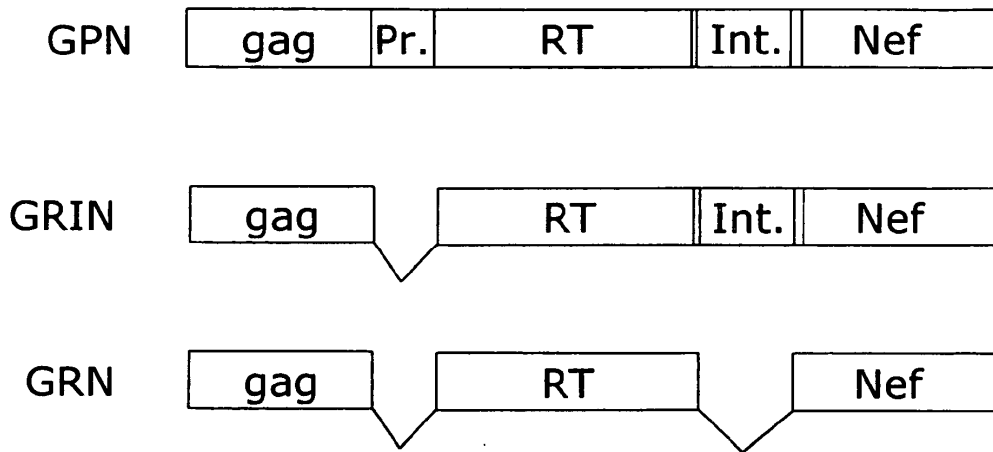


FIG. 8



*FIG. 9*

**GAG**

AY253305 8766 bp ADN lineal VRL 26-AGOSTO-2004

Cepa 01TZA173 del VIH-1 de Tanzania: genes de la proteína gag (gag) y la proteína pol (pol), cds parciales; y genes de la proteína vif (vif), la proteína vpr (vpr), la proteína tat (tat), la proteína rev (rev), la proteína vpu (vpu), la glucoproteína de la envoltura (env) y la proteína (nef), cds completos

\*MG ARASILSGGKLD AWEKIRLRPGGKKYRLKHLVWASRELDRFAL  
 NPSLLETTEGCQQIMNQLQPAVKVTGTEEEKSLFNVTATLYCVHORIDVVDITKEALDKI  
 EEIQNKSKQKTQQAADTGDSSKVSQNYPIIQNAQQQMIHQNLSPRTLNAWVKVIEEK  
 AFSPEVIPMFSALSEGATPQDLNVMNLNIVGGHQAAMQMLKDTINEEAAEWDRLHPVQA  
 GPIPPGQIREPRGSDIAGTTS TPQEQLQWMTGNPPIPVGNLYKRWILGLNKIVRMY S  
 PVSILDIKQGPKEPFRDYVDRFFKALRAEQATQDVKGWMTETLLVQNANPDCKSILKA  
 LSGGATLEEMMTACQVGGPGHKARVLAEAMSAQQTNIMMQRGNFRGQKRIKCFNCG  
 KEGHLARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSSKGRPGNFPQSRPE  
 PTAPPAELFGMGEIASLPKQEQKDREQVPPLVSLKSLFGNDPLSQ

\*MG falta en la entrada de Genbank, artefacto del iniciador amplicon

**FIG. 10**

## POL

AF457081 8827 bp ADN lineal VRL 11-octubre-2002  
 Cepa 00KE\_MSA4070 del VIH-1 de Kenya, genoma parcial

PQJTLWQRPLVTVKIGGQKKEALLDTGADDTVLEDNLPGKWKPRM  
 IGGIGGFIKVKQYDQDILIEICGKKAIGTVLVGPTVNIGRNMLTQIGCTLNFPISPI  
 ETVPVTLKPGMDGPKVKQWPLTEEKIKALTEICTEMEKESKIGPENPNYTPIFAI  
 KKKDSTKWRKLVDFRELNKRITQDFWEVQLGPHPAGLKKKKSVTVLDVGDAYFSVPLD  
 ENFRKYTAFITPSTNNEITPGVRYQYNVLPQGWKGSFAIFQSSMTKILEPFRSKNPEII  
 IYQYMDLTVGSDLEIGQHRTKIEELRAHLLSWGFTTPDKKHQKEPFLWMGYELHPD  
 KWTVPIMLPDKESWTVNDIQKLVGKLNWASQYAGIKVKQLCRLLRGAKALDIDVTL  
 TEEAELELAENREILKDPVHGYYDPSKDLVAEIQKQGDQWYQYQEPFKNLKTGK  
 YARKRSAHINDVRQLAEVVQKVAMESIVWGTPKFKLPIQKETWETWWMDDYVQATWI  
 PEWEFVNTPLVKLWYQLEKDPILGAETFYVDGAANRETCLGKAGYVTDGRQKVVSL  
 TETTQKTELHAILLALQDSGSEVNIIVDSQYALGIIQAQPDRSESELVNIQIEKLLIG  
 KDKIYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEDHERYHSNWRMTMSD  
 FNLPPVAKAIVASCDCQQLKGEAMHGQVDCSPGIWQLDCTHLEKGVILVAHVASGY  
 IEAEVIPAETGQETAYFLKLAGRWPVKVHTDNGSNFTSAAVKAACWVANIQQEFGI  
 PYNPQSQGVVESHMKNELKKIIGQVRDQAEHLKTAVQMAVFIHFKRKGIGGYSAGER  
 IIDHATDIQTKELQKITKIQNFRVYYRDSRDPWKGPAKLLWKGEHAVVIQDNDSI  
 KVVPRRKAKILRDYGGKQMGAGDDCVAGRQDED

FIG. 11

NEF

AF457081 8827 bp ADN lineal VRL 11-octubre-2002  
Cepa 00KE\_MSA4070 del VIH-1 de Kenya, genoma parcial

MGKWSKGSIVGWPEIRERMRRAPA AAPGVGAVSQDLDKHGAI  
SSNINNPSCVWLEAQEEEEVGFVVRPQVPLRPMTYKGAFDLSHFLKKEKGGLDGLJYSR  
KRQEILDWVYHTQGYFPDWNQNYTPGPGVRYPLTFGWCFKLVPMEPDEVEKATEGENN  
SLLHPTCQHGMDDEREVLIWKFSRLALKHRAQELHPEFYKDC

FIG. 12

**ENV**

AY253314 8758 BP ADN LINEAL\_VRL 26-AGOSTO-2004

Cepa 01TZA173 del VIH-1 de Tanzania: genes de la proteína gag (gag) y la proteína pol (pol), cds parciales; y genes de la proteína vif (vif), la proteína vpr (vpr), la proteína tat (tat), la proteína rev (rev), la proteína vpu (vpu), la glucoproteína de la envoltura (env) y la proteína (nef), cds completos

La secuencia Env gp140 siguiente NO incluye la región transmembrana

```

MRVMEIQRNCQHLLRWGIMLGMIIICSTADNLWVTVYYGVPVW
RDAETTLFCASDAKAYSTEKHNVWATHACVPTDNPQEIPLDNVTEEFNMWKNMVDQ
MHEDIISLWDQSLKPCVQLTPLCVTLNCSNARVNAFENSTEDREGMKNC SFNMITTELK
DKKQQVYSLFYRLDIEKINSNNSEYRLVNCNTSAITQACPKVTFEPIPIHYCAPAG
FAILKCNDFEFGTGPCKNVSTVQCETHGKIPVSTQLLLNGSLAEREVIRSENIANN
AKNIIVQFASPVKINCRPNNTNRKSYRIGPGQIFYATDIVGDIRQAHCVSRIDWNN
TLRLVANQLRKYFSNKTHFTNSSGGDLEITHSENCGGEFFYCNSTSGLFNSTWTTNN
MQESNDTNGTITLPCRKQIIRMWQRVGQAMYAPPIEGVTRCESNITGLILTRDGGN
NNSANETFRPGGDIRDNWRSELYKYKVVKIEPLGVAPTRAKRRVVEREKRAVVGIGAV
FLGFLGAAGSTMGAASITLTVQARQLLSGIVQQSNLLRAIEAQQQLLKLTVWGKQL
QARVLAVERYLRDQQLLGIWGC SGKLICTTNPWNSSWSNKS YDDIHWQNNMTWLQWDKE
ISNYTDIHSYLSIEESQNEQEKNEQDLLALDKWANLWNWFDISKWLWYI

```

**FIG. 13**



**Secuencia ensamblada del inserto de GRIN (DZU36984)**

AGTCTTCTGTTTTTACGTAGGTGTCAGCCTAGGTGGTCAATATTGGCCATTAG  
 CCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTG  
 CATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAAC  
 ATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA  
 CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACG  
 GTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAA  
 TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAA  
 TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATC  
 ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTG  
 GCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT  
 ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAAT  
 GGGCGTGGATAGCGGTTTACTCACGGGGATTTCCAAGTCTCCACCCCATTG  
 ACGTCAATGGGAGTTTGTGTTTGGCACCAAATCAACGGGACTTTCAAAAATG  
 TCGTAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGG  
 GAGGTCTATATAAGCAGAGCTCGTTTGTGAAACCGTCAGATCGCCTGGAGAC  
 GCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCT  
 CCGCGGCCGGGAACGGTGCATTGGAAGCTTGGCCGCCACCATGGCCGCCAGA  
 GCCAGACTCCTGAGCGGGGCAAGCTGGACGCCTGGGAGAAGATCAGACTG  
 AGGCTGGCGGCAAGAAGAAGTACCGGCTGAAGCACCTGGTGTGGGCCAGC  
 AGAGAGCTGGATCGCTTCGCCCTGAATCCTAGCCTGCTGGAGACCACCGAGG  
 GCTGCCAGCAGATCATGAACCAGCTGCAGCCCCGCCGTGAAAACCGGCACCGA  
 GGAGATCAAGAGCCTGTTCAACACCGTGGCCACCCTGTAAGTGCCTGCACCAG  
 CGGATCGACGTGAAGGATACCAAGGAGGCCCTGGACAAGATCGAGGAGATC  
 CAGAACAAGAGCAAGCAGAAAACCCAGCAGGCCGCTGCCGACACCGGCGAC  
 AGCAGCAAAGTGAGCCAGAACTACCCCATCATCCAGAATGCCAGGGCCAG  
 ATGATCCACCAGAACCTGAGCCCCAGAACCCTGAATGCCTGGGTGAAAGTGA  
 TCGAGGAAAAGGCCTTCAGCCCCGAAGTGATCCCTATGTTTCAGCGCCCTGAG  
 CGAGGGCGCCACCCCCAGGACCTGAACGTGATGCTGAACATTGTGGGCGGA  
 CACCAGGCCCATGCAGATGCTGAAGGACACCATCAATGAGGAGGCCGCC  
 GAGTGGGACAGACTGCACCCCGTGCAGGCCGACCCATCCCCCTGGCCAGA  
 TCAGAGAGCCCAGAGGCAGCGACATCGCCGGCACCACTCCACCCCTCAAGA  
 ACAGCTGCAGTGGATGACCGGCAACCCTCCATCCCTGTGGGCAACATCTAC  
 AAGCGGTGGATCATCCTGGGCCTGAACAAGATTGTGCGGATGTACAGCCCCG  
 TGTCCATCCTGGATATCAAGCAGGGCCCCAAGGAGCCCTTCAGAGACTACGT  
 GGACCGGTTCTTCAAGGCCCTGAGAGCCGAGCAGGCCACCCAGGACGTGAA  
 GGGCTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAG  
 AGCATCCTGAAGGCCCTGGGCAGCGGCCCACTGGAGGAGATGATGACC  
 GCCTGCCAGGGAGTGGGCGGACCCGGCCACAAGGCCAGAGTGCTGGCCGAG  
 GCCATGAGCCAGGCCAGCAGACCAACATCATGATGCAGCGGGGCAACTTCA  
 GAGGCCAGAAGCGGATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACCTGG  
 CCAGAACTGCAGAGCCCCAGGAAGAAGGGCTGCTGGAAGTGTGGCAAGG  
 AAGGGCACCAGATGAAGGACTGCACCGAGAGGCAGGCCAATTCCTGGGCA  
 AGATTTGGCCTAGCAGCAAGGGCAGACCCGGCAATTTCCCCCAGAGCAGACC

