

19

OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 525 135**

21 Número de solicitud: 201330705

51 Int. Cl.:

**C12N 9/12** (2006.01)**C12Q 1/68** (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

**17.05.2013**

43 Fecha de publicación de la solicitud:

**17.12.2014**

56 Se remite a la solicitud internacional:

**PCT/ES2014/070389**

71 Solicitantes:

**CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTÍFICAS (CSIC) (100.0%)  
SERRANO, 117  
28006 MADRID ES**

72 Inventor/es:

**MENÉNDEZ ARIAS, Luis;  
MATAMOROS GRANDE, Tania;  
ABIA HOLGADO, David y  
BARRIOLUENGO FERNÁNDEZ, Verónica**

74 Agente/Representante:

**PONS ARIÑO, Ángel**54 Título: **RETROTRANSCRIPTASAS DEL VIH TIPO 1 GRUPO O, ACTIVAS A TEMPERATURAS  
ELEVADAS**

57 Resumen:

Retrotranscriptasas del VIH tipo 1 grupo O, activas a temperaturas elevadas.

La presente invención se encuadra dentro del campo de la biotecnología. Más concretamente, se refiere a retrotranscriptasas expresadas y purificadas en bacterias y que tienen la secuencia de aminoácidos de la retrotranscriptasa de un virus de la inmunodeficiencia humana de tipo 1 (VIH-1) del grupo O modificada en las posiciones 358, 359 y 360; y de variantes de esta enzima que contienen cambios adicionales en las posiciones 355 y 357 o en la 478 o en la posición 69 (en este caso acompañada de una inserción de dos aminoácidos). Estas polimerasas presentan mayor actividad que la enzima no mutada, a temperaturas elevadas (superiores a 60°C). Además retienen capacidad de síntesis de ADN a temperaturas superiores a 70°C. Por otro lado, la fidelidad de copia de estas enzimas no difiere significativamente de la presentada por la retrotranscriptasa no mutada.

ES 2 525 135 A1

**DESCRIPCIÓN****RETROTRANSCRIPTASAS DEL VIH TIPO 1 GRUPO O, ACTIVAS A TEMPERATURAS ELEVADAS**

5

**Sector de la técnica**

La presente invención se encuadra dentro del campo de la biotecnología. Más concretamente, se refiere a retrotranscriptasas expresadas y purificadas en bacterias y que  
10 tienen la secuencia de aminoácidos de la retrotranscriptasa de un virus de la inmunodeficiencia humana de tipo 1 (VIH-1) del grupo O modificada en varias posiciones. Estas polimerasas presentan mayor actividad que la enzima no mutada, a temperaturas elevadas (superiores a 60°C). Además retienen capacidad de síntesis de ADN a temperaturas superiores a 70°C. Por otro lado, la fidelidad de copia de estas enzimas no  
15 difiere significativamente de la presentada por la retrotranscriptasa no mutada.

**Estado de la técnica**

En retrovirus, la retrotranscriptasa (RT) (o transcriptasa inversa) es la enzima responsable  
20 de la replicación del genoma viral. La RT convierte el ARN genómico monocatenario en ADN bicatenario capaz de integrarse en el genoma de la célula hospedadora [revisado por Le Grice. *J Biol Chem*, 2012; 287: 40850-40857]. Se trata de una polimerasa capaz de sintetizar ADN, utilizando como molde ARN ó ADN, indistintamente. Además, la RT posee actividad endonucleasa (ribonucleasa H), que le permite degradar el molde ARN durante el  
25 proceso de síntesis de ADN dependiente de ARN.

Las RTs de retrovirus son enzimas útiles para la obtención de ADN complementario (ADNc) a partir de ARN mensajero (ARNm) o microARNs (miARN), que una vez amplificado por técnicas convencionales (PCR, por ejemplo) puede utilizarse para detectar expresión génica  
30 en organismos o tejidos. La eficacia de la retrotranscripción es importante en muchas aplicaciones biotecnológicas. Por ejemplo, para la detección de ARNs mensajeros, para su cuantificación mediante PCR en tiempo real, para el análisis de la expresión génica utilizando “microarrays” y también en estudios de transcriptómica utilizando técnicas de secuenciación masiva. Sin embargo, la presencia de estructuras secundarias en el ARN puede disminuir la  
35 eficacia de estas técnicas. Disponer de RTs capaces de sintetizar ADN a temperaturas elevadas puede resultar útil para mejorar el rendimiento de los procesos de amplificación.

Desde un punto de vista metodológico las RTs más utilizadas comercialmente en reacciones de amplificación son las del virus de la mieloblastosis de aves (AMV), la del virus Moloney de la leucemia de ratón (MLV) (Coté y Roth. *Virus Res* 2008; 134: 186-202) y variantes de la RT del virus de la inmunodeficiencia humana tipo 1 (VIH-1) (revisado en Hizi y Herschhorn. *Virus Res* 2008; 134: 203-220). Las RTs de AMV y MLV son las más utilizadas en ensayos de RT-PCR, aunque la de AMV presenta una mayor estabilidad térmica a temperaturas entre 42 y 52°C (Gerard *et al. Nucleic Acids Res.* 2002; 30: 3118-3129). Existen variantes de la RT del MLV tales como AffinityScript (Agilent) o Super Script III (Invitrogen) que se comercializan como enzimas activas a temperaturas más elevadas (Arezi y Hogrefe. *Nucleic Acids Res.* 2009; 37: 473-481).

Estudios llevados a cabo con RTs del VIH-1 clasificadas como pertenecientes al grupo M (subtipo B) han demostrado que estas enzimas presentan mayor actividad y estabilidad térmica que la RT del MLV, si bien son superadas por una variante RT “wild-type” derivada de un VIH-1 filogenéticamente distinto y clasificado como de grupo O (Álvarez *et al. J Mol Biol* 2009; 392: 872-884; patente WO2010130864 (A1)). En ensayos de RT-PCR aplicados a la expresión de ARN mensajero de tubulina, la RT del MLV no dio lugar a amplificación a temperaturas superiores a 52°C, si bien las RTs “wild-type” de grupo O o de subtipo B lo hacían a temperaturas de hasta 64°C, aunque solo la primera de ellas amplificaba a 66-68°C (Álvarez *et al. J Mol Biol* 2009; 392: 872-884). Las RTs del VIH-1 son heterodímeros compuestos por dos subunidades, una de 560 aminoácidos (denominada p66) y otra de 440 aminoácidos (denominada p51). La secuencia de p51 es idéntica a la de los aminoácidos 1-440 de p66. Las RTs del VIH-1 del grupo O se caracterizan por presentar alrededor de un 20% de aminoácidos distintos cuando se comparan con secuencias prototipo “wild-type” de subtipo B (Quiñones-Mateu *et al. Virology* 1997; 236: 364-373). Se han caracterizado varias RTs de grupo O que presentan mayor fidelidad de copia que la RT “wild-type”, como por ejemplo las portadoras de los cambios V75I, K65R y K65R/V75I. Estas RTs tienen una estabilidad térmica y eficacia catalítica a temperaturas elevadas similar a la exhibida por la RT “wild-type” (Barrioluengo *et al. Biochem J* 2011; 436: 599-607; patente WO2012080541 (A1)).

30

### **Descripción breve de la invención**

La presente invención se refiere a retrotranscriptasas aisladas de un virus de la inmunodeficiencia humana de tipo 1 (VIH-1) de grupo O y modificadas en una o en más posiciones, que presentan una mayor termoestabilidad que la enzima original, manteniendo

35

su fidelidad de copia, así como a su uso para llevar a cabo la retrotranscripción, amplificación o secuenciación de un ácido nucleico molde.

Un primer objeto de la invención se refiere al polipéptido que codifica una proteína con actividad RT aislada de un VIH-1 de grupo O y que presenta una mayor estabilidad a altas temperaturas, y que posee mayor actividad que la enzima WT a temperaturas superiores a 75°C, manteniendo su fidelidad de copia, y caracterizado porque su secuencia de aminoácidos presenta una identidad de al menos un 50% con respecto a la secuencia parental SEQ ID NO 1, y porque comprende alteraciones en su secuencia de aminoácidos (como pueden ser, por ejemplo, sustituciones, deleciones y/o inserciones) en las siguientes posiciones:

- la posición homóloga a la posición 358 de dicha secuencia, que sustituye el aminoácido lisina (K) original por el aminoácido arginina (R) (mutación K358R).
- la posición homóloga a la posición 359 de dicha secuencia, que sustituye el aminoácido alanina (A) original por el aminoácido glicina (G) (mutación A359G).
- la posición homóloga a la posición 360 de dicha secuencia, que sustituye el aminoácido serina (S) original por el aminoácido alanina (A) (mutación S360A).

También se describen otras variantes de retrotranscriptasas que contienen cambios adicionales a la combinación K358R/A359G/S360A (común a todas ellas), y que mejoran la termoestabilidad de la RT WT, es decir su capacidad para sintetizar ADN a elevadas temperaturas utilizando ARN como molde. Estos cambios de aminoácido e inserciones pertenecen, a título ilustrativo y sin que limite el alcance de la invención, al siguiente grupo:

- a) La sustitución del aminoácido treonina (T) por la inserción de dos aminoácidos serina y una glicina (SSG) en la posición homóloga a la posición 69 de la SEQ ID NO 1 (mutación T69SSG).
- b) La sustitución del aminoácido treonina (T) por el aminoácido alanina (A) en la posición homóloga a la posición 355 de la SEQ ID NO 1 (mutación T355A).
- c) La sustitución del aminoácido glutamina (Q) por el aminoácido metionina (M) en la posición homóloga a la posición 357 de la SEQ ID NO 1 (mutación Q357M).
- d) La sustitución del ácido glutámico (E) por el aminoácido glutamina (Q) en la posición homóloga a la posición 478 de la SEQ ID NO 1 (E478Q).

En esta invención se incluyen como objeto particular de la invención varias retrotranscriptasas variantes del VIH-1 del grupo O que se han usado como punto de partida y que presentan las mutaciones anteriormente descritas, y cuyas secuencias polipeptídicas particulares se corresponden con la SEQ ID NO 3, SEQ ID NO 5, SEQ ID NO 7, y SEQ ID

NO 9; así como sus correspondientes secuencias nucleotídicas (SEQ ID NO 4, SEQ ID NO 6, SEQ ID NO 8, SEQ ID NO 10).

Otro objeto de la invención son los vectores que comprenden los polinucleótidos que codifican las retrotranscriptasas de la invención.

Otro objeto de la invención es la célula que comprende el polinucleótido que codifica la retrotranscriptasa de la invención. Preferentemente dicha célula es una bacteria, y más preferentemente aún es una *Escherichia coli*.

Otro objeto de la presente invención es el método de obtención del polipéptido de la invención, que comprende los siguientes pasos:

- 1) Introducir el vector de la invención en una célula hospedadora adecuada (célula hospedadora de la invención),
- 2) Cultivar la célula hospedadora de la invención en un medio adecuado, y,
- 3) Purificar el polipéptido de la invención con actividad RT.

Otro objeto de la invención se refiere al uso del polinucleótido de la invención para la obtención del polipéptido de la invención con actividad retrotranscriptasa.

Otro objeto de la invención se refiere al uso de la célula hospedadora de la invención para la obtención del polipéptido de la invención.

Otro objeto de la invención es el uso del polipéptido de la invención en métodos de retrotranscripción, amplificación y secuenciación de ácidos nucleicos. Preferentemente, el ARNm.

Otro objeto de la presente invención se refiere a un kit que comprende los elementos necesarios para llevar a cabo cualquiera de los métodos descritos en la presente descripción, y que, preferentemente, comprende:

- a) la RT de la invención, y
- b) al menos, un elemento de la lista que comprende:
  - i) un tampón,
  - ii) un cebador,
  - iii) una ADN polimerasa dependiente de ADN, y
  - iv) un nucleótido.

**Descripción detallada de la invención**

Las retrotranscriptasas (RT) de retrovirus son enzimas útiles, por ejemplo, para la obtención de ADN complementario a partir de ARN mensajero, que una vez amplificado puede ser clonado en un vector. Es deseable que las RTs que se utilicen con este fin presenten elevada estabilidad térmica, ya que en estas condiciones disminuirían los niveles de estructura secundaria en el ARN y mejoraría el proceso de amplificación. Sin embargo, las RTs son enzimas que carecen de actividad exonucleasa correctora de errores y por tanto, tienen una tasa de error relativamente alta, por lo que un aumento de su fidelidad sería deseable para poder ser empleadas con este fin.

A partir de la existencia de homología de secuencia entre las RTs de VIH-1 grupo M subtipo B (por ejemplo, BH10 o HXB2) y las RTs de grupo O, los inventores han construido un modelo molecular de esta última unida a un complejo molde-iniciador, basándose en la estructura cristalográfica depositada en el Protein Data Bank, en el fichero 1RTD (www.pdb.org). Esta estructura es la única disponible en la que una RT del VIH-1 (en este caso, de la cepa HXB2) aparece formando un complejo ternario con ADN de doble cadena (molde-iniciador) y un nucleótido entrante (dTTP, en este caso). Además se trata de una estructura completa y con buena resolución. A partir del modelado molecular basado en homología de secuencia, se identificaron posiciones clave en la interacción con el molde-iniciador en ambas enzimas. Esta información se utilizó para diseñar variantes de la RT de grupo O (RTO) con mayor afinidad por ácidos nucleicos. Así, los inventores han obtenido por mutagénesis dirigida cuatro nuevas variantes de la RT del VIH-1 grupo O, denominadas RTO\_3M, RTO\_5M, RTO\_T69SSG\_3M, RTO\_E478Q\_3M y que son portadoras de las mutaciones K358R/A359G/S360A, T355A/Q357M/K358R/A359G/S360A, T69SSG/K358R/A359G/S360A y K358R/A359G/S360A/E478Q, respectivamente.

Estas enzimas se han expresado en *Escherichia coli* y purificado para su posterior caracterización, siguiendo un procedimiento similar al descrito en WO2010130864. Se observa que en comparación con la RT "wild-type" de grupo O (RTO\_WT), las cuatro enzimas presentan mayor actividad ADN polimerasa a temperaturas superiores a 75°C en reacciones de amplificación de ARN mediante RT-PCR tanto en ensayos cualitativos (Figura 1), como cuantitativos (Figura 2). Los ensayos de fidelidad llevados a cabo con estas enzimas demuestran que su fidelidad de copia se mantiene con respecto a la RTO\_WT (Tablas 2 y 3; Figura 3).

La presente invención se refiere a RTs aisladas de un virus de la inmunodeficiencia humana de tipo 1 (VIH-1) de grupo O y modificadas en una o en más posiciones, que presentan una mayor termoestabilidad que la enzima WT, manteniendo su fidelidad de copia, así como a su uso para llevar a cabo la retrotranscripción, amplificación o secuenciación de un ácido nucleico molde.

El punto de partida fue una RT identificada en un paciente que no había sido tratado con fármacos antirretrovirales y que estaba infectado por un VIH-1 de grupo O (cepa ESP49) (Quiñones-Mateu *et al. Virology* 1997; 236: 364-373; Menéndez-Arias *et al. J. Biol. Chem.* 2001; 276: 27470-27479) (RTO\_WT con secuencia polipeptídica SEQ ID NO 1). La secuencia de nucleótidos de la subunidad p66 de esta RT se clonó en un vector de expresión para la obtención y purificación de la enzima que contenía sitios de restricción adecuados y una cola de histidinas en su extremo C-terminal seguida de un codón de terminación (Álvarez *et al. J Mol Biol* 2009; 392: 872-884) (con secuencia nucleotídica SEQ ID NO 2). A partir de esta construcción y tras un proceso de mutagénesis dirigida se han obtenido 4 variantes que presentan una mayor termoestabilidad en comparación con la enzima de partida (la RTO\_WT) y que mantienen su fidelidad de copia. Esta adaptación de las RTs es potencialmente útil para la amplificación de ARNs difíciles, es decir, aquellos que contengan estructuras secundarias y/o secuencias ricas en pares de bases G:C.

Un resumen de las secuencias de polinucleótidos y polipéptidos descritos en la presente invención se observa en la Tabla 1.

**Tabla 1.**

Nombre	Mutaciones	Polipéptido	Nucleótido
RTO_WT		SEQ ID NO 1	SEQ ID NO 2
RTO_3M	K358R/A359G/S360A	SEQ ID NO 3	SEQ ID NO 4
RTO_5M	T355A/Q357M/K358R/A359G/S360A	SEQ ID NO 5	SEQ ID NO 6
RTO_T69SSG_3M	T69SSG/K358R/A359G/S360A	SEQ ID NO 7	SEQ ID NO 8
RTO_E478Q_3M	K358R/A359G/S360A/E478Q	SEQ ID NO 9	SEQ ID NO 10
RTO_WT*		SEQ ID NO 11	SEQ ID NO 12
RTO_3M*	K358R/A359G/S360A	SEQ ID NO 13	SEQ ID NO 14
RTO_5M*	T355A/Q357M/K358R/A359G/S360A	SEQ ID NO 15	SEQ ID NO 16
RTO_T69SSG_3M*	T69SSG/K358R/A359G/S360A	SEQ ID NO 17	SEQ ID NO 18
RTO_E478Q_3M*	K358R/A359G/S360A/E478Q	SEQ ID NO 19	SEQ ID NO 20

Por tanto, un primer objeto de la invención se refiere al polipéptido que codifica una proteína con actividad RT aislada de un VIH-1 de grupo O y que presenta una mayor estabilidad a altas temperaturas, en adelante polipéptido de la invención, y que posee mayor actividad que la enzima WT a temperaturas superiores a 75°C, manteniendo su fidelidad de copia, y caracterizado porque su secuencia de aminoácidos presenta una identidad de al menos un 50% con respecto a la secuencia parental SEQ ID NO 1, y porque comprende alteraciones en su secuencia de aminoácidos (como pueden ser, por ejemplo, sustituciones, deleciones y/o inserciones) en las siguientes posiciones:

- la posición homóloga a la posición 358 de dicha secuencia, que sustituye el aminoácido lisina (K) original por el aminoácido arginina (R) (mutación K358R).
- la posición homóloga a la posición 359 de dicha secuencia, que sustituye el aminoácido alanina (A) original por el aminoácido glicina (G) (mutación A359G).
- la posición homóloga a la posición 360 de dicha secuencia, que sustituye el aminoácido serina (S) original por el aminoácido alanina (A) (mutación S360A).

Así, en un objeto preferido de la invención, las sustituciones presentes en el polipéptido de la invención son K358R, A359G y S360A (retrotranscriptasa RTO\_3M). En una realización particular de la invención, el polipéptido de la invención se corresponde con la SEQ ID NO 3.

La temperatura a la que la RT presenta su máxima actividad ADN polimerasa (dependiente de ARN o de ADN) se denomina temperatura óptima. Por encima de esa temperatura, la actividad catalítica decrece, en parte debido a la desnaturalización térmica de la RT. El término "termostabilidad" se refiere a la estabilidad mostrada por una RT cuando se encuentra sometida a una temperatura elevada, por ejemplo, típicamente a una temperatura de, al menos 50°C, preferiblemente de, al menos 63°C, más preferiblemente de, al menos, 68°C y aún más preferiblemente de, al menos 75°C.

La termostabilidad de la RT de la presente invención puede determinarse mediante distintos tipos de ensayos. Puede estimarse, por ejemplo, pero sin limitarse, analizando la cantidad del producto obtenido durante la síntesis de ADN utilizando como molde ARN mensajero, en una RT-PCR, que puede ser cualitativa o cuantitativa. Para ello, se realiza primero la reacción de retrotranscripción a una temperatura elevada, y después el ADN complementario obtenido se puede amplificar mediante PCR. La cantidad de producto obtenido en estas reacciones constituye una medida de la estabilidad de la RT, a la temperatura en la que se ha llevado a cabo la reacción de retrotranscripción. El análisis del producto en geles de agarosa nos permite hacer una valoración cualitativa de la eficacia de

la reacción. De la misma forma podemos determinar el rendimiento de la etapa de retrotranscripción mediante PCR en tiempo real determinando el valor de  $\Delta Ct$ , como  $\Delta Ct = Ct - Ct(\text{ref})$ , donde  $Ct$  es el ciclo en el que se observa una amplificación significativa y  $Ct(\text{ref})$  es la media de los valores de  $Ct$  obtenidos para una RT utilizada de referencia.

5

Una “reacción de síntesis a temperatura elevada”, tal y como se utiliza en la presente descripción, se refiere a una reacción y, más preferiblemente, una reacción de retrotranscripción (o transcripción inversa) que se realiza a una temperatura de al menos 50°C, preferiblemente de, al menos 63°C, más preferiblemente de, al menos, 68°C, más preferiblemente de, al menos 75°C, y aún más preferiblemente de, al menos 78°C.

10

Una RT con termostabilidad “incrementada” o “aumentada” se define como una RT que tiene un incremento o un aumento significativo (aplicando criterios estadísticos) en la termostabilidad, típicamente de, al menos, 1,5 veces, más preferiblemente de, al menos, 2 veces, y aún más preferiblemente de, al menos, 3 veces, y aún más preferiblemente de, al menos, 4 veces, con respecto a la RT con la que se efectúe la comparación.

