

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 795 287**

51 Int. Cl.:

C12N 9/02 (2006.01)

C12Q 1/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.05.2010** E 18184699 (9)

97 Fecha y número de publicación de la concesión europea: **15.04.2020** EP 3409764

54 Título: **Luciferasas de *Oplophorus* sintéticas con mayor emisión de luz**

30 Prioridad:

01.05.2009 US 174838 P

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

23.11.2020

73 Titular/es:

**PROMEGA CORPORATION (100.0%)
2800 Woods Hollow Road
Madison, WI 53711-5399, US**

72 Inventor/es:

**ENCELL, LANCE P.;
WOOD, KEITH V.;
WOOD, MONIKA G.;
HALL, MARY P.;
OTTO, PAUL;
VIDUGIRIS, GEDIMINAS y
ZIMMERMAN, KRISTOPHER**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 795 287 T3

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

DESCRIPCIÓN

Luciferasas de *Oplophorus* sintéticas con mayor emisión de luz

5 **Referencia cruzada a solicitudes relacionadas**

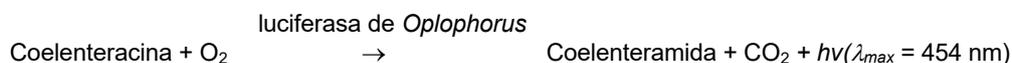
Esta solicitud reivindica la prioridad de la solicitud provisional de EEUU n.º 61/174.838, presentada el 1 de mayo de 2009.

10 **Antecedentes**

La presente invención se refiere a luciferasas de *Oplophorus* sintéticas que tienen mejores propiedades, comparado con la luciferasa de *Oplophorus* de tipo salvaje.

15 El camarón abisal *Oplophorus gracilirostris* expulsa una nube luminosa azul desde la base de sus antenas cuando se estimula, al igual que diversos otros camarones decápodos luminiscentes que incluyen los géneros *Heterocarpus*, *Systemaspis* y *Acanthephyra* (Herring, *J. Mar. Biol. Assoc. UK.*, 156:1029 (1976)). El mecanismo subyacente a la luminiscencia de *Oplophorus* implica la oxidación de la luciferina de *Oplophorus* (coelenteracina) con oxígeno molecular, que es catalizada por la luciferasa de *Oplophorus* como sigue:

20



25 La coelenteracina, un compuesto de imidazopirazinona, está implicada en la bioluminiscencia de una amplia diversidad de organismos como una luciferina o como el resto funcional de fotoproteínas. Por ejemplo, la luciferina del pensamiento de mar *Renilla* es la coelenteracina (Inoue et al., *Tetrahed. Lett.*, 18:2685 (1977)), y la fotoproteína sensible al calcio aequorina de la medusa *Aequorea* también contiene coelenteracina como resto funcional (Shimomura et al., *Biochem.*, 17:994 (1978); Head et al., *Nature.*, 405:372 (2000)).

30 INOUE et al., *FEBS LETT.*, 481(1):19-25 (2000), se refiere a la clonación de ADNc de una luciferasa 1 de imidazopirazinona secretiva del camarón luminoso *Oplophorus gracilirostris*.

FUJII et al., *Analytical Biochemistry*, 366(2):131-136 (2007), se refiere a un aumento en la intensidad de la bioluminiscencia de la luciferasa de luciérnaga empleando una modificación genética.

35 LOENING et al., *Protein Engineering, Design and Selection* 19(9): 391-400 se refiere a la mutagénesis guiada consenso de la luciferasa de *Renilla* con el fin de proporcionar una estabilidad y emisión de luz mejoradas.

40 El documento WO9518853 se refiere a mutantes no naturales activos de luciferasas de escarabajo y a ADN que codifica dichos mutantes, en el que una luciferasa mutante se diferencia de la luciferasa de tipo salvaje correspondiente por la producción de bioluminiscencia con una longitud de onda con una intensidad de pico que se diferencia al menos en 1 nm de la longitud de onda con una intensidad de pico de la bioluminiscencia producida por la enzima de tipo salvaje. Las luciferasas mutantes y los ADN del documento WO9518853 se emplean en diversas aplicaciones de biosensores.

45

Sumario

La invención proporciona un método que comprende: (a) expresar un polipéptido de luciferasa en una célula, en el que el polipéptido de luciferasa es una luciferasa modificada con al menos 60% de identidad de secuencia de aminoácidos con una luciferasa de *Oplophorus* de tipo salvaje, y que comprende al menos una sustitución de un aminoácido en una posición seleccionada de las posiciones 2, 4, 11, 20, 23, 28, 33, 34, 44, 45, 51, 54, 68, 72, 75, 76, 77, 89, 90, 92, 99, 104, 115, 124, 135, 138, 139, 143, 144, 164, 166, 167 o 169 correspondientes a SEQ ID NO:1, en el que la luciferasa modificada tiene al menos una de luminiscencia mayor, estabilidad de la señal mayor y estabilidad de la proteína mayor con relación a la correspondiente luciferasa de *Oplophorus* de tipo salvaje; (b) exponer dicho polipéptido de luciferasa a un sustrato; y (c) detectar luminiscencia.

55

Otros aspectos de la invención se definen en las reivindicaciones.

Breve descripción de los dibujos

60

La figura 1 muestra alineamientos de la estructura secundaria de proteínas de unión a ácidos grasos (FABP) y OgLuc.

La figura 2 muestra alineamientos de la estructura secundaria de la luciferasa de dinoflagelados, FABP y OgLuc.

65 La figura 3 muestra un alineamiento de las secuencias de aminoácidos de OgLuc y diversas FABP (SEQ ID NO:1, 3, 4, 5 y 17-20, respectivamente) basándose en la superposición de las estructuras tridimensionales de FABP.

- 5 Las figuras 4A-D muestran la emisión de luz (es decir, la luminiscencia) a lo largo del tiempo de variantes de OgLuc modificadas con una combinación de dos o más sustituciones de aminoácidos en OgLuc, comparado con la variante de OgLuc N166R y la luciferasa de *Renilla*. 4A-4B) Luminiscencia (“lum”) en unidades relativas de luz (RLU) empleando un ensayo de luminiscencia “Destello” (“Flash”) mostrado en dos escalas diferentes de luminiscencia a lo largo del tiempo en minutos. 4C-4D) Luminiscencia (“lum”) en RLU empleando un ensayo de luminiscencia de tergitol al 0,5% “Glo”, mostrada en dos escalas diferentes de luminiscencia a lo largo del tiempo en minutos.
- 10 Las figuras 5A-C resumen el promedio de luminiscencia en RLU de las diversas variantes de OgLuc descritas en el ejemplo 7 (“Muestra”) en T = 0 (“Promedio”) con la desviación estándar (“DesvEst”) y el coeficiente de varianza (“CV”) comparado con OgLuc WT (de tipo salvaje) empleando un tampón de ensayo de tergitol al 0,5%.
- 15 Las figuras 6A-B resumen el aumento en número de veces en la luminiscencia en T = 0 de las variantes de OgLuc frente a OgLuc WT determinada a partir de los datos del tampón de ensayo de tergitol al 0,5% mostrados en las figuras 5A-C.
- 20 Las figuras 7A-C resumen el promedio de luminiscencia en RLU de las variantes de OgLuc (“Muestra”) en T = 0 (“Promedio”) con la desviación estándar (“DesvEst”) y el coeficiente de varianza (“CV”) comparado con OgLuc WT (de tipo salvaje) empleando RLAB.
- La figura 8 resume el aumento en número de veces en la luminiscencia en T = 0 de las variantes de OgLu frente a OgLuc WT determinada a partir de los datos de RLAB mostrados en las figuras 7A-C.
- 25 Las figuras 9A-D muestran la estabilidad de la señal de las variantes de OgLuc comparado con OgLuc WT empleando un tampón de ensayo de tergitol al 0,5%. 9A-9C) Emisión de luz a lo largo del tiempo de las variantes de OgLuc (“clon”), midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos. 9D) Semivida de la señal en minutos de las variantes de OgLuc determinada a partir de los datos de emisión de luz a lo largo del tiempo mostrados en las figuras 9A-C.
- 30 Las figuras 10A-C muestran la emisión de luz a lo largo del tiempo (es decir, la estabilidad de la señal) de las variantes de OgLuc comparado con OgLu WT, empleando RLAB, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos.
- 35 Las figuras 11A-B muestran la semivida de la señal en minutos de los variantes de OgLuc comparado con OgLuc WT determinada a partir de los datos de emisión de luz a lo largo del tiempo mostrados en las figuras 10A-C.
- Las figuras 12A-B muestran la estabilidad de las proteínas a 22°C como la semivida en minutos de las variantes de OgLuc, comparado con OgLuc WT.
- 40 Las figuras 13A-B resumen el promedio de luminiscencia en RLU de las variantes de OgLuc A33K y F68Y en T = 0 (“Promedio”) con el coeficiente de varianza (“% cv”), comparado con OgLuc WT, empleando el tampón de ensayo de tergitol al 0,5% (13A) o RLAB (13B).
- 45 Las figuras 14A-B resumen el aumento en número de veces en la luminiscencia a T = 0 de las variantes de OgLuc A33K y F68Y frente a OgLuc WT, determinado a partir de los datos mostrados en las figuras 13A-B para ensayos que emplean el tampón de ensayo de tergitol al 0,5% (14A) o RLAB (14B), respectivamente.
- 50 Las figuras 15A-B muestran la estabilidad de la señal de los variantes de OgLuc A33K y F68Y comparado con OgLuc WT empleando el tampón de ensayo de tergitol al 0,5%. 15A) Emisión de luz a lo largo del tiempo de las variantes de OgLuc A33K y F68Y, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos. 15B) Semivida de la señal en minutos de las variantes de OgLuc A33K y F68Y determinada a partir de los datos de emisión de luz a lo largo del tiempo mostrados en las figuras 15A.
- 55 Las figuras 16A-B muestran la estabilidad de la señal de las variantes de OgLuc A33K y F68Y comparado con OgLuc WT empleando RLAB. 16A) Emisión de luz a lo largo del tiempo de las variantes de OgLuc A33K y F68Y, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos. 16B) Semivida de la señal en minutos de las variantes de OgLuc A33K y F68Y determinada a partir de los datos de emisión de luz a lo largo del tiempo mostrados en las figuras 16A.
- 60 La figura 17 muestra la estabilidad de las proteínas a 22°C como la semivida en minutos de las variantes de OgLuc A33K y F68Y.
- 65 Las figuras 18A-B muestran la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de la combinación central de OgLuc, comparado con la variante de OgLuc N166R y la luciferasa de *Renilla* empleando el tampón de ensayo de tergitol al 0,5%, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos.

La figura 19 muestra la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de combinación central de OgLuc, comparado con la variante de OgLuc N166R y la luciferasa de *Renilla* empleando RLAB, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos.

5 Las figuras 20A-B muestran la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de OgLuc C1+C2+A4E y C1+A4E, comparado con OgLuc WT ("Og-Luc") y la luciferasa de *Renilla* ("hRL"), y las variantes T2T y A54F, empleando el tampón de ensayo de tergitol al 0,5% (20A) o RLAB (20B), midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos.

10 La figura 21 muestra la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de OgLuc C1+C2+A4E y C1+A4E, comparado con OgLuc WT ("Og-Luc") y la luciferasa de *Renilla* ("hRL"), y las variantes T2T y A54F, empleando el tampón de ensayo de tergitol al 0,25%, midiéndose la luminiscencia en RLU a lo largo del tiempo en minutos.

15 La figura 22 muestra la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de OgLuc C1+C2+A4E y C1+A4E, comparado con OgLuc WT ("Og-Luc") y la luciferasa de *Renilla* ("hRL"), y las variantes T2T y A54F, en células HEK 293 con tampón RLAB, normalizado a la luciérnaga.

20 La figura 23 muestra la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de OgLuc C1+C2+A4E y C1+A4E, comparado con OgLuc WT ("Og-Luc") y la luciferasa de *Renilla* ("hRL"), en células HEK 293 empleando el tampón de tergitol al 0,25%, normalizado a la luciérnaga.

25 La figura 24 muestra la estabilidad de las proteínas como la semivida en minutos de las variantes de OgLuc C1, C1+A4E, C1+C2+A4E, y C1+C3+A4E comparado con OgLuc WT, la luciferasa de *Renilla* y la variante N166R a diversas temperaturas, tales como 22, 37, 42, 50 y 54°C.

30 La figura 25 muestra la emisión de luz a lo largo del tiempo (es decir, estabilidad de la señal) de las variantes de OgLuc C1, C1+A4E, C1+C2+A4E, y C1+C3+A4E comparado con OgLuc WT ("Og-Luc") y la luciferasa de *Renilla* ("hRL"), empleando RLAB, midiéndose la luminiscencia en RLU ("lum") a lo largo del tiempo en minutos, y la semivida en minutos determinada a partir de los datos a lo largo del tiempo.

35 La figura 26 muestra la longitud de onda óptima en nm con la mayor luminiscencia, empleando coelenteracina como sustrato para las variantes N166R, C1+A4E y C1+C2+A4E, comparado con la luciferasa de *Renilla*, normalizado al valor de RLU más alto en el espectro.

Las figuras 27A-B resumen el aumento en número de veces en la luminiscencia a T = 0 de las variantes mutagenizadas aleatoriamente de C1+A4E ("muestra ID") frente a la correspondiente variante de partida C1+A4E, con el cambio en el aminoácido indicado, utilizando tampón de tergitol al 0,5%.

40 La figura 28 resume el aumento en la luminiscencia en número de veces a T = 0 de las variantes L92 de C1+A4E frente a la correspondiente variante de partida C1+A4E, con el cambio en el aminoácido indicado, utilizando tampón de tergitol al 0,5%.

45 La figura 29 resume el aumento en la luminiscencia en número de veces a T = 0 de las variantes de combinación de C1+A4E ("muestra ID") frente al correspondiente variante de partida C1+A4E, con los cambios en los aminoácidos indicados, utilizando tampón de tergitol al 0,5%.

50 La figura 30 muestra la emisión de luz a lo largo del tiempo del valor del logaritmo natural (ln) de la luminiscencia medida en RLU a lo largo del tiempo en minutos, y la semivida en minutos de la variante C1+A4E+F54I, comparado con el correspondiente OgLuc C1+A4E de partida a 50°C.

55 La figura 31 muestra el alineamiento de la secuencia de aminoácidos de SEQ ID NO:10 (NATIVA), SEQ ID NO:13 (WT Sintético), SEQ ID NO:15 (N166R), SEQ ID NO:25 (C1), SEQ ID NO:27 (C1+C2), SEQ ID NO:23 (C1+A4E), SEQ ID NO:29 (C1+C2+A4E), y SEQ ID NO:31 (C1+C3+A4E) con la secuencia consenso.

La figura 32 muestra el alineamiento de la secuencia de nucleótidos de SEQ ID NO:12 (NATIVA), SEQ ID NO:2 (WT Sintético), SEQ ID NO:14 (N166R), SEQ ID NO:18 (C1), SEQ ID NO:20 (C1+C2), SEQ ID NO:16 (C1+A4E), SEQ ID NO:22 (C1+C2+A4E), y SEQ ID NO:24 (C1+C3+A4E) con la secuencia consenso.

60 La figura 33A resume el aumento en la luminiscencia en número de veces a T = 0 de las variantes de OgLuc frente a N166R determinado a partir de los datos del tampón de ensayo de tergitol al 0,5% mostrados en las figuras 5A-C y 14A, normalizados a la variante N166R.

65 La figura 33B resume el aumento en la luminiscencia en número de veces a T = 0 de las variantes de OgLuc frente a N166R determinado a partir de los datos de RLAB mostrados en las figuras 7A-C y 14B, normalizados a la variante N166R.

La figura 33C resume la semivida de la señal en minutos de las variantes de OgLuc determinada a partir de los datos de emisión de luz a lo largo del tiempo mostrados en las figuras 9A-C y 15B (tampón de ensayo de tergitol al 0,5%) y 10A-C y 16B (RLAB), normalizados a la variante N166R.

La figura 33D resume la estabilidad de las proteínas a 22°C como la semivida en minutos de las variantes de OgLuc, comparado con OgLuc WT mostrado en las figuras 12A-B y 17, normalizado a la variante N166R.

La figura 33E resume el aumento en la luminiscencia en número de veces, la semivida de la señal y la semivida a 22°C mostrados en las figuras 33A-D.

La figura 34A muestra los resultados de la luminiscencia de lisados de *E. coli* que contienen la variante IV ("IV"), luciferasa de *Renilla* ("Renilla") y C1+A4E ("C1A4E") ensayados con tergitol al 0,5%.

La figura 34B muestra la estabilidad de las proteínas a 50°C como la semivida en minutos de las variantes VI ("VI") y la luciferasa de *Renilla* ("Renilla").

Descripción detallada

En la siguiente descripción de los métodos de la invención, las etapas del proceso se realizan a temperatura ambiente (aproximadamente 22°C) y a presión atmosférica a menos que se indique lo contrario. También debe entenderse, de modo específico, que cualquier intervalo numérico indicado en la presente incluye todos los valores desde el valor más bajo hasta el valor más alto. Por ejemplo, si se menciona que un intervalo de concentración o un intervalo de efecto beneficioso es del 1% al 50%, esto significa que valores tales como del 2% al 40%, del 10% al 30%, o del 1% al 3%, etc., se enumeran expresamente en esta memoria descriptiva. De modo similar, si se da un intervalo de identidad de secuencia entre, por ejemplo, 60% a <100%, esto pretende significar que 65%, 75%, 90%, etc., se enumeran expresamente en esta memoria descriptiva. Estos son solo ejemplos de lo que se pretende específicamente, y se considera que todos los posibles valores numéricos desde el valor más bajo hasta el valor más alto se mencionan expresamente en la solicitud.

En realizaciones de la presente invención, se emplean diversas técnicas, tal como se describe en la presente, para identificar sitios para que una sustitución de un aminoácido produzca un polipéptido de luciferasa de *Oplophorus* sintético mejorado. Se emplearon otras técnicas para optimizar los codones de los polinucleótidos que codifican los diversos polipéptidos para mejorar la expresión de los polipéptidos. Se encontró que cuando se realizan una o más sustituciones de aminoácidos, solas o en diversas combinaciones, se producen luciferasas de tipo de *Oplophorus* sintéticas que presentan al menos uno de una mayor luminiscencia, una mayor estabilidad de la señal, y una mayor estabilidad de las proteínas. Además, mediante la inclusión de una o más sustituciones de optimización de codones en los polinucleótidos que codifican los diversos polipéptidos se produce una mayor expresión de los polipéptidos en diversos sistemas de expresión eucariotas y procariotas.

La luminiscencia se refiere a la emisión de luz del polipéptido de luciferasa bajo condiciones apropiadas, por ejemplo, en presencia de un sustrato adecuado, tal como coelenteracina. La emisión de luz puede medirse como una medición instantánea o casi instantánea de emisión de luz (que a veces se denomina luminiscencia en "T = 0" o "destello") tras el inicio de la reacción de luminiscencia, que puede iniciarse tras la adición del sustrato de coelenteracina. La reacción de luminiscencia en diversas realizaciones se realiza en una disolución que contiene un lisado, por ejemplo, procedente de las células en un sistema de expresión procariota o eucariota; en otras realizaciones, la expresión se produce en un sistema *in vitro*, o la proteína de luciferasa se segrega hacia un medio extracelular, de modo que, en este último caso, no es necesario producir un lisado. En algunas realizaciones, la reacción se inicia inyectando materiales apropiados, por ejemplo, coelenteracina, en una cámara de reacción (por ejemplo, un pocillo de una placa de múltiples pocillos, tal como una placa de 96 pocillos) que contiene la proteína de luciferasa. La cámara de reacción puede estar situada en un dispositivo de lectura que puede medir la emisión de luz, por ejemplo, empleando un luminómetro o un fotomultiplicador. La emisión de luz o luminiscencia también puede medirse a lo largo del tiempo, por ejemplo, en la misma cámara de reacción durante un periodo de segundos, minutos, horas, etc. La emisión de luz o luminiscencia puede indicarse como el promedio a lo largo del tiempo, la semivida de descomposición de la señal, la suma de la señal a lo largo de un periodo de tiempo, o como la emisión máxima.

Una mayor luminiscencia incluye una emisión de luz o luminiscencia aumentada, determinada mediante la comparación adecuada de mediciones obtenidas de modo comparable. Tal como se describe en la presente, una o más sustituciones de aminoácidos adecuadas en la secuencia de la luciferasa de *Oplophorus* sintética producen polipéptidos de luciferasa modificados que muestran una mayor luminiscencia. Los cambios en la secuencia de nucleótidos a partir de la secuencia de nucleótidos de *Oplophorus* de tipo salvaje pueden contribuir a una mayor luminiscencia al conducir a una sustitución de un aminoácido y/o por medio de la expresión de las proteínas mejorada.

Una mayor estabilidad de la señal incluye un aumento en el tiempo en el que la señal de una luciferasa sigue emitiendo luminiscencia, por ejemplo, según se mide mediante la semivida de la descomposición de la señal a lo largo del tiempo.

Una mayor estabilidad de las proteínas incluye la estabilidad térmica (por ejemplo, estabilidad a temperaturas elevadas) y la estabilidad química (por ejemplo, estabilidad en presencia de desnaturizantes, tales como detergentes, que incluyen, por ejemplo, Triton X-100).

5 El término "OgLuc" se refiere a la subunidad de 19 kDa madura del complejo de la proteína de la luciferasa de *Oplophorus*, es decir, sin una secuencia señal; la forma nativa de la secuencia del polipéptido de OgLuc maduro se indica en SEQ ID NO:1. La expresión "variante de OgLuc" se refiere a una OgLuc sintética con una o más sustituciones de aminoácidos. Por ejemplo, "variante N166R de OgLuc" y "OgLuc+N166R" se refieren a una OgLuc sintética que
10 tiene una sustitución de un aminoácido de N a R en la posición 166 con relación a SEQ ID NO:1. Los términos "WT", "OgLuc WT" y "OgLuc de tipo salvaje" se refieren a una proteína de OgLuc madura sintética codificada por un polinucleótido sintético con ACC en la posición 2, con relación a SEQ ID NO:1. El término "T2T" se refiere a una proteína de OgLuc madura sintética codificada por un polinucleótido sintético con ACA en la posición 2, con relación a SEQ ID NO:1. Para los datos presentados a continuación en los ejemplos, la proteína de tipo salvaje que se sintetizó es la proteína de tipo salvaje sintética de SEQ ID NO:13, que es codificada por la secuencia de nucleótidos de SEQ
15 ID NO:2.

La numeración de los aminoácidos empleada a lo largo de esta solicitud para identificar los restos sustituidos se indica con relación a las posiciones en la secuencia del polipéptido de OgLuc de tipo salvaje maduro de SEQ ID NO:1. La secuencia de OgLuc de tipo salvaje natural puede sintetizarse inicialmente con otros aminoácidos que después se escinden, dando como resultado la generación de un polipéptido de tipo salvaje maduro, tal como se muestra en SEQ
20 ID NO:1. Por ejemplo, una secuencia señal (por ejemplo, para dirigir la proteína naciente hacia un orgánulo concreto, tal como el retículo endoplásmico y/o para dirigir a la proteína para la secreción) puede estar presente al comienzo de la proteína naciente y después puede escindirse para producir la proteína de tipo salvaje madura.

25 La especificidad de sustrato de la luciferasa de *Oplophorus* es inesperadamente amplia (Inouye y Shimomura. BBRC 223:349 (1997)). Por ejemplo, la bisdesoxicoelenteracina, un análogo de la coelenteracina, es un excelente sustrato para la luciferasa de *Oplophorus* comparable a la coelenteracina (Nakamura et al., Tetrahed. Lett., 38:6405 (1997)). Además, la luciferasa de *Oplophorus* es una enzima segregada, al igual que la luciferasa del ostrácodo marino *Cypridina (Vargula) hilgendorffii* (Johnson y Shimomura, Meth. Enzyme, 57:331 (1978)), que también emplea una
30 luciferina de tipo imidazopirazinona para emitir luz.

Se ha indicado que el peso molecular de la luciferasa de *Oplophorus* es de 130 kDa (mediante filtración en gel) para el complejo de la proteína nativa, y de 31 kDa después de un tratamiento con SDS (Shimomura et al., Biochem.,
35 17:1994 (1978)). La luciferasa también muestra un peso molecular de aproximadamente 106 kDa en filtración en gel, y se ha descubierto que la molécula se separa en proteínas de 35 kDa y 19 kDa tras un análisis mediante electroforesis en gel de dodecilsulfato de sodio-poliacrilamida (SDS-PAGE) (Inouye et al., FEBS Lett., 481:19 (2000)). Inouye et al. (2000) describen la clonación molecular de los ADNc que codifican las proteínas de 35 kDa y 19 kDa, y la identificación del componente de la proteína que cataliza la reacción de luminiscencia. Los ADNc que codifican las proteínas se expresaron en células bacterianas y de mamífero como una proteína de 19 kDa que es capaz de catalizar la oxidación
40 luminiscente de la coelenteracina (Inouye et al., 2000). La secuencia primaria de la proteína de 35 kDa reveló una secuencia repetida rica en leucina, mientras que la proteína catalítica de 19 kDa no compartía homología con ninguna de las luciferasas conocidas, incluyendo diversas luciferasas de imidazopirazinona (Inouye et al., 2000).

45 La proteína de 19 kDa (OgLuc) de la luciferasa de *Oplophorus* parece ser el componente catalítico más pequeño que tiene función luciferasa, y su estructura primaria no presenta homología significativa con ninguna de las luciferasas descritas, incluyendo las luciferasas de imidazopirazinona (Lorenz et al., PNAS USA, 88:4438 (1991); Thompson et al., PNAS USA, 86:6567 (1989)). Inouye et al. (2000) indicaron que la secuencia de aminoácidos global de la proteína de 19 kDa parecía similar a la de una amina oxidasa de *E. coli* (757 restos aminoácidos; pir 140924) en la región de los restos 217- 392 (dominio de D3-S1) (Parson et al., Structure, 3:1171 (1995)), mientras que la región amino-terminal (3-49) de la misma proteína es homóloga con la región amino-terminal (1-47) de la proteína de unión a ácidos grasos (132 restos aminoácidos; GenBank, L23322) (Becker et al., Gene, 148:321 (1994)).
50

La formación de modelos de homología requiere la identificación de al menos un molde de estructura tridimensional adecuado, habitualmente una estructura tridimensional determinada de modo experimental de una proteína homóloga
55 con una similitud de secuencia significativa con la proteína diana. La OgLuc no presenta una similitud de secuencia significativa con otras proteínas conocidas. Por tanto, se emplearon métodos de reconocimiento de plegamiento diseñados para identificar homólogos distantes de OgLuc, tales como proteínas con baja similitud de secuencia con OgLuc. Esta estrategia produjo varios moldes de estructura terciaria potenciales que pertenecen a la familia de proteínas de unión a ácidos grasos (FABP), que es parte de la superfamilia de proteínas de calicina. El modelo mostró
60 que la firma estructural de plegamiento de la calicina, que une de modo eficaz los N- y C-terminales entre sí a través de enlaces de hidrógeno, y que está presente en al menos tres FABP, no está completamente conservada en OgLuc. El resto Asn166 de OgLuc (cerca del C-terminal) es incapaz de formar un enlace de hidrógeno con los carbonilos de la cadena principal cerca del N-terminal. Sin embargo, los modelos de mutantes que contienen Arg o Lys en la posición 166 de OgLuc sugieren que el restablecimiento de este motivo estructural podría mejorar la estabilidad estructural de
65 OgLuc y su expresión/actividad en células.

En la presente se describe una luciferasa modificada (variante) sintética, así como sus fragmentos, por ejemplo, los que son útiles en ensayos de complementación, que tiene al menos una sustitución de un aminoácido con relación a la correspondiente luciferasa de tipo salvaje en una región que es estructuralmente homóloga con un miembro de la superfamilia de proteínas de la calicina, por ejemplo, la familia de las proteínas de unión a ácidos grasos. En la presente se describe una luciferasa de crustáceo modificada, por ejemplo, una luciferasa de decápodo modificada, así como sus fragmentos, por ejemplo, los que son útiles en ensayos de complementación, que tiene al menos una sustitución de un aminoácido con relación a la correspondiente luciferasa de crustáceo de tipo salvaje en una región que es estructuralmente homóloga con un miembro de la superfamilia de proteínas de la calicina, por ejemplo, la familia de las proteínas de unión a ácidos grasos. En la presente se describe una luciferasa modificada de un flagelado unicelular eucariota, así como sus fragmentos, por ejemplo, los que son útiles en ensayos de complementación, que tiene al menos una sustitución de un aminoácido con relación a la correspondiente luciferasa de flagelado unicelular eucariota de tipo salvaje, por ejemplo, luciferasas de *Dinoflagellata*, que incluyen *Dinophyceae*, *Noctilucliphyceae*, o *Syndiniophyceae*, en una región que es estructuralmente homóloga con un miembro de la superfamilia de proteínas de la calicina, por ejemplo, la familia de las proteínas de unión a ácidos grasos. Una molécula de ácido nucleico que codifica la luciferasa modificada puede o no codificar un péptido señal secretor unido a la luciferasa modificada.

En la presente, una sustitución en la luciferasa modificada sintética, o uno de sus fragmentos, es en un resto aminoácido en la correspondiente posición en la región que es estructuralmente homóloga con un miembro de la superfamilia de proteínas de la calicina, por ejemplo, la familia de las proteínas de unión a ácidos grasos, resto que puede participar en la formación de enlaces de hidrógeno o iónicos intramoleculares, y está asociado con una mayor luminiscencia en la luciferasa modificada. Una mayor luminiscencia incluye, pero no se limita a una mayor emisión de luz, una cinética alterada de la emisión de luz, por ejemplo, mayor estabilidad de la intensidad de luz, o un color alterado de la luminiscencia, por ejemplo, un desplazamiento hacia longitudes de onda más cortas o más largas, o una de sus combinaciones. En una realización, el resto en la luciferasa modificada sintética en la correspondiente posición puede interactuar con un resto en una región que se corresponde con los restos 1 a 10 o 144 a 148 de OgLuc, por ejemplo, uno que tiene SEQ ID NO:1 (nótese que la numeración de estas posiciones se basa en un Phe en el resto 1 de la secuencia madura, no una Met; sin embargo, otros restos pueden preceder al Phe, tal como una Val en la posición -1, que pueden ser introducidos mediante la inserción de un sitio de clonación) o un resto con átomos que están dentro de 4 a 8 Å, por ejemplo, dentro de 6 Å, del resto en la correspondiente posición (posición 166). Las posiciones correspondientes pueden identificarse mediante el alineamiento de las secuencias empleando, por ejemplo, programas de alineamiento de secuencias, programas de predicción de la estructura secundaria, o métodos de reconocimiento de plegamiento, o una de sus combinaciones. La luciferasa modificada según la invención puede incluir otras sustituciones de aminoácidos que alteren el color de la luminiscencia, por ejemplo, una o más sustituciones que produzcan una luminiscencia desplazada al rojo, que alteren la estabilidad de la señal, que alteren la estabilidad de las proteínas, o cualquiera de sus combinaciones.

En la presente se describe una luciferasa de decápodo modificada con tiene una mayor luminiscencia con relación a la correspondiente luciferasa de decápodo de tipo salvaje, o una luciferasa de decápodo modificada que emplea coelenteracina. Las coelenteracinas incluyen, pero no se limitan a las coelenteracinas naturales, así como a sus derivados (análogos), tales como las descritas en la patente de EEUU n.º 7.118.878, así como EnduRen, ViviRen, coelenteracina n, coelenteracina h, coelenteracina c, coelenteracina cp, coelenteracina e, coelenteracina f, coelenteracina fcp, coelenteracina hh, coelenteracina i, coelenteracina icp, 2-metil-coelenteracina, y las descritas en el documento WO/040100 y en la solicitud de EE:UU. n.º de serie 12/056.073, cuyas descripciones se incorporan por referencia a la presente.

Una luciferasa modificada según la presente descripción puede tener un resto distinto de la asparagina en una posición que se corresponde con el resto 166 en SEQ ID NO:1, que da como resultado una mayor luminiscencia y, opcionalmente, un ácido aspártico en una posición que se corresponde con el resto 5 en SEQ ID NO:1, una glicina en una posición que se corresponde con el resto 8 en SEQ ID NO:1, un ácido aspártico en una posición que se corresponde con el resto 9 en SEQ ID NO:1, un triptófano, una tirosina o una fenilalanina en una posición que se corresponde con el resto 10 en SEQ ID NO:1, una asparagina en una posición que se corresponde con el resto 144 en SEQ ID NO:1 y/o una glicina en una posición que se corresponde con el resto 147 en SEQ ID NO:1, o cualquiera de sus combinaciones. En una realización, el resto en la luciferasa modificada que corresponde al resto 166 en SEQ ID NO:1 es lisina. En otra realización, el resto en la luciferasa modificada que corresponde al resto 166 en SEQ ID NO:1 es arginina. En una realización, el resto en la luciferasa modificada que corresponde al resto 166 en SEQ ID NO:1 es capaz de formar uno o más enlaces de hidrógeno o iónicos intramoleculares con los carbonilos o la cadena lateral en una posición que se corresponde con el resto 9 en SEQ ID NO:1 cerca del N-terminal de la luciferasa modificada. En una realización, la luciferasa modificada carece de una secuencia de péptido señal. En una realización, la luciferasa modificada tiene al menos 60%, por ejemplo, al menos 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99%, pero menos del 100% de identidad de secuencia de aminoácidos con SEQ ID NO:1.

En una realización, la correspondiente luciferasa de tipo salvaje es una luciferasa de *Oplophorus*, por ejemplo, *Oplophorus gracilirostris*, *Oplophorus grimaldii*, *Oplophorus spinicauda*, *Oplophorus foliaceus*, *Oplophorus noraezeelandiae*, *Oplophorus typus*, *Oplophorus noraezeelandiae* u *Oplophorus spinous*, una luciferasa de *Heterocarpus*, una luciferasa de *Systemella* o una luciferasa de *Acanthephyra*. En una realización, la luciferasa modificada tiene al menos 2 veces o más, por ejemplo, al menos 4 veces, una emisión de luminiscencia aumentada

en una célula procariota y/o una célula eucariota con relación a la luciferasa de tipo salvaje correspondiente.

5 En otra realización, una luciferasa de dinoflagelado modificada puede tener una mayor luminiscencia con relación a la correspondiente luciferasa de dinoflagelado de tipo salvaje, por ejemplo, una luciferasa de dinoflagelado, tal como una luciferasa de *Lingulodinium polyedrum*, una luciferasa de *Pyrocystis lunula*, o una luciferasa que tiene la SEQ ID NO:21. La luciferasa modificada puede tener un resto distinto a la asparagina en una posición que corresponde al resto 166 en SEQ ID NO:1, por ejemplo, una arginina, y opcionalmente una prolina en una posición que corresponde al resto 5 en SEQ ID NO:1, una glicina en una posición que corresponde al resto 8 en SEQ ID NO:1, una arginina en una posición que corresponde al resto 9 en SEQ ID NO:1, un triptófano, una tirosina o una fenilalanina en una posición que corresponde al resto 10 en SEQ ID NO:1, una fenilalanina en una posición que corresponde al resto 144 en SEQ ID NO:1 y/o una treonina en una posición que corresponde al resto 147 en SEQ ID NO:1, o cualquiera de sus combinaciones. En una realización, el resto en la luciferasa modificada que se corresponde con el resto 166 en SEQ ID NO:1 es lisina. En otra realización, el resto en la luciferasa modificada que corresponde al resto 166 en SEQ ID NO:1 es arginina. En una realización, el resto en la luciferasa modificada que corresponde al resto 166 en SEQ ID NO:1 es capaz de formar uno o más enlaces de hidrógeno o iónicos intramoleculares con los carbonilos o la cadena lateral en una posición que se corresponde con el resto 9 en SEQ ID NO:1 cerca del N-terminal de la luciferasa modificada. En una realización, la luciferasa modificada carece de una secuencia de péptido señal.

20 En una realización, la luciferasa modificada tiene al menos 60%, por ejemplo, al menos 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99%, pero menos del 100% de identidad de secuencia de aminoácidos con SEQ ID NO:21. La luciferasa modificada, que incluye una con otras sustituciones de aminoácidos que alteran el color de la luminiscencia, puede emplearse con una luciferina modificada en una reacción luminogénica que produzca un color alterado de la luminiscencia.

25 También se describe una luciferasa modificada que tiene un dominio estructural tridimensional relacionado con el barril-beta de FABP, teniéndola luciferasa modificada una sustitución que da como resultado la unión no covalente, por ejemplo, a través de enlaces de hidrógeno o iónicos intramoleculares, de las láminas beta terminales del barril beta, y opcionalmente con otros enlaces no covalentes, por ejemplo, a través de enlaces de hidrógeno o iónicos intramoleculares, con estructuras secundarias adyacentes.

30 También se describe una luciferasa de decápodo o de dinoflagelado modificada que tiene mayor luminiscencia y una arginina, lisina, alanina, leucina, prolina, glutamina o serina en una posición que corresponde al resto 166 en SEQ ID NO:1, y al menos una sustitución de un aminoácido con relación a la correspondiente luciferasa de decápodo o de dinoflagelado de tipo salvaje. En una realización, la al menos una sustitución de un aminoácido en la luciferasa modificada es una sustitución en una posición que se corresponde con el resto 4, 11, 33, 44, 45, 54, 75, 104, 115, 124, 135, 138, 139, 167 o 169, o una de sus combinaciones, en SEQ ID NO:1, por ejemplo, una sustitución que resulta en una mayor luminiscencia con relación a una luciferasa modificada que tiene una mayor luminiscencia y una arginina, lisina, alanina, leucina, prolina, glutamina o serina en una posición que corresponde al resto 166 en SEQ ID NO:1.

40 En una realización, la luciferasa modificada tiene una o más secuencias de aminoácidos heterólogas en el N-terminal, C-terminal o ambos (un polipéptido de fusión, tal como uno con un marcador de epitopo o de fusión), que opcionalmente interacciona, directa o indirectamente, con una molécula de interés. En una realización, la presencia de la secuencia(s) heteróloga(s) no altera sustancialmente la luminiscencia de la luciferasa modificada antes o después de la interacción con la molécula de interés. En una realización, la secuencia de aminoácidos heteróloga es un marcador de epitopo. En otra realización, la secuencia de aminoácidos heteróloga es una que durante o después de la interacción con una molécula de interés sufre un cambio conformacional, que a su vez altera la actividad de la luciferasa, por ejemplo, una Ogluc modificada con dicha secuencia de aminoácidos es útil para detectar interacciones alostéricas. La luciferasa modificada o una fusión con la luciferasa modificada, o uno de sus fragmentos, puede emplearse como indicador.

50 En una realización, un fragmento de una luciferasa de la invención se fusiona con una secuencia de aminoácidos heteróloga, formando con ello la fusión un barril-beta, siendo capaz de generar la proteína de fusión luminiscencia a partir de una luciferina natural o uno de sus derivados.

55 También se describe un polinucleótido que codifica una luciferasa modificada de la invención, o una fusión de esta, una célula hospedante aislada que tiene el polinucleótido o la luciferasa modificada, o una fusión de esta, y métodos para utilizar el polinucleótido, la luciferasa modificada, o una fusión de esta, o la célula hospedante de la invención.