**FIG. 14A**



CAGCCCCATCGAGACCGTGCCCCGTGACCCTGAAGCCCGGCATGGATGGCCCC  
 AAAGTGA AACAGTGGCCCCGTGACCGAGGAGAAGATTAAGGCCCTGACCGAA  
 ATCTGTACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAG  
 AACCCCTACAACACCCCATCTTCGCCATCAAGAAGAAGGACAGCACCAAGT  
 GGCGGAAACTGGTGGACTTCCGGGAGCTGAACAAGAGGACCCAGGACTTCTG  
 GGAAGTGCAGCTGGGCATCCCCACCCTGCCGGCCTGAAGAAGAAGAAGTCC  
 GTGACAGTGTGGATGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACGAGA  
 ACTTCAGGAAGTACACCGCCTTCACCATCCCCAGCACCAACAACGAGACCCC  
 CGGAGTGAGATAACAGTACAACGTGCTGCCTCAGGGCTGGAAGGGCAGCCCC  
 GCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGGAGCAAGA  
 ACCCCGAGATCATCATCTACCAGTACATGGCCGCCCTGTATGTGGGCAGCGA  
 TCTGGAGATCGGCCAGCACAGGACCAAGATCGAAGAGCTGAGGGGCCACCT  
 GCTGAGCTGGGGCTTCAACACCCCGATAAGAAGCACCAGAAGGAGCCCCCT  
 TTCCTGTGGATGGGCTACGAGCTGCACCCCGATAAGTGGACCGTGCAGCCCA  
 TCATGTGTCCCGATAAGGAGAGCTGGACCGTGAACGACATCCAGAACTGGT  
 GGGCAAGCTGAATTGGGCCAGCCAAATCTACGCCGGCATTAAAGTGAAGCAG  
 CTGTGCAGGCTGCTGAGAGCGCCAAAGCCCTGACAGACATCGTGACACTGA  
 CAGAGGAGGCCGAGCTGGAGCTGGCCGAGAACAGGGAGATCCTGAAGGACC  
 CCGTGCACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATTCA  
 GAAGCAGGGCCAGGACCAGTGGACCTACCAATCTACCAGGAGCCTTTCAAG  
 AACCTGAAAACCGGGAAGTACGCCAGGAAGAGAAGCGCCACACCAACGAT  
 GTGAGGCAGCTGGCCGAAGTGGTGCAGAAAGTGGCTATGGAGAGCATCGTG  
 ATCTGGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGG  
 AAACCTGGTGGATGGACTACTGGCAGGCCACCTGGATTCTGAGTGGGAGTT  
 CGTGAACACCCCCCTCTGGTGAAGCTGTGGTATCAGCTGGAGAAGGACCCC  
 ATCCTGGGCGCCGAGACCTTCTACGTGGACGGAGCCGCCAATAGAGAGACCA  
 AGCTGGGCAAGGCCGGCTACGTGACCGACAGAGGCAGACAGAAAGTGGTGT  
 CTCTGACCGAGACAACCAACCAGAAAACCGAGCTGCACGCCATCCTGTGGC  
 CCTGCAGGACACGGCCAGCGAAGTGAACATCGTGACCGACTCCCAGTACGCC  
 CTGGGCATCATTAGGCCAGCCGATAGAAGCGAGAGCGAGCTGGTGAACC  
 AGATCATCGAGAAGCTGATCGGCAAGGACAAAATCTACCTGAGCTGGGTGCC  
 CGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGTCCAG  
 CGGCATCCGAAAGTGCTGTTTCTGGACGGCATCGACAAGGCCAGGAGGAC  
 CACGAGAGATAACACAGCAACTGGCGGACAATGGCCAGCGACTTCAACCTGC  
 CTCCCATCGTGGCCAAGGAGATCGTGGCCAGCTGCGATAAGTGTGAGCTGAA  
 GGGCGAGGCCATGCACGGCCAGGTGGACTGCAGCCCTGGCATCTGGCAGCTG  
 GCCTGCACCCACCTGGAGGGCAAAGTGATTCTGGTGGCCGTGCACGTGGCCA  
 GCGGCTACATCGAGGCCGAAGTGATTCCC GCCGAGACCGGCCAGGAGACCG  
 CCTACTTCTGTGAAGCTGGCCGGCAGATGGCCCGTGAAGTGGTGCACAC  
 CGCCAACGGCAGCAACTTACCTCTGCCGCCGTGAAGGCCGCCTGTTGGTGG  
 GCCAATATCCAGCAGGAGTTCGGCATCCCCTACAACCCTCAGAGCCAGGGCG  
 TGGTGGCCAGCATGAACAAGGAGCTGAAGAAGATCATCGGCCAGGTGAGGG  
 ACCAGGCCGAGCACCTGAAAACAGCCGTGCAGATGGCCGTGTTTATCCACAA  
 CTTCAAGCGGAAGGGCGGCATTGGCGGCTACAGCGCCGGAGAGCGGATCAT  
 CGACATCATCGCCACCGATATCCAGACCAAGGAACTGCAGAAGCAGATCACC  
 AAGATTCAAGAACTTCAAGAGTGTACTACCGGGACAGCAGGGACCCCATCTGGA

**FIG. 14B**

AGGATGAGGACAGATCTATGGGCGGCAAGTGGTCCAAGGGCAGCATTGTGG  
GCTGGCCCAGATCCGGGAGAGAATGAGAAGAGCCCCTGCCGCCGCTCCT  
GGAGTGGGCGCCGTGTCTCAGGATCTGGATAAGCACGGCGCCATCACCAG  
CAGCAACATCAACAACCCAGCTGTGTGGCTGGAGGCCAGGAAGAGGA  
GGAAGTGGGCTTCCCTGTGAGACCCAGGTGCCCTGAGACCCATGACCTA  
CAAGGGCGCCTTCGACCTGAGCCACTTCCTGAAGGAGAAGGGCGGCCTGG  
ACGGCCTGATCTACAGCCGGAAGCGGCAGGAGATCCTGGATCTGTGGGTGT  
ACACACCCAGGGCTACTTCCCCGACTGGCAGAATTACACCCCTGGCCCTG  
GAGTGC GG TATCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTGCCTATGG  
AGCCCGACGAAGTGGAGAAGGCCACAGAGGGCGAGAACAACAGCCTGCTG  
CACCCATCTGCCAGCACGGCATGGACGATGAGGAGCGGGAAGTGCTGATC  
TGGAAGTTCGACAGCAGGCTGGCCCTGAAGCACAGAGCCCAGGAACTGCAC  
CCAGAGTTCTACAAGGACTGCTGATGATCATAATAATCTAGACGAGATCCGA  
ACTTGTATTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT  
TCACAAATAAAGCATTTCCTGACTGCATTCTAGTTGTGGTTTGTCCAACTCA  
TCAATGTATCTTATCATGTCTAGATCTGAGGTATGATGATACGAGATCGAGGG  
TGCGCGCATGCGAATGCGGAGGCAAGCATGCCAGGTTCCAGC

Nota: los sitios de clonación están subrayados y en negrita

**FIG. 14C**

Apéndice 2: Secuencia ensamblada del inserto de Env (DSP33447\_01)

ATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA  
 TGTATTTAgAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG  
 TGCCACCTGACGTCTAAGAAACCATATTATCATgACATTAACCTATAAAAAATA  
 GCGIATCACGAGGCCCTTCGTCTTCAAGAATTGGTCGATGGCAAACAGCT  
 ATATGGGTATTATGGGTTTCAATTAATTAATCGACATCATCAATAATATACCT  
 TATAGATGGAATGGTGCCAATATGTAAATGAGGTGATTTTAAAAAGTGTGGG  
 CCGTGTGGTGATTGGCTGTGGGGTTAACGGTTAAAAGGGGCGGCGCGGCC  
 GTGGGAAAATGACGTTTTATGGGGGTGGAGTTTTTTTTGCAAGTTGTCGCGG  
 GAAATGTTACGCATAAAAAGGCTTCTTTTCTCACGGAACACTTAGTTTTCCC  
 ACGGTATTTAACAGGAAATGAGGTAGTTTTGACCGGATGCAAGTGAAAATTG  
 CTGATTTTCGCGCGAAAAGTGAATGAGGAAGTGTCTGAATAATGTGGT  
 ATTTATGGCAGGTGGAGTATTTGTTTCAGGGCCAGGTAGACTTTGACCCATT  
 ACGTGGAGTTTTCGATTACCGTGTTTTTTACCTGAATTTCCGCGTACCGTGT  
 CAAGTCTTCTGTTTTACGTAGGTGTCAGCCTAGGTGGTCAATATTGGCCA  
 TTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCC  
 ATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTC  
 CAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCA  
 ATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC  
 TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCATTG  
 ACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATT  
 GACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC  
 AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATG  
 GCCGCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTG  
 GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG  
 CAGTACATCAATGGGCGTGGATAGCGTTTTGACTCACGGGGATTTCCAAGT  
 CTCCACCCCATGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGG  
 GACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAAATGGGCGGT  
 AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT  
 CAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACAC  
 CGGGACCGATCCAGCCTCCGCGGCCGGAACGGTGCATTGGAAGCTTGCC  
 GCCACCATGAGGGTGTATGGAGATCCAGCGGAACTGCCAGCACCTGCTGAG  
 ATGGGGCATCATGATCCTGGGCATGATTATCATCTGCAGCACCGCCGACAA  
 CCTGTGGGTGACCGTGTACTACGGCGTGCCTGTGTGGAGAGATGCCGAGA  
 CCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACAGCACCGAGAAGCAC  
 AATGTGTGGGCCACCCACGCCTGCGTGCCTACCGATCCCAACCCTCAGGA  
 GATCCCCCTGGACAACGTGACCGAGGAGTTCAACATGTGGAAGAACAACAT  
 GTGGACCAGATGCACGAGGACATCATCAGCCTGTGGGACCA

FIG. 15A

GAGCCTGAAGCCCTGCGTGCAGCTGACCCCCCTGTGCGTGACCCTGAACT  
GCAGCAACGCCAGAGTGAACGCCACCTTCAACTCCACCGAGGACAGGGAG  
GGCATGAAGAACTGCAGCTTCAACATGACCACCGAGCTGCGGGATAAGAAG  
CAGCAGGTGTACAGCCTGTTCTACCGGCTGGACATCGAGAAGATCAACAGC  
AGCAACAACAACAGCGAGTACCGGCTGGTGAAGTGAATACCAGCGCCATC  
ACCCAGGCCTGCCCTAAGGTGACCTTCGAGCCCATCCCCATCCACTACTGC  
GCCCCTGCCGGCTTCGCCATCCTGAAGTGAACGACACCGAGTTCAATGG  
CACCGGCCCCCTGCAAGAATGTGAGCACCGTGCAGTGCACCCACGGCATCA  
AGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAGA  
GAAGTGCGGATCAGGAGCGAGAACATCGCCAACAACGCCAAGAACATCATC  
GTGCAGTTCGCCAGCCCCGTGAAGATCAACTGCATCCGGCCCCAACAACAAT  
ACCCGGAAGAGCTACAGAATCGGCCCTGGCCAGACCTTCTACGCCACCGA  
CATTGTGGGCGACATCAGACAGGCCCACTGCAACGTGTCCAGGACCGACT  
GGAACAACACCCTGAGACTGGTGGCCAACCAGCTGCGGAAGTACTTCAGC  
AACAAGACCATCATCTTCAACCAACAGCAGCGGCGGAGACCTGGAGATCACC  
ACCCACAGCTTCAATTGTGGCGGCGAGTTCTTCTACTGCAACACCTCCGGC  
CTGTTCAATAGCACCTGGACCACCAACAACATGCAGGAGTCCAACGCACACC  
AGCAACGGCACCATCACCTGCCCTGCCGGATCAAGCAGATCATCCGGAT  
GTGGCAGCGCTGGGGCCAGGCCATGTACGCCCTCCCATCGAGGGCGTG  
ATTCGCTGCGAGCAACATCACCGGCTGATCCTGACCAGAGATGGCGG  
CAACAACAATTCCGCCAACGAGACCTTCAGACCTGGCGGCGGAGATATCCG  
GGACAACCTGGCGGAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGC  
CCCTGGGCGTGGCCCCCACCAGAGCCAAGAGAAGAGTGGTGGAGCGGGA  
GAAGAGAGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCG  
CCGGATCTACAATGGGAGCCGCCAGCATCACCTGACCGTGCAGGCCAGA  
CAGCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAATCTGCTGAGAGCCAT  
CGAGGCCCAGCAGCAGCTGCTGAAGCTGACAGTGTGGGGCATCAAGCAGC  
TGCAGGCCAGGGTGTGGCCGTGGAGAGATACCTGAGGGACCAGCAGCTC  
CTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCAACGTGCC  
CTGGAATAGCAGCTGGAGCAACAAGAGCTACGACGACATCTGGCAGAACAT  
GACCTGGCTGCAGTGGGACAAGGAGATCAGCAACTACACCGACATCATCTA  
CAGCCTGATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGCAGGATC  
TGCTGGCCCTGGACAAGTGGGCCAACCTGTGGAAGTGGTTTCGACATCAGC  
AAGTGGCTGTGGTACATCAGATCTTGATAATCTAGACGAGATCCGAACCTTGT  
TTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACA  
AATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA  
TGTATCTTATCATGTCTAGATCTGAGGTATGATGATACGAGATCGAGGGTGC  
GCGCATGCGAATGCGGAGGCAAGCATGCCAGGTTCCAGCCGGTGTGTGTA  
GATGTGACCGAAGATCTCAGACCGGATCATTTGGTTATTGCCCGCACTGGA  
GCAGAGTTCGGATCCAGTGGAGAAGAACTGACTAAGGTGAGTATTGGGAA  
AACTTTG