15

El término “fidelidad de copia” se refiere a la exactitud del proceso de polimerización de ADN catalizado por la RT, en el que influye su capacidad para discriminar entre sustratos correctos e incorrectos ya sean nucleótidos o complejos molde-iniciador, durante la síntesis de ADN complementario a un ácido nucleico que sirve de molde.

20

La fidelidad de la RT de la presente invención puede analizarse mediante distintos tipos de ensayos, como por ejemplo, pero sin limitarnos, ensayos de fidelidad en cultivos celulares o ensayos de fidelidad “in vitro” (genéticos o bioquímicos).

25

En los ensayos bioquímicos se utiliza la RT purificada para la determinación de constantes cinéticas sobre un molde ARN o ADN, bajo unas condiciones determinadas (pH, concentración de sustrato, etc...). De esta manera se pueden obtener los parámetros cinéticos de la fidelidad de copia de ADN (dependiente de ARN o ADN) de la RT, tanto en estado estacionario como preestacionario. Los ensayos bioquímicos de incorporación errónea en el estado estacionario, se basan en la determinación de las constantes cinéticas ( $k_{cat}$  y  $K_m$ ) para la incorporación de nucleótidos en el extremo 3' de un cebador y proporcionan una estimación de la selectividad de la RT por el nucleótido. La determinación de los parámetros cinéticos se lleva a cabo midiendo la incorporación de nucleótidos en el extremo 3' del cebador, previamente marcado en su extremo 5' con  $[\gamma^{32}P]ATP$ , en presencia

30

35

- de distintas concentraciones de dNTP, una vez formado el complejo binario RT/molde-  
 cebador. Los productos resultantes son analizados mediante electroforesis en geles de  
 poliacrilamida. Los datos obtenidos se ajustan a la ecuación de Michaelis-Menten, y los  
 parámetros  $k_{cat}$  (tasa de incorporación) y  $K_m$  (constante de Michaelis-Menten) son  
 5 determinados para los nucleótidos correctos e incorrectos. La eficiencia de incorporación  
 errónea ( $f_{inc}$ ) se define como la relación entre la eficacia catalítica ( $k_{cat}/K_m$ ) obtenida para el  
 nucleótido incorrecto y la eficacia catalítica obtenida para el nucleótido correcto. Así, una  
 mayor fidelidad de la RT implica una menor eficiencia de incorporación errónea.
- 10 Para que se produzca la fijación de un error en el ADN naciente, no es suficiente la  
 incorporación de un nucleótido incorrecto, la RT debe ser capaz además, de extender el  
 extremo desapareado que se genera como consecuencia de esta incorporación errónea.  
 Esta medida de fidelidad se lleva a cabo mediante ensayos de extensión de extremos  
 desapareados. En estos ensayos se calculan las constantes cinéticas en estado  
 15 estacionario para la incorporación de un nucleótido correcto sobre dos tipos de complejo  
 molde-cebador: el complejo con el extremo 3' correctamente apareado y el mismo complejo  
 con el extremo 3' desapareado. La eficiencia de extensión de extremos desapareados ( $f_{ext}$ )  
 se define como la relación entre  $k_{cat}/K_m$  obtenida para la extensión del extremo desapareado  
 y la obtenida para la extensión del extremo correctamente apareado.
- 20 Los ensayos bioquímicos en el estado pre-estacionario examinan la capacidad de la RT para  
 unirse e incorporar el dNTP a tiempos muy cortos (como por ejemplo, del orden de milisegundos).  
 De esta forma, es posible el cálculo de la constante de afinidad ( $K_d$ ) para la interacción entre el  
 dNTP y el complejo binario RT/molde-cebador, y de la constante de polimerización ( $k_{pol}$ ). La  
 25 eficiencia de incorporación errónea y de extensión de extremos desapareados se determina a  
 partir de los valores de  $k_{pol}/K_d$  obtenidos para la incorporación de nucleótidos correctos o  
 incorrectos, o los obtenidos para la incorporación de nucleótidos correctos sobre complejos molde-  
 cebador que contienen un extremo 3'-OH apareado o desapareado.
- 30 Los ensayos genéticos más utilizados, denominados "forward mutation assays", se suelen llevar  
 a cabo usando como molde-cebador de la reacción de síntesis de ADN, el ADN bicatenario del  
 fago M13mp2, del cual se ha eliminado la secuencia correspondiente al gen *lacZ* de una de las  
 hebras. La reacción de síntesis de ADN se realiza en presencia de la RT y de concentraciones  
 altas de los dNTPs. Tras la transformación bacteriana con el producto de la reacción, los  
 35 mutantes se identifican como placas azules/blancas en medio de cultivo que contiene X-Gal (5-  
 bromo-4-cloro-3-indol- $\beta$ -galactopiranosido) e IPTG (isopropil- $\beta$ -tio-galactopiranosido). De este

modo, si en la reacción de síntesis de ADN no se producen errores, el resultado es una placa azul oscuro. Por el contrario, la introducción de uno o varios errores implica la pérdida parcial o total de la  $\alpha$ -complementación, lo que se traduce en placas azul claro o blancas. El ADN recuperado de estas placas puede ser secuenciado, para determinar exactamente el número, el tipo y la posición en el genoma de las mutaciones introducidas por la RT.

Una RT con una fidelidad de copia “incrementada” o “aumentada” se define como una RT que tiene un incremento o un aumento significativo (aplicando criterios estadísticos) en la fidelidad de copia con respecto a la RT no modificada, típicamente de, al menos, 1,5 veces, más preferiblemente de, al menos, 2 veces, y aún más preferiblemente de, al menos, 3 veces, y aún más preferiblemente de, al menos, 4 veces. Por ejemplo, en ensayos bioquímicos de incorporación de nucleótido, se considera un aumento de fidelidad cuando el valor obtenido para la RT modificada es significativamente superior al de la no modificada (aplicando criterios estadísticos), típicamente, de al menos, 1,5 veces, preferiblemente, de al menos, 2 veces, más preferiblemente de, al menos, 3 veces, y aún más preferiblemente de, al menos, 4 veces. Por ejemplo, en ensayos genéticos (“forward mutation assays”) se considera fidelidad incrementada cuando hay un aumento significativo (aplicando criterios estadísticos) en la frecuencia de mutante obtenida por la RT modificada, típicamente de, al menos, un 50% (1,5 veces), preferiblemente, de al menos, 2 veces, más preferiblemente de, al menos, 3 veces, y aún más preferiblemente de, al menos, 4 veces.

Junto a al RT portadora de los cambios K358R, A359G y S360A, la RTO\_3M, se describen otras variantes de la RT del VIH-1 grupo O que contienen cambios adicionales a la combinación K358R/A359G/S360A (común a todas ellas), y que mejoran la termoestabilidad de la RT WT, es decir su capacidad para sintetizar ADN a elevadas temperaturas utilizando ARN como molde. Estos cambios de aminoácido e inserciones pertenecen, a título ilustrativo y sin que limite el alcance de la invención, al siguiente grupo:

- e) La sustitución del aminoácido treonina (T) por la inserción de dos aminoácidos serina y una glicina (SSG) en la posición homóloga a la posición 69 de la SEQ ID NO 1 (mutación T69SSG).
- f) La sustitución del aminoácido treonina (T) por el aminoácido alanina (A) en la posición homóloga a la posición 355 de la SEQ ID NO 1 (mutación T355A).
- g) La sustitución del aminoácido glutamina (Q) por el aminoácido metionina (M) en la posición homóloga a la posición 357 de la SEQ ID NO 1 (mutación Q357M).
- h) La sustitución del ácido glutámico (E) por el aminoácido glutamina (Q) en la posición homóloga a la posición 478 de la SEQ ID NO 1 (E478Q).

Así, en otro objeto preferido de la invención, el polipéptido de la invención presenta los cambios de aminoácido K358R/A359G/S360A y además de las sustituciones T355A y Q357M (retrotranscriptasa RTO\_5M). En una realización particular, el polipéptido de la invención se corresponde con la SEQ ID NO 5.

5

En otro objeto preferido de la invención, el polipéptido de la invención presenta los cambios de aminoácido K358R/A359G/S360A y además de T69SSG (retrotranscriptasa RTO\_T69SSG\_3M). En una realización particular, el polipéptido de la invención se corresponde con la SEQ ID NO 7.

10

En otro objeto preferido de la invención, el polipéptido de la invención presenta los cambios de aminoácido K358R/A359G/S360A, y además la sustitución E478Q (retrotranscriptasa RTO\_ E478Q \_3M). En una realización particular, el polipéptido de la invención se corresponde con la SEQ ID NO 9.

15

Adicionalmente, el polipéptido de la invención puede hallarse flanqueado de pequeños fragmentos polipeptídicos cuya presencia es necesaria y/o ventajosa para la expresión del polipéptido en un vector adecuado y que son conocidos en el estado de la técnica. Dentro de éstos se encuentran tres aminoácidos en el extremo N-terminal (MNS, es decir Metionina-Asparagina-Serina), cuya secuencia se incluye para facilitar la iniciación de la traducción y acomodar sitios de reconocimiento de endonucleasas de restricción, útiles para la construcción de los vectores de expresión que contienen la retrotranscriptasa. También se incluyen aquí secuencias que permiten una mejora en la purificación del polipéptido, tales como una cola de residuos de histidinas en el extremo C-terminal, cuando ésta, preferiblemente, consiste en, al menos, 6 residuos de histidina. Estos residuos flanqueantes del polipéptido de la invención dan lugar a nuevas secuencias polipeptídicas que contienen el polipéptido de la invención y que mantienen su actividad retrotranscriptasa.

20

25

Así, en otro objeto preferido de la invención, las retrotranscriptasas RTO\_3M, RTO\_5M, RTO\_T69SSG\_3M y RTO\_E478Q\_3M presentan las secuencias flanqueantes MNS en el extremo N-terminal y una cola de histidinas en el C-terminal, dando lugar a los polipéptidos con actividad retrotranscriptasa: RTO\_3M\*, RTO\_5M\*, RTO\_T69SSG\_3M\* y RTO\_E478Q\_3M\*, que en un objeto particular se corresponden, respectivamente, con las secuencias polipeptídicas SEQ ID NO 13, SEQ ID NO 15, SEQ ID NO 17 y SEQ ID NO 19.

35

Las mutaciones aquí descritas introducidas en la secuencia polipeptídica "wild-type" de una RT aislada de un VIH-1 del grupo O pueden obtenerse mediante técnicas de ingeniería genética o ADN recombinante, como por ejemplo mutando la secuencia codificante de la RT mediante mutagénesis dirigida o al azar, o bien pueden obtenerse a partir de la síntesis química de la secuencia de nucleótidos que codifique para la subunidad p66 de la RT portadora de las mutaciones.

Tal y como se utiliza en la presente descripción, el término "mutación" se refiere a una sustitución de un aminoácido por otro aminoácido distinto. Los aminoácidos individuales en una secuencia se representan aquí como XN, en la que X es el aminoácido en la secuencia (designado mediante el código de una letra universalmente aceptado en la nomenclatura de aminoácidos), y N es la posición en la secuencia. Las mutaciones puntuales tipo sustitución en una secuencia de aminoácidos se representan aquí como X1NX2, en la que X1 es el aminoácido en la secuencia de proteína no mutada, X2 es el/los aminoácido/s nuevos de la secuencia de proteína mutada, y N es la posición en la secuencia de aminoácidos.

Además, con la información suministrada, un experto en la materia es capaz de combinar las mutaciones anteriormente descritas en la presente invención para generar nuevas variantes de la RT con actividad similar o mejorada a temperaturas elevadas. Una posibilidad es la sustitución conservativa de los aminoácidos en las posiciones anteriormente comentadas. Así, por ejemplo, se entiende por sustitución conservativa aquella que mantiene las características de polaridad y carga del aminoácido sustituido. Por ejemplo, lisina y arginina son aminoácidos cuyas cadenas laterales están cargadas positivamente a pH neutro, por lo que se acepta que los cambios de lisina por arginina o viceversa representan cambios conservativos. Se han clasificado los 20 aminoácidos que constituyen la base de todas las proteínas naturales de acuerdo a su conservatividad en grupos: (i) aminoácidos aromáticos (fenilalanina, tirosina, triptófano); (ii) aminoácidos alifáticos (glicina, alanina, valina, leucina, isoleucina y metionina); (iii) aminoácidos ionizables básicos (histidina, lisina y arginina); (iv) aminoácidos ionizables ácidos (ácidos aspártico y glutámico); (v) amidas de aminoácidos ácidos (asparagina y glutamina); y (vi) aminoácidos hidroxilados (serina y treonina). Algunos autores incluirían la cisteína en este último grupo.

Los polinucleótidos que codifican para los polipéptidos de la invención descritos en la invención corresponden a variantes de los mismos obtenidos mediante mutagénesis dirigida de la región codificante de la RT aislada de una cepa del VIH-1 de grupo O (SEQ ID NO 2).

Dichos polinucleótidos se corresponden con la secuencia de nucleótidos, que constituye la secuencia codificante del polipéptido de la invención, denominados de ahora en adelante polinucleótidos de la invención.

- 5 Los términos "polinucleótido", "secuencia nucleotídica", "secuencia de nucleótidos", "ácido nucleico" y "oligonucleótido" se usan aquí de manera intercambiable, y se refieren a una forma polimérica de nucleótidos de cualquier longitud, que pueden estar, o no, química o bioquímicamente modificados.
- 10 Así, un segundo objeto de la invención hace referencia a la secuencia de polinucleótidos que codifica el polipéptido de la invención, de ahora en adelante polinucleótido de la invención, y que presenta elevada mayor actividad que la enzima WT a temperaturas superiores a 68°C manteniendo la fidelidad de copia de la enzima original y caracterizada porque su secuencia presenta una identidad de al menos un 50% con respecto a la
- 15 secuencia parental SEQ ID NO 2, y porque el polipéptido que codifica presenta los cambios en su secuencia de aminoácido descritos anteriormente.

Así, un objeto preferido de la invención es el polinucleótido que codifica el polipéptido de la invención que presenta los cambios de aminoácido K358R, A359G y S360A

20 (retrotranscriptasa RTO\_3M). En una realización particular de la invención, el polinucleótido de la invención se corresponde con la SEQ ID NO 4.

Otro objeto preferido de la invención es el polinucleótido que codifica el polipéptido de la invención que presenta los cambios de aminoácido K358R, A359G y S360A y además las

25 sustituciones T355A y Q357M (retrotranscriptasa RTO\_5M). En una realización particular, el polinucleótido de la invención se corresponde con la SEQ ID NO 6.

Así, otro objeto preferido de la invención, es el polinucleótido que codifica el polipéptido de la invención que presenta los cambios de aminoácido K358R, A359G y S360A y además

30 T69SSG (retrotranscriptasa RTO\_T69SSG\_3M). En una realización particular, el polinucleótido de la invención se corresponde con la SEQ ID NO 8.

Otro objeto preferido de la invención es el polinucleótido que codifica el polipéptido de la invención que presenta los cambios de aminoácido K358R, A359G y S360A y además la

35 sustitución E478Q (retrotranscriptasa RTO\_E478Q\_3M). En una realización particular, el polinucleótido de la invención se corresponde con la SEQ ID NO 10.

Así, en otro objeto preferido de la invención, las retrotranscriptasas RTO\_3M, RTO\_5M, RTO\_T69SSG\_3M y RTO\_E478Q\_3M presentan las secuencias de nucleótidos que codifican para las secuencias polipeptídicas flanqueantes MNS en el extremo N-terminal y una cola de histidina en el C-terminal, dando lugar a los polinucleótidos que codifican las retrotranscriptasas RTO\_3M\*, RTO\_5M\*, RTO\_T69SSG\_3M\* y RTO\_E478Q\_3M\*, que en un objeto particular se corresponden, respectivamente, con las secuencias de nucleótidos SEQ ID NO 14, SEQ ID NO 16, SEQ ID NO 18 y SEQ ID NO 20.

Considerando que las RTs de distintas cepas y aislados circulantes del VIH puedan ser evolutivamente similares, es de esperar que la identidad global de los genes que las codifican sea igual o mayor de un 50%, y más concretamente al nivel de la secuencia polinucleotídica correspondiente a la SEQ ID NO 2 (la RTO\_WT), sea del 60% o mayor. Además, el grado de identidad o la homología existente entre las secuencias de aminoácidos de la(s) RT(s) objeto de la invención y las secuencias de otras RTs similares pueden determinarse por métodos conocidos en estado de la técnica. Por ejemplo, a través del alineamiento de la secuencias de aminoácido de la RT putativa y la correspondiente a la RTO\_3M de esta memoria.

El término “homología”, tal y como se utiliza en esta memoria, hace referencia a la semejanza entre dos estructuras debida a una ascendencia evolutiva común, y más concretamente a la semejanza o identidad entre los nucleótidos de posiciones equivalentes en dos o más polinucleótidos.

El término “identidad”, tal y como se utiliza en esta memoria, hace referencia a la proporción de nucleótidos idénticos entre dos polinucleótidos que se comparan. Los métodos de comparación de secuencias son conocidos en el estado de la técnica, e incluyen, aunque sin limitarse a ellos, el programa BLASTP o BLASTN, ClustalW y FASTA. Puesto que dos proteínas se consideran homólogas si tienen el mismo origen evolutivo o si tienen función y estructura similares, en general, se asume que valores superiores de similitud o identidad del 30% indican estructuras homólogas. Podemos considerar, por tanto, que porcentajes de identidad de, al menos, un 80% mantendrán las mismas propiedades de dicho polipéptido.

Los virus de la inmunodeficiencia humana (VIH-1 y VIH-2) son los agentes etiológicos del SIDA en el hombre. Pertenecen al género *Lentivirus* dentro de la familia Retroviridae (retrovirus), que poseen como una de sus principales características una enorme diversidad genética. El VIH-1 se ha clasificado en cuatro grupos: M, O, N y P. El primer aislado del VIH-

1 grupo O se obtuvo de pacientes infectados en 1987, y su secuencia de nucleótidos se publicó tres años más tarde (De Leys et al. J. Virol. 1990; 64: 1207-1216). Actualmente, se pueden obtener variantes del VIH-1 grupo O (por ejemplo, la cepa MVP5180/91) a través del NIH AIDS Research & Reference Reagent Program (<http://www.aidsreagent.org>) (Germantown, Maryland, EE.UU.). La información contenida en la presente invención, también permite para un experto en el estado de la técnica generar variantes de RT con mayor termoestabilidad que el enzima original, partiendo de la secuencia de RTs de diferentes cepas.

El término “aislado”, tal y como se utiliza en esta memoria, hace referencia a nucleótidos o péptidos que: 1) se encuentran sustancialmente libres de componentes que normalmente acompañan o interaccionan con él en la naturaleza, o 2) si se encuentran en su medio natural, han sido sintéticamente (no naturalmente) alterados por la intervención humana y/o introducidos en una célula que no los posee de forma nativa. Por ejemplo, un polinucleótido natural se convierte en “aislado” si ha sido alterado por medio de la intervención humana (por medio de, por ejemplo pero sin limitarnos, mutagénesis dirigida, añadiendo inserciones y/o deleciones, etc.). De la misma manera, un polinucleótido natural se convierte en “aislado” si se introduce por medios no naturales en un organismo no nativo a dicho polinucleótido (transfección). Por tanto, el término “aislado” en este último caso, es equivalente al término “heterólogo”.

Con la información suministrada en la presente invención un experto en la materia es capaz de identificar secuencias de nucleótidos homólogas a las descritas en la presente invención y que codifican para RTs con características idénticas a las descritas para la RT de la invención. Por tanto, el polinucleótido de la invención constituye la secuencia codificante de una variante de la RT aislada de un VIH-1 del grupo O con la actividad mejorada descrita, cuya secuencia de nucleótidos se corresponde a:

- a) moléculas de ácido nucleico de la secuencia polinucleotídica aislada o en su cadena complementaria,
- b) moléculas de ácido nucleico cuya cadena complementaria es capaz de hibridar con una secuencia polinucleotídica de (a), o
- c) moléculas de ácido nucleico cuya secuencia difiere de (a) y/o (b) debido a la degeneración del código genético.

El polinucleótido de la invención puede encontrarse aislado como tal o formando parte de los vectores que permiten la propagación de dichos polinucleótidos en células hospedadoras

adecuadas. Por lo tanto, en otro objeto, la invención se refiere a un vector, en adelante vector de la invención, que comprende el polinucleótido de la invención como se describe anteriormente.

- 5 El vector puede ser, por ejemplo un vector de clonación o un vector de expresión. Preferiblemente, dicho vector es un plásmido apropiado para la expresión y purificación de la RT de la invención.

10 El término "vector de clonación", tal y como se utiliza en la presente descripción, se refiere a una molécula de ADN en la que se puede integrar otro fragmento de ADN, sin que pierda la capacidad de autorreplicación. Ejemplos de vectores de expresión son, pero sin limitarse, plásmidos, cósmidos, fagos de ADN o cromosomas artificiales de levadura.

