

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 399 563**

51 Int. Cl.:

**C07K 14/43** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.10.2006 E 06809023 (2)**

97 Fecha y número de publicación de la concesión europea: **14.11.2012 EP 1954713**

54 Título: **Proteínas verdes fluorescentes modificadas y método para la utilización de las mismas**

30 Prioridad:

**04.11.2005 US 733429 P**

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

**02.04.2013**

73 Titular/es:

**EVROGEN, JSC (100.0%)  
MIKLUKHO-MAKLAYA 16/10  
MOSCOW 117997, RU**

72 Inventor/es:

**LUKYANOV, SERGEY A.**

74 Agente/Representante:

**DURÁN MOYA, Luis Alfonso**

**ES 2 399 563 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**

Proteínas verdes fluorescentes modificadas y método para la utilización de las mismas

**5 Sector de la invención**

La presente invención se refiere, en general, al campo de la biología y la química. De manera más particular, la presente invención se refiere a proteínas fluorescentes.

**10 ANTECEDENTES DE LA INVENCION**

Se descubrió que la proteína verde fluorescente (GFP, por sus siglas en inglés) de la hidromedusa *Aequorea victoria* (sinónimo de *A. A.*), dada a conocer por Johnson y otros en *J Cell Comp Physiol.* (1962), 60:85-104, formaba parte del sistema bioluminiscente de la medusa, en el que la GFP jugaba el papel de emisor secundario transformando la luz azul de la fotoproteína aequorina en luz verde.

El ADNc que codificaba la GFP de *A. victoria* fue clonado por Prasher y otros (*Gene*, 1992, V. 111(2), pág. 229-233). Resultó que este gen se puede expresar manera heterológica prácticamente en cualquier organismo debido a la capacidad única de la GFP para formar un fluoróforo por sí mismo (Chalfie y otros, *Gene* (1992), 111(2):229-233). Este descubrimiento abre amplias perspectivas para la utilización de la GFP en biología celular como marcador fluorescente codificado genéticamente.

Se está realizando mucha investigación para mejorar las propiedades de la GFP y para producir reactivos de GFP útiles y se está optimizando para una serie de objetivos de investigación. Se han desarrollado nuevas versiones de GFP, tales como un ADN de GFP "humanizado", cuyo producto proteico ha incrementado su síntesis en las células de mamíferos (Haas y otros, *Current Biology* 1996, V. 6, pág. 315-324; Yang, y otros, *Nucleic Acids Research* 1996, V. 24, pág. 4592-4593). Una de dichas proteínas humanizadas es la "proteína verde fluorescente mejorada" (EGFP, por sus siglas en inglés). Otras mutaciones en la GFP han dado lugar a versiones que emiten luz azul, cian y verde amarillo. Además, se han obtenido variantes de GFP con plegamiento mejorado y fluorescencia celular bajo incubación a 37°C. Los mutantes de GFP de *A. victoria* útiles se dan a conocer en detalle en las patentes de Estados Unidos 5.491.084, 5.625.048, 5.777.079, 5.804.387, 6.090.919, 5.874.304, 5.968.750, 6.020.192, 6.027.881, 6.046.925, 6.054.321, 6.066.476, 6.096.865, 6.146.826, 6.414.119, 6.638.732, 6.699.687, 6.803.188, 6.077.707, 6.124.128, 6.172.188, 6.818.443, 6.194.548, 6.265.548, 6.319.669, 6.403.374, 6.593.135, 6.800.733, 6.780.975, 6.852.849, y 6.919.186.

Se aislaron homólogos de GFP de diferentes especies, incluyendo Anthozoa y Arthropoda, (Matz y otros, *Nature Biotechnol.* 1999, V. 17, pág. 969-973; Shagin y otros, *Mol Biol Evol.* 2004, V. 21(5), pág. 841-850). Un conjunto de aplicaciones biológicas y biomédicas de estas proteínas se dan a conocer en detalle por Lippincott-Schwartz y Patterson en *Science*, 2003, V. 300(5616), pág. 87-91. Además, se aislaron homólogos próximos de la GFP de *A. victoria* de otras medusas del género *Aequorea*, incluyendo la proteína verde fluorescente de *A. macrodactyla*, GFPxm (Xia y otros, *Mar Biotechnol* 2002, V. 4(2), pág. 155-62) y la proteína de tipo GFP de *A. coerulea*, AcGFPL (Gurskaya y otros, *Biochem J.* (2003), 373(Pt 2): 403-408).

La GFPxm de *A. macrodactyla* comparte el 83% de identidad con la GFP de *A. victoria*. La GFPxm de tipo natural no es útil como marcador fluorescente en ensayos de base celular debido a una velocidad de maduración baja a 37°C. La modificación de GFPxm para optimizar su velocidad de maduración a temperaturas de 35-39°C proporciona un medio para detectar el informador en células de mamífero a niveles inferiores de expresión y/o la sensibilidad incrementada en relación con la GFPxm de tipo natural. Esto mejora ampliamente la utilidad de la GFPxm en el estudio de funciones celulares en células vivas.

El documento CN 1438239 se refiere a una nueva proteína fluorescente y también da a conocer la secuencia de nucleótidos que codifica la proteína, la cual mejora el polipéptido de la proteína verde fluorescente natural de medusa grande "multibarrel" para obtener la proteína fluorescente, cuyo espectro de excitación y de emisión es obviamente diferente del de la proteína natural.

**55 CARACTERÍSTICAS DE LA INVENCION**

La presente invención da a conocer proteínas fluorescentes funcionales modificadas con una velocidad de maduración incrementada a una temperatura de 20°C o superior en comparación con la proteína verde fluorescente de *A. Macrodactyla* (GFPxm) de tipo natural, en las que dichas proteínas fluorescentes funcionales modificadas tienen, como mínimo, el 95% de identidad con la secuencia de aminoácidos seleccionada del grupo que comprende las SEQ ID 18, 20, 22, y 24 y difiere de la secuencia de aminoácidos de la proteína verde fluorescente de *A. macrodactyla* (GFPxm) (SEQ ID NO: 2) en una sustitución de aminoácido F220L.

En una realización preferente, la presente invención da a conocer una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica una proteína fluorescente funcional, cuya secuencia de aminoácidos es similar, de manera sustancial, a la secuencia de aminoácidos de la proteína verde fluorescente de *A. Macrodactyla* (GFPxm) (SEQ ID NO: 2) y difiere de la SEQ ID NO:2, como mínimo, en una sustitución de aminoácido F220L. Dicha proteína fluorescente funcional presenta una velocidad de maduración incrementada a una temperatura de 20°C o superior en comparación con la GFPxm.

Además, la presente invención se refiere a un ácido nucleico aislado, que comprende una secuencia de ácidos nucleicos que codifica una proteína fluorescente modificada genéticamente, cuya la secuencia de aminoácidos tiene, como mínimo, el 80% de identidad de la secuencia con la secuencia de aminoácidos de la SEQ ID NO: 2, y cuya secuencia de aminoácidos difiere de la secuencia de aminoácidos de la SEQ ID NO: 2, como mínimo, en una sustitución de aminoácido F220L, y comprende una secuencia de aminoácidos que tiene, como mínimo, el 95% de identidad con la secuencia de aminoácidos seleccionada del grupo que comprende las SEQ ID 18, 20, 22, y 24.

En una realización preferente, una molécula de ácido nucleico de la presente invención codifica una proteína fluorescente que también comprende sustituciones adicionales de aminoácidos seleccionadas del grupo que comprende K3G, E6D, T9A, P58T, F99L, F99H, M128K, M128E, I136M, Y151H, N144S, K162E, K156M, T214A, G228C, G228S, y K238R, en la que dicha proteína fluorescente funcional tiene una velocidad de maduración incrementada a una temperatura de 20°C o superior e n comparación con la GFPxm de *A. macrodactyla* de tipo natural.

En realizaciones preferentes, una molécula de ácido nucleico de la presente invención codifica una proteína fluorescente funcional que es similar, de manera sustancial, con la secuencia de aminoácidos de GFPxm y comprende una o más sustituciones adicionales de aminoácidos que alteran sus propiedades fluorescentes y/o optimizan el plegamiento, tal como se muestra, por ejemplo, en las SEQ ID NOs: 18-24.

En otra realización preferente, la presente invención da a conocer una proteína fluorescente funcional mutante, cuya secuencia de aminoácidos es similar, de manera sustancial, con la secuencia de aminoácidos de GFPxm de *A. macrodactyla* GFPxm (SEQ ID NO: 2) y que difiere de la SEQ ID NO: 2, como mínimo, en una sustitución de aminoácido F220L. Dicha proteína fluorescente funcional mutante tiene una velocidad de maduración mejorada a una temperatura de 20°C o superior en comparación con la GFPxm. También se dan a conocer ejemplos de proteínas fluorescentes mutantes que tienen composiciones de aminoácidos seleccionadas del grupo que comprende las SEQ ID NO 4-24, en los que dichas proteínas fluorescentes mutantes tienen una velocidad de maduración mejorada a una temperatura de 20°C o superior en comparación con la GFPxm.

En otras realizaciones, se dan a conocer vectores que comprenden un ácido nucleico de la presente invención. Además, la presente invención da a conocer un casete de expresión que comprende un ácido nucleico de la presente invención y elementos reguladores necesarios para la expresión del ácido nucleico en la célula, de manera más específica una región de inicio de la transcripción que es funcional en un huésped de expresión y una región de terminación de la transcripción funcional en dicho huésped de expresión.

De manera adicional, se dan a conocer células huésped, líneas celulares estables, animales transgénicos y plantas transgénicas que comprende ácidos nucleicos, vectores o casetes de expresión.

De manera adicional, se dan a conocer kits que comprenden ácidos nucleicos o vectores o casetes de expresión que comprenden dichos ácidos nucleicos, o la proteína de la presente invención.

### **DESCRIPCIÓN BREVE DE LOS DIBUJOS**

La figura 1 muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de la proteína fluorescente GFPxm.

La figura 2 muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de la proteína fluorescente Mut 2.

La figura 3 muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de la proteína fluorescente Mut-g9.

La figura 4 muestra el brillo relativo de colonias de *E. coli* que expresan las proteínas fluorescentes GFPxm, Mut 2, o Mut-g9 después del crecimiento a diferentes temperaturas. Las condiciones de temperatura y el tiempo de incubación se indican en la base del histograma. Todos los datos se normalizan al brillo de las colonias que expresan Mut-g9 después de 36 horas de crecimiento a 20°C.

La figura 5 muestra curvas de la fluorescencia del crecimiento de colonias de *E. coli* que expresan GFPxm (línea 1), Mut 2 (línea 2) o Mut-g9 (línea 3) durante 6 horas después de la inducción.

5 La figura 6 muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de tagGFP.

La figura 7A muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de tagCFP.

La figura 7B muestra los espectros normalizados de excitación (línea 1) y emisión (línea 2) de tagYFP1.

10

### **DESCRIPCIÓN DETALLADA**

Tal como se utiliza en el presente documento, el término "proteína fluorescente" significa una proteína que es fluorescente; por ejemplo, puede mostrar una fluorescencia baja, media o intensa después de la radiación con luz de una longitud de onda de excitación apropiada. La característica fluorescente de la proteína fluorescente es aquella que aparece del fluoróforo, en la que el fluoróforo es el resultado de la ciclación autocatalítica de dos o más residuos de aminoácidos en el esqueleto del polipéptido. Por lo tanto, las proteínas fluorescentes de la presente invención no incluyen proteínas que muestran fluorescencia sólo a partir de los residuos que actúan por sí mismos como fluorescentes intrínsecos, es decir, el triptófano, la tirosina y la fenilalanina.

15

20

Tal como se utiliza en el presente documento, "propiedad fluorescente" se refiere al coeficiente de extinción molar a una longitud de onda de excitación apropiada, la eficacia cuántica de la fluorescencia, la forma del espectro de excitación o del espectro de emisión, la longitud de onda máxima de excitación y la longitud de onda máxima de emisión, la proporción de las amplitudes de excitación a dos longitudes de onda diferentes, la proporción de las amplitudes de emisión a dos longitudes de onda diferentes, el tiempo de vida del estado excitado o la anisotropía de fluorescencia. Una diferencia medible en cualquiera de estas propiedades entre la GFPxm de tipo natural y la forma mutante es útil. Se puede determinar una diferencia medible como la cantidad de cualquier propiedad fluorescente cuantitativa, por ejemplo, la cantidad de fluorescencia a una longitud de onda particular o la integral de fluorescencia sobre el espectro de emisión.

25

30

Tal como se utiliza en el presente documento, "tasa de maduración" o "velocidad de maduración" se refiere a la velocidad de formación de proteínas fluorescentes maduras (es decir, una proteína fluorescente capaz de producir fluorescencia) después de la traducción. La velocidad de maduración se puede caracterizar con el semitiempo de maduración. Se ha descubierto que la maduración de proteína fluorescente incluye dos etapas: (i) el plegamiento de proteínas que significa la formación de una proteína de barril beta con una hélice alfa central que contiene los aminoácidos que formarán el cromóforo. Esta etapa se caracteriza habitualmente con una constante de velocidad de aproximadamente  $10^{(-2)}s^{(-1)}$  o un semitiempo de varios segundos a decenas de segundos; (ii) la maduración del cromóforo que es la ciclación y deshidratación del esqueleto de la proteína. Esta etapa se caracteriza habitualmente con una constante de velocidad de aproximadamente  $10^{(-4)}s^{(-1)}$  o un semitiempo de aproximadamente varios minutos. Por lo tanto, esta etapa más lenta es la etapa limitante en la maduración de la proteína verde fluorescente (Reid BG, Flynn GC. *Biochemistry*. 1997 V. 36(22), pág. 6786-6791).

35

40

Tal como se utiliza en el presente documento, el término "GFP" se refiere a la proteína verde fluorescente de *A. victoria*, que incluye versiones del estado de la técnica anterior de GFP modificada para proporcionar una mayor fluorescencia o una fluorescencia en diferentes colores. La secuencia de la GFP de tipo natural se ha dado a conocer en Prasher y otros, *Gene* 111 (1992), 229-33.

45

Tal como se utiliza en el presente documento, el término "GFPxm" se refiere a la proteína verde fluorescente de tipo natural de *A. macrodactyla*.

50

Tal como se utiliza en el presente documento, el término "aislada" significa una molécula o una célula que está en un medio diferente del que se encuentra la molécula o la célula de forma natural.

La referencia a una secuencia de nucleótidos "que codifica" un polipéptido significa que la secuencia, después de la transcripción y traducción del ARNm, produce el polipéptido. Esto incluye tanto la hebra codificante, cuya secuencia de nucleótidos es idéntica al ARNm y cuya secuencia se dispone normalmente en el listado de secuencias, como su hebra complementaria, que se utiliza como la plantilla para la transcripción. Tal como cualquier experto en la materia sabe, esto también incluye todas las secuencias de nucleótidos degeneradas que codifican la misma secuencia de aminoácidos. Las secuencias de nucleótidos que codifica un polipéptido incluyen secuencias que contienen intrones.

55

60

Tal como se utiliza en el presente documento, el término "mutante" se refiere a una proteína dada a conocer en la presente invención, en la que se añaden y/o sustituyen y/o se eliminan y/o se insertan uno o más aminoácidos en el extremo N-terminal y/o el extremo C-terminal, y/o en las secuencias de aminoácidos nativas de las proteínas de la

presente invención. Tal como se utiliza en el presente documento, el término "mutante" se refiere a una molécula de ácidos nucleicos que codifica una proteína mutante. Además, el término "mutante" se refiere a cualquier versión más corta o más larga de la proteína o el ácido nucleico del presente documento.

5 Tal como se utiliza en el presente documento, "homólogo u homología" es un término utilizado en la técnica para describir el parentesco de una secuencia de nucleótidos o de péptidos con otra secuencia de nucleótidos o de péptidos, que se determina mediante el grado de identidad y/o la similitud entre dichas secuencias comparadas.

10 Tal como se utiliza en el presente documento, una secuencia de aminoácidos o una secuencia de nucleótidos es "idéntica, de manera sustancial" a una secuencia de referencia si la secuencia de aminoácidos o la secuencia de nucleótidos tiene, como mínimo, el 95% de identidad de la secuencia (por ejemplo, el 95%, el 97%, el 98%, el 99% o el 100% de identidad de la secuencia) con la secuencia de referencia sobre una ventana de comparación determinada. Tal como se utiliza en el presente documento, una secuencia de aminoácidos o una secuencia de nucleótidos es "similar, de manera sustancial" a una secuencia de referencia si la secuencia de aminoácidos o la  
 15 secuencia de nucleótidos tiene, como mínimo, el 80% de identidad de la secuencia (por ejemplo, el 80%, el 85%, el 90%, el 95%, el 97%, el 98%, el 99% ó el 100% de identidad de la secuencia) con la secuencia de referencia sobre una ventana de comparación determinada. La identidad de la secuencia se calcula en base a una secuencia de referencia. Los algoritmos para el análisis de secuencia son conocidos en la técnica, tales como BLAST, dado a conocer en Altschul y otros, J. Mol. Biol., 215, pág. 403-10 (1990).

20 Tal como se resume anteriormente, la presente invención se refiere a moléculas de ácidos nucleicos que comprenden secuencias de nucleótidos que codifican proteínas fluorescentes mutantes, así como proteínas codificadas por estos ácidos nucleicos. Las proteínas de interés son idénticas, de manera sustancial, con la proteína verde fluorescente de tipo natural de *A. macrodactyla* GFPxm (SEQ ID NO: 2) y comprenden, como mínimo, una  
 25 sustitución de aminoácido F220L. Dichos mutantes son proteínas fluorescentes funcionales que tienen una velocidad de maduración mejorada a una temperatura de 20°C o superior en comparación con GFPxm.

30 En una realización, dicho mutante comprende sólo una sustitución F220L. Los inventores de la presente invención han descubierto que la sustitución F220L da lugar a un incremento medible de la velocidad de maduración de la GFPxm a una temperatura de 20°C o superior en comparación con la GFPxm de tipo natural. Los inventores de la presente invención han descubierto adicionalmente que la sustitución F220L altera las propiedades fluorescentes de la proteína en comparación con la GFPxm de *A. macrodactyla*.

35 En otra realización preferente, dicho mutante también comprende sustituciones adicionales de aminoácidos que incrementan de manera adicional la velocidad de maduración de la proteína a una temperatura de 20°C o superior, por ejemplo, se da a conocer un mutante que tiene una secuencia de aminoácidos seleccionada del grupo que comprende las SEQ ID NO 6, 8, 10, 12, 14, 16, y 18.

40 Las mutaciones indicadas anteriormente en GFPxm se pueden combinar con mutaciones que incrementan de manera adicional el plegamiento, reducen la oligomerización o influyen en las propiedades espectrales de la GFPxm y sus mutantes, tal como se muestra, por ejemplo, en las SEQ ID NO: 18-24.

45 En otras realizaciones, se dan a conocer vectores que comprenden un ácido nucleico de la presente invención. Además, la presente invención da a conocer un casete de expresión que comprende un ácido nucleico de la presente invención y elementos reguladores necesarios para la expresión del ácido nucleico en la célula.

50 También son de interés proteínas y ácidos nucleicos que son similares, de manera sustancial, a las proteínas y ácidos nucleicos específicos referenciados anteriormente, o derivados u homólogos o mutantes de los mismos. Además, se dan a conocer células huésped, líneas celulares estables y organismos transgénicos que comprenden moléculas de ácido nucleico referenciados anteriormente. Las composiciones de proteínas y ácidos nucleicos de la presente invención son útiles en un conjunto de aplicaciones y métodos diferentes, en particular aplicaciones de marcaje de células y proteínas. Finalmente, se dan a conocer kits para la utilización en dichos métodos y aplicaciones.

55 Moléculas de ácido nucleico

La presente invención da a conocer moléculas de ácido nucleico que comprenden secuencias de nucleótidos que codifican proteínas fluorescentes mutantes que son idénticas, de manera sustancial, a la proteína verde fluorescente de *A. Macroductyla* GFPxm de tipo natural (SEQ ID NO: 2) y comprenden, como mínimo, una sustitución de  
 60 aminoácido F220L.

Una molécula de ácido nucleico, tal como se utiliza en el presente documento, es una molécula de ADN, tal como moléculas de ADN genómico o moléculas de ADNc, o una molécula de ARN, tal como moléculas de ARNm.

En particular, dichas moléculas de ácido nucleico son moléculas de ADN que comprenden un marco de lectura abierto que codifica una proteína fluorescente de la invención. Los ácidos nucleicos de la presente invención están presentes en un medio diferente a su medio natural; por ejemplo, están aislados, presentes en cantidades enriquecidas, o están presentes o se expresan *in vitro* o en una célula u organismo diferente de su medio natural. En una realización preferente, las moléculas de ácido nucleico de la presente invención se modifican, es decir, se obtienen a partir de una proteína natural, por ejemplo, la proteína verde fluorescente de *A. Macrodactyla* GFP<sub>xm</sub> de tipo natural, mediante modificaciones.

Las modificaciones, así como las adiciones o eliminaciones, se pueden introducir mediante cualquier método conocido en la técnica (véase, por ejemplo, Gustin y otros, *Biotechniques* (1993) 14: 22; Barany, *Gene* (1985) 37: 111-123; y Colicelli y otros, *Mol. Gen. Genet.* (1985) 199:537-539, Sambrook y otros, *Molecular Cloning: A Laboratory Manual*, (1989), CSH Press, pág. 15.3-15.108), que incluye la PCR propensa a errores, transposiciones ("shuffling"), mutagénesis dirigida por oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis de casete, mutagénesis de conjunto recursiva, mutagénesis de conjunto exponencial, mutagénesis dirigida a sitio, mutagénesis aleatoria, reensamblaje de genes, mutagénesis saturada en sitios de genes (GSSM), reensamblaje de unión sintética (SLR), o una combinación de los mismos. Las modificaciones, adiciones o eliminaciones también se pueden introducir mediante un método que comprende la recombinación, la recombinación de secuencia recursiva, mutagénesis de ADN modificada con fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis de doble cadena espaciada, mutagénesis de reparación en desapareamientos puntuales, mutagénesis de cepa huésped deficiente en la reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis por eliminación, mutagénesis por restricción-selección, mutagénesis por restricción-purificación, síntesis artificial de genes, mutagénesis de conjunto, creación de multímeros de ácidos nucleicos quiméricos o una combinación de los mismos.

Las moléculas de ácido nucleico específicas de interés comprenden secuencias de nucleótidos que codifican las siguientes proteínas fluorescentes: Mut 2 (SEQ ID NO 4); Mut 235 (SEQ ID NO 6); Mut 235-1 (SEQ ID NO 8); Mut 235-2 (SEQ ID NO 10); Mut 235-4 (SEQ ID NO 12); Mut-g9 (SEQ ID NO 14); Mut 235-4G6 (SEQ ID NO 16). También son de interés las moléculas de ácido nucleico que comprenden secuencias de ácidos nucleico que codifican mutantes Mut-g9, tagGFP (también denominada macGFP, SEQ ID NO: 18), tagCFP (SEQ ID NO:20), tagYFP1 (SEQ ID NO: 22) y tagYFP2 (SEQ ID NO:24), en las que las propiedades fluorescentes de estos mutantes están alteradas en comparación con la proteína Mut-g9.

En las SEQ ID NO 3-23 se muestran ejemplos de secuencias de nucleótidos que codifican las proteínas anteriores.

Cada una de estos tipos particulares de moléculas de ácido nucleico de interés se dan a conocer con más detalle de forma individual en la sección posterior de "Ejemplos".

También se dan a conocer ácidos nucleicos que se hibridan a los ácidos nucleicos descritos anteriormente bajo condiciones severas, de manera preferente bajo condiciones de severidad elevada (es decir, complementos de los ácidos nucleicos descritos previamente). Un ejemplo de condiciones severas es la hibridación a 50°C o superior y 0,1xSSC (cloruro de sodio 15 mM/citrato de sodio 1,5 mM). Otro ejemplo de condiciones de hibridación de severidad elevada es la incubación durante una noche a 42°C en una solución de formamida al 50%, 5xSSC (NaCl 150 mM, citrato de trisodio 15 mM), fosfato de sodio 50 mM (pH 7,6), 5 x solución de Denhardt, sulfato de dextrano al 10%, y ADN de esperma de salmón fragmentado desnaturalizado 20 µg/ml, seguido de lavado en 0,1xSSC a aproximadamente 65°C. En la técnica se conocen otras condiciones de hibridación de severidad elevada y también se pueden utilizar para identificar ácidos nucleicos de la presente invención.

Además, también se dan a conocer variantes degeneradas de los ácidos nucleicos que codifican las proteínas de la presente invención. Las variantes degeneradas de ácidos nucleicos comprenden sustituciones de los codones del ácido nucleico con otros codones que codifican los mismos aminoácidos. En particular, las variantes degeneradas de los ácidos nucleicos se generan para incrementar su expresión en una célula huésped. En esta realización, los codones del ácido nucleico que no son preferentes o son menos preferentes en los genes en la célula huésped se sustituyen con los codones sobrerrepresentados en las secuencias codificantes en genes en la célula huésped, en las que dichos codones sustituidos codifican el mismo aminoácido. En una realización preferente, los ácidos nucleicos de la presente invención están humanizados. Tal como se utiliza en el presente documento, el término "humanizado" se refiere a cambios realizados a la secuencia de ácidos nucleicos para optimizar los codones para la expresión de la proteína en células de mamífero (humano) (Yang y otros, *Nucleic Acids Research* (1996) 24: 4592-4593). Véase también la patente de Estados Unidos No. 5.795.737 que da a conocer la humanización de proteínas.

Los ácidos nucleicos de la presente invención, los ADNc correspondientes, los genes completos y las construcciones se pueden generar de forma sintética mediante un conjunto de protocolos diferentes conocidos por los expertos en la

materia. Las construcciones adecuadas de ácidos nucleicos se purifican utilizando técnicas estándar de ADN recombinante tal como se dan a conocer en, por ejemplo, Sambrook y otros, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, y bajo las regulaciones descritas en, por ejemplo, el Departamento de Salud y Servicios Humanos de Estados Unidos, Instituto Nacional de Salud (NIH), Directrices para la Investigación de ADN Recombinante.

Se ha descubierto que las proteínas fluorescentes se pueden fusionar genéticamente a otras proteínas diana y se pueden utilizar como marcadores para identificar la localización y la cantidad de la proteína diana producida. Por consiguiente, la presente invención da a conocer ácidos nucleicos que codifican proteínas de fusión que comprenden una proteína fluorescente y secuencias de aminoácidos adicionales. Dichas secuencias pueden tener, por ejemplo, hasta aproximadamente 15, hasta aproximadamente 100, hasta aproximadamente 200 o hasta aproximadamente 1000 aminoácidos de largo. Las proteínas de fusión poseen la capacidad de producir fluorescencia que se determina mediante una parte de la proteína fluorescente.

También se dan a conocer construcciones de vectores y otros ácidos nucleicos que comprenden los ácidos nucleicos de la presente invención. Entre los vectores adecuados se incluyen vectores virales y no virales, plásmidos, cósmidos, fagos, etc., de manera preferente plásmidos, y se utilizan para la clonación, amplificación, expresión, transferencia, etc., de la secuencia de ácidos nucleicos de la presente invención en el huésped apropiado. La elección del vector apropiado es bien conocida en la técnica y existen muchos de dichos vectores disponibles comercialmente. Para preparar las construcciones, el ácido nucleico parcial o completo se inserta en un vector habitualmente mediante la unión de la ADN ligasa a un sitio cortado por enzimas de restricción en el vector. De manera alternativa, la secuencia de nucleótidos deseada se puede insertar mediante recombinación homóloga *in vivo*, habitualmente mediante la unión de regiones de homología al vector en los flancos de la secuencia de nucleótidos deseada. Las regiones de homología se añaden mediante la unión de oligonucleótidos o mediante la reacción en cadena de la polimerasa utilizando cebadores que comprenden tanto la región de homología como una parte de la secuencia de nucleótidos deseada, por ejemplo.

También se dan a conocer casetes de expresión o sistemas utilizados, entre otras cosas, para la producción de las proteínas fluorescentes de la presente invención o proteínas de fusión de las mismas o para la replicación de las moléculas de ácido nucleico de la presente invención. El casete de expresión puede existir como un elemento extracromosómico o puede estar incorporado en el genoma de la célula como resultado de la introducción de dicho casete de expresión en la célula. Para la expresión, el producto génico codificado por el ácido nucleico de la presente invención se expresa en cualquier sistema de expresión conveniente, que incluye, por ejemplo, sistemas bacterianos, de levadura, de insectos, de anfibios o de mamíferos. En el vector de expresión, un ácido nucleico de la presente invención está unido de manera operativa a una secuencia reguladora que puede incluir promotores, potenciadores, terminadores, operadores, represores e inductores. Los métodos para la preparación de casetes o sistemas de expresión capaces de expresar el producto deseado son conocidos para un experto en la materia.

Las líneas celulares, que expresan de manera estable las proteínas de la presente invención, se pueden seleccionar mediante los métodos conocidos en la técnica (por ejemplo, la cotransfección con un marcador seleccionable, tal como dhfr, gpt, neomicina o higromicina, permite la identificación y el aislamiento de las células transfectadas que contienen el gen incorporado en el genoma).

Los sistemas de expresión descritos anteriormente se pueden utilizar en huéspedes procariotas o eucariotas. Las células huésped, tales como *E. coli*, *B. subtilis*, *S. cerevisiae*, células de insectos en combinación con vectores de baculovirus, o células de un organismo superior, tal como vertebrados, por ejemplo, células COS 7, HEK 293, CHO, ovocitos de *Xenopus*, etc., se pueden utilizar para la producción de la proteína.

Cuando cualquiera de las células huésped referenciadas anteriormente, u otras células huésped apropiadas, se utilizan para replicar y/o expresar los ácidos nucleicos de la presente invención, el ácido nucleico replicado, la proteína o el polipéptido expresados resultantes se encuentran dentro del alcance de la presente invención como productos de la célula u organismo huésped. El producto se puede recuperar mediante un medio apropiado conocido en la técnica.

#### Proteínas

La presente invención también da a conocer proteínas fluorescentes funcionales mutantes cuyas secuencias de aminoácidos son idénticas, de manera sustancial, a la secuencia de aminoácidos de GFPxm de *A. macrodactyla* (SEQ ID NO: 2) y que difieren de la SEQ ID NO: 2, como mínimo, en una sustitución de aminoácido F220L. Dichas proteínas fluorescentes funcionales mutantes tienen una velocidad de maduración mejorada a una temperatura de 20°C o superior en comparación con la GFPxm.

En una realización preferente, una proteína fluorescente de la presente invención comprende sólo la sustitución

F220L en comparación con la SEQ ID NO:2 y tiene una velocidad de maduración incrementada en comparación con la GFPxm de *A. macrodactyla*. En una realización preferente, esta proteína fluorescente también tiene propiedades fluorescentes alteradas en comparación con la GFPxm de *A. Macrodactyla*.

5 En otra realización preferente, la sustitución F220L se combina con otras mutaciones para mejorar las propiedades de la proteína. Por ejemplo, diferentes combinaciones de sustituciones de aminoácidos seleccionadas del grupo que comprende K3G, E6D, T9A, P58T, F99L, F99H, M128K, M128E, I136M, Y151H, N144S, K162E, K156M, T214A, G228C, G228S y K238R incrementan adicionalmente la velocidad de maduración de la proteína a una temperatura de 20°C o superior tal como se muestra en la sección de "Ejemplo".

10 En muchas realizaciones, las proteínas de la presente invención presentan una absorbancia máxima que varía de aproximadamente 300 a 700 nm, normalmente de aproximadamente 350 a 650 nm y más normalmente de aproximadamente 400 a 600 nm. Las proteínas de la presente invención son proteínas fluorescentes, por lo que se entiende que se pueden excitar a una longitud de onda de luz, después de lo cual emitirán luz a otra longitud de onda. Los espectros de excitación de las proteínas de la presente invención varían habitualmente de aproximadamente 300 a 700 nm. Las proteínas de la presente invención presentan, en general, un coeficiente máximo de extinción que varía de aproximadamente 25.000 a 150.000 y normalmente de aproximadamente 45.000 a 129.000. Las proteínas de la presente invención varían habitualmente en longitud de aproximadamente 150 a 300 aminoácidos y normalmente de aproximadamente 200 a 300 residuos de aminoácidos, y, en general, presentan un peso molecular que varía de aproximadamente 15 a 35 kDa, normalmente de aproximadamente 17,5 a 32,5 kDa.

15 En ciertas realizaciones, las proteínas de la presente invención son brillantes, entendiendo por brillo que la fluorescencia de la proteína se puede detectar mediante métodos habituales (por ejemplo, cribado visual, espectrofotometría, espectrofluorometría, microscopía fluorescente, mediante máquinas de FACS, etc.). El brillo de la fluorescencia de proteínas fluorescentes particulares se determina por su rendimiento cuántico multiplicado por el coeficiente máximo de extinción.

20 En ciertas realizaciones, las proteínas de la presente invención presentan una velocidad de maduración incrementada a una temperatura de 20°C o superior en comparación con la GFPxm. La velocidad de maduración se puede estimar por el tiempo requerido para las proteínas para conseguir su estructura terciaria que da lugar a su calidad de fluorescencia en un cierto periodo de tiempo. En otras palabras, la velocidad de maduración de una proteína fluorescente se puede estimar mediante la intensidad de fluorescencia de células huésped que expresan la proteína de la presente invención después de cierto periodo de tiempo después de la transfección de células huésped con una construcción de expresión capaz de expresar dicha proteína fluorescente.

25 En ciertas realizaciones, las proteínas de la presente invención tienen una velocidad de maduración incrementada a una temperatura de 20°C o superior, de manera preferente de 30°C o superior, de la manera más preferente a una temperatura que varía de 35°C a 39°C, por ejemplo a 37°C. Se sabe que muchas células, incluyendo las células de mamífero, se incuban a aproximadamente 37°C a efectos de asegurar un crecimiento óptimo y/o un crecimiento fisiológicamente relevante. Las líneas celulares que se originan de diferentes organismos o tejidos pueden tener diferentes temperaturas relevantes que varían de aproximadamente 35°C para fibroblastos a aproximadamente 38°C-39°C para células beta de ratones.

30 Por ejemplo, para comparar las velocidades de maduración de proteínas fluorescentes a diferentes temperaturas, se pueden utilizar la siguiente estrategia: las células huésped (por ejemplo, células bacterianas, de manera preferente células de *E. Coli*) se transfectan con un vector de expresión que codifica una proteína fluorescente bajo el control de un promotor adecuado. En una cierta realización, la expresión de la proteína fluorescente comienza inmediatamente después de la transfección (cuando se utiliza un promotor constitutivo o debido a la pérdida de un promotor inducible). En otra realización, la expresión de la proteína fluorescente es inducida por el método conocido en la técnica. Las células huésped se desarrollan en una placa de petri a 20, 30 ó 37°C durante ciertos periodos de tiempo (por ejemplo, 36, 24 y 12 horas después del inicio de la expresión de las proteínas fluorescentes), se detecta la fluorescencia de las colonias de *E. coli* mediante los métodos habituales (por ejemplo, cribado visual, espectrofotometría, espectrofluorometría, microscopía fluorescente, mediante máquinas de FACS, etc.) y se calcula el brillo de su fluorescencia.

35 Las proteínas específicas de interés son las proteínas verdes fluorescentes mutantes: Mut 2 (SEQ ID NO 4); Mut 235 (SEQ ID NO 6); Mut 235-1 (SEQ ID NO 8); Mut 235-2 (SEQ ID NO 10); Mut 235-4 (SEQ ID NO 12); Mut-g9 (SEQ ID NO 14); y Mut 235-4G6 (SEQ ID NO 16). Las proteínas específicas de interés presentan una velocidad de maduración a una temperatura de 20°C o superior más elevada que la proteína GFPxm.

40 Las proteínas específicas de interés se describen con mayor detalle de forma individual posteriormente en la sección "Ejemplos".

También se dan a conocer proteínas que son similares o idénticas, de manera sustancial, a las secuencias de aminoácidos específicas de la presente invención, es decir las SEQ ID NO: 4-16. La identidad de secuencia se calcula en base a una secuencia de referencia determinada utilizando MegAlign, algoritmo clustal de DNASTAR descrito en D.G. Higgins y P.M. Sharp, "Fast and Sensitive multiple Sequence Alignments on a Microcomputer," (5) ("Alineaciones rápidas y sensibles de secuencias múltiples en un microordenador") CABIOS, 5 pág. 151-3 (1989) (utilizando los parámetros ktuple 1, gap penalty 3, window 5 y diagonals saved 5). En muchas realizaciones, las secuencias de aminoácidos de interés tienen una identidad de la secuencia mucho más elevada, por ejemplo, del 93%, 95%, 97%, 99%, 100%, de manera particular para las secuencias de los aminoácidos que proporcionan las regiones funcionales de la proteína.

(10) También se dan a conocer proteínas que son mutantes de las proteínas descritas anteriormente. Los mutantes pueden retener propiedades biológicas de las proteínas de origen o pueden tener propiedades biológicas que difieren de las proteínas de tipo natural. El término "propiedad biológica" de las proteínas de la presente invención se refiere, aunque sin limitarse a las mismas, a propiedades fluorescentes; propiedades bioquímicas, tales como la estabilidad *in vivo* y/o *in vitro* (por ejemplo, semivida); velocidad de maduración, tendencia de agregación y tendencia de oligomerización y otras de dichas propiedades. Entre las mutaciones se incluyen cambios de sólo un aminoácido, eliminaciones inserciones de uno o más aminoácidos, truncamientos o extensiones en el extremo N-terminal, truncamientos o extensiones en el extremo C-terminal, y similares.

(15) Los mutantes se pueden generar utilizando técnicas estándar de biología molecular tal como se describe en detalle en la sección anterior "Moléculas de ácido nucleico". Los mutantes descritos en el presente documento incluyen:

- (25) (1) un mutante de la Mut-g9 con propiedades fluorescentes aumentadas que comprende las sustituciones I167T, F223S, S65C, y F64L en comparación con Mut-g9 (SEQ ID NO:14). Dicho mutante también posee una velocidad de maduración incrementada en comparación las proteínas GFPxm y Mut-9. La secuencia de aminoácidos de este mutante denominado tagGFP (también macGFP) se muestra en la SEQ ID NO: 18;
- (30) (2) un mutante de la tagGFP con desplazamiento a cian en los espectros de fluorescencia que comprende las sustituciones C65A, Y66W, L99H, I123V, K128E, D129G, F145A, N146I, H148D, V163A, T167I, T203C, T205S, C227Y en comparación con la tagGFP. La secuencia de aminoácidos de este mutante denominado tagCFP se muestra en la SEQ ID NO: 20;
- (35) (3) un mutante de la tagGFP con desplazamiento a Amarillo en los espectros de fluorescencia que comprende las sustituciones C65T, I68V, E76K, M153T, F224V, C228S y T203Y en comparación con la tagGFP. La secuencia de aminoácidos de este mutante denominado tagYFP se muestra en la SEQ ID NO: 22.

(40) Dada la directriz proporcionada en los ejemplos, y utilizando técnicas estándar, los expertos en la materia pueden generar fácilmente una amplia variedad de mutantes adicionales y analizar si se ha alterado una propiedad biológica (por ejemplo, bioquímica, espectral, etc.). Por ejemplo, la intensidad de fluorescencia se puede medir utilizando un espectrofotómetro a varias longitudes de onda de excitación.

(45) Las proteínas de la presente invención están presentes en forma aislada, por lo que se entiende que la proteína está libre, de manera sustancial, de otras proteínas y otras moléculas biológicas naturales, tales como oligosacáridos, ácidos nucleicos y fragmentos de los mismos, y similares, en las que el término "libre de manera sustancial" en este caso significa que menos del 70%, normalmente menos del 60% y más normalmente menos del 50% de la composición que contiene la proteína aislada es alguna otra molécula biológica natural. En ciertas realizaciones, las proteínas están presentes en forma purificada, de manera sustancial, en las que "forma purificada, de manera sustancial" significa, como mínimo, el 95%, normalmente, como mínimo, el 97% y más normalmente, como mínimo, el 99% puras.

(50) En una realización preferente, las proteínas de la presente invención se producen de manera sintética, por ejemplo, mediante la expresión de una secuencia codificante de ácido nucleico recombinante que codifica la proteína de interés en un huésped adecuado, tal como se ha descrito anteriormente. Se puede utilizar cualquier procedimiento de purificación de proteínas conveniente, en el que las metodologías de purificación de proteínas adecuadas se describen en la Guide to Protein Purification, (Deutscher ed.) (Academic Press, 1990). Por ejemplo, se puede (55) preparar un lisado a partir de la fuente original y se puede purificar utilizando HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía de afinidad, y similares.

(60) También se dan a conocer proteínas de fusión que comprenden una proteína de la presente invención, o fragmentos funcionales de la misma, fusionada, por ejemplo, a una secuencia de degradación, una secuencia de localización subcelular (por ejemplo, una señal de localización nuclear, una señal de reconocimiento peroximal, una secuencia de reconocimiento del aparato de Golgi, una secuencia de reconocimiento mitocondrial, etc.), un péptido señal, o cualquier proteína o polipéptido de interés. Las proteínas de fusión pueden comprender, por ejemplo, una proteína fluorescente de la presente invención y un segundo polipéptido ("el compañero de la fusión") fusionado en el marco a

los extremos N-terminal y/o C-terminal de la proteína fluorescente. Entre los compañeros de fusión se incluyen, aunque sin limitarse a los mismos, polipéptidos que se pueden unir a anticuerpos específicos al compañero de fusión (por ejemplo, marcadores de epítomos), anticuerpos o fragmentos de unión de los mismos, polipéptidos que proporcionan una función catalítica o inducen una respuesta celular, ligandos o receptores o miméticos de los mismos, y similares.

También se dan a conocer anticuerpos que se unen de manera específica a las proteínas fluorescentes de la presente invención. Los anticuerpos adecuados se pueden producir utilizando las técnicas conocidas en el sector. Por ejemplo, se pueden obtener anticuerpos policlonales tal como se da a conocer en (Harlow y Lane, *Antibodies: A Laboratory Manual* ("Anticuerpos: un manual de laboratorio"), (1988) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York) y se pueden obtener anticuerpos monoclonales tal como se da a conocer en (Goding *Monoclonal Antibodies: Principles and Practice: Production and Application of Monoclonal Antibodies in Cell Biology* ("Anticuerpos monoclonales: principios y práctica: producción y aplicación de anticuerpos monoclonales en biología celular"), *Biochemistry and Immunology*; 3ª edición, (1996) Academic Press). También son de interés los anticuerpos quiméricos que incluyen anticuerpos humanizados, así como anticuerpos monocatenarios y fragmentos de anticuerpos, tales como Fv, F(ab')<sub>2</sub> y Fab.

#### Transgénicos

Los ácidos nucleicos de la presente invención se pueden utilizar para generar organismos transgénicos o modificaciones específicas de sitio de genes en líneas celulares. Las células transgénicas de la presente invención incluyen uno o más ácidos nucleicos, según la presente invención, presentes como un transgén. Para los objetivos de la presente invención, se puede utilizar cualquier célula huésped adecuada, que incluye células huésped procariontas (por ejemplo, *Escherichia coli*, especie *Streptomyces*, *Bacillus subtilis*, *Lactobacillus acidophilus*, etc.) o células huésped eucariotas. Los organismos transgénicos de la presente invención pueden ser organismos procariontas o eucariotas que incluyen bacterias, cianobacterias, hongos, plantas y animales, en los que una o más de las células del organismo contienen ácido nucleico heterólogo de la presente invención introducido mediante intervención humana, tal como mediante técnicas transgénicas conocidas en el sector.

El ácido nucleico aislado de la presente invención se puede introducir en el huésped mediante métodos conocidos en la técnica, por ejemplo, infección, transfección, transformación o transconjugación. Las técnicas para transferir las moléculas de ácido nucleico (es decir, ADN) en dichos organismos son ampliamente conocidas y se dan a conocer en referencias, tales como Sambrook y otros (*Molecular Cloning: A Laboratory Manual* ("Clonación molecular: un manual de laboratorio"), 3ª edición, (2001) Cold Spring Harbor Press, Cold Spring Harbor, NY).

En una realización, el organismo transgénico puede ser un organismo procarionta. Los métodos de transformación de huéspedes procariontas están bien documentados en la técnica (por ejemplo, véase, Sambrook y otros, *Molecular Cloning: A Laboratory Manual* ("Clonación molecular: un manual de laboratorio"), 2ª edición (1989) Cold Spring Harbor Laboratory Press y Ausubel y otros, *Current Protocols in Molecular Biology* ("Protocolos actuales en biología molecular") (1995) John Wiley & Sons, Inc).

En otra realización, el organismo transgénico puede ser un hongo, por ejemplo, levadura. La levadura se utiliza ampliamente como vehículo para la expresión de genes heterólogos (por ejemplo, véase, Goodey y otros, *Yeast biotechnology* ("Biotecnología de levaduras"), D R Berry y otros, eds, (1987) Allen y Unwin, Londres, pág. 401-429, y King y otros, *Molecular and Cell Biology of Yeasts* ("Biología molecular y celular de las levaduras"), E. F. Walton y G. T. Yarronton, eds, Blackie, Glasgow (1989) pág. 107-133). Existen varios tipos de vectores de levaduras disponibles, que incluyen vectores integrativos, que requieren la recombinación el genoma huésped para su mantenimiento y vectores plásmidos que se replican de manera autónoma.

Otro organismo huésped es un animal. Los animales transgénicos se pueden obtener mediante técnicas transgénicas conocidas en el sector y dadas a conocer en referencias, tales como Pinkert, *Transgenic Animal Technology: a Laboratory Handbook* ("Tecnología de animales transgénicos: un manual de laboratorio"), 2ª edición (2003) San Diego: Academic Press; Gersenstein y Vintersten, *Manipulating the Mouse Embryo: A Laboratory Manual* ("Manipulación de embriones de ratón: un manual de laboratorio"), 3ª edición, (2002) Nagy A. (Ed), Cold Spring Harbor Laboratory; Blau y otros, *Laboratory Animal Medicine* ("Medicina de animales de laboratorio"), 2ª edición, (2002) Fox J.G., Anderson L.C., Loew F.M., Quimby F.W. (Eds), American Medical Association, American Psychological Association; Gene Targeting: A Practical Approach ("Reconocimiento génico: una estrategia práctica") por Alexandra L. Joyner (Ed.) Oxford University Press; 2ª edición (2000). Por ejemplo, los animales transgénicos se pueden obtener a través de recombinación homóloga, en la que se altera el locus endógeno. De manera alternativa, se incorpora de manera aleatoria una construcción de ácido nucleico en el genoma. Entre los vectores para la incorporación estable se incluyen plásmidos, retrovirus y otros virus de animales, YAC, y similares.

El ácido nucleico se puede introducir en la célula, de manera directa o indirecta, mediante la introducción en un

precursor de la célula, por medio de manipulación genética deliberada, tal como mediante microinyección o mediante infección con un virus recombinante o con un vector viral recombinante y similar. El término manipulación genética no incluye el cruce clásico o la fecundación *in vitro*, sino que se refiere a la introducción de una molécula de ácido nucleico recombinante. Esta molécula de ácido nucleico se puede incorporar en un cromosoma, o puede ser ADN que se replica de manera extracromosómica.

Las construcciones de ADN para la recombinación homóloga comprenderá, como mínimo, una parte de un ácido nucleico de la presente invención, en las que el gen tiene la modificación o modificaciones genéticas deseadas, e incluye regiones de homología al locus diana. Las construcciones de ADN para la incorporación aleatoria no necesitan incluir regiones de homología para mediar en la recombinación. De manera conveniente, se pueden incluir marcadores para la selección positiva y negativa. Los métodos para generar células que tienen modificaciones génicas reconocidas a través de recombinación homóloga son conocidos en la técnica. Para varias técnicas para transfectar células de mamífero, véase Keown y otros, *Meth. Enzymol.* (1990) 185:527-537.

Para células madre embrionarias (ES), se puede utilizar una línea celular ES, o se pueden obtener células embrionarias nuevas de un huésped, tal como un ratón, una rata, una cobaya, etc. Dichas células se desarrollan en una capa adecuada de alimentación de fibroblastos o se desarrollan en presencia de un factor inhibidor de leucemia (LIF). Las ES transformadas o las células embrionarias se pueden utilizar para producir animales transgénicos utilizando la técnica apropiada descrita en el sector.

Los animales transgénicos pueden ser cualquier animal no humano, que incluye mamíferos no humanos (por ejemplo, ratón, rata), un ave o un anfibio, etc., y se pueden utilizar en estudios funcionales, cribado de fármacos y similares. Entre los ejemplos representativos de la utilización de animales transgénicos se incluyen los descritos a continuación.

También se pueden producir plantas transgénicas. Los métodos de preparación de células vegetales transgénicas y plantas se dan a conocer en las patentes de Estados Unidos No. 5.767.367, 5.750.870, 5.739.409, 5.689.049, 5.689.045, 5.674.731, 5.656.466, 5.633.155, 5.629.470, 5.595.896, 5.576.198, 5.538.879, y 5.484.956, la materia de las cuales se incorpora en el presente documento por referencia. Los métodos de producción de plantas transgénicas también se revisan en *Plant Biochemistry and Molecular Biology* (eds. Lea y Leegood, John Wiley & Sons) (1993) pág. 275-295 y en *Plant Biotechnology and Transgenic Plants* (eds. Oksman-Caldentey y Barz), (2002) pág. 719.

Por ejemplo, se pueden utilizar explantes embriogénicos que comprenden células somáticas para la preparación del huésped transgénico. Después de la recogida de células o tejido, se introduce el ADN exógeno de interés en las células de la planta, donde para dicha introducción existen un conjunto de técnicas diferentes disponibles. Con protoplastos aislados, surge la oportunidad para la introducción mediante los protocolos de transferencia de genes mediada por ADN, que incluyen la incubación de los protoplastos con ADN desnudo, tal como plásmidos que comprenden la secuencia codificante exógena de interés en presencia de cationes polivalentes (por ejemplo, PEG o PLO); o la electroporación de los protoplastos en presencia de ADN desnudo que comprende la secuencia exógena de interés. A continuación, se seleccionan los protoplastos que han captado de manera satisfactoria el ADN exógeno, se desarrollan en un callo, y finalmente, en una planta transgénica a través del contacto con las cantidades y las proporciones apropiadas de factores estimuladores, tales como auxinas y citoquinas.

Se pueden utilizar otros métodos adecuados para producir plantas, tales como la estrategia de "pistola génica" ("gene-gun") o la transformación mediada por *Agrobacterium* disponible para los expertos en la materia.

#### Métodos de utilización

Las proteínas fluorescentes de la presente invención (así como otros componentes de la presente invención dados a conocer anteriormente) son útiles en un conjunto de aplicaciones diferentes. Las utilizaciones representativas para cada uno de estos tipos de proteínas se describirán a continuación, en las que las utilizaciones descritas en el presente documento son meramente a modo de ejemplo y, de ningún modo, pretenden limitar la utilización de las proteínas de la presente invención a las descritas.

En una realización preferente, en relación al método para marcar una proteína, célula u orgánulo celular, las proteínas de la presente invención son útiles como marcadores *in vivo* (o moléculas informadoras) en ensayos celulares y de biología molecular. Entre los ensayos de interés se incluyen, aunque sin limitarse a los mismos, ensayos para la expresión génica, localización y co-localización de proteínas, interacciones de proteína-proteína, interacciones de proteína-ácido nucleico, interacciones de ácido nucleico-ácido nucleico, localización e interacciones de células y orgánulos celulares, etc. Las proteínas fluorescentes de la presente invención son útiles como marcadores de proteínas, o marcadores de orgánulos celulares en células vivas y fijadas, como marcadores en fusiones celulares o de orgánulos, como marcadores de la incorporación en una célula u orgánulo, como marcadores

de una transfección (por ejemplo, como marcadores para la selección de células transfectadas que contienen un vector de expresión que codifica, como mínimo, una proteína fluorescente de la presente invención), y como sondas en tiempo real que trabajan a concentraciones casi fisiológicas, etc.

5 Por ejemplo, las proteínas de la presente invención son útiles para identificar y/o medir la expresión de una proteína o polipéptidos de interés en un material biológico. Este método comprende: i) introducir en una célula una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica una proteína fluorescente, según la presente invención, en la que dicha molécula de ácido nucleico está unida de manera operativa a una secuencia de control de la expresión y bajo el control de la misma, que controla la expresión de la proteína o el polipéptido de interés; ii) expresión de dicho ácido nucleico bajo condiciones adecuadas; y iii) detectar la emisión de fluorescencia de la proteína fluorescente como medio de medición de la expresión de la proteína de interés.

10 Además, las proteínas de la presente invención son útiles para la localización de una proteína o polipéptido de interés en material biológico. Este método comprende: i) introducir en una célula una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica una proteína fluorescente, según la presente invención, en la que dicha molécula de ácido nucleico se fusiona con una secuencia que codifica una proteína o polipéptido de interés y se une de manera operativa a una secuencia de control de la expresión y bajo el control de la misma; ii) cultivar la célula bajo condiciones adecuadas para la expresión de la proteína de interés; y iii) detectar la emisión de fluorescencia de la proteína fluorescente como medio de medición de la localización de la proteína de interés.

15 Entre las aplicaciones de interés se incluyen la utilización de las proteínas de la presente invención en métodos de transferencia de energía por resonancia de fluorescencia (FRET). En estos métodos, las proteínas de la presente invención actúan como dadoras y/oceptoras en combinación con una segunda proteína fluorescente o colorante, por ejemplo, un proteína fluorescente tal como se da a conocer en Matz y otros, Nature Biotechnology 17: 969-973 (1999); otros colorantes fluorescentes, tales como cumarina y sus derivados, 7-amino-4-metilcumarina y aminocumarina; colorantes bodipy; azul cascada; o fluoresceína y sus derivados, tales como isotiocianato de fluoresceína y verde Oregón; colorantes de rodamina, tales como rojo Texas, tetrametilrodamina, eosinas y eritrosinas; colorantes de cianina, tales como Cy3 y Cy5; quelatos macrocíclicos de iones lantánidos, tales como un colorante cuántico; y colorantes quimioluminiscentes, tales como luciferasas, que incluyen los descritos en las patentes de Estados Unidos N° 5.843.746, 5.700.673, 5.674.713, 5.618.722, 5.418.155, 5.330.906, 5.229.285, 5.221.623, y 5.182.202, cuya materia se incorpora en el presente documento por referencia.

20 Entre los ejemplos específicos de cuando los ensayos de FRET utilizan proteínas fluorescentes de la presente invención se incluyen, aunque sin limitarse a los mismos, los descritos en: las patentes de Estados Unidos N° 6.008.373, 5.998.146, 5.981.200, 5.945.526, 5.945.283, 5.911.952, 5.869.255, 5.866.336, 5.863.727, 5.728.528, 5.707.804, 5.688.648, y 5.439.797, cuya materia se incorpora en el presente documento por referencia.

25 Las proteínas fluorescentes de la presente invención son útiles en un método para la detección de los efectos de una sustancia de prueba en la regulación de la expresión y/o la translocación de una o más proteínas de interés en una célula. De manera alternativa, son útiles en un método para la detección de la expresión de una proteína de interés y la actividad simultánea de una secuencia de control de la expresión en respuesta a una sustancia de prueba. Las proteínas fluorescentes también son útiles en un método para comparar la actividad de dos o más secuencias de control de la expresión en una célula en respuesta a una sustancia de prueba. Dichos métodos se pueden realizar en presencia y en ausencia de una sustancia de prueba, cuyo efecto en el proceso debe medirse.

30 Las proteínas fluorescentes de la presente invención también son útiles en aplicaciones que implican el cribado automatizado de conjuntos de células que expresan grupos informadores fluorescentes mediante la utilización de análisis de imagen microscópica y análisis electrónico. El cribado se puede utilizar para el descubrimiento de fármacos y en el campo de la genómica funcional, en el que las proteínas de la presente invención se utilizan como marcadores de las células completas para detectar cambios en la reorganización y migración multicelular, por ejemplo, en la formación de túbulos multicelulares (formación de vasos sanguíneos) por las células endoteliales, la migración de células a través del sistema de inserción Fluoroblok (Becton Dickinson Co.), la curación de heridas o el crecimiento de neuritas. El cribado también se puede utilizar cuando la proteínas de la presente invención se utilizan como marcadores fusionados a péptidos (tales como, secuencias de reconocimiento) o proteínas que detectan cambios en la localización intracelular como indicadores de la actividad celular, por ejemplo, en la transducción de señales, tales como la translocación de quinasas y el factor de transcripción después de estímulos. Entre los ejemplos se incluyen quinasa C, proteína quinasa A, factor de transcripción NFκB y NFAT; proteínas del ciclo celular, tales como ciclina A, ciclina B1 y ciclina E; la separación por proteasas con el posterior movimiento del sustrato separado; fosfolípidos, con marcadores para estructuras intracelulares, tales como el retículo endoplasmático, el aparato de Golgi, las mitocondrias, los peroxisomas, el núcleo, el nucléolo, la membrana plasmática, las histonas, los endosomas, los lisosomas o los microtúbulos.

35 Las proteínas de la presente invención también se pueden utilizar en cribados de alto contenido para detectar la

5 co-localización de otras proteínas de fusión fluorescentes con marcadores de localización como indicadores del movimiento de las proteínas/péptidos fluorescentes intracelulares o como marcadores solos. Entre los ejemplos de solicitudes que implican el cribado automatizado de conjuntos de células en que las proteínas fluorescentes de la presente invención son útiles se incluyen la patente de Estados Unidos N° 5.989.835, sí como los documentos WO 0017624, WO 00/26408, WO 00/17643 y WO 00/03246, cuya materia se incorpora en el presente documento por referencia.

10 Las proteínas fluorescentes de la presente invención también son útiles en ensayos de cribado de alto rendimiento. Las proteínas fluorescentes de la presente invención son proteínas estables con semividas de más de 24 horas. También se dan a conocer versiones desestabilizadas de las proteínas fluorescentes de la presente invención con semividas disminuidas que se pueden utilizar como informadoras de la transcripción para el descubrimiento de fármacos. Por ejemplo, una proteína, según la presente invención, se puede fusionar con una posible secuencia señal proteolítica derivada de una proteína con una semivida más corta, tal como una secuencia PEST del gen de la ornitín descarboxilasa de ratón, una caja de destrucción de la ciclina B1 de ratón o ubiquitina, etc. Para una descripción de proteínas desestabilizadas y vectores que se pueden utilizar para producir las mismas, véase, por ejemplo, la patente de Estados Unidos N° 6.130.313, cuya materia se incorpora en el presente documento por referencia. Se pueden detectar promotores en los mecanismos de transducción de señales utilizando versiones desestabilizadas de las proteínas fluorescentes de la presente invención para el cribado de fármacos, tales como, por ejemplo, AP1, NFAT, NFkB, Smad, STAT, p53, E2F, Rb, myc, CRE, ER, GR y TRE, y similares.

20 Las proteínas de la presente invención se pueden utilizar como detectores de segundos mensajeros mediante la fusión de las proteínas de la presente invención a dominios específicos, tales como el dominio de unión a Ca de PKCgamma, el dominio de unión a DAG de PKCgamma, el dominio SH2 o el dominio SH3, etc.

25 Las formas secretadas de las proteínas de la presente invención, que a su vez se pueden utilizar en un conjunto de aplicaciones diferentes se pueden preparar mediante la fusión de secuencias líder secretadas a las proteínas de la presente invención.

30 Las proteínas de la presente invención también son útiles en aplicaciones de separación celular activada por fluorescencia (FACS). En dichas aplicaciones, la proteína fluorescente de la presente invención se utiliza como un marcador para marcar una población de células y, a continuación, la población marcada resultante de células se separa con un dispositivo de separación celular activada por fluorescencia, tal como se conoce en la técnica. Los métodos FACS se dan a conocer en las patentes de Estados Unidos N° 5.968.738 y 5.804.387, cuya materia se incorpora en el presente documento por referencia.

35 Las proteínas de la presente invención también son útiles como marcadores *in vivo* en animales transgénicos. Por ejemplo, la expresión de la proteína de la presente invención puede estar impulsada por promotores específicos de tejido, en la que dichos métodos son útiles en la investigación para terapia génica, tal como la evaluación de la eficacia de la expresión transgénica, entre otras aplicaciones. Una aplicación representativa de proteínas fluorescentes en animales transgénicos que ilustra dichas aplicaciones se encuentra en el documento WO 00/02997, cuya materia se incorpora en el presente documento por referencia.

45 Entre las aplicaciones adicionales de las proteínas de la presente invención se incluyen como marcadores, después de la inyección en células o animales y en calibración para mediciones cuantitativas, como marcadores o informadores en dispositivos biosensores de oxígeno para hacer el seguimiento de la viabilidad celular, y como marcadores para animales, animales domésticos, juguetes, alimentos, y similares.

50 Las proteínas fluorescentes de la presente invención también son útiles como biosensores en células procariontas y eucariotas, tales como un indicador de iones de Ca<sup>2+</sup>, un indicador de pH, un indicador de fosforilación, o como un indicador de otros iones, tales como magnesio, sodio, potasio, cloruro y haluros. Los métodos de utilización de proteínas fluorescentes como biosensores también incluyen los dados a conocer en las patentes de Estados Unidos N° 5.972.638, 5.824.485, y 5.650.135 (así como las referencias citadas en las mismas), cuya materia se incorpora en el presente documento por referencia.

55 Las proteínas fluorescentes de la presente invención también son útiles como fuente de proteínas fluorescentes permutadas de forma circular y los biosensores de las mismas. Los métodos de preparación y la utilización de proteínas fluorescentes permutadas de forma circular se dan a conocer en Nagai y otros, Proc Natl Acad Sci USA, 2001, V. 98(6), pág. 3197-3202, Nagai y otros, Proc Natl Acad Sci USA, 2004, V. 101(29), pág. 10554-10559, Filippin y otros, J Biol Chem., 2003, V. 278(40), pág. 39224-34, y las patentes de Estados Unidos N° 6.469.154 y 6.699.687, cuya materia se incorpora en el presente documento por referencia.

60 Los anticuerpos de la presente invención, descritos anteriormente, también son útiles en un conjunto de aplicaciones, que incluyen la diferenciación de las proteínas de la presente invención de otras proteínas

fluorescentes.

Kits

5 La presente invención también da a conocer kits para la utilización en la práctica de una o más de las aplicaciones  
 descritas anteriormente. En realizaciones preferentes, los kits se pueden utilizar para el marcaje de proteínas. Los  
 kits incluyen habitualmente la propia proteína de la presente invención, o un ácido nucleico que codifica la misma, de  
 10 manera preferente con los elementos para la expresión de las proteínas de la presente invención, por ejemplo, una  
 construcción, tal como un vector que comprende un ácido nucleico que codifica la proteína de la presente invención.  
 Los componentes del kit están presentes habitualmente en un medio de almacenamiento adecuado, tal como una  
 15 solución tamponada, habitualmente en un recipiente adecuado. En los kits también pueden estar presentes  
 anticuerpos específicos a la proteína proporcionada. En ciertas realizaciones, el kit comprende un conjunto de  
 vectores diferentes que codifican cada uno la proteína de la presente invención, en el que los vectores se diseñan  
 para la expresión en medios diferentes y/o bajo condiciones diferentes, por ejemplo, la expresión constitutiva en la  
 que el vector incluye un promotor fuerte para la expresión en células de mamífero o un vector sin promotor con un  
 sitio de clonación múltiple para una inserción habitual de un promotor y la expresión ajustada, etc.

Además de los componentes anteriores, los kits de la presente invención incluirán además instrucciones para  
 20 realizar los métodos de la presente invención. Estas instrucciones pueden estar presentes en los kits de la presente  
 invención en una variedad de formas, una o más de las cuales pueden estar presentes en el kit.

El siguiente ejemplo se ofrece a modo de ilustración y no a modo de limitación.

**Ejemplos**

25

Ejemplo 1

Generación de ácidos nucleicos que codifican proteínas fluorescentes mutantes de GFPxm

30 Se produjo de manera sintética un ácido nucleico que codificaba la GFPxm de *A. macrodactyla* de tipo natural. Para  
 aumentar la producción de proteínas en sistemas de expresión eucariotas, se realizó una vez la humanización del  
 gen de GFPxm. Las composiciones de nucleótidos y aminoácidos para la GFPxm humanizada se muestran en las  
 SEQ ID NO: 1, 2.

35 Se realizó una mutagénesis aleatoria adicional para obtener una biblioteca de variantes de GFPxm mutadas de  
 forma aleatoria utilizando el kit de Mutagénesis aleatoria por PCR para diversidad (CLONTECH), bajo condiciones  
 óptimas para 3-4 mutaciones por cada 1000 pb. El producto de la PCR se clonó en el vector pQE30 (Qiagen) y se  
 transformó en *E. coli* (cepa XL1-azul). Las colonias de *E. coli* que expresaban proteínas mutantes se desarrollaron a  
 40 37°C y se cribaron visualmente con un estereomicroscopio fluorescente SZX-12 (Olympus) después de 12-24 horas  
 de crecimiento.

En la primera ronda, se seleccionó el clon que poseía la fluorescencia más brillante después de 18 horas de  
 crecimiento celular. La secuencia de la inserción de ácido nucleico de este clon mostró que comprende una  
 45 sustitución F220L en comparación con la proteína GFPxm. Las composiciones de nucleótidos y aminoácidos de esta  
 proteína denominada Mut 2 se muestran en las SEQ ID NO: 3, 4.

El ácido nucleico de Mut 2 se sometió a varias rondas adicionales de mutagénesis aleatoria dando lugar a los  
 siguientes mutantes: (i) segunda ronda: Mut 235 (SEQ ID NO: 5, 6); (ii) tercera ronda: -Mut 235-1 (SEQ ID NO: 7, 8);  
 Mut 235-2 (SEQ ID NO: 9, 10); Mut 235-4 (SEQ ID NO: 11, 12); (iii) cuarta ronda: Mut-g9 (SEQ ID NO: 13, 14); Mut  
 50 235-4G6 (SEQ ID NO: 15, 16). Según los datos de cribado visual, el mutante Mut-g9 que comprende las  
 sustituciones de aminoácidos F220U K3G/T9A/ F99L/M128K/N144S/K162E/T214A/G228C/K238R (en comparación  
 con GFPxm) madura más rápido a 37°C que otros mutantes analizados.

Ejemplo 2

55

Caracterización de proteínas fluorescentes mutantes

Los ácidos nucleicos que codifican las proteínas GFPxm, Mut 2 y Mut-g9 se obtuvieron tal como se describe en el  
 ejemplo 1. Tal como se describe anteriormente, estos ácidos nucleicos se clonaron en un vector de expresión  
 60 pQE30 (Qiagen), de manera que la proteína recombinante contenía una etiqueta de seis histidinas en su extremo  
 N-terminal. Después de la expresión en *E. coli*, las proteínas se purificaron a través de una resina de afinidad a  
 metal TALON (Clontech). Se obtuvieron los espectros de excitación-emisión utilizando el Espectrofotómetro de  
 fluorescencia Varian Cary Eclipse. Los espectros de excitación-emisión para estas proteínas se muestran en las

figuras 1-3. Se observó que la mutación F220L altera las propiedades fluorescentes de la proteína fluorescente.

Las velocidades de maduración de estas proteínas se caracterizaron en dos sistemas *in vivo*. En el primer experimento, se transformaron células de *E. coli* (cepa XL1-azul) con pQE30 (Qiagen) que codificaba las correspondientes proteínas fluorescentes bajo el control del promotor T5 y se desarrollaron en una placa de petri a 20, 30 ó 37°C durante 36, 24 y 12 horas, respectivamente. En el sistema utilizado, la proteína fluorescente se expresa de manera constante y madura durante el crecimiento en *E. Coli* debido a la pérdida del promotor. Después del crecimiento celular bajo las condiciones mencionadas, las colonias fluorescentes se fotografiaron utilizando un microscopio estéreo fluorescente Olympus US SZX12 completado con una cámara Olympus DP50. El brillo de las colonias se calculó utilizando el software ImageJ. Los resultados de medición se muestran en un histograma en la figura 4.

En el otro experimento, se desarrollaron en un medio LB complementado con glucosa al 2% y ampicilina 100 µg/ml durante 5 horas colonias individuales de *E. coli* que portaban vectores que codifican la proteína fluorescente, se centrifugaron y se colocaron en el tampón TrisHCl, pH 7,5 que contenía NaCl 100 mM. Se indujo la expresión intensa de una proteína fluorescente a 37°C mediante la adición de IPTG hasta una concentración final de 1 mM. Se hizo el seguimiento de la señal de fluorescencia a 37°C debido a la expresión y maduración de la proteína fluorescente sintetizada utilizando el Espectrofotómetro de fluorescencia Varian Cary Eclipse en software Kinetics durante 6 horas (figura 5).

En ambos sistemas experimentales, la velocidad de maduración de las proteínas se incrementa en el orden mostrado: GFPxm < Mut 2 < Mut-g9 (figuras 4, 5).

Ejemplo 3

#### Mutagénesis de Mut-g9

El ácido nucleico que codifica la proteína Mut-g9 se obtuvo tal como se describe en el ejemplo 1 y se sometió a mutagénesis dirigida de sitio para obtener variantes con propiedades fluorescentes alteradas y una capacidad baja de formar dímeros. Como resultado, se obtuvo la proteína tagGFP (SEQ ID NO: 17, 18) que contenía las siguientes sustituciones de aminoácidos (en comparación con Mut-g9): I167T, F223S, S65C, F64L. Los espectros de excitación-emisión para esta proteína se muestran en la figura 6. La velocidad de maduración de esta proteína fue superior que la de las proteínas Mut-2 y Mut-g9 de GFPxm. La velocidad de maduración se analizó tal como se describe en el ejemplo 2.

De manera adicional, las variantes de tagGFP con espectros alterados de fluorescencia producidas mediante mutagénesis dirigida a sitio de T203 e Y66 dieron lugar a una variante desplazada al amarillo (excitación/emisión con picos máximos a 502/521 nm) que comprendía las sustituciones T203Y y F224V, y una variante desplazada a cian (excitación/emisión con picos máximos a 430/470 nm) que comprendía una sustitución Y66W. Se utilizaron los ácidos nucleicos que codificaban estas variantes espectrales para mutagénesis aleatoria para mejorar el plegamiento de las proteínas (tal como se muestra después de la expresión en *E. Coli*, cepa XL1-Azul). Éstas dieron lugar la proteína fluorescente cian tagCFP con las secuencias de nucleótidos y aminoácidos mostradas en las SEQ ID NO: 19, 20 y la proteína fluorescente amarilla tagYFP1 con las secuencias de nucleótidos y aminoácidos mostradas en las SEQ ID NO: 21, 22.

En comparación con la tagGFP, la tagCFP comprende la sustitución Y66W en combinación con C65A, L99H, I123V, K128E, D129G, F145A, N146I, H148D, V163A, T167I, T203C, T205S, y C227Y, mientras que la tagYFP1 comprende las sustituciones T203Y, F224V en combinación con las sustituciones C65T, I68V, E76K, M153T, y C228S. Los espectros de excitación y emisión para estas proteínas se muestran en las figuras 7A y 7B.

Además, también se generó un mutante de tagYFP con tendencia de oligomerización reducida, mediante mutagénesis dirigida a sitio del residuo A206, denominado tagYFP2 (SEQ ID: 23, 24). Esta proteína existe como monómero incluso a concentraciones elevadas (5 mg/ml), tal como se ha observado mediante la filtración en gel.

Ejemplo 4

#### Marcaje de células de mamífero utilizando tagGFP, tagCFP y tagYFP1.

Para el marcaje fluorescente de células eucariotas, se clonaron por separado en el vector pEGFP-C1 (CLONTECH) entre los sitios de restricción *AgeI* y *BglII* (en lugar de la región codificante de EGFP) los ácidos nucleicos que codificaban tagGFP, tagCFP y tagYFP1, preparados tal como se describe anteriormente en el ejemplo 3. Se utilizaron las siguientes líneas celulares: células epiteliales de riñón humano 293T, fibroblastos de embriones de ratón 3T3, fibroblastos subcutáneos murinos L929, células epiteliales de riñón de mono verde africano Vero y

fibroblastos de riñón de mono verde africano COS1. Las células se transfectaron utilizando el reactivo LipofectAMINE (Invitrogen) y se analizaron 20 horas después de la transfección. Para la obtención de imágenes de las células se utilizó un microscopio de fluorescencia Olympus CK40 equipado con una cámara CCD (DP-50, Olympus). La expresión de estas proteínas en diferentes líneas celulares dio lugar a señales fluorescentes brillantes sin agregación. La fluorescencia era claramente detectable a las 24 horas después de la transfección. No se observó toxicidad celular.

Ejemplo 5

10 Marcaje de proteínas orgánulos utilizando tagGFP y tagCFP.

Los ácidos nucleicos que codificaban tagGFP y tagCFP, preparados tal como se describe anteriormente en el ejemplo 3, se unieron de manera operativa con ácidos nucleicos que codificaban beta-actina citoplasmática humana, alfa-tubulina, fibrilarina o la secuencia dirigida a las mitocondrias del precursor de la subunidad VIII de la citocromo C oxidasa humana. La transfección de células humanas 293T y HeLa con los plásmidos indicados anteriormente que expresaban fusiones de proteínas fluorescentes con proteínas celulares huésped y/o señales de localización dieron lugar a una fluorescencia brillante que revelaba patrones que estaban estrechamente de acuerdo con los observados para la fusiones con EGFP.

20 Todas las publicaciones y solicitudes de patentes citadas en este documento se incorporan en el presente documento por referencia como si cada publicación o solicitud de patente individual se incorporara de manera específica e individual por referencia. La citación de cualquier publicación es para proporcionar contexto y entendimiento de la presente invención y no debe interpretarse como una admisión de que cualquiera de dichas publicaciones es del estado de la técnica anterior.

25

LISTADO DE SECUENCIAS

<110> Lukyanov, Sergey

30 <120> PROTEÍNAS VERDES FLUORESCENTES MODIFICADAS Y MÉTODO PARA LA UTILIZACIÓN DE LAS MISMAS

<130> EVRO0012\_PCT.doc

35 <150> US 60/733.429

<151> 2005-11-04

<160> 24

40 <170> Patent In version 3.1

<210> 1

<211> 717

<212> ADN

45 <213> Secuencia artificial

<220>

<223> Secuencia de nucleótidos con el uso de codones humanizados que codifican la proteína verde fluorescente de *Aequorea macrodactyla* GFPxm

50

<400> 1

ES 2 399 563 T3

|   |     |
|---|-----|
| atgagcaagg gcgaggagct gttcaccggc atcgtgcccg tgctgatcga gctggacggc | 60  |
| gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc | 120 |
| aagctggaga tcaagttcat ctgcaccacc ggcaagctgc cegtgcctg gccaccctg   | 180 |
| gtgaccacct tcagctacgg catccagtgc ttcgcccgt accccgagca catgaagatg  | 240 |
| aacgacttct tcaagagcgc catgcccag ggctacatcc aggagcgcac catcttcttc  | 300 |
| caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg | 360 |
| aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatcct gggccacaag | 420 |
| ctggagtaca acttcaacag ccacaacgtg tacatcatgc ccgacaaggc caacaacggc | 480 |
| ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac | 540 |
| cactaccaga ccaacgtgcc cctggggcgc ggecccgtgc tgatcccat caaccactac  | 600 |
| ctgagcacc agaccgcat cagcaaggac cgcaacgaga cccgcgacca catggtgttc   | 660 |
| <br>  |     |
| ctggagtctc tcagcgcctg cggccacacc cacggcatgg acgagctgta caagtga    | 717 |

<210> 2

5 <211> 238

<212> PRT

<213> Aequorea macrodactyla

<400> 2

10

ES 2 399 563 T3

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Ile Val Pro Val Leu Ile  
1 5 10 15

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
20 25 30

Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
35 40 45

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe  
50 55 60

Ser Tyr Gly Ile Gln Cys Phe Ala Arg Tyr Pro Glu His Met Lys Met  
65 70 75 80

Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg  
85 90 95

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

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Met  
115 120 125

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn  
130 135 140

Phe Asn Ser His Asn Val Tyr Ile Met Pro Asp Lys Ala Asn Asn Gly  
145 150 155 160

Leu Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Gly Gly Gly Val  
165 170 175

Gln Leu Ala Asp His Tyr Gln Thr Asn Val Pro Leu Gly Asp Gly Pro  
180 185 190

Val Leu Ile Pro Ile Asn His Tyr Leu Ser Thr Gln Thr Ala Ile Ser  
195 200 205

Lys Asp Arg Asn Glu Thr Arg Asp His Met Val Phe Leu Glu Phe Phe  
210 215 220

Ser Ala Cys Gly His Thr His Gly Met Asp Glu Leu Tyr Lys  
225 230 235

# ES 2 399 563 T3

<210> 3  
<211> 717  
<212> ADN  
<213> Secuencia artificial

5

<220>  
<223> Mut 2 de la proteína GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

10

<400> 3

|    |   |            |
|----|---|------------|
| 15 | <b>atgagcaagg gcgaggagct gttcaccggc atcgtgcccg tgctgatcga gctggacggc</b>  | <b>60</b>  |
|    | <b>gacgtgcacg gccacaagtt cagcgtgcbc ggcgagggcg agggcgacgc cgactacggc</b>  | <b>120</b> |
|    | <b>aagctggaga tcaagtcat ctgcaccacc ggcaagctgc ccgtgccttg gccaccctg</b>    | <b>180</b> |
| 20 | <b>gtgaccacot tcagctacgg catccagtgc ttcgcccgt accccgagca catgaagatg</b>   | <b>240</b> |
|    | <b>aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac catcttcttc</b> | <b>300</b> |
| 25 | <b>caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctgggtg</b> | <b>360</b> |
|    | <b>aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatcct gggccacaag</b>  | <b>420</b> |
|    | <b>ctggagtaca acttcaacag ccacaacgtg tacatcatgc ccgacaaggc caacaacggc</b>  | <b>480</b> |
| 30 | <b>ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac</b>  | <b>540</b> |
|    | <b>cactaccaga ccaacgtgcc cctgggacgac ggccccgtgc tgatccccat caaccactac</b> | <b>600</b> |
| 35 | <b>ctgagcacc agaccgcat cagcaaggac cgcaacgaga cccgcgacca catggtgctc</b>    | <b>660</b> |
|    | <b>ctggagtctc tcagcgctg cggccacacc cacggcatgg acgagctgta caagtga</b>      | <b>717</b> |

<210> 4  
<211> 238  
<212> PRT  
<213> Secuencia artificial

40

<220>  
<223> Proteína Mut 2 derivada de GFPxm de Aequorea macrodactyla

45

<400> 4

50

55

60

ES 2 399 563 T3

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

ES 2 399 563 T3

<210> 5  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

5 <220>  
 <223> Proteína Mut 235 derivada de GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

10 <400> 5

```

atgagcgggg gcgaggagct gttcaccggc atcgtgcccg tgctgatcga gctggacggc      60
gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc      120
aagctggaga tcaagttcat ctgcaccacc ggcaagctgc ccgtagccctg gccaccctg      180
gtgaccacct tcagctacgg catccagtgc ttcgcccgtt accccgagca catgaagatg      240
aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac catcttcttc      300
caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg      360
aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatcct gggccacaag      420
ctggagtaca gcttcaacag ccacaacgtg tacatcatgc ccgacaaggc caacaacggc      480
ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac      540
cactaccaga ccaacgtgcc cctggggcgc ggccccgtgc tgatcccatc caaccactac      600
ctgagcacc agaccgccat cagcaaggac cgcaacgagg cccgcgacca catggtgctc      660
ctggagttct tcagcgcctg cggccacacc cacggcatgg acgagctgta caggtga      717

```

40 <210> 6  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

45 <220>  
 <223> Proteína Mut 235 derivada de GFPxm de Aequorea macrodactyla

<400> 6

50

```

Met Ser Gly Gly Glu Glu Leu Phe Thr Gly Ile Val Pro Val Leu Ile
1          5          10          15
Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu
20          25          30
Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys

```

60

ES 2 399 563 T3

|    | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |     |
|----|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5  | Thr   | Thr | Gly | Lys | Leu | Pro | Val | Pro | Trp | Pro | Thr | Leu | Val | Thr | Thr | Phe |
|    |   | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 10 | Ser   | Tyr | Gly | Ile | Gln | Cys | Phe | Ala | Arg | Tyr | Pro | Glu | His | Met | Lys | Met |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| 15 | Asn   | Asp | Phe | Phe | Lys | Ser | Ala | Met | Pro | Glu | Gly | Tyr | Ile | Gln | Glu | Arg |
|    |   |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| 20 | Thr   | Ile | Phe | Phe | Gln | Asp | Asp | Gly | Lys | Tyr | Lys | Thr | Arg | Gly | Glu | Val |
|    |   |     |     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |
| 25 | Lys   | Phe | Glu | Gly | Asp | Thr | Leu | Val | Asn | Arg | Ile | Glu | Leu | Lys | Gly | Met |
|    |   |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 30 | Asp   | Phe | Lys | Glu | Asp | Gly | Asn | Ile | Leu | Gly | His | Lys | Leu | Glu | Tyr | Ser |
|    |   | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 35 | Phe   | Asn | Ser | His | Asn | Val | Tyr | Ile | Met | Pro | Asp | Lys | Ala | Asn | Asn | Gly |
|    | 145   |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| 40 | Leu   | Lys | Val | Asn | Phe | Lys | Ile | Arg | His | Asn | Ile | Glu | Gly | Gly | Gly | Val |
|    |   |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| 45 | Gln   | Leu | Ala | Asp | His | Tyr | Gln | Thr | Asn | Val | Pro | Leu | Gly | Asp | Gly | Pro |
|    |   |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 50 | Val   | Leu | Ile | Pro | Ile | Asn | His | Tyr | Leu | Ser | Thr | Gln | Thr | Ala | Ile | Ser |
|    |   |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| 55 | Lys   | Asp | Arg | Asn | Glu | Ala | Arg | Asp | His | Met | Val | Leu | Leu | Glu | Phe | Phe |
|    |   | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 60 | Ser   | Ala | Cys | Gly | His | Thr | His | Gly | Met | Asp | Glu | Leu | Tyr | Arg |     |     |
|    | 225   |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     |     |
|    | <210> 7<br><211> 717<br><212> ADN<br><213> Secuencia artificial |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

ES 2 399 563 T3

<220>

<223> Proteína Mut 235-1 derivada de GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

5 <400> 7

atgagcgggg gcgaggagct gttcgccggc atcgtgcccg tgctgatcga gctggacggc 60

10 gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc 120

aagctggaga tcaagttcat ctgcaccacc ggcaagctgc cegtgccttg gcccaccctg 180

15 gtgaccacct tcagctacgg catccagtgc ttcgcccgtt accccgagca catgaagatg 240

aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac cctcctcttc 300

caggacgacg gcaagtataa gaccgcggc gaggtgaagt tcgagggcga caccctggtg 360

20 aaccgcatcg agctgaaggg caaggacttc aaggaggacg gcaacatcct gggccacaag 420

ctggagtaca gcttcaacag ccacaacgtg tacatcatgc ccgacaaggc caacaacggc 480

25 ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540

cactaccaga ccaacgtgcc cctggggcgac ggccccgtgc tgatcccat caaccactac 600

ctgagcacc cagaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660

30 ctggagttct tcagcgctg cggccacacc cacggcatgg acgagctgta caggtga 717

<210> 8

<211> 238

35 <212> PRT

<213> Secuencia artificial

<220>

<223> Proteína Mut 235-1 derivada de GFPxm de Aequorea macrodactyla

40 <400> 8

45 Met Ser Gly Gly Glu Glu Leu Phe Ala Gly Ile Val Pro Val Leu Ile  
1 5 10 15

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
20 25 30

50 Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
35 40 45

55 Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe  
50 55 60

60 Ser Tyr Gly Ile Gln Cys Phe Ala Arg Tyr Pro Glu His Met Lys Met  
65 70 75 80

Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg

ES 2 399 563 T3

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     |     | 95  |
| 5  | Thr | Ile | Leu | Phe | Gln | Asp | Asp | Gly | Lys | Tyr | Lys | Thr | Arg | Gly | Glu | Val |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 10 | Lys | Phe | Glu | Gly | Asp | Thr | Leu | Val | Asn | Arg | Ile | Glu | Leu | Lys | Gly | Lys |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 15 | Asp | Phe | Lys | Glu | Asp | Gly | Asn | Ile | Leu | Gly | His | Lys | Leu | Glu | Tyr | Ser |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 20 | Phe | Asn | Ser | His | Asn | Val | Tyr | Ile | Met | Pro | Asp | Lys | Ala | Asn | Asn | Gly |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| 25 | Leu | Lys | Val | Asn | Phe | Lys | Ile | Arg | His | Asn | Ile | Glu | Gly | Gly | Gly | Val |
|    |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| 30 | Gln | Leu | Ala | Asp | His | Tyr | Gln | Thr | Asn | Val | Pro | Leu | Gly | Asp | Gly | Pro |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 35 | Val | Leu | Ile | Pro | Ile | Asn | His | Tyr | Leu | Ser | Thr | Gln | Thr | Ala | Ile | Ser |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| 40 | Lys | Asp | Arg | Asn | Glu | Ala | Arg | Asp | His | Met | Val | Leu | Leu | Glu | Phe | Phe |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 45 | Ser | Ala | Cys | Gly | His | Thr | His | Gly | Met | Asp | Glu | Leu | Tyr | Arg |     |     |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     |     |

<210> 9  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <223> Proteína Mut 235-2 derivada de GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

<400> 9

|    |            |            |             |            |            |            |     |
|----|------------|------------|-------------|------------|------------|------------|-----|
| 55 | atgagcgggg | gcgaggagct | gttcaccggc  | atcgtgcccg | tgctgatcga | gctggacggc | 60  |
|    | gacgtgcacg | gccacaagtt | cagcgtgcgc  | ggcgagggcg | agggcgacgc | cgactacggc | 120 |
| 60 | aagctggaga | tcaagttcat | ctgcaccacc  | ggcaagctgc | ccgtgccttg | gaccaccctg | 180 |
|    | gtgaccacct | tcagctacgg | catccagtgc  | ttcgcccget | accccgagca | catgaagatg | 240 |
|    | aacgacttct | tcaagagcgc | catgcccagag | ggctacatcc | aggagcgcac | catcttcttc | 300 |

ES 2 399 563 T3

caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg 360  
 aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatgct gggccacaag 420  
 5 ctggagtaca gcttcaacag ccacaacgtg tacatcatgc ccgacatggc caacaacggc 480  
 ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540  
 cactaccaga ccaacgtgcc cctggggcgac ggccccgtgc tgatcccat caaccactac 600  
 10 ctgagcacc agaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660  
 ctggagtctc tcagcgctg cggccacacc cacggcatgg acgagctgta caggtga 717

<210> 10  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Proteína Mut 235-2 derivada de GFPxm de Aequorea macrodactyla

<400> 10

25 Met Ser Gly Gly Glu Glu Leu Phe Thr Gly Ile Val Pro Val Leu Ile  
 1 5 10 15

30 Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
 20 25 30

35 Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
 35 40 45

40 Thr Thr Gly Lys Leu Pro Val Pro Trp Thr Thr Leu Val Thr Thr Phe  
 50 55 60

45 Ser Tyr Gly Ile Gln Cys Phe Ala Arg Tyr Pro Glu His Met Lys Met  
 65 70 75 80

50 Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg  
 85 90 95

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

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Met  
 115 120 125

60 Asp Phe Lys Glu Asp Gly Asn Met Leu Gly His Lys Leu Glu Tyr Ser  
 130 135 140

ES 2 399 563 T3

**Phe Asn Ser His Asn Val Tyr Ile Met Pro Asp Met Ala Asn Asn Gly**  
**145 150 155 160**

5  
**Leu Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Gly Gly Gly Val**  
**165 170 175**

10  
**Gln Leu Ala Asp His Tyr Gln Thr Asn Val Pro Leu Gly Asp Gly Pro**  
**180 185 190**

15  
**Val Leu Ile Pro Ile Asn His Tyr Leu Ser Thr Gln Thr Ala Ile Ser**  
**195 200 205**

20  
**Lys Asp Arg Asn Glu Ala Arg Asp His Met Val Leu Leu Glu Phe Phe**  
**210 215 220**

25  
**Ser Ala Cys Gly His Thr His Gly Met Asp Glu Leu Tyr Arg**  
**225 230 235**

<210> 11

<211> 717

30 <212> ADN

<213> Secuencia artificial

<220>

35 <223> Proteína Mut 235-4 derivada de GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

<400> 11

**atgagcgggg gcgaggagct gttcaccggc atcgtgcccg tgctgatcga gctggacggc 60**  
**gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc 120**  
**aagctggaga tcaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gccaccctg 180**  
**gtgaccacct tcagctacgg catccagtgc ttcgcccgt accccgagca tatgaagatg 240**  
**aacgacttct tcaagagcgc catgcccag gagctacatcc aggagcgcac catcttcttc 300**  
**caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg 360**  
**aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatcct gggccacaag 420**  
**ctggagtaca gttcaacag ccacaacgtg cacatcatgc ccgacaaggc caacaacggc 480**  
**ctgaaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540**  
**cactaccaga ccaacgtgcc cctggggcgc ggcccctgctc tgatccccat caaccactac 600**  
**ctgagcacc agaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660**  
**ctggagttct tcagcgctc cggccacacc cacggcatgg acgagctgta caggtga 717**

ES 2 399 563 T3

<210> 12  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> Proteína Mut 235-4 derivada de GFPxm de Aequorea macrodactyla

<400> 12

10

Met Ser Gly Gly Glu Glu Leu Phe Thr Gly Ile Val Pro Val Leu Ile  
 1 5 10 15

15

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
 20 25 30

20

Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
 35 40 45

25

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe  
 50 55 60

30

Ser Tyr Gly Ile Gln Cys Phe Ala Arg Tyr Pro Glu His Met Lys Met  
 65 70 75 80

35

Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg  
 85 90 95

40

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

45

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Met  
 115 120 125

50

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Ser  
 130 135 140

55

Phe Asn Ser His Asn Val His Ile Met Pro Asp Lys Ala Asn Asn Gly  
 145 150 155 160

60

Leu Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Gly Gly Gly Val  
 165 170 175

Gln Leu Ala Asp His Tyr Gln Thr Asn Val Pro Leu Gly Asp Gly Pro  
 180 185 190

ES 2 399 563 T3

Val Leu Ile Pro Ile Asn His Tyr Leu Ser Thr Gln Thr Ala Ile Ser  
 195 200 205

5 Lys Asp Arg Asn Glu Ala Arg Asp His Met Val Leu Leu Glu Phe Phe  
 210 215 220

10 Ser Ala Cys Gly His Thr His Gly Met Asp Glu Leu Tyr Arg  
 225 230 235

15 <210> 13  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <223> Proteína Mut-g9 derivada de GFPxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

<400> 13

25 atgagcgggg gcgaggagct gttcgccggc atcgtgcccg tgctgatcga gctggacggc 60  
 gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc 120  
 30 aagctggaga tcaagtcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctg 180  
 gtgaccacct tcagctacgg catccagtgc ttcgcccgtt accccgagca catgaagatg 240  
 aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac catcctcttc 300  
 35 caggacgacg gcaagtataa gacccgcggc gaggtgaagt tcgagggcga caccctggtg 360  
 aaccgcatcg agctgaaggg caaggacttc aaggaggacg gcaacatcct gggccacaag 420  
 40 ctggagtaca gttcaacag ccacaacgtg tacatcatgc ccgacaaggc caacaacggc 480  
 ctggaggtga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540  
 cactaccaga ccaacgtgcc cctgggcgac ggccccgtgc tgatccccat caaccactac 600  
 45 ctgagcaccc agaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660  
 ctggagttct tcagcgctg ctgccacacc cacggcatgg acgagctgta caggtga 717

50 <210> 14  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

55 <220>  
 <223> Proteína Mut-g9 derivada de GFPxm de Aequorea macrodactyla

<400> 14

60

ES 2 399 563 T3

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

ES 2 399 563 T3

<210> 15  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

5 <220>  
 <223> Proteína 235-4G6 derivada de GFPdnaxm de Aequorea macrodactyla, secuencia de nucleótidos con el uso de codones humanizados

10 <400> 15

```

atgagcgggg gcgaggacct gttcaccggc atcgtgcccg tgctgatcga gctggacggc      60
gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc      120
15 aagctggaga tcaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctg      180
gtgaccacct tcagctacgg catccagtgc ttgcgccgct accccgagca tatgaagatg      240
20 aacgacttct tcaagagcgc catgcccgag ggctacatcc aggagcgcac catcttcttc      300
caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg      360
25 aaccgcatcg agctgaaggg catggacttc aaggaggacg gcaacatcct gggccacaag      420
ctggagtaca gtttcaacag ccacaacgtg cacatcatgc ccgacaaggc caacaacggc      480
ctgaaggtga acttcaagat ccgccacaac atcgagggcg ggggogtgca gctggccgac      540
30 cactaccaga ccaacgtgcc cctgggcgac ggccccgtgc tgatccccat caaccactac      600
ctgagcaccc agaccgccat cagcaaggac cgcaacgagg cccgcgacca catggtgctc      660
35 ctggagttct tcagcgcctg cggccacacc cacggcatgg acgagctgta caggtga      717
    
```

<210> 16  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

40 <220>  
 <223> Proteína 235-4G6 derivada de GFPxm de Aequorea macrodactyla

45 <400> 16

```

Met Ser Gly Gly Glu Asp Leu Phe Thr Gly Ile Val Pro Val Leu Ile
50 1           5           10           15

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu
55           20           25           30
    
```

60

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

ES 2 399 563 T3

<210> 17  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> tagGFP, secuencia de nucleótidos con el uso de codones humanizados

<400> 17

10

```

atgagcgggg gcgaggagct gttcgccggc atcgtgcccg tgctgatcga gctggacggc      60
gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc      120
aagctggaga tcaagttcat ctgcaccacc ggcaagctgc cegtgccttg gccaccctg      180
gtgaccacc tctgctacgg catccagtgc ttcgcccgt accccgagca catgaagatg      240
aacgacttct tcaagagcgc catgcccgag ggctacatcc aggagcgcac catcctcttc      300
caggacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctggtg      360
aaccgcatcg agctgaaggg caaggacttc aaggaggacg gcaacatcct gggccacaag      420
ctggagtaca gttcaacag ccacaacgtg tacatcatgc cgcacaaggc caacaacggc      480
ctggaggtga acttcaagac ccgccacaac atcgagggcg gcggcgtgca gctggccgac      540
cactaccaga ccaacgtgcc cctggggcgac ggcccctgctc tgatcccat caaccactac      600
ctgagcacc agaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc      660
ctggagtcct tcagcgctg ctgccacacc cacggcatgg acgagctgta caggtga      717
    
```

35

<210> 18  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

40

<220>  
 <223> Secuencia de aminoácidos de tagGFP

<400> 18

45

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

50

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
 20 25 30

55

Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
 35 40 45

60

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu  
 50 55 60

Cys Tyr Gly Ile Gln Cys Phe Ala Arg Tyr Pro Glu His Met Lys Met  
 65 70 75 80

Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg

ES 2 399 563 T3

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|    |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |     |  |
| 5  | Thr | Ile | Leu | Phe | Gln | Asp | Asp | Gly | Lys | Tyr | Lys | Thr | Arg | Gly | Glu | Val |  |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |
| 10 | Lys | Phe | Glu | Gly | Asp | Thr | Leu | Val | Asn | Arg | Ile | Glu | Leu | Lys | Gly | Lys |  |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |
| 15 | Asp | Phe | Lys | Glu | Asp | Gly | Asn | Ile | Leu | Gly | His | Lys | Leu | Glu | Tyr | Ser |  |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |
| 20 | Phe | Asn | Ser | His | Asn | Val | Tyr | Ile | Met | Pro | Asp | Lys | Ala | Asn | Asn | Gly |  |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |  |
| 25 | Leu | Glu | Val | Asn | Phe | Lys | Thr | Arg | His | Asn | Ile | Glu | Gly | Gly | Gly | Val |  |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |  |
| 30 | Gln | Leu | Ala | Asp | His | Tyr | Gln | Thr | Asn | Val | Pro | Leu | Gly | Asp | Gly | Pro |  |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |  |
| 35 | Val | Leu | Ile | Pro | Ile | Asn | His | Tyr | Leu | Ser | Thr | Gln | Thr | Ala | Ile | Ser |  |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |  |
| 40 | Lys | Asp | Arg | Asn | Glu | Ala | Arg | Asp | His | Met | Val | Leu | Leu | Glu | Ser | Phe |  |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |
| 45 | Ser | Ala | Cys | Cys | His | Thr | His | Gly | Met | Asp | Glu | Leu | Tyr | Arg |     |     |  |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     |     |  |

<210> 19

<211> 717

<212> ADN

<213> Secuencia artificial

<220>

<223> tagCFP, secuencia de nucleótidos con el uso de codones humanizados

<400> 19

|    |            |            |            |             |            |            |     |
|----|------------|------------|------------|-------------|------------|------------|-----|
| 50 | atgagcgggg | gcgaggagct | gttcgctggc | atcgtgcccg  | tgctgatcga | gctggacggc | 60  |
|    | gacgtgcacg | gccacaagtt | cagcgtgcgc | ggtgaggggag | agggcgacgc | cgactacggc | 120 |
| 55 | aagctggaga | tcaagttcat | ctgcaccacc | ggcaagctgc  | ccgtgccctg | gcccaccctg | 180 |
|    | gtgaccacce | tcgcctgggg | catccagtgc | ttcgcccgct  | accccgagca | catgaagatg | 240 |
| 60 | aacgacttct | tcaagagcgc | catgcccgag | ggctacatcc  | aggagcgcac | catccacttc | 300 |
|    | caggacgacg | gcaagtacaa | gaccgcggc  | gaggtgaagt  | tcgagggcga | caccctggtg | 360 |

ES 2 399 563 T3

aaccgctcg agctgaaggg cgagggcttc aaggaggacg gcaacatcct gggccacaag 420  
 ttggagtaca gcgccatcag cgacaacgtg tacatcatgc ccgacaaggc caacaacggc 480  
 5 ctggaggcga acttcaagat ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540  
 cactaccaga ccaacgtgcc cctggggcgat ggccccgtgc tgatccccat caaccactac 600  
 10 ctgagctgcc agtccgcat cagcaaggac cgcaacgaag cccgcgacca catggtgctc 660  
 ctggagtctc tcagcgcta ctgccacacc cacggcatgg acgagctgta ccgctaa 717

15 <210> 20  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

20 <220>  
 <223> Secuencia de aminoácidos de tagCFP

25 <400> 20

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

60

ES 2 399 563 T3

5 Ala Ile Ser Asp Asn Val Tyr Ile Met Pro Asp Lys Ala Asn Asn Gly  
 145 150 155 160

Leu Glu Ala Asn Phe Lys Ile Arg His Asn Ile Glu Gly Gly Gly Val  
 165 170 175

10 Gln Leu Ala Asp His Tyr Gln Thr Asn Val Pro Leu Gly Asp Gly Pro  
 180 185 190

15 Val Leu Ile Pro Ile Asn His Tyr Leu Ser Cys Gln Ser Ala Ile Ser  
 195 200 205

20 Lys Asp Arg Asn Glu Ala Arg Asp His Met Val Leu Leu Glu Ser Phe  
 210 215 220

25 Ser Ala Tyr Cys His Thr His Gly Met Asp Glu Leu Tyr Arg  
 225 230 235

25 <210> 21  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <223> tagYFP1, secuencia de nucleótidos con el uso de codones humanizados

35 <400> 21

atgagcgggg gcgaggagct gttcgccggc atcgtgcccg tgctgatcga gctggacggc 60

gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc 120

40 aagctggaga tcaagttcat ctgcaccacc ggcaagctgc cegtgccttg gcccaccctg 180

gtgaccaccc tcacctacgg cgtacagtgc ttcgcccgtc accccaagca catgaagatg 240

45 aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac catcctcttc 300

caagacgacg gcaagtacaa gacccgcggc gaggtgaagt tcgagggcga caccctgggtg 360

50 aaccgcatcg agctgaaggg caaggacttc aaggaggacg gcaacatcct gggccacaag 420

ctggagtaca gtttcaacag ccacaacgtc tacatcacc ccgacaaggc caacaacggc 480

ctggaggtga acttcaagac ccgccacaac atcgagggcg gcggcgtgca gctggccgac 540

55 cactaccaga ccaacgtgcc cctggggcgac ggccccgtgc tgatccccat caaccactac 600