*FIG. 15B*

**GGGTGGGATTTTCAGATGGACAGATTGAGTAAAAATTTGTTTTTCTGTCTTG**  
**CAGCTGACATGACT**  
**GGAAATGCTTCTTTTAAGGGGGGGAGTCTTCAGCCCTTATCTGACAGGGCG**  
**TCTCCCATCCTGGGCAGGAGTTCGT**

Nota: los sitios de clonación están subrayados y en negrita

*FIG. 15C*

```

HindIII      NcoI      BstNI
1  AAGCTTGCCGCCACCATGGCCGCCAGAGCCAGCATCCTGAGCGGGGGCAAGCTGGACGCC
   -----+-----+-----+-----+-----+-----+-----+
TTCGAACGGCGGTGGTACCGGCGGTCTCGGTCGTAGGACTCGCCCCGTTGACCTGCGG
      M_A_A_R_A_S_I_L_S_G_G_K_L_D_A_

                StuI
                BstNI      BstNI
61  TGGGAGAAGATCAGACTGAGGCCTGGCGGCAAGAAGAAGTACCGGCTGAAGCACCTGGTG
   -----+-----+-----+-----+-----+-----+-----+
ACCTCTTCTAGTCTGACTCCGGACCGCGTTCTTCTTCATGGCCGACTTCGTGGACCAC
W_E_K_I_R_L_R_P_G_G_K_K_K_Y_R_L_K_H_L_V_

                                HinfI      BsaI
121  TGGGCCAGCAGAGAGCTGGATCGTTCGCCCTGAATCCTAGCCTGCTGGAGACCACCGAG
   -----+-----+-----+-----+-----+-----+-----+
ACCCGGTCGTCTCTCGACCTAGCGAAGCGGGACTTAGGATCGGACGACCTCTGGTGGCTC
W_A_S_R_E_L_D_R_F_A_L_N_P_S_L_L_E_T_T_E_

                PvuII
                PstI
181  GGCTGCCAGCAGATCATGAACCAGCTGCAGCCCGCCGTGAAAACCGGCACCGAGGAGATC
   -----+-----+-----+-----+-----+-----+-----+
CCGACGGTCGTCTAGTACTTGGTCGACGTCGGGCGGCACTTTTGGCCGTGGCTCCTCTAG
G_C_Q_Q_I_M_N_Q_L_Q_P_A_V_K_T_G_T_E_E_I_

241  AAGAGCCTGTTCAACACCGTGGCCACCCTGTA CTGCGTGCACCAGCGGATCGACGTGAAG
   -----+-----+-----+-----+-----+-----+-----+
TTCTCGGACAAGTTGTGGCACCGGTGGGACATGACGCACGTGGTCGCCTAGCTGCACTTC
K_S_L_F_N_T_V_A_T_L_Y_C_V_H_Q_R_I_D_V_K_

                BstNI
301  GATACCAAGGAGGCCCTGGACAAGATCGAGGAGATCCAGAACAAGAGCAAGCAGAAAACC
   -----+-----+-----+-----+-----+-----+-----+
CTATGGTTCCTCCGGGACCTGTTCTAGCTCCTCTAGGTCTTGTTCGTTTCGTTCTTTTGG
D_T_K_E_A_L_D_K_I_E_E_I_Q_N_K_S_K_Q_K_T_

361  CAGCAGGCCGCTGCCGACACCGGCGACAGCAGCAAAGTGAGCCAGAACTACCCCATCATC
   -----+-----+-----+-----+-----+-----+-----+
GTCGTCCGGGACGGCTGTGGCCGCTGTCGTCGTTTCACTCGGTCTTGATGGGGTAGTAG
Q_Q_A_A_A_D_T_G_D_S_S_K_V_S_Q_N_Y_P_I_I_

                BstNI      BstNI
421  CAGAATGCCCAGGGCCAGATGATCCACCAGAACCTGAGCCCCAGAACCTGAATGCCTGG
   -----+-----+-----+-----+-----+-----+-----+
GTCTTACGGGTCCCAGTCTACTAGGTGGTCTTGGACTCGGGTCTTGGGACTTACGGACC
Q_N_A_Q_G_Q_M_I_H_Q_N_L_S_P_R_T_L_N_A_W_

```

FIG. 16A

```

                                StuI                                HaeII
481  GTGAAAGTGATCGAGGAAAAGCCTTCAGCCCCGAAGTGATCCCTATGTTTCAGCGCCCTG
-----+-----+-----+-----+-----+-----+-----+
    CACTTTCACTAGCTCCTTTTCCGGAAGTCGGGGCTTCACTAGGGATAACAAGTCGCGGGAC
    V_K_V_I_E_E_K_A_F_S_P_E_V_I_P_M_F_S_A_L_

                                NarI
                                KasI
                                HaeII          BstNI                                BstNI
541  AGCGAGGGCGCCACCCCCAGGACCTGAACGTGATGCTGAACATTGTGGGCGGACACCAG
-----+-----+-----+-----+-----+-----+-----+
    TCGCTCCC CGGTGGGGGCTCTGGACTTGCCTACGACTTGTAAACACCCGCCTGTGGTC
    S_E_G_A_T_P_Q_D_L_N_V_M_L_N_I_V_G_G_H_Q_

601  GCCGCCATGCAGATGCTGAAGGACACCATCAATGAGGAGGCCCGAGTGGGACAGACTG
-----+-----+-----+-----+-----+-----+-----+
    CGGCGGTACGTCTACGACTTCTGTGGTAGTTACTCCTCCGGCGGCTCACCTGTCTGAC
    A_A_M_Q_M_L_K_D_T_I_N_E_E_A_A_E_W_D_R_L_

                                BstNI
661  CACCCCGTGCAGGCCGACCCATCCCCCTGGCCAGATCAGAGAGCCCAGAGGCAGCGAC
-----+-----+-----+-----+-----+-----+-----+
    GTGGGGCACGTCCGGCCTGGGTAGGGGGGACCGGTCTAGTCTCTCGGGTCTCCGTCGCTG
    H_P_V_Q_A_G_P_I_P_P_G_Q_I_R_E_P_R_G_S_D_

                                PvuII
                                PstI                                BstXI
721  ATCGCCGGCACCACCTCCACCCCTCAAGAACAGCTGCAGTGGATGACCGGCAACCCTCCC
-----+-----+-----+-----+-----+-----+-----+
    TAGCGCCCGTGGTGGAGGTGGGGAGTTCTTGTGCGACGTACCTACTGGCCGTTGGGAGGG
    I_A_G_T_T_S_T_P_Q_E_Q_L_Q_W_M_T_G_N_P_P_

                                BstNI
781  ATCCCTGTGGGCAACATCTACAAGCGGTGGATCATCCTGGGCCTGAACAAGATTGTGCGG
-----+-----+-----+-----+-----+-----+-----+
    TAGGGACACCCGTTGTAGATGTTGCGCCACCTAGTAGGACCCGGACTTGTCTAACACGCC
    I_P_V_G_N_I_Y_K_R_W_I_I_L_G_L_N_K_I_V_R_

                                EcoRV
                                BstNI          ApaI
841  ATGTACAGCCCCGTGTCCATCCTGGATATCAAGCAGGGCCCCAAGGAGCCCTTCAGAGAC
-----+-----+-----+-----+-----+-----+-----+
    TACATGTCGGGGCACAGGTAGGACCTATAGTTCGTCCCGGGGTTCTCGGGAAGTCTCTG
    M_Y_S_P_V_S_I_L_D_I_K_Q_G_P_K_E_P_F_R_D_

                                AgeI                                BstNI
901  TACGTGGACCGGTTCTTCAAGGCCCTGAGAGCCGAGCAGGCCACCCAGGACGTGAAGGGC
-----+-----+-----+-----+-----+-----+-----+
    ATGCACCTGGCCAAGAAGTTCGGGACTCTCGGCTCGTCCGGTGGGTCTGCACTTCCCG
    Y_V_D_R_F_F_K_A_L_R_A_E_Q_A_T_Q_D_V_K_G_

```

FIG. 16B

ES 2 396 915 T3

**BsaI**  
 961 TGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGAGCATCCTGAAG  
 -----+-----+-----+-----+-----+-----+-----+  
 ACCTACTGGCTCTGGGACGACCACGTCTTGGGGTTGGGGCTGACGTTCTCGTAGGACTTC  
 W\_M\_T\_E\_T\_L\_L\_V\_Q\_N\_A\_N\_P\_D\_C\_K\_S\_I\_L\_K\_

**NarI**  
**KasI**  
**BstNI**                      **HaeII**    **PfI**MI  
**BstNI**  
 1021 GCCCTGGGCAGCGGCCACACTGGAGGAGATGATGACCGCCTGCCAGGGAGTGGGCGGA  
 -----+-----+-----+-----+-----+-----+-----+  
 CGGGACCCGTCGCCGCGGTGTGACCTCCTCTACTACTGGCGGACGGTCCCTCACCCGCCT  
 A\_L\_G\_S\_G\_A\_T\_L\_E\_E\_M\_M\_T\_A\_C\_Q\_G\_V\_G\_G\_

**BstXI**  
 1081 CCCGGCCACAAGGCCAGAGTGTGCGCCGAGGCCATGAGCCAGGCCCAGCAGACCAACATC  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGCCGGTGTTCGCGTCTCAGACCGGTCCGGTACTCGGTCCGGTTCGTCTGGTTGTAG  
 P\_G\_H\_K\_A\_R\_V\_L\_A\_E\_A\_M\_S\_Q\_A\_Q\_Q\_T\_N\_I\_

**BstNI**  
 1141 ATGATGCAGCGGGCAACTTCAGAGGCCAGAAGCGGATCAAGTGCTTCAACTGCGGCAAG  
 -----+-----+-----+-----+-----+-----+-----+  
 TACTACGTGCGCCCGTTGAAGTCTCCGGTCTTCGCCTAGTTCACGAAGTTGACGCCGTTT  
 M\_M\_Q\_R\_G\_N\_F\_R\_G\_Q\_K\_R\_I\_K\_C\_F\_N\_C\_G\_K\_

**BstNI**                      **PstI**                      **BstNI**  
 1201 GAGGGCCACCTGGCCAGAACTGCAGAGCCCCAGGAAGAAGGGCTGCTGGAAGTGTGGC  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCCTCGGTGGACCGGTCTTTGACGTCTCGGGGTCTTCTTCCCGACGACCTTCACACCG  
 E\_G\_H\_L\_A\_R\_N\_C\_R\_A\_P\_R\_K\_K\_G\_C\_W\_K\_C\_G\_

**BstXI**    **BstNI**  
 1261 AAGGAAGGGCACCAGATGAAGGACTGCACCGAGAGGCAGGCCAATTTCTGGGCAAGATT  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCCTTCCCGTGGTCTACTTCTGACGTGGCTCTCCGTCCGGTTAAAGGACCCGTTCTAA  
 K\_E\_G\_H\_Q\_M\_K\_D\_C\_T\_E\_R\_Q\_A\_N\_F\_L\_G\_K\_I\_

**BstXI**    **BstNI**  
 1321 TGGCCTAGCAGCAAGGGCAGACCCGGCAATTTCCCCCAGAGCAGACCCGAGCCCACCGCC  
 -----+-----+-----+-----+-----+-----+-----+  
 ACCGGATCGTCGTTCCCGTCTGGGCGGTTAAAGGGGTCTCGTCTGGGCTCGGGTGGCGG  
 W\_P\_S\_S\_K\_G\_R\_P\_G\_N\_F\_P\_Q\_S\_R\_P\_E\_P\_T\_A\_

**BstXI**    **BstNI**  
 1381 CCTCCCGCCGAGCTGTTCCGGCATGGGCGAGGGCATCGCCAGCCTGCCAAGCAGGAGCAG  
 -----+-----+-----+-----+-----+-----+-----+  
 GGAGGGCGGCTCGACAAGCCGTACCCGCTCCCGTAGCGGTTCGGACGGGTTCGTCCTCGTC  
 P\_P\_A\_E\_L\_F\_G\_M\_G\_E\_G\_I\_A\_S\_L\_P\_K\_Q\_E\_Q\_