15 El término "vector de expresión", tal y como se utiliza en la presente descripción, se refiere a un vector de clonación adecuado para expresar un ácido nucleico que ha sido clonado en el mismo tras ser introducido en una célula, denominada célula huésped. Dicho ácido nucleico se encuentra, por lo general, unido operativamente a secuencias de control.

20 El término "expresión" se refiere al proceso por el cual se sintetiza un polipéptido a partir de un polinucleótido. Incluye la transcripción del polinucleótido en un ARN mensajero (ARNm) y la traducción de dicho ARNm en una proteína o un polipéptido.

25 El término "célula hospedadora" o "célula huésped", tal y como se utiliza en la presente descripción se refiere a cualquier organismo procarionta o eucariota que es recipiente de un vector de expresión, de clonación o de cualquier otra molécula de ADN.

Los vectores adecuados para la inserción del polinucleótido de la invención son plásmidos utilizados para la expresión de proteínas en procariotas tales, a título ilustrativo, como pUC18, pUC19, Bluescript y sus derivados, mp18, mp19, pBR322, pMB9, Co1E1, pCR1, 30 RP4, fagos y vectores "lanzadera", tales como pSA3 y pAT28; vectores de expresión en levaduras tales como el plásmido de 2 micras de *Saccharomyces cerevisiae*), plásmidos de integración, vectores YEP, plásmidos centromeros y similares; vectores de expresión en células de insectos tales como los vectores de la serie pAC y de la pVL vectores de expresión; vectores de expresión en células de plantas tales como piBi, pEarleyGate, PAVA, 35 pCAMBIA, PGSA, PGWB, PMDC, PMY, serie de poros y similares, y otros plásmidos de expresión de proteínas utilizados en células eucariotas, incluyendo baculovirus adecuados

para la transfección de células de insecto usando cualquier sistema disponible en el estado de la técnica.

5 Tal y como se utiliza en la presente memoria, una “célula hospedadora” o “célula huésped” incluye cualquier célula cultivable que puede ser modificada mediante la introducción de ADN no contenido de manera natural en la célula, de aquí en adelante célula hospedadora de la invención.

10 Preferiblemente, una célula hospedadora es aquélla en la que el polinucleótido de la invención puede ser expresado, dando lugar a un polipéptido estable, modificado post-traduccionalmente y localizado en el compartimento subcelular apropiado. La elección de una célula hospedadora adecuada puede también estar influida por la elección de la señal de detección. Por ejemplo, el uso de construcciones con genes reporteros (por ejemplo, *lacZ*, luciferasa, timidina quinasa o la proteína verde fluorescente “GFP”) puede proporcionar  
15 una señal seleccionable mediante la activación o inhibición de la transcripción del gen de interés en respuesta a una proteína reguladora de la transcripción. De cara a conseguir una selección o “screening” óptimo, el fenotipo de la célula hospedadora deberá ser considerado.

20 Una célula hospedadora de la presente invención incluye células procariontas y eucariotas. Entre las procariontas se incluyen organismos gram negativos (por ejemplo, *Escherichia coli*) o gram positivos (por ejemplo, bacterias del género *Bacillus*). Las células procariontas se usarán, preferiblemente, para la propagación de la secuencia del control de la transcripción del vector que contiene el(los) polinucleótido(s) objeto(s) de la invención, lo que permitirá  
25 conseguir un mayor número de copias del vector conteniendo el(los) polinucleótido(s) objeto(s) de la invención. Entre las células hospedadoras procariontas adecuadas para la transformación de este vector se encuentran, por ejemplo, pero sin limitarse a, *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium*, y otras especies dentro de los géneros *Pseudomonas*, *Streptomyces* y *Staphylococcus*. Las células eucariotas incluyen, entre otras,  
30 células de levadura, células de plantas, células de hongos, células de insectos, células de mamífero, y células de organismos parásitos (por ejemplo, *Trypanosomas*). Tal y como se emplea en esta memoria, el término levadura no incluye sólo levadura en el sentido taxonómico estricto, es decir, organismos unicelulares, sino también hongos multicelulares similares a las levaduras u hongos filamentosos. Ejemplos de especies son *Kluyveromyces*  
35 *lactis*, *Schizosaccharomyces pombe*, y *Ustilago maydis*, con *Saccharomyces cerevisiae* y *Pichia pastoris* como organismos preferidos. Otras levaduras que pueden utilizarse en la

producción de la(s) secuencia(s) poliaminoacídica(s) de la presente invención son *Neurospora crassa*, *Aspergillus niger*, *Aspergillus nidulans*, *Candida tropicalis*, y *Hansenula polymorpha*. Los sistemas de cultivo con células hospedadora de mamífero incluyen líneas celulares establecidas como las células COS, células L, células 3T3, células de ovario de hámster chino (CHO), células madre embrionarias, con las células BHK, HeK o HeLa como células preferidas. Las células eucariotas son, preferiblemente, utilizadas para la expresión del gen recombinante mediante la aplicación de la secuencia de regulación de la transcripción o el vector de expresión de la presente invención.

10 Así otro objeto de la invención se refiere a un método de obtención del polipéptido de la invención, que comprende:

- 1) Introducir el vector de la invención en una célula hospedadora adecuada (célula hospedadora de la invención),
- 2) Cultivar la célula hospedadora de la invención en un medio adecuado, y,
- 15 3) Purificar el polipéptido de la invención con actividad RT.

Un cultivo de células hospedadoras se refiere al proceso de mantener y crecer las células hospedadoras. Los cultivos celulares necesitan condiciones controladas de temperatura, pH, porcentajes de gases (oxígeno y dióxido de carbono), así como la presencia de los nutrientes adecuados para permitir la viabilidad y la división celular. Los cultivos celulares pueden desarrollarse en sustratos sólidos como el agar, o en medio líquido, lo que permite cultivar grandes cantidades de células en suspensión.

El término “purificar” tal y como se emplea en la descripción, se refiere al aislamiento del polipéptido de la invención y a su concentración, del resto de polipéptidos presentes en el medio de cultivo de la célula hospedadora de la invención. El aislamiento de la RT puede llevarse a cabo mediante técnicas de solubilidad diferencial, cromatografía, electroforesis o isoelectroenfoque. Las técnicas de cromatografía pueden estar basadas en el peso molecular, la carga iónica (basada en el estado de ionización de los aminoácidos en las condiciones de trabajo), la afinidad de la proteína por determinadas matrices o columnas cromatográficas, o mediante etiquetas de purificación, y puede realizarse en columna, en papel o en placa. El aislamiento de la proteína puede realizarse, por ejemplo, mediante precipitación con sulfato amónico, cromatografía líquida rápida (FPLC, del inglés “Fast Protein Liquid Chromatography”) o cromatografía líquida de alta eficacia (HPLC, del inglés “High Performance Liquid Chromatography”), empleando sistemas automatizados que

reducen notablemente el tiempo de purificación e incrementan el rendimiento de la purificación.

5 La expresión "etiqueta de purificación" o "etiqueta de afinidad", tal y como se utiliza en la presente descripción se refiere a una secuencia de aminoácidos que ha sido incorporada (generalmente, por ingeniería genética) a una proteína para facilitar su purificación. La etiqueta, que puede ser otra proteína o una secuencia corta de aminoácidos, permite la purificación de la proteína, por ejemplo, mediante cromatografía de afinidad. Etiquetas de purificación conocidas en el estado de la técnica son, por ejemplo, pero sin limitarse a, el péptido de unión a 10 calmodulina (CBP), la enzima glutatión-S-transferasa (GST) o una cola de residuos de histidina.

Otro objeto de la invención se refiere al método de obtención del polinucleótido de la invención, de ahora en adelante método de la invención, que podría hacerse, por ejemplo, pero sin limitarse a, por mutagénesis dirigida o al azar a partir de un polinucleótido no mutado, mediante 15 síntesis química del polinucleótido completo, o a través del ensamblado de fragmentos de ADN que codifiquen para distintas porciones de la secuencia que se desea obtener.

Otro objeto de la invención se refiere al uso del polinucleótido de la invención para la obtención del polipéptido de la invención con actividad RT. 20

Otro objeto de la invención se refiere al uso de la célula hospedadora de la invención para la obtención del polipéptido de la invención. Preferentemente, la célula hospedadora de la invención es una bacteria, más preferentemente *Escherichia coli*.

25 Las RT estables a temperaturas elevadas, como es el caso del polipéptido de la invención, son útiles en aplicaciones tales como la amplificación de ARNs difíciles, es decir, aquellos que contengan estructuras secundarias y/o secuencias ricas en pares de bases G:C.

Así, otro objeto de la invención es el uso del polipéptido de la invención en cualquiera de las 30 aplicaciones o métodos ya conocidos en el estado de la técnica.

Un objeto particular de la presente invención se refiere a un método de retrotranscripción de un ácido nucleico molde, preferiblemente ARNm o miARN, que comprende:

- a) mezclar dicho ácido nucleico molde con la RT de la invención, e
- 35 b) incubar la mezcla del paso (a) en condiciones que permitan la síntesis de ADN complementario al ácido nucleico molde.

Otro objeto particular de la presente invención se refiere a un método de amplificación de un ácido nucleico molde, preferiblemente ARNm o miARN, que comprende:

- a) mezclar dicho ácido nucleico con la RT de la invención y con, al menos, una ADN polimerasa dependiente de ADN, e
- 5 b) incubar la mezcla del paso (a) en condiciones que permitan la amplificación de ADN complementario al ácido nucleico molde.

Otro objeto particular de la presente invención se refiere a un método de secuenciación de un ácido nucleico, preferiblemente ARNm o miARN, que comprende:

- 10 a) poner en contacto dicho ácido nucleico con la RT de la invención,
- b) incubar dicha mezcla en condiciones que permitan la síntesis de una población de moléculas de ADN complementario al ácido nucleico molde, y
- c) separar dicha población de moléculas de ADN complementario para determinar la secuencia de nucleótidos.

15

El término "retrotranscripción" o "transcripción inversa", tal y como se utiliza en la presente descripción, se refiere a la síntesis de un ADN complementario a un ARN.

El término "amplificación", tal y como se utiliza en la presente descripción, se refiere al aumento del número de copias de un ácido nucleico molde. En una realización preferida, la amplificación tiene lugar mediante PCR.

El término "secuenciación", tal y como se utiliza en la presente descripción, se refiere a la determinación del orden de los nucleótidos de un ácido nucleico molde.

25

El término "ácido nucleico molde" o "molde" tal y como se utiliza en la presente descripción se refiere a una molécula de ácido nucleico de cadena simple o de doble cadena que va a ser retrotranscrita, amplificada o secuenciada.

La expresión "condiciones que permitan la síntesis de ADN complementario" se refiere a las condiciones en las que puede tener lugar la incorporación de los nucleótidos a un ADN naciente mediante complementariedad de bases con el ácido nucleico molde.

Generalmente las condiciones en las que tiene lugar la síntesis de ADN incluyen: (a) poner en contacto dicho ácido nucleico molde con la RT de la invención en una mezcla que además comprende un cebador, un catión bivalente, por ejemplo, Mg<sup>2+</sup>, y nucleótidos, y (b)

35

someter dicha mezcla a una temperatura suficiente para que una polimerasa de ADN, por ejemplo, la RT de la invención, inicie la incorporación de los nucleótidos al cebador mediante complementariedad de bases con el ácido nucleico molde, y de lugar una población de moléculas de ADN complementario de diferente tamaño. La separación de dicha población de moléculas de ADN complementario permite determinar la secuencia de nucleótidos del ácido nucleico molde.

La incorporación de nucleótidos mal apareados durante la síntesis del ADN complementario puede resultar en una o más bases desapareadas. Por tanto, la cadena de ADN sintetizada puede no ser exactamente complementaria al ácido nucleico molde.

La expresión "condiciones que permitan la síntesis de una población de moléculas de ADN complementario al ácido nucleico molde" se refiere a las condiciones en las cuáles se realiza la secuenciación, y que generalmente incluyen (a) poner en contacto dicho ácido nucleico molde con la RT de la invención en una mezcla que además comprende un cebador, un catión bivalente, (por ejemplo,  $Mg^{2+}$ ), y nucleótidos, generalmente, dNTPs y, al menos, un ddNTP, y (b) someter dicha mezcla a una temperatura suficiente para que una polimerasa de ADN, por ejemplo, la RT de la invención, inicie la incorporación de los nucleótidos al cebador mediante complementariedad de bases con el ácido nucleico molde, y de lugar a una población de moléculas de ADN complementario de diferente tamaño. La separación de dicha población de moléculas de ADN complementario, generalmente, mediante electroforesis, permite determinar la secuencia de nucleótidos.

El término "cebador", como se utiliza aquí, se refiere a un oligonucleótido capaz de actuar como punto de inicio de la síntesis de ADN cuando hibrida con el ácido nucleico molde. Preferiblemente, el cebador es un oligonucleótido de desoxirribosa.

Los cebadores pueden prepararse mediante cualquier método adecuado, incluyendo, por ejemplo, pero sin limitarse a, la clonación y restricción de secuencias apropiadas y la síntesis química directa. Los cebadores pueden diseñarse para hibridar con secuencias específicas de nucleótidos en el ácido nucleico molde (cebadores específicos) o pueden ser sintetizados al azar (cebadores arbitrarios).

El término "cebador específico", tal y como se utiliza en la presente descripción, se refiere a un cebador cuya secuencia es complementaria a una secuencia específica de nucleótidos en el ácido nucleico molde que se quiere retrotranscribir, amplificar o secuenciar.

El término "cebador arbitrario" se refiere a un cebador cuya secuencia es sintetizada al azar y que se usa para iniciar la síntesis del ADN en posiciones aleatorias del ácido nucleico molde que se quiere retrotranscribir, amplificar o secuenciar. Con frecuencia se emplea una población de distintos cebadores arbitrarios. El término "cebadores arbitrarios" se refiere a un conjunto de cebadores cuya secuencia es sintetizada al azar y que se usa para iniciar la síntesis del ADN en posiciones aleatorias del ácido nucleico molde que se quiere retrotranscribir, amplificar o secuenciar.

El término "hibridación", tal y como se utiliza en la presente descripción, se refiere al apareamiento de dos moléculas de ácido nucleico (de ADN y/o ARN) de cadena simple complementarias para dar una molécula de doble cadena. Preferiblemente, la complementariedad es del 100%. Esto es, en la región de complementariedad cada nucleótido de una de las dos moléculas de ácido nucleico puede formar enlaces de hidrógeno con un nucleótido presente en la otra molécula de ácido nucleico. Sin embargo, aquellos con una experiencia normal en el campo reconocerán que dos moléculas de ácido nucleico que posean una región con complementariedad menor al 100% también pueden hibridar.

El término "nucleótido", tal y como se utiliza en la presente descripción, se refiere a una molécula orgánica formada por la unión covalente de una pentosa, una base nitrogenada y un grupo fosfato. El término nucleótido incluye desoxirribonucleósidos trifosfato como, por ejemplo, pero sin limitarse, dATP, dCTP, dITP, dUTP, dGTP, dTTP, o derivados de los mismos. El término nucleótido incluye también dideoxirribonucleósidos trifosfato (ddNTPs), como por ejemplo, ddATP, ddCTP, ddGTP, ddITP, ddTTP o derivados de los mismos. De acuerdo con la presente invención un "nucleótido" o un "cebador" puede ser marcado o etiquetado mediante técnicas bien conocidas en el estado de la técnica. Etiquetas detectables incluyen, por ejemplo, isótopos radiactivos, etiquetas fluorescentes, etiquetas quimioluminiscentes, etiquetas bioluminiscentes o etiquetas enzimáticas.

El término "ADN polimerasa dependiente de ADN", tal y como se utiliza en la presente descripción, se refiere a una ADN polimerasa capaz de catalizar la polimerización de desoxirribonucleótidos utilizando ADN como ácido nucleico molde. Ejemplos de ADN polimerasa dependientes de ADN que pueden ser empleadas en el método de amplificación de la siguiente invención son, pero sin limitarnos, las ADN polimerasas de *Thermus thermophilus* (Tth), *Thermus aquaticus* (Taq), *Thermotoga neapolitana* (Tne), *Thermotoga maritima* (Tma), *Thermococcus litoralis* (Tli o Vent™), *Pyrococcus furiosus* (Pfu),

5 *Pyrococcus* sp. GB-D (Deep Vent™), *Pyrococcus waasii* (Pwo), *Bacillus stearothermophilus* (Bst), *Bacillus caldophilus* (Bca), *Sulfolobus acidocaldarius* (Sac), *Thermoplasma acidophilum* (Tac), *Thermus flavus* (Tfl/Tub), *Thermus ruber* (Tru), *Thermus brockianus* (DyNAzyme™), *Methanobacterium thermoautotrophicum* (Mth) o *Mycobacterium* sp. (Mtb, Mlep).

Otro objeto de la presente invención se refiere a un kit que comprende los elementos necesarios para llevar a cabo cualquiera de los métodos descritos anteriormente en la presente descripción, de ahora en adelante kit de la invención.

10 Una realización preferida de este objeto de la invención, se refiere a un kit de la invención para llevar a cabo cualquiera de los métodos descritos anteriormente en la presente descripción, y que comprende:

a) la RT de la invención, y

15 b) al menos, un elemento de la lista que comprende:

i) un tampón,

ii) un cebador,

iii) una ADN polimerasa dependiente de ADN, y

iv) un nucleótido.

20 Un objeto particular de la invención se refiere al uso del kit de la invención para la retrotranscripción, la amplificación o la secuenciación de un ácido nucleico molde donde preferiblemente éste es ARN mensajero (ARNm) o un microARN (miARN) o un precursor de miARN.

25 A lo largo de la descripción y las reivindicaciones la palabra “comprende” y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración de la invención y no se pretende que sean limitativos de la presente invención.

### **Descripción de las figuras**

35 **Figura 1.** Amplificación por RT-PCR de fragmentos de ARN codificante para actina (de aproximadamente 500 y 950 pares de bases), a partir de ARN total de hígado de ratón. La

amplificación se llevó a cabo con las enzimas indicadas: (A) RT de VIH-1 grupo O “wild-type” (RTO\_WT\*) (pocillo 7), y mutantes K358R/A359G/S360A (RTO\_3M\*) (pocillo 1), T355A/Q357M/K358R/A359G/S360A (RTO\_5M\*) (pocillo 2), K65R/K358R/A359G/S360A (pocillo 3), K65R/V75I/K358R/A359G/S360A (pocillo 4), K358R/A359G/S360A/E478Q (RTO\_E478Q\_3M\*) (pocillo 5) y K65R/K358R/A359G/S360A/E478Q (pocillo 6). (B) Mutantes de la RT del VIH-1 grupo O: K358R/A359G/S360A (RTO\_3M\*) (pocillo 1), T69SSG/K358R/A359G/S360A (RTO\_T69SSG\_3M\*) (pocillo 2), V148I/K358R/A359G/S360A (pocillo 3) y F61A/K358R/A359G/S360A (pocillo 4). Las temperaturas indicadas se refieren a la reacción de síntesis del ADN copia. Los pocillos m y c muestran los marcadores de peso molecular (ADN del fago phi29 digerido con HindIII) y un control negativo (sin ADNc), respectivamente. Todas las mutaciones indicadas se han introducido en el contexto de secuencia de la RTO\_WT.

**Figura 2.** Muestra la eficacia de la retrotranscripción estimada por PCR a tiempo real para las RTs wild-type del VIH-1 grupo M subtipo B aislado BH10 (BH10) y de grupo O (RTO\_WT\*; indicada como RTO en la figura), junto con los mutantes objeto de esta patente RTO\_3M\* (3M), RTO\_E478Q\_3M\* (3Q) y RTO\_T69SSG\_3M\* (SG) y un mutante de referencia portador de los cambios característicos de 3M, junto con K65R y V75I (K65R/V75I/K358R/A359G/S360A, KV). (ACTB) representa amplificaciones de ARN mensajero de actina, y (GAPDH) de ARN mensajero de gliceraldehído 3-fosfato deshidrogenasa. En la parte superior derecha de cada histograma se indica la temperatura a la que se llevó a cabo la reacción de retrotranscripción.

**Figura 3.** Muestra la eficiencia de extensión de extremos desapareados (G:T, G:G y G:A) de las RTs mutantes RTO\_3M\* (3M), RTO\_5M\* (5M), RTO\_T69SSG\_3M\* (SG) y RTO\_E478Q\_3M\* (3Q) en comparación con la RTO\_WT\* (O\_WT).