60 También se describe un método para identificar posiciones de aminoácidos en una proteína de interés que se encuentran en diferentes estructuras secundarias, por ejemplo, estructuras separadas por 5 o más aminoácidos que no son parte de cualquiera de las estructuras secundarias, y que son capaces de formar enlaces de hidrógeno o iónicos entre sí. El método incluye comparar las estructuras secundarias predichas para la secuencia de aminoácidos de una proteína de interés con las estructuras secundarias de una o más proteínas sin similitud de secuencia global, por ejemplo, menos del 30% de identidad con la proteína de interés. La una o más proteínas tienen una estructura tridimensional definida y al menos una de las proteínas posee un primer resto asociado con al menos una primera estructura secundaria que forma un enlace de hidrógeno o iónico, por ejemplo, puentes salinos, entre las cadenas

65

laterales o entre una cadena lateral de un carbonilo o un carbonilo de la cadena principal cerca o a una distancia de 5 o 10 restos de un segundo resto asociado con una segunda estructura secundaria, respectivamente. En una realización, la primera estructura secundaria es C-terminal con respecto a la segunda estructura secundaria. En otra realización, la primera estructura secundaria es N-terminal con respecto a la segunda estructura secundaria. Después se determina si la proteína de interés posee una o más estructuras secundarias que se corresponden con al menos la primera estructura secundaria en la una o más proteínas, y si es así, se determinan las posiciones de los aminoácidos en la proteína de interés que se corresponden con el primer resto, el segundo resto, o ambos, en la una o más proteínas. En una realización, una estructura secundaria es una hélice 3_{10} o un barril-beta. En una realización, la proteína de interés es una luciferasa. En una realización, el primer resto es capaz de formar un enlace de hidrógeno o iónico con uno o más carbonilos de la cadena principal a 5 restos del segundo resto. En una realización, la una o más proteínas son proteínas de unión a ácidos grasos.

Definiciones

Los restos aminoácidos en las luciferasas modificadas de la invención pueden estar en configuración L, en configuración D o ser aminoácidos no naturales, tales como norleucina, L-etionina, β -2-tienilalanina, 5-metiltriptófano-norvalina, L-canavanina, p-fluorofenilalanina, p-(4-hidroxibenzoil)fenilalanina, ácido 2-ceto-4-(metiltio)butírico, beta-hidroxi-leucina, gamma-cloronorvalina, gamma-metil-D-leucina, beta-D-L-hidroxileucina, ácido 2-amino-3-clorobutírico, N-metil-D-valina, 3,4-difluoro-L-fenilalanina, 5,5,5-trifluoroleucina, 4,4,4,-trifluoro-L-valina, 5-fluoro-L-triptófano, 4-azido-L-fenilalanina, 4-bencil-L-fenilalanina, tiaprolina, 5,5,5-trifluoroleucina, 5,5,5,5',5',5'-hexafluoroleucina, ácido 2-amino-4-metil-4-pentenoico, ácido 2-amino-3,3,3-trifluoro-metilpentanoico, ácido 2-amino-3-metil-5,5,5-trifluoro-pentanoico, ácido 2-amino-3-metil-4-pentenoico, trifluorovalina, hexafluorovalina, homocisteína, hidroxilisina, ornitina, y aquellos con enlaces peptídicos opcionalmente reemplazados por un enlace tal como $-\text{CH}_2\text{NH}-$, $-\text{CH}_2\text{S}-$, $-\text{CH}_2-\text{CH}_2-$, $-\text{CH}=\text{CH}-$ (cis y trans), $-\text{COCH}_2-$, $-\text{CH}(\text{OH})\text{CH}_2-$, y $-\text{CH}_2\text{SO}-$, mediante métodos conocidos en la técnica. Sin apartarse de la nomenclatura convencional de polipéptidos, las abreviaturas para los restos aminoácidos naturales son como se muestran en la siguiente tabla de correspondencia.

Tabla de correspondencia

1 letra	3 letras	AMINOÁCIDO
Y	Tyr	L-tirosina
G	Gly	L-glicina
F	Phe	L-fenilalanina
M	Met	L-metionina
A	Ala	L-alanina
S	Ser	L-serina
I	Ile	L-isoleucina
L	Leu	L-leucina
T	Thr	L-treonina
V	Val	L-valina
P	Pro	L-prolina
K	Lys	L-lisina
H	His	L-histidina
Q	Gln	L-glutamina
E	Glu	ácido L-glutámico
W	Trp	L-triptófano
R	Arg	L-arginina
D	Asp	ácido L-aspártico
N	Asn	L-asparagina
C	Cys	L-cisteína

Una mayor luminiscencia, tal como se emplea en la presente, puede incluir cualquiera de lo siguiente: una mayor emisión de luz, una cinética alterada de la emisión de luz, por ejemplo, mayor estabilidad de la intensidad de luz, o un color alterado de la luminiscencia, por ejemplo, un desplazamiento hacia longitudes de onda más cortas o más largas.

El término “homología” se refiere a un grado de complementariedad entre dos o más secuencias. Puede existir una homología parcial o una homología completa (es decir, identidad). La homología a menudo se mide empleando un programa de análisis de secuencias (por ejemplo, “GCG” y el paquete informático de análisis de secuencias “Seqweb”, anteriormente comercializado por the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Estos programas aparean secuencias similares asignando grados de homología a diversas sustituciones, deleciones, inserciones y otras modificaciones. Las sustituciones conservativas generalmente incluyen sustituciones dentro de los siguientes grupos: glicina, alanina; valina, isoleucina, leucina; ácido aspártico, ácido glutámico, asparagina, glutamina; serina, treonina; lisina, arginina; y fenilalanina, tirosina.

El término “aislado”, cuando se emplea en relación con un ácido nucleico o un polipéptido, tal como en “oligonucleótido aislado”, “polinucleótido aislado”, “proteína aislada” o “polipéptido aislado”, se refiere a un ácido nucleico o una secuencia de aminoácidos que se identifica y se separa de al menos un contaminante con el que normalmente está asociado en su origen. Así, un ácido nucleico aislado o un polipéptido aislado está presente en una forma o en un emplazamiento que es diferente del que se encuentra en la naturaleza. En contraste, los ácidos nucleicos no aislados (por ejemplo, ADN y ARN) o los polipéptidos no aislados (por ejemplo, proteínas y enzimas) se encuentran en el estado en el que existen en la naturaleza. Por ejemplo, una secuencia dada de ADN (por ejemplo, un gen) se encuentra en el cromosoma de la célula hospedante cerca de genes vecinos; las secuencias de ARN (por ejemplo, una secuencia de ARNm específica que codifica una proteína específica) se encuentran en la célula como una mezcla con numerosos otros ARNm que codifican una multitud de proteínas. Sin embargo, un ácido nucleico aislado incluye, por ejemplo, dicho ácido nucleico en células que normalmente expresan ese ácido nucleico, en donde el ácido nucleico está en una localización cromosómica diferente que la de las células naturales, o bien está flanqueado por una secuencia de ácido nucleico diferente que la que se encuentra en la naturaleza. El ácido nucleico u oligonucleótido aislado puede estar presente en forma monocatenaria o bicatenaria. Cuando un ácido nucleico u oligonucleótido aislado se va a utilizar para expresar una proteína, el oligonucleótido contiene, como mínimo, la hebra sentido o codificadora (es decir, un ácido nucleico monocatenario), pero puede contener hebras sentido y antisentido (es decir, un ácido nucleico bicatenario).

Los términos “molécula de ácido nucleico”, “polinucleótido” o “secuencia de ácido nucleico”, tal como se emplean en la presente, se refieren a un ácido nucleico, ADN o ARN, que comprende las secuencias codificadoras necesarias para la producción de un polipéptido o precursor de proteína. El polipéptido codificado puede ser un polipéptido de longitud completa, uno de sus fragmentos (más pequeño que la longitud completa), o una fusión del polipéptido de longitud completa, o de uno de sus fragmentos, con otro polipéptido, produciendo un polipéptido de fusión.

La “luciferasa de *Oplophorus*” es un complejo de proteínas de 35 kDa y 19 kDa nativas. La proteína de 19 kDa es el componente catalítico más pequeño (GenBank n.º de registro BAB13776, 196 aminoácidos). Tal como se emplea en la presente, OgLuc es la proteína de 19 kDa sin el péptido señal (169 aminoácidos, restos 28 a 196 de BAB13776).

Por “péptido”, “proteína” y “polipéptido” se quiere incitar cualquier cadena de aminoácidos, independientemente de la longitud o la modificación postraduccional (por ejemplo, glicosilación o fosforilación). Las moléculas de ácidos nucleicos de la invención codifican una variante de una proteína natural, o un fragmento polipeptídico de esta, que tiene una secuencia de aminoácidos que tiene al menos 60%, por ejemplo, al menos 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99%, pero menos del 100% de identidad de secuencia de aminoácidos con la secuencia de aminoácidos de la proteína natural (nativa o de tipo salvaje) a partir de la cual se deriva. La expresión “polipéptido de fusión” o “proteína de fusión” se refiere a una proteína quimérica que contiene una proteína de referencia (por ejemplo, luciferasa) unida en el N- y/o C-terminal a una o más secuencias heterólogas (por ejemplo, un polipéptido que no es de luciferasa).

La estructura primaria de la proteína (secuencia primaria, secuencia peptídica, secuencia de proteína) es la secuencia de aminoácidos. En general se indica comenzando a partir del extremo amino-terminal (N) hasta el extremo carboxilo-terminal (C). La estructura secundaria de la proteína puede describirse como la conformación local de la cadena peptídica, independiente del resto de la proteína. Existen elementos de la estructura secundaria “regulares” (por ejemplo, hélices, láminas o cadenas) que en general están estabilizados mediante interacciones de enlaces de hidrógeno entre los átomos del esqueleto de los restos participantes, y elementos de la estructura secundaria “irregulares” (por ejemplo, curvas, vueltas, bucles, espirales, segmentos desordenados o no estructurados). La estructura secundaria de la proteína puede predecirse con diferentes métodos/programas, por ejemplo, PSIPRED (McGuffin et al., *Bioinformatics*, 16:404 (2000)), PORTER (Pollastrì et al., *Bioinformatics*, 21:1719 (2005)), DSC (King y Sternberg, *Protein Sci.*, 5:2298 (1996)), véase <http://www.expasy.org/tools/#secondary> para obtener una lista. La estructura terciaria de una proteína es la estructura tridimensional (3D) global de la cadena peptídica. Se describe mediante las posiciones atómicas en el espacio tridimensional, y puede implicar interacciones entre grupos que están distantes en la estructura primaria. Las estructuras terciarias de las proteínas se clasifican en plegamientos, que son disposiciones tridimensionales específicas de elementos de la estructura secundaria. A veces no existe una similitud de secuencia discernible entre proteínas que tienen el mismo plegamiento.

El término “de tipo salvaje” o el término “nativo”, tal como se emplean en la presente, se refieren a un gen o un producto génico que tiene las características de ese gen o producto génico aislado a partir de una fuente natural. Un gen de

tipo salvaje es el que se observa con más frecuencia en una población y, por tanto, se denomina de modo arbitrario la forma “de tipo salvaje” del gen. En contraste, el término “mutante” se refiere a un gen o un producto génico que muestra modificaciones en la secuencia y/o las propiedades funcionales (es decir, características alteradas) cuando se compara con el gen o el producto génico de tipo salvaje. Se señala que los mutantes naturales pueden aislarse; estos se identifican por el hecho de que tienen características alteradas cuando se comparan con el gen o el producto génico de tipo salvaje.

I. Polinucleótidos y proteínas ilustrativos

La presente descripción incluye una luciferasa modificada, o uno de los fragmentos de su proteína, por ejemplo, los que tienen deleciones, por ejemplo, una deleción de 1 a aproximadamente 5 restos, y sus quimeras (fusiones) (véanse las solicitudes de EE.UU. números de serie 60/985.585 y 11/732.105, cuyas descripciones se incorporan por referencia a la presente) que tienen al menos una sustitución de un aminoácido con relación a una luciferasa de tipo salvaje, y dicha sustitución da como resultado que la luciferasa modificada tiene una mayor estabilidad, una mayor luminiscencia, por ejemplo, una mayor emisión de luminiscencia, una mayor estabilidad de la cinética de la luminiscencia, o un color alterado de la luminiscencia, o ambos. Las secuencias de luciferasa de una luciferasa modificada son sustancialmente las mismas que la secuencia de aminoácidos de la correspondiente luciferasa de tipo salvaje. Un polipéptido o un péptido que tenga sustancialmente la misma secuencia significa que una secuencia de aminoácidos es, en gran parte pero no totalmente, la misma y conserva la actividad funcional de la secuencia con la que está relacionada. En general, dos secuencias de aminoácidos son sustancialmente la misma o son sustancialmente homólogas si tienen al menos 60%, por ejemplo, al menos 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99%, pero menos del 100% de identidad de secuencia de aminoácidos. En una realización, la luciferasa modificada está codificada por un polinucleótido recombinante.

La homología o identidad a menudo puede medirse empleando un programa de análisis de secuencias. Estos programas aparean las secuencias similares asignando grados de homología a diversas deleciones, sustituciones, y otras modificaciones. Los términos “homología” e “identidad”, en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, se refieren a dos o más secuencias o subsecuencias que son las mismas o que tienen un porcentaje especificado de restos aminoácidos o nucleótidos que son los mismos cuando se comparan y se alinean para la máxima correspondencia a través de una ventana de comparación o una región designada, según se mide empleando cualquier número de algoritmos de comparación de secuencias o mediante alineamiento manual e inspección visual.

Para la comparación de las secuencias, generalmente una secuencia actúa como secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se designan las coordenadas de subsecuencias, si es necesario, y se designan los parámetros del programa del algoritmo de la secuencia. Pueden emplearse los parámetros por defecto del programa, o pueden designarse otros parámetros. El algoritmo de comparación de secuencias calcula después el porcentaje de identidades de secuencia para las secuencias de ensayo, con relación a la secuencia de referencia, basándose en los parámetros del programa.

Los métodos de alineamiento de secuencias para la comparación son bien conocidos en la técnica. El alineamiento óptimo de las secuencias para la comparación puede ser llevado a cabo mediante el algoritmo de homología local de Smith et al. (1981), el algoritmo de alineamiento de homología de Needleman et al. (*J. Mol. Biol.*, 48:443 (1970)), el método de búsqueda de similitud de Person et al. (*Proc. Natl. Acad. Sci. USA*, 85, 2444 (1988)), las aplicaciones informáticas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el paquete Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante alineamiento manual e inspección visual.

Las aplicaciones informáticas de estos algoritmos matemáticos pueden utilizarse para la comparación de secuencias para determinar la identidad de las secuencias. Estas aplicaciones incluyen, pero no se limitan a: CLUSTAL en el programa PC/Gene (disponible en Intelligenetics, Mountain View, California); el programa ALIGN (versión 2.0) y GAP, BESTFIT, BLAST, FASTA, y TFASTA en el paquete Wisconsin Genetics Software Package, versión 8 (disponible en Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, EEUU). Empleando estos programas, los alineamientos pueden realizarse utilizando los parámetros por defecto. El programa CLUSTAL está bien descrito por Higgins et al., *Gene*, 73:237 (1988); Higgins et al., *CABIOS*, 5:157 (1989); Corpet et al., *Nucl. Acids Res.*, 16:1088 (1988); Huang et al., *CABIOS*, 8:155 (1992); y Pearson et al., *Methods Mol. Biol.s*, 24:307 (1994). El programa ALIGN se basa en el algoritmo de Myers y Miller, *LABIOS*, 4:11 (1988). Los programas BLAST de Altschul et al. (*J. Mol. Biol.*, 215:403 (1990)) se basan en el algoritmo de Karlin y Altschul (*PNAS USA*, 90:5873 (1993)).

Los programas informáticos para realizar los análisis BLAST están disponibles al público en the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar, en primer lugar, las parejas de secuencia de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia problema, que se aparean o cumplen con alguna puntuación umbral de valor positivo T cuando se alinean con una palabra con la misma longitud en una secuencia de una base de datos. T se denomina el umbral de puntuación de palabra del vecindario (Altschul et al., *J. Mol. Biol.*, 215:403 (1990)). Estos descubrimientos de palabras del vecindario iniciales actúan como

semillas para iniciar búsquedas para encontrar HSP más largas que las contengan. Los descubrimientos de las palabras después se extienden en ambas direcciones a lo largo de cada secuencia hasta el máximo que pueda aumentar la puntuación de alineamiento acumulada. Las puntuaciones acumuladas se calculan empleando, para secuencias de nucleótidos, los parámetros M (puntuación de premio para una pareja de restos apareados; siempre >0) y N (puntuación de penalización para restos desapareados; siempre <0). Para secuencias de aminoácidos se emplea una matriz de puntuación para calcular la puntuación acumulada. La extensión de los descubrimientos de las palabras en cada dirección se detiene cuando la puntuación de alineamiento acumulada disminuye en una cantidad X desde su valor máximo logrado, cuando la puntuación acumulada llega hasta cero o menor debido a la acumulación de uno o más alineamientos de restos de puntuación negativa, o cuando se alcanza el final de cualquiera de las secuencias.

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, PNAS USA, 90:5873 (1993)). Una medición de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma menor (P(N)), que proporciona una indicación de la probabilidad con la que un apareamiento entre dos secuencias de nucleótidos o de aminoácidos se produciría por azar. Por ejemplo, una secuencia de ácido nucleico de ensayo se considera similar a una secuencia de referencia si la probabilidad de suma menor en una comparación de la secuencia del ácido nucleico de ensayo con la secuencia de ácido nucleico de referencia es menor que aproximadamente 0,1, más preferiblemente menor que aproximadamente 0,01, y lo más preferiblemente es menor que aproximadamente 0,001.

Para obtener los alineamientos con huecos para el propósito de comparación, puede utilizarse Gapped BLAST (en BLAST 2.0), según se describe en Altschul et al. (Nuc. Acids Res., 25:3389 (1997)). Como alternativa, puede emplearse PSI-BLAST (en BLAST 2.0) para realizar una búsqueda iterada que detecte relaciones distantes entre moléculas. Véase Altschul et al., *supra*. Cuando se emplea BLAST, Gapped BLAST, PSI-BLAST, pueden emplearse los parámetros por defecto de los respectivos programas (por ejemplo, BLASTN para secuencias de nucleótidos, BLASTX para proteínas). El programa BLASTN (para secuencias de nucleótidos) emplea, como parámetros por defecto, una longitud de palabra (W) de 11, una esperanza (E) de 10, un punto de corte de 100, M = 5, N = -4, y una comparación de ambas hebras. Para secuencias de aminoácidos, el programa BLASTP emplea, como parámetros por defecto, una longitud de palabra (W) de 3, una esperanza (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, PNAS USA, 89:10915 (1989)). Véase "www.ncbi.nlm.nih.gov."

En particular, un polipéptido puede estar sustancialmente relacionado con otro polipéptido (de referencia), excepto por una variación conservativa o no conservativa. Una variación conservativa indica la sustitución de un resto aminoácido por otro resto biológicamente similar, que incluye restos aminoácidos naturales o no naturales. Los ejemplos de variaciones conservativas incluyen la sustitución de un resto hidrófobo, tal como isoleucina, valina, leucina o metionina, por otro, o la sustitución de un resto polar por otro, tal como la sustitución de arginina por lisina, ácido glutámico por ácido aspártico, o glutamina por asparagina, y similares. Otros ejemplos ilustrativos de sustituciones conservativas incluyen los cambios de: alanina a serina; arginina a lisina; asparagina a glutamina o histidina; aspartato a glutamato; cisteína a serina; glutamina a asparagina; glutamato a aspartato; glicina a prolina; histidina a asparagina o glutamina; isoleucina a leucina o valina; leucina a valina o isoleucina; lisina a arginina, glutamina, o glutamato; metionina a leucina o isoleucina; fenilalanina a tirosina, leucina o metionina; serina a treonina; treonina a serina; triptófano a tirosina; tirosina a triptófano o fenilalanina; valina a isoleucina a leucina. Una luciferasa modificada de la invención tiene una sustitución conservativa o no conservativa que da como resultado una mayor estabilidad, luminiscencia, o ambas.

Las proteínas de luciferasa modificadas o sus proteínas de fusión de la invención pueden prepararse por métodos recombinantes o por métodos de síntesis química de péptidos en fase sólida. Estos métodos son conocidos en la técnica.

II. Vectores y células hospedantes que codifican la luciferasa modificada o sus fusiones

Tras haber preparado una molécula de ácido nucleico deseable que codifica una luciferasa modificada, uno de sus fragmentos, tal como un fragmento con actividad de luminiscencia o que puede ser complementado por otra molécula para dar como resultado actividad de luminiscencia, o una fusión de esta con actividad de luminiscencia, puede prepararse una casete de expresión que codifica la luciferasa modificada, uno de sus fragmentos, por ejemplo, un fragmento para la complementación, o una fusión de esta con actividad de luminiscencia. Por ejemplo, una molécula de ácido nucleico que comprende una secuencia de ácido nucleico que codifica una luciferasa modificada opcionalmente se une operablemente a secuencias reguladoras de la transcripción, por ejemplo, uno o más potenciadores, un promotor, una secuencia de terminación de la transcripción o una de sus combinaciones, para formar un módulo de expresión. La molécula de ácido nucleico o la casete de expresión puede introducirse en un vector, por ejemplo, un vector plasmídico o vírico, que opcionalmente incluye un gen de un marcador seleccionable, y el vector se introduce en una célula de interés, por ejemplo, una célula procariota, tal como *E. coli*, *Streptomyces* spp., *Bacillus* spp., *Staphylococcus* spp. y similares, así como células eucariotas que incluyen células vegetales (de dicotiledóneas o de monocotiledóneas), hongos, levaduras, por ejemplo, *Pichia*, *Saccharomyces* o *Schizosaccharomyces*, o una célula de mamífero, sus lisados, o como una mezcla de transcripción/traducción *in vitro*. Las células de mamífero incluyen, pero no se limitan a células bovinas, caprinas, ovinas, felinas, de primate no humano, por ejemplo de simio, y humanas. Las líneas de células de mamífero incluyen, pero no se limitan a células

CHO, COS, 293, HeLa, CV-1, SH-SY5Y, HEK293, y NIH3T3.

La expresión de una luciferasa modificada codificada puede ser controlada por cualquier promotor capaz de expresarse en células procariotas o en células eucariotas, incluyendo promotores sintéticos. Los promotores de procariotas incluyen, pero no se limitan los promotores SP6, T7, T5, *tac*, *bla*, *trp*, *gal*, *lac* o de maltosa, e incluyen cualquier fragmento que tenga actividad de promotor. Los promotores de eucariotas incluyen, pero no se limitan a promotores constitutivos, por ejemplo, promotores víricos, tales como promotores de CMV, SV40 y RSV, así como promotores regulables, por ejemplo, un promotor inducible o reprimible, tal como el promotor tet, el promotor hsp70 y un promotor sintético regulado por CRE, incluyendo cualquier fragmento que tenga actividad de promotor. La molécula de ácido nucleico, el casete de expresión y/o el vector de la invención puede introducirse en una célula mediante cualquier método que incluye, pero no se limita a una transformación mediada por calcio, una electroporación, una microinyección, una lipofección y similares.

III. Secuencias optimizadas y vectores y células hospedantes que codifican la luciferasa modificada

También se proporciona una molécula de ácido nucleico aislada (polinucleótido) que comprende una secuencia de ácido nucleico que codifica una luciferasa modificada de la invención, uno de sus fragmentos o una fusión de esta. En una realización, la molécula de ácido nucleico aislada comprende una secuencia de ácido nucleico que está optimizada para la expresión en al menos un hospedante seleccionado. Las secuencias optimizadas incluyen secuencias que tienen codones optimizados, es decir, los codones que se emplean con más frecuencia en un organismo con relación a otro organismo, por ejemplo, un organismo con parentesco lejano, así como modificaciones para añadir o modificar secuencias Kozak y/o intrones y/o para eliminar secuencias no deseables, por ejemplo, sitios potenciales de unión a factores de transcripción. Estas secuencias optimizadas pueden producir una mayor expresión, por ejemplo, un mayor nivel de expresión de proteínas, cuando se introducen en una célula hospedante.

En una realización, el polinucleótido incluye una secuencia de ácido nucleico que codifica una luciferasa modificada de la invención, estando optimizada la secuencia de ácido nucleico para la expresión en una célula hospedante de mamífero. En una realización, un polinucleótido optimizado ya no se hibrida con la correspondiente secuencia no optimizada, por ejemplo, no se hibrida con la secuencia no optimizada bajo condiciones de rigurosidad alta o intermedia. El término "rigurosidad" se emplea para referirse a condiciones de temperatura, fuerza iónica y la presencia de otros compuestos, bajo las cuales se realizan las hibridaciones de los ácidos nucleicos. Con unas condiciones de "alta rigurosidad", el apareamiento de las bases de los ácidos nucleicos solo se producirá entre fragmentos de ácidos nucleicos que tengan una alta frecuencia de secuencias de bases complementarias. Así, a menudo son necesarias condiciones "altas" o "intermedias" de rigurosidad cuando se desea que unos ácidos nucleicos que no son completamente complementarios entre sí se hibriden o se asocien entre ellos. En la técnica se conocen numerosas condiciones equivalentes que pueden emplearse para obtener condiciones de rigurosidad intermedias o bajas.

En otra realización, el polinucleótido tiene menos del 90%, por ejemplo, menos del 80% de identidad de secuencia de ácido nucleico con la correspondiente secuencia no optimizada, y opcionalmente codifica un polipéptido que tiene al menos 60%, por ejemplo, al menos 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99%, pero menos del 100% de identidad de secuencia de aminoácidos con el polipéptido codificado por la secuencia no optimizada. También se proporcionan construcciones, por ejemplo, casetes de expresión, y vectores que comprenden la molécula de ácido nucleico aislada, por ejemplo, con una secuencia de ácido nucleico optimizada, así como kits que comprenden la molécula de ácido nucleico aislada, la construcción o el vector.

Una molécula de ácido nucleico que comprende una secuencia de ácido nucleico que codifica una luciferasa modificada de la invención, uno de sus fragmentos o una fusión de esta, está opcionalmente optimizada para la expresión en una célula hospedante concreta y también está unida opcionalmente operablemente a secuencias reguladoras de la transcripción, por ejemplo, uno o más potenciadores, un promotor, una secuencia de terminación de la transcripción o una de sus combinaciones, para formar un casete de expresión.

En una realización, una secuencia de ácido nucleico que codifica una luciferasa modificada de la invención, uno de sus fragmentos o una fusión de esta, se optimiza mediante la sustitución de codones, por ejemplo, al menos 25% de los codones, en una secuencia de luciferasa de tipo salvaje, por codones que se emplean preferentemente en una célula concreta (seleccionada). Los codones preferidos tienen una frecuencia de uso de los codones relativamente alta en una célula seleccionada, y preferiblemente su introducción resulta en la introducción de relativamente pocos sitios de unión a factores de transcripción para los factores de transcripción presentes en la célula hospedante seleccionada, y relativamente pocos otros atributos estructurales no deseables. Así, el producto de ácido nucleico optimizado puede presentar un nivel mejorado de expresión debido a una mayor frecuencia de uso de codones, y un menor riesgo de desarrollar un comportamiento transcripcional inapropiado debido a un número reducido de secuencias reguladoras de la transcripción no deseables.

Una molécula de ácido nucleico aislada y optimizada puede tener una composición de codones que difiera de la de la correspondiente secuencia de ácido nucleico de tipo salvaje en más del 30%, 35%, 40%, o más del 45%, por ejemplo, 50%, 55%, 60% o más de los codones. Los ejemplos de codones para uso en la invención son los codones que se emplean con más frecuencia que al menos otro codón para el mismo aminoácido en un organismo concreto y, en una

realización, tampoco son codones de bajo uso en ese organismo y tampoco son codones de bajo uso en el organismo empleado para clonar o seleccionar la expresión de la molécula de ácido nucleico. Además, los codones para ciertos aminoácidos (es decir, aquellos aminoácidos que tienen tres o más codones) pueden incluir dos o más codones que se emplean con más frecuencia que el otro codón (codones) (no preferido(s)). La presencia de codones en la molécula de ácido nucleico que se emplean con más frecuencia en un organismo que en otro organismo da como resultado una molécula de ácido nucleico que, cuando se introduce en las células del organismo que emplea estos codones con más frecuencia, se expresa en esas células a un nivel que es mayor que la expresión de la secuencia del ácido nucleico de origen o de tipo salvaje en esas células.

En una realización de la invención, los codones que son diferentes son los que se emplean con más frecuencia en un mamífero, mientras que en otra realización, los codones que son diferentes son los que se emplean con más frecuencia en una planta. Los codones preferidos para diferentes organismos son conocidos en la técnica, por ejemplo, véase www.kazusa.or.jp/codon/. Un tipo concreto de mamífero, por ejemplo, un ser humano, puede tener un conjunto diferente de codones preferidos que otro tipo de mamífero. De modo similar, un tipo concreto de planta puede tener un conjunto diferente de codones preferidos que otro tipo de planta. En una realización de la invención, la mayoría de los codones que difieren son aquellos que son los codones preferidos en una célula hospedante deseada. Los codones preferidos para organismos, que incluyen mamíferos (por ejemplo, seres humanos) y plantas, son conocidos en la técnica (por ejemplo, Wada et al., *Nucl. Acids Res.*, 18:2367 (1990); Murray et al., *Nucl. Acids Res.*, 17:477 (1989)).

IV. Luciferasa para la potenciación de la estabilidad ilustrativa

Se ha mostrado que la luciferasa segregada por el camarón abisal *Oplophorus gracilirostris* posee muchas características interesantes, tales como una alta actividad, un alto rendimiento cuántico y una amplia especificidad de sustrato (coelenteracina, análogos de la coelenteracina). La reacción bioluminiscente de *Oplophorus* se produce cuando la oxidación de la coelenteracina (la luciferina) con oxígeno molecular es catalizada por la luciferasa de *Oplophorus*, lo que da como resultado una luz de una intensidad máxima a 462 nm y en los productos CO₂ y coelenteramida (Shimomura et al., *Biochemistry*, 17:994 (1978); esto difiere de Inouye 2000 que menciona 454 nm). La luminiscencia óptima se produce a pH 9 en presencia de NaCl 0,05-0,1 M a 40°C y, debido a la resistencia poco común de esta enzima al calor, la luminiscencia visible aparece a unas temperaturas por encima de 50°C cuando se emplea la enzima altamente purificada, o a más de 70°C cuando se emplea la enzima parcialmente purificada. A pH 8,7, la luciferasa nativa tiene un peso molecular de aproximadamente 130.000, y aparentemente comprende 4 monómeros de 31.000; a pH menores, la luciferasa nativa tiende a polimerizar.

La proteína madura consiste en proteínas de 19 kDa y 35 kDa (heterotetramero que consiste en dos componentes de 19 kDa y dos componentes de 35 kDa). La proteína de 19 kDa (OgLuc) se ha sobreexpresado como un monómero en *E. coli* y ha mostrado ser activa, aunque se produce predominantemente en forma de cuerpos de inclusión. Es probable que la formación de cuerpos de inclusión sea debida a la inestabilidad de la proteína dentro de la célula.

La estructura tridimensional de OgLuc no está disponible. Además, no están disponibles modelos conocidos basados en una homología, puesto que OgLuc no presenta homología de secuencia con otras luciferasas ni tampoco una similitud de secuencia global significativa con otras proteínas conocidas. Para generar un modelo se empleó un método de reconocimiento de plegamiento para identificar proteínas homólogas distantes. Empleando esta estrategia, tal como se describe a continuación, se identificó un conjunto de proteínas de unión a ácidos grasos (FABP) que pertenecen a la superfamilia de proteínas de la calicina, y se generó un modelo de homología de OgLuc basándose en las estructuras tridimensionales de tres de estas FABP.

Las calicinas son una superfamilia de proteínas cuyos miembros comparten estructuras en barril-β similares. Los miembros incluyen, pero no se limitan a proteínas de unión a ácidos grasos (FABP) y lipocalinas. La familia de proteínas FABP tiene una estructura en barril-β discontinua de diez cadenas; los barriles de la avidina y MPI, aunque tienen ocho cadenas, son más circulares en sección transversal que los de las lipocalinas y no tienen una hélice C-terminal o cadena I; mientras que la triabina tiene una geometría de barril similar pero posee una topología modificada. Las cadenas N- y C-terminales de las FABP y las lipocalinas pueden superponerse con gran similitud, siendo necesaria la pérdida (de FABP a lipocalina) o la ganancia (de lipocalina a FABP) de dos cadenas centrales necesarias para efectuar la transformación de una en otra (Flower et al., *Protein Science*, 2:753 (1993)). Además, más allá de alguna similitud funcional (unión a ligandos hidrófobos y/o interacción macromolecular), estas familias se caracterizan por un patrón de plegamiento similar (un barril-β antiparalelo dominado por una topología mayoritariamente +1), dentro del cual amplias partes de sus estructuras pueden ser consideradas equivalentes desde el punto de vista estructural, aunque las familias no comparten una similitud de secuencia global.

Estudios previos (Flower, *Protein Pept. Lett.*, 2:341 (1995)) han mostrado que los miembros de la superfamilia de la calicina también comparten un patrón estructural diferenciado. Un resto de arginina o lisina (procedente de la última cadena del barril-β) forma enlaces de hidrógeno con los grupos carbonilo de la cadena principal de la hélice similar a 3₁₀ N-terminal y se empaqueta sobre un triptófano conservado (procedente de la primera cadena del barril-β). Este patrón puede observarse en las estructuras de las lipocalinas del kernel, que también comparten una interacción conservada del bucle L6, y en las lipocalinas externas, más diversas desde el punto de vista estructural. También es

evidente en las otras cuatro familias que comprenden las calicinas. El examen de las estructuras disponibles de la estreptavidina y la avidina de pollo, del inhibidor de metaloproteinasa de *Erwinia chrysanthemi*, y de la estructura de la triabina, revela una disposición muy similar de restos en interacción. La mayoría de las FABP conocidas tienen una disposición de las interacciones de cadenas laterales similar a la descrita anteriormente, en la que un triptófano, procedente de la primera cadena del barril de FABP, se empaqueta contra una arginina procedente de una posición cercana al extremo de este último. Sin embargo, esta característica está ausente en un grupo de FABP más altamente divergentes, ejemplificadas por las FABP de músculo de insecto.

El modelo de homología de OgLuc muestra que la firma estructural de plegamiento de la calicina, que une con eficacia el N- y C-terminal entre sí mediante enlaces de hidrógeno, y que está presente en las tres FABP, no está completamente conservada en OgLuc. La firma estructural diferenciada (en la que una arginina o una lisina, que es capaz de formar una serie de enlaces de hidrógeno potenciales con los carbonilos de la cadena principal de una hélice 3_{10} corta, se empaqueta sobre un triptófano conservado de una manera no aleatoria y estructuralmente superponible) se corresponde con unos determinantes de secuencia comunes a los miembros de la familia de la calicina: un patrón de secuencia N-terminal característico, que muestra la conservación de restos clave, y un motivo C-terminal más débil. La conservación de restos e interacciones concretas a través de los miembros de las familias apoya la opinión de que hubo un origen evolutivo común, aunque muy lejano, para la superfamilia de la calicina. El presente modelo de OgLuc predice que el resto Asn166 de OgLuc cerca del C-terminal no es capaz de unirse con enlaces de hidrógeno a los carbonilos de la cadena principal cerca del N-terminal. Sin embargo, los modelos de mutantes que contienen Arg o Lys en la posición 166 sugieren que el restablecimiento de este motivo estructural podría mejorar la estabilidad estructural de OgLuc y su expresión/actividad en células.

Ejemplo 1

Las desventajas de OgLuc pueden solucionarse mediante ingeniería de proteínas, pero para que esto pueda hacerse de una manera eficaz se necesitaría conocer la estructura tridimensional (3D) de OgLuc. No existen modelos de la estructura terciaria ni estructuras terciarias experimentales publicados de OgLuc. Se empleó la formación de modelos de homología para generar un modelo de estructura terciaria de OgLuc. La construcción del modelo de homología comprende varias etapas, que incluyen la identificación de uno o más moldes estructurales tridimensionales, el alineamiento de la secuencia diana (por ejemplo, OgLuc) y una o más estructuras de molde, la construcción del modelo, y la evaluación de la calidad del modelo. La identificación de uno o más moldes estructurales tridimensionales para OgLuc no es intuitiva, porque los métodos de búsqueda de secuencias convencionales no identifican una similitud global significativa con ninguna proteína con una estructura terciaria conocida. Para solucionar este problema se emplearon dos estrategias para identificar homólogos de OgLuc remotos con estructura terciaria conocida.

Estrategia 1:

Se empleó una búsqueda en un banco de moldes basada en un modelo oculto de Markov (HMM) (Karplus et al., *Bioinformatics*, 14:846 (1998)) para detectar estructuras de molde lejanamente relacionadas empleando la herramienta de identificación de moldes SWISS-MODEL en <http://swissmodel.expasy.org/SWISS-MODEL.html> (Arnold et al., *Bioinformatics*, 22:195 (2006)).

El mejor molde de estructura tridimensional (mayor puntuación del valor E) identificado para OgLuc empleando esta estrategia fue una proteína de unión a ácidos grasos (FABP) (Protein Data Bank (PDB) n.º de registro 1VYF) (Angelucci et al., *Biochemistry*, 43:13000 (2004)). También se identificaron otras FABP con menor puntuación, que incluyen PDB n.º de registro 1PMP y 1CRB.

A continuación se muestran alineamientos ilustrativos de la secuencia diana (OgLuc, los restos 1-2 y 168-169 se han omitido) y las secuencias de los moldes de estructura tridimensional identificados (1VYF, 1PMP, 1CRB). Nótese que, debido a la baja similitud de secuencia, la colocación de los huecos en el alineamiento puede variar.

```

lvyf 1 GSMSSFLGKWKLSSEHNFDVMSKLGVSWATRQIGNTVTPTVFTTMDGDK.. 50
      F G W      N D V      G S      G VTP      G
Diana 3 --LADFVGDWQQTAGYNQDQVLEQGGSSSLFQALGVSVTPIQKVVLSENGl 52

lvyf 51 .....MTMLTESTFKN..LSCTFKF.....GEEF 72
      S F      FK      G
Diana 53 kadihviIPYEGLSGFQMGliEMIFKVvypvddhhfkiiIhygtlvidGVTP 104

lvyf 73 DEKTS DGRNVKSVVEKNSESKLTQTQVDPKNTTVIVREV.DGDTMKT TVTVG 123
      GR      N      R      VT
Diana 105 NMIDYFGRPYPGIAVFDGKQITVTGT LWNGNKIYDERLI nPDGSLLFRVTIN 156

lvyf 124 DVTAIRNYKRLS 135 (SEQ ID NO:5)
      VT R
Diana 157 GVTGWRLCENI 167 (SEQ ID NO:7)

lpmf 3 SNKFLGTWKLVSSENFDEYMKALGVGLATRKLGNLAKPRVIISKKGI.... 48
      F G W      N D      G      LG P      G
Diana 3 LADFVGDWQQTAGYNQDQVLEQGGSSSLFQALGVSVTPIQKVVLSENGlka 54

lpmf 49 .....ITIRTESPFKNTEISFKL.....GOEFEE 72
      P      L      G
Diana 55 dihvii pyeglsqfmgliEMIFKVvypvddhhfkiiIhygtlvidGVTPNM 106

lpmf 73 TTADNRKTKSTVTLARGSLNQQVQK.WNGNETTIKRKL.VDGK MVVECKMKDV 122
      R      WNGN      R      DG      V
Diana 107 IDYFGRPYPGIAVFDGKQITVTGT LWNGNKIYDERLI nPDGSLLFRVTINGV 158

lpmf 123 VCTRIYEKV 131 (SEQ ID NO:3)
      R E
Diana 159 TGWRLCENI 167 (SEQ ID NO:7)

lcrb 1 PVDFNGYWKMLSNENFEEYL RALDVNVALRKIANLLKPDKEIVQDGDH.... 48
      DF G W      N L      P V G
Diana 3 LADFVGDWQQTAGYNQDQVLEQGGSSSLFQALGVSVTPIQKVVLSENGlka 54

lcrb 49 .....MIIRTLSTFRNYIMDFQV.....GKEFEE 72
      MI      G
Diana 55 dihvii pyeglsqfmgliEMIFKVvypvddhhfkiiIhygtlvidGVTPNM 106

lcrb 73 DLTGIDDRKCMTTVSWDGDQLQCVQK.GEKEGRGWTQWI.EGDELHLEMRAE 122
      R      DG      I L
Diana 107 --IDYFGRPYPGIAVFDGKQITVTGT LWNGNKIYDERLI nPDGSLLFRVTIN 156

lcrb 123 GVTCKQVFKKVH 134 (SEQ ID NO:4)
      GVT
Diana 157 GVTGWRLCENI- 165 (SEQ ID NO:7)

```

Estrategia 2:

5 También se empleó un método de reconocimiento de plegamiento empleando el meta-servidor GeneSilico en <https://genesilico.pl/meta2> (Kurowski et al., *Nucl. Acids Res.*, 31:3305 (2003)) para identificar homólogos de Ogluc remotos con estructura terciaria conocida.