**BspMI**                      **BstNI**  
 1441 AAGGACAGAGAGCAGGTGCCCCCCCTGGTGTCCCTGAAGTCCCTGTTCTGGCAACGATCCT  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCCTGTCTCTCGTCCACGGGGGGACCACAGGGACTTCAGGGACAAGCCGTTGCTAGGA  
 K\_D\_R\_E\_Q\_V\_P\_P\_L\_V\_S\_L\_K\_S\_L\_F\_G\_N\_D\_P\_

**FIG. 16C**



BstNI NcoI BstNI  
 BamHI BsaI BstEII  
 1501 CTGAGCCAGGGATCCATGGCCCCCAGATCACCCCTGTGGCAGAGACCCCTGGTGACCGTG  
 -----+-----+-----+-----+-----+-----+-----+  
 GACTCGGTCCCTAGGTACCGGGGGTCTAGTGGGACACCGTCTCTGGGGACCACTGGCAC  
 L\_S\_Q\_G\_S\_M\_A\_P\_Q\_I\_T\_L\_W\_Q\_R\_P\_L\_V\_T\_V\_

NarI  
 KasI  
 HaeII  
 PvuII  
 1561 AAGATCGGCGGCCAGCTGAAGGAAGCCCTGCTGGATACAGGCGCCGATGATACCGTGCTG  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCTAGCCGCCGGTGCAGTTCCTTCGGGACGACCTATGTCCGCGGCTACTATGGCACGAC  
 K\_I\_G\_G\_Q\_L\_K\_E\_A\_L\_L\_D\_T\_G\_A\_D\_D\_T\_V\_L\_

BspMI  
 1621 GAGGACATCAACCTGCCCGCAAGTGAAGCCCTAGAATGATCGGCGGCATCGGGGGCTTC  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCCTGTAGTTGGACGGGCCGTTACCTTCGGATCTTACTAGCCGCCGTAGCCCCGAAG  
 E\_D\_I\_N\_L\_P\_G\_K\_W\_K\_P\_R\_M\_I\_G\_G\_I\_G\_G\_F\_

1681 ATCAAAGTGAAGCAGTACGACCAGATCCTGATCGAGATTTGCGGGAAGAAGGCCATCGGC  
 -----+-----+-----+-----+-----+-----+-----+  
 TAGTTTCACTTCGTCATGCTGGTCTAGGACTAGCTCTAAACGCCCTTCTTCCGGTAGCCG  
 I\_K\_V\_K\_Q\_Y\_D\_Q\_I\_L\_I\_E\_I\_C\_G\_K\_K\_A\_I\_G\_

ApaI EagI  
 1741 ACCGTGCTGGTGGGCCCCACCCCTGTGAATATCATCGGCCGGAACATGCTGACCCAGATC  
 -----+-----+-----+-----+-----+-----+-----+  
 TGGCAGCACCACCCGGGGTGGGACACTTATAGTAGCCGGCCTTGTACGACTGGGTCTAG  
 T\_V\_L\_V\_G\_P\_T\_P\_V\_N\_I\_I\_G\_R\_N\_M\_L\_T\_Q\_I\_

BsaI  
 1801 GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGACCCTGAAGCCC  
 -----+-----+-----+-----+-----+-----+-----+  
 CCGACGTGGGACTTGAAGGGTAGTCTGGGGTAGCTCTGGCACGGGCACTGGGACTTCGGG  
 G\_C\_T\_L\_N\_F\_P\_I\_S\_P\_I\_E\_T\_V\_P\_V\_T\_L\_K\_P\_

1861 GGCATGGATGGCCCCAAAGTGAACAGTGGCCCCTGACCGAGGAGAAGATTAAGGCCCTG  
 -----+-----+-----+-----+-----+-----+-----+  
 CCGTACCTACCGGGGTTTCACTTTGTACCGGGGACTGGCTCCTTCTTAATTCCGGGAC  
 G\_M\_D\_G\_P\_K\_V\_K\_Q\_W\_P\_L\_T\_E\_E\_K\_I\_K\_A\_L\_

1921 ACCGAAATCTGTACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAAC  
 -----+-----+-----+-----+-----+-----+-----+  
 TGGCTTTAGACATGGCTCTACCTCTTCTCCCGTTCTAGTCGTTCTAGCCGGGGCTCTTG  
 T\_E\_I\_C\_T\_E\_M\_E\_K\_E\_G\_K\_I\_S\_K\_I\_G\_P\_E\_N\_

1981 CCCTACAACACCCCATCTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGGAAACTG  
 -----+-----+-----+-----+-----+-----+-----+  
 GGGATGTTGTGGGGGTAGAAGCGGTAGTTCTTCTTCTGTCGTGGTTACCGCCTTTGAC  
 P\_Y\_N\_T\_P\_I\_F\_A\_I\_K\_K\_K\_D\_S\_T\_K\_W\_R\_K\_L\_

FIG. 16D

```

                                     BstNI          PvuII
2041  GTGGACTTCCGGGAGCTGAACAAGAGGACCCAGGACTTCTGGGAAGTGCAGCTGGGCATC
      -----+-----+-----+-----+-----+
      CACCTGAAGGCCCTCGACTTGTTCCTGGGTCCTGAAGACCCTTCACGTCGACCCGTAG
      V_D_F_R_E_L_N_K_R_T_Q_D_F_W_E_V_Q_L_G_I_

2101  CCCCACCCTGCCGGCCTGAAGAAGAAGAAGTCCGTGACAGTGCTGGATGTGGGCGACGCC
      -----+-----+-----+-----+-----+
      GGGGTGGGACGGCCGACTTCTTCTTCTCAGGCACTGTCACGACCTACACCCGCTGCGG
      P_H_P_A_G_L_K_K_K_K_S_V_T_V_L_D_V_G_D_A_

                                     BstNI
2161  TACTTCAGCGTGCCCTGGACGAGAACTTCAGGAAGTACACCGCCTTCACCATCCCCAGC
      -----+-----+-----+-----+-----+
      ATGAAGTCGCACGGGGACCTGCTCTTGAAGTCCTTCATGTGGCGGAAGTGGTAGGGGTCG
      Y_F_S_V_P_L_D_E_N_F_R_K_Y_T_A_F_T_I_P_S_

                                     BsaI
2221  ACCAACAAACGAGACCCCCGGAGTGAGATACCAGTACAACGTGCTGCCTCAGGGCTGGAAG
      -----+-----+-----+-----+-----+
      TGGTTGTTGCTCTGGGGGCCTCACTCTATGGTCATGTTGCACGACGGAGTCCCGACCTTC
      T_N_N_E_T_P_G_V_R_Y_Q_Y_N_V_L_P_Q_G_W_K_

                                     BstXI  BstNI
2281  GGCAGCCCCGCCATCTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGGGAGCAAG
      -----+-----+-----+-----+-----+
      CCGTCGGGGCGGTAGAAGGTCTCGTCGTAAGTCTAGGACCTCGGGAAGGCCCTCGTTC
      G_S_P_A_I_F_Q_S_S_M_T_K_I_L_E_P_F_R_S_K_

                                     PflMI
2341  AACCCCGAGATCATCATCTACCAGTACATGGCCGCCCTGTATGTGGGACGCGATCTGGAG
      -----+-----+-----+-----+-----+
      TTGGGGCTCTAGTAGTAGATGGTCATGTACCGGGCGGGACATACACCCGTCGCTAGACCTC
      N_P_E_I_I_I_Y_Q_Y_M_A_A_L_Y_V_G_S_D_L_E_

                                     ApaI  BspMI
2401  ATCGGCCAGCACAGGACCAAGATCGAAGAGCTGAGGGCCACCTGCTGAGCTGGGGCTTC
      -----+-----+-----+-----+-----+
      TAGCCGGTCGTGCTCCTGGTTCTAGCTTCTCGACTCCCGGGTGGACGACTCGACCCCGAAG
      I_G_Q_H_R_T_K_I_E_E_L_R_A_H_L_L_S_W_G_F_

2461  ACCACCCCGATAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTG
      -----+-----+-----+-----+-----+
      TGGTGGGGCTATTCTTCGTGGTCTTCTCGGGGAAAGGACACCTACCCGATGCTCGAC
      T_T_P_D_K_K_H_Q_K_E_P_P_F_L_W_M_G_Y_E_L_

2521  CACCCCGATAAGTGGACCGTGCAGCCCATCATGCTGCCCGATAAGGAGAGCTGGACCGTG
      -----+-----+-----+-----+-----+
      GTGGGGCTATTACCTGGCAGTCGGGTAGTACGACGGGCTATTCTCTCGACCTGGCAC
      H_P_D_K_W_T_V_Q_P_I_M_L_P_D_K_E_S_W_T_V_

```

FIG. 16E

PF1MI  
 AACGACATCCAGAACTGGTGGGCAAGCTGAATTGGGCCAGCCAAATCTACGCCGGCATT  
 2581 -----+-----+-----+-----+-----+-----+-----+  
 TTGCTGTAGGTCTTTGACCACCCGTTTCGACTTAACCCGGTCGGTTTAGATGCCGGCCGTAA  
 N\_D\_I\_Q\_K\_L\_V\_G\_K\_L\_N\_W\_A\_S\_Q\_I\_Y\_A\_G\_I\_

NarI  
 KasI  
 HaeII  
 PvuII  
 AAAGTGAAGCAGCTGTGCAGGCTGCTGAGAGGCGCCAAAGCCCTGACAGACATCGTGACA  
 2641 -----+-----+-----+-----+-----+-----+-----+  
 TTTCACTTCGTCGACACGTCCGACGACTCTCCGCGGTTTCGGGACTGTCTGTAGCACTGT  
 K\_V\_K\_Q\_L\_C\_R\_L\_L\_R\_G\_A\_K\_A\_L\_T\_D\_I\_V\_T\_

CTGACAGAGGAGGCCGAGCTGGAGCTGGCCGAGAACAGGGAGATCCTGAAGGACCCCGTG  
 2701 -----+-----+-----+-----+-----+-----+-----+  
 GACTGTCTCCTCCGGCTCGACCTCGACCGGCTTTGTCCCTCTAGGACTTCTGGGGCAC  
 L\_T\_E\_E\_A\_E\_L\_E\_L\_A\_E\_N\_R\_E\_I\_L\_K\_D\_P\_V\_

BstNI                    HinfI                    BstNI  
 CACGGCGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATTGAGAAGCAGGGCCAG  
 2761 -----+-----+-----+-----+-----+-----+-----+  
 GTGCCGCACATGATGCTGGGGTCGTTCTGGACCACCGGCTCTAAGTCTTCGTCCCGGTC  
 H\_G\_V\_Y\_Y\_D\_P\_S\_K\_D\_L\_V\_A\_E\_I\_Q\_K\_Q\_G\_Q\_

BstNI  
 GACCAGTGGACCTACCAAATCTACCAGGAGCCTTTCAAGAACCTGAAAACCGGGAAGTAC  
 2821 -----+-----+-----+-----+-----+-----+-----+  
 CTGGTCACCTGGATGGTTTAGATGGTCCTCGGAAAGTTCTTGGACTTTTGGCCCTTCATG  
 D\_Q\_W\_T\_Y\_Q\_I\_Y\_Q\_E\_P\_F\_K\_N\_L\_K\_T\_G\_K\_Y\_

BstNI                    HaeII                    PvuII  
 GCCAGGAAGAGAAGCGCCACACCAACGATGTGAGGCAGCTGGCCGAAGTGGTGAGAAA  
 2881 -----+-----+-----+-----+-----+-----+-----+  
 CGGTCTTCTCTTCGCGGGTGTGGTTGCTACACTCCGTGACCGGCTTACCACGTCTTT  
 A\_R\_K\_R\_S\_A\_H\_T\_N\_D\_V\_R\_Q\_L\_A\_E\_V\_V\_Q\_K\_

GTGGCTATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAG  
 2941 -----+-----+-----+-----+-----+-----+-----+  
 CACCGATACCTCTCGTAGCACTAGACCCCGTTCTGGGGTTCAAGTTCGACGGGTAGGTC  
 V\_A\_M\_E\_S\_I\_V\_I\_W\_G\_K\_T\_P\_K\_F\_K\_L\_P\_I\_Q\_

BstNI                    HinfI  
 BsaI                    BstNI                    BstNI  
 AAGGAGACCTGGGAAACCTGGTGGATGGACTACTGGCAGGCCACCTGGATTCTGAGTGG  
 3001 -----+-----+-----+-----+-----+-----+-----+  
 TTCCTCTGGACCCTTTGGACCACCTACCTGATGACCGTCCGGTGGACCTAAGGACTCACC  
 K\_E\_T\_W\_E\_T\_W\_W\_M\_D\_Y\_W\_Q\_A\_T\_W\_I\_P\_E\_W\_