## **EJEMPLOS**

**30 Efecto de la temperatura de la reacción de retrotranscripción sobre la eficacia de la reacción de RT-PCR utilizando distintos mutantes de la RT del VIH-1 grupo O.**

La eficiencia de la reacción de retrotranscripción a diferentes temperaturas se determinó tras amplificación por PCR del ADN copia obtenido. Los mutantes K358R/A359G/S360A (RTO\_3M\*), K358R/A359G/S360A/E478Q (RTO\_E478Q\_3M\*), T355A/Q357M/K358R/A359G/S360A (RTO\_5M\*) y T69SSG/K358R/A359G/S360A

(RTO\_T69SSG\_3M\*) fueron más eficaces que la RT “wild-type” de VIH-1 grupo O (RTO\_WT\*) y otras RTs mutantes en la amplificación de fragmentos de ARN de aproximadamente 500 y 950 pares de bases derivados del gen de actina, a partir de reacciones de retrotranscripción llevadas a cabo a distintas temperaturas en el rango de 37 a 78°C (Figura 1).

Estos resultados fueron consistentes con los obtenidos en la amplificación de secuencias derivadas de ARNs mensajeros de actina o de gliceraldehído-3-fosfato deshidrogenasa (GAPDH), en reacciones de PCR a tiempo real. En ambos casos se observó que la eficacia de retrotranscripción de los mutantes RTO\_3M\* y RTO\_E478Q\_3M\* fue sensiblemente superior a la de la RT “wild-type” a temperaturas de 75 y 78°C (Figura 2). En el caso del mutante RTO\_T69SSG\_3M\*, se observó que su eficacia de amplificación fue superior a la de la enzima “wild-type” en el caso del ARN de GAPDH, y similar en el caso del ARN de actina.

15

#### **Fidelidad de copia de los mutantes con mayor actividad a temperaturas elevadas**

Las retrotranscriptasas RTO\_3M\*, RTO\_5M\* y RTO\_E478Q\_3M\* presentaron una actividad catalítica similar a la de la enzima “wild-type” (RTO\_WT\*) en ensayos de incorporación de un solo nucleótido (Tabla 2), si bien la eficacia catalítica ( $k_{pol}/K_d$ ) del mutante RTO\_T69SSG\_3M\* fue algo inferior a la de la enzima wild-type. No se detectaron diferencias significativas entre las enzimas estudiadas en los ensayos de extensión del extremo 3' desapareado G:T. En la extensión de extremos G:A y G:G todas las RTs presentaron una eficacia similar a la de la enzima “wild-type”, si bien, el mutante RTO\_T69SSG\_3M\* fue más proclive que el resto a extender extremos G:G o G:A (Figura 3).

25

**Tabla 2.** Parámetros cinéticos de extensión de extremos desapareados sobre el complejo molde-iniciador 31T/21P para la RT del VIH-1 grupo O “wild-type” (RTO\_WT\*) y los mutantes RTO\_3M\*, RTO\_5M\*, RTO\_T69SSG\_3M\* y RTO\_E478Q\_3M\*, determinados en el estado pre-estacionario. Los números indicados entre paréntesis en la última columna muestran el aumento de fidelidad con respecto a la RTO\_WT\*, obtenido de los cocientes entre las eficiencias de extensión de los extremos desapareados estudiados.

30

Tabla 2. (Cont.)

Enzima	Par de bases en el extremo 3'	$k_{\text{pol}}$ (s <sup>-1</sup> )	$K_d$ (μM)	$k_{\text{pol}}/K_d$ (μM <sup>-1</sup> s <sup>-1</sup> )	Eficiencia de extensión de extremos desapareados ( $f_{\text{ext}}$ )
<b>RTO_WT*</b>	G:C	32.3 ± 1.5	11.4 ± 1.7	2.83 ± 0.43	
	G:T	13.8 ± 2.8	8148 ± 3206	(1.70 ± 0.75) × 10 <sup>-3</sup>	(6.00 ± 2.80) × 10 <sup>-4</sup>
	G:G	0.56 ± 0.07	1132 ± 316	(4.94 ± 1.51) × 10 <sup>-4</sup>	(1.74 ± 0.59) × 10 <sup>-4</sup>
	G:A	0.021 ± 0.003	7817 ± 2334	(2.71 ± 0.89) × 10 <sup>-6</sup>	(9.57 ± 3.46) × 10 <sup>-7</sup>
<b>RTO_3M*</b>	G:C	16.4 ± 0.6	5.0 ± 0.7	3.28 ± 0.50	
	G:T	13.5 ± 1.3	7615 ± 1519	(1.77 ± 0.39) × 10 <sup>-3</sup>	(5.40 ± 1.45) × 10 <sup>-4</sup> <b>(1.1)</b>
	G:G	0.41 ± 0.04	1325 ± 249	(3.09 ± 0.64) × 10 <sup>-4</sup>	(9.42 ± 2.42) × 10 <sup>-5</sup> <b>(1.8)</b>
	G:A	0.013 ± 0.001	5516 ± 1061	(2.28 ± 0.47) × 10 <sup>-6</sup>	(6.95 ± 1.78) × 10 <sup>-7</sup> <b>(1.4)</b>
<b>RTO_5M*</b>	G:C	23.5 ± 1.4	9.1 ± 1.9	2.57 ± 0.55	
	G:T	6.08 ± 0.38	3249 ± 445	(1.87 ± 0.28) × 10 <sup>-3</sup>	(7.28 ± 1.90) × 10 <sup>-4</sup> <b>(0.8)</b>
	G:G	0.57 ± 0.05	1003 ± 211	(5.65 ± 1.27) × 10 <sup>-4</sup>	(2.19 ± 0.68) × 10 <sup>-4</sup> <b>(0.8)</b>
	G:A	0.029 ± 0.003	7512 ± 1552	(3.91 ± 0.89) × 10 <sup>-6</sup>	(1.52 ± 0.47) × 10 <sup>-6</sup> <b>(0.6)</b>
<b>RTO_E478Q_3M*</b>	G:C	27.7 ± 1.0	12.5 ± 1.6	2.21 ± 0.30	
	G:T	11.1 ± 2.0	7566 ± 2725	(1.47 ± 0.59) × 10 <sup>-3</sup>	(6.65 ± 2.81) × 10 <sup>-4</sup> <b>(0.9)</b>
	G:G	0.54 ± 0.01	489 ± 16	(1.10 ± 0.04) × 10 <sup>-3</sup>	(4.97 ± 0.70) × 10 <sup>-4</sup> <b>(0.4)</b>
	G:A	0.029 ± 0.002	6362 ± 1110	(4.54 ± 0.86) × 10 <sup>-6</sup>	(2.05 ± 0.48) × 10 <sup>-6</sup> <b>(0.5)</b>
<b>RTO_T69SSG_3M*</b>	G:C	21.1 ± 1.8	14.0 ± 4.4	1.51 ± 0.49	
	G:T	3.15 ± 0.36	4632 ± 1070	(6.80 ± 1.75) × 10 <sup>-4</sup>	(4.50 ± 1.86) × 10 <sup>-4</sup> <b>(1.3)</b>
	G:G	0.26 ± 0.01	2080 ± 250	(1.23 ± 0.16) × 10 <sup>-4</sup>	(8.14 ± 2.85) × 10 <sup>-5</sup> <b>(2.1)</b>
	G:A	(7.6 ± 1.6) × 10 <sup>-3</sup>	5854 ± 1877	(1.30 ± 0.50) × 10 <sup>-6</sup>	(8.61 ± 4.33) × 10 <sup>-7</sup> <b>(1.1)</b>

También se midió la fidelidad de copia de algunas de estas enzimas en ensayos de complementación que utilizan derivados del fago M13mp2 portadores del gen *lacZ*. Se determinó la frecuencia con que se obtenían mutantes cuando el proceso de síntesis se realizaba con las distintas RTs recombinantes. En estos ensayos tampoco se observaron diferencias importantes entre la fidelidad de copia de los mutantes y la de la enzima “wild-type” (Tabla 3). En su conjunto, estos datos indican que el aumento de eficacia de la retrotranscripción a altas temperaturas, que presentan los mutantes aquí descritos no tiene un efecto negativo sobre la fidelidad de copia de la enzima.

**Tabla 3.** Fidelidad de los mutantes RTO\_3M\*, RTO\_T69SSG\_3M\* y RTO\_E478Q\_3M\*, comparada con la de RTO\_WT\* y estimada mediante ensayos genéticos de complementación (M13mp2 *lacZ*α “forward mutation assay”). Los datos de la RTO\_WT\*, obtenidos de dos experimentos independientes, han sido publicados previamente (Álvarez *et al. Nucleic Acids Res* 2013; 41: 4601-4612).

RT	Placas totales	Placas mutante	Frecuencia de mutantes	Incremento de la fidelidad respecto a la RTO_WT
RTO_WT* (experimento 1)	7579	63	0,00831	--
(experimento 2)	3957	38	0,00960	
RTO_3M*	8394	74	0,00881	1,0
RTO_T69SSG_3M*	14262	79	0,00554	1,6
RTO_E478Q_3M*	7802	52	0,00666	1,3

## MATERIAL Y MÉTODOS

### Expresión y purificación de la RT de grupo O y de sus mutantes

La expresión y purificación de las RTs se llevó a cabo con una versión modificada del plásmido p66RTB (Boretto *et al. Anal. Biochem.* 2001; 292: 139-147; Matamoros *et al. J. Mol. Biol.* 2005; 349: 451-463), que contiene el gen de resistencia a ampicilina y en el que se clonó la región codificante de la subunidad p66 de la RT de un aislado del VIH-1 de grupo O (Menéndez-Arias *et al. J. Biol. Chem.* 2001; 276: 27470-27479; Álvarez *et al. J Mol Biol* 2009; 392: 872-884; patente WO2010130864). La purificación de las RTs se lleva a cabo siguiendo el procedimiento descrito por Boretto *et al. (Anal. Biochem.* 2001; 292: 139-147),

en el que se incluye una etapa de lisis bacteriana y homogenización, seguida de cromatografías de intercambio iónico (en fosfocelulosa) y de afinidad (en columnas de Ni<sup>2+</sup>-nitriloacético-agarosa).

5 **Construcción de plásmidos portadores de mutaciones en el contexto de la RT del VIH-1 grupo O**

Los plásmidos para la expresión de las RTs mutantes RTO\_3M\*, RTO\_E478Q\_3M\*, RTO\_5M\* y RTO\_T69SSG\_3M\* se obtuvieron por mutagénesis dirigida utilizando el kit  
10 “Quik-Change Site-Directed Mutagenesis” de Stratagene, siguiendo las instrucciones del fabricante.

Como oligonucleótidos mutagénicos se emplearon:

- a) para introducir las mutaciones K358R/A359G/S360A:  
15 5'-GGGAAATATACTAGGCAAAGGGGCGCCACACAAATGAC-3'  
5'-GTCATTTGTGTGGGCGCCCCTTTCCTAGTATATTTCCC-3'
- b) para T355A,  
5'-ACAGGGAAATATGCTAGGATGAGGGGCGCC-3' y  
5'-GGCGCCCCTCATCCTAGCATATTTCCCTGT-3'
- 20 c) para Q357M  
5'-GGGAAATATACTAGGATGAGGGGCGCCACACAAATGAC-3' y  
5'-GTCATTTGTGTGGGCGCCCCTCATCCTAGTATATTTCCC-3'
- d) para E478Q  
5'-CCAATCAAAGGCTCAATTAATGGCAG-3' y  
25 5'-CTGCCATTAATTGAGCCTTTTGATTGG-3'
- e) para introducir el cambio T69S, junto con la inserción Ser-Gly  
5'-GCTATAAAAAGAAAGATAGTAGTTCCGGGAAGTGGAGAAAGCTGGTAGAC-3'  
5'-GTCTACCAGCTTTCTCCACTTCCCGGAAGTACTATCTTTCTTTTTATAGC-3'

30 Como molde para la introducción de K358R/A359G/S360A, se empleó el plásmido portador de la secuencia que codifica para la p66 de VIH-1 (grupo O) (“wild-type”), descrito anteriormente (Álvarez *et al. J Mol Biol* 2009; 392: 872-884; patente WO2010130864). Las mutaciones T355A y E478Q y la mutación T69S (asociada con la inserción SG) se introdujeron por separado en el plásmido portador de K358R/A359G/S360A, para obtener  
35 los mutantes: T355A/K358R/A359G/S360A, K358R/A359G/S360A/E478Q y T69SSG/K358R/A359G/S360A. Finalmente el cambio Q357M se introdujo en el plásmido de

expresión portador del cambio T355A/K358R/A359G/S360A, para obtener el mutante T355A/Q357M/K358R/A359G/S360A. En todos los casos, tras la mutagénesis, se comprobó por secuenciación que la región codificante de p66 en dichos plásmidos era correcta y que contenía únicamente las mutaciones introducidas.

5

### **Efectos de mutaciones sobre la eficiencia de la reacción de retrotranscripción acoplada con amplificación por PCR**

Se llevaron a cabo reacciones de retrotranscripción a diferentes temperaturas, y posteriormente los productos de reacción (ADNc) se amplificaron por PCR en condiciones estándar (Álvarez *et al.* *J Mol Biol* 2009; 392: 872-884). Típicamente, la reacción de retrotranscripción se llevó a cabo en un volumen de 20  $\mu$ l [4  $\mu$ l de tampón Tris-HCl 250 mM (pH 8,3 at 25°C) que contiene KCl 375 mM, MgCl<sub>2</sub> 15 mM y ditiotreitól 50 mM; 1  $\mu$ l de ARN total aislado de hígado de ratón (1  $\mu$ g/ $\mu$ l); 4  $\mu$ l de una mezcla de los 4 dNTPs (a 2,5 mM cada uno de ellos); 1  $\mu$ l de oligo(dT)<sub>16</sub> (100  $\mu$ M); 0,5  $\mu$ l de inhibidor de ribonucleasa (40 unidades/ $\mu$ l); la RT a una concentración aproximada de 150 nM y el resto hasta 20  $\mu$ l de agua]. Inicialmente se incubó el ARN y el oligo dT a 68°C durante 3 min. Después se añaden los demás componentes de la reacción (incluida la RT) y se incubó durante 1 hora a la temperatura deseada para obtener el ADNc. Finalmente, la reacción se detiene incubando 10 min a 92°C, para inactivar a la enzima. El ADNc se amplifica por PCR en condiciones estándar, utilizando Taq polimerasa u otras enzimas similares (por ejemplo, Expand High Fidelity ADN polimerasa).

### **PCR en tiempo real**

25

La eficiencia de la retrotranscripción de las RTs se determinó a varias temperaturas (37, 50, 68, 75 y 78°C) mediante PCR en tiempo real. Para ello, se hicieron tres reacciones independientes en cada experimento. Todas las reacciones se llevaron a cabo en un volumen de 20  $\mu$ l en tampón Tris-HCl 50 mM (pH 8,3) que contenía KCl 75 mM, MgCl<sub>2</sub> 3 mM, ditiotreitól 10 mM, 1 U/ $\mu$ l de inhibidor de ribonucleasa (RNasin® Plus, Promega), 500  $\mu$ M de cada dNTP, 5  $\mu$ M de oligo(dT)<sub>16</sub>, 50 ng/ $\mu$ l de ARN total de hígado de ratón (Stratagene) y la RT correspondiente a 150 nM. La hibridación del ARN molde con el oligo(dT)<sub>16</sub> y la reacción de síntesis de ADNc se llevó a cabo en las condiciones descritas en el apartado anterior.

35 La eficiencia de la retrotranscripción se determinó mediante PCR cuantitativa (qPCR), calculando la cantidad relativa de ADNc producida a partir de los ARNs mensajeros de  $\beta$ -

actina y gliceraldehido 3-fosfato deshidrogenasa (GAPDH). Para ello se usaron los oligonucleótidos: 5'-CTAAGGCCAACCGTGAAAAG-3' y 5'-ACCAGAGGCATACAGGGACA-3' para actina, y 5'-CTCCCACTCTTCCACCTTCG-3' y 5'-CATACCAGGAAATGAGCTTGACAA-3' para GAPDH. Las reacciones de amplificación por PCR se hicieron por triplicado en un volumen final de 10 µl con una cantidad de ADNc aproximadamente igual a 5 ng de ARN total, 250 nM de cada oligonucleótido ("primer") y 5 µl de Power Sybr Green PCR Master Mix (Applied Biosystems PN 4367659), que incluye AmpliTaq Gold® ADN polimerasa, dNTPs y el resto de reactivos necesarios para llevar a cabo la PCR. Se emplearon placas de 384 pocillos MicroAmp® Optical 384-well Reaction Plates con código de barras (Applied Biosystems PN 4309849). Tras una etapa inicial de desnaturalización a 95°C (10 min), las muestras se sometieron a 40 ciclos de amplificación (15 s a 95°C más 1 min a 60°C). Se incluyeron curvas de desnaturalización de 60 a 95°C (2% de pendiente) al final del programa para verificar la especificidad de la PCR. La fluorescencia se midió durante los pasos a 60°C y de desnaturalización, en un instrumento de PCR cuantitativa ABI 7900HT (Applied Biosystems). En todas las placas se incluyó un control negativo y una curva control de eficiencia de la amplificación génica.

El análisis de datos se hizo con el programa SDS 2.2.1 (Applied Biosystems). El valor de  $\Delta C_t$  se calculó para cada muestra como  $\Delta C_t = C_t - C_{t_{ref}}$ , donde  $C_t$  es el ciclo en el que se observa una amplificación significativa y  $C_{t_{ref}}$  es la media de los valores de  $C_t$  obtenidos para la RT WT del clon BH10. Las cantidades relativas de ADNc obtenido se calcularon como  $2^{\Delta C_t}$ , y los valores se expresan como la media  $\pm$  desviación estándar de los tres valores calculados en cada experimento.

## 25 **Ensayos de fidelidad**

La fidelidad de copia de las RTs se ha determinado mediante ensayos cinéticos de extensión de extremos desapareados, todos ellos llevados a cabo en condiciones de estado pre-estacionario, con el objetivo de determinar la capacidad de las distintas RTs para extender complejos molde-iniciador correcta o incorrectamente apareados (Matamoros *et al. J. Mol. Biol.* 2008; 375: 1234-1248; Barrioluengo *et al. Biochem J* 2011; 436: 599-607). Además se llevaron a cabo ensayos genéticos basados en la expresión del gen *lacZ $\alpha$*  en el contexto del fago M13mp2 (Bebenek y Kunkel. *Methods Enzymol.* 1995; 262: 217-232; Barrioluengo *et al. Biochem J* 2011; 436: 599-607).

35

## REIVINDICACIONES

1.- Polipéptido que codifica una proteína con actividad retrotranscriptasa aislada de un VIH-1 de grupo O, que presenta una mayor estabilidad a altas temperaturas, y que posee mayor actividad que la enzima de la que procede a temperaturas superiores a 75°C, manteniendo su fidelidad de copia, y caracterizado porque su secuencia de aminoácidos presenta una identidad de al menos un 50% con respecto a la secuencia parental SEQ ID NO 1, y porque comprende, al menos, los siguientes cambios de aminoácido:

- la sustitución del aminoácido lisina (K) original por el aminoácido arginina (R) en la posición homóloga a la posición 358 de SEQ ID NO 1 (mutación K358R),
- la sustitución del aminoácido alanina (A) original por el aminoácido glicina (G) en la posición homóloga a la posición 359 de SEQ ID NO 1 (mutación A359G), y
- la sustitución del aminoácido serina (S) original por el aminoácido alanina (A) en la posición homóloga a la posición 360 de SEQ ID NO 1 (mutación S360A).

2.- Polipéptido según la reivindicación 1 caracterizado porque su secuencia de aminoácidos posee además una de las siguientes mutaciones adicionales o cualesquiera de sus combinaciones:

- a) La sustitución del aminoácido treonina (T) original por la inserción de dos aminoácidos serina y una glicina (SSG) en la posición homóloga a la posición 69 de la SEQ ID No 1 (mutación T69SSG);
- b) La sustitución del aminoácido treonina (T) original por el aminoácido alanina (A) en la posición homóloga a la posición 355 de la SEQ ID NO 1 (mutación T355A);
- c) La sustitución del aminoácido glutamina (Q) original por el aminoácido metionina (M) en la posición homóloga a la posición 357 de la SEQ ID NO 1 (mutación Q357M);
- d) La sustitución del aminoácido ácido glutámico (E) original por el aminoácido glutamina (Q) en la posición homóloga a la posición 478 de la SEQ ID NO 1 (E478Q).

3.- Polipéptido según la reivindicación 1 caracterizado por que su secuencia se corresponde con la SEQ ID NO 3.

4.- Polipéptido según las reivindicaciones 1 y 2 caracterizado por que su secuencia se corresponde con la SEQ ID NO 5.

5.- Polipéptido según las reivindicaciones 1 y 2 caracterizado por que su secuencia se corresponde con la SEQ ID NO 7.