10 El plegamiento de una proteína es una clasificación estructural tridimensional. Las proteínas que comparten el mismo plegamiento tienen una disposición similar de estructuras secundarias regulares, pero no siempre muestran pruebas de una relación evolutiva al nivel de secuencia de la proteína.

Empleando este método se identificaron tres moldes de estructura tridimensional con la mayor puntuación (PDB n.º

de registro 1VYF, 1PMP, y 1CRB). A continuación se muestran alineamientos de la secuencia diana (OgLuc) ilustrativos y las secuencias de los moldes de estructura tridimensional (1VYF, 1PMP, 1CRB). Nótese que, debido a la baja similitud de secuencia, la colocación exacta de los huecos en el alineamiento es difícil de predecir con confianza.

5

OgLuc y 1PMP:

```
--SNKFLGTWKLVSSENFDEYMKALGVGLATRKLGNLAKPRVIISKKG-----DIITIRTE-----
FTLADFVGDWQQTAGYNQDQVLEQGGSSLFQALGVSVTPIQKVVLSGENGLKADIHVIIPYEGLSGFQMGLIEMIFKVV

-----SPFKNTEISFKLGQEFEEETAD-----NRKTKSTVTLARGSLNQV-QKWNGNETTIKRKLV-DGKMVVECKMKDV
YPVDDHHFKIILHYGTL--VIDGVTPNMI DYFGRPYPGIAVFDGKQITVTGTLWNGNKIYDERLINPDGSLLFRVTINGV

VCTRIYEKV-- (1PMP) (SEQ ID NO:3)
TGWRLCENILA (OgLuc) (SEQ ID NO:1)
```

OgLuc y FABP:

10

```
--SNKFLGTWKLVSSENFDEYMKALGVGLATRKLGNLAKPRVIISKKG-----DIITIRTESP-----
--PVDFNGYWKMLSNENFEEYLRALDVNVALRKIANLLKPDKEIVQDG-----DHMIIRTLST-----
GSMSSFLGKWKLSSESHNFDAVMSKLGVSWATRQIGNTVTPTVTFTMDG-----DKMTMLTEST-----
FTLADFVGDWQQTAGYNQDQVLEQGGSSLFQALGVSVTPIQKVVLSGENGLKADIHVIIPYEGLSGFQMGLIEMIFKVV

-----FKNTEISFKLGQEFEEETA-----DNRKTKSTVTLAR-GSLNQV-QKWNGNETTIKRKLV-DGKMVVECKMKD
-----FRNYIMDFQVGKEFEEDLT---GIDDRKCMTTVSWDG-DKLQCV-QKGEKEGRGWTQWIE-GDELHLEMRAEG
-----FKNLSCTFKFGEEFDEKTS-----DGRNVKSVVEKNSKLTQT-QVDPKNTTVIVREVD-GDTMKTTVTVGD
YPVDDHHFKIILHYGTL--VIDGVTPNMI DYFGRPYPGIAVFDG-KQITVTGTLWNGNKIYDERLINPDGSLLFRVTING

VVCTRIYEKV-- (1PMP) (SEQ ID NO:3)
VTCKQVFKKVH- (1CRB) (SEQ ID NO:4)
VTAIRNYKRLS- (1VYF) (SEQ ID NO:5)
VTGWRLCENILA (OgLuc) (SEQ ID NO:1)
```

Empleando la información generada en las anteriores estrategias, se generaron modelos de homología de OgLuc basándose en tres moldes de estructura tridimensional de FABP (1PMP, 1CRB, y 1VYF) empleando los programas informáticos Discovery Studio y MODELER (Accelrys Software Inc.).

15

La figura 1 también muestra los alineamientos de la estructura secundaria de FABP y OgLuc. 1PMP, 1CRB, 1VYF son los códigos de registro de Protein Data Bank (www.rcsb.org) de las secuencias FABP ilustrativas con estructura tridimensional conocida. "PDB" significa la asignación de estructura secundaria proporcionada por los autores que depositaron la información de la estructura tridimensional en Protein Data Bank. "DSC" significa la predicción de la estructura secundaria basándose en el método DSC (King et al., *Protein Science*, 5:2298 (1996)). "Kabsch y Sander" significa la predicción de la estructura secundaria basándose en el método de Kabsch y Sander (Kabsch y Sander, *Biopolymers*, 22:2577 (198)). Los recuadros rojos indican la extensión aproximada de los elementos de hélice de la estructura secundaria, las flechas azules indican la extensión aproximada de los elementos de lámina-beta de la estructura secundaria, y las barras grises indican otras estructuras secundarias distintas de hélices o láminas-beta. En los alineamientos pueden observarse los motivos de secuencia centrados en los restos conservados de la firma estructural de la calicina (Flower et al., *Biochem. Biophys. Acta.*, 16:1088(2000)). El motivo MOTIF1 N-terminal más altamente conservado incluye el resto Trp10 de OgLuc, y el motivo MOTIF2 C-terminal menos bien conservado incluye el resto N166 de OgLuc. Para el segundo alineamiento, el porcentaje de identidades de secuencia de proteína apareadas aproximadas son: OgLuc-1PMP 14%, OgLuc-1CRB 9%, y OgLuc-1VYF 15%.

20

La figura 2 muestra los alineamientos de la estructura secundaria de la luciferasa de dinoflagelados, FABP y OgLuc. 1VPR y 1HMR son los códigos de registro de Protein Data Bank (www.rcsb.org) para secuencias con estructura tridimensional conocida. 1VPR es el dominio 3 de la luciferasa de dinoflagelados, y 1HMR es la FABP de músculo

25

30

humano, la proteína más relacionada con la luciferase de dinoflagelados (Schultz et al., *PNAS USA*, 102:1378 (2005)). “Kabasch y Sander” significa la predicción de la estructura secundaria basándose en el método de Kabasch y Sander (Kabasch y Sander, *Biopolymers*, 22:2577 (1983)). Los recuadros rojos indican la extensión aproximada de los elementos de hélice de la estructura secundaria, las flechas azules indican la extensión aproximada de los elementos de lámina-beta de la estructura secundaria, y las barras grises indican otras estructuras secundarias distintas de hélices o láminas-beta. 1VPR tiene la SEQ ID NO:21; 1HMR tiene la SEQ ID NO:22.

La figura 3 muestra el alineamiento de las secuencias de aminoácidos de OgLuc y diversas FABP (SEQ ID NO:1, 3, 4, 5, y 17-20, respectivamente), basándose en la superposición de las estructuras tridimensionales de las FABP.

Ejemplo 2

Las proteínas de unión a ácidos grasos (FABP) pertenecen a la superfamilia de proteínas de la calicina. Las calicinas no presentan una similitud global significativa al nivel de secuencia, sino que comparten una estructura de barril-beta relacionada con una firma estructural diferenciada: una arginina o una lisina (cerca del C-terminal) que es capaz de formar una serie de enlaces de hidrógeno potenciales con los carbonilos de la cadena principal de una hélice 3_{10} corta y se empaqueta sobre un triptófano conservado (cerca del N-terminal) (Flower et al., *Biochem. Biophys. Acta*, 1482:9 (2000)). En el modelo de OgLuc generado en el ejemplo 1, la firma estructural de calicina solo está presente parcialmente. El triptófano conservado (Trp10) cerca del N-terminal (tal como un resto en una lámina-beta N-terminal de un barril-beta) se empaqueta sobre una asparagina (Asn166) en lugar de una arginina o una lisina cerca del C-terminal (tal como un resto en una lámina-beta C-terminal de un barril-beta). El presente modelo predice que la cadena lateral de la asparagina, más corta, parece incapaz de formar enlaces de hidrógeno con restos cerca del N-terminal (en la lámina-beta N-terminal del barril-beta). Los modelos de OgLuc en los que se realizaron las sustituciones Asn166Arg y Asn166Lys demostraron que las cadenas laterales de la arginina y la lisina, más largas, en OgLuc deberían poder formar uno o más enlaces, por ejemplo, uno o más enlaces de hidrógeno, con los carbonilos de la cadena principal y/o las cadenas laterales de los restos cerca del N-terminal. Por ejemplo, pueden formar uno o más enlaces de hidrógeno con los restos Asp9 y/o Gly8 y/o Asp5 de OgLuc cerca del N-terminal. Además, pueden formar uno o más enlaces de hidrógeno con uno o más restos en otros elementos de la estructura secundaria que están en proximidad espacial cercana con la posición 166, por ejemplo, Asn144 y/o Gly147. Así, el restablecimiento de la firma estructural de calicina en OgLuc con una mutación Asn166Arg o Asn166Lys puede unir entre sí de forma eficaz los dos terminales del barril-beta (o láminas-beta terminales del barril-beta) y quizás otros elementos de la estructura secundaria. Esto puede mejorar la estabilidad global de la estructura de la proteína y, por tanto, la actividad de OgLuc.

Una secuencia ilustrativa de la proteína OgLuc es:

```
FTLADFGVDW QQTAGYNQDQ VLEQGGLSSL FQALGVSVTP IQKVVLSENGEN
GLKADIHVII PYEGLSGFQM GLIEMIFKVV YPVDHDFKI ILHYGTLVID
GVTPNMIDYF GRPYPGIAVF DGKQITVTGT LWNGNKIYDE RLINPDGSL
FRVTINGVTG WRLCENILA (SEQ ID NO:1; 169 aminoácidos, Asn166 subrayado en negrita).
```

Una secuencia ilustrativa de nucleótidos de OgLuc es:

```
atggtgtttacccttggcagatttcgttggagactggcaacagacagctggatacaaccaagatcaagtgttagaacaaggaggattgtctagtct
gttccaagccctgggagtgctcagtcacccaatccagaaagtgtgctgctgtgggagaatgggttaaagctgatattcatcatccctta
cgagggactcagtggtttcaaattggctgattgaaatgatctcaaagtgtttaccagtgatgatcatcattcaagattattctccaitatggt
acactcgttattgacggtgtgacaccaacatgattgactactttggacgccctaccctggaattgctgtgttgacggcaagcagatcacagtta
ctggaactctgtggaacggcaacaagatctatgatgagcgcctgatcaaccagatgggtcactcctctccgcgttactatcaatggagtcacc
ggatggcgcctttgcgagAACattctgcc (SEQ ID NO:2).
```

El codón AAC de SEQ ID NO:2, que está escrito en mayúsculas en el anterior listado, se corresponde con la posición del aminoácido 166 en la secuencia de OgLuc madura de tipo salvaje de SEQ ID NO:1. La secuencia de nucleótidos de SEQ ID NO:2 también incluye un codón ATG (metionina/señal de inicio) y un codón GTG (valina) al principio para facilitar el uso en sistemas de expresión. No obstante, la numeración de los aminoácidos empleada a lo largo de esta solicitud para identificar restos sustituidos se ofrece con relación a la secuencia del polipéptido de OgLuc maduro de tipo salvaje de SEQ ID NO:1. La secuencia de OgLuc de tipo salvaje natural puede sintetizarse inicialmente con otros aminoácidos que luego se escinden, lo que da como resultado la generación de un polipéptido maduro de tipo salvaje, tal como el que aparece en SEQ ID NO:1. Por ejemplo, puede estar presente una secuencia señal (por ejemplo, para dirigir a la proteína naciente hacia un orgánulo concreto, tal como el retículo endoplásmico y/o para dirigir la proteína para la secreción) al principio de la proteína naciente, y después puede escindirse para producir la proteína madura

de tipo salvaje.

A continuación se muestra un alineamiento de OgLuc y tres FABP ilustrativo:

```
--SNKFLGTWKLVSSENFDEyMKALGVGLATRKLGNLAKPRV IISKKG-----DIITIRTESP-----
--PVDFNGYWKMLSNENFEEYLRALDVNVALRKIANLLKPDKEIVQDG-----DHMIIRTLST-----
GSMSSFLGKWKLSESHNFDAVMSKLGVSWATRQIGNTVTPTVTFTMDG-----DKMTMLTEST-----
FTLADFVGDWQQTAGYNQDQVLEQGGSSLFQALGVSVTPIQKVVLSENGLKADIHVIIPYEGLSGFQMG LIE
      11              33              44              54
```

```
-----FKNTEISFKLQGFEETTA-----DNRKTKSTVTLAR-GSLNQV-QKWNGNETTIKRKLV-
-----FRNYIMDFQVGKEFEEDLT---GIDDRKCMTTVSWDG-DKLQCV-QKGEKEGRGWTQWIE-
-----FKNLSCTFKFGEEFDEKTS-----DGRNVKSVVEKNSESKLTQT-QVDPKNTTVIVREVD-
MIFKVVYPVDDHHFKIILHYGTL--VIDGVT_PNMIDYFGRPY_PGIAVFDG-KQITVTGTLWNGNKIYDERLINP
75              114              115              124              135
```

5

DGKMVVECKMKDVVCTRIYEKV-- (SEQ ID NO:3)

GDELHLEMRAEGVTCKQVFKKVH- (SEQ ID NO:4)

GDTMKTTVTVDVTAIRNYKRLS- (SEQ ID NO:5)

KGSLLFRVTINGVTGWRLCENILA (SEQ ID NO:1)

Ejemplo 3:

10

Generación de variantes de luciferasa modificada con luminiscencia aumentada

A menos que se indique lo contrario, se generaron variantes de una secuencia de OgLuc de partida con sustituciones aleatorias empleando el sistema basado en PCR mutagénico propenso a errores GeneMorph II Random Mutagenesis Kit (Stratagene; Daughtery, PNAS USA, 97(5):2029 (2000)), según las instrucciones del fabricante, y una saturación NNK según se conoce en la técnica. Los variantes resultantes se construyeron en el contexto del vector pF1K Flexi® para la expresión basada en T7 (Promega Corp.) y se emplearon para transformar *E. coli* KRX empleando técnicas conocidas en la técnica. El banco resultante se expresó en *E. coli* y se seleccionó para las variantes que presentaban una mayor emisión de luz, comparado con la proteína OgLuc de partida. Se utilizaron técnicas de secuenciación convencionales conocidas en la técnica para identificar las sustituciones de aminoácidos en cada clon de interés.

Se generaron variantes de una secuencia de OgLuc de partida con mutaciones específicas empleando el kit de mutagénesis dirigida a sitio basada en oligos QuikChange Site-Directed Mutagenesis Kit (Stratagene; Kunkel, PNAS USA, 82(2):488 (1985)), según las instrucciones del fabricante.

25

Ejemplo 4:

Métodos para medir la emisión de luz y la estabilidad de la señal

Se cultivaron clones de *E. coli* que contenían el ADN plasmídico que codifica las variantes de luciferasa modificada con sustituciones de aminoácidos en OgLuc en una placa de 96 pocillos y se indujeron con una inducción de alejamiento, es decir, una autoinducción (Shagat et al., "KRX Autoinduction Protocol: A Convenient Method for Protein Expression," Promega Notes, 98:17 (2008)) durante 17 horas. Cada variante y la correspondiente luciferasa de partida tenía 6 pocillos replicados. Las células se lisaron empleando un tampón de lisis que consiste en HEPES 150 mM, pH 8,0, tiourea 100 mM, 0,1x PLB (Promega Corp., n.º de catálogo E194A), lisozima 0,1 mg/ml y ADNasa RQ1 0,001 U/μl, y se midió la luminiscencia empleando reactivos de sustrato de luciferasa de *Renilla* (Promega Corp.) en un luminómetro Infinite 500 Tecan. Las mediciones se tomaron inmediatamente después de la adición con una inyección de un tampón de ensayo de tergitol al 0,5% "Glo" ("tergitol al 0,5%"), que contiene KCl 150 mM, CDTA 1 mM, DTT 10 mM, tergitol al 0,5%, coelenteracina 20 μM (Promega Corp.), o un tampón RLAB "Destello" ("Flash") (Promega Corp.) que contiene coelenteracina 20 μM (Promega Corp.) ("RLAB") a la muestra de lisado. Esta medición de la luminiscencia, tomada inmediatamente después de la adición, es la medición en el momento "T = 0" y, en diversas realizaciones, se toma como la medición de la emisión de luz total (luminiscencia) generada por la muestra. El

40

promedio de la luminiscencia de los 6 replicados se comparó entre los variantes con la de la correspondiente luciferasa de partida. En diversas realizaciones, las mediciones de luminiscencia se normalizaron a la correspondiente luciferasa de partida de interés, por ejemplo, OgLuc sintética, y se indican en ciertas realizaciones como una mejora, un aumento o similares “en X veces” (es decir, en 2 veces, en 3 veces, en 4,5 veces, etc.).

La estabilidad de la señal de un clon variante se determinó volviendo a leer la placa multiples veces después de la adición del tampón de ensayo a la muestra, por ejemplo, midiendo la luminiscencia cada 30 segundos o cada minuto, durante una longitud de tiempo. La semivida de la señal se determinó empleando estas mediciones, y el promedio de los 6 pocillos replicados se comparó entre las variantes con la correspondiente luciferasa de partida. La semivida que indica la estabilidad de la señal se normalizó a la correspondiente luciferasa de partida de interés, por ejemplo, OgLuc.

Ejemplo 5:

Método para medir la estabilidad de la proteína, concretamente la termoestabilidad

Se prepararon muestras de lisados a partir de cultivos inducidos tal como se describe en el ejemplo 4. Las muestras de lisado en placas de 96 pocillos replicadas se incubaron a diversas temperaturas, que incluyen, por ejemplo a 22, 30, 37, 42, 50 o 54°C. En diferentes momentos, las placas se cultivaron a -70°C. Antes de medir la luminiscencia según se describe en el ejemplo 4, cada placa se descongeló a temperatura ambiente, es decir, 22°C, durante 10 minutos. Las muestras se ensayaron con el tampón de ensayo de tergitol al 0,5% descrito en el ejemplo 4. La medición “T = 0”, según se describe en el ejemplo 4, para cada placa del momento del tiempo, se empleó para determinar la semivida de la proteína. La semivida, que indica la estabilidad de la proteína, se normalizó a la correspondiente luciferasa de partida de interés, por ejemplo, OgLuc.

Ejemplo 6:

Generación de una luciferasa modificada con mayor emisión de luz

Para estudiar si el restablecimiento de la firma estructural de la calcina en OgLuc podría mejorar la actividad y la estabilidad globales de las proteínas, se diseñaron versiones sintéticas de la secuencia de OgLuc. Las versiones sintéticas incluyen el uso optimizado de codones para *E. coli* y células de mamífero, y codones para Arg o Lys que sustituyen a Asn en la posición 166. Tal como se mencionó previamente, la numeración se basa en SEQ ID NO:1. La optimización de codones (para *E. coli*) y los cambios de nucleótidos para el codón 166 a Arg o Lys se realizaron por medios sintéticos (Gene Dynamics, LLC). En el clon OgLuc+N166R, el codón AAC se cambió a CGT (que codifica Arg). En el clon OgLuc+N166K, el codón AAC se cambió a AAA (que codifica Lys).

Los genes de OgLuc sintéticos se subclonaron en un vector adecuado para la sobreexpresión en bacterias o lisados de reticulocitos de conejo TnT® (Promega Corp.; vector pF1K Flexi® para sistemas de expresión basados en T7), y se empleó para transformar *E. coli* KRX. Se seleccionaron colonias individuales, se cultivaron, se indujeron con ramnosa, se lisaron empleando lisozima y una única congelación-descongelación, y se midieron para la luminiscencia empleando reactivos de sustrato de luciferasa de *Renilla* (Promega Corp.) en un luminómetro Veritas. Las reacciones de reticulocitos de conejo TnT® se realizaron según los protocolos del fabricante (Promega Corp.) y se midieron de la misma forma que los lisados bacterianos.

Los mutantes se compararon con la proteína OgLuc parental (es decir, de partida) sintética para la producción de emisión de luz total (luminiscencia). En *E. coli*, se observó una mejora en 5 veces y 10 veces (N166K y N166R, respectivamente) en la luminiscencia con coelenteracina como sustrato. En los lisados de TnT®, la mejora fue entre 4 veces y 7 veces (N166K y N166R). Estas secuencias (que contienen Arg o Lys en la posición 166) representan variantes de OgLuc que producen una mayor estabilidad.

Se analizaron diversos variantes de OgLuc con una sustitución de un aminoácido en la posición 166 para el brillo, es decir, se seleccionaron variantes que eran al menos 1,2 veces más brillantes que la OgLuc de tipo salvaje. Las siguientes sustituciones produjeron un variante que era al menos 1,2 veces más brillante que OgLuc de tipo salvaje: N166K; N166R; N166A; N166L; N166P; N166Q; y N166S (véase la tabla 1). La tabla 1 muestra la variante más brillante, indicado por la mejora en número de veces con respecto a OgLuc de tipo salvaje, que tiene la sustitución del aminoácido N166R.

Tabla 1 - Resumen de la mejora en número de veces de la luminiscencia en variantes de OgLuc con una sustitución de un aminoácido en la posición 166 frente a OgLuc de tipo salvaje

Sustitución del aminoácido en la posición 166	Mejora en el plegamiento
R	10
K	4
A	3
L	3
P	2
Q	2
S	2

5 Una mutagénesis empleando PCR propensa a errores y saturación NNK, según se describe en el ejemplo 3, del variante OgLuc+N166R produjo variantes con mayor brillo, por ejemplo, al menos 1,2 veces más brillantes, con relación al variante OgLuc+N166R. La tabla 2 resume estos variantes que comprenden la sustitución N166R, así como una de las siguientes sustituciones en los restos 2 (S), 4 (E, S, R, G, D, T o L), 11 (R, V, I, L, K o T), 33 (K), 44 (I o L), 45 (E), 54 (F, T, V, G, W, S, o L), 68 (V, Y), 75 (R, K, Q, G, T o A), 104 (L), 115 (E, I, Q, L, V, G, H, R, S, C, A, o T), 124 (K), 135 (K), 138 (V, I, N, T, L, C, R, M o K), 139 (E), 167 (V), o 169 (L). La tabla 2 muestra la mejora del plegamiento en forma de la mejora en número de veces de la luminiscencia del variante frente al correspondiente variante OgLuc+N166R de partida empleando RLAB y utilizando el promedio de la señal en el intervalo de 4-6 minutos después de comenzar la reacción, por ejemplo, después de la inyección del sustrato. Para cada sustitución de aminoácido listada, la sustitución que produce la mayor mejora se lista en primer lugar y la sustitución que produce la menor mejora se lista al final. Los variantes que muestran la mayor mejora incluyen variantes que contienen una sustitución en el resto 4, 54, o 138.

Tabla 2 - Resumen de la mejora en número de veces de la luminiscencia de los variantes de OgLuc+N166R frente al correspondiente OgLuc+N166R de partida

Posición	Aminoácido	Codón	Mejora en n.º de veces del brillo (RLAB), promedio de 4-6 min (con rel. a N166R)
2	S	TCC	9
4	E	GAG	20
4	S	AGT	7
4	R	AGG	6
4	G	GGG	4
4	D	GAT	4
4	T	ACG	3
4	L	CTG	3
11	R	CGG	13
11	V	GTG	6
11	I	ATT	6
11	L	CTT	3
11	K	AAG	3
11	T	ACT	2
33	K	AAG	10
44	I	ATT	25
44	L	CTT	2
45	E	GAG	2
54	F	TTT	10
54	T	ACT	8

ES 2 795 287 T3

54	V	GTT	6
54	G	GGG	5
54	S	AGT	4
54	W	TGG	3
54	L	TTG	2
68	V	GTT	2
68	Y	TAT	3
72	Q	CAG	3
75	R	AGG	6
75	K	AAG	5
75	Q	CAG	5
75	G	GGT	4
75	T	ACG	4
75	A	GCG	4
104	L	CTT	10
115	E	GAG	20
115	I	ATT	4
115	Q	CAG	3
115	L	CTT	3
115	V	GTT	3
115	G	GGG	3
115	H	CAT	3
115	R	CGG	2
115	S	AGT	2
115	C	TGT	2
115	A	GCT	2
124	K	AAA	8
135	K	AAG	10
138	V	GTG	10
138	I	ATT	8
138	T	ACG	6
138	L	CTG	5
138	C	TGT	6
138	R	CGG	5
138	M	ATG	4
138	K	AAG	3
139	E	GAG	13
167	V	GTT	40
169	L	TTG	10

Otras variantes del variante OgLuc+N166R presentan más de una sustitución de un aminoácido. Estas otras variantes se listan en la tabla 2 en la que se listan las sustituciones de aminoácidos y la mejora en número de veces de la luminiscencia del variante OgLuc+N166R frente al correspondiente OgLuc N166R de partida. Se encontraron otras variantes que incluían mutaciones silenciosas, es decir, cambios en nucleótidos que no alteran el aminoácido codificado por ese codón.

5

Tabla 3 - Resumen del aumento en número de veces de la luminiscencia de los variantes OgLuc+N166R con más de una sustitución de un aminoácido y/o mutaciones silenciosas frente al correspondiente OgLuc+N166R de partida

N.º de veces frente a N166R	Cambio del aminoácido con respecto a N166R (codones)
6	E23V (gta), S28P (cct), I143V (ctc)
15	A4S (gca), L34M (atg), I76V (gtc)
2	G51V (ggt), I99V (ggt)
13	L3L (tta), S37S (tcg), V44V (gta)
5	L3L (tta), L27M (atg)
5	L3L (tta)
4	L3L (tta), Q32L (cta), K43R (aga)
3	L72Q (cag), G10G (ggt)
2	N144K (aag), A54A (gca)

5 Ejemplo 7:

Evaluación de las sustituciones específicas en luciferasas modificadas

Se generaron otras variantes de OgLuc mediante mutagénesis específica dirigida a sitio según se describe en el ejemplo 3 para que tuviesen una sustitución en una de las siguientes posiciones: 2, 4, 11, 44, 54, 90, 115, 124 o 138 con relación a SEQ ID NO:1. En el ejemplo 6 se demostró que las sustituciones en estas posiciones en combinación con N166R presentaban una mayor emisión de luz total (luminiscencia), comparado con OgLuc WT. En las figuras 5A-5C, 6A-6C, 7A-7C, 8, 9A-9D, 10A-10C, 11A-11B, 12A-12B y 33A-33E, "WT," "N166R," y "T2T" se refieren a las proteínas codificadas por SEQ ID NO:2, 14 y 32, respectivamente, "T2T+N166R" se refiere a la proteína codificada por SEQ ID NO:32, que tiene una sustitución en N166R, "A4E," "Q11R," "V44I," "A54F," "A54F+N166R," "A54I," "P115E," "P155E+N166R," "Y138I," "Q124K," "Y138C+N166R," y "I90V" se refieren cada uno a la proteína codificada por SEQ ID NO:2 que tiene una sustitución en los respectivos restos indicados en la columna "Muestra" en la figura 5A. Estos variantes se evaluaron midiendo la luminiscencia según se describe en el ejemplo 4. Las figuras 5A-5C y 7A-7C resumen el promedio de luminiscencia en T = 0 de los variantes de OgLuc WT empleando tergitol al 0,5% (figuras 5A-5C) o RLAB (figuras 7A-7C). El aumento en número de veces de la luminiscencia de los variantes frente a OgLuc WT se muestra en las figuras 6A-B (tergitol al 0,5%) y la figura 8 (RLAB). El aumento en número de veces de la luminiscencia de los variantes frente al variante N166R se muestra en las figuras 33A (tergitol al 0,5%) y 33B (RLAB). Las figuras 5B, 6B, y 7B muestran los mismos datos que las figuras 5C, 6C, y 7C, respectivamente, pero a diferente escala para permitir ver las barras más pequeñas con más claridad.

Para determinar si las sustituciones de aminoácidos en los diferentes variantes también tenían efecto sobre la estabilidad de la señal, se midió la estabilidad de la señal para cada variante. La estabilidad de la señal de los variantes se midió como se describió en el ejemplo 4 y se muestra en las figuras 9A-9C (tergitol al 0,5%) y las figuras 10A-10C (RLAB) como la emisión de luz total (luminiscencia) a lo largo del tiempo. La semivida de la señal de cada variante se determinó a partir de estos datos y se muestra en la figura 9D (tergitol al 0,5%) y las figuras 11A-11B (RLAB). La semivida de la señal de cada variante se normalizó al variante N166R y se muestra en la figura 33C.

Para determinar si las sustituciones de aminoácidos en los diferentes variantes también tenían efecto sobre la estabilidad de las proteínas (concretamente, la termoestabilidad) se midió la estabilidad de las proteínas de cada variante a 22°C según se describió en el ejemplo 5 y se muestra en las figuras 12A-12B. A 22°C, la proteína del variante OgLuc A54F+N166R tiene una semivida de 178 minutos, mientras que el variante OgLuc P115E+N166R tiene una semivida de casi 120 minutos, comparado con OgLuc WT, que tiene una semivida de 38 minutos.

La figura 33D resume la semivida en minutos a 22°C de los variantes de OgLuc, comparado con OgLuc WT que aparece en las figuras 12A-B y 17 normalizado al variante N166R.

La figura 33E resume el aumento en número de veces de la luminiscencia, la semivida de la señal y la semivida a 22°C mostrados en las figuras A-D.

45 Ejemplo 8:

Evaluación de sustituciones específicas en luciferasas modificadas

Se generaron otras variantes de OgLuc sintéticos con sustituciones en los sitios 33 y 68. De modo específico, se realizaron las sustituciones A33K y F68Y en la secuencia de OgLuc WT (identificadas como "WT A33K" y "WT F68Y")

en las figuras 13A-13B, 14A-14B, 15A-15B, 16A-16B, 17, y 33A-33E) y de la variante OgLuc+N166R (identificadas como "N166R A33K" y "N166R F68Y" en las figuras 13A-13B, 14A-14B, 15A-15B, 16A-16B, 17, y 33A-33E) y se compararon con la correspondiente OgLuc WT de partida (identificada como "WT" en las figuras 13A-13B, 14A-14B, 15A-15B, 16A-16B, 17, y 33A-33E) y la variante OgLuc+N166R (identificada como "N166R" en las figuras 13A-13B, 14A-14B, 15A-15B, 16A-16B, 17, y 33A-33E). El promedio de la luminiscencia en T = 0 de las variantes de OgLuc A33K y F68Y empleando tergitol al 0,5% y RLAB se muestra en las figuras 13A y 13B, respectivamente. Las variantes A33K y F68Y tienen mayor luminiscencia, comparado con la respectiva y correspondiente OgLuc de partida, tal como se muestra por el aumento en número de veces de la luminiscencia de las variantes frente a OgLuc WT en las figuras 14A (tergitol al 0,5%) y 14B (RLAB). Los A33K y F68Y por separado sobre el fondo de tipo salvaje mostraron un aumento en 1,6 y 1,7 veces frente a WT empleando RLAB (véase la figura 14B) y en 3,8 y 3,9 veces frente a WT con tergitol al 0,5% (figura 14A). Los A33K y F68Y por separado sobre el fondo de OgLuc+N166R mostraron un aumento en 5,1 y 3,3 veces frente a OgLuc WT empleando RLAB (véase la figura 14B) y un aumento en 9,2 y 5 veces frente a OgLuc WT empleando tergitol al 0,5% (figura 14A).

El aumento en número de veces de la luminiscencia de los variantes frente al variante OgLuc+N166R se muestra en las figuras 33A (RLAB) y 33B (tergitol al 0,5%). La sustitución A33K sobre el fondo de tipo salvaje mostró un aumento en 2,6 (tergitol al 0,5%) y en 0,6 (RLAB) veces de la luminiscencia frente a la variante OgLuc+N166R (véase la figura 33A y 33B). La sustitución F68Y sobre el fondo de tipo salvaje mostró un aumento en 2,7 (tergitol al 0,5%) y en 0,7 (RLAB) veces frente al variante OgLuc+N166R (véase la figura 33A y 33B). La sustitución A33K sobre el fondo de la variante OgLuc+N166R mostró un aumento en 6,3 (tergitol al 0,5%) y en 2,0 (RLAB) veces frente al variante OgLuc+N166R (véase la figura 33A y 33B). La sustitución F68Y sobre el fondo de OgLuc+N166R mostró un aumento en 3,4 (tergitol) y en 1,3 (RLAB) veces frente a N166R (véase la figura 33A y 33B).

Se midió la estabilidad de la señal de las variantes A33K y F68Y según se describió en el ejemplo 4 empleando tergitol al 0,5% (figuras 15A-15B) y RLAB (figuras 16A-16B). La semivida de la señal del variante A33K sobre el fondo de OgLuc WT fue mayor que la semivida de OgLuc WT, pero menor que sobre el fondo de la variante OgLuc+N166R cuando se emplea tergitol al 0,5% (figura 15B) o RLAB (figura 16B). La semivida de la señal del variante F68Y sobre el fondo de OgLuc WT fue mayor que la semivida de OgLuc WT empleando tergitol al 0,5% (figura 16B), pero menor que sobre cualquier fondo empleando RLAB (figura 15B).

La estabilidad de las proteínas (concretamente, la termoestabilidad) de las variantes A33K y F68Y se midió como se describió en el ejemplo 5 a 22°C y se muestra en la figura 17. Las sustituciones A33K y F68Y sobre el fondo de la variante N166R presentan una mayor semivida, de modo específico 72 y 78 minutos, comparado con OgLuc WT y el variante N166R, que es de 55 y 67 minutos, respectivamente (figura 17). Las sustituciones A33K y F68Y sobre el fondo de OgLuc WT presentan unas semividas de 58 y 57 minutos, respectivamente (figura 17).

Ejemplo 9:

Evaluación de combinaciones centrales específicas de sustituciones en luciferasas modificadas - emisión de luz

Para determinar si una combinación de dos o más sustituciones de aminoácidos en OgLuc proporciona una mayor mejora en la luminiscencia, se generaron diferentes variantes (denominados C1-C3) de OgLuc que contenían las siguientes sustituciones de aminoácidos: C1: N166R, Q11R, A33K, A54F, P115E, Q124K, Y138I y V44I (el resto 44 puede ponerse en contacto con el sustrato), C2: V45E, N135K, I167V, P104L, y D139E (nótese que dos de estas están en sitios que pueden ponerse en contacto con el sustrato); C3: S28P, L34M, G51V, I99V, e I143L. Estas variantes de combinación central se generaron mutando T2T OgLuc mediante mutagénesis específica dirigida a sitio según se describió en el ejemplo 3. La variante C1 se volvió a mutar para que contuviese una sustitución del aminoácido A4E para crear la variante C1+A4E. También se crearon combinaciones de estas variantes con la sustitución A4E, por ejemplo, C1+C2+A4E y C1+C3+A4E. Estos clones recombinantes se construyeron empleando una mutagénesis específica dirigida a sitio basada en oligonucleótidos, seguido de una subclonación en el vector pF4Ag (que contiene los promotores de T7 y CMV, pF4A disponible en el mercado modificado para que contenga un sitio de unión a ribosomas de *E. coli*). Todos los variantes se seleccionaron en células de *E. coli*. Brevemente, los clones se sobreexpresaron en *E. coli* KRX, tras lo cual las células se lisaron y se midió la luminiscencia empleando colenteracina como un sustrato. También se seleccionaron la variante OgLuc N166R y la luciferasa de *Renilla*. Las variantes C1, C1+A4E y C1+C3+A4E fueron aproximadamente 4 log más brillantes que la variante OgLuc N166R y al menos tan brillantes como la luciferasa de *Renilla* (figura 4A-4D). La emisión de luz total (es decir, la luminiscencia) de estas variantes de combinaciones centrales en T = 0 se midió según se describió en el ejemplo 4 empleando tergitol al 0,5% "Destello" ("Flash") (figura 4A) y RLAB "Glo" (figura 4B).

Se muestra un alineamiento de las secuencias de las proteínas (figura 31) y de nucleótidos (figura 32) de la proteína nativa, WT, N166R, C1, C1+C2, C1+A4E, C1+C2+A4E y C1+C3+A4E.

Se introdujo otra sustitución en C1+A4E y C1+C3+A4E. De modo específico, el resto A54F en estas variantes se cambió a F54T. Estas variantes, C1+A4E+F54T y C1+C3+A4E+F54T, se compararon con las correspondientes C1+A4E y C1+C3+A4E de partida, así como con las luciferasas de *Renilla* y OgLuc WT empleando el método del

ejemplo 4. Tal como puede observarse en las figuras 18A, 18B y 19, las variantes con la sustitución F54T presentan una disminución del 50-75% con tergitol al 0,5% y un aumento en aproximadamente 2-5 veces en la luminiscencia con RLAB, comparado con WT (véase la medición en T = 0 en las figuras 18A y 19, respectivamente). La adición de la sustitución F54T mostró un aumento en la emisión de luz total con RLAB, pero con una disminución más rápida a lo largo del tiempo (figura 19). Con tergitol al 0,5%, la disminución a lo largo del tiempo es similar a C1+A4E, pero las RLU son menores comparado con C1+A4E (figura 18A-18B).

Se midió la luminiscencia de las variantes C1, C1+A4E, C1+C2, y C1+C2+A4E, comparado con la luciferasa de *Renilla*, OgLuc WT, la variante T2T y la variante A54F, empleando el método descrito en el ejemplo 4 (figura 20A y 20B). Las variantes C1+A4E y C1+C2+A4E presentan un aumento de 4 y 2 log, respectivamente, frente a WT empleando tergitol al 0,5% (figura 20A). Los variantes C1+A4E, C1+C2+A4E y C1+C3+A4E presentan un aumento de 3, 1,5 y 3 log, respectivamente, frente a WT empleando RLAB (figura 20B). Se empleó un tampón de tergitol al 0,25% en lugar de tergitol al 0,5% para determinar la estabilidad de la señal, que no depende del tergitol. La figura 21 muestra que las variantes C1, C1+A4E, C1+C2, y C1+C2+A4E presentan un aumento de 4, 4, 2 y 2 log, respectivamente, frente a WT empleando tergitol al 0,25%.