FIG. 16F

```

                                PvuII          BstNI
3061  GAGTTCGTGAACACCCCCCTCTGGTGAAGCTGTGGTATCAGCTGGAGAAGGACCCCATC
-----+-----+-----+-----+-----+-----+-----+
CTCAAGCACTTGTGGGGGGGAGACCACTTCGACACCATAGTCGACCTCTTCCTGGGGTAG
E_F_V_N_T_P_P_L_V_K_L_W_Y_Q_L_E_K_D_P_I_

      NarI
      KasI
      HaeII BsaI          BsaI
3121  CTGGGCGCCGAGACCTTCTACGTGGACGGAGCCGCAATAGAGAGACCAAGCTGGGCAAG
-----+-----+-----+-----+-----+-----+
GACCCGCGGCTCTGGAAGATGCACCTGCCTCGGCGTTATCTCTCTGGTTCGACCCGTTTC
L_G_A_E_T_F_Y_V_D_G_A_A_N_R_E_T_K_L_G_K_

3181  GCCGGTACGTGACCGACAGAGGCAGACAGAAAGTGGTGTCTCTGACCGAGACAACCAAC
-----+-----+-----+-----+-----+-----+
CGGCCGATGCACTGGCTGTCTCCGTCTGTCTTTCACCACAGAGACTGGCTCTGTTGGTTG
A_G_Y_V_T_D_R_G_R_Q_K_V_V_S_L_T_E_T_T_N_

                                BstXI          PstI
3241  CAGAAAACCGAGCTGCACGCCATCCTGCTGGCCCTGCAGGACAGCGGCAGCGAAGTGAAC
-----+-----+-----+-----+-----+-----+
GTCTTTTGGCTCGACGTGCGGTAGGACGACCGGGACGTCTGTGCGCCGTCGCTTCACTTG
Q_K_T_E_L_H_A_I_L_L_A_L_Q_D_S_G_S_E_V_N_

      HinfI          BstNI
3301  ATCGTGACCGACTCCCAGTACGCCCTGGGCATCATTGAGCCAGCCCGATAGAAGCGAG
-----+-----+-----+-----+-----+-----+
TAGCACTGGCTGAGGGTCATGCGGGACCCGTAGTAAGTCCGGGTCGGGCTATCTTCGCTC
I_V_T_D_S_Q_Y_A_L_G_I_I_Q_A_Q_P_D_R_S_E_

3361  AGCGAGCTGGTGAACCAGATCATCGAGAAGCTGATCGGCAAGGACAAAATCTACCTGAGC
-----+-----+-----+-----+-----+-----+
TCGCTCGACCACTTGGTCTAGTAGCTCTTCGACTAGCCGTTCTGTTTTAGATGGACTCG
S_E_L_V_N_Q_I_I_E_K_L_I_G_K_D_K_I_Y_L_S_

                                BspMI
3421  TGGGTGCCCCGCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGTCCAGC
-----+-----+-----+-----+-----+-----+
ACCCACGGGCGGGTGTCCCGTAGCCGCCGTTGCTCGTCCACCTGTTTCGACCACAGGTCG
W_V_P_A_H_K_G_I_G_G_N_E_Q_V_D_K_L_V_S_S_

                                BstNI
3481  GGCATCCGGAAAGTGCTGTTTCTGGACGGCATCGACAAGGCCAGGAGGACCACGAGAGA
-----+-----+-----+-----+-----+-----+
CCGTAGGCCTTTCACGACAAAGACCTGCCGTAGCTGTTCCGGGTCCTCCTGGTGCTCTCT
G_I_R_K_V_L_F_L_D_G_I_D_K_A_Q_E_D_H_E_R_

```

FIG. 16G

**BspMI**

3541 TACCACAGCAACTGGCGGACAATGGCCAGCGACTTCAACCTGCCTCCCATCGTGGCCAAG  
 -----+-----+-----+-----+-----+-----+-----+  
 ATGGTGTCTGTTGACCGCCTGTTACCGGTCTGCTGAAGTTGGACGGAGGGTAGCACCGGTTCC  
 Y\_H\_S\_N\_W\_R\_T\_M\_A\_S\_D\_F\_N\_L\_P\_P\_I\_V\_A\_K

**PvuII                      PvuII                      BstNI**

3601 GAGATCGTGGCCAGCTGCGATAAGTGTCAAGTGAAGGGCGAGGCCATGCACGGCCAGGTG  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCTAGCACCGGTGACGCTATTCACAGTCTGACTTCCCGTCCGGTACGTGCCGGTCCAC  
 E\_I\_V\_A\_S\_C\_D\_K\_C\_Q\_L\_K\_G\_E\_A\_M\_H\_G\_Q\_V

**PstI    BstNI                      PvuII                      BstNI                      HinfI**

3661 GACTGCAGCCCTGGCATCTGGCAGCTGGCCTGCACCCACCTGGAGGGCAAAGTGATTCTG  
 -----+-----+-----+-----+-----+-----+-----+  
 CTGACGTCGGGACCGTAGACCGTCTGACCGGACGTGGGTGGACCTCCCGTTTCACTAAGAC  
 D\_C\_S\_P\_G\_I\_W\_Q\_L\_A\_C\_T\_H\_L\_E\_G\_K\_V\_I\_L

**HinfI                      BsaI    BstNI**

3721 GTGGCCGTGCACGTGGCCAGCGGCTACATCGAGGCCGAAGTGATTCCCGCCGAGACCGGC  
 -----+-----+-----+-----+-----+-----+-----+  
 CACCGGCACGTGACCGGTGCGCGATGTAGTCCGGCTTCACTAAGGGCGGCTCTGGCCG  
 V\_A\_V\_H\_V\_A\_S\_G\_Y\_I\_E\_A\_E\_V\_I\_P\_A\_E\_T\_G

**BsaI**

3781 CAGGAGACCGCCTACTTCTGCTGAAGTGTGGCCGGCAGATGGCCCGTGAAGTGGTGCAC  
 -----+-----+-----+-----+-----+-----+-----+  
 GTCCTCTGGCGGATGAAGGACGACTTCGACCGGCCGTCTACCGGGCACTTTCACCACGTG  
 Q\_E\_T\_A\_Y\_F\_L\_L\_K\_L\_A\_G\_R\_W\_P\_V\_K\_V\_V\_H

3841 ACCGCCAACGGCAGCAACTTCACCTCTGCCGCCGTGAAGGCCGCCTGTTGGTGGGCCAAT  
 -----+-----+-----+-----+-----+-----+-----+  
 TGGCGGTTGCCGTCGTTGAAGTGGAGACGGCGGCACTTCCGGCGGACAACCAACCGGTTA  
 T\_A\_N\_G\_S\_N\_F\_T\_S\_A\_A\_V\_K\_A\_A\_C\_W\_W\_A\_N

**Pf1MI**

**BstNI**

3901 ATCCAGCAGGAGTTCGGCATCCCCTACAACCCTCAGAGCCAGGGCGTGGTGGCCAGCATG  
 -----+-----+-----+-----+-----+-----+-----+  
 TAGTCTGCTCCTCAAGCCGTAGGGGATGTTGGGAGTCTCGGTCCCGCACCAACCGGTCGTAC  
 I\_Q\_Q\_E\_F\_G\_I\_P\_Y\_N\_P\_Q\_S\_Q\_G\_V\_V\_A\_S\_M

**BstNI                      BstNI**

3961 AACAAAGGAGCTGAAGAAGATCATCGGCCAGGTGAGGGACCAGGCCGAGCACCTGAAAACA  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGTTCTCTGACTTCTTCTAGTAGCCGGTCCACTCCCTGGTCCGGCTCGTGGACTTTTGT  
 N\_K\_E\_L\_K\_K\_I\_I\_G\_Q\_V\_R\_D\_Q\_A\_E\_H\_L\_K\_T

4021 GCCGTGCAGATGGCCGTGTTTCATCCACAACCTCAAGCGGAAGGGCGGCATTGGCGGCTAC  
 -----+-----+-----+-----+-----+-----+-----+  
 CGGCACGTCTACCGGACAAGTAGGTGTTGAAGTTCGCCTTCCCGCCGTAACCGCCGATG  
 A\_V\_Q\_M\_A\_V\_F\_I\_H\_N\_F\_K\_R\_K\_G\_G\_I\_G\_G\_Y

**FIG. 16H**

**HaeII** **EcoRV** **PstI**  
 4081 AGCGCCGGAGAGCGGATCATCGACATCATCGCCACCGATATCCAGACCAAGGAACCTGCAG  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGCGGCTCTCGCCTAGTAGCTGTAGTAGCGGTGGCTATAGGTCTGGTTCTTGACGTC  
 S\_A\_G\_E\_R\_I\_I\_D\_I\_I\_A\_T\_D\_I\_Q\_T\_K\_E\_L\_Q\_

**HinFI**  
 4141 AAGCAGATCACCAAGATTCAAGACTTCAGAGTGTACTACCGGGACAGCAGGGACCCCATC  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCGTCTAGTGGTTCTAAGTCTTGAAGTCTCACATGATGGCCCTGTCTGCTCCCTGGGGTAG  
 K\_Q\_I\_T\_K\_I\_Q\_N\_F\_R\_V\_Y\_Y\_R\_D\_S\_R\_D\_P\_I\_

**NarI** **KaSI**  
**ApaI** **HaeII** **BstNI**  
 4201 TGGGAAGGGCCCTGCCAAGCTGCTGTGGAAGGGCGAAGGCGCCGTGGTGATCCAGGACAAC  
 -----+-----+-----+-----+-----+-----+-----+  
 ACCTTCCCGGGACGGTTTCGACGACACCTTCCCGCTTCCGCGGCACCACTAGGTCCTGTTG  
 W\_K\_G\_P\_A\_K\_L\_L\_W\_K\_G\_E\_G\_A\_V\_V\_I\_Q\_D\_N\_

**HinFI**  
 4261 AGCGACATCAAAGTGGTGCCCCGAGGAAGGCCAAGATTCTGCGGGACTACGGCAAACAG  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGCTGTAGTTTCACCACGGGGCCTCCTTCCGGTTCTAAGACGCCCTGATGCCGTTTGTG  
 S\_D\_I\_K\_V\_V\_P\_R\_R\_K\_A\_K\_I\_L\_R\_D\_Y\_G\_K\_Q\_

**BglII**  
 4321 ATGGCCGGCGATGACTGCGTGGCCGGCAGGCAGGATGAGGACAGATCTATGGGCGGCAAG  
 -----+-----+-----+-----+-----+-----+-----+  
 TACCGGCCGCTACTGACGCACCGGCCGTCCGTCTACTCCTGTCTAGATACCCGCCGTTG  
 M\_A\_G\_D\_D\_C\_V\_A\_G\_R\_Q\_D\_E\_D\_R\_S\_M\_G\_G\_K\_

4381 TGGTCCAAGGGCAGCATTGTGGGCTGGCCCCGAGATCCGGGAGAGAATGAGAAGAGCCCCT  
 -----+-----+-----+-----+-----+-----+-----+  
 ACCAGGTTCCCGTCGTAACACCCGACCGGGCTCTAGGCCCTCTTACTCTTCTCGGGGA  
 W\_S\_K\_G\_S\_I\_V\_G\_W\_P\_E\_I\_R\_E\_R\_M\_R\_R\_A\_P\_

**NarI** **NarI**  
**BstNI** **KaSI** **KaSI**  
**HaeII** **HaeII**  
 4441 GCCGCCGCTCCTGGAGTGGGCGCCGTGTCTCAGGATCTGGATAAGCACGGCGCCATCACC  
 -----+-----+-----+-----+-----+-----+-----+  
 CGGCGGCGAGGACCTCACCCGCGGCACAGATCCTAGACCTATTCGTGCCGCGGTAGTGG  
 A\_A\_A\_P\_G\_V\_G\_A\_V\_S\_Q\_D\_L\_D\_K\_H\_G\_A\_I\_T\_

**PvuII** **BstNI**  
 4501 AGCAGCAACATCAACAACCCAGCTGTGTGGCTGGAGGCCAGGAAGAGGAGGAAGTG  
 -----+-----+-----+-----+-----+-----+-----+  
 TCGTCGTTGTAGTTGTTGGGTCGACACACACCGACCTCCGGGTCCTTCTCCTCTCAC  
 S\_S\_N\_I\_N\_N\_P\_S\_C\_V\_W\_L\_E\_A\_Q\_E\_E\_E\_V\_