- 6.- Polipéptido según las reivindicaciones 1 y 2 caracterizado por que su secuencia se corresponde con la SEQ ID NO 9.
- 7.- Polipéptido según las reivindicaciones 1 a la 6 que presenta además las secuencias  
5 flanqueantes MNS en el extremo N-terminal y una cola de histidina en el extremo C-terminal, caracterizado porque su secuencia se corresponde con alguna de las siguientes: SEQ ID NO 13, SEQ ID NO 15, SEQ ID NO 17, SEQ ID NO 19.
- 8.- Polinucleótido que codifica un polipéptido con actividad retrotranscriptasa aislada de un  
10 VIH-1 de grupo O, que presenta una mayor estabilidad a altas temperaturas, y que posee mayor actividad que la enzima de la que procede a temperaturas superiores a 75°C, manteniendo su fidelidad de copia caracterizado por que codifica alguna de las secuencias nucleotídicas según las reivindicaciones 1 a 7.
- 15 9.- Polinucleótido según la reivindicación 8 caracterizado por que su secuencia se corresponde con alguna de las siguientes: SEQ ID NO 4, SEQ ID NO 6, SEQ ID NO 8, SEQ ID NO 10, SEQ ID NO 14, SEQ ID NO 16, SEQ ID NO 18, SEQ ID NO 20.
- 10.- Vector caracterizado por que comprende el polinucleótido según las reivindicaciones 8 y 9.  
20
- 11.- Célula hospedadora caracterizada porque comprende el polinucleótido según las reivindicaciones 8 y 9 o el vector según la reivindicación 10 y es capaz de producir el polipéptido según las reivindicaciones 1 a 7.
- 25 12.- La célula hospedadora según la reivindicación 11 caracterizada por ser una bacteria.
- 13.- La célula hospedadora según la reivindicación 12 caracterizada por ser *Escherichia coli*.
- 14.- Método de obtención del polipéptido según las reivindicaciones 1 a 7 caracterizado  
30 porque comprende las siguientes etapas:  
a. introducir un vector según la reivindicación 10 en una célula hospedadora adecuada,  
b. cultivar la célula hospedadora en un medio adecuado, y,  
c. purificar el polipéptido con actividad retrotranscriptasa.
- 35 15.- Uso del polinucleótido según las reivindicaciones 8 y 9 para la obtención del polipéptido según las reivindicaciones 1 a 7.

**16.-** Uso de la célula hospedadora según las reivindicaciones 11 a 13 para la obtención del polipéptido según las reivindicaciones 1 a 7.

5 **17.-** Uso del polipéptido según las reivindicaciones 1 a 7 para la retrotranscripción de un ácido nucleico molde.

**18.-** Uso del polipéptido según las reivindicaciones 1 a 7 para la amplificación de un ácido nucleico molde.

10 **19.-** Uso del polipéptido según las reivindicaciones 1 a 7 para la secuenciación de un ácido nucleico molde.

**20.-** Uso del polipéptido según las reivindicaciones 17 a 19 donde el ácido nucleico molde es ARNm o miARN.

15

**21.-** Método de retrotranscripción de un ácido nucleico molde que comprende:

- a) mezclar dicho ácido nucleico molde con el polipéptido según las reivindicaciones 1 a 7,
- b) incubar la mezcla del paso (a) en condiciones que permitan la síntesis de ADN complementario al ácido nucleico molde.

20

**22.-** Método de amplificación de un ácido nucleico molde que comprende:

- a) mezclar dicho ácido nucleico con el polipéptido según las reivindicaciones 1 a 7 y con, al menos, una ADN polimerasa dependiente de ADN, e
- b) incubar la mezcla del paso (a) en condiciones que permitan la amplificación de ADN complementario al ácido nucleico molde.

25

**23.-** Método de secuenciación de un ácido nucleico que comprende:

- a) poner en contacto dicho ácido nucleico con el polipéptido según las reivindicaciones 1 a 7,
- 30 b) incubar dicha mezcla en condiciones que permitan la síntesis de una población de moléculas de ADN complementario al ácido nucleico molde, y
- c) separar dicha población de moléculas de ADN complementario para determinar la secuencia de nucleótidos.

35 **24.-** Método según las reivindicaciones 21 a 23 donde el ácido nucleico molde es ARNm o miARN.

**25.-** Kit que comprende los elementos necesarios para llevar a cabo un método según las reivindicaciones 21 a 24 que comprende:

- a) el polipéptido según las reivindicaciones 1 a 7, y
- b) al menos, un elemento de la lista que comprende:

- 5           i) un tampón,
- ii) un cebador,
- iii) una ADN polimerasa dependiente de ADN, y
- iv) un nucleótido.

10 **26.-** Uso del kit según la reivindicación 25 para llevar a cabo un método según las reivindicaciones 21 a 24.

Figura 1

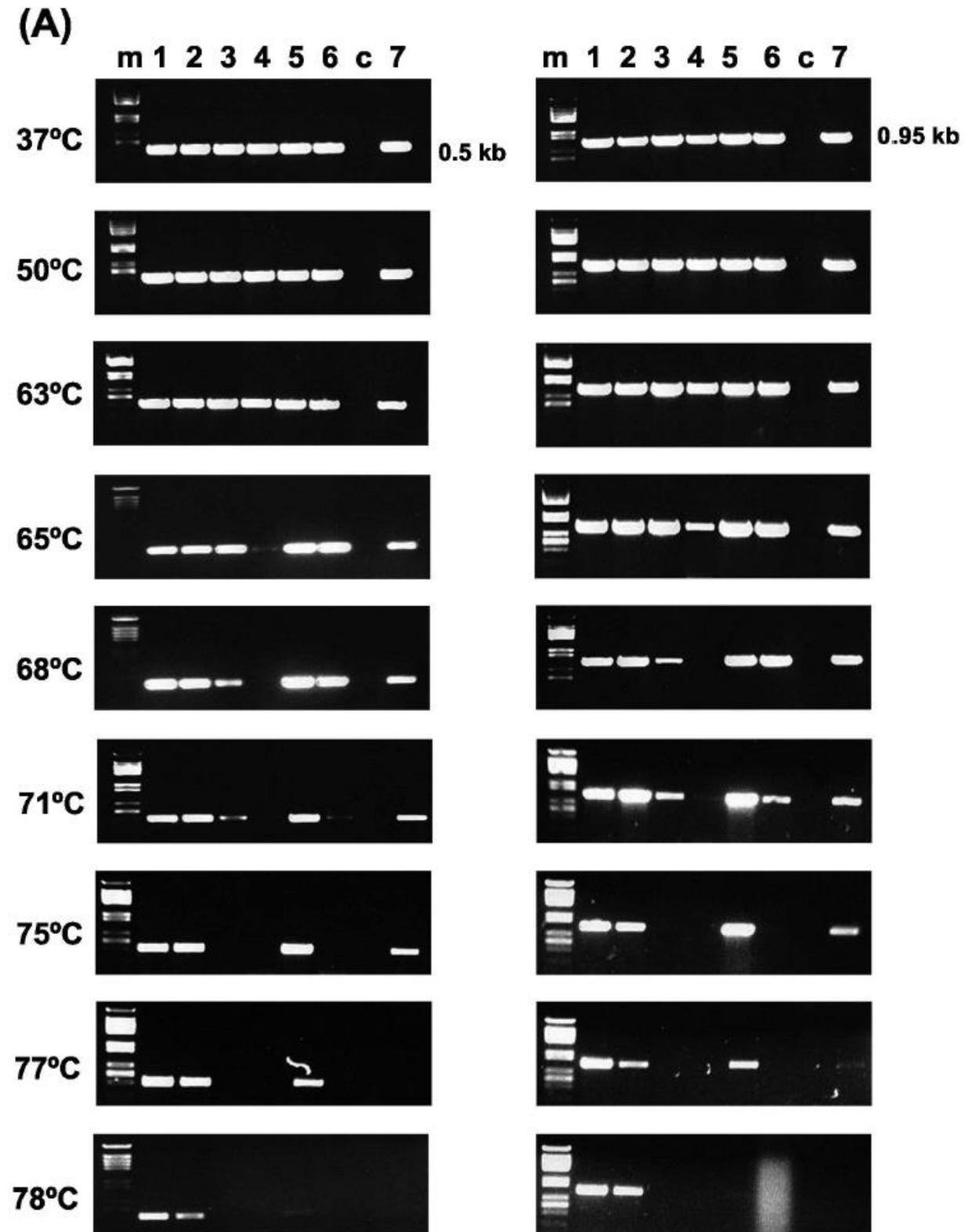


Figura 1 (Cont.)

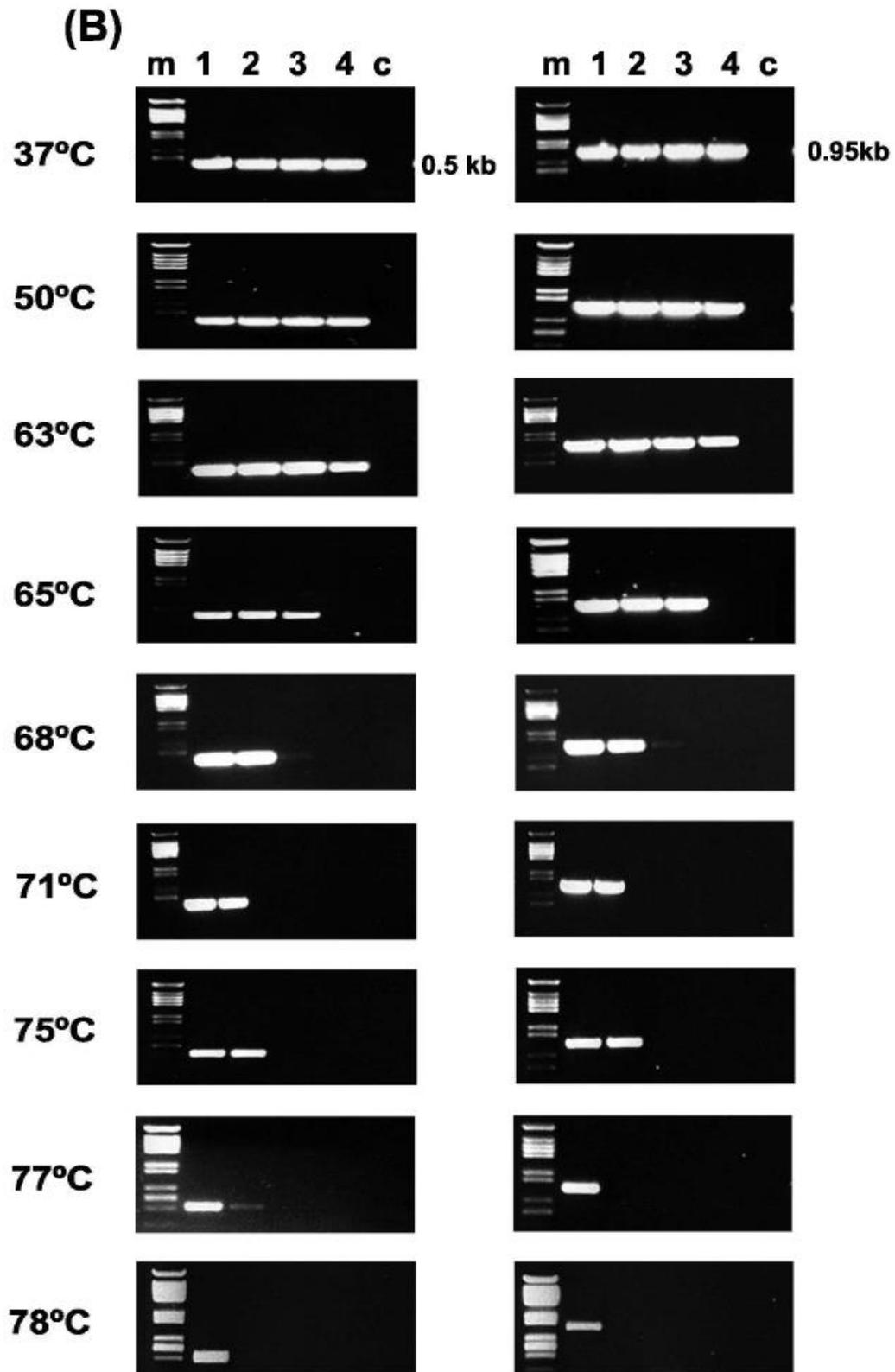


Figura 2

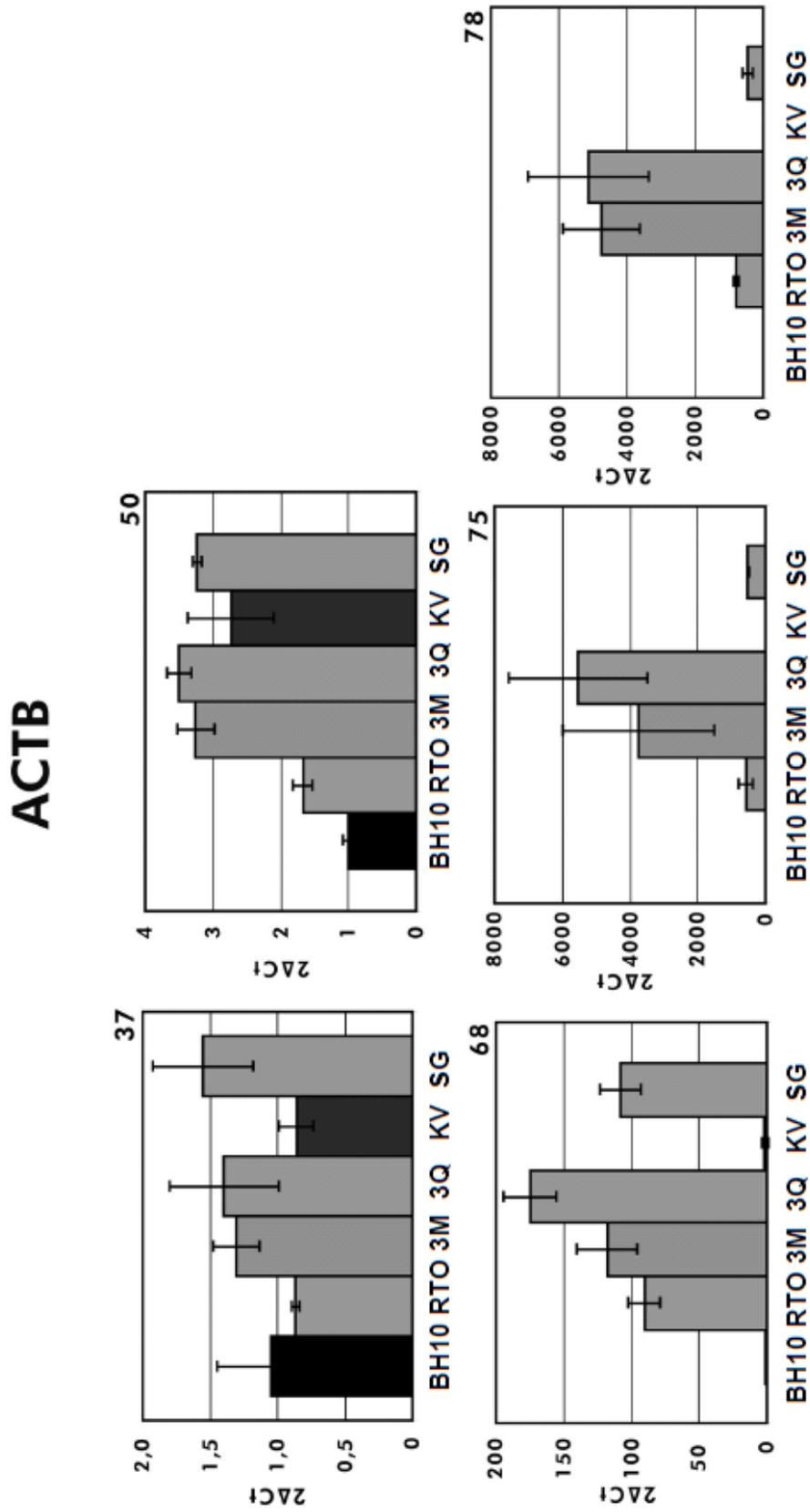


Figura 2 (Cont.)