Las variantes C1, C1+A4E, C1+C2, y C1+C2+A4E, comparadas con la luciferasa de *Renilla*, OgLuc WT, las variantes T2T y OgLuc+A54F, también se evaluaron en células HEK 293. Brevemente, las células HEK 293, cultivadas en placas a 15.000 células/pocillo en una placa de 96 pocillos, se transfectaron transitoriamente empleando TransIT-LTI (mirus Bio) con ADN plasmídicos que codifican las diversas variantes y/o secuencias de control. Los mismos plásmidos también portan un gen para la expresión constitutiva de la luciferasa de luciérnaga para actuar como control de la transfección. Brevemente, las células se cultivaron, se lisaron y se trataron como se describió en el ejemplo 4. Las células se cotransfectaron con pGL4.13 para el control de la transfección con luciferasa de luciérnaga (empleando 0,04 ug/transfección o 10% del ADN transfectado total). La luminiscencia se midió según se describió en el ejemplo 4 empleando RLAB (figura 22) o tergitol al 0,25% (figura 23). Todos los datos de la luciferasa modificada después se normalizaron para la eficacia de la transfección empleando la luminiscencia de la luciferasa de luciérnaga (sustrato de luciferina) (figuras 22 y 23). Las variantes C1, C1+A4E, C1+C2 y C1+C2+A4E presentaban mayor luminiscencia comparado con OgLuc en tergitol al 0,5% (figura 22). Las variantes C1+A4E y C1+C2+A4E también presentaron mayor luminiscencia comparado con OgLuc en tergitol al 0,25% (figura 23).

Ejemplo 10:

Evaluación de combinaciones específicas de sustituciones en luciferasas modificadas - estabilidad de las proteínas

Para determinar si las sustituciones de aminoácidos en las diferentes variantes también tenían un efecto sobre la estabilidad de las proteínas, las diferentes variantes se seleccionaron a diferentes temperaturas y se midió el efecto sobre la estabilidad. Tal como se muestra en la figura 24, a temperatura ambiente (aproximadamente 22°C), la OgLuc de tipo salvaje mostró una semivida de la proteína de 1 hora, mientras que la variante C1 mostró una semivida de la proteína de 9,4 horas. Tal como se muestra en la figura 24, a 30°C, la variante OgLuc N166R mostró una semivida de la proteína de 21 minutos, mientras que la variante C1+A4E mostró en este caso una descomposición después de 6 horas. A 30°C, la semivida de la proteína para la luciferasa de *Renilla* fue de 7,9 horas. La clasificación de la estabilidad a 30°C es de OgLuc C1+A4E > luciferasa de *Renilla* > OgLuc N166R. Tal como se muestra en la figura 24, a 37°C, la semivida de la proteína de la variante OgLuc N166R fue de 2 minutos, mientras que no se observó descomposición en la variante C1+A4E. A 54°C, las semividas de las proteínas de las diferentes variantes fueron las siguientes: C1: 7 minutos, C1+A4E: 8 minutos, C1+C2+A4E: 128 minutos, y C1+C3+A4E: 24 minutos. Las semividas de OgLuc de tipo salvaje y de la variante OgLuc N166R no pudieron determinarse a 54°C debido a que eran demasiado inestables.

Ejemplo 11:

Evaluación de combinaciones específicas de sustituciones en luciferasas modificadas - estabilidad de la señal

Para determinar si las sustituciones de aminoácidos en las diferentes variantes también tenían efecto sobre la estabilidad de la señal, las diferentes variantes se seleccionaron para la estabilidad de la señal. La estabilidad de la señal se midió como se describió en el ejemplo 4 empleando RLAB. Se determinaron las siguientes semividas de las señales para las diferentes variantes: OgLuc de tipo salvaje: 1,8 minutos, luciferasa de *Renilla*: 0,8 minutos, C1: 1,7 minutos, C1+A4E: 1,7 minutos, C1+C2+A4E: 12,6 minutos, y C1+C3+A4E: 3,3 minutos (figura 25).

Ejemplo 12:

Evaluación de combinaciones específicas de sustituciones en luciferasas modificadas - color de la luminiscencia

Se determinó la longitud de onda óptima con la mayor luminiscencia empleando como sustrato coelenteracina (Promega Corp.) para las variantes OgLuc+N166R, C1+A4E, y C1+C2+A4E, comparado con la luciferasa de *Renilla*. Las muestras se prepararon como se describió en el ejemplo 4. Se determinó el pico espectral midiendo la

luminiscencia en incrementos de 5 nm de la longitud de onda empleando un luminómetro Varioskan y tergitol al 0,5%. Los datos se normalizaron al valor más alto de RLU en el espectro. Tal como se muestra en la figura 26, la luciferasa de *Renilla* presenta un pico espectral de 480 nm, mientras que OgLuc+N166R, C1+A4E y C1+C2+A4E presentan un pico espectral a 465 nm, que representa un desplazamiento desde la OgLuc nativa, que previamente ha sido indicado que se encuentra a 455 nm (Inouye, FEBS Letters, 481(1):19-25 (2000)).

Ejemplo 13:

Generación de una luciferasa modificada con mayor luminiscencia

Se generaron otras variantes mediante mutagénesis aleatoria como se describió en el ejemplo 3 a partir de la variante C1+A4E. Se midió la emisión de luz total como se describió en el ejemplo 4. Se listan las variantes ilustrativas de C1+A4E (es decir, las que son al menos 1,2 veces más brillantes que C1+A4E), aunque no se limitan a estas, en la figura 27A y 27B según el número identificativo de la muestra y la sustitución de aminoácidos. Las variantes de C1+A4E con sustituciones de aminoácidos en las posiciones 20, 54, 72, 77, 79, 89, 90 o 164 con relación a SEQ ID NO:1, muestran un aumento en al menos 1,9 veces de la luminiscencia frente al correspondiente variante C1+A4E de partida.

El clon 29H7, que contiene la variante C1+A4E+F54I, también fue ensayado para la estabilidad de las proteínas a 50°C empleando el método descrito en el ejemplo 5. El clon 29H7 tiene una semivida más larga que el correspondiente variante C1+A4E de partida (figura 30).

Se analizaron para el brillo diversas variantes de C1+A4E con una sustitución de un aminoácido en la posición 92, por ejemplo, se seleccionaron variantes que eran al menos 1,2 veces más brillantes que la variante C1+A4E. Las siguientes sustituciones produjeron un variante que era al menos 1,2 veces más brillante que C1+A4E: L92G; L92Q; L92S; y L92A, y presentan un aumento en 2,2, 2, 2,9 y 2,5 veces frente a C1+A4E, respectivamente (véase la figura 28).

Se generaron otras variantes mediante mutagénesis específica dirigida a sitio, descrita en el ejemplo 3, de la variante C1+A4E para que contuviesen combinaciones específicas de las sustituciones F54I, F68S, M75K e I90V. Tal como se muestra en la figura 29, que lista las variantes ("Muestra ID") y las sustituciones de aminoácidos que se encuentran en cada variante, estas combinaciones de sustituciones muestran un aumento significativo en la luminiscencia en al menos 17,5-19,3 veces frente al correspondiente variante C1+A4E de partida.

Todas las publicaciones, patentes y solicitudes de patente se incorporan en la presente como referencia. Aunque en la anterior memoria descriptiva esta invención se ha descrito con relación a ciertas realizaciones preferidas de la misma, y muchos detalles se han ofrecido como ilustración, será evidente para los expertos en la técnica que la invención es susceptible de otras realizaciones y que ciertos detalles de la presente pueden variar considerablemente sin apartarse de los principios básicos de la invención. Se generó otra variante de combinación específica de C1+A4E para que incluyese I90V y F54I ("IV"). Tal como se muestra en la figura 34A, IV presenta un aumento en aproximadamente 20 veces de la luminiscencia comparado con el correspondiente variante C1+A4E de partida, según se mide empleando el método del ejemplo 4. Tal como se muestra en la figura 34B, la proteína IV es más estable que la luciferasa de *Renilla* a 50°C, puesto que la semivida para IV era de 27,2 minutos, comparado con la luciferasa de *Renilla*, que era de 9,6 minutos empleando el método del ejemplo 5.

Listado de secuencias

<110> Promega Corporation

<120> LUCIFERASAS DE *OPLOPHORUS* SINTÉTICAS CON MAYOR EMISIÓN DE LUZ

<130> 206736

<150> US 61/174838

<151> 01-05-2009

<160> 32

<170> PatentIn versión 3.5

<210> 1

<211> 169

<212> PRT

<213> *Oplophorus gracilirostris*

<400> 1

ES 2 795 287 T3

Phe Thr Leu Ala Asp Phe Val Gly Asp Trp Gln Gln Thr Ala Gly Tyr
 1 5 10 15

Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu Phe Gln
 20 25 30

Ala Leu Gly Val Ser Val Thr Pro Ile Gln Lys Val Val Leu Ser Gly
 35 40 45

Glu Asn Gly Leu Lys Ala Asp Ile His Val Ile Ile Pro Tyr Glu Gly
 50 55 60

Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys Val Val
 65 70 75 80

Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr Gly Thr
 85 90 95

Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe Gly Arg
 100 105 110

Pro Tyr Pro Gly Ile Ala Val Phe Asp Gly Lys Gln Ile Thr Val Thr
 115 120 125

Gly Thr Leu Trp Asn Gly Asn Lys Ile Tyr Asp Glu Arg Leu Ile Asn
 130 135 140

Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val Thr Gly
 145 150 155 160

Trp Arg Leu Cys Glu Asn Ile Leu Ala

165

5 <210> 2
 <211> 513
 <212> ADN
 <213> Secuencia Artificial

10 <220>
 <223> Sintético

<400> 2

```

atggtgttta ccttggcaga ttctgttggg gactggcaac agacagctgg atacaaccaa      60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aagccctggg agtgtcagtc      120
acccaatcc agaaagttgt gctgtctggg gagaatgggt taaaagctga tattcatgtc      180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaaa      240
gttgtttacc cagtggatga tcatcatttc aagattatcc tccattatgg tacactcgtt      300
attgacggtg tgacaccaa catgattgac tactttggac gcccttacct tgggaattgct      360
gtgtttgacg gcaagcagat cacagttact ggaactctgt ggaacggcaa caagatctat      420
gatgagcgcc tgatcaaccc agatggttca ctctcttcc gcgttactat caatggagtc      480
accggatggc gcctttgcga gaacattctt gcc                                     513

```

<210> 3
 <211> 131
 <212> PRT
 <213> *Bos taurus*

5

<400> 3

```

Ser Asn Lys Phe Leu Gly Thr Trp Lys Leu Val Ser Ser Glu Asn Phe
1           5           10           15

Asp Glu Tyr Met Lys Ala Leu Gly Val Gly Leu Ala Thr Arg Lys Leu
          20           25           30

Gly Asn Leu Ala Lys Pro Arg Val Ile Ile Ser Lys Lys Gly Asp Ile
          35           40           45

Ile Thr Ile Arg Thr Glu Ser Pro Phe Lys Asn Thr Glu Ile Ser Phe
          50           55           60

Lys Leu Gly Gln Glu Phe Glu Glu Thr Thr Ala Asp Asn Arg Lys Thr
65           70           75           80

Lys Ser Thr Val Thr Leu Ala Arg Gly Ser Leu Asn Gln Val Gln Lys
          85           90           95

Trp Asn Gly Asn Glu Thr Thr Ile Lys Arg Lys Leu Val Asp Gly Lys
          100           105           110

Met Val Val Glu Cys Lys Met Lys Asp Val Val Cys Thr Arg Ile Tyr
          115           120           125

Glu Lys Val
          130

```

<210> 4
 <211> 134
 <212> PRT
 <213> *Rattus rattus*

5

<400> 4

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

10

<210> 5
 <211> 135
 <212> PRT
 <213> *Schistosoma mansoni*

15

<400> 5

Gly Ser Met Ser Ser Phe Leu Gly Lys Trp Lys Leu Ser Glu Ser His
 1 5 10 15

Asn Phe Asp Ala Val Met Ser Lys Leu Gly Val Ser Trp Ala Thr Arg
 20 25 30

Gln Ile Gly Asn Thr Val Thr Pro Thr Val Thr Phe Thr Met Asp Gly
 35 40 45

Asp Lys Met Thr Met Leu Thr Glu Ser Thr Phe Lys Asn Leu Ser Cys
 50 55 60

Thr Phe Lys Phe Gly Glu Glu Phe Asp Glu Lys Thr Ser Asp Gly Arg
 65 70 75 80

Asn Val Lys Ser Val Val Glu Lys Asn Ser Glu Ser Lys Leu Thr Gln
 85 90 95

Thr Gln Val Asp Pro Lys Asn Thr Thr Val Ile Val Arg Glu Val Asp
 100 105 110

Gly Asp Thr Met Lys Thr Thr Val Thr Val Gly Asp Val Thr Ala Ile
 115 120 125

Arg Asn Tyr Lys Arg Leu Ser
 130 135

<210> 6
 <211> 135
 <212> PRT
 <213> *Schistosoma mansoni*

5

<400> 6

Gly Ser Met Ser Ser Phe Leu Gly Lys Trp Lys Leu Ser Glu Ser His
 1 5 10 15

Asn Phe Asp Ala Val Met Ser Lys Leu Gly Val Ser Trp Ala Thr Arg
 20 25 30

Gln Ile Gly Asn Thr Val Thr Pro Thr Val Thr Phe Thr Met Asp Gly
 35 40 45

Asp Lys Met Thr Met Leu Thr Glu Ser Thr Phe Lys Asn Leu Ser Cys
 50 55 60

Thr Phe Lys Phe Gly Glu Glu Phe Asp Glu Lys Thr Ser Asp Gly Arg
 65 70 75 80

Asn Val Lys Ser Val Val Glu Lys Asn Ser Glu Ser Lys Leu Thr Gln
 85 90 95

ES 2 795 287 T3

Thr Gln Val Asp Pro Lys Asn Thr Thr Val Ile Val Arg Glu Val Asp
 100 105 110

Gly Asp Thr Met Lys Thr Thr Val Thr Val Gly Asp Val Thr Ala Ile
 115 120 125

Arg Asn Tyr Lys Arg Leu Ser
 130 135

<210> 7

- 5 <211> 165
- <212> PRT
- <213> Secuencia Artificial

<220>

- 10 <223> Sintético

<400> 7

Leu Ala Asp Phe Val Gly Asp Trp Gln Gln Thr Ala Gly Tyr Asn Gln
 1 5 10 15

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

Gly Val Ser Val Thr Pro Ile Gln Lys Val Val Leu Ser Gly Glu Asn
 35 40 45

Gly Leu Lys Ala Asp Ile His Val Ile Ile Pro Tyr Glu Gly Leu Ser
 50 55 60

Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys Val Val Tyr Pro
 65 70 75 80

Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr Gly Thr Leu Val
 85 90 95

Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe Gly Arg Pro Tyr
 100 105 110

Pro Gly Ile Ala Val Phe Asp Gly Lys Gln Ile Thr Val Thr Gly Thr
 115 120 125

Leu Trp Asn Gly Asn Lys Ile Tyr Asp Glu Arg Leu Ile Asn Pro Asp
 130 135 140

Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val Thr Gly Trp Arg
 145 150 155 160

Leu Cys Glu Asn Ile
165

<210> 8
<211> 131
<212> PRT
5 <213> *Bos taurus*

<400> 8

Ser Asn Lys Phe Leu Gly Thr Trp Lys Leu Val Ser Ser Glu Asn Phe
1 5 10 15

Asp Glu Tyr Met Lys Ala Leu Gly Val Gly Leu Ala Thr Arg Lys Leu
20 25 30

Gly Asn Leu Ala Lys Pro Arg Val Ile Ile Ser Lys Lys Gly Asp Ile
35 40 45

Ile Thr Ile Arg Thr Glu Ser Pro Phe Lys Asn Thr Glu Ile Ser Phe
50 55 60

Lys Leu Gly Gln Glu Phe Glu Glu Thr Thr Ala Asp Asn Arg Lys Thr
65 70 75 80

Lys Ser Thr Val Thr Leu Ala Arg Gly Ser Leu Asn Gln Val Gln Lys
85 90 95

Trp Asn Gly Asn Glu Thr Thr Ile Lys Arg Lys Leu Val Asp Gly Lys
100 105 110

Met Val Val Glu Cys Lys Met Lys Asp Val Val Cys Thr Arg Ile Tyr
115 120 125

Glu Lys Val
130

10 <210> 9
<211> 134
<212> PRT
15 <213> *Rattus rattus*

<400> 9

Pro Val Asp Phe Asn Gly Tyr Trp Lys Met Leu Ser Asn Glu Asn Phe
1 5 10 15

Glu Glu Tyr Leu Arg Ala Leu Asp Val Asn Val Ala Leu Arg Lys Ile
20 25 30

Ala Asn Leu Leu Lys Pro Asp Lys Glu Ile Val Gln Asp Gly Asp His
 35 40 45

Met Ile Ile Arg Thr Leu Ser Thr Phe Arg Asn Tyr Ile Met Asp Phe
 50 55 60

Gln Val Gly Lys Glu Phe Glu Glu Asp Leu Thr Gly Ile Asp Asp Arg
 65 70 75 80

Lys Cys Met Thr Thr Val Ser Trp Asp Gly Asp Lys Leu Gln Cys Val
 85 90 95

Gln Lys Gly Glu Lys Glu Gly Arg Gly Trp Thr Gln Trp Ile Glu Gly
 100 105 110

Asp Glu Leu His Leu Glu Met Arg Ala Glu Gly Val Thr Cys Lys Gln
 115 120 125

Val Phe Lys Lys Val His
 130

<210> 10

<211> 196

5 <212> PRT

<213> *Oplophorus gracilirostris*

<400> 10

Met Ala Tyr Ser Thr Leu Phe Ile Ile Ala Leu Thr Ala Val Val Thr
 1 5 10 15

Gln Ala Ser Ser Thr Gln Lys Ser Asn Leu Thr Phe Thr Leu Ala Asp
 20 25 30

Phe Val Gly Asp Trp Gln Gln Thr Ala Gly Tyr Asn Gln Asp Gln Val
 35 40 45

Leu Glu Gln Gly Gly Leu Ser Ser Leu Phe Gln Ala Leu Gly Val Ser
 50 55 60

Val Thr Pro Ile Gln Lys Val Val Leu Ser Gly Glu Asn Gly Leu Lys
 65 70 75 80

Ala Asp Ile His Val Ile Ile Pro Tyr Glu Gly Leu Ser Gly Phe Gln
 85 90 95

Met Gly Leu Ile Glu Met Ile Phe Lys Val Val Tyr Pro Val Asp Asp
 100 105 110

ES 2 795 287 T3

His His Phe Lys Ile Ile Leu His Tyr Gly Thr Leu Val Ile Asp Gly
 115 120 125

Val Thr Pro Asn Met Ile Asp Tyr Phe Gly Arg Pro Tyr Pro Gly Ile
 130 135 140

Ala Val Phe Asp Gly Lys Gln Ile Thr Val Thr Gly Thr Leu Trp Asn
 145 150 155 160

Gly Asn Lys Ile Tyr Asp Glu Arg Leu Ile Asn Pro Asp Gly Ser Leu
 165 170 175

Leu Phe Arg Val Thr Ile Asn Gly Val Thr Gly Trp Arg Leu Cys Glu
 180 185 190

Asn Ile Leu Ala
 195

<210> 11

<211> 131

5 <212> PRT

<213> *Bos taurus*

<400> 11

Ser Asn Lys Phe Leu Gly Thr Trp Lys Leu Val Ser Ser Glu Asn Phe
 1 5 10 15

Asp Glu Tyr Met Lys Ala Leu Gly Val Gly Leu Ala Thr Arg Lys Leu
 20 25 30

Gly Asn Leu Ala Lys Pro Arg Val Ile Ile Ser Lys Lys Gly Asp Ile
 35 40 45

Ile Thr Ile Arg Thr Glu Ser Pro Phe Lys Asn Thr Glu Ile Ser Phe
 50 55 60

Lys Leu Gly Gln Glu Phe Glu Glu Thr Thr Ala Asp Asn Arg Lys Thr
 65 70 75 80

Lys Ser Thr Val Thr Leu Ala Arg Gly Ser Leu Asn Gln Val Gln Lys
 85 90 95

Trp Asn Gly Asn Glu Thr Thr Ile Lys Arg Lys Leu Val Asp Gly Lys
 100 105 110

Met Val Val Glu Cys Lys Met Lys Asp Val Val Cys Thr Arg Ile Tyr
 115 120 125

Glu Lys Val
130

5 <210> 12
<211> 588
<212> ADN
<213> *Oplophorus gracilirostris*

<400> 12

atggcgtact	ccactctggt	cataattgca	ttgaccgccg	ttgtcactca	agcttctca	60
actcaaaaat	ctaacctaac	ttttacggtg	gcagatttcg	ttggagactg	gcaacagaca	120
gctggataca	accaagatca	agtgttagaa	caaggaggat	tgtctagtct	gttccaagcc	180
ctgggagtgt	cagtcacgcc	catacagaaa	gttgtactgt	ctggggagaa	tgggttaaaa	240
gctgatattc	atgtcataat	accttacgag	ggactcagtg	gttttcaaat	gggtctaatt	300
gaaatgatct	tcaaagttgt	ttaccccggtg	gatgatcatc	atttcaagat	tattctccat	360
tatggtacac	tcggtattga	cggtgtaaca	cccaacatga	ttgactactt	tggaagacct	420
taccctggaa	ttgctgtatt	tgacggcaag	cagatcacag	ttactggaac	tctgtggaac	480
ggcaacaaga	tctatgatga	gaggctaatc	aaccctgatg	gttcactcct	cttcagagtt	540
actatcaatg	gagtcacggg	atggaggcct	tgcgagaaca	ttcttgcc		588

10

<210> 13

15 <211> 171
<212> PRT
<213> Secuencia Artificial

<220>

20 <223> Sintético

<400> 13

ES 2 795 287 T3

Met Val Phe Thr Leu Ala Asp Phe Val Gly Asp Trp Gln Gln Thr Ala
 1 5 10 15

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
 20 25 30

Phe Gln Ala Leu Gly Val Ser Val Thr Pro Ile Gln Lys Val Val Leu
 35 40 45

Ser Gly Glu Asn Gly Leu Lys Ala Asp Ile His Val Ile Ile Pro Tyr
 50 55 60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
 65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
 85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Pro Gly Ile Ala Val Phe Asp Gly Lys Gln Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Asn Lys Ile Tyr Asp Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Asn Ile Leu Ala
 165 170

5 <210> 14
 <211> 513
 <212> ADN
 <213> Secuencia Artificial

10 <220>
 <223> Sintético

<400> 14

ES 2 795 287 T3

```

atggtgttta ccttggcaga ttctggtgga gactggcaac agacagctgg atacaaccaa      60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aagccctggg agtgtcagtc      120
accocaatcc agaaagttgt gctgtctggg gagaatgggt taaaagctga tattcatgct      180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaaa      240
gttgtttacc cagtggatga tcatcatttc aagattatcc tccattatgg tacactcgtt      300
attgacggtg tgacaccaaa catgattgac tactttggac gcccttacct tggattgct      360
gtgtttgacg gcaagcagat cacagttact ggaactctgt ggaacggcaa caagatctat      420
gatgagcgcc tgatcaacct agatggttca ctctcttcc gcgttactat caatggagtc      480
accggatggc gcctttgcga gcgtattctt gcc                                  513

```

5 <210> 15
 <211> 171
 <212> PRT
 <213> Secuencia Artificial

10 <220>
 <223> Sintético
 <400> 15

```

Met Val Phe Thr Leu Ala Asp Phe Val Gly Asp Trp Gln Gln Thr Ala
1           5           10           15

```

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
 20 25 30

Phe Gln Ala Leu Gly Val Ser Val Thr Pro Ile Gln Lys Val Val Leu
 35 40 45

Ser Gly Glu Asn Gly Leu Lys Ala Asp Ile His Val Ile Ile Pro Tyr
 50 55 60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
 65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
 85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Pro Gly Ile Ala Val Phe Asp Gly Lys Gln Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Asn Lys Ile Tyr Asp Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Arg Ile Leu Ala
 165 170

<210> 16
 <211> 513
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintético

<400> 16

5

10

```

atggtgttta cattggagga ttctggtgga gactggcggc agacagctgg atacaaccaa      60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aaaagctggg agtgtcagtc      120
acccaatcc agaaaattgt gctgtctggg gagaatgggt taaaatttga tattcatgtc      180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaaa      240
gttgtttacc cagtggatga tcatcatttc aagattattc tccattatgg tacactcgtt      300
attgacggtg tgacaccaa catgattgac tactttggac gcccttacga ggaattgct      360
gtgtttgacg gcaagaagat cacagttact ggaactctgt ggaacggcaa caagatcatt      420

gatgagcggc tgatcaacce agatggttca ctctcttcc gcgttactat caatggagtc      480
accggatggc gcctttgoga gcgtattctt gcc                                  513

```

<210> 17
 <211> 133
 <212> PRT
 <213> *Locusta migratoria*

<400> 17

```

Val Lys Glu Phe Ala Gly Ile Lys Tyr Lys Leu Asp Ser Gln Thr Asn
1           5           10           15

Phe Glu Glu Tyr Met Lys Ala Ile Gly Val Gly Ala Ile Glu Arg Lys
          20           25           30

Ala Gly Leu Ala Leu Ser Pro Val Ile Glu Leu Glu Val Leu Asp Gly
          35           40           45

Asp Lys Phe Lys Leu Thr Ser Lys Thr Ala Ile Lys Asn Thr Glu Phe
          50           55           60

Thr Phe Lys Leu Gly Glu Glu Phe Asp Glu Asp Thr Leu Asp Gly Arg
          65           70           75           80

Lys Val Lys Ser Ile Ile Thr Gln Asp Gly Pro Asn Lys Leu Val His
          85           90           95

Glu Gln Lys Gly Asp His Pro Thr Ile Ile Ile Arg Glu Phe Ser Lys
          100          105          110

Glu Gln Cys Val Ile Thr Ile Lys Leu Gly Asp Leu Val Ala Thr Arg
          115          120          125

Ile Tyr Lys Ala Gln
          130

```

ES 2 795 287 T3

<210> 18
 <211> 125
 <212> PRT
 <213> *Bufo arenarum*

5

<400> 18

Ala Phe Asn Gly Thr Trp Asn Val Tyr Ala Gln Glu Asn Tyr Glu Asn
 1 5 10 15

Phe Leu Arg Thr Val Gly Leu Pro Glu Asp Ile Ile Lys Val Ala Lys
 20 25 30

Asp Val Asn Pro Val Ile Glu Ile Glu Gln Asn Gly Asn Glu Phe Val
 35 40 45

Val Thr Ser Lys Thr Pro Lys Gln Thr His Ser Asn Ser Phe Thr Val
 50 55 60

Gly Lys Glu Ser Glu Ile Thr Ser Met Asp Gly Lys Lys Ile Lys Val
 65 70 75 80

Thr Val Gln Leu Glu Gly Gly Lys Leu Ile Cys Lys Ser Asp Lys Phe
 85 90 95

Ser His Ile Gln Glu Val Asn Gly Asp Glu Met Val Glu Lys Ile Thr
 100 105 110

Ile Gly Ser Ser Thr Leu Thr Arg Lys Ser Lys Arg Val
 115 120 125

10

<210> 19
 <211> 134
 <212> PRT
 <213> *Echinococcus granulosus*

15

<220>
 <221> característica_misc
 <222> (1)..(1)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

20

<220>
 <221> característica_misc
 <222> (64)..(64)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

25

<400> 19

ES 2 795 287 T3

Xaa Met Glu Ala Phe Leu Gly Thr Trp Lys Met Glu Lys Ser Glu Gly
1 5 10 15

Phe Asp Lys Ile Met Glu Arg Leu Gly Val Asp Phe Val Thr Arg Lys
20 25 30

Met Gly Asn Leu Val Lys Pro Asn Leu Ile Val Thr Asp Leu Gly Gly
35 40 45

Gly Lys Tyr Lys Met Arg Ser Glu Ser Thr Phe Lys Thr Thr Glu Xaa
50 55 60

Ser Phe Lys Leu Gly Glu Lys Phe Lys Glu Val Thr Pro Asp Ser Arg
65 70 75 80

Glu Val Ala Ser Leu Ile Thr Val Glu Asn Gly Val Met Lys His Glu
85 90 95

Gln Asp Asp Lys Thr Lys Val Thr Tyr Ile Glu Arg Val Val Glu Gly
100 105 110

Asn Glu Leu Lys Ala Thr Val Lys Val Asp Glu Val Val Cys Val Arg
115 120 125

Thr Tyr Ser Lys Val Ala
130

- <210> 20
- 5 <211> 190
- <212> PRT
- <213> *Mycobacterium tuberculosis*
- <400> 20

Ala Ala Val Glu Arg Ala Lys Ala Thr Ala Ala Arg Asn Ile Pro Ala
 1 5 10 15

Phe Asp Asp Leu Pro Val Pro Ala Asp Thr Ala Asn Leu Arg Glu Gly
 20 25 30

Ala Asp Leu Asn Asn Ala Leu Leu Ala Leu Leu Pro Leu Val Gly Val
 35 40 45

Trp Arg Gly Glu Gly Glu Gly Arg Gly Pro Asp Gly Asp Tyr Arg Phe
 50 55 60

Gly Gln Gln Ile Val Val Ser His Asp Gly Gly Asp Tyr Leu Asn Trp
 65 70 75 80

Glu Ser Arg Ser Trp Arg Leu Thr Ala Thr Gly Asp Tyr Gln Glu Pro
 85 90 95

Gly Leu Arg Glu Ala Gly Phe Trp Arg Phe Val Ala Ile Glu Leu Leu
 100 105 110

Leu Ala His Ser Ala Gly Tyr Val Glu Leu Phe Tyr Gly Arg Pro Arg
 115 120 125

Thr Gln Ser Ser Trp Glu Leu Val Thr Asp Ala Leu Ala Arg Ser Arg
 130 135 140

Ser Gly Val Leu Val Gly Gly Ala Lys Arg Leu Tyr Gly Ile Val Glu
 145 150 155 160

Gly Gly Asp Leu Ala Tyr Val Glu Glu Arg Val Asp Ala Asp Gly Gly
 165 170 175

Leu Val Pro His Leu Ser Ala Arg Leu Ser Arg Phe Val Gly
 180 185 190

5 <210> 21
 <211> 351
 <212> PRT
 <213> *Lingulodinium polyedrum*

10 <220>
 <221> característica_misc
 <222> (31)..(31)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

15 <220>
 <221> característica_misc
 <222> (45)..(45)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

<220>

<221> característica_misc
 <222> (90)..(90)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

5 <220>
 <221> característica_misc
 <222> (123)..(123)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

10 <220>
 <221> característica_misc
 <222> (203)..(203)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

15 <220>
 <221> característica_misc
 <222> (265)..(265)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

20 <220>
 <221> característica_misc
 <222> (286)..(286)
 <223> Xaa puede ser cualquier aminoácido que se produce de forma natural

25 <400> 21

Glu Lys Gly Phe Glu Ala Gly Asp Asn Lys Leu Gly Gly Ala Leu Asn
1 5 10 15

Ala Lys His Val Glu Lys Tyr Gly Asp Asn Phe Lys Asn Gly Xaa His
20 25 30

Lys Pro Glu Phe His Glu Asp Gly Leu His Lys Pro Xaa Glu Val Gly
35 40 45

Gly Lys Lys Phe Glu Ser Gly Phe His Tyr Leu Leu Glu Cys His Glu
 50 55 60

Leu Gly Gly Lys Asn Ala Ser Gly Gly Tyr Gly Gly Pro Leu Cys Glu
 65 70 75 80

Asp Pro Tyr Gly Ser Glu Val Gln Ala Xaa Thr Glu Lys Leu Leu Lys
 85 90 95

Glu Ala Asp Ser Asp Arg Thr Leu Cys Phe Asn Asn Phe Gln Asp Pro
 100 105 110

Cys Pro Gln Leu Thr Lys Glu Gln Val Ala Xaa Cys Lys Gly Phe Asp
 115 120 125

Tyr Gly Asp Lys Thr Leu Lys Leu Pro Cys Gly Pro Leu Pro Trp Pro
 130 135 140

Ala Gly Leu Pro Glu Pro Gly Tyr Val Pro Lys Thr Asn Pro Leu His
 145 150 155 160

Gly Arg Trp Ile Thr Val Ser Gly Gly Gln Ala Ala Phe Ile Lys Glu
 165 170 175

Ala Ile Lys Ser Gly Met Leu Gly Ala Ala Glu Ala Asn Lys Ile Val
 180 185 190

Ala Asp Thr Asp His His Gln Thr Gly Gly Xaa Tyr Leu Arg Ile Asn
 195 200 205

Gln Phe Gly Asp Val Cys Thr Val Asp Ala Ser Val Ala Lys Phe Ala
 210 215 220

Arg Ala Lys Arg Thr Trp Lys Ser Gly His Tyr Phe Tyr Glu Pro Leu
 225 230 235 240

Val Ser Gly Gly Asn Leu Leu Gly Val Trp Val Leu Pro Glu Glu Tyr
 245 250 255

Arg Lys Ile Gly Phe Phe Trp Glu Xaa Glu Ser Gly Arg Cys Phe Arg
 260 265 270

Ile Glu Arg Arg Ala Phe Pro Val Gly Pro Tyr Thr Phe Xaa Arg Gln
 275 280 285

Ala Thr Glu Val Gly Gly Lys Ile Ser Phe Val Phe Tyr Val Lys Val
 290 295 300

ES 2 795 287 T3

Ser Asn Asp Pro Glu Ser Asp Pro Ile Pro Leu Gln Ser Arg Asp Tyr
 305 310 315 320

Thr Ala Leu Ala Gly Arg Asp Asn Ala Pro Thr Asn Leu Gly Lys Pro
 325 330 335

Tyr Pro Thr Leu Ala Lys Asp Leu Asp Tyr Pro Lys Lys Arg Asp
 340 345 350

5 <210> 22
 <211> 131
 <212> PRT
 <213> *Homo sapiens*

10 <400> 22

Val Asp Ala Phe Leu Gly Thr Trp Lys Leu Val Asp Ser Lys Asn Phe
 1 5 10 15

Asp Asp Tyr Met Lys Ser Leu Gly Val Gly Phe Ala Thr Arg Gln Val
 20 25 30

Ala Ser Met Thr Lys Pro Thr Thr Ile Ile Glu Lys Asn Gly Asp Ile
 35 40 45

Leu Thr Leu Lys Thr His Ser Thr Phe Lys Asn Thr Glu Ile Ser Phe
 50 55 60

Lys Leu Gly Val Glu Phe Asp Glu Thr Thr Ala Asp Asp Arg Lys Val
 65 70 75 80

Lys Ser Ile Val Thr Leu Asp Gly Gly Lys Leu Val His Leu Gln Lys
 85 90 95

Trp Asp Gly Gln Glu Thr Thr Leu Val Arg Glu Leu Ile Asp Gly Lys
 100 105 110

Leu Ile Leu Thr Leu Thr His Gly Thr Ala Val Cys Thr Arg Thr Tyr
 115 120 125

Glu Lys Glu
 130

15 <210> 23
 <211> 171
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Sintético

<400> 23

Met Val Phe Thr Leu Glu Asp Phe Val Gly Asp Trp Arg Gln Thr Ala
 1 5 10 15

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
 20 25 30

Phe Gln Lys Leu Gly Val Ser Val Thr Pro Ile Gln Lys Ile Val Leu
 35 40 45

Ser Gly Glu Asn Gly Leu Lys Phe Asp Ile His Val Ile Ile Pro Tyr
 50 55 60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
 65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
 85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Glu Gly Ile Ala Val Phe Asp Gly Lys Lys Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Asn Lys Ile Ile Asp Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Arg Ile Leu Ala
 165 170

5

<210> 24

<211> 513

<212> ADN

10 <213> Secuencia Artificial

<220>

<223> Sintético

15 <400> 24

ES 2 795 287 T3

atggtgttta cattggcaga tttcgttga gactggcggc agacagctgg atacaaccaa	60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aaaagctggg agtgtcagtc	120
acccaatcc agaaaattgt gctgtotggg gagaatgggt taaaatttga tattcatgtc	180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaa	240
gttgtttacc cagtggatga tcatcattc aagattatc tccattatgg tacactogtt	300
attgacggtg tgacaccaa catgattgac tactttggac gcccttacga ggaattgct	360
gtgtttgacg gcaagaagat cacagttact ggaactctgt ggaacggcaa caagatcatt	420
gatgagcgcc tgatcaacc agatggttca ctctcttcc gcgttactat caatggagtc	480
accggatggc gcctttgca gcgtattctt gcc	513

5 <210> 25
 <211> 171
 <212> PRT
 <213> Secuencia Artificial

10 <220>
 <223> Sintético

<400> 25

ES 2 795 287 T3

Met Val Phe Thr Leu Ala Asp Phe Val Gly Asp Trp Arg Gln Thr Ala
 1 5 10 15

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
 20 25 30

Phe Gln Lys Leu Gly Val Ser Val Thr Pro Ile Gln Lys Ile Val Leu
 35 40 45

Ser Gly Glu Asn Gly Leu Lys Phe Asp Ile His Val Ile Ile Pro Tyr
 50 55 60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
 65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
 85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Pro Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Glu Gly Ile Ala Val Phe Asp Gly Lys Lys Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Asn Lys Ile Ile Asp Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Arg Ile Leu Ala

165

170

<210> 26

5 <211> 513
 <212> ADN
 <213> Secuencia Artificial

<220>
 10 <223> Sintético

<400> 26

ES 2 795 287 T3

```

atggtgttta cattggcaga tttcgttggg gactggcggc agacagctgg atacaaccaa      60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aaaagctggg agtgtcagtc      120
accccaatcc agaaaattga gctgtctggg gagaatgggt taaaatttga tattcatgtc      180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaa      240
gttgtttacc cagtggatga tcatcatttc aagattatc tccattatgg tacactogtt      300
attgacgggtg tgacacttaa catgattgac tactttggac gcccttacga ggaattgct      360
gtgtttgacg gcaagaagat cacagttact ggaactctgt ggaacggcaa gaagatcatt      420
gaggagcgcc tgatcaacc agatggttca ctctcttcc gcgttactat caatggagtc      480
accggatggc gcctttgcga gcgtgttctt gcc                                     513

```

<210> 27
 <211> 171
 <212> PRT
 <213> Secuencia Artificial

5

<220>
 <223> Sintético

10

<400> 27

```

Met Val Phe Thr Leu Ala Asp Phe Val Gly Asp Trp Arg Gln Thr Ala
 1           5           10           15

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
          20           25           30

Phe Gln Lys Leu Gly Val Ser Val Thr Pro Ile Gln Lys Ile Glu Leu
          35           40           45

Ser Gly Glu Asn Gly Leu Lys Phe Asp Ile His Val Ile Ile Pro Tyr
 50           55           60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
65           70           75           80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
          85           90           95

```

ES 2 795 287 T3

Gly Thr Leu Val Ile Asp Gly Val Thr Leu Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Glu Gly Ile Ala Val Phe Asp Gly Lys Lys Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Lys Lys Ile Ile Glu Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Arg Val Leu Ala
 165 170

- 5 <210> 28
- <211> 513
- <212> ADN
- <213> Secuencia Artificial

- 10 <220>
- <223> Sintético

<400> 28

```

atggtgttta cattggagga ttcggtgga gactggcggc agacagctgg atacaaccaa      60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aaaagctggg agtgtcagtc      120
acccaatcc agaaaattga gctgtctggg gagaatgggt taaaatttga tattcatgtc      180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaaa      240
gttgtttacc cagtggatga tcatcatttc aagattattc tccattatgg tacactcgtt      300
attgacggtg tgacacttaa catgattgac tactttggac gcccttacga ggaattgct      360
gtgtttgacg gcaagaagat cacagttact ggaactctgt ggaacggcaa gaagatcatt      420
gaggagcggc tgatcaacce agatggttca ctctcttcc gcgttactat caatggagtc      480
accgatggc gcctttgcga gcgtgttctt gcc                                          513
    
```