FIG. 16I

```

                                     BsaI  BstNI      BsaI      NarI
                                     KasI
                                     HaeII
4561  GGCTTCCCTGTGAGACCCCAGGTGCCCCTGAGACCCATGACCTACAAGGGCGCCTTCGAC
-----+-----+-----+-----+-----+-----+
      CCGAAGGGACACTCTGGGGTCCACGGGGACTCTGGGTACTGGATGTTCCCGCGGAAGCTG
      G_F_P_V_R_P_Q_V_P_L_R_P_M_T_Y_K_G_A_F_D_

                                     BstNI
4621  CTGAGCCACTTCCTGAAGGAGAAGGGCGGCCTGGACGGCCTGATCTACAGCCGGAAGCGG
-----+-----+-----+-----+-----+
      GACTCGGTGAAGGACTTCTCTTCCC GCCGGACCTGCCGGACTAGATGTCGGCCTTCGCC
      L_S_H_F_L_K_E_K_G_G_L_D_G_L_I_Y_S_R_K_R_

      BstNI      BstNI
4681  CAGGAGATCCTGGATCTGTGGGTGTACCACACCCAGGGCTACTTCCCCGACTGGCAGAAT
-----+-----+-----+-----+-----+
      GTCCTCTAGGACCTAGACACCCACATGGTGTGGGTCCCGATGAAGGGGCTGACCGTCTTA
      Q_E_I_L_D_L_W_V_Y_H_T_Q_G_Y_F_P_D_W_Q_N_

      BstNI  BstNI
4741  TACACCCCTGGCCCTGGAGTGCGGTATCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTG
-----+-----+-----+-----+-----+
      ATGTGGGGACCGGGACCTCACGCCATAGGGGACTGGAAGCCGACCACGAAGTTCGACCAC
      Y_T_P_G_P_G_V_R_Y_P_L_T_F_G_W_C_F_K_L_V_

4801  CCTATGGAGCCCGACGAAGTGGAGAAGGCCACAGAGGGCGAGAACAACAGCCTGCTGCAC
-----+-----+-----+-----+-----+
      GGATACCTCGGGCTGCTTCACTCTTCCGGTGTCTCCCGCTCTTGTGTCGGACGACGTG
      P_M_E_P_D_E_V_E_K_A_T_E_G_E_N_N_S_L_L_H_

4861  CCTATCTGCCAGCACGGCATGGACGATGAGGAGCGGGAAGTCTGATCTGGAAGTTCGAC
-----+-----+-----+-----+-----+
      GGATAGACGGTCGTGCCGTACCTGCTACTCCTCGCCCTTACGACTAGACCTTCAAGCTG
      P_I_C_Q_H_G_M_D_D_E_E_R_E_V_L_I_W_K_F_D_

                                     BstNI
4921  AGCAGGCTGGCCCTGAAGCACAGAGCCCAGGAAGTGCACCCAGAGTTCTACAAGGACTGC
-----+-----+-----+-----+-----+
      TCGTCCGACCGGGACTTCGTGTCTCGGGTCTTACGTGGGTCTCAAGATGTTCTGACG
      S_R_L_A_L_K_H_R_A_Q_E_L_H_P_E_F_Y_K_D_C_

      BclI      xbaI
4981  TGATGATCATAATAATCTAGAA
-----+-----+
      ACTACTAGTATTATTAGATCTT
      *
  
```

FIG. 16J

ES 2 396 915 T3

```

HindIII                                     BspMI
1  AAGCTTGCCGCCACCATGAGGGTGATGGAGATCCAGCGGAACTGCCAGCACCTGCTGAGA
   +-----+-----+-----+-----+-----+
TTCGAACGGCGGTGGTACTCCCACTACCTCTAGGTCGCCTTGACGGTCGTGGACGACTCT
   M_R_V_M_E_I_Q_R_N_C_Q_H_L_L_R_

                                     BstNI           PstI           BstEII
61  TGGGGCATCATGATCCTGGGCATGATTATCATCTGCAGCACCGCCGACAACCTGTGGGTG
   +-----+-----+-----+-----+-----+
ACCCCGTAGTACTAGGACCCGTAATAAGTAGACGTCGTGGCGGCTGTTGGACACCCAC
   W_G_I_M_I_L_G_M_I_I_I_C_S_T_A_D_N_L_W_V_

                                     BsaI
121 ACCGTGTA CTACGGCGTGCCTGTGTGGAGAGATGCCGAGACCACCTGTTCTGCGCCAGC
   +-----+-----+-----+-----+-----+
TGGCACATGATGCCGCACGGACACACCTCTCTACGGCTCTGGTGGGACAAGACGCGGTGCG
   T_V_Y_Y_G_V_P_V_W_R_D_A_E_T_T_L_F_C_A_S_

                                     StuI
181 GACGCCAAGGCCTACAGCACCGAGAAGCACAATGTGTGGGCCACCCACGCCTGCGTGCCT
   +-----+-----+-----+-----+-----+
CTGCGGTTCCGGATGTCGTGGCTCTTCGTGTTACACACCCGGTGGGTGCGGACGCACGGA
   D_A_K_A_Y_S_T_E_K_H_N_V_W_A_T_H_A_C_V_P_

                                     BstNI
241 ACCGATCCCAACCCTCAGGAGATCCCCCTGGACAACGTGACCGAGGAGTTCAACATGTGG
   +-----+-----+-----+-----+-----+
TGGCTAGGGTTGGGAGTCCCTCTAGGGGGACCTGTTGCACTGGCTCCTCAAGTTGTACACC
   T_D_P_N_P_Q_E_I_P_L_D_N_V_T_E_E_F_N_M_W_

301 AAGAACAACATGGTGGACCAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTG
   +-----+-----+-----+-----+-----+
TTCTTGTGTACCACCTGGTCTACGTGCTCCTGTAGTAGTCGGACACCCTGGTCTCGGAC
   K_N_N_M_V_D_Q_M_H_E_D_I_I_S_L_W_D_Q_S_L_

                                     PvuII           PstI
361 AAGCCCTGCGTGCAGCTGACCCCCTGTGCGTGACCCTGAACTGCAGCAACGCCAGAGTG
   +-----+-----+-----+-----+-----+
TTCGGGACGCACGTCGACTGGGGGGACACGCACTGGGACTTGACGTCGTTGCGGTCTCAC
   K_P_C_V_Q_L_T_P_L_C_V_T_L_N_C_S_N_A_R_V_

                                     PstI
421 AACGCCACCTTCAACTCCACCGAGGACAGGGAGGGCATGAAGAACTGCAGCTTCAACATG
   +-----+-----+-----+-----+-----+
TTGCGGTGGAAGTTGAGGTGGCTCCTGTCCCTCCCGTACTTCTTGACGTCGAAGTTGTAC
   N_A_T_F_N_S_T_E_D_R_E_G_M_K_N_C_S_F_N_M_

                                     BspMI
481 ACCACCGAGCTGCGGGATAAGAAGCAGCAGGTGTACAGCCTGTTCTACCGGCTGGACATC
   +-----+-----+-----+-----+-----+
TGGTGGCTCGACGCCCTATTCTTCGTGTCACATGTCGGACAAGATGGCCGACCTGTAG
   T_T_E_L_R_D_K_K_Q_Q_V_Y_S_L_F_Y_R_L_D_I_

```

FIG. 17A



HaeII

541 GAGAAGATCAACAGCAGCAACAACAACAGCGAGTACCGGCTGGTGAAGTGAATACCAGC  
 -----+-----+-----+-----+-----+-----+-----+  
 CTCTTCTAGTTGTGTCGTCGTTGTTGTTGTCGCTCATGGCCGACCACTTGACGTTATGGTCG  
 E\_K\_I\_N\_S\_S\_N\_N\_N\_S\_E\_Y\_R\_L\_V\_N\_C\_N\_T\_S\_

StuI  
 BstNI                      BstEII

601 GCCATCACCCAGGCTGCCCTAAGGTGACCTTCGAGCCCATCCCCATCCACTACTGCGCC  
 -----+-----+-----+-----+-----+-----+-----+  
 CGGTAGTGGGTCCGGACGGGATTCCACTGGAAGCTCGGGTAGGGGTAGGTGATGACGCGG  
 A\_I\_T\_Q\_A\_C\_P\_K\_V\_T\_F\_E\_P\_I\_P\_I\_H\_Y\_C\_A\_

661 CCTGCCGGCTTCGCCATCCTGAAGTGCAACGACACCGAGTTCAATGGCACCGGCCCTGC  
 -----+-----+-----+-----+-----+-----+-----+  
 GGACGGCCGAAGCGGTAGGACTTCACGTTGCTGTGGCTCAAGTTACCGTGGCCGGGGACG  
 P\_A\_G\_F\_A\_I\_L\_K\_C\_N\_D\_T\_E\_F\_N\_G\_T\_G\_P\_C\_

PvuII

721 AAGAATGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTG  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCTTACACTCGTGGCACGTCACGTGGGTGCCGTAGTTCGGGCACCACAGGTGGGTGCAC  
 K\_N\_V\_S\_T\_V\_Q\_C\_T\_H\_G\_I\_K\_P\_V\_V\_S\_T\_Q\_L\_

BstNI

781 CTGCTGAACGGCAGCCTGGCCGAGAGAGAAGTGCGGATCAGGAGCGAGAACATCGCCAAC  
 -----+-----+-----+-----+-----+-----+-----+  
 GACGACTTGCCGTCCGACCGGCTCTCTCTTACGCCTAGTCCTCGCTCTTGTAGCGGTTG  
 L\_L\_N\_G\_S\_L\_A\_E\_R\_E\_V\_R\_I\_R\_S\_E\_N\_I\_A\_N\_

841 AACGCCAAGAACATCATCGTGCAGTTCGCCAGCCCCGTGAAGATCAACTGCATCCGGCCC  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGCGGTTCTTGTAGTAGCACGTCAAGCGGTCGGGGCACTTCTAGTTGACGTAGGCCGGG  
 N\_A\_K\_N\_I\_I\_V\_Q\_F\_A\_S\_P\_V\_K\_I\_N\_C\_I\_R\_P\_

HinfI      BstNI

901 AACAAATAACCCGGAAGAGCTACAGAATCGGCCCTGGCCAGACCTTCTACGCCACCGAC  
 -----+-----+-----+-----+-----+-----+-----+  
 TTGTTGTTATGGGCCTTCTCGATGTCTTAGCCGGGACCGGTCTGGAAGATGCGGTGGCTG  
 N\_N\_N\_T\_R\_K\_S\_Y\_R\_I\_G\_P\_G\_Q\_T\_F\_Y\_A\_T\_D\_

BstNI

961 ATTGTGGGCGACATCAGACAGGCCCACTGCAACGTGTCCAGGACCGACTGGAACAACACC  
 -----+-----+-----+-----+-----+-----+-----+  
 TAACACCCGCTGTAGTCTGTCCGGGTGACGTTGCACAGGTCCTGGCTGACCTTGTGTGG  
 I\_V\_G\_D\_I\_R\_Q\_A\_H\_C\_N\_V\_S\_R\_T\_D\_W\_N\_N\_T\_

PvuII      ScaI

1021 CTGAGACTGGTGGCCAACCAGCTGCGGAAGTACTTCAGCAACAAGACCATCATCTTCACC  
 -----+-----+-----+-----+-----+-----+-----+  
 GACTCTGACCACCGGTTGGTCGACGCCTTCATGAAGTCGTTGTTCTGGTAGTAGAAGTGG  
 L\_R\_L\_V\_A\_N\_Q\_L\_R\_K\_Y\_F\_S\_N\_K\_T\_I\_I\_F\_T\_