**GAPDH**

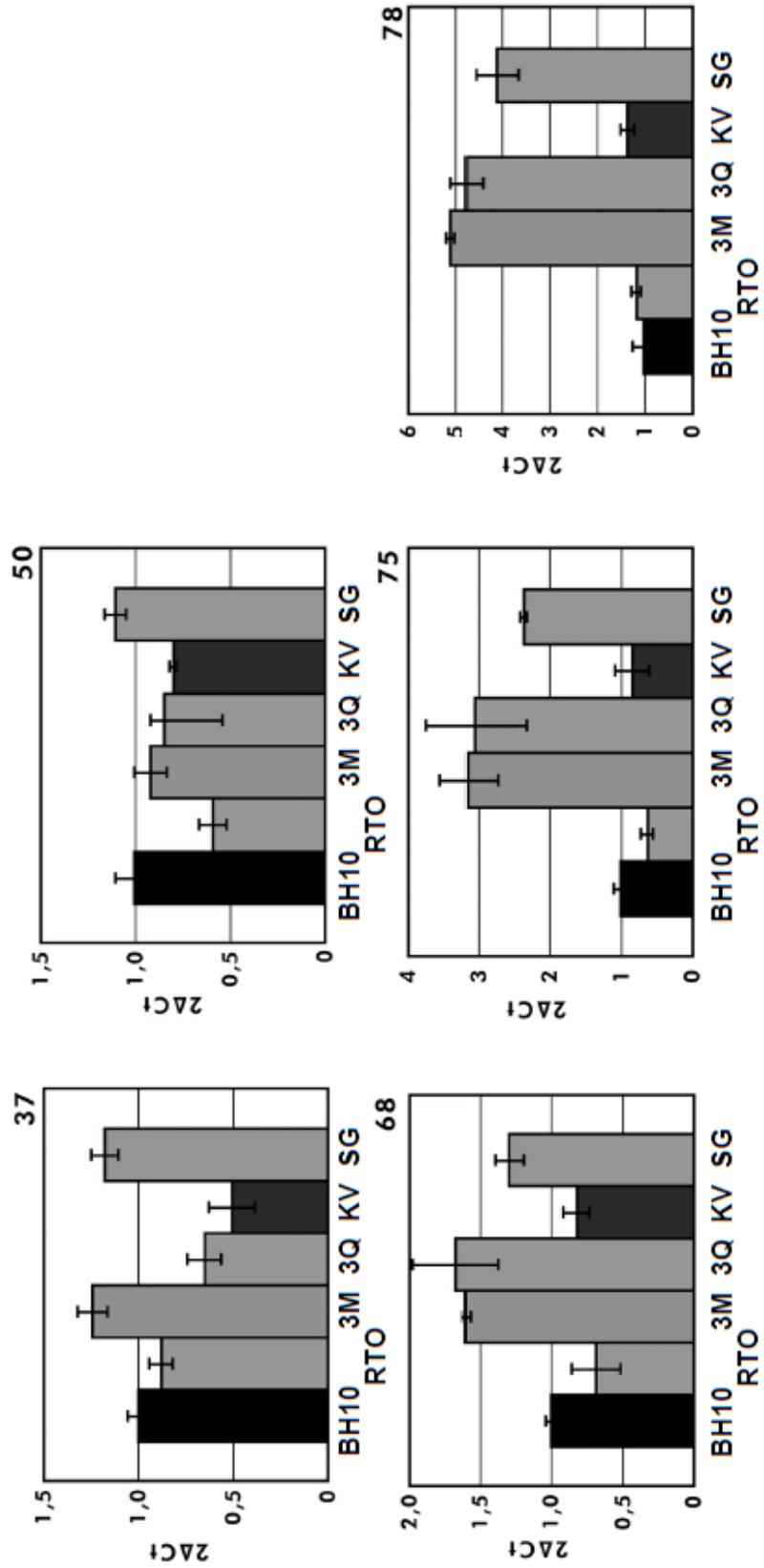


Figura 3

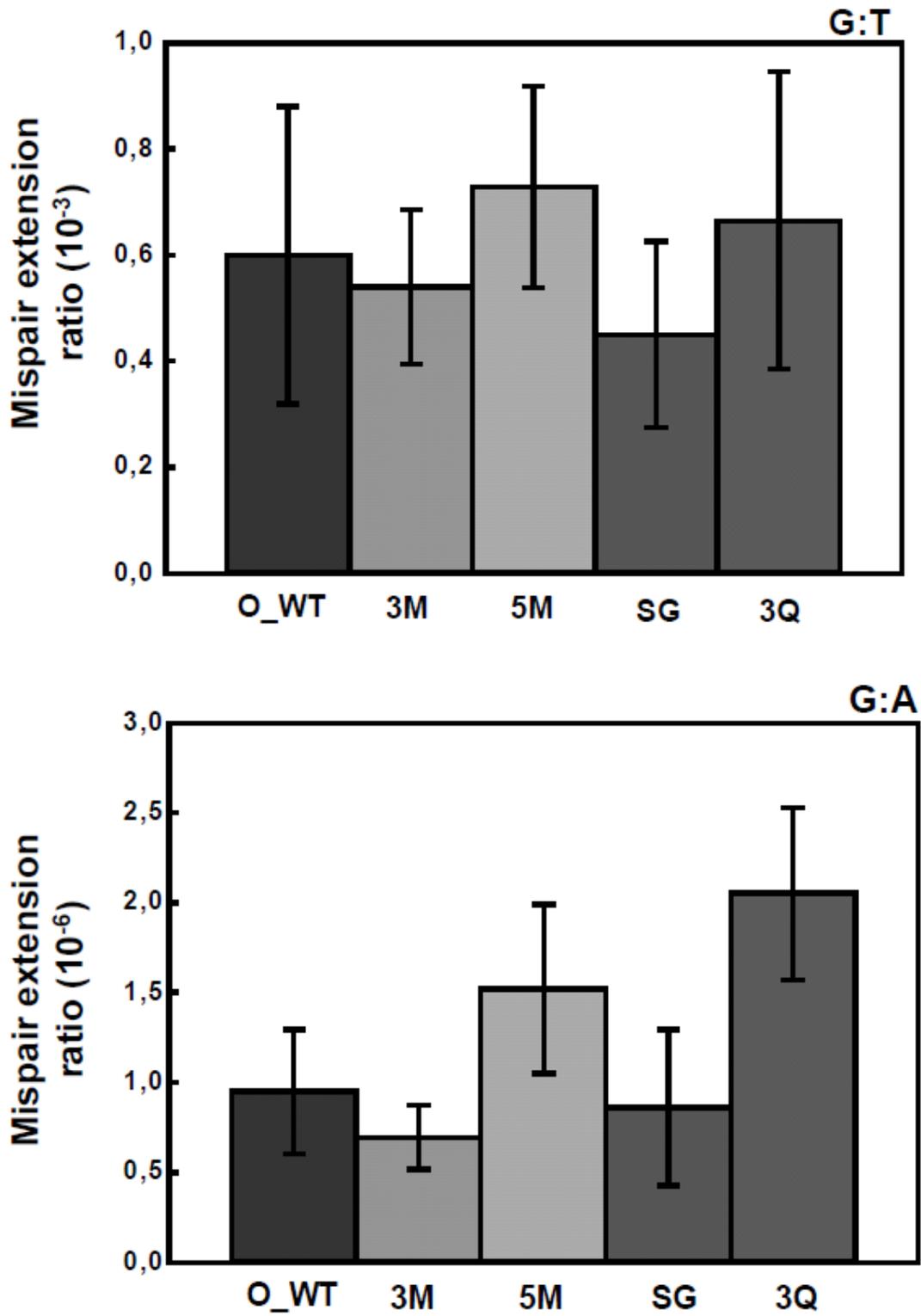
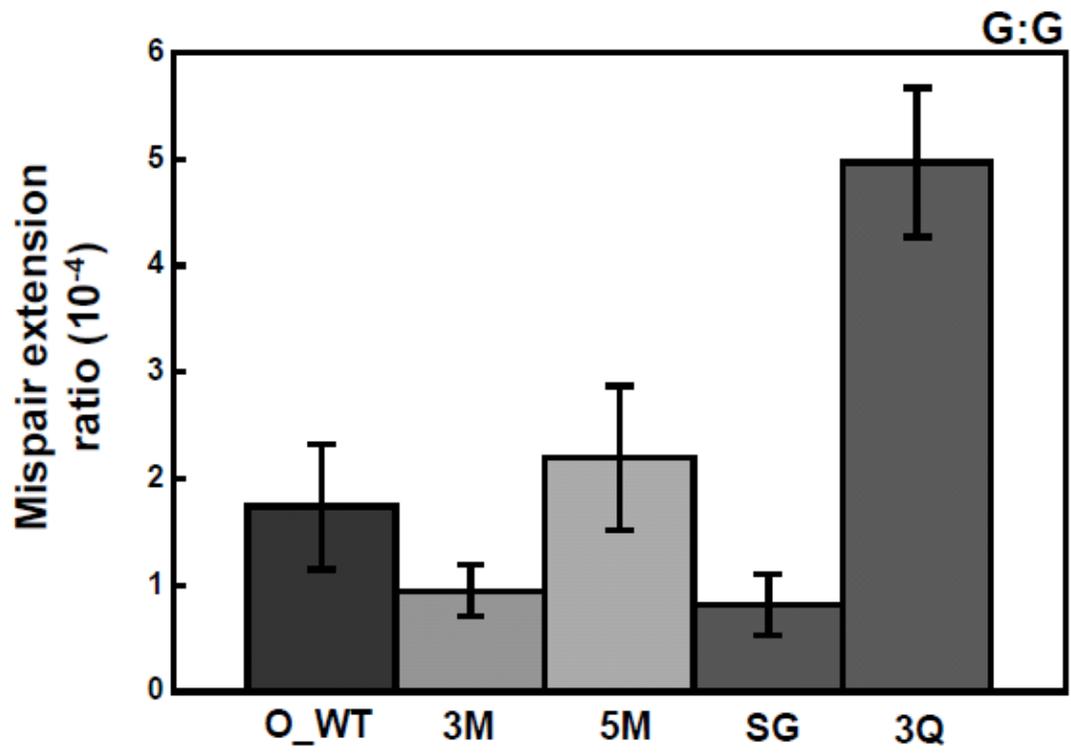


Figura 3 (Cont.)



# ES 2 525 135 A1

## Lista de Secuencias

- <110> Consejo Superior de Investigaciones Científicas
- <120> Retrotranscriptasas del VIH tipo 1 grupo O, activas a temperaturas elevadas
- <130> 2013-0288
- <160> 20
- <170> BiSSAP 1.2
- <210> 1
- <211> 560
- <212> PRT
- <213> HIV-1 group O

<400> 1  
 Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys Pro Gly Met  
 1 5 10 15  
 Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu Lys Ile Glu  
 20 25 30  
 Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly Lys Ile Ser  
 35 40 45  
 Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys  
 50 55 60  
 Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu  
 65 70 75 80  
 Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His  
 85 90 95  
 Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu Asp Val Gly  
 100 105 110  
 Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg Lys Tyr Thr  
 115 120 125  
 Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly Ile Arg Tyr  
 130 135 140  
 Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe  
 145 150 155 160  
 Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys Asp Asn Pro  
 165 170 175  
 Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp  
 180 185 190  
 Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu Arg Glu His  
 195 200 205  
 Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu  
 210 215 220  
 Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr  
 225 230 235 240  
 Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr Val Asn Asp  
 245 250 255  
 Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Gln  
 260 265 270  
 Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly Thr Lys Ser  
 275 280 285  
 Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met Glu Leu Glu  
 290 295 300  
 Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val Tyr Tyr Gln  
 305 310 315 320  
 Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly Glu Gly Gln  
 325 330 335  
 Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu Lys Thr Gly  
 340 345 350  
 Lys Tyr Thr Arg Gln Lys Ala Ser His Thr Asn Asp Ile Arg Gln Leu  
 355 360 365  
 Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val Ile Trp Gly  
 370 375 380  
 Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr Trp Glu Thr  
 385 390 395 400

ES 2 525 135 A1

Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Asp Tyr  
 Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu Glu Ser Glu  
 Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala Ala Asn Arg  
 Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln Gly Lys Gln  
 Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala Glu Leu Met  
 Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val Asn Ile Val  
 Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln Pro Thr Gln  
 Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu Thr Lys Lys  
 Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly Ile Gly Gly  
 Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg Arg Val Leu

<210> 2  
 <211> 1680  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1680  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 2  
 cctataagcc ccatagcccc agtgccagta aagctaaaac caggaatgga tggaccaaaa 60  
 gtaaaacaat ggcccctatc taaagaaaaa atagaagcct taacagcaat atgccaggaa 120  
 atggaacaag aaggaaaaat ttcaagaata ggacctgaaa atccttataa tacacctatc 180  
 tttgctataa aaaagaaaga tagtactaag tggagaaagc tggtagactt tagggaatta 240  
 aacaagagaa cacaagattt ttgggaggtta cagttaggta tcccacatcc ggggggttta 300  
 aagcaaaagc gatctgttac agtcttagat gtaggagatg cctatttctc atgccctta 360  
 gaccagatt ttagaaaata tactgctttc actattccta gtgtaaacia tgagacccca 420  
 ggaataagat accagtacaa tgtcctcccg caaggatgga aaggttcgcc agccatattc 480  
 caaagttcaa tgacaaaaat tttagatcca tttaggaaag acaaccaga attagaaatt 540  
 tgtcagtaca tggatgacct atatgtagga tcagatttac ccctggcaga acatagaaaa 600  
 agggttgaat tgcttagaga acacttatat cagtggggat tcactacccc tgataaaaag 660  
 catcaaaagg aacctccctt tatgtggatg gggtatgagc tccatccaga taaatggaca 720  
 gtacagccca tcaaattgcc taacaaagat gtgtggacag taaatgatat acaaaaacta 780  
 ataggaaagt taaattgggc aagtcaaadc tatcaaggaa ttagagtaag agaattgtgt 840  
 aagttaatta gaggaccaa gtcattaaca gaagtagtac cattaagcaa agaggcagag 900  
 atggaattag aggagaacag agagaaatta aaagaaccaa tgcatggtgt atactatcaa 960  
 cctgataaag acttatgggt taatattcag aagcaaggag aagggcaatg gacttaccag 1020  
 atatatcagg atgaacataa gaacctcaaa acagggaaat atactaggca aaaggcctcc 1080

ES 2 525 135 A1

cacacaaatg acataagaca attagcagaa gtaatccaga aggtgtctca agaatctata 1140  
 gttatctggg gaaaattgcc taaatttaag ctgccagtca ctagagaaac ttgggaaaca 1200  
 tgggtgggcag attattggca agccacctgg atcccagaat gggattatgt cagcacaccc 1260  
 ccattgatca aattatggta ccggttagaa agtgaaccta ttaggggggc agaaacctat 1320  
 tatgtagatg gagcagctaa tagagataca aaattaggaa aagcaggata tgttacagaa 1380  
 caagggaaac agaaaataat aaaattagat gagaccacca atcaaaaggc tgaattaatg 1440  
 gcagtattat tagccttaca ggattccaaa gaaaaagtaa atatagtaac agactcacia 1500  
 tatgtattgg gcattatctc ctcccagcct acacagagtg aatcccctat agttcagcag 1560  
 ataatagagg aactaacaaa aaaggaacag gtgtatctta catgggttcc tgctcataaa 1620  
 ggcataggag gaaatgaaaa aatagataaa ttagtaagca aggatattag aagagtcctc 1680

<210> 3  
 <211> 560  
 <212> PRT  
 <213> HIV-1 group O

<400> 3  
 Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys Pro Gly Met  
 1 5 10 15  
 Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu Lys Ile Glu  
 20 25 30  
 Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly Lys Ile Ser  
 35 40 45  
 Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys  
 50 55 60  
 Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu  
 65 70 75 80  
 Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His  
 85 90 95  
 Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu Asp Val Gly  
 100 105 110  
 Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg Lys Tyr Thr  
 115 120 125  
 Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly Ile Arg Tyr  
 130 135 140  
 Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe  
 145 150 155 160  
 Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys Asp Asn Pro  
 165 170 175  
 Glu Leu Glu Ile Cys Gln Tyr Met Asp Leu Tyr Val Gly Ser Asp  
 180 185 190  
 Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu Arg Glu His  
 195 200 205  
 Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu  
 210 215 220  
 Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr  
 225 230 235 240  
 Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr Val Asn Asp  
 245 250 255  
 Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Gln  
 260 265 270  
 Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly Thr Lys Ser  
 275 280 285  
 Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met Glu Leu Glu  
 290 295 300  
 Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val Tyr Tyr Gln  
 305 310 315 320  
 Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly Glu Gly Gln

ES 2 525 135 A1

Trp	Thr	Tyr	Gln	Ile	Tyr	Gln	Asp	Glu	His	Lys	Asn	Leu	Lys	Thr	Gly
			340					345					350		
Lys	Tyr	Thr	Arg	Gln	Arg	Gly	Ala	His	Thr	Asn	Asp	Ile	Arg	Gln	Leu
		355					360					365			
Ala	Glu	Val	Ile	Gln	Lys	Val	Ser	Gln	Glu	Ser	Ile	Val	Ile	Trp	Gly
	370					375					380				
Lys	Leu	Pro	Lys	Phe	Lys	Leu	Pro	Val	Thr	Arg	Glu	Thr	Trp	Glu	Thr
385					390					395					400
Trp	Trp	Ala	Asp	Tyr	Trp	Gln	Ala	Thr	Trp	Ile	Pro	Glu	Trp	Asp	Tyr
			405						410					415	
Val	Ser	Thr	Pro	Pro	Leu	Ile	Lys	Leu	Trp	Tyr	Arg	Leu	Glu	Ser	Glu
			420					425					430		
Pro	Ile	Arg	Gly	Ala	Glu	Thr	Tyr	Val	Asp	Gly	Ala	Ala	Asn	Arg	
		435					440				445				
Asp	Thr	Lys	Leu	Gly	Lys	Ala	Gly	Tyr	Val	Thr	Glu	Gln	Gly	Lys	Gln
	450					455					460				
Lys	Ile	Ile	Lys	Leu	Asp	Glu	Thr	Thr	Asn	Gln	Lys	Ala	Glu	Leu	Met
465					470					475					480
Ala	Val	Leu	Leu	Ala	Leu	Gln	Asp	Ser	Lys	Glu	Lys	Val	Asn	Ile	Val
				485					490					495	
Thr	Asp	Ser	Gln	Tyr	Val	Leu	Gly	Ile	Ile	Ser	Ser	Gln	Pro	Thr	Gln
			500					505					510		
Ser	Glu	Ser	Pro	Ile	Val	Gln	Gln	Ile	Ile	Glu	Glu	Leu	Thr	Lys	Lys
		515					520					525			
Glu	Gln	Val	Tyr	Leu	Thr	Trp	Val	Pro	Ala	His	Lys	Gly	Ile	Gly	Gly
	530					535					540				
Asn	Glu	Lys	Ile	Asp	Lys	Leu	Val	Ser	Lys	Asp	Ile	Arg	Arg	Val	Leu
545					550					555					560

<210> 4  
 <211> 1680  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1680  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 4  
 cctataagcc ccatagcccc agtgccagta aagctaaaac caggaatgga tggaccaaaa 60  
 gtaaaacaat ggcccctatc taaagaaaa atagaagcct taacagcaat atgccaggaa 120  
 atggaacaag aaggaaaaat ttcaagaata ggacctgaaa atccttataa tacacctatc 180  
 tttgctataa aaaagaaaga tagtactaag tggagaaagc tggtagactt tagggaatta 240  
 aacaagagaa cacaagattt ttgggaggta cagttaggta tcccacatcc ggggggttta 300  
 aagcaaaagc gatctgttac agtcttagat gtaggagatg cctatttctc atgccctta 360  
 gaccagatt ttagaaaata tactgctttc actattccta gtgtaaaca tgagaccca 420  
 ggaataagat accagtacaa tgcctcccg caaggatgga aaggttcgcc agccatattc 480  
 caaagttcaa tgacaaaaat ttagatcca ttaggaaag acaaccaga attagaaatt 540  
 tgtcagtaca tggatgacct atatgtagga tcagatttac ccctggcaga acatagaaaa 600  
 agggttgaat tgcttagaga acacttatat cagtgggat tctactaccc tgataaaaag 660  
 catcaaaagg aacctccctt tatgtggatg gggtatgagc tccatccaga taaatggaca 720  
 gtacagcca tcaaattgcc taacaaagat gtgtggacag taaatgatat acaaaaacta 780

ES 2 525 135 A1

ataggaaagt taaattgggc aagtcaaadc tatcaaggaa ttagagtaag agaattgtgt 840  
aagttaatta gaggaccaa gtcattaaca gaagtagtac cattaagcaa agaggcagag 900  
atggaattag aggagaacag agagaaatta aaagaaccaa tgcattggtgt atactatcaa 960  
cctgataaag acttatgggt taatattcag aagcaaggag aagggcaatg gacttaccag 1020  
atatatcagg atgaacataa gaacctcaaa acagggaaat atactaggca aaggggcgcc 1080  
cacacaaatg acataagaca attagcagaa gtaatccaga aggtgtctca agaactata 1140  
gttatctggg gaaaattgcc taaatttaag ctgccagtca ctagagaaac ttgggaaaca 1200  
tgggtgggcag attattggca agccacctgg atcccagaat gggattatgt cagcacaccc 1260  
ccattgatca aattatggta ccggttagaa agtgaacctt ttaggggggc agaaacctat 1320  
tatgtagatg gagcagctaa tagagataca aaattaggaa aagcaggata tgttacagaa 1380  
caagggaaac agaaaataat aaaattagat gagaccacca atcaaaaggc tgaattaatg 1440  
gcagtattat tagccttaca ggattccaaa gaaaaagtaa atatagtaac agactcacia 1500  
tatgtattgg gcattatctc ctcccagcct acacagagtg aatcccctat agttcagcag 1560  
ataatagagg aactaacaaa aaaggaacag gtgtatctta catgggttcc tgctcataaa 1620  
ggcataggag gaaatgaaaa aatagataaa ttagtaagca aggatattag aagagtcctc 1680

<210> 5  
<211> 560  
<212> PRT  
<213> HIV-1 group O

<400> 5  
Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys Pro Gly Met  
1 5 10 15  
Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu Lys Ile Glu  
20 25 30  
Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly Lys Ile Ser  
35 40 45  
Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys  
50 55 60  
Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu  
65 70 75 80  
Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His  
85 90 95  
Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu Asp Val Gly  
100 105 110  
Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg Lys Tyr Thr  
115 120 125  
Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly Ile Arg Tyr  
130 135 140  
Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe  
145 150 155 160  
Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys Asp Asn Pro  
165 170 175  
Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp  
180 185 190  
Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu Arg Glu His  
195 200 205  
Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu  
210 215 220  
Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr  
225 230 235 240  
Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr Val Asn Asp  
245 250 255

ES 2 525 135 A1

Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Gln  
 260 265 270  
 Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly Thr Lys Ser  
 275 280 285  
 Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met Glu Leu Glu  
 290 295 300  
 Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val Tyr Tyr Gln  
 305 310 315  
 Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly Glu Gly Gln  
 325 330 335  
 Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu Lys Thr Gly  
 340 345 350  
 Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Ile Arg Gln Leu  
 355 360 365  
 Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val Ile Trp Gly  
 370 375 380  
 Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr Trp Glu Thr  
 385 390 395 400  
 Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Asp Tyr  
 405 410 415  
 Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu Glu Ser Glu  
 420 425 430  
 Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala Ala Asn Arg  
 435 440 445  
 Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln Gly Lys Gln  
 450 455 460  
 Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala Glu Leu Met  
 465 470 475 480  
 Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val Asn Ile Val  
 485 490 495  
 Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln Pro Thr Gln  
 500 505 510  
 Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu Thr Lys Lys  
 515 520 525  
 Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly Ile Gly Gly  
 530 535 540  
 Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg Arg Val Leu  
 545 550 555 560

<210> 6  
 <211> 1680  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1680  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 6  
 cctataagcc ccatagcccc agtgccagta aagctaaaac caggaatgga tggaccaaaa 60  
 gtaaaacaat ggcccctatc taaagaaaa atagaagcct taacagcaat atgccaggaa 120  
 atggaacaag aaggaaaaat ttcaagaata ggacctgaaa atccttataa tacacctatc 180  
 tttgctataa aaaagaaaga tagtactaag tggagaaagc tggtagactt tagggaatta 240  
 aacaagagaa cacaagatth ttgggaggta cagttaggta tcccacatcc ggggggttta 300  
 aagcaaaagc gatctgttac agtcttagat gtaggagatg cctatttctc atgccctta 360  
 gaccagatt ttagaaaata tactgcttc actattccta gtgtaaaca tgagaccca 420  
 ggaataagat accagtacaa tgcctccc caaggatgga aaggttcgcc agccatattc 480  
 caaagttcaa tgacaaaaat ttagatcca ttaggaaag acaaccaga attagaaatt 540

ES 2 525 135 A1

tgtcagtaca tggatgacct atatgtagga tcagatttac ccctggcaga acatagaaaa 600  
 agggttgaat tgcttagaga acacttatat cagtggggat tcactacccc tgataaaaag 660  
 catcaaaagg aacctccctt tatgtggatg gggtatgagc tccatccaga taaatggaca 720  
 gtacagccca tcaaattgcc taacaaagat gtgtggacag taaatgatat acaaaaacta 780  
 ataggaaagt taaattgggc aagtcaaatc tatcaaggaa ttagagtaag agaattgtgt 840  
 aagttaatta gaggcaccaa gtcattaaca gaagtagtac cattaagcaa agaggcagag 900  
 atggaattag aggagaacag agagaaatta aaagaaccaa tgcattggtgt atactatcaa 960  
 cctgataaag acttatgggt taatattcag aagcaaggag aaggccaatg gacttaccag 1020  
 atatatcagg atgaacataa gaacctcaaa acagggaaat atgctaggat gaggggcgcc 1080  
 cacacaaatg acataagaca attagcagaa gtaatccaga aggtgtctca agaactata 1140  
 gttatctggg gaaaattgcc taaatttaag ctgccagtca ctagagaaac ttgggaaaca 1200  
 tgggtgggcag attattggca agccacctgg atcccagaat gggattatgt cagcacacc 1260  
 ccattgatca aattatggta ccggttagaa agtgaaccta ttaggggggc agaaacctat 1320  
 tatgtagatg gacagctaa tagagataca aaattaggaa aagcaggata tgttacagaa 1380  
 caagggaac agaaaataat aaaattagat gagaccacca atcaaaaggc tgaattaatg 1440  
 gcagtattat tagccttaca ggattccaaa gaaaaagtaa atatagtaac agactcacia 1500  
 tatgtattgg gcattatctc ctcccagcct acacagagtg aatcccctat agttcagcag 1560  
 ataatagagg aactaacaaa aaaggaacag gtgtatctta catgggttcc tgctcataaa 1620  
 ggcataggag gaaatgaaaa aatagataaa ttagtaagca aggatattag aagagtcctc 1680