- 15 <210> 29
- <211> 171
- <212> PRT
- <213> Secuencia Artificial

- 20 <220>
- <223> Sintético

<400> 29

Met Val Phe Thr Leu Glu Asp Phe Val Gly Asp Trp Arg Gln Thr Ala
 1 5 10 15

ES 2 795 287 T3

Gly Tyr Asn Gln Asp Gln Val Leu Glu Gln Gly Gly Leu Ser Ser Leu
 20 25 30

Phe Gln Lys Leu Gly Val Ser Val Thr Pro Ile Gln Lys Ile Glu Leu
 35 40 45

Ser Gly Glu Asn Gly Leu Lys Phe Asp Ile His Val Ile Ile Pro Tyr
 50 55 60

Glu Gly Leu Ser Gly Phe Gln Met Gly Leu Ile Glu Met Ile Phe Lys
 65 70 75 80

Val Val Tyr Pro Val Asp Asp His His Phe Lys Ile Ile Leu His Tyr
 85 90 95

Gly Thr Leu Val Ile Asp Gly Val Thr Leu Asn Met Ile Asp Tyr Phe
 100 105 110

Gly Arg Pro Tyr Glu Gly Ile Ala Val Phe Asp Gly Lys Lys Ile Thr
 115 120 125

Val Thr Gly Thr Leu Trp Asn Gly Lys Lys Ile Ile Glu Glu Arg Leu
 130 135 140

Ile Asn Pro Asp Gly Ser Leu Leu Phe Arg Val Thr Ile Asn Gly Val
 145 150 155 160

Thr Gly Trp Arg Leu Cys Glu Arg Val Leu Ala
 165 170

<210> 30
 <211> 513
 5 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Sintético

10 <400> 30

ES 2 795 287 T3

atggtgttta cattggagga tttegttga gactggcggc agacagctgg atacaaccaa	60
gatcaagtgt tagaacaagg aggattgcct agtctgttcc aaaagatggg agtgtcagtc	120
acccaatcc agaaaattgt gctgtctggg gagaatgtgt taaaatttga tattcatgtc	180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaaa	240
gttgtttacc cagtggatga tcatcatttc aagattattc tccattatgg tacactcgtt	300
gttgacggtg tgacaccaaa catgattgac tactttggac gcccttacga ggaattgct	360
gtgtttgacg gcaagaagat cacagttact ggaactctgt ggaacggcaa caagatcatt	420
gatgagcggc tgctcaacco agatggttca ctctcttcc gcgttactat caatggagtc	480
acgggatggc gcctttgoga gcgtattctt gcc	513

5 <210> 31
 <211> 171
 <212> PRT
 <213> Secuencia Artificial

10 <220>
 <223> Sintético

<400> 31

ES 2 795 287 T3

atggtgttta cattggcaga tttcgttga gactggcaac agacagctgg atacaaccaa	60
gatcaagtgt tagaacaagg aggattgtct agtctgttcc aagccctggg agtgtcagtc	120
acccaatcc agaaagtgt gctgtctggg gagaatgggt taaaagctga tattcatgtc	180
atcatccctt acgagggact cagtggtttt caaatgggtc tgattgaaat gatcttcaa	240
gttgtttacc cagtggatga tcatcattc aagattatc tccattatgg tacactcgtt	300
attgacggtg tgacaccaa catgattgac tactttggac gcccttacc tggaattgct	360
gtgtttgacg gcaagcagat cacagttact ggaactctgt ggaacggcaa caagatctat	420
gatgagcgc tgatcaacc agatggttca ctctcttcc gcgttactat caatggagtc	480
accggatggc gcctttgcga gaacattctt gcc	513

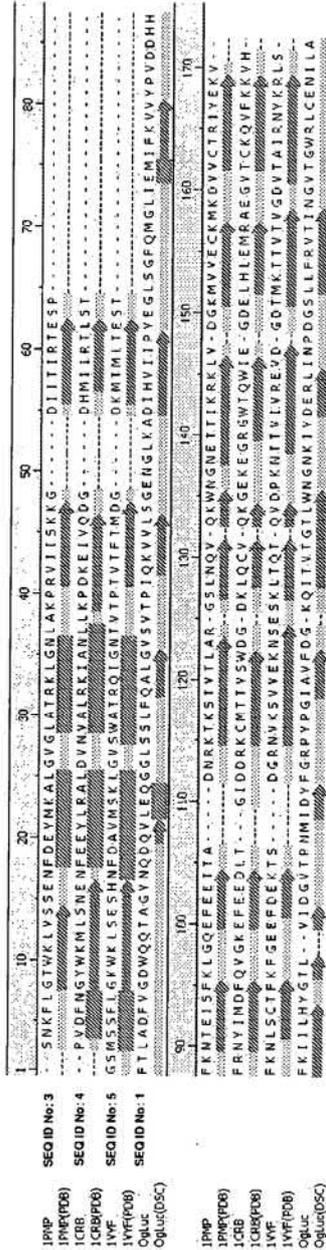
REIVINDICACIONES

1. Método que comprende:
 - 5 (a) expresar un polipéptido de luciferasa en una célula, en el que el polipéptido de luciferasa es una luciferasa modificada con al menos 60% de identidad de secuencia de aminoácidos con una luciferasa de *Oplophorus* de tipo salvaje, y que comprende al menos una sustitución de un aminoácido en una posición seleccionada de las posiciones 2, 4, 11, 20, 23, 28, 33, 34, 44, 45, 51, 54, 68, 72, 75, 76, 77, 89, 90, 92, 99, 104, 115, 124, 135, 138, 139, 143, 144, 164, 166, 167 o 169 correspondientes a SEQ ID NO:1, en el
 - 10 que la luciferasa modificada tiene al menos una de luminiscencia mayor, estabilidad de la señal mayor y estabilidad de la proteína mayor con relación a la luciferasa de *Oplophorus* de tipo salvaje;
 - (b) exponer dicho polipéptido de luciferasa a un sustrato; y
 - 15 (c) detectar luminiscencia.
 2. El método de la reivindicación 1, en el que el sustrato es coelenteracina o un derivado de coelenteracina.
 3. El método de la reivindicación 2, en el que la coelenteracina se selecciona de coelenteracina, coelenteracina n, coelenteracina h, coelenteracina c, coelenteracina cp, coelenteracina e, coelenteracina f, coelenteracina fcp, coelenteracina i, coelenteracina icp o coelenteracina 2-metilo.
 4. El método de la reivindicación 1, en el que el polipéptido de luciferasa se expresa como una fusión con un polipéptido de interés.
 - 25 5. El método de la reivindicación 1, que comprende además una etapa antes de la etapa (a) de introducir en la célula un vector que comprende una secuencia de ácido nucleico que codifica el polipéptido de luciferasa.
 6. El método de la reivindicación 1, en el que la luciferasa modificada tiene codones optimizados para la expresión en la célula hospedante.
 - 30 7. El método de la reivindicación 5, en el que la secuencia de ácido nucleico tiene codones optimizados para la expresión de la luciferasa modificada en la célula hospedante.
 8. El método de la reivindicación 6, en el que la luciferasa modificada tiene codones optimizados para la expresión en una célula hospedante de mamífero.
 9. El método de la reivindicación 7, en el que la secuencia de ácido nucleico tiene codones optimizados para la expresión de la luciferasa modificada en una célula hospedante de mamífero.

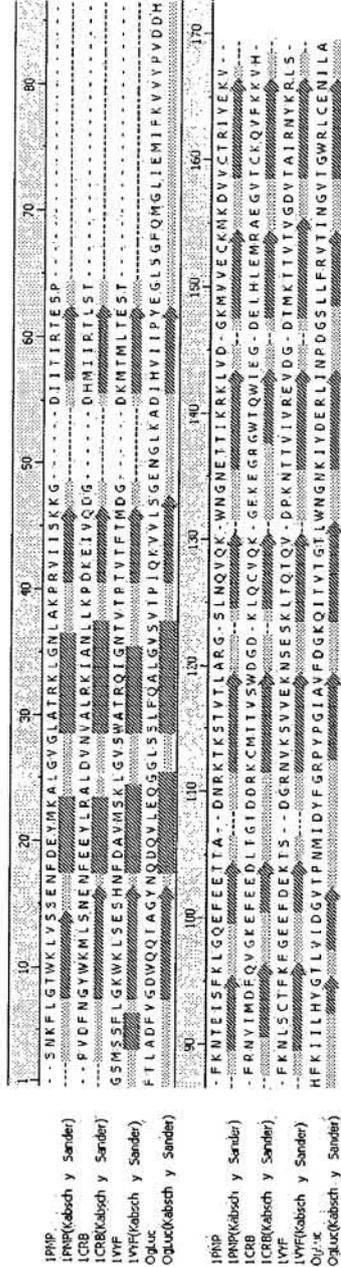
40

- extensión aproximada de los elementos de hélice de la estructura secundaria
- extensión aproximada de los elementos de lámina-beta de la estructura secundaria
- estructura secundaria distinta de hélice o lámina-beta

Alineamiento de la estructura secundaria n.º 1



Alineamiento de la estructura secundaria n.º 2



En los alineamientos pueden observarse los motivos de secuencia centrados en los restos conservados de la firma estructural de la calcina. El motivo MOTIF1 N-terminal, más altamente conservado, incluye el resto de Ogluc Trp10, y el motivo MOTIF2 C-terminal, menos altamente conservado, incluye el resto de Ogluc N166. Para el alineamiento n.º 2, el porcentaje de identidades de secuencia de proteína apareadas aproximado es: Ogluc-IPMP 14%, Ogluc-ICRB 9%, y Ogluc-IYVF 15%.

FIG. 1

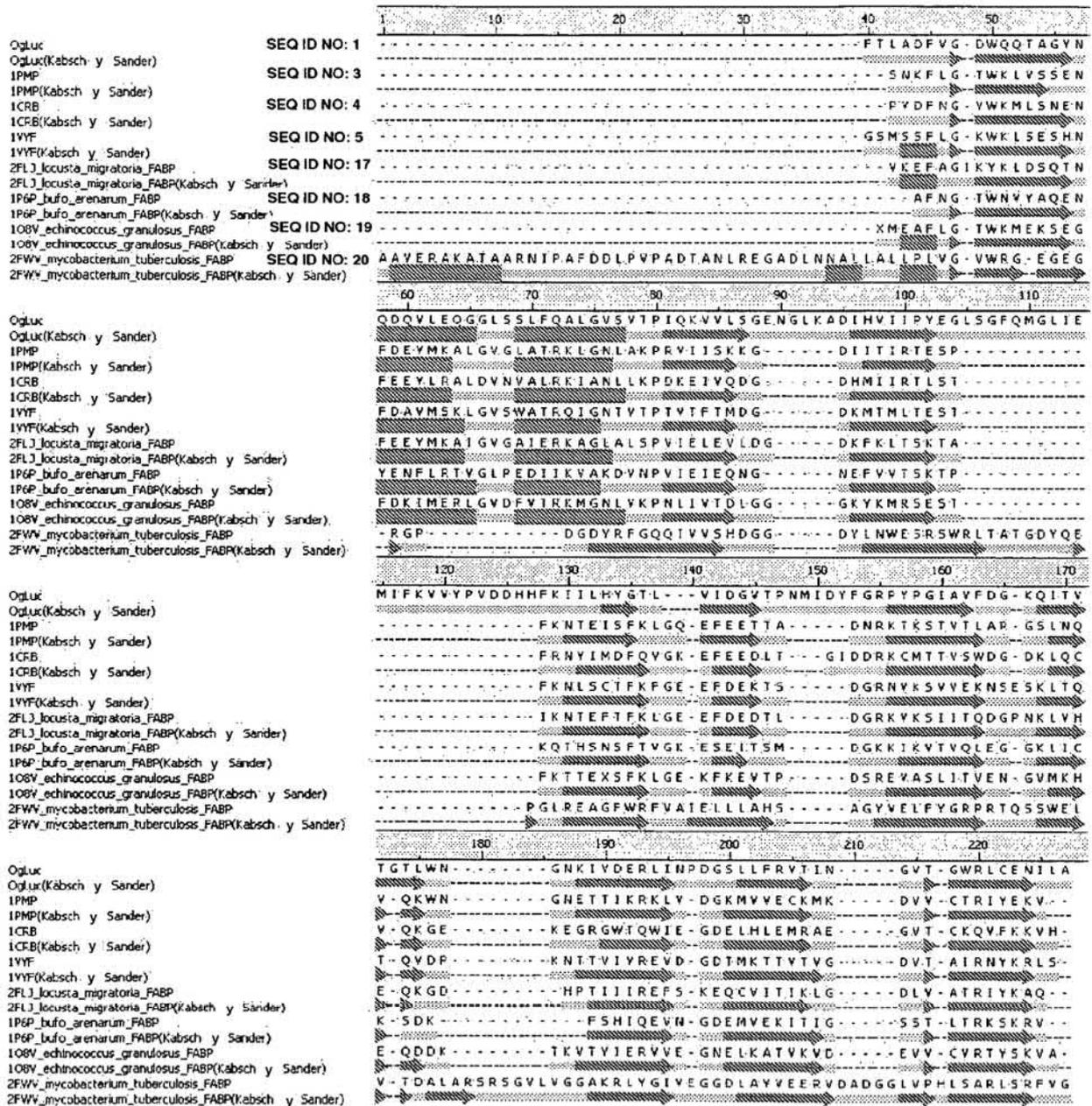


FIG. 3

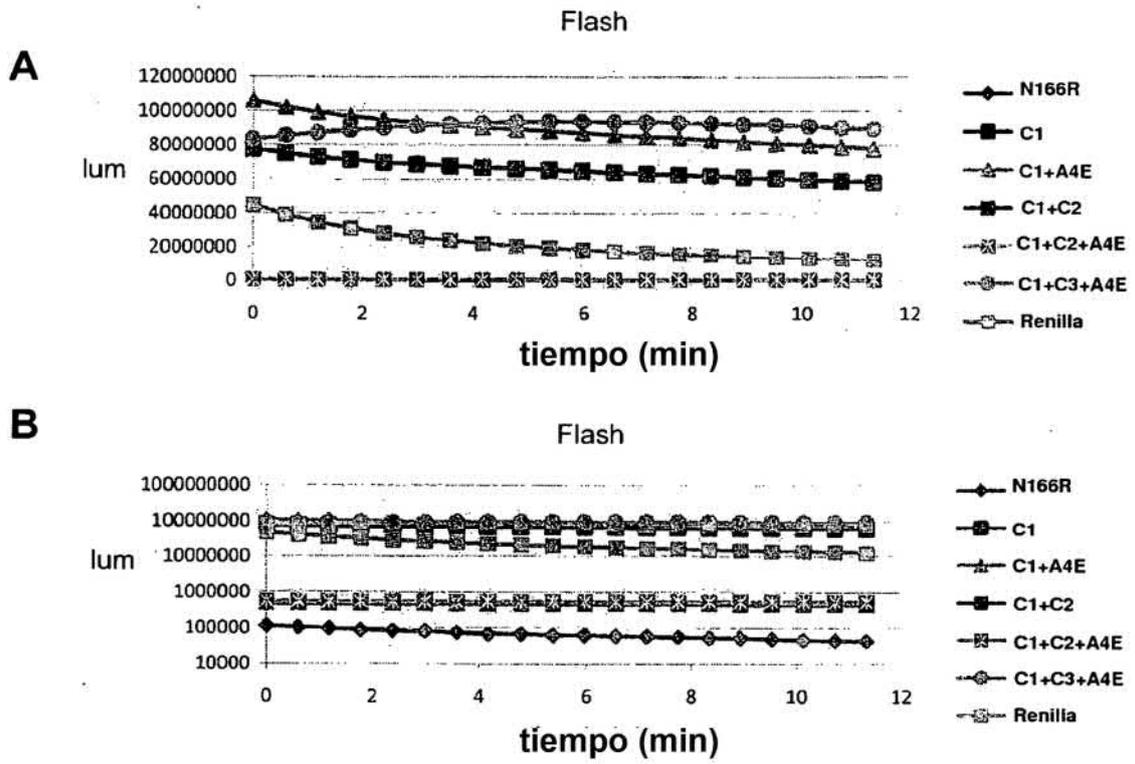
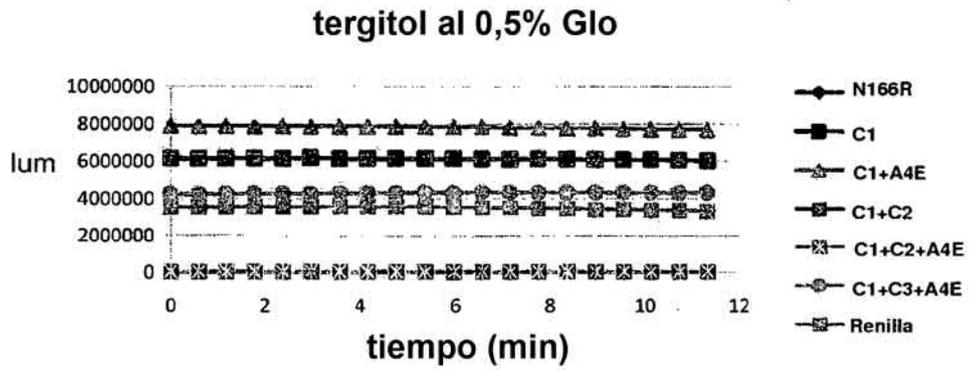


FIG. 4

C



D

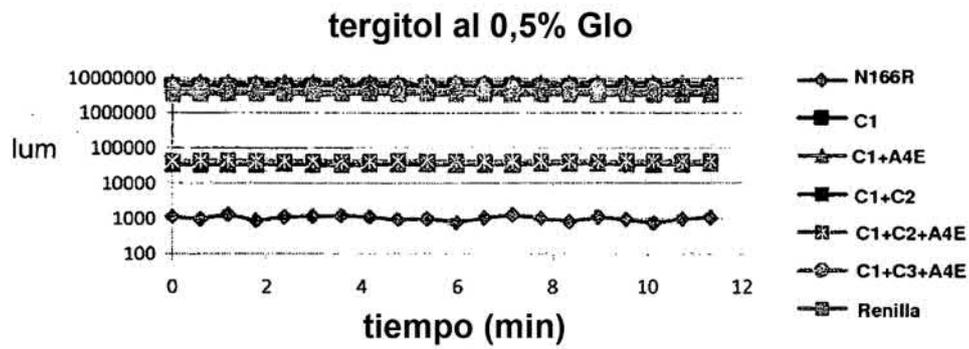


FIG. 4 cont.

T=0 (inyector) tampón de ensayo de tergitol al 0,5%

Muestra	Promedio	DesvEst	C V
W T	116.132	13.032	11,2%
N166R	160.334	18.014	11,2%
T2T	259.374	17.488	6,7%
T2T+N166R	367.468	30.406	8,3%
A4E	193.837	10.272	5,3%
Q11R	172.425	16.389	9,5%
V44I	2.155.303	231.323	10,7%
A54F	1.559.763	121.554	7,8%
A54F+N166R	8.052.686	981.637	12,2%
A54I	15.676.789	2.266.808	14,5%
P115E	406.670	18.160	4,5%
P115E+N166R	489.722	26.228	5,4%
Y138I	1.119.166	128.394	11,5%
Q124K	322.475	30.313	9,4%
Y138C+N166R	590.665	40.138	6,8%
I90V	387.765	29.162	7,5%

FIG. 5A

T=0 (inyector) tampón de ensayo de tergitol al 0,5%

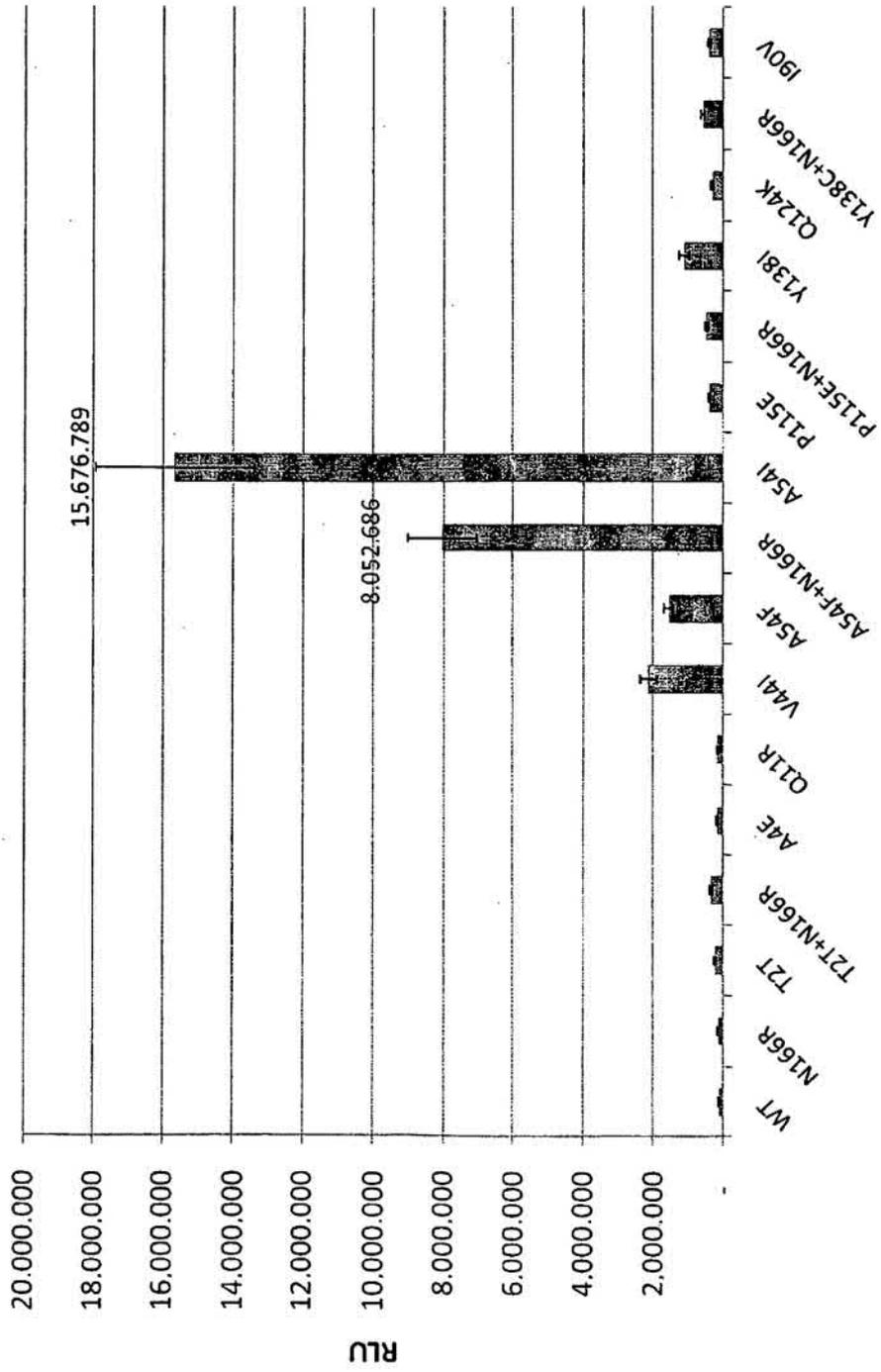


FIG. 5B

T=0 (inyector) tampón de ensayo de tergitol al 0,5%

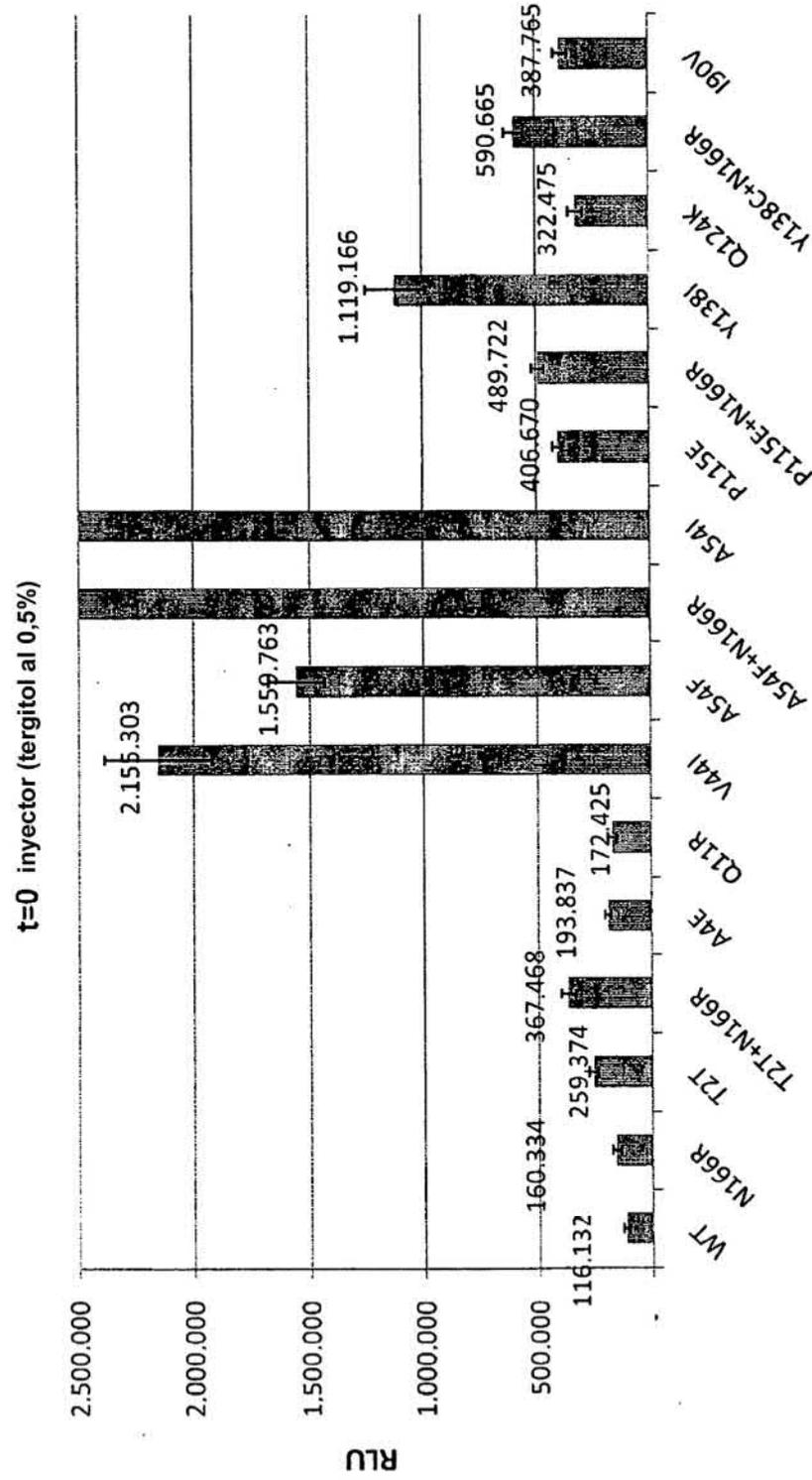


FIG. 5C

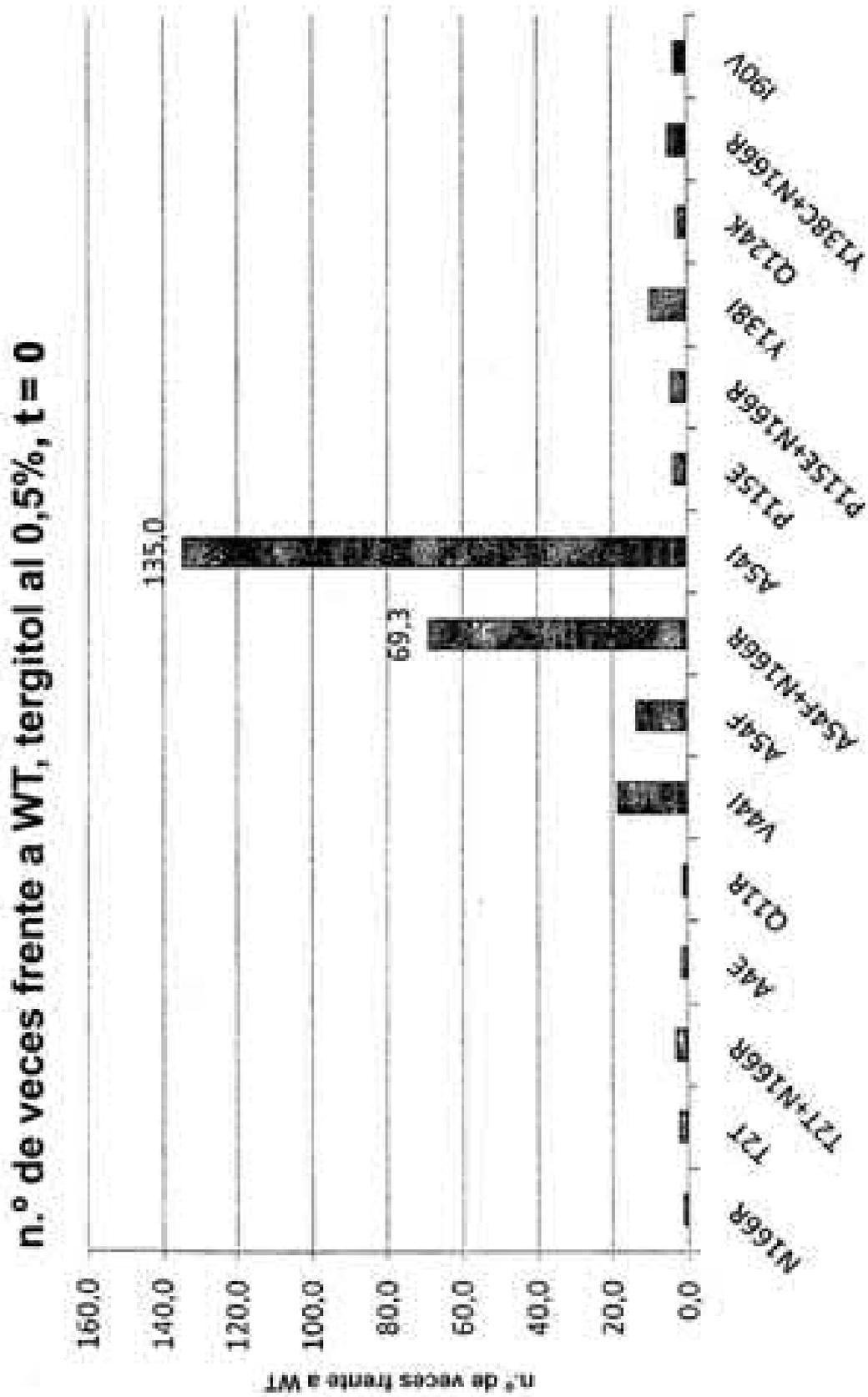


FIG. 6A

n.º de veces frente a WT, tergitol al 0,5%, t = 0

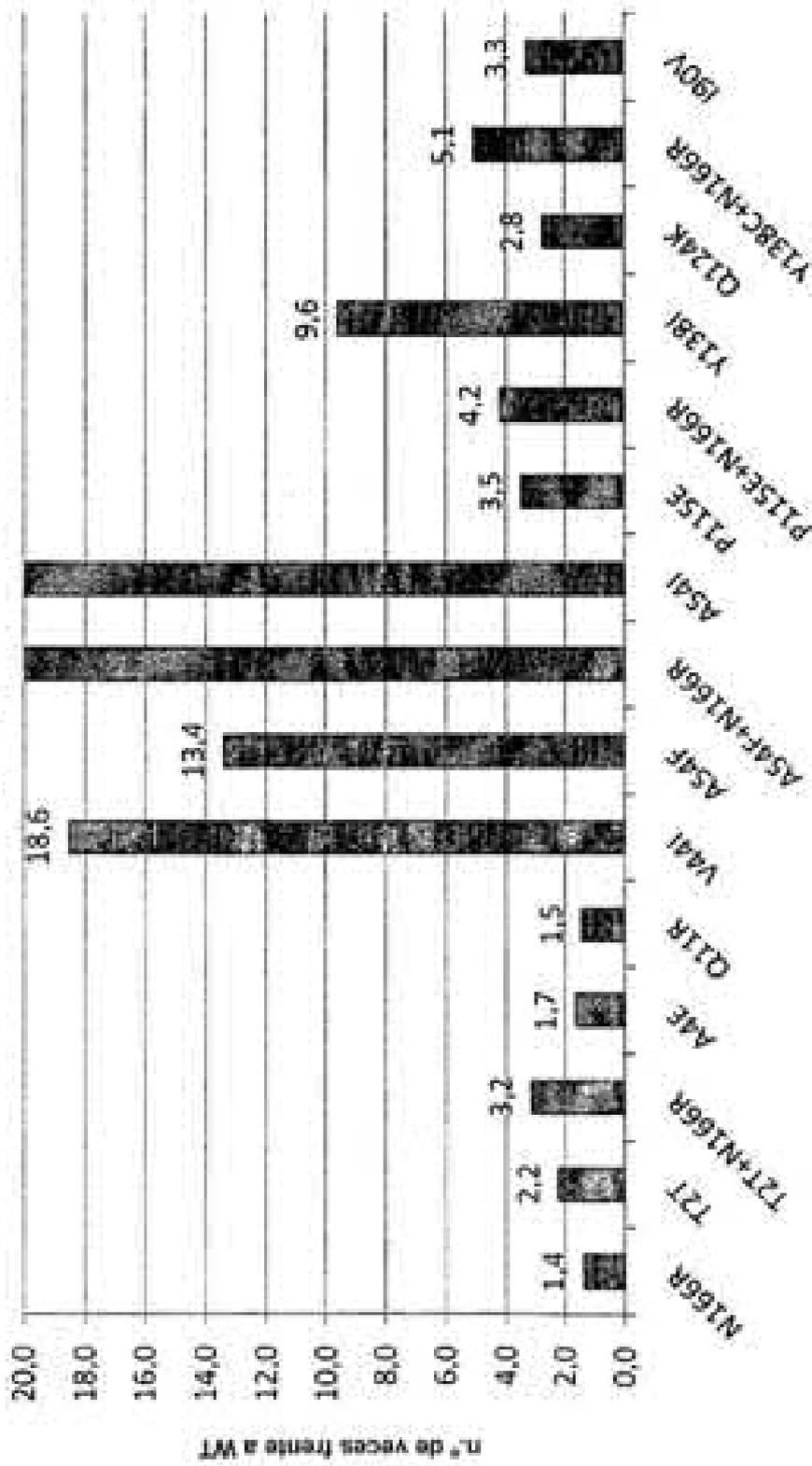


FIG. 6B

T=0 RLAB con inyector

Muestra	Promedio	DesvEst	CV
W T	2.266.822	222.850	9,8%
N166R	5.582.331	514.426	9,2%
T2T	3.599.341	244.441	6,8%
T2T+N166R	9.823.129	457.592	4,7%
A4E	2.888.005	231.055	8,0%
Q11R	4.043.391	179.037	4,4%
V44I	12.935.691	546.014	4,2%
A54F	8.572.288	464.442	5,4%
A54F+N166R	79.773.644	3.157.533	4,0%
A54I	32.309.914	2.594.059	8,0%
P115E	7.400.192	416.225	5,6%
P115E+N166R	16.167.230	481.714	3,0%
Y138I	14.573.569	619.909	4,3%
Q124K	5.473.417	517.799	9,5%
Y138C+N166R	15.516.040	1.147.103	7,4%
I90V	4.474.794	297.720	6,7%

FIG. 7A

T=0 RLAB con inyector

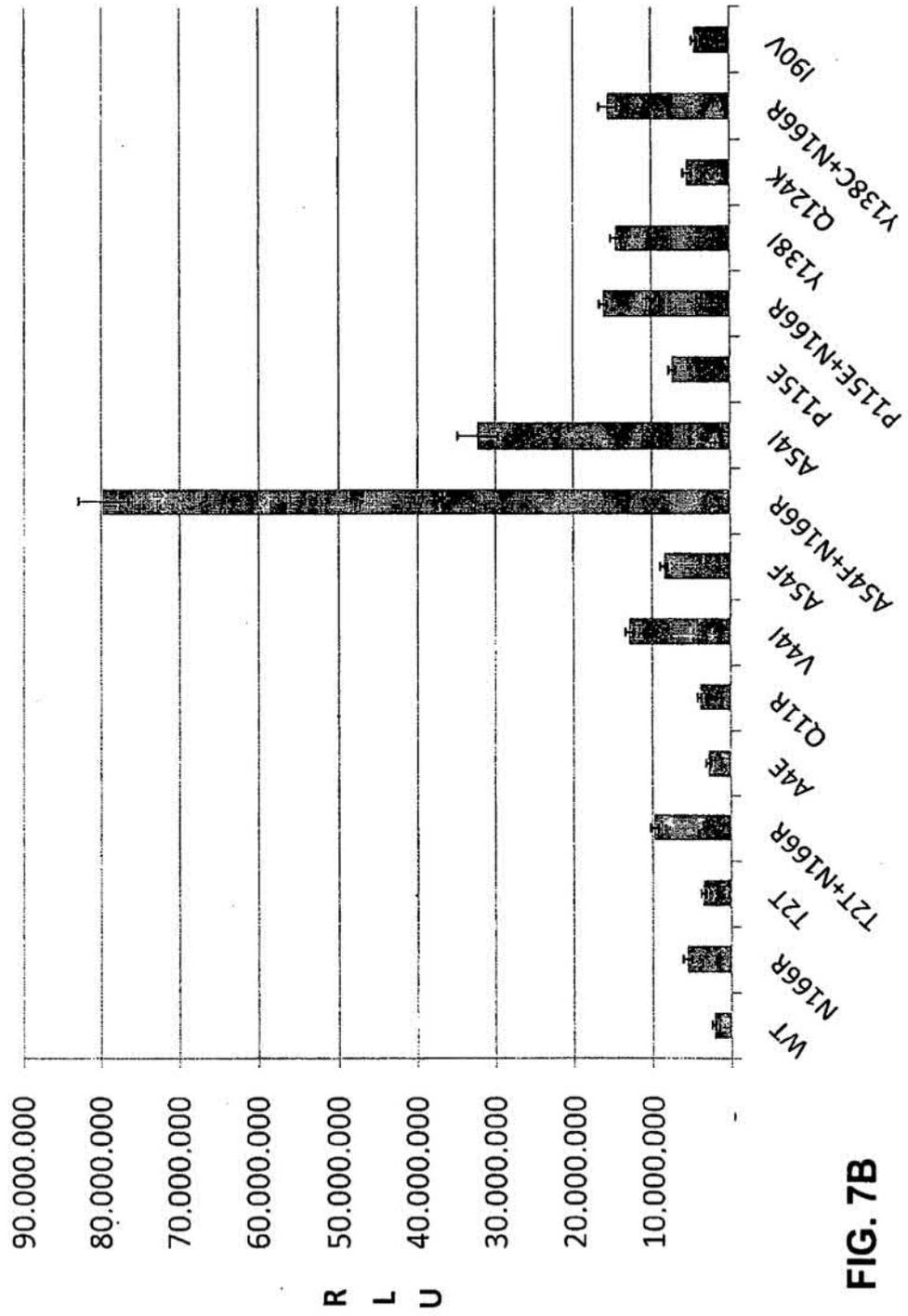


FIG. 7B

T=0 RLAB con inyector

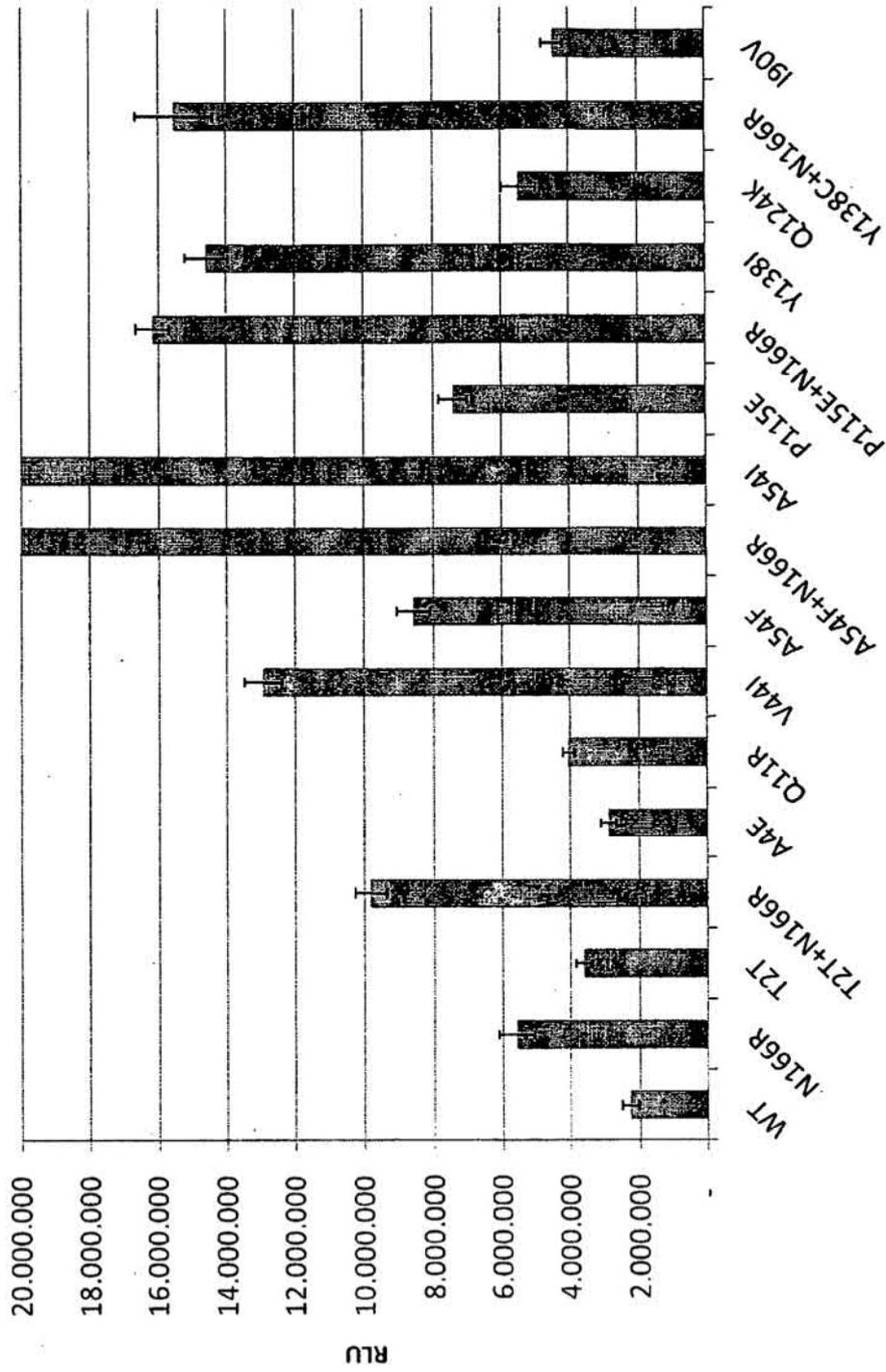


FIG. 7C

RLAB t=0 (con inyector)
n.º de veces frente a WT

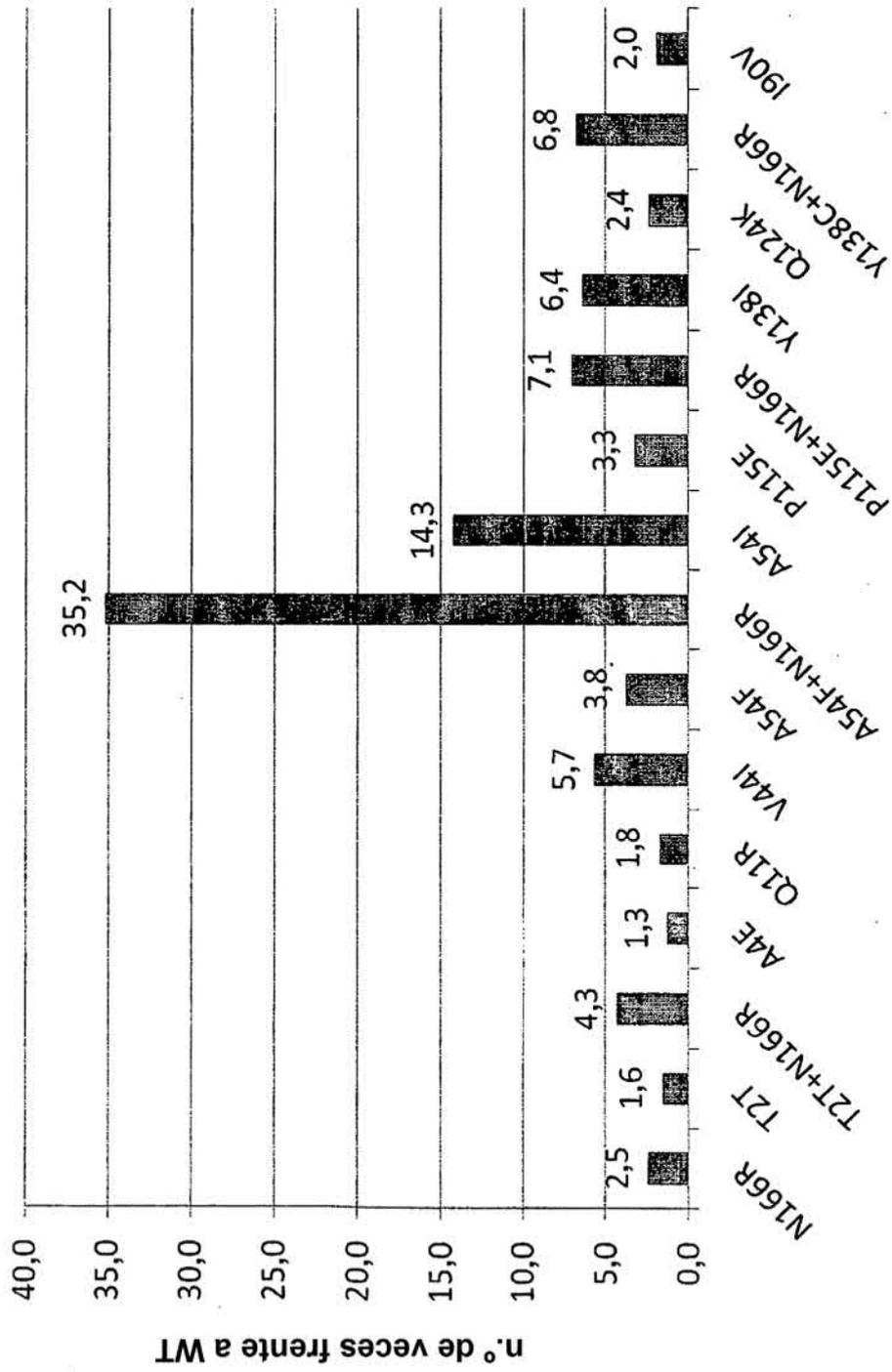


FIG. 8

Señal de clones de Og-luc con tampón de ensayo de tergitol al 0,5%

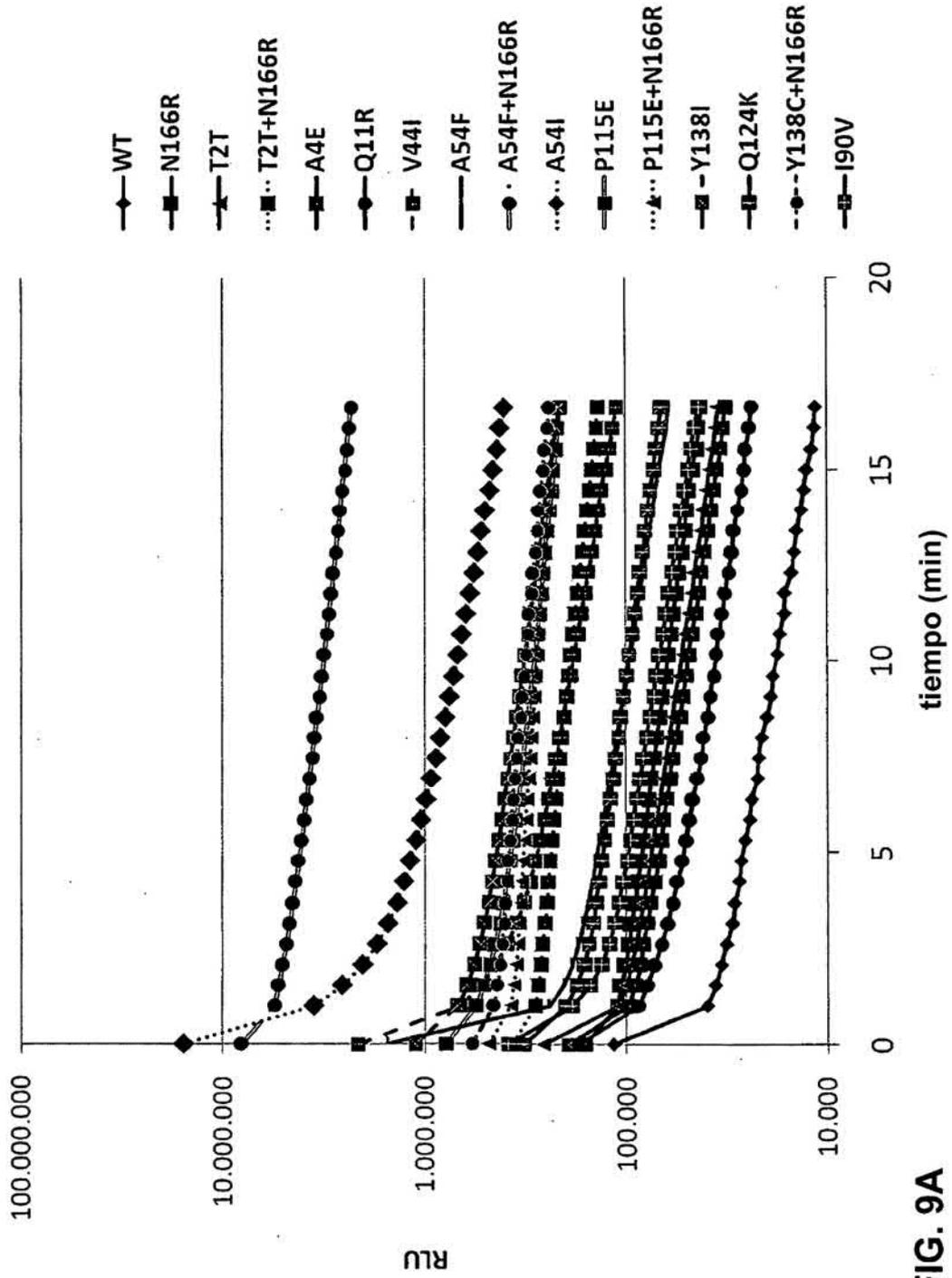


FIG. 9A

Estabilidad de la señal en tampón de tergitol al 0,5%

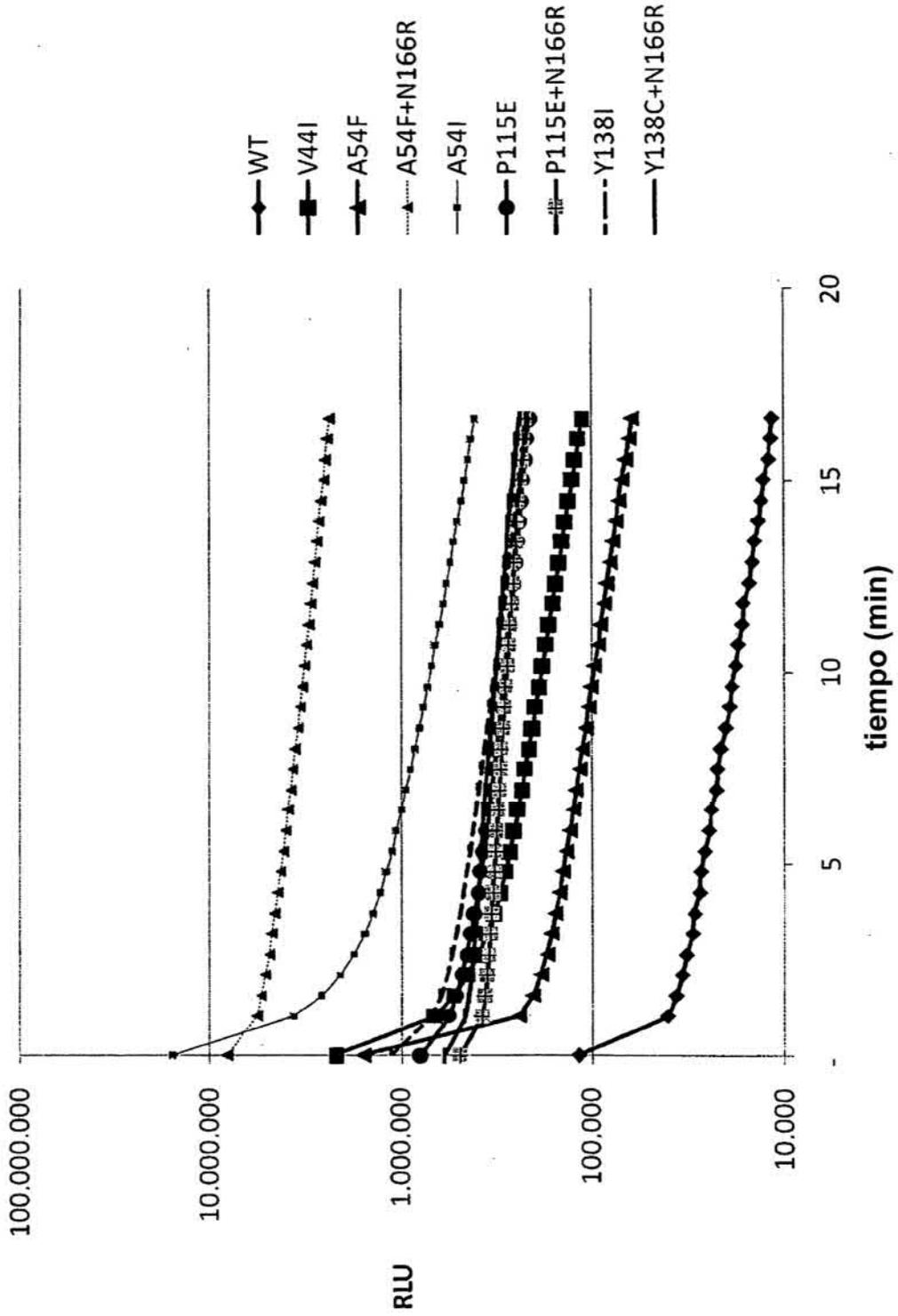


FIG. 9B

Estabilidad de la señal en tampón de tergitol al 0,5%

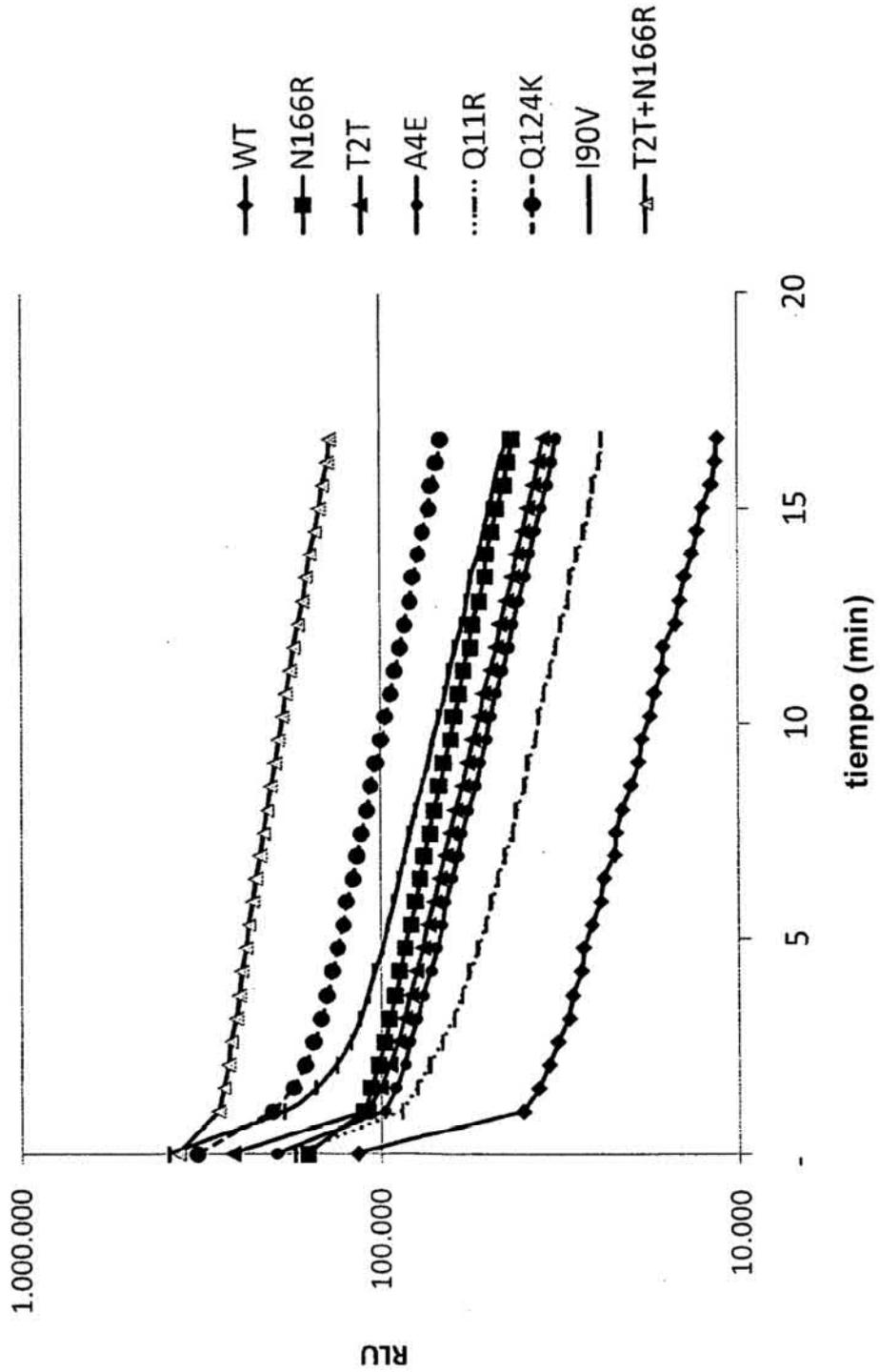


FIG. 9C

Semivida de la señal (tampón de ensayo de tergitol al 0,5%)

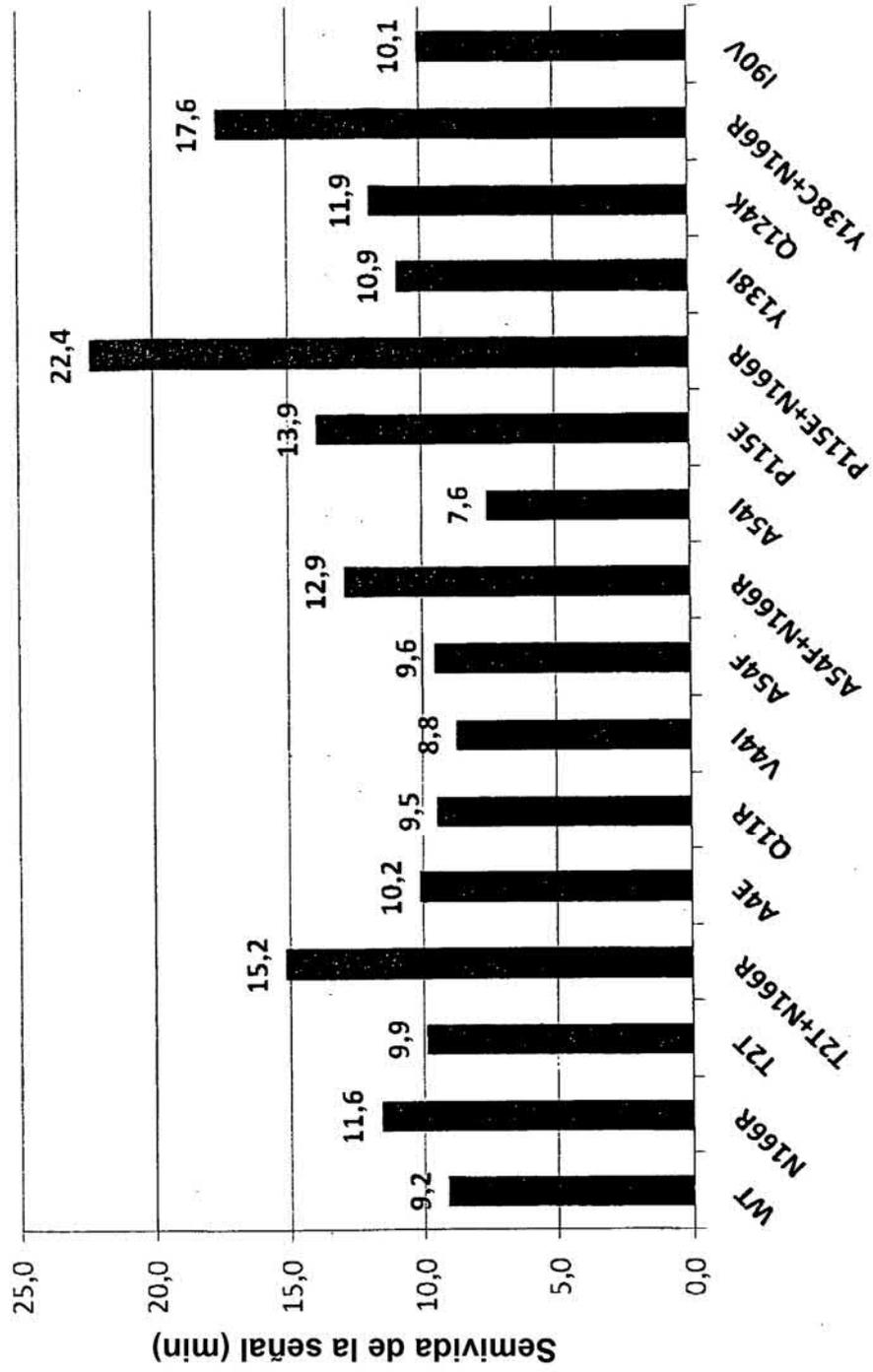


FIG. 9D

Estabilidad de la señal en RLAB

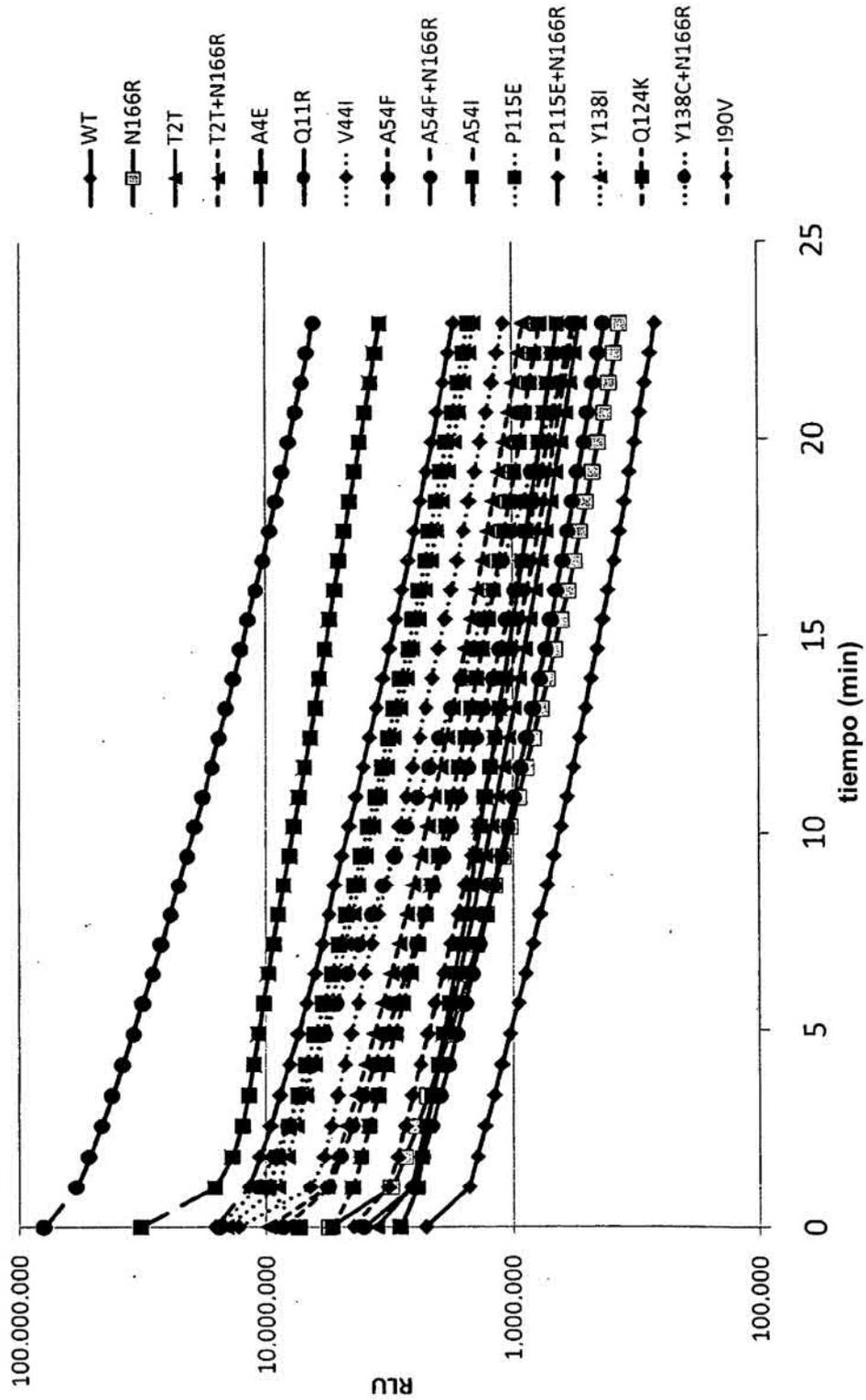


FIG. 10A

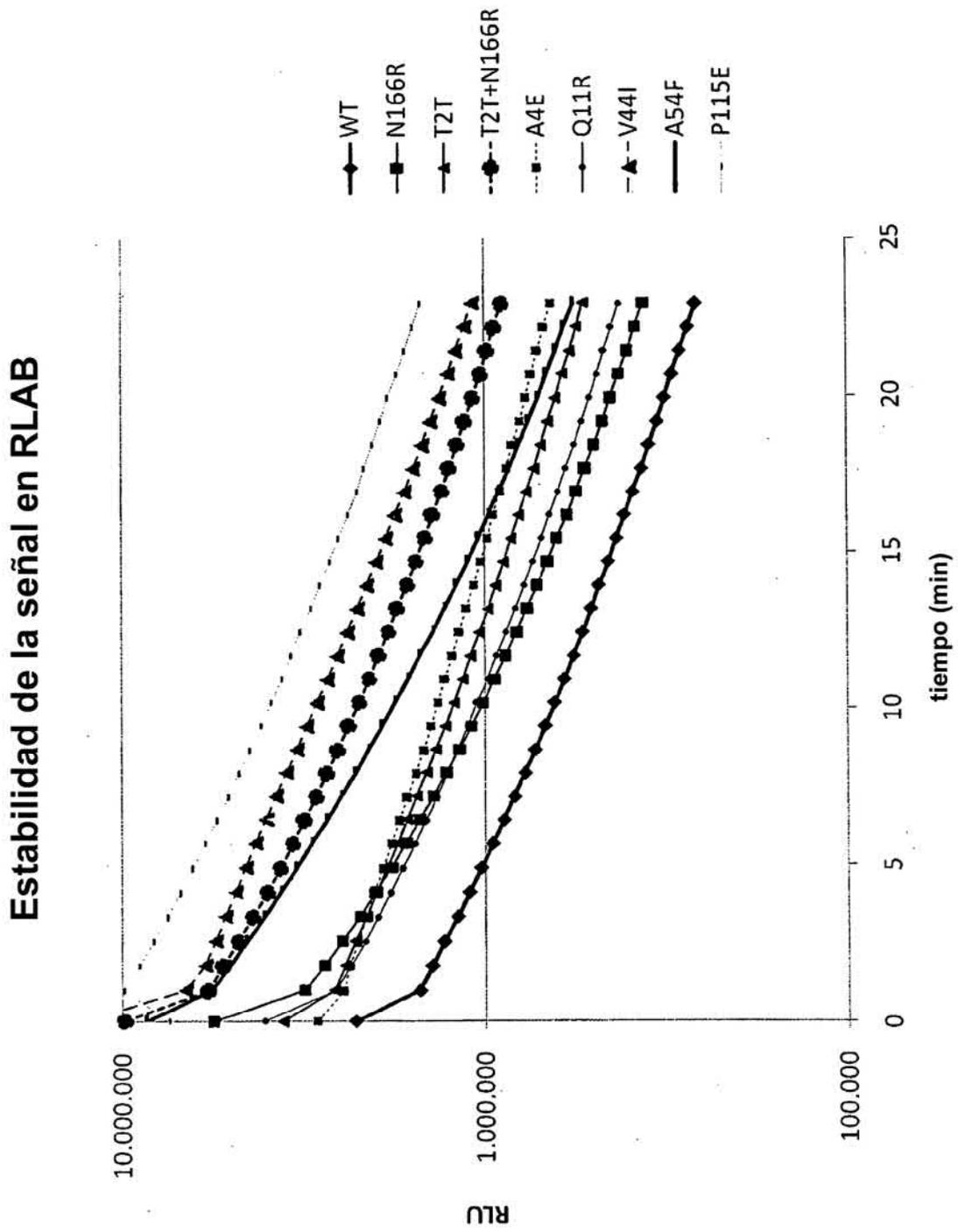


FIG. 10B

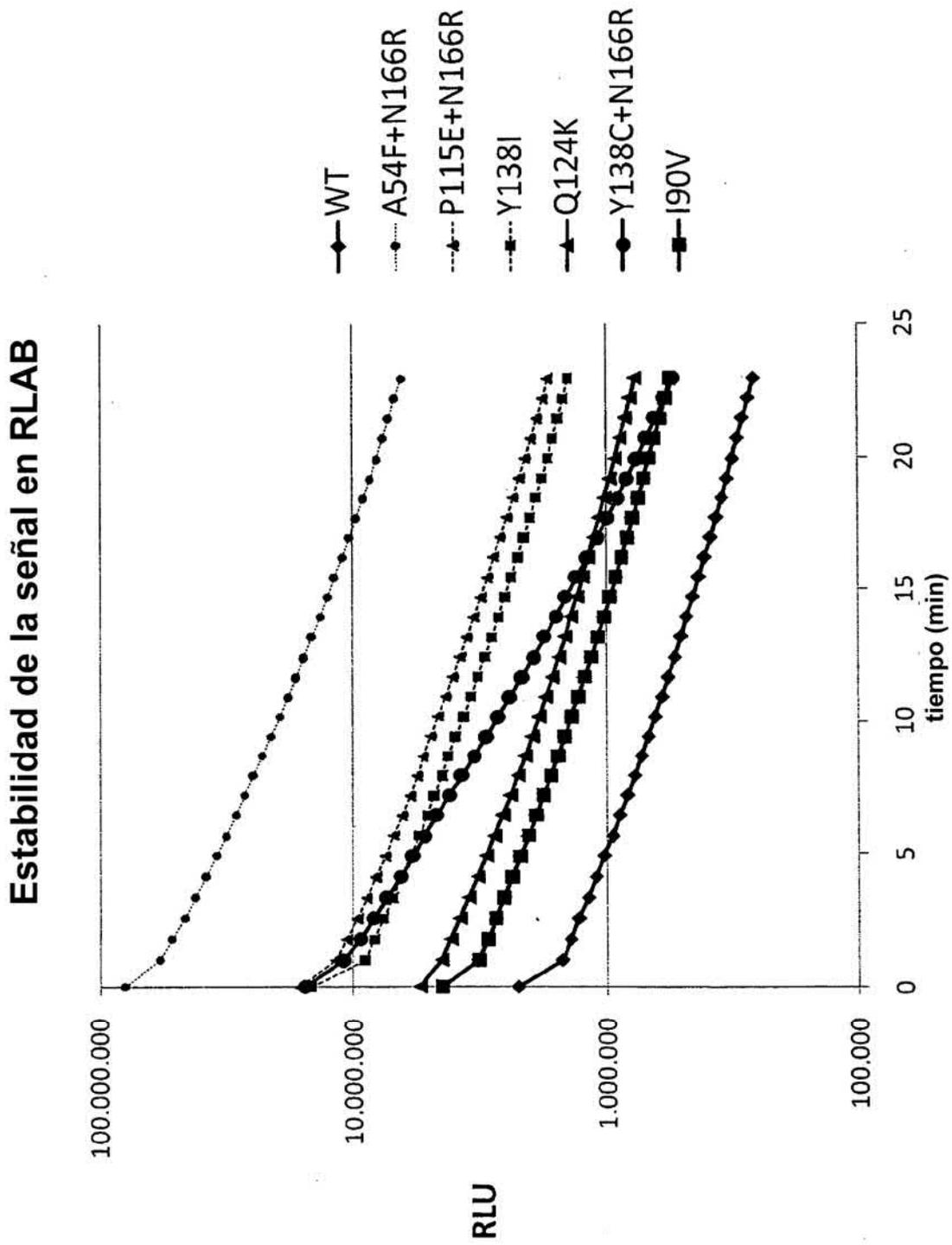


FIG. 10C

Semivida de la señal (minutos) RLAB

Muestra	Semivida de la señal (minutos) RLAB
W T	9, 0 3
N 1 6 6 R	8, 1 7
T 2 T	9, 7 9
T 2 T + N 1 6 6 R	8, 7 4
A 4 E	1 1, 4 8
Q 1 1 R	8, 6 9
V 4 4 I	8, 5 4
A 5 4 F	6, 8 0
A 5 4 F + N 1 6 6 R	7, 3 0
A 5 4 I	1 0, 8 8
P 1 1 5 E	8, 3 1
P 1 1 5 E + N 1 6 6 R	8, 1 6
Y 1 3 8 I	8, 3 6
Q 1 2 4 K	8, 7 7
Y 1 3 8 C + N 1 6 6 R	4, 9 7
I 9 0 V	8, 8 3

FIG. 11A

Semivida de la señal (minutos) RLAB

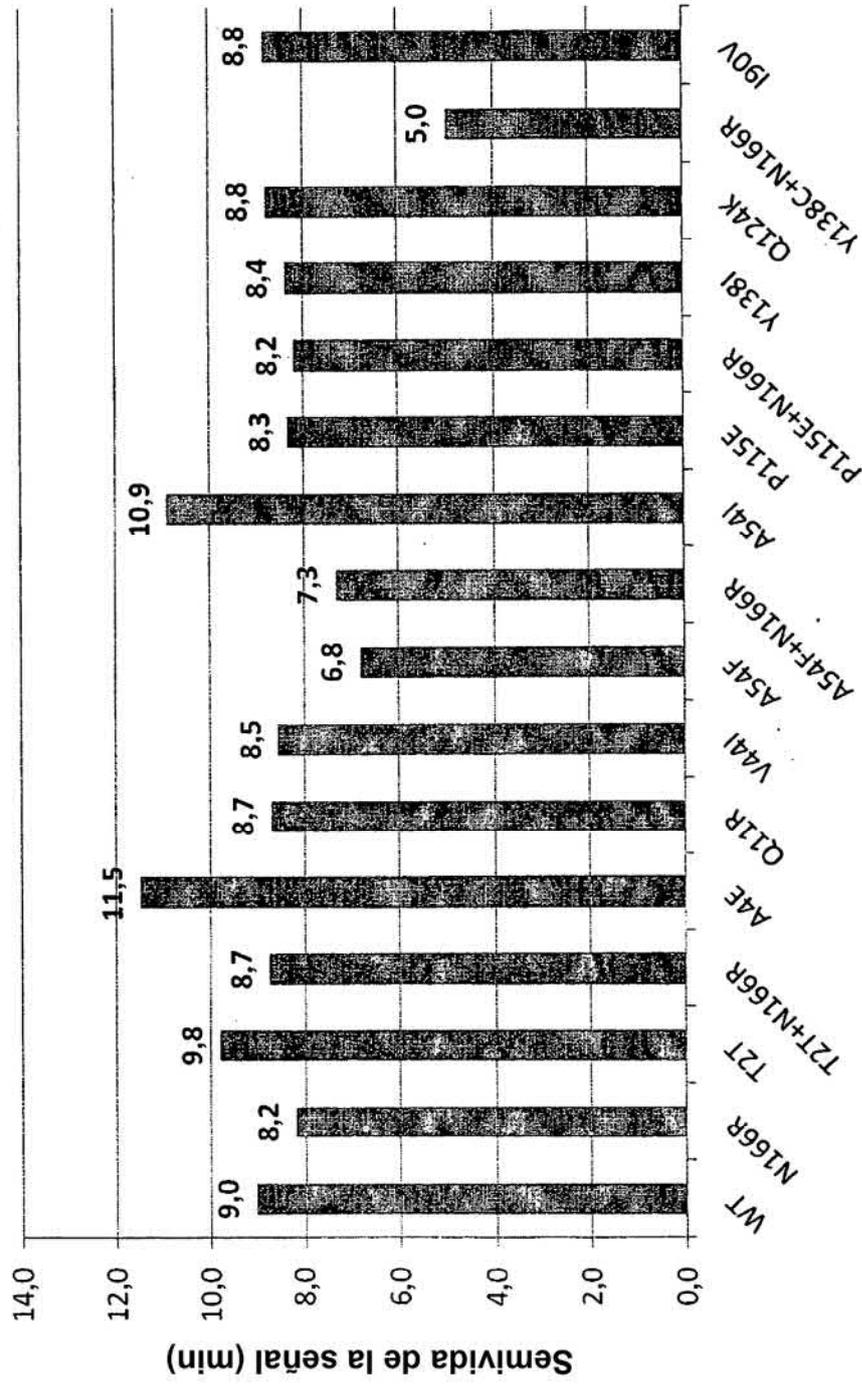


FIG. 11B

Estabilidad a 22 °C (semivida en minutos)

Muestra	Estabilidad (semivida a 22 °C)
W T	38,1
N 166 R	61,3
T 2 T	43,1
T 2 T + N 166 R	68,0
A 4 E	38,5
Q 11 R	45,9
V 44 I	49,5
A 54 F	56,8
A 54 F + N 166 R	177,7
A 54 I	63,0
P 115 E	49,9
P 115 E + N 166 R	119,5
Y 138 I	55,9
Q 124 K	43,9
Y 138 C + N 166 R	35,5
I 90 V	41,3

FIG. 12A

Estabilidad a 22 °C (semivida en minutos)

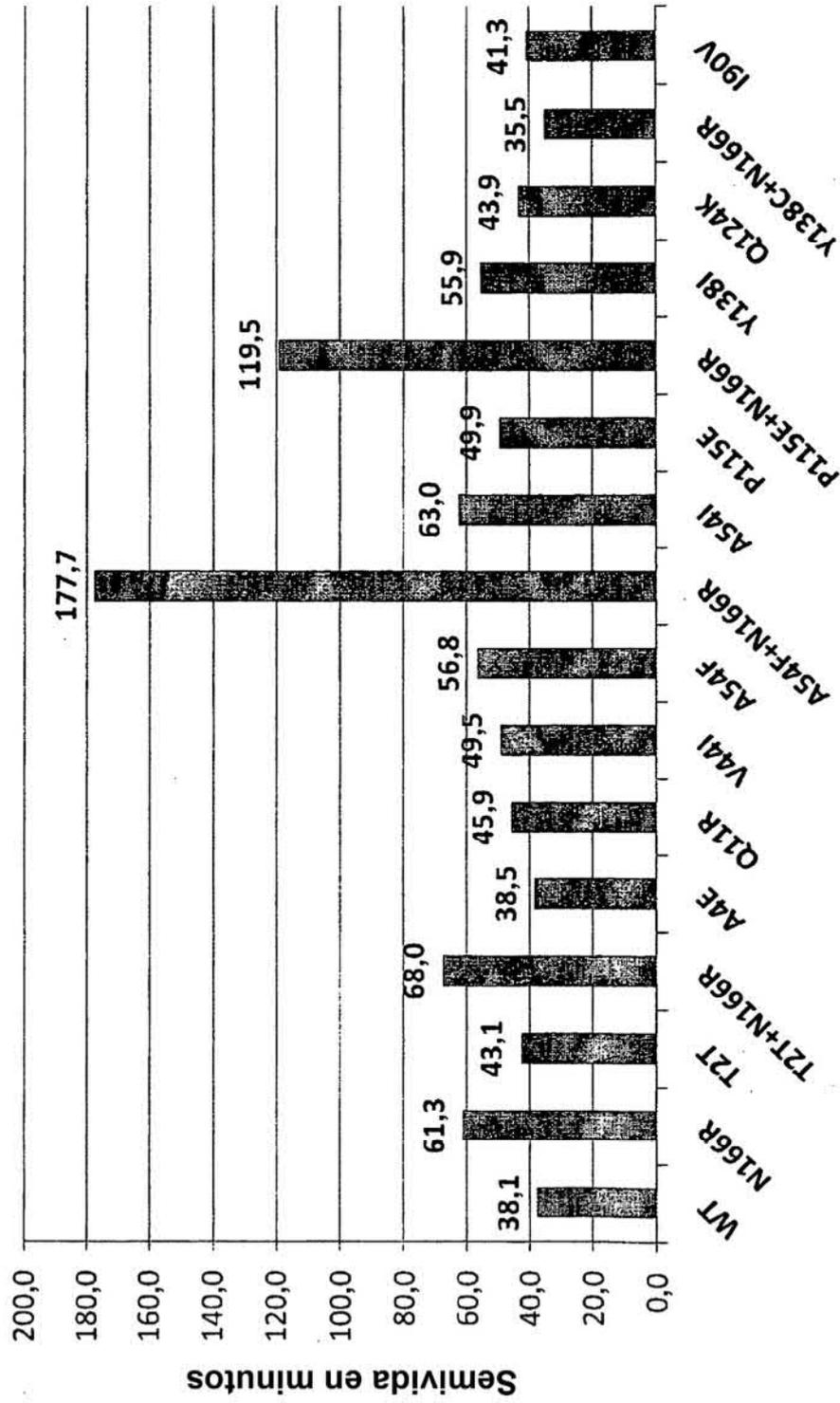


FIG. 12B

T = 0 tampón de ensayo de tergitol al 0,5%

Muestra	Promedio	% c v
W T	1 1 5 1 9 0 , 4	8 , 6 %
N 1 6 6 R	1 6 6 6 4 7	1 2 , 2 %
W T A 3 3 K	4 3 8 4 5 5	7 , 8 %
W T F 6 8 Y	4 4 9 0 8 3 , 3	6 , 5 %
N 1 6 6 R A 3 3 K	1 0 5 4 3 8 7	8 , 0 %
N 1 6 6 R F 6 8 Y	5 7 1 3 4 7 , 1	4 , 7 %

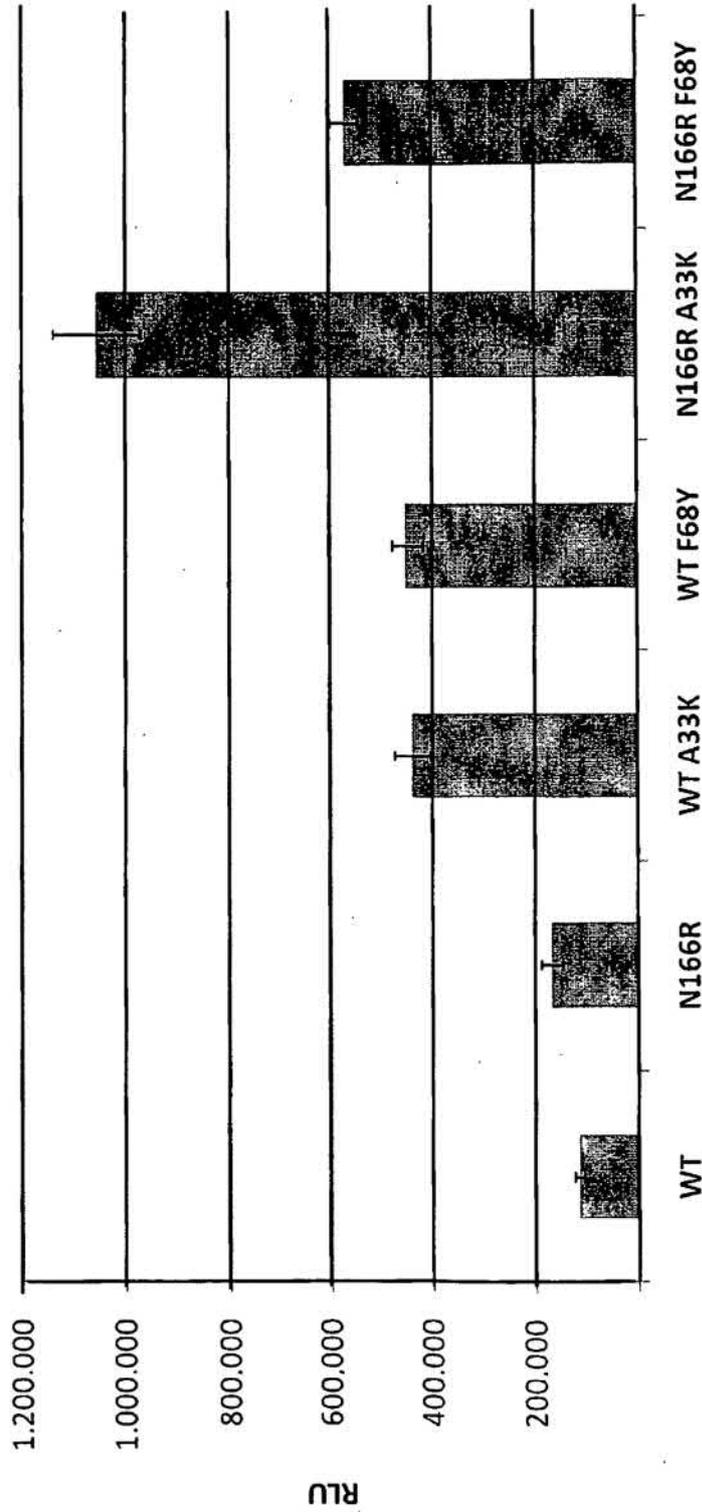


FIG. 13A

RLAB t=0 RLU

Muestra	Promedio	c v
W T	27.306.288	8,9%
N166R	69.445.779	9,2%
W T A 33K	44.140.390	14,2%
W T F 68Y	45.593.349	10,3%
N166R A 33K	138.736.058	8,0%
N166R F 68Y	89.324.499	12,9%

RLAB t=0

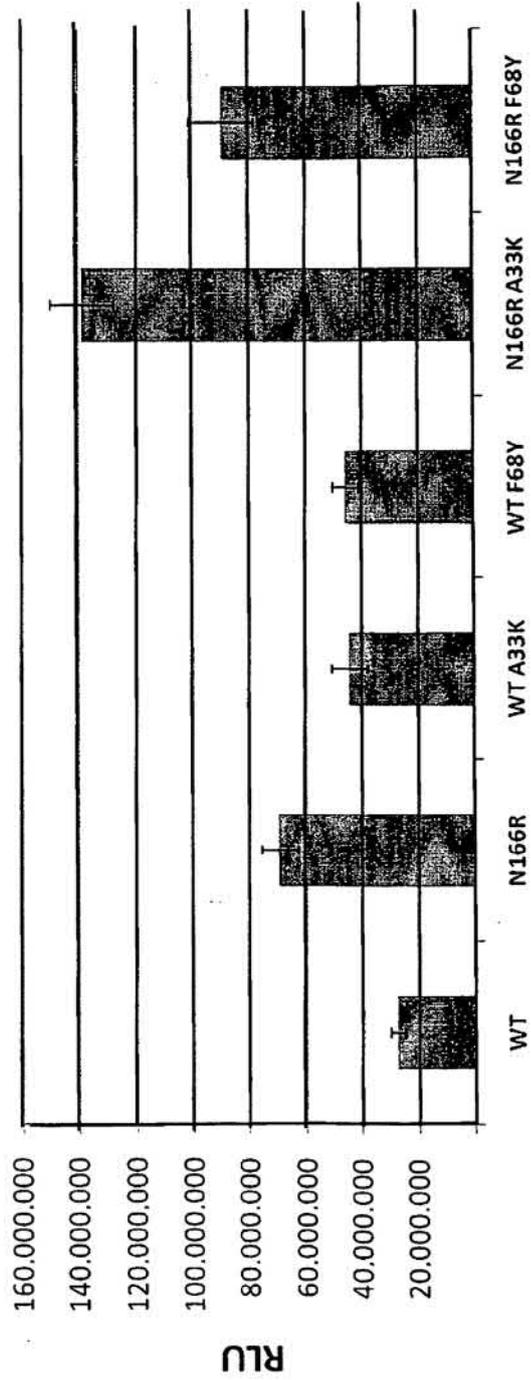


FIG. 13B

n.º de veces frente a WT (tampón de ensayo de tergitol al 0,5%)

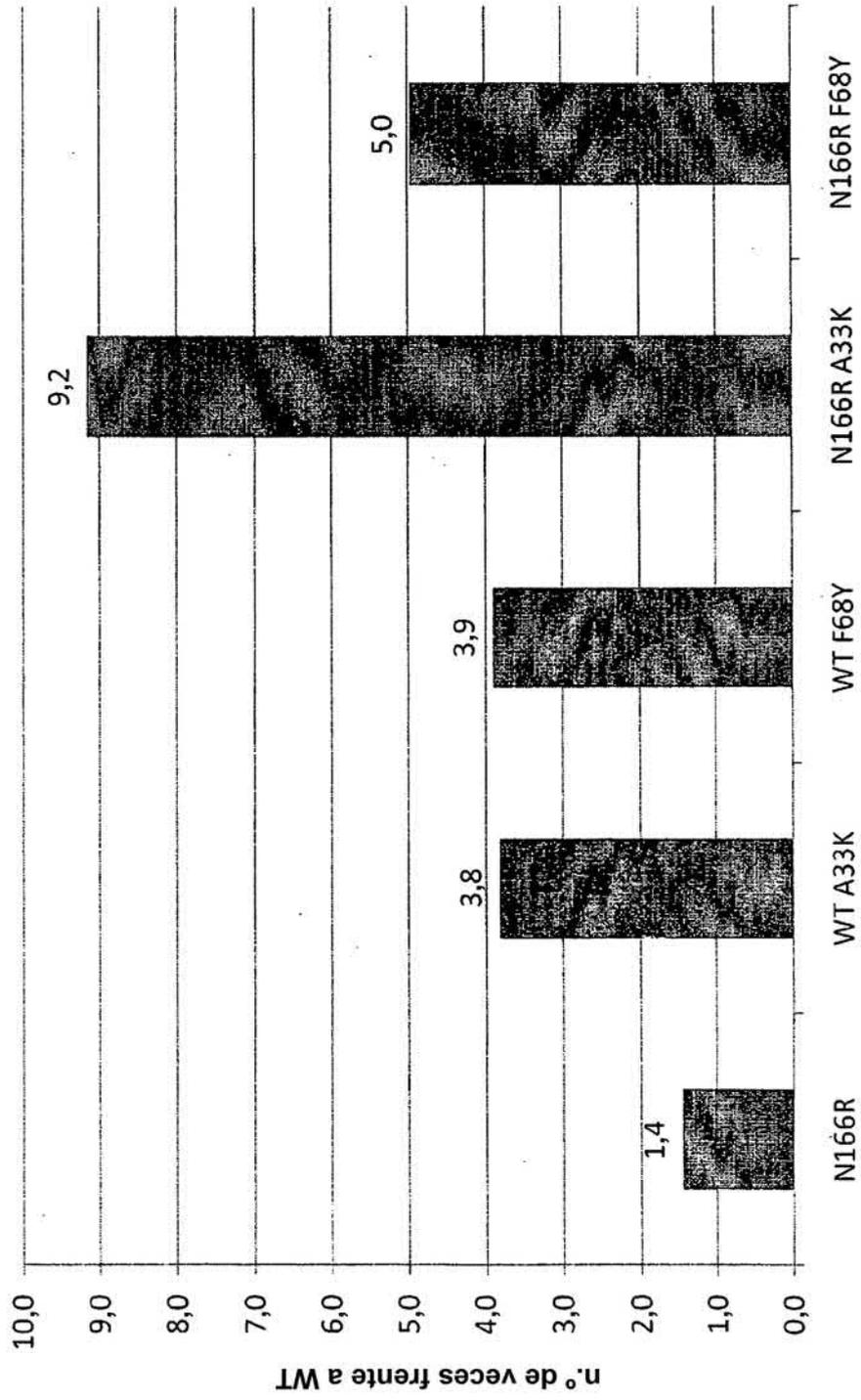


FIG. 14A

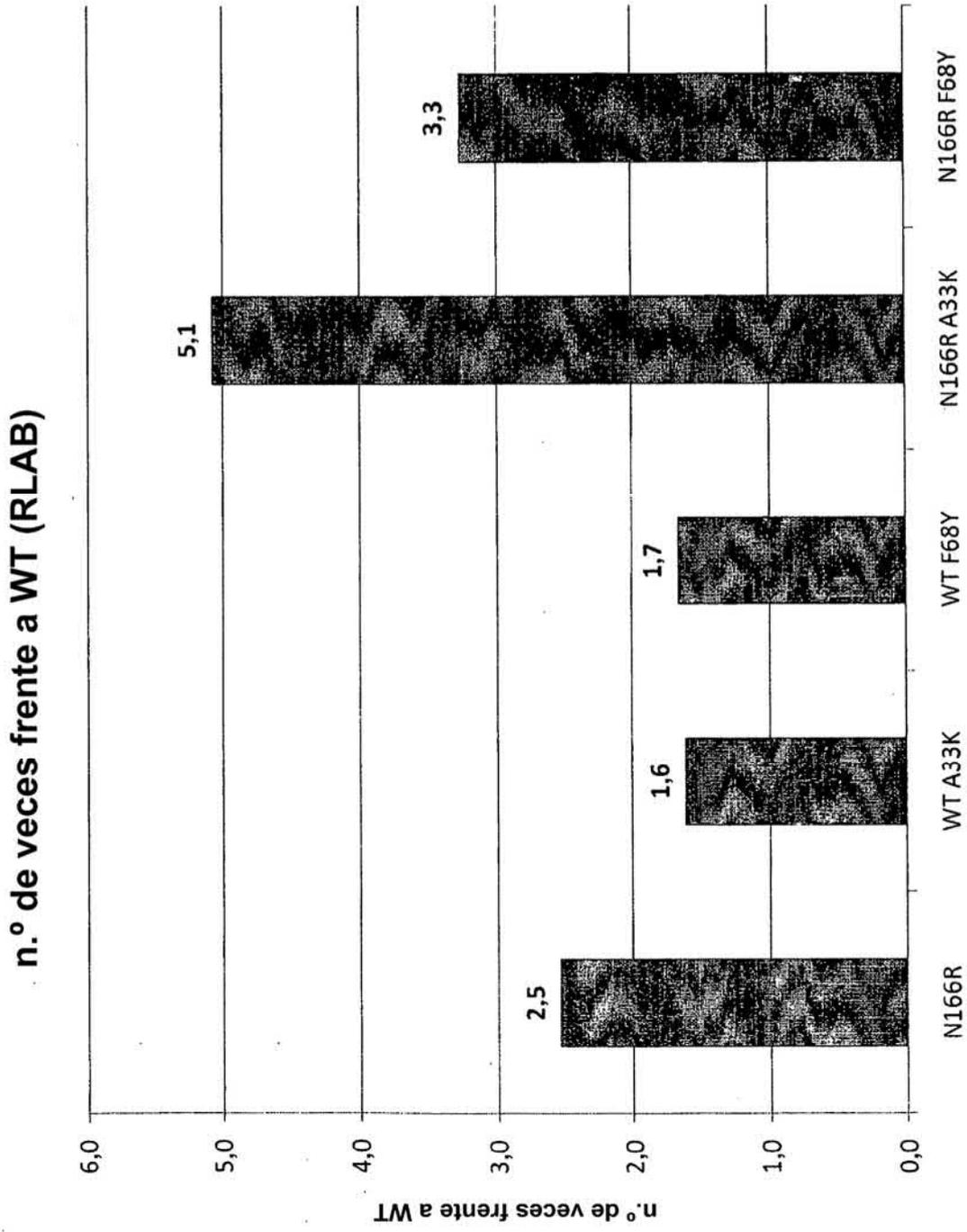


FIG. 14B

Cinética de la señal (tampón de ensayo de tergitol al 0,5%)

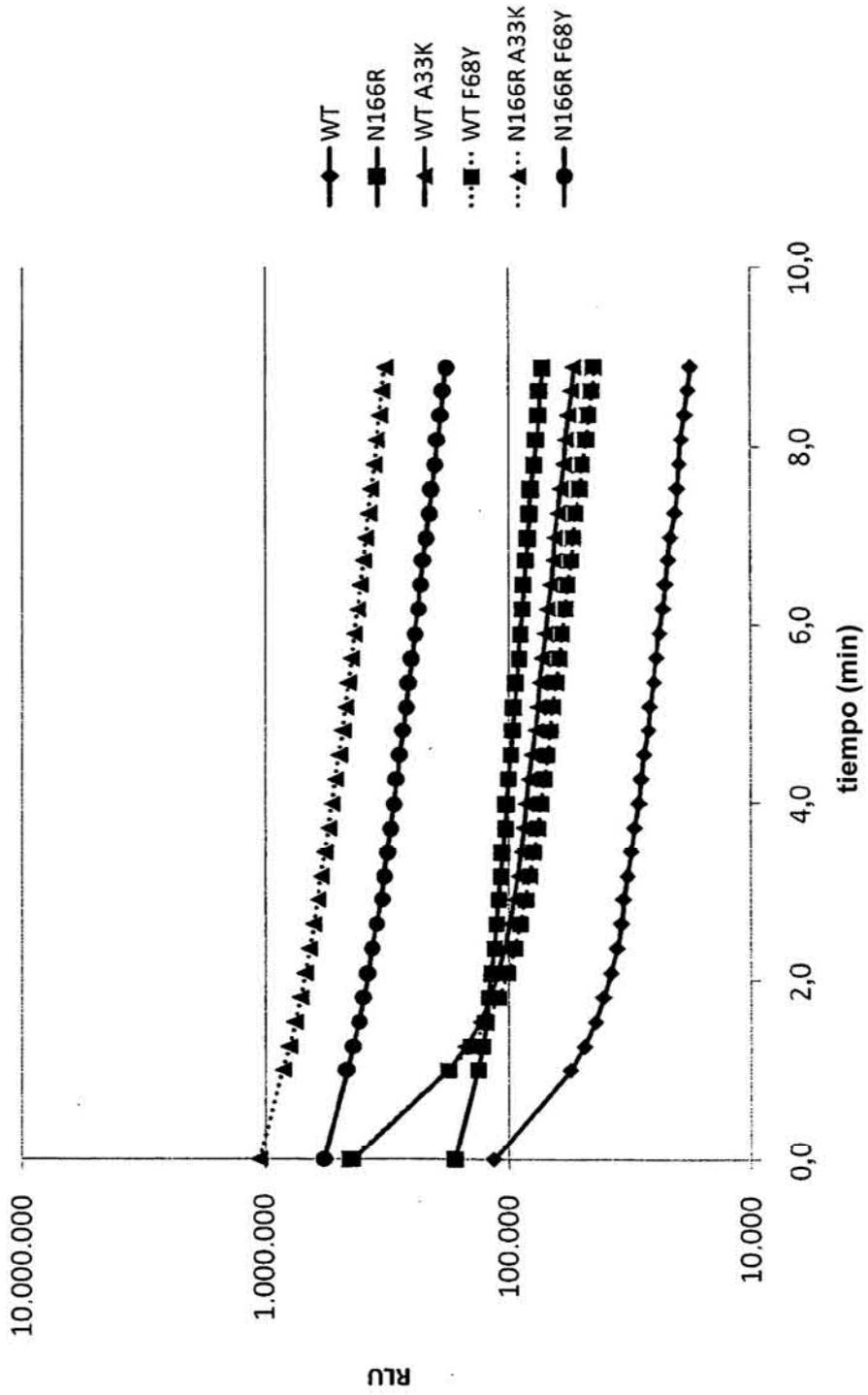


FIG. 15A

Estabilidad de la señal, tampón de ensayo de tergitol al 0,5%

Muestra	Semivida de la señal (tampón de ensayo de tergitol al 0,5%)
WT	6,5
N166R	9,6
WTA33K	7,4
WTF68Y	6,3
N166R A33K	6,3
N166R F68Y	6,2

FIG. 15B

Cinética de la señal (RLAB)

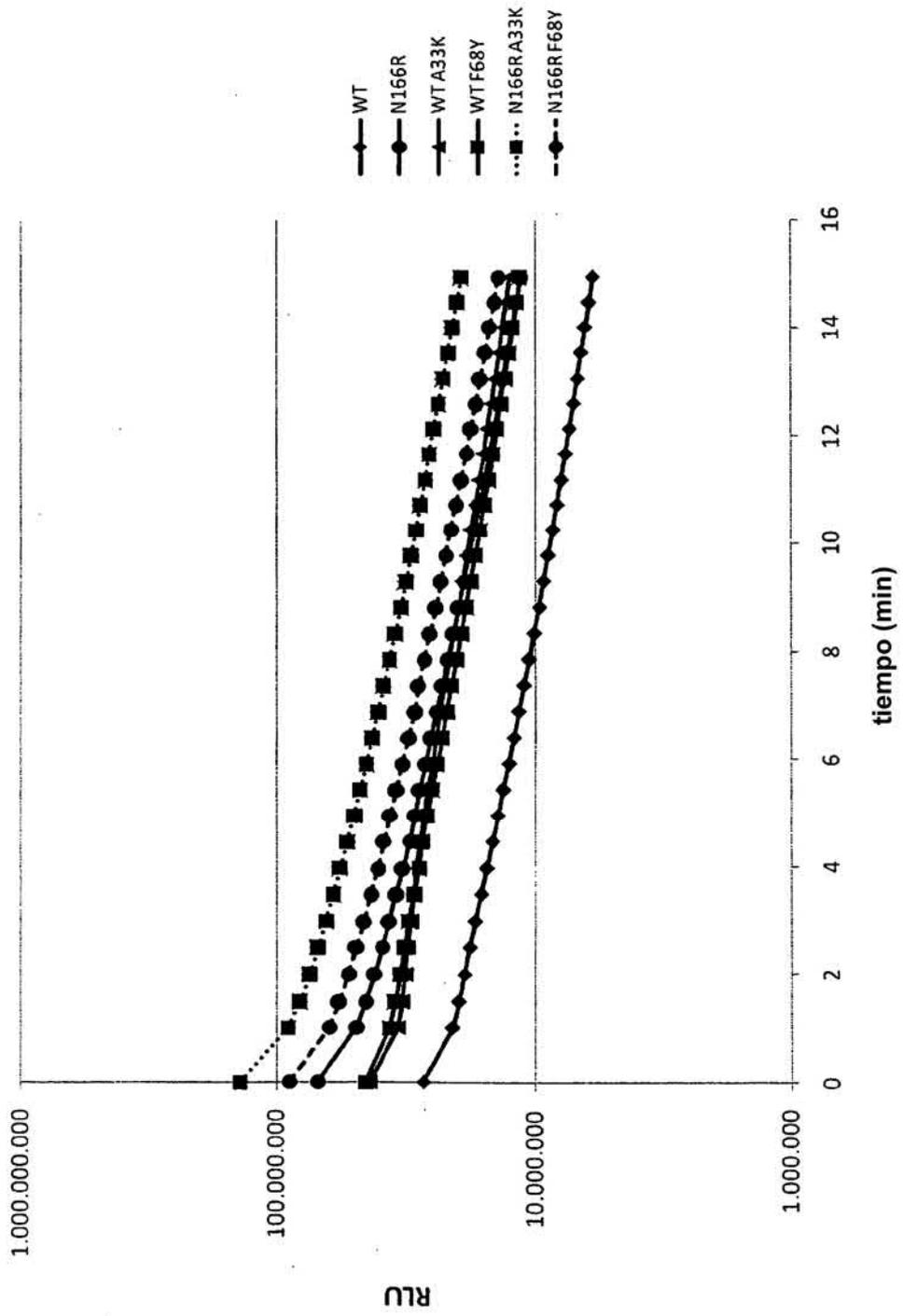


FIG. 16A

Estabilidad de la señal RLAB

Muestra	Semivida de la señal, minutos
WT	7,64
N166R	6,73
A33K	9,22
F68Y	8,13
N166R/A33K	6,58
N166R/F68Y	6,60

FIG. 16B

Semivida de la estabilidad a 22 °C

Muestra	Semivida (min), 22 °C
WT	54,6
N166R	67,3
A33K	58,2
F68Y	57,3
N166R, A33K	72,2
N166R, F68Y	77,9

FIG. 17

Tampón de ensayo de tergitol al 0,5%

tiempo (min)	N166R	C1+A4E	C1+A4E+F54T	C1+C2+A4E	C1+C3+A4E	C1+C3+A4E+F54T	Renilla
0,0	1.110	7.910.000	3.583.750	43.371	4.217.500	3.322.500	3.542.500
0,6	970	7.876.250	3.587.500	43.902	4.217.500	3.326.250	3.543.750
1,2	1.277	7.888.750	3.576.250	45.133	4.243.750	3.351.250	3.557.500
1,8	872	7.868.750	3.566.250	43.553	4.257.500	3.338.750	3.563.750
2,4	1.110	7.887.500	3.575.000	42.476	4.260.000	3.366.250	3.571.250
3,0	1.137	7.880.000	3.576.250	43.036	4.266.250	3.397.500	3.562.500
3,6	1.193	7.860.000	3.577.500	43.637	4.276.250	3.386.250	3.586.250
4,2	1.095	7.855.000	3.556.250	42.896	4.293.750	3.406.250	3.568.750
4,8	928	7.862.500	3.556.250	44.755	4.290.000	3.415.000	3.565.000
5,4	998	7.852.500	3.538.750	44.112	4.315.000	3.422.500	3.568.750
6,0	774	7.843.750	3.542.500	44.196	4.311.250	3.433.750	3.551.250
6,6	1.026	7.841.250	3.520.000	42.938	4.332.500	3.431.250	3.525.000
7,2	1.277	7.786.250	3.516.250	43.217	4.311.250	3.445.000	3.512.500
7,8	1.012	7.783.750	3.511.250	43.986	4.318.750	3.430.000	3.486.250
8,4	816	7.775.000	3.491.250	43.483	4.317.500	3.440.000	3.467.500
9,0	1.110	7.771.250	3.471.250	42.854	4.317.500	3.450.000	3.416.250
9,6	928	7.750.000	3.460.000	44.713	4.308.750	3.465.000	3.422.500
10,1	732	7.705.000	3.457.500	44.168	4.320.000	3.462.500	3.386.250
10,7	970	7.706.250	3.451.250	43.245	4.296.250	3.433.750	3.363.750
11,3	1.068	7.687.500	3.430.000	43.902	4.315.000	3.450.000	3.333.750

FIG. 18A

Tampón de ensayo de tergitol al 0,5%

lisados de *E. coli*

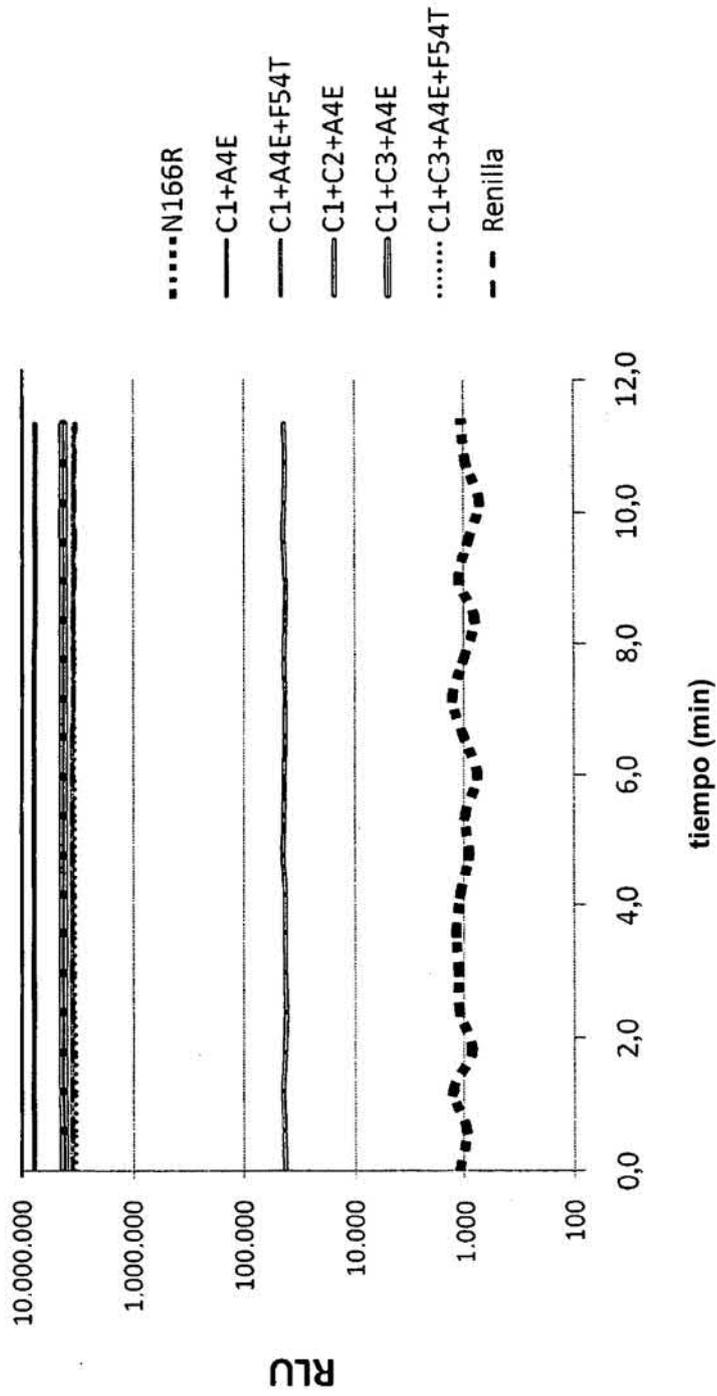


FIG. 18B

RLAB con inyector, lisados de *E. coli*

tiempo (min)	C1+A4E	C1+A4E+F54T	C1+C3+A4E	C1+C3+A4E+F54T	renilla	WT
0	6.288.773	10.015.723	3.998.483	19.218.247	5.721.221	10.170
1	3.880.142	1.479.839	3.501.544	6.033.827	2.111.378	3.490
2	2.515.425	242.450	2.948.188	1.422.742	827.548	1.753
3	1.739.960	99.503	2.378.597	526.400	437.314	1.326
4	1.269.359	66.080	1.926.802	273.345	277.539	966
5	957.041	51.848	1.587.709	172.383	195.944	843
6	734.186	41.485	1.320.817	120.617	147.764	729
7	573.084	34.985	1.106.277	89.734	116.523	609
8	453.838	29.690	944.146	69.148	95.257	469
9	363.570	25.892	803.433	55.316	78.655	513
10	296.648	22.868	696.959	45.239	66.302	423
11	244.349	19.838	606.222	38.464	55.581	296
12	203.980	17.906	533.878	33.612	48.518	289
13	171.296	16.269	469.833	29.523	40.784	253
14	145.640	14.579	420.908	26.163	36.657	236

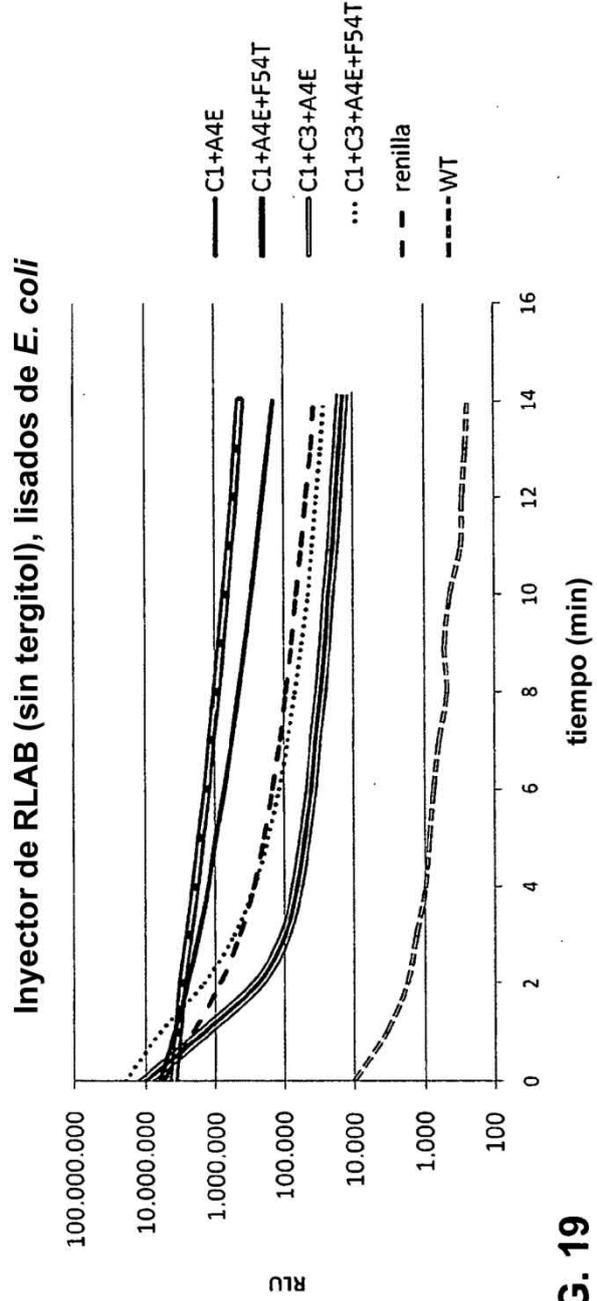


FIG. 19

Compuestos de Og luc, tergitol al 0,5%

tiempo (min)	C1+C2+A4E	C1+A4	hRL	Og-Luc	T2T	A54F
0	3.283.757	705.604.545	82.499.609	40.946	129.209	3.196.722
1	3.272.774	655.089.672	81.682.809	43.269	144.141	3.053.210
2	3.263.848	607.514.297	80.390.118	39.618	141.818	3.051.850
3	3.323.641	561.553.077	78.958.229	45.923	164.050	2.972.811
4	3.251.173	520.696.293	77.087.348	29.000	165.045	2.863.212
5	3.259.501	481.737.466	75.545.993	54.218	178.318	2.773.522
6	3.332.268	443.689.477	73.595.608	38.955	168.031	2.774.186
7	3.290.426	408.149.817	71.668.815	37.296	184.622	2.630.011
8	3.389.076	375.439.634	69.682.029	44.264	180.641	2.614.748
9	3.320.986	346.313.863	67.570.182	42.936	192.588	2.575.859

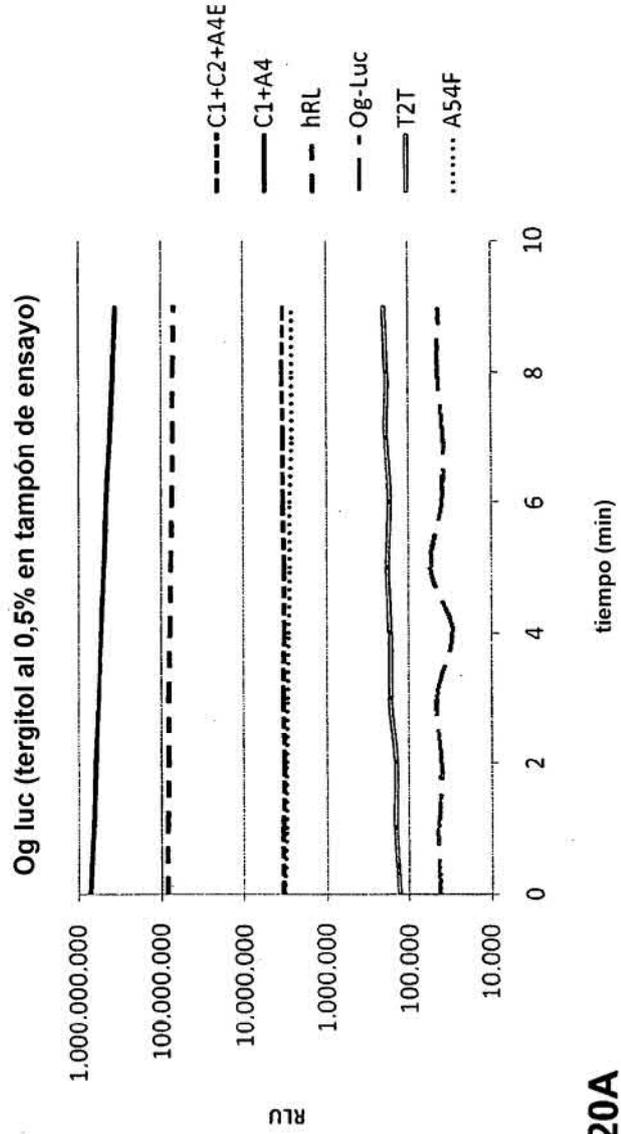


FIG. 20A

lisados de *E. coli*, inyector de RLAB

tiempo (min)	WT	Ren	C1+A4E	C1+C2+A4E	C1+C3+A4E
0	1.516	5.224.047	7.334.767	41.049	3.886.812
1	1.146	1.400.228	4.282.362	35.602	3.162.835
2	690	628.021	2.940.985	31.649	2.451.582
3	493	362.496	2.127.271	29.607	1.842.268
4	393	243.436	1.606.017	27.491	1.412.564
5	266	176.676	1.248.100	26.635	1.117.276
6	253	137.402	998.258	25.333	906.594
7	203	111.480	820.954	24.741	751.344
8	200	90.889	681.497	23.637	634.050
9	156	78.227	573.912	23.222	542.510
10	156	66.966	487.929	22.152	469.571
11	110	58.172	422.367	21.129	413.198
12	103	51.130	368.901	21.209	363.911
13	83	44.578	321.104	21.032	327.210