**FIG. 17B**

```

          BstNI
        BsaI
1081 AACAGCAGCGCGGAGACCTGGAGATCACCACCCACAGCTTCAATTGTGGCGGCGAGTTC
-----+-----+-----+-----+-----+
      TTGTCGTCGCCGCTCTGGACCTCTAGTGGTGGGTGTGGAAGTTAACACCGCGCTCAAG
      N_S_S_G_G_D_L_E_I_T_T_H_S_F_N_C_G_G_E_F_

          BstNI
          Hinfi
1141 TTCTACTGCAACACCTCCGGCCTGTTCAATAGCACCTGGACCACCAACAACATGCAGGAG
-----+-----+-----+-----+-----+
      AAGATGACGTTGTGGAGGCCGGACAAGTTATCGTGGACCTGGTGGTTGTTGTACGTCCTC
      F_Y_C_N_T_S_G_L_F_N_S_T_W_T_T_N_N_M_Q_E_

1201 TCCAACGACACCAGCAACGGCACCATCACCTGCCCTGCCGATCAAGCAGATCATCCGG
-----+-----+-----+-----+-----+
      AGGTTGCTGTGGTCTGTGCCGTGGTAGTGGGACGGGACGGCCTAGTTCGTCTAGTAGGCC
      S_N_D_T_S_N_G_T_I_T_L_P_C_R_I_K_Q_I_I_R_

          BstNI
          Hinfi
1261 ATGTGGCAGCGCGTGGGCCAGGCCATGTACGCCCTCCCATCGAGGGCGTGATTGCTGCTGC
-----+-----+-----+-----+-----+
      TACACCGTCGCGCACCCGGTCCGGTACATGCGGGGAGGGTAGCTCCCGCACTAAGCGACG
      M_W_Q_R_V_G_Q_A_M_Y_A_P_P_I_E_G_V_I_R_C_

1321 GAGAGCAACATCACCGCCTGATCCTGACCAGAGATGGCGGCAACAACAATTCCGCCAAC
-----+-----+-----+-----+-----+
      CTCTCGTTGTAGTGGCCGACTAGGACTGGTCTCTACCGCGTTGTTGTTAAGGCGGTTG
      E_S_N_I_T_G_L_I_L_T_R_D_G_G_N_N_N_S_A_N_

        BsaI      BstNI      EcoRV
1381 GAGACCTTCAGACCTGGCGGCGGAGATATCCGGGACAACCTGGCGGAGCGAGCTGTACAAG
-----+-----+-----+-----+-----+
      CTCTGGAAGTCTGGACCGCCGCTCTATAGGCCCTGTTGACCGCCTCGCTCGACATGTTCT
      E_T_F_R_P_G_G_G_D_I_R_D_N_W_R_S_E_L_Y_K_

          BstNI
1441 TACAAGGTGGTGAAGATCGAGCCCCTGGGCGTGGCCCCACCAGAGCCAAGAGAAGAGTG
-----+-----+-----+-----+-----+
      ATGTTCCACCACTTCTAGCTCGGGGACCCGCACCGGGGGTGGTCTCGGTTCTCTCTCAC
      Y_K_V_V_K_I_E_P_L_G_V_A_P_T_R_A_K_R_R_V_

          NarI
          KasI
          HaeII
          BstNI
1501 GTGGAGCGGGAGAAGAGAGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
-----+-----+-----+-----+-----+
      CACCTCGCCCTTCTCTCGGCACCCGTAGCCGCGGCACAAAGACCCGAAGGACCCCTCGG
      V_E_R_E_K_R_A_V_G_I_G_A_V_F_L_G_F_L_G_A_

```

FIG. 17C

PvuII

1561 GCCGGATCTACAATGGGAGCCGCCAGCATCACCCCTGACCGTGCAGGCCAGACAGCTGCTG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 CGGCCTAGATGTTACCCTCGGCGGTCGTAGTGGGACTGGCACGTCCGGTCTGTCTGACGAC  
 A\_G\_S\_T\_M\_G\_A\_A\_S\_I\_T\_L\_T\_V\_Q\_A\_R\_Q\_L\_L

PvuII

1621 AGCGGCATCGTG CAGCAGCAGAGCAATCTGCTGAGAGCCATCGAGGCCAGCAGCAGCTG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TCGCCGTAGCAGCTCGTCGTCTCGTTAGACGACTCTCGGTAGCTCCGGGTCGTCTGTCGAC  
 S\_G\_I\_V\_Q\_Q\_Q\_S\_N\_L\_L\_R\_A\_I\_E\_A\_Q\_Q\_Q\_L

PvuII      BstXI  
 PstI      BstNI

1681 CTGAAGCTGACAGTGTGGGCATCAAGCAGCTGCAGGCCAGGGTGTGGCCGTGGAGAGA  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 GACTTCGACTGTACACCCCGTAGTTCGTGCGACTCCGGTCCCACGACCGGCACCTCTCT  
 L\_K\_L\_T\_V\_W\_G\_I\_K\_Q\_L\_Q\_A\_R\_V\_L\_A\_V\_E\_R

BstNI      PstI

1741 TACCTGAGGGACCAGCAGCTCCTGGGCATCTGGGGCTGCAGCGCAAGCTGATCTGCACC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ATGGACTCCCTGGTCTGTCGAGGACCCGTAGACCCCGACGTCGCCGTTGACTAGACGTGG  
 Y\_L\_R\_D\_Q\_Q\_L\_L\_G\_I\_W\_G\_C\_S\_G\_K\_L\_I\_C\_T

BstNI      PvuII

1801 ACCAACGTGCCCTGGAATAGCAGCTGGAGCAACAAGAGCTACGACGACATCTGGCAGAAC  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TGGTTGCACGGACCTTATCGTCGACCTCGTTGTTCTCGATGCTGCTGTAGACCGTCTTG  
 T\_N\_V\_P\_W\_N\_S\_S\_W\_S\_N\_K\_S\_Y\_D\_D\_I\_W\_Q\_N

PstI

BstNI

1861 ATGACCTGGCTGCAGTGGGACAAGGAGATCAGCAACTACACCGACATCATCTACAGCCTG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TACTGGACCGACGTCAACCCTGTTCTCTAGTCGTTGATGTGGCTGTAGTAGATGTCCGGAC  
 M\_T\_W\_L\_Q\_W\_D\_K\_E\_I\_S\_N\_Y\_T\_D\_I\_I\_Y\_S\_L

BstNI

1921 ATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGCAGGATCTGCTGGCCCTGGACAAG  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 TAGCTCCTCTCGGTCCTGGTCCTCTTCTTGTCTGCTCCTAGACGACCGGGACCTGTTT  
 I\_E\_E\_S\_Q\_N\_Q\_Q\_E\_K\_N\_E\_Q\_D\_L\_L\_A\_L\_D\_K

PflMI      BglII

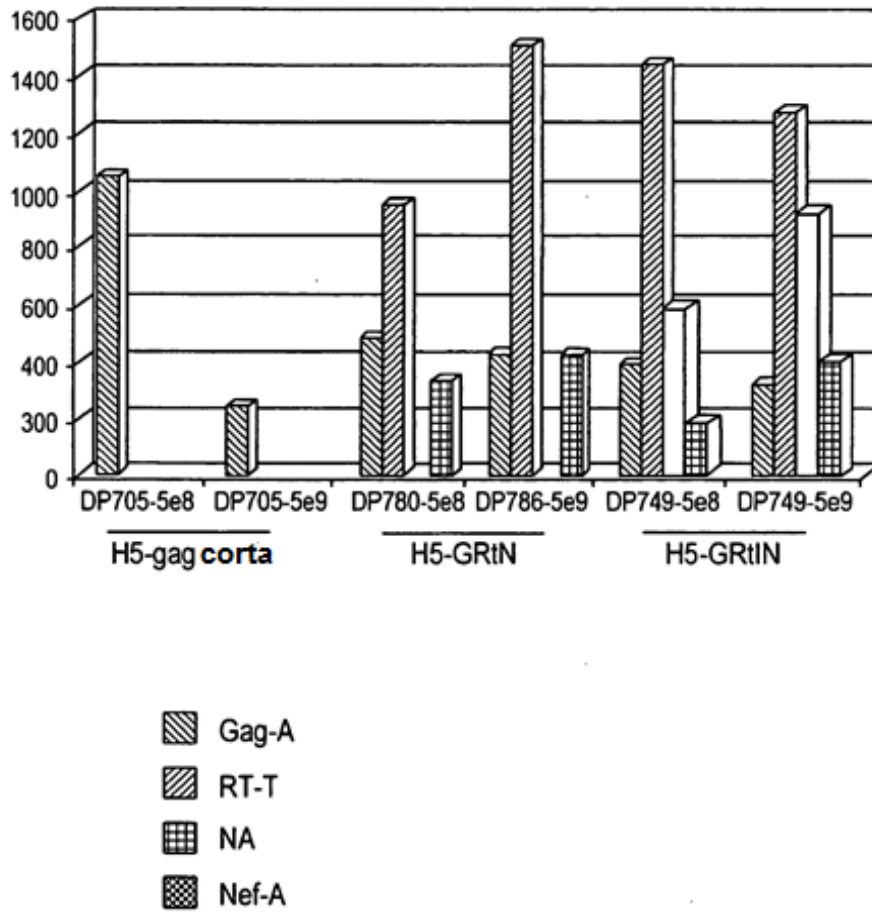
1981 TGGGCCAACCTGTGGAAGCTGGTTCGACATCAGCAAGTGGCTGTGGTACATCAGATCTTGA  
 -----+-----+-----+-----+-----+-----+-----+-----+  
 ACCCGGTTGGACACCTTGACCAAGCTGTAGTCGTTCCACCGACACCATGTAGTCTAGA  
 W\_A\_N\_L\_W\_N\_W\_F\_D\_I\_S\_K\_W\_L\_W\_Y\_I\_R\_S\_\*

XbaI

2041 TAATCTAGAA  
 -----+  
 ATTAGATCTT

*FIG. 17D*

**IAVA – Evaluación de los constructos AdH5  
clonados en CB6F1; DÍA 10 – Elispot**



**FIG. 18**

FCC/05.002: Media de las respuestas a las mezclas de la biblioteca de péptidos del VIH del clado A por IFN-gamma Elispot 10 días post-primovacunación

Cada barra representa la media de 4 u 8 muestras (cada una analizada por triplicado) +/- 1 Error Estándar

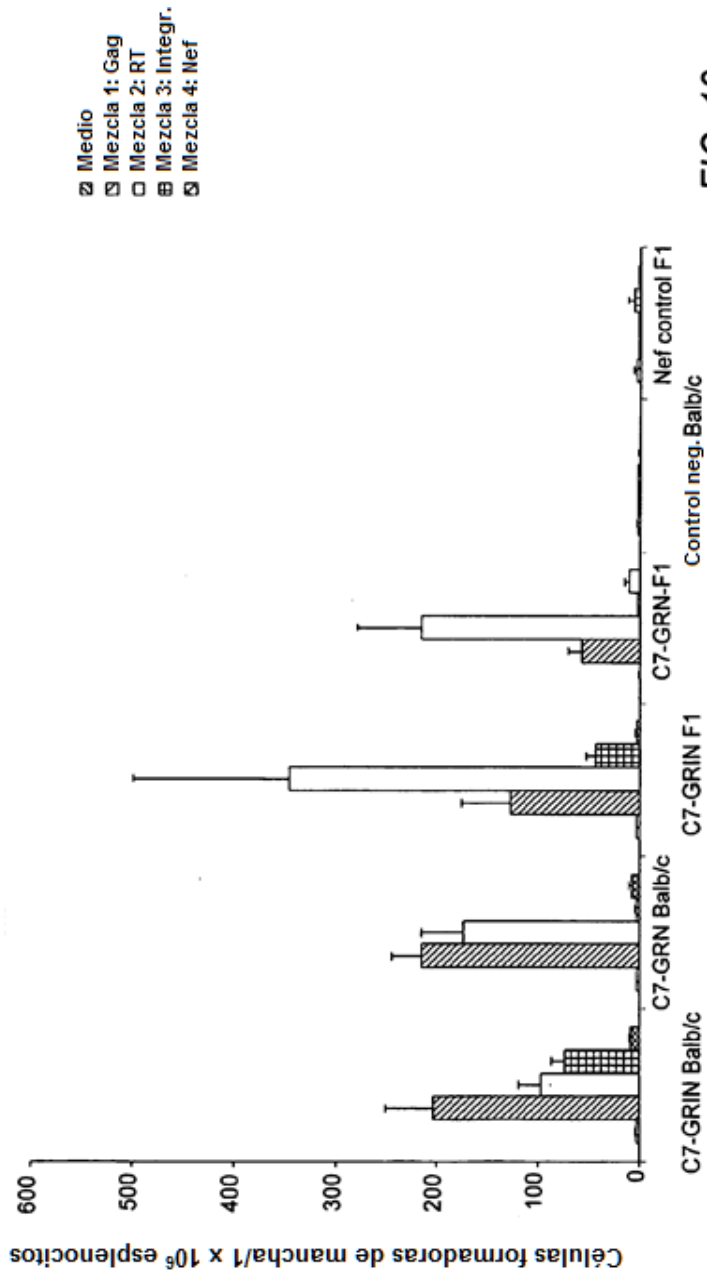


FIG. 19

FCC/02.005: Media de las respuestas a las mezclas de la biblioteca de péptidos del VIH del clado A por IL-2 Elispot 10 días post-primovacunación

Cada barra representa la media de 4 u 8 muestras (cada una analizada por triplicado) +/- 1 Error Estándar