<210> 7  
 <211> 562  
 <212> PRT  
 <213> HIV-1 group 0

<400> 7  
 Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys Pro Gly Met  
 1 5 10 15  
 Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu Lys Ile Glu  
 20 25 30  
 Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly Lys Ile Ser  
 35 40 45  
 Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys  
 50 55 60  
 Lys Lys Asp Ser Ser Ser Gly Lys Trp Arg Lys Leu Val Asp Phe Arg  
 65 70 75 80  
 Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile  
 85 90 95  
 Pro His Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu Asp  
 100 105 110  
 Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg Lys  
 115 120 125  
 Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly Ile  
 130 135 140  
 Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala  
 145 150 155 160  
 Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys Asp  
 165 170 175  
 Asn Pro Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val Gly

ES 2 525 135 A1

Ser Asp Leu 180 Pro Leu Ala Glu His 185 Arg Lys Arg Val Glu 190 Leu Leu Arg  
 195 200  
 Glu His Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln  
 210 215 220  
 Lys Glu Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp Lys  
 225 230 235 240  
 Trp Thr Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr Val  
 245 250 255  
 Asn Asp Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln Ile  
 260 265 270  
 Tyr Gln Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly Thr  
 275 280 285  
 Lys Ser Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met Glu  
 290 295 300  
 Leu Glu Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val Tyr  
 305 310 315 320  
 Tyr Gln Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly Glu  
 325 330 335  
 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu Lys  
 340 345 350  
 Thr Gly Lys Tyr Thr Arg Gln Arg Gly Ala His Thr Asn Asp Ile Arg  
 355 360 365  
 Gln Leu Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val Ile  
 370 375 380  
 Trp Gly Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr Trp  
 385 390 395 400  
 Glu Thr Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp  
 405 410 415  
 Asp Tyr Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu Glu  
 420 425 430  
 Ser Glu Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala Ala  
 435 440 445  
 Asn Arg Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln Gly  
 450 455 460  
 Lys Gln Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala Glu  
 465 470 475 480  
 Leu Met Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val Asn  
 485 490 495  
 Ile Val Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln Pro  
 500 505 510  
 Thr Gln Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu Thr  
 515 520 525  
 Lys Lys Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly Ile  
 530 535 540  
 Gly Gly Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg Arg  
 545 550 555 560  
 Val Leu

<210> 8  
 <211> 1686  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1686  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 8  
 cctataagcc ccatagcccc agtgccagta aagctaaaac caggaatgga tggaccaaaa 60  
 gtaaaacaat ggcccctatc taaagaaaaa atagaagcct taacagcaat atgccaggaa 120  
 atggaacaag aaggaaaaat ttcaagaata ggacctgaaa atccttataa tacacctatc 180  
 tttgctataa aaaagaaaga tagtagttcc ggaagtgga gaaagctggt agactttagg 240

ES 2 525 135 A1

gaattaaaca agagaacaca agatTTTTGG gaggtacagt taggtatccc acatccgggg 300  
 ggTTTaaagc aaaagcgatc tgTTacagtc ttagatgtag gagatgccta tttctcatgc 360  
 cccttagacc cagatTTtag aaaatatact gctttcacta ttcctagtgt aaacaatgag 420  
 accccaggaa taagatacca gtacaatgtc ctcccgcaag gatggaaagg ttcgccagcc 480  
 atattccaaa gttcaatgac aaaaatttta gatccattta ggaaagacaa cccagaatta 540  
 gaaatttgtc agtacatgga tgacctatat gtaggatcag atttaccctt gccagaacat 600  
 agaaaaaggg ttgaattgct tagagaacac ttatatcagt ggggattcac taccctgat 660  
 aaaaagcatc aaaaggaacc tccctttatg tggatggggg atgagctcca tccagataaa 720  
 tggacagtac agcccatcaa attgcctaac aaagatgtgt ggacagtaaa tgatatacaa 780  
 aaactaatag gaaagttaa ttgggcaagt caaatctatc aaggaattag agtaagagaa 840  
 ttgtgtaagt taattagagg caccaagtca ttaacagaag tagtaccatt aagcaaagag 900  
 gcagagatgg aattagagga gaacagagag aaattaaaag aaccaatgca tgggtgtatac 960  
 tatcaacctg ataaagactt atgggttaat attcagaagc aaggagaagg gcaatggact 1020  
 taccagatat atcaggatga acataagaac ctcaaaacag ggaaatatac taggcaaagg 1080  
 ggcgcccaca caaatgacat aagacaatta gcagaagtaa tccagaaggt gtctcaagaa 1140  
 tctatagtta tctggggaaa attgcctaaa tttagctgc cagtcactag agaaacttgg 1200  
 gaaacatggt gggcagatta ttggcaagcc acctggatcc cagaatggga ttatgtcagc 1260  
 acacccccat tgatcaaatt atggtaccgg ttagaaagtg aacctattag gggggcagaa 1320  
 acctattatg tagatggagc agctaataga gatacaaat taggaaaagc aggatatggt 1380  
 acagaacaag ggaaacagaa aataataaaa ttagatgaga ccaccaatca aaaggctgaa 1440  
 ttaatggcag tattattagc cttacaggat tccaaagaaa aagtaaata agtaacagac 1500  
 tcacaatatg tattgggcat tatctcctcc cagcctacac agagtgaatc ccctatagtt 1560  
 cagcagataa tagaggaact acaaaaaag gaacaggtgt atcttacatg ggttcctgct 1620  
 cataaaggca taggaggaaa tgaaaaaata gataaattag taagcaagga tattagaaga 1680  
 gtcttc 1686

<210> 9  
 <211> 560  
 <212> PRT  
 <213> HIV-1 group O

<400> 9  
 Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys Pro Gly Met  
 1 5 10 15  
 Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu Lys Ile Glu  
 20 25 30  
 Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly Lys Ile Ser  
 35 40 45  
 Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys  
 50 55 60  
 Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu  
 65 70 75 80  
 Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His  
 85 90 95  
 9

ES 2 525 135 A1

Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu Asp Val Gly  
 100 105 110  
 Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg Lys Tyr Thr  
 115 120 125  
 Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly Ile Arg Tyr  
 130 135 140  
 Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe  
 145 150 155 160  
 Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys Asp Asn Pro  
 165 170 175  
 Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp  
 180 185 190  
 Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu Arg Glu His  
 195 200 205  
 Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu  
 210 215 220  
 Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr  
 225 230 235 240  
 Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr Val Asn Asp  
 245 250 255  
 Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Gln  
 260 265 270  
 Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly Thr Lys Ser  
 275 280 285  
 Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met Glu Leu Glu  
 290 295 300  
 Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val Tyr Tyr Gln  
 305 310 315 320  
 Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly Glu Gly Gln  
 325 330 335  
 Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu Lys Thr Gly  
 340 345 350  
 Lys Tyr Thr Arg Gln Arg Gly Ala His Thr Asn Asp Ile Arg Gln Leu  
 355 360 365  
 Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val Ile Trp Gly  
 370 375 380  
 Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr Trp Glu Thr  
 385 390 395 400  
 Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Asp Tyr  
 405 410 415  
 Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu Glu Ser Glu  
 420 425 430  
 Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala Ala Asn Arg  
 435 440 445  
 Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln Gly Lys Gln  
 450 455 460  
 Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala Gln Leu Met  
 465 470 475 480  
 Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val Asn Ile Val  
 485 490 495  
 Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln Pro Thr Gln  
 500 505 510  
 Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu Thr Lys Lys  
 515 520 525  
 Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly Ile Gly Gly  
 530 535 540  
 Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg Arg Val Leu  
 545 550 555 560

<210> 10  
 <211> 1680  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1680  
 <223> /mol\_type="unassigned DNA"

ES 2 525 135 A1

/organism="HIV-1 group O"

<400> 10  
cctataagcc ccatagcccc agtgccagta aagctaaaac caggaatgga tggaccaaaa 60  
gtaaaacaat ggcccctatc taaagaaaa atagaagcct taacagcaat atgccaggaa 120  
atggaacaag aaggaaaaat ttcaagaata ggacctgaaa atccttataa tacacctatc 180  
tttgctataa aaaagaaaga tagtactaag tggagaaagc tggtagactt tagggaatta 240  
aacaagagaa cacaagatTT ttgggaggta cagttaggta tcccacatcc ggggggttta 300  
aagcaaaagc gatctgttac agtcttagat gtaggagatg cctatttctc atgccctta 360  
gaccagatt ttagaaaata tactgctttc actattccta gtgtaaaca tgagaccca 420  
ggaataagat accagtacaa tgcctcccg caaggatgga aaggttcgcc agccatattc 480  
caaagttcaa tgacaaaaat tttagatcca tttaggaaag acaaccaga attagaaatt 540  
tgtcagtaca tggatgacct atatgtagga tcagatttac ccctggcaga acatagaaaa 600  
agggttgaat tgcttagaga acacttatat cagtggggat tcactacccc tgataaaaag 660  
catcaaaaag aacctccctt tatgtggatg gggatgagc tccatccaga taaatggaca 720  
gtacagcca tcaattgcc taacaagat gtgtggacag taaatgatat acaaaaacta 780  
ataggaaagt taaattggc aagtcaaatc tatcaaggaa ttagagtaag agaattgtgt 840  
aagttaatta gaggcacca gtcattaaca gaagtagtac cattaagcaa agaggcagag 900  
atggaattag aggagaacag agagaaatta aaagaacca tgcatggtgt atactatcaa 960  
cctgataaag acttatgggt taatattcag aagcaaggag aagggaatg gacttaccag 1020  
atatatcagg atgaacataa gaacctcaa acagggaaat atactaggca aagggcgcc 1080  
cacacaaatg acataagaca attagcagaa gtaatccaga aggtgtctca agaactata 1140  
gttatctggg gaaaattgcc taaatttaag ctgccagtca ctagagaaac ttgggaaaca 1200  
tgggtggcag attattggca agccacctgg atcccagaat gggattatgt cagcacacc 1260  
ccattgatca aattatggta ccggttagaa agtgaacctt ttaggggggc agaaacctat 1320  
tatgtagatg gagcagctaa tagagataca aaattaggaa aagcaggata tgttacagaa 1380  
caagggaaac agaaaataat aaaattagat gagaccacca atcaaaaggc tcaattaatg 1440  
gcagtattat tagccttaca ggattccaaa gaaaaagtaa atatagtaac agactcaca 1500  
tatgtattgg gcattatctc ctcccagcct acacagatg aatcccctat agttcagcag 1560  
ataatagagg aactaacaaa aaaggaacag gtgtatctta catgggttcc tgctcataaa 1620  
ggcataggag gaaatgaaaa aatagataaa ttagtaagca aggatattag aagagtcctc 1680

<210> 11  
<211> 572  
<212> PRT  
<213> HIV-1 group O

<400> 11  
Met Asn Ser Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys  
1 5 10 15  
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu  
11

ES 2 525 135 A1

			20					25				30			
Lys	Ile	Glu	Ala	Leu	Thr	Ala	Ile	Cys	Gln	Glu	Met	Glu	Gln	Glu	Gly
		35					40					45			
Lys	Ile	Ser	Arg	Ile	Gly	Pro	Glu	Asn	Pro	Tyr	Asn	Thr	Pro	Ile	Phe
		50				55					60				
Ala	Ile	Lys	Lys	Lys	Asp	Ser	Thr	Lys	Trp	Arg	Lys	Leu	Val	Asp	Phe
65					70					75				80	
Arg	Glu	Leu	Asn	Lys	Arg	Thr	Gln	Asp	Phe	Trp	Glu	Val	Gln	Leu	Gly
				85					90					95	
Ile	Pro	His	Pro	Gly	Gly	Leu	Lys	Gln	Lys	Arg	Ser	Val	Thr	Val	Leu
			100					105					110		
Asp	Val	Gly	Asp	Ala	Tyr	Phe	Ser	Cys	Pro	Leu	Asp	Pro	Asp	Phe	Arg
		115					120					125			
Lys	Tyr	Thr	Ala	Phe	Thr	Ile	Pro	Ser	Val	Asn	Asn	Glu	Thr	Pro	Gly
		130				135					140				
Ile	Arg	Tyr	Gln	Tyr	Asn	Val	Leu	Pro	Gln	Gly	Trp	Lys	Gly	Ser	Pro
145					150					155					160
Ala	Ile	Phe	Gln	Ser	Ser	Met	Thr	Lys	Ile	Leu	Asp	Pro	Phe	Arg	Lys
				165					170					175	
Asp	Asn	Pro	Glu	Leu	Glu	Ile	Cys	Gln	Tyr	Met	Asp	Asp	Leu	Tyr	Val
			180					185					190		
Gly	Ser	Asp	Leu	Pro	Leu	Ala	Glu	His	Arg	Lys	Arg	Val	Glu	Leu	Leu
		195					200					205			
Arg	Glu	His	Leu	Tyr	Gln	Trp	Gly	Phe	Thr	Thr	Pro	Asp	Lys	Lys	His
		210				215					220				
Gln	Lys	Glu	Pro	Pro	Phe	Met	Trp	Met	Gly	Tyr	Glu	Leu	His	Pro	Asp
225					230				235						240
Lys	Trp	Thr	Val	Gln	Pro	Ile	Lys	Leu	Pro	Asn	Lys	Asp	Val	Trp	Thr
				245					250					255	
Val	Asn	Asp	Ile	Gln	Lys	Leu	Ile	Gly	Lys	Leu	Asn	Trp	Ala	Ser	Gln
			260					265					270		
Ile	Tyr	Gln	Gly	Ile	Arg	Val	Arg	Glu	Leu	Cys	Lys	Leu	Ile	Arg	Gly
		275					280					285			
Thr	Lys	Ser	Leu	Thr	Glu	Val	Val	Pro	Leu	Ser	Lys	Glu	Ala	Glu	Met
		290				295					300				
Glu	Leu	Glu	Glu	Asn	Arg	Glu	Lys	Leu	Lys	Glu	Pro	Met	His	Gly	Val
305				310						315					320
Tyr	Tyr	Gln	Pro	Asp	Lys	Asp	Leu	Trp	Val	Asn	Ile	Gln	Lys	Gln	Gly
				325					330					335	
Glu	Gly	Gln	Trp	Thr	Tyr	Gln	Ile	Tyr	Gln	Asp	Glu	His	Lys	Asn	Leu
			340					345					350		
Lys	Thr	Gly	Lys	Tyr	Thr	Arg	Gln	Lys	Ala	Ser	His	Thr	Asn	Asp	Ile
		355					360					365			
Arg	Gln	Leu	Ala	Glu	Val	Ile	Gln	Lys	Val	Ser	Gln	Glu	Ser	Ile	Val
		370				375					380				
Ile	Trp	Gly	Lys	Leu	Pro	Lys	Phe	Lys	Leu	Pro	Val	Thr	Arg	Glu	Thr
385					390					395					400
Trp	Glu	Thr	Trp	Trp	Ala	Asp	Tyr	Trp	Gln	Ala	Thr	Trp	Ile	Pro	Glu
				405					410					415	
Trp	Asp	Tyr	Val	Ser	Thr	Pro	Pro	Leu	Ile	Lys	Leu	Trp	Tyr	Arg	Leu
			420					425					430		
Glu	Ser	Glu	Pro	Ile	Arg	Gly	Ala	Glu	Thr	Tyr	Tyr	Val	Asp	Gly	Ala
		435					440					445			
Ala	Asn	Arg	Asp	Thr	Lys	Leu	Gly	Lys	Ala	Gly	Tyr	Val	Thr	Glu	Gln
		450				455					460				
Gly	Lys	Gln	Lys	Ile	Ile	Lys	Leu	Asp	Glu	Thr	Thr	Asn	Gln	Lys	Ala
465				470					475						480
Glu	Leu	Met	Ala	Val	Leu	Leu	Ala	Leu	Gln	Asp	Ser	Lys	Glu	Lys	Val
				485					490					495	
Asn	Ile	Val	Thr	Asp	Ser	Gln	Tyr	Val	Leu	Gly	Ile	Ile	Ser	Ser	Gln
			500					505					510		
Pro	Thr	Gln	Ser	Glu	Ser	Pro	Ile	Val	Gln	Gln	Ile	Ile	Glu	Glu	Leu
		515					520					525			
Thr	Lys	Lys	Glu	Gln	Val	Tyr	Leu	Thr	Trp	Val	Pro	Ala	His	Lys	Gly
		530				535					540				
Ile	Gly	Gly	Asn	Glu	Lys	Ile	Asp	Lys	Leu	Val	Ser	Lys	Asp	Ile	Arg
545					550					555					560
Arg	Val	Leu	Glu	Ser	Thr	His									
				565					570						

ES 2 525 135 A1

<210> 12  
 <211> 1722  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1722  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

```

<400> 12
atgaattctc ctataagccc catagcccca gtgccagtaa agctaaaacc aggaatggat      60
ggaccaaaaag taaaacaatg gcccctatct aaagaaaaaa tagaagcctt aacagcaata      120
tgccaggaaa tggaacaaga aggaaaaatt tcaagaatag gacctgaaa tccttataat      180
acacctatct ttgctataaa aaagaaagat agtactaagt ggagaaagct ggtagacttt      240
aggaattaa acaagagaac acaagatfff tgggaggtag agttaggtat cccacatccg      300
gggggtttaa agcaaaagcg atctgttaca gtcttagatg taggagatgc ctatttctca      360
tgccccttag acccagatff tagaaaatat actgctttca ctattcctag tgtaacaat      420
gagaccccag gaataagata ccagtacaat gtcctcccgc aaggatggaa aggttcgcca      480
gccatattcc aaagttcaat gacaaaaatt ttagatccat ttaggaaaga caaccagaa      540
ttagaaatff gtcagtacat ggatgaccta tatgtaggat cagatttacc cctggcagaa      600
catagaaaaa gggttgaatt gcttagagaa cacttatatc agtggggatt cactaccct      660
gataaaaagc atcaaaaagga acctccctff atgtggatgg ggtatgagct ccatccagat      720
aatggacag tacagcccat caaattgcct aacaaagatg tgtggacagt aatgatata      780
caaaaactaa taggaaagtt aaattgggca agtcaaactc atcaaggaat tagagtaaga      840
gaattgtgta agttaattag aggaccaag tcattaacag aagtagtacc attaagcaaa      900
gaggcagaga tggaattaga ggagaacaga gagaaattaa aagaaccaat gcatggtgta      960
tactatcaac ctgataaaga cttatgggff aatattcaga agcaaggaga agggcaatgg     1020
acttaccaga tatatcagga tgaacataag aacctcaaaa cagggaata tactaggcaa     1080
aaggcctccc acacaaatga cataagacaa ttagcagaag taatccagaa ggtgtctcaa     1140
gaatctatag ttatctgggg aaaattgcct aaatttaagc tgccagtcac tagagaaact     1200
tgggaaacat ggtgggcaga ttattggcaa gccacctgga tcccagaatg ggattatgfc     1260
agcacacccc cattgatcaa attatggtag cgggttagaaa gtgaacctat taggggggca     1320
gaaacctatt atgtagatgg agcagctaaf agagatacaa aattaggaaa agcaggatat     1380
gttacagaac aagggaaaca gaaaataata aaattagatg agaccaccaa tcaaaaggct     1440
gaattaatgg cagtattatt agccttacag gattccaaag aaaaagtaaa tatagtaaca     1500
gactcacaat atgtattggg cattatctcc tcccagccta cacagagtga atcccctata     1560
gttcagcaga taatagagga actaacaaaa aaggaacagg tgtatcttac atgggttcct     1620
gctcataaag gcataggagg aatgaaaaa atagataaat tagtaagcaa ggatattaga     1680
agagtcctcg agtcgactca ccaccaccac caccactgat aa                          1722
  