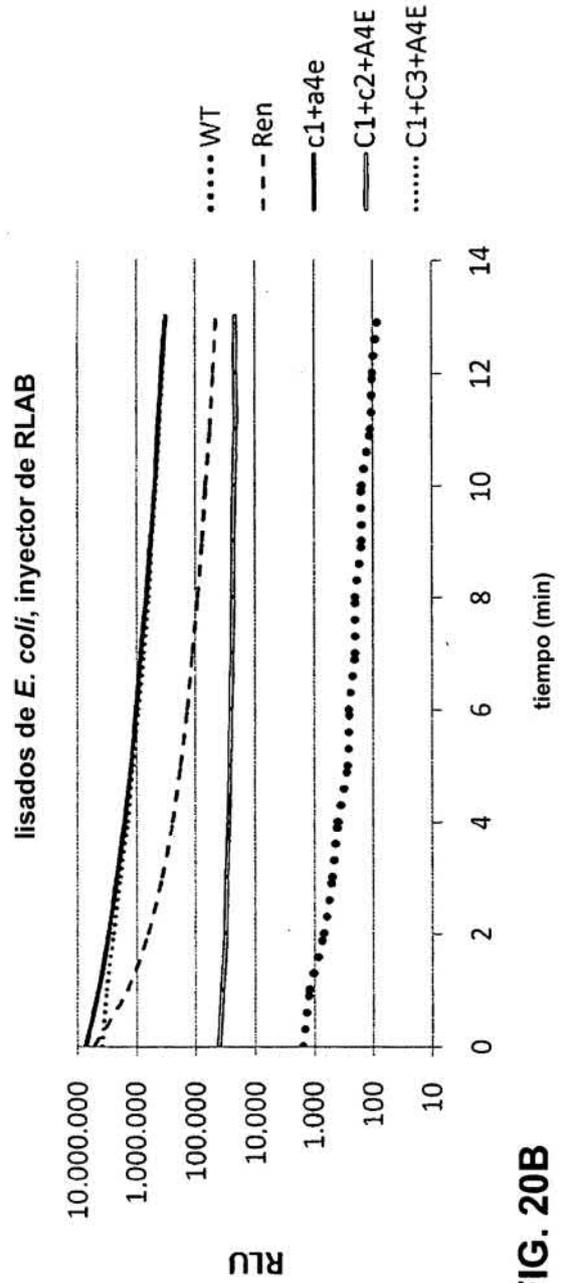


FIG. 20B

Resultados en *E. coli* (tergitol en tampón de ensayo para estabilizar la señal)

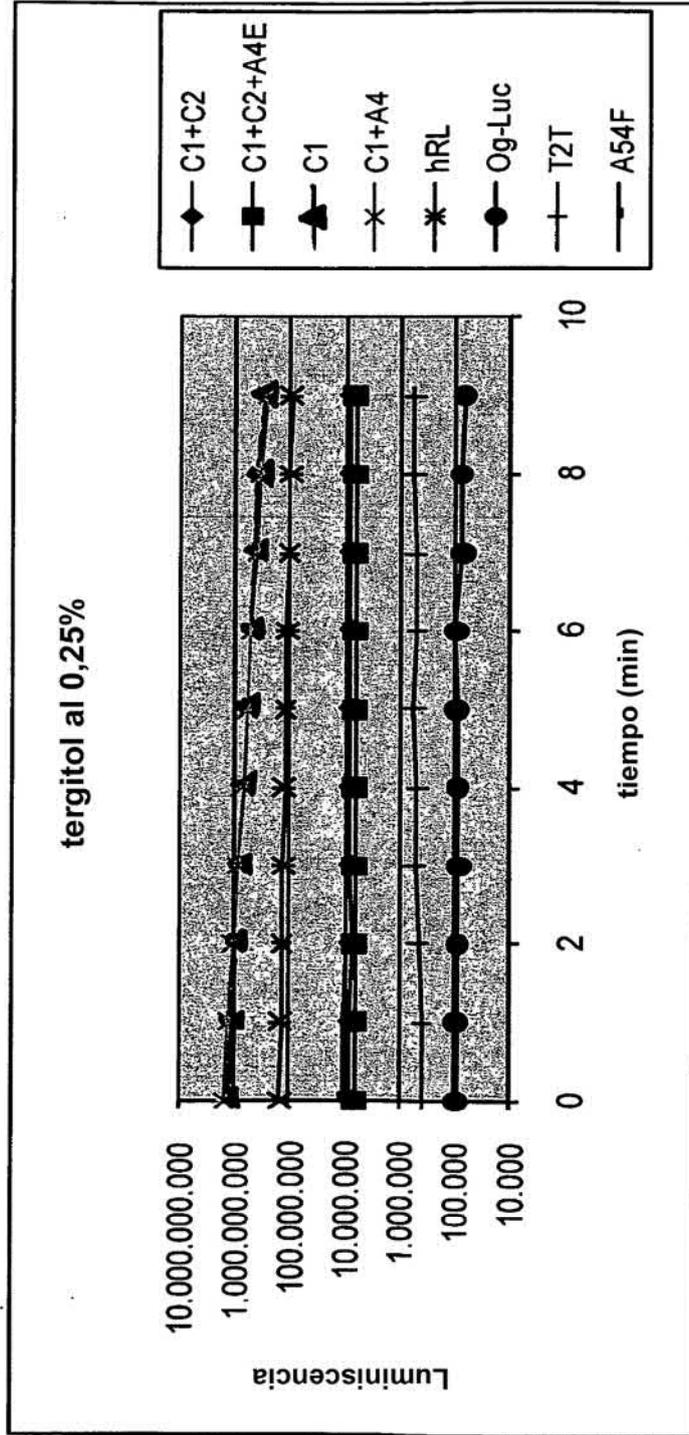


FIG. 21

Compuestos de Og-Luc en células 293

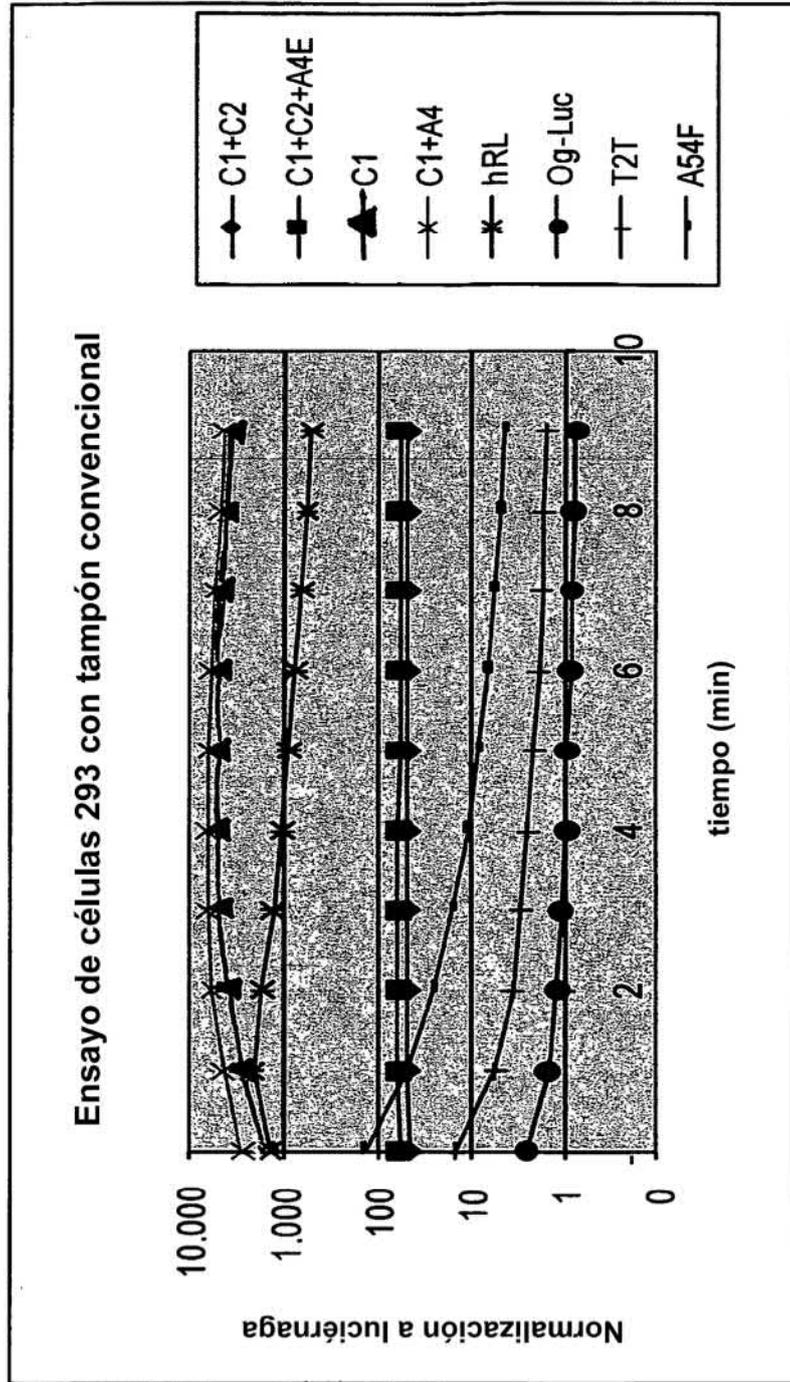


FIG. 22

Células 293 normalizadas a luciérnaga

Células 293 normalizadas a luciérnaga, tergitol al 0,25%

tiempo (min)	C1+C2+A4E	C1+A4	hRL	Og-Luc (wt)
0	2,71	741,28	344,96	0,07
1	2,78	731,25	325,17	0,07
2	2,65	722,31	306,02	0,08
3	2,90	715,06	288,69	0,05
4	2,73	706,89	269,87	0,07
5	2,92	697,22	254,60	0,05
6	2,69	693,01	239,46	0,08
7	2,61	687,04	228,25	0,07
8	2,75	676,30	215,62	0,04
9	2,76	671,85	204,51	0,04

Células 293 normalizadas a luciérnaga, tergitol al 0,25%

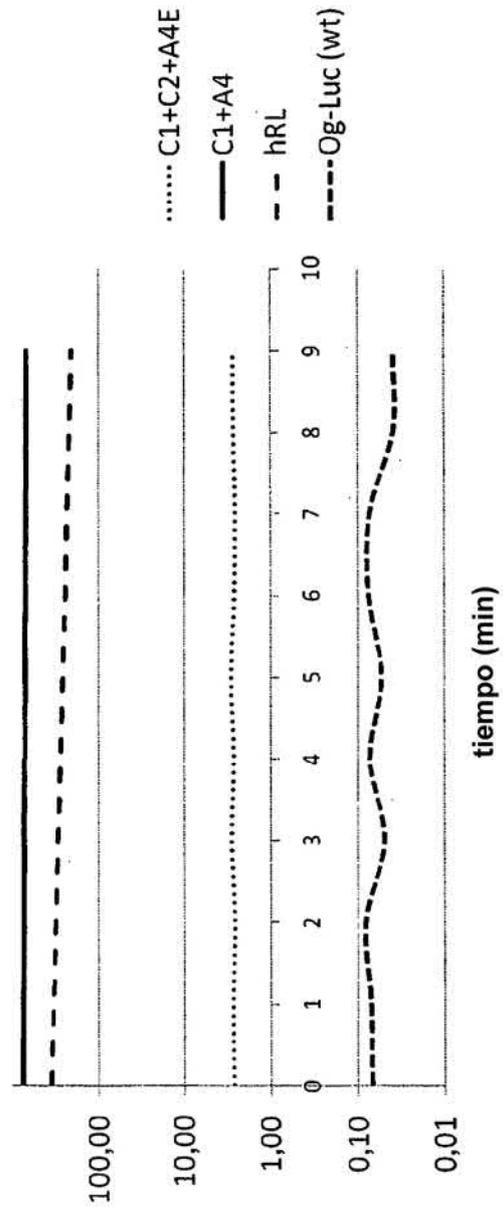


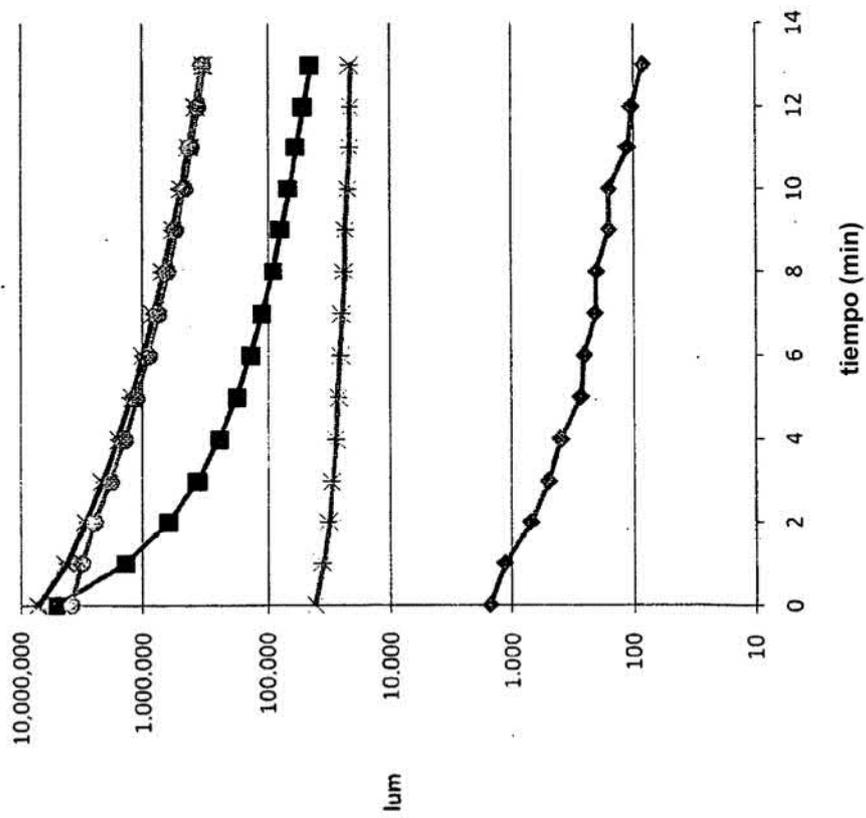
FIG. 23

Estabilidad de los clones de Og-Luc

Temp.	Clon	Semivida
RT (22°C)	WT	1 hora
RT (22°C)	og-luc C1	9,4 horas
RT (22°C)	Renilla	aprox. 24 horas
30°C	N166R	30 horas
30°C	og-luc C1+A4E	Sin descom. en 6 horas
30°C	Renilla	7,9 horas
37°C	N166R	2 min
37°C	og-luc C1+A4E	Sin pérdida en 3 horas
42°C	og-luc C1+A4E	6,4 horas
54°C	og-luc C1	7,1 min
54°C	og-luc C1+A4E	8 (7,1, 7,7) min
54°C	og-luc C1 +C2+A4E	128 min
54°C	og-luc C1 +C3+A4E	23,7 min

FIG. 24

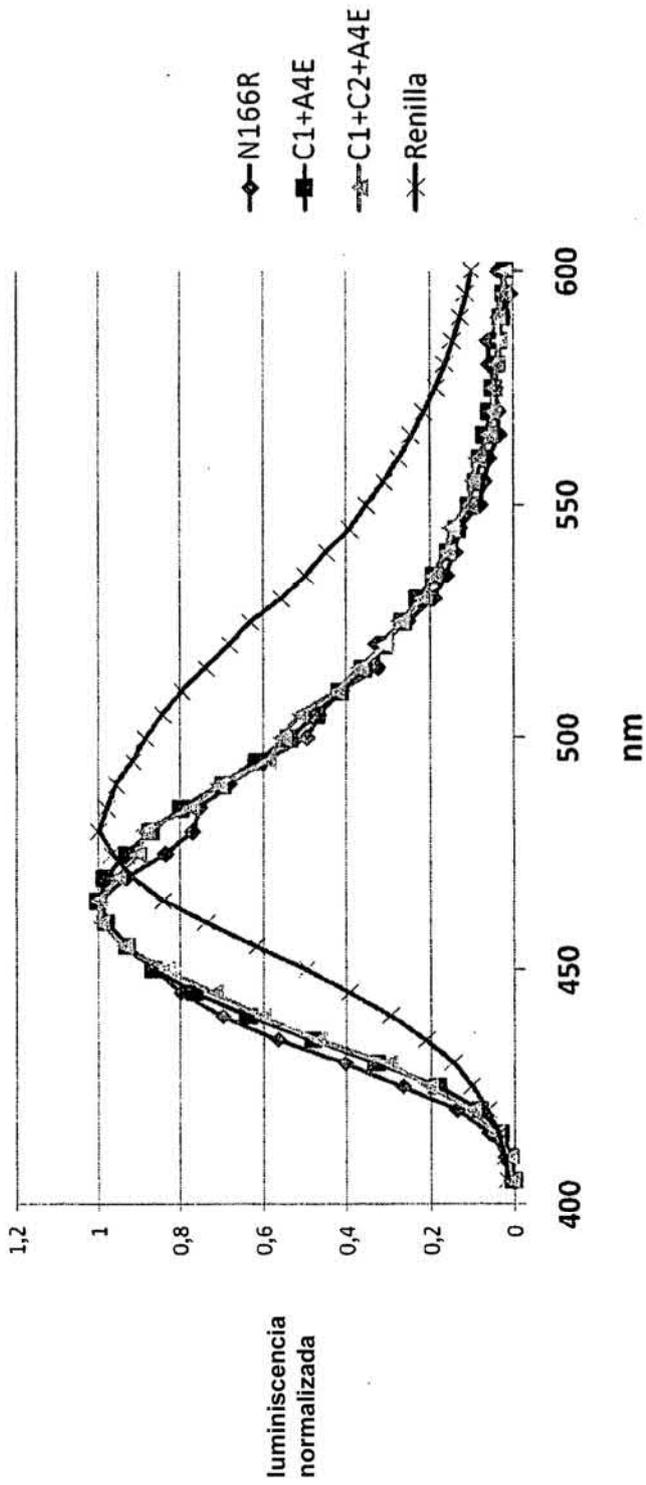
Semivida de la señal o estabilidad de la señal



Muestra	Semivida (min)
WT	1,8
Ren	0,8
c1	1,7
c1+a4e	1,7
C1+c2+A4E	12,6
C1+C3+A4E	3,3

FIG. 25

Espectros de clones de Og-luc



Muestra	Pico espectral
N166R	465
C1+A4E	465
C1+C2+A4E	465
Renilla	480

FIG. 26

Mutaciones: mutagénesis aleatoria de C1+A4E

Mutagénesis aleatoria de C1+A4E	Cambio de AA	n.º de veces frente a C1+A4E
04g7	L72Q	3
10f2	I90T	1,9
16c5	Q20R	4,1
25a1	C164S	3,1
28b2	V79I	2,9
29h7	F54I	10,3
35d11	K89E	2,2

FIG. 27A

Mutaciones: mutagénesis aleatoria de C1+A4E

Cambio de aminoácido	n.º de veces frente a C1+A4E
I90V	1,5
F77W	2

FIG. 27B

Cambios en la posición 92

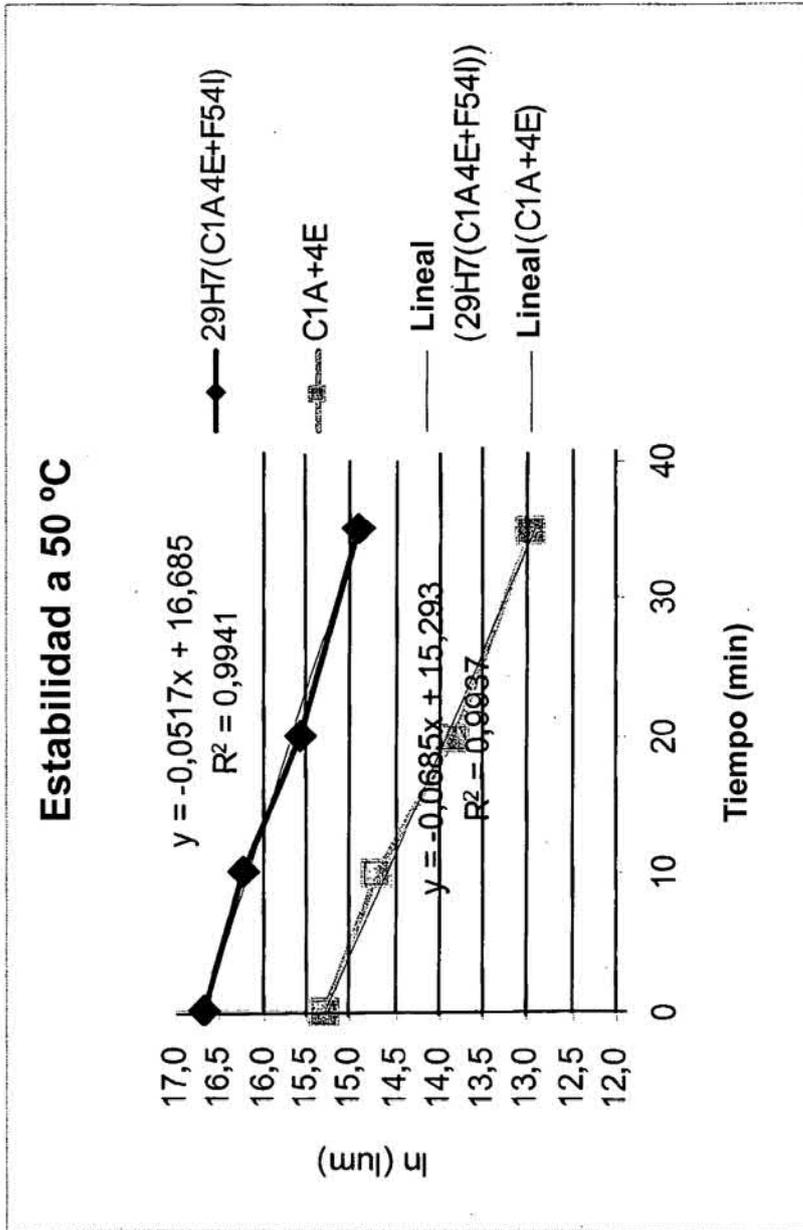
Cambio de aminoácido	n.º de veces frente a C1A4E (coelenteracina 20 uM)
L92G	2,2
L92Q	2
L92S	2,9
L92A	2,5

FIG. 28

Combinaciones que mejoran la emisión de luz

Muestra ID (experimento de combinación)	n.º de veces frente a C1A4E	Cambio de aminoácido con respecto a C1A4E
17A6	19,3	F54I, F68S, I90V
14A12	17,7	F54I, M75K
1G4	17,5	F54I, M75K, I90V
3C6	17,5	F54I, F68S, M75K
14B7	17,5	F54I, I90V

FIG. 29



Muestra	Semivida a 50 °C (min)
C1A4E	10,1
29H7(C1A4E+F54I)	13,3

FIG. 30

Mutación	n.º de veces frente a N166R
T2T	0,6
T2T+N166R	1,8
A4E	0,5
Q11R	0,7
V44I	2,3
A54F	1,5
A54F+N166R	14,3
A54I	5,8
P115E	1,3
P115E+N166R	2,9
Y138I	2,6
Q124K	1,0
Y138C+N166R	2,8
I90V	0,8
A33K	0,6
F68Y	0,7
N166R+A33K	2,0
N166R+F68Y	1,3

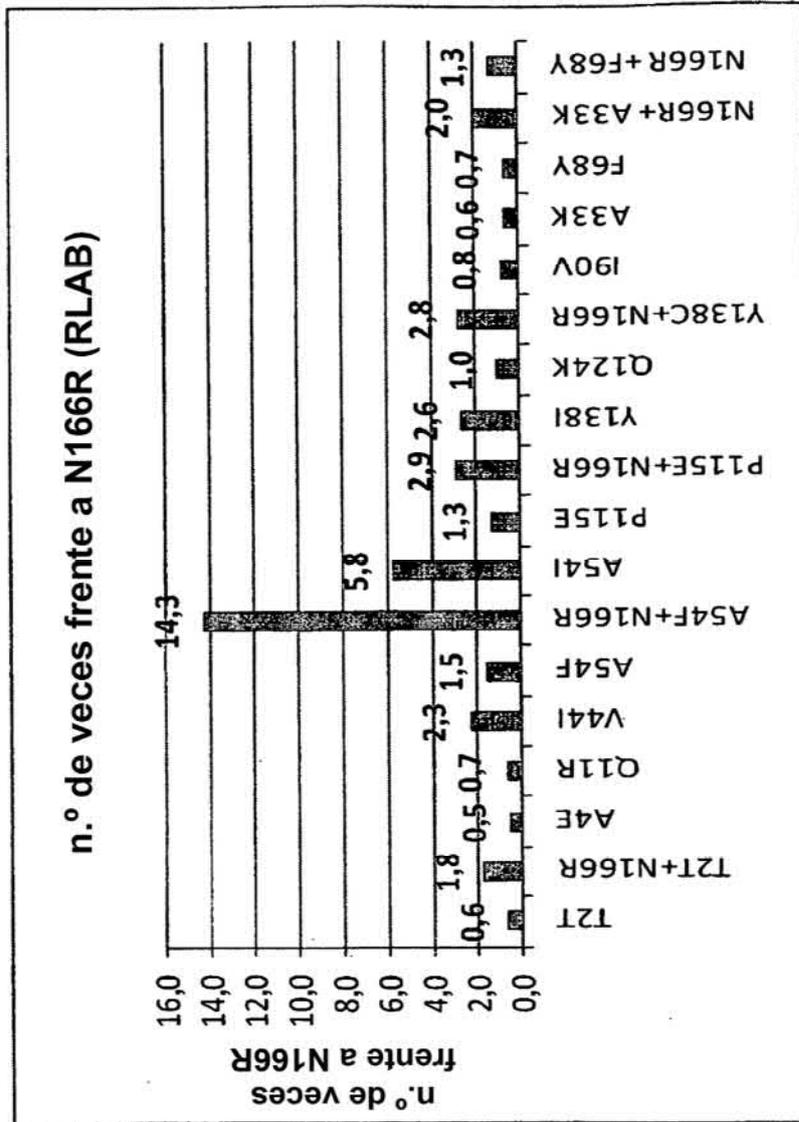


FIG. 33A

Mutación	n.º de veces frente a N166R
T2T	1,6
T2T+N166R	2,3
A4E	1,2
Q11R	1,1
V44I	13,4
A54F	9,7
A54F+N166R	50,2
A54I	97,8
P115E	2,5
P115E+N166R	3,1
Y138I	7,0
Q124K	2,0
Y138C+N166R	3,7
I90V	2,4
A33K	2,6
F68Y	2,7
N166R+ A33K	6,3
N166R +F68Y	3,4

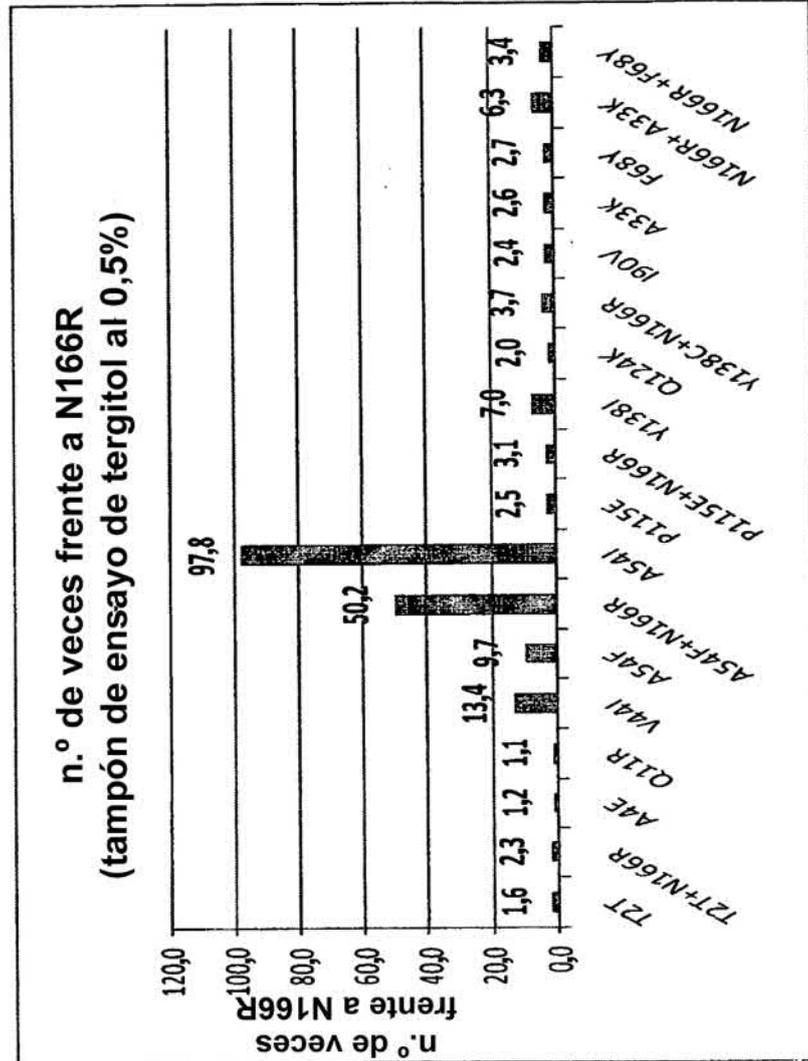


FIG. 33B

Muestra	Estabilidad de la señal, tampón de tergitol al 0,5%, normalizado a N166R
T2T	0,9
T2T+N166R	1,3
A4E	0,9
Q11R	0,8
V44I	0,8
A54F	0,8
A54F+N166R	1,1
A54I	0,7
P115E	1,2
P115E+N166R	1,9
Y138I	0,9
Q124K	1,0
Y138C+N166R	1,5
I90V	0,9
A33K	0,8
F68Y	0,9
N166R/A33K	1,0
N166R/F68Y	1,0

Muestra	Semivida de la señal (min), RLAB, normalizado a N166R
T2T	1,2
T2T+N166R	1,1
A4E	1,4
Q11R	1,1
V44I	1,0
A54F	0,8
A54F+N166R	0,9
A54I	1,3
P115E	1,0
P115E+N166R	1,0
Y138I	1,0
Q124K	1,1
Y138C+N166R	0,6
I90V	1,1
A33K	1,4
F68Y	1,2
N166R/A33K	1,0
N166R/F68Y	1,0

FIG. 33C

Estabilidad a 22 °C, normalizado a N166R

Muestra	Estabilidad (semivida a 22 °C) normalizado a N166R
T2T	0,7
T2T+N166R	1,1
A4E	0,6
Q11R	0,7
V44I	0,8
A54F	0,9
A54F+N166R	2,9
A54I	1,0
P115E	0,8
P115E+N166R	1,9
Y138I	0,9
Q124K	0,7
Y138C+N166R	0,6
I90V	0,7

FIG. 33D

Resumen

Muestra	Semivida de la señal (min), tampón de tergitol al 0,5%, normaliz. a N166R	Semivida de la señal (min), RLAB, normaliz. a N166R	n.º de veces frente a N166R, tampón de ensayo de tergitol al 0,5%	n.º de veces frente a N166R, RLAB	Estabilidad (semivida a 22 °C), normaliz. a N166R
T2T	0,9	1,2	1,6	0,6	0,7
T2T+N166R	1,3	1,1	2,3	1,8	1,1
A4E	0,9	1,4	1,2	0,5	0,6
Q11R	0,8	1,1	1,1	0,7	0,7
V44I	0,8	1,0	13,4	2,3	0,8
A54F	0,8	0,8	9,7	1,5	0,9
A54F+N166R	1,1	0,9	50,2	14,3	2,9
A54I	0,7	1,3	97,8	5,8	1,0
P115E	1,2	1,0	2,5	1,3	0,8
P115E+N166R	1,9	1,0	3,1	2,9	1,9
Y138I	0,9	1,0	7,0	2,6	0,9
Q124K	1,0	1,1	2,0	1,0	0,7
Y138C+N166R	1,5	0,6	3,7	2,8	0,6
I90V	0,9	1,1	2,4	0,8	0,7
A33K	0,8	1,4	2,6	0,6	0,9
F68Y	0,9	1,2	2,7	0,7	0,9
N166R+A33K	1,0	1,0	6,3	2,0	1,1
N166R+F68Y	1,0	1,0	3,4	1,3	1,2

FIG. 33E

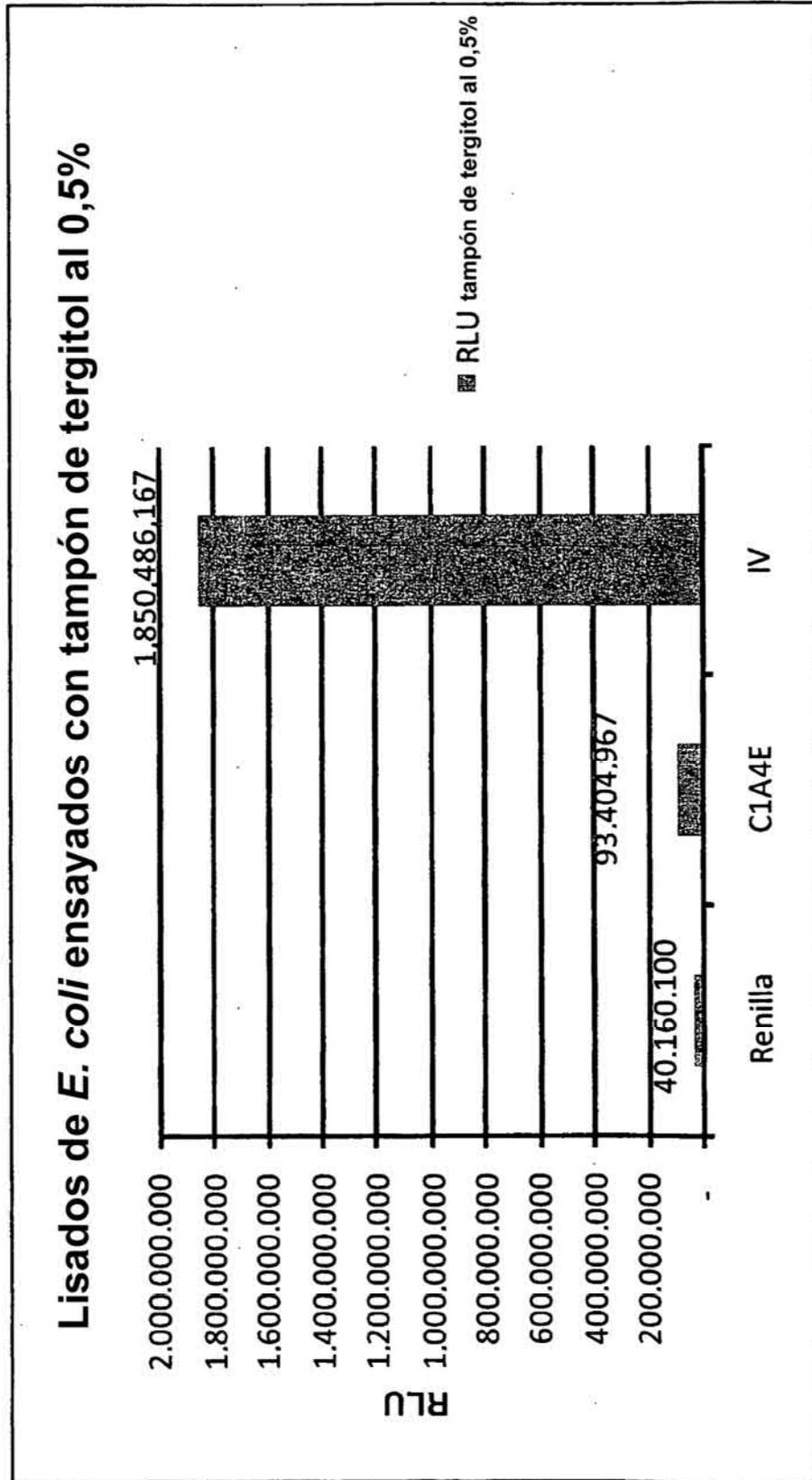
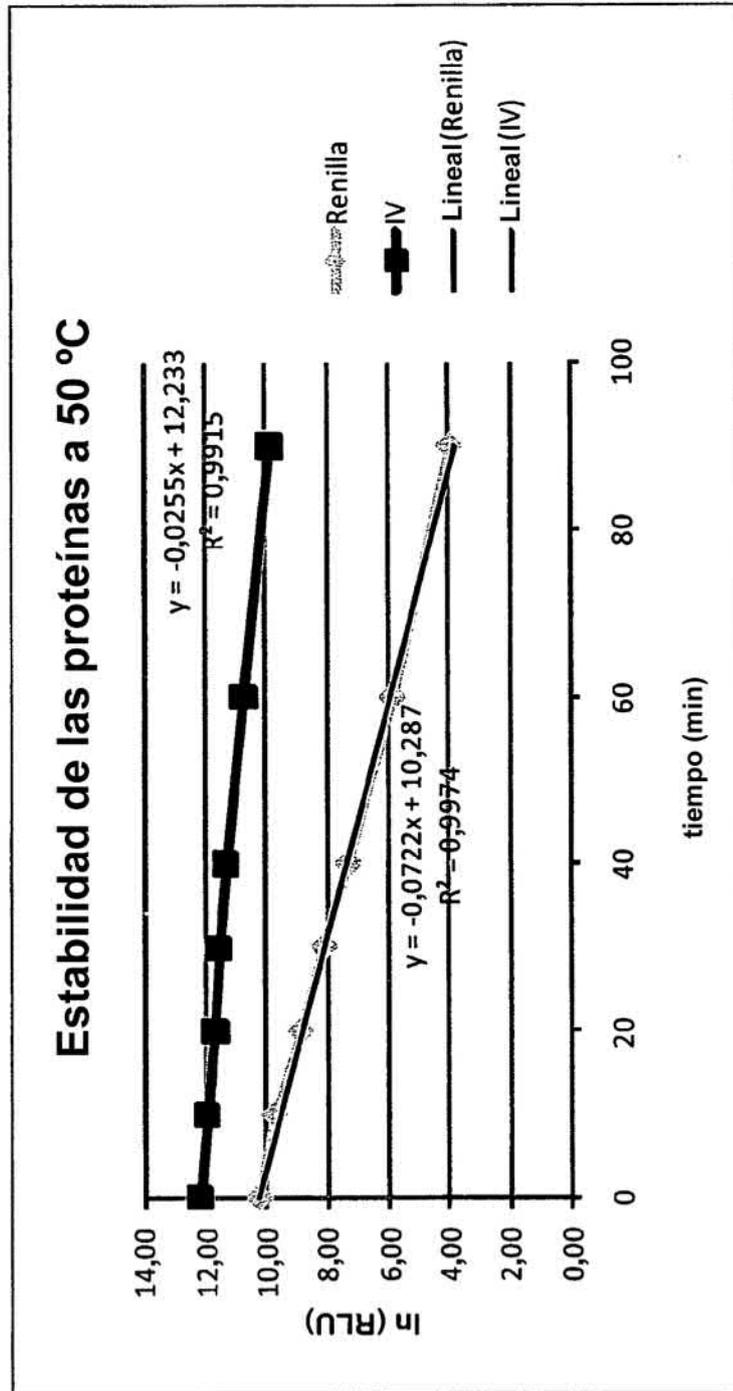


FIG. 34A



Muestra Semivida (min)

Renilla	9,6
IV	27,2

FIG. 34B