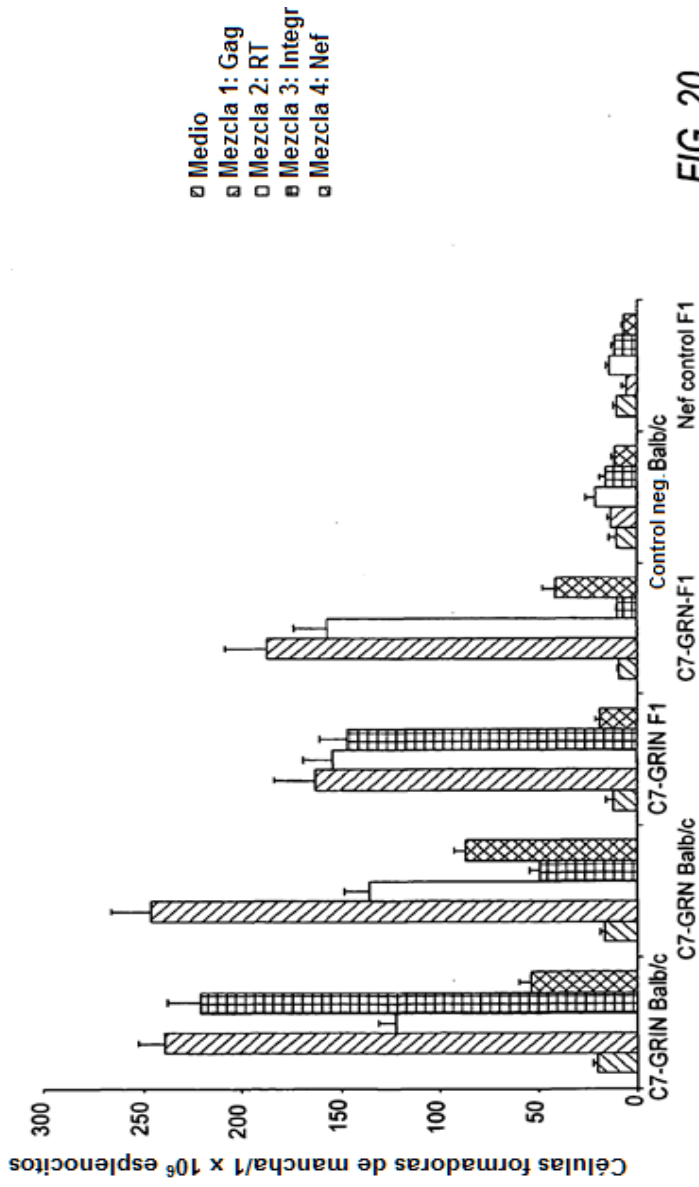


FIG. 20

FCC/05.003: Media de las respuestas a las mezclas de la biblioteca de péptidos del VIH del clado A por IFN-gamma Elispot 10 días post-primovacunación

Cada barra representa la media de 8 muestras (cada una analizada por triplicado) +/- 1 Error Estándar

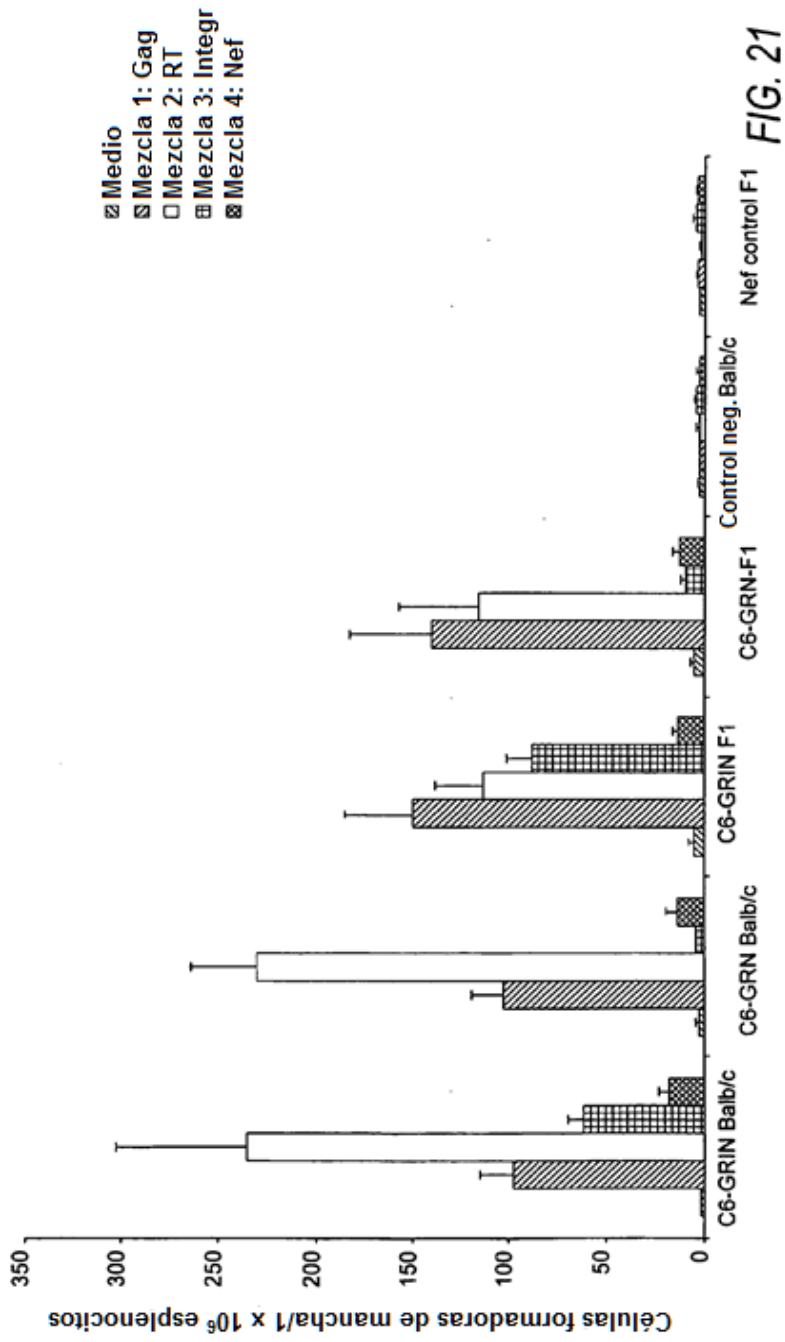


FIG. 21

FCC/05.003: Media de las respuestas a las mezclas de la biblioteca de péptidos del VIH del clado A por IL-2 Elispot 10 días post-primovacunación

Cada barra representa la media de 4 muestras (cada una analizada por triplicado) +/- Error Estándar

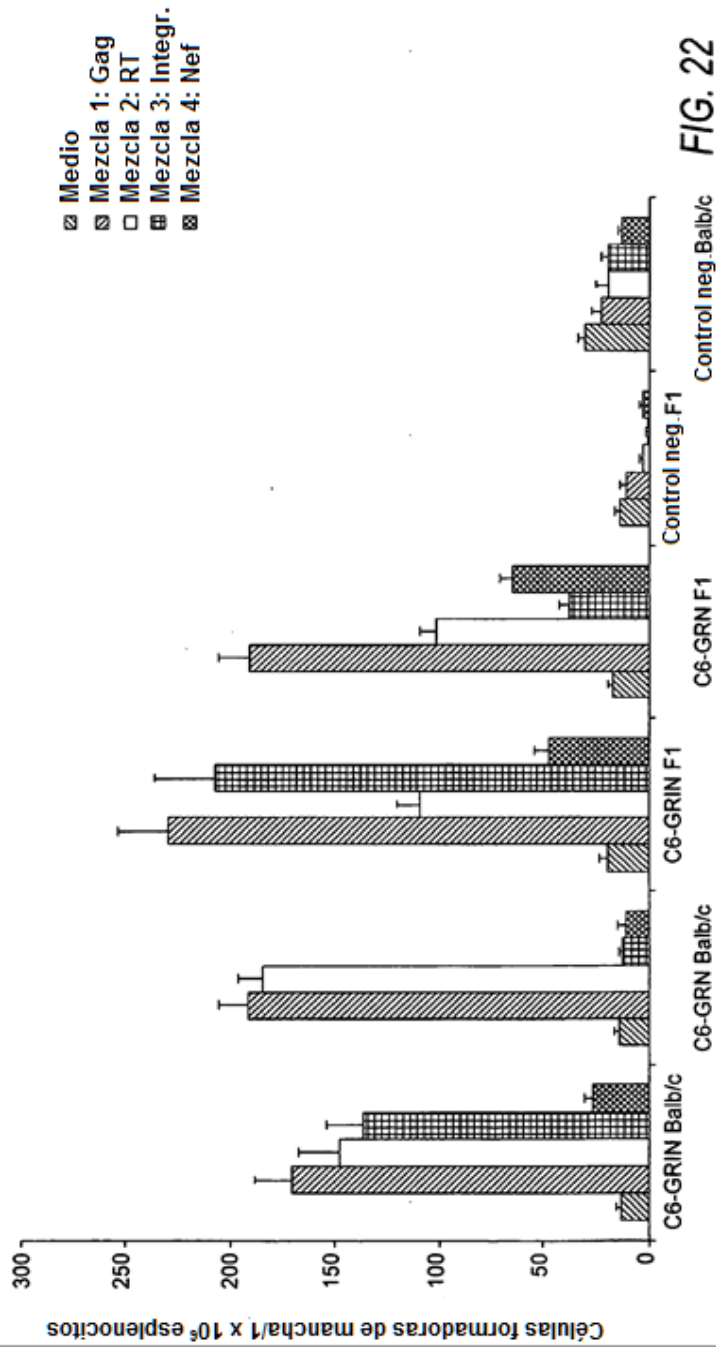


FIG. 22



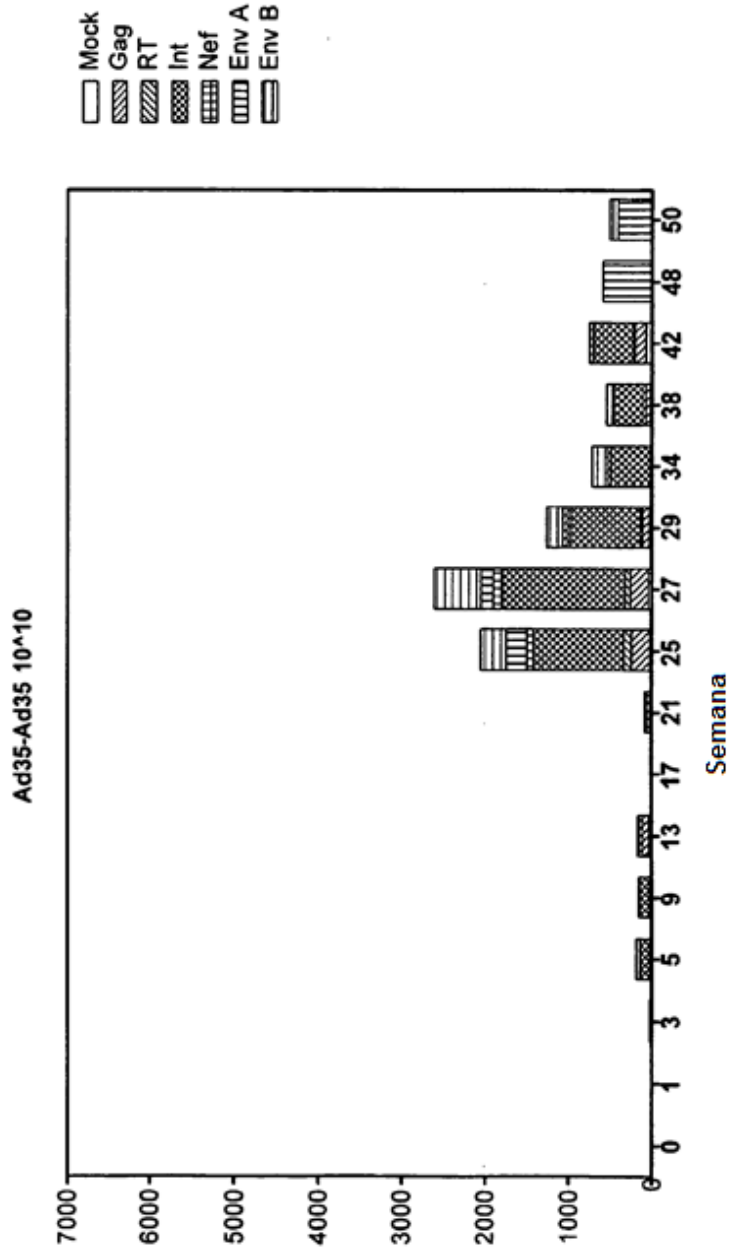


FIG. 23A