```

ES 2 525 135 A1

<210> 13  
 <211> 572  
 <212> PRT  
 <213> HIV-1 group O

<400> 13  
 Met Asn Ser Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys  
 1 5 10 15  
 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu  
 20 25 30  
 Lys Ile Glu Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly  
 35 40 45  
 Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe  
 50 55 60  
 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe  
 65 70 75 80  
 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly  
 85 90 95  
 Ile Pro His Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu  
 100 105 110  
 Asp Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg  
 115 120 125  
 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly  
 130 135 140  
 Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro  
 145 150 155 160  
 Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys  
 165 170 175  
 Asp Asn Pro Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val  
 180 185 190  
 Gly Ser Asp Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu  
 195 200 205  
 Arg Glu His Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His  
 210 215 220  
 Gln Lys Glu Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp  
 225 230 235 240  
 Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr  
 245 250 255  
 Val Asn Asp Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln  
 260 265 270  
 Ile Tyr Gln Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly  
 275 280 285  
 Thr Lys Ser Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met  
 290 295 300  
 Glu Leu Glu Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val  
 305 310 315 320  
 Tyr Tyr Gln Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly  
 325 330 335  
 Glu Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu  
 340 345 350  
 Lys Thr Gly Lys Tyr Thr Arg Gln Arg Gly Ala His Thr Asn Asp Ile  
 355 360 365  
 Arg Gln Leu Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val  
 370 375 380  
 Ile Trp Gly Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr  
 385 390 395 400  
 Trp Glu Thr Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu  
 405 410 415  
 Trp Asp Tyr Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu  
 420 425 430  
 Glu Ser Glu Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala  
 435 440 445  
 Ala Asn Arg Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln  
 450 455 460  
 Gly Lys Gln Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala  
 465 470 475 480  
 Glu Leu Met Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val  
 485 490 495

ES 2 525 135 A1

Asn Ile Val Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln  
 500 505 510  
 Pro Thr Gln Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu  
 515 520 525  
 Thr Lys Lys Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly  
 530 535 540  
 Ile Gly Gly Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg  
 545 550 555 560  
 Arg Val Leu Glu Ser Thr His His His His His His  
 565 570

<210> 14  
 <211> 1722  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1722  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 14  
 atgaattctc ctataagccc catagcccca gtgccagtaa agctaaaacc aggaatggat 60  
 ggaccaaaaag taaaacaatg gcccctatct aaagaaaaaa tagaagcctt aacagcaata 120  
 tgccaggaaa tggaacaaga aggaaaaatt tcaagaatag gacctgaaa tccttataat 180  
 acacctatct ttgctataaa aaagaaagat agtactaagt ggagaaagct ggtagacttt 240  
 agggaattaa acaagagaac acaagatfff tgggaggtac agttaggtat cccacatccg 300  
 gggggtttaa agcaaaagcg atctgttaca gtcttagatg taggagatgc ctatttctca 360  
 tgccccttag acccagattt tagaaaatat actgctttca ctattcctag tgtaacaat 420  
 gagaccccag gaataagata ccagtacaat gtcctcccgc aaggatggaa aggttcgcca 480  
 gccatattcc aaagttcaat gacaaaaatt ttagatccat ttaggaaaga caaccagaa 540  
 ttagaaatff gtcagtacat ggatgaccta tatgtaggat cagatttacc cctggcagaa 600  
 catagaaaaa gggttgaatt gcttagagaa cacttatatc agtggggatt cactaccct 660  
 gataaaaagc atcaaaaagga acctcccttt atgtggatgg ggtatgagct ccatccagat 720  
 aaatggacag tacagcccat caaattgcct aacaaagatg tgtggacagt aaatgatata 780  
 caaaaactaa taggaaagtt aaattgggca agtcaaact atcaaggaat tagagtaaga 840  
 gaattgtgta agttaattag aggcaccaag tcattaacag aagtagtacc attaagcaaa 900  
 gaggcagaga tggaattaga ggagaacaga gagaaattaa aagaaccaat gcatggtgta 960  
 tactatcaac ctgataaaga cttatggggt aatattcaga agcaaggaga agggcaatgg 1020  
 acttaccaga tatatcagga tgaacataag aacctcaaaa cagggaata tactaggcaa 1080  
 aggggcgccc acacaaatga cataagacaa ttagcagaag taatccagaa ggtgtctcaa 1140  
 gaatctatag ttatctgggg aaaattgcct aaatttaagc tgccagtcac tagagaaact 1200  
 tgggaaacat ggtgggcaga ttattggcaa gccacctgga tcccagaatg ggattatgtc 1260  
 agcacacccc cattgatcaa attatggtac cggttagaaa gtgaacctat taggggggca 1320  
 gaaacctatt atgtagatgg agcagctaag agagatacaa aattaggaaa agcaggatat 1380  
 gttacagaac aagggaaaca gaaaataata aaattagatg agaccaccaa tcaaaaggct 1440

ES 2 525 135 A1

gaattaatgg cagtattatt agccttacag gattccaaag aaaaagtaaa tatagtaaca 1500  
gactcacaat atgtattggg cattatctcc tcccagccta cacagagtga atcccctata 1560  
gttcagcaga taatagagga actaacaaaa aaggaacagg tgtatcttac atgggttcct 1620  
gctcataaag gcataggagg aaatgaaaa atagataaat tagtaagcaa ggatattaga 1680  
agagtcctcg agtcgactca ccaccaccac caccactgat aa 1722

<210> 15  
<211> 572  
<212> PRT  
<213> HIV-1 group O

<400> 15  
Met Asn Ser Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys  
1 5 10 15  
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu  
20 25 30  
Lys Ile Glu Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly  
35 40 45  
Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe  
50 55 60  
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe  
65 70 75 80  
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly  
85 90 95  
Ile Pro His Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu  
100 105 110  
Asp Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg  
115 120 125  
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly  
130 135 140  
Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro  
145 150 155 160  
Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys  
165 170 175  
Asp Asn Pro Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val  
180 185 190  
Gly Ser Asp Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu  
195 200 205  
Arg Glu His Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His  
210 215 220  
Gln Lys Glu Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp  
225 230 235 240  
Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val Trp Thr  
245 250 255  
Val Asn Asp Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala Ser Gln  
260 265 270  
Ile Tyr Gln Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile Arg Gly  
275 280 285  
Thr Lys Ser Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala Glu Met  
290 295 300  
Glu Leu Glu Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His Gly Val  
305 310 315 320  
Tyr Tyr Gln Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys Gln Gly  
325 330 335  
Glu Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys Asn Leu  
340 345 350  
Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Ile  
355 360 365  
Arg Gln Leu Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser Ile Val  
370 375 380  
Ile Trp Gly Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg Glu Thr  
385 390 395 400  
Trp Glu Thr Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu  
16

ES 2 525 135 A1

```

          405                      410                      415
Trp Asp Tyr Val Ser Thr Pro Pro Leu Ile Lys Leu Trp Tyr Arg Leu
          420                      425                      430
Glu Ser Glu Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp Gly Ala
          435                      440                      445
Ala Asn Arg Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Glu Gln
          450                      455                      460
Gly Lys Gln Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln Lys Ala
465                      470                      475
Glu Leu Met Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu Lys Val
          485                      490                      495
Asn Ile Val Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser Ser Gln
          500                      505                      510
Pro Thr Gln Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu Glu Leu
          515                      520                      525
Thr Lys Lys Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His Lys Gly
          530                      535                      540
Ile Gly Gly Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp Ile Arg
545                      550                      555                      560
Arg Val Leu Glu Ser Thr His His His His His His
          565                      570

```

<210> 16  
 <211> 1722  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1722  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

```

<400> 16
atgaattctc ctataagccc catagcccca gtgccagtaa agctaaaacc agaatggat      60
ggacaaaaag taaaacaatg gccctatct aaagaaaaaa tagaagcctt aacagcaata      120
tgccaggaaa tggaacaaga aggaaaaatt tcaagaatag gacctgaaaa tccttataat      180
acacctatct ttgctataaa aaagaaagat agtactaagt ggagaaagct ggtagacttt      240
aggaattaa acaagagaac acaagatttt tgggaggtag agttaggtag cccacatccg      300
gggggtttaa agcaaaagcg atctgttaca gtcttagatg taggagatgc ctatttctca      360
tgccccttag acccagattt tagaaaatat actgctttca ctattcctag tgtaacaat      420
gagaccccag gaataagata ccagtacaat gtcctcccg aaggatggaa aggttcgcca      480
gccatattcc aaagttcaat gacaaaaatt ttagatccat ttaggaaaga caaccagaa      540
ttagaaattt gtcagtacat ggatgaccta tatgtaggat cagatttacc cctggcagaa      600
catagaaaaa gggttgaatt gcttagagaa cacttatatc agtggggatt cactaccct      660
gataaaaagc atcaaaaagga acctcccttt atgtggatgg ggtatgagct ccatccagat      720
aatggacag tacagcccat caaattgcct aacaaagatg tgtggacagt aatgatata      780
caaaaactaa taggaaagtt aaattgggca agtcaaact atcaaggaat tagagtaaga      840
gaattgtgta agttaattag aggcaccaag tcattaacag aagtagtacc attaagcaaa      900
gaggcagaga tggaattaga ggagaacaga gagaaattaa aagaaccaat gcatggtgta      960
tactatcaac ctgataaaga cttatgggtt aatattcaga agcaaggaga agggcaatgg     1020
acttaccaga tatatcagga tgaacataag aacctcaaaa cagggaata tgctaggatg     1080

```

ES 2 525 135 A1

aggggcgccc acacaaatga cataagacaa ttagcagaag taatccagaa ggtgtctcaa 1140  
 gaatctatag ttatctgggg aaaattgcct aaatttaagc tgccagtcac tagagaaact 1200  
 tgggaaacat ggtgggcaga ttattggcaa gccacctgga tcccagaatg ggattatgtc 1260  
 agcacacccc cattgatcaa attatggtac cggttagaaa gtgaacctat taggggggca 1320  
 gaaacctatt atgtagatgg agcagctaata agagatacaa aattaggaaa agcaggatat 1380  
 gttacagaac aagggaaaca gaaaataata aaattagatg agaccaccaa tcaaaggct 1440  
 gaattaatgg cagtattatt agccttacag gattccaaag aaaaagtaaa tatagtaaca 1500  
 gactcacaat atgtattggg cattatctcc tcccagccta cacagagtga atcccctata 1560  
 gttcagcaga taatagagga actaacaaaa aaggaacagg tgtatcttac atgggttcct 1620  
 gctcataaag gcataggagg aaatgaaaaa atagataaat tagtaagcaa ggatattaga 1680  
 agagtcctcg agtcgactca ccaccaccac caccactgat aa 1722

<210> 17  
 <211> 574  
 <212> PRT  
 <213> HIV-1 group O

<400> 17  
 Met Asn Ser Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys  
 1 5 10 15  
 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu  
 20 25 30  
 Lys Ile Glu Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly  
 35 40 45  
 Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe  
 50 55 60  
 Ala Ile Lys Lys Lys Asp Ser Ser Ser Gly Lys Trp Arg Lys Leu Val  
 65 70 75 80  
 Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln  
 85 90 95  
 Leu Gly Ile Pro His Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr  
 100 105 110  
 Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp  
 115 120 125  
 Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr  
 130 135 140  
 Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly  
 145 150 155 160  
 Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe  
 165 170 175  
 Arg Lys Asp Asn Pro Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu  
 180 185 190  
 Tyr Val Gly Ser Asp Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu  
 195 200 205  
 Leu Leu Arg Glu His Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys  
 210 215 220  
 Lys His Gln Lys Glu Pro Phe Met Trp Met Gly Tyr Glu Leu His  
 225 230 235 240  
 Pro Asp Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Asn Lys Asp Val  
 245 250 255  
 Trp Thr Val Asn Asp Ile Gln Lys Leu Ile Gly Lys Leu Asn Trp Ala  
 260 265 270  
 Ser Gln Ile Tyr Gln Gly Ile Arg Val Arg Glu Leu Cys Lys Leu Ile  
 275 280 285  
 Arg Gly Thr Lys Ser Leu Thr Glu Val Val Pro Leu Ser Lys Glu Ala  
 290 295 300  
 Glu Met Glu Leu Glu Glu Asn Arg Glu Lys Leu Lys Glu Pro Met His  
 305 310 315 320

ES 2 525 135 A1

Gly Val Tyr Tyr Gln Pro Asp Lys Asp Leu Trp Val Asn Ile Gln Lys  
 325 330  
 Gln Gly Glu Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Asp Glu His Lys  
 340 345 350  
 Asn Leu Lys Thr Gly Lys Tyr Thr Arg Gln Arg Gly Ala His Thr Asn  
 355 360 365  
 Asp Ile Arg Gln Leu Ala Glu Val Ile Gln Lys Val Ser Gln Glu Ser  
 370 375 380  
 Ile Val Ile Trp Gly Lys Leu Pro Lys Phe Lys Leu Pro Val Thr Arg  
 385 390 395 400  
 Glu Thr Trp Glu Thr Trp Trp Ala Asp Tyr Trp Gln Ala Thr Trp Ile  
 405 410 415  
 Pro Glu Trp Asp Tyr Val Ser Thr Pro Leu Ile Lys Leu Trp Tyr  
 420 425 430  
 Arg Leu Glu Ser Glu Pro Ile Arg Gly Ala Glu Thr Tyr Tyr Val Asp  
 435 440 445  
 Gly Ala Ala Asn Arg Asp Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr  
 450 455 460  
 Glu Gln Gly Lys Gln Lys Ile Ile Lys Leu Asp Glu Thr Thr Asn Gln  
 465 470 475 480  
 Lys Ala Glu Leu Met Ala Val Leu Leu Ala Leu Gln Asp Ser Lys Glu  
 485 490 495  
 Lys Val Asn Ile Val Thr Asp Ser Gln Tyr Val Leu Gly Ile Ile Ser  
 500 505 510  
 Ser Gln Pro Thr Gln Ser Glu Ser Pro Ile Val Gln Gln Ile Ile Glu  
 515 520 525  
 Glu Leu Thr Lys Lys Glu Gln Val Tyr Leu Thr Trp Val Pro Ala His  
 530 535 540  
 Lys Gly Ile Gly Gly Asn Glu Lys Ile Asp Lys Leu Val Ser Lys Asp  
 545 550 555 560  
 Ile Arg Arg Val Leu Glu Ser Thr His His His His His  
 565 570

<210> 18  
 <211> 1728  
 <212> DNA  
 <213> HIV-1 group O

<220>  
 <221> source  
 <222> 1..1728  
 <223> /mol\_type="unassigned DNA"  
 /organism="HIV-1 group O"

<400> 18  
 atgaattctc ctataagccc catagcccca gtgccagtaa agctaaaacc aggaatggat 60  
 ggaccaaaaag taaaacaatg gcccttatct aaagaaaaaa tagaagcctt aacagcaata 120  
 tgccaggaaa tggaacaaga aggaaaaatt tcaagaatag gacctgaaa tccttataat 180  
 acacctatct ttgctataaa aaagaaagat agtagttccg ggaagtggag aaagctggta 240  
 gactttaggg aattaaacaa gagaacacaa gatttttggg aggtacagtt aggtatccca 300  
 catccggggg gtttaaagca aaagcgatct gttacagtct tagatgtagg agatgcctat 360  
 ttctcatgcc ccttagacc agattttaga aaatatactg ctttactat tcctagtgta 420  
 aacaatgaga ccccaggaat aagataccag tacaatgtcc tcccgcaagg atggaaaggt 480  
 tcgccagcca tattccaaag ttcaatgaca aaaattttag atccatttag gaaagacaac 540  
 ccagaattag aaatttgtca gtacatggat gacctatatg taggatcaga tttaccctg 600  
 gcagaacata gaaaaagggt tgaattgctt agagaacact tatatcagtg gggattcact 660  
 acccctgata aaaagcatca aaaggaacct ccctttatgt ggatggggta tgagctccat 720  
 ccagataaat ggacagtaca gcccatcaaa ttgcctaaca aagatgtgtg gacagtaaat 780

ES 2 525 135 A1

gatatacaaa aactaatagg aaagttaa at tgggcaagtc aaatctatca aggaattaga 840  
 gtaagagaat tgtgtaagtt aattagaggc accaagtcat taacagaagt agtaccatta 900  
 agcaaagagg cagagatgga attagaggag aacagagaga aattaaaga accaatgcat 960  
 ggtgtatact atcaacctga taaagactta tgggttaata ttcagaagca aggagaaggg 1020  
 caatggactt accagatata tcaggatgaa cataagaacc tcaaacagg gaaatatact 1080  
 aggcaaaggg gcgcccacac aatgacata agacaattag cagaagtaat ccagaaggtg 1140  
 tctcaagaat ctatagttat ctggggaaaa ttgcctaaat ttaagctgcc agtcactaga 1200  
 gaaacttggg aaacatggtg ggcagattat tggcaagcca cctggatccc agaatgggat 1260  
 tatgtcagca caccaccatt gatcaaatta tggtagcggg tagaaagtga acctattagg 1320  
 ggggcagaaa cctattatgt agatggagca gctaataagag atacaaaatt aggaaaagca 1380  
 ggatatgta cagaacaagg gaaacagaaa ataataaaat tagatgagac caccaatcaa 1440  
 aaggctgaat taatggcagt attattagcc ttacaggatt ccaagaaaa agtaaatata 1500  
 gtaacagact cacaatatgt attgggcatt atctcctccc agcctacaca gagtgaatcc 1560  
 cctatagttc agcagataat agaggaacta acaaaaaagg aacaggtgta tcttacatgg 1620  
 gttcctgctc ataaaggcat aggaggaaat gaaaaaatag ataaattagt aagcaaggat 1680  
 attagaagag tcctcgagtc gactcaccac caccaccacc actgataa 1728

<210> 19  
 <211> 572  
 <212> PRT  
 <213> HIV-1 group O

<400> 19  
 Met Asn Ser Pro Ile Ser Pro Ile Ala Pro Val Pro Val Lys Leu Lys  
 1 5 10 15  
 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Ser Lys Glu  
 20 25 30  
 Lys Ile Glu Ala Leu Thr Ala Ile Cys Gln Glu Met Glu Gln Glu Gly  
 35 40 45  
 Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe  
 50 55 60  
 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe  
 65 70 75 80  
 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly  
 85 90 95  
 Ile Pro His Pro Gly Gly Leu Lys Gln Lys Arg Ser Val Thr Val Leu  
 100 105 110  
 Asp Val Gly Asp Ala Tyr Phe Ser Cys Pro Leu Asp Pro Asp Phe Arg  
 115 120 125  
 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Val Asn Asn Glu Thr Pro Gly  
 130 135 140  
 Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro  
 145 150 155 160  
 Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Asp Pro Phe Arg Lys  
 165 170 175  
 Asp Asn Pro Glu Leu Glu Ile Cys Gln Tyr Met Asp Asp Leu Tyr Val  
 180 185 190  
 Gly Ser Asp Leu Pro Leu Ala Glu His Arg Lys Arg Val Glu Leu Leu  
 195 200 205  
 Arg Glu His Leu Tyr Gln Trp Gly Phe Thr Thr Pro Asp Lys Lys His  
 210 215 220  
 Gln Lys Glu Pro Pro Phe Met Trp Met Gly Tyr Glu Leu His Pro Asp  
 20



ES 2 525 135 A1

gagaccccag gaataagata ccagtacaat gtcctcccgc aaggatggaa aggttcgcca 480  
 gccatattcc aaagttcaat gacaaaaatt ttagatccat ttaggaaaga caaccagaa 540  
 ttagaaatth gtcagtacat ggatgaccta tatgtaggat cagatttacc cctggcagaa 600  
 catagaaaaa gggttgaatt gcttagagaa cacttatatc agtggggatt cactaccct 660  
 gataaaaagc atcaaaagga acctcccttt atgtggatgg ggtatgagct ccatccagat 720  
 aaatggacag tacagcccat caaattgcct aacaaagatg tgtggacagt aaatgatata 780  
 caaaaactaa taggaaagtt aaattgggca agtcaaact atcaaggaat tagagtaaga 840  
 gaattgtgta agttaattag aggcaccaag tcattaacag aagtagtacc attaagcaaa 900  
 gaggcagaga tggaattaga ggagaacaga gagaaattaa agaaccaat gcatggtgta 960  
 tactatcaac ctgataaaga cttatggggtt aatattcaga agcaaggaga agggcaatgg 1020  
 acttaccaga tatatcagga tgaacataag aacctcaaaa cagggaaata tactaggcaa 1080  
 aggggcgccc acacaaatga cataagacaa ttagcagaag taatccagaa ggtgtctcaa 1140  
 gaatctatag ttatctgggg aaaattgcct aaatttaagc tgccagtcac tagagaaact 1200  
 tgggaaacat ggtgggcaga ttattggcaa gccacctgga tcccagaatg ggattatgtc 1260  
 agcacacccc cattgatcaa attatggtac cggttagaaa gtgaacctat taggggggca 1320  
 gaaacctatt atgtagatgg agcagctaag agagatacaa aattaggaaa agcaggatat 1380  
 gttacagaac aagggaaaca gaaaataata aaattagatg agaccaccaa tcaaaaggct 1440  
 caattaatgg cagtattatt agccttacag gattccaaag aaaaagtaaa tatagtaaca 1500  
 gactcacaat atgtattggg cattatctcc tcccagccta cacagagtga atcccctata 1560  
 gttcagcaga taatagagga actaacaaaa aaggaacagg tgtatcttac atgggttcct 1620  
 gctcataaag gcataggagg aaatgaaaaa atagataaat tagtaagcaa ggatattaga 1680  
 agagtcctcg agtcgactca ccaccaccac caccactgat aa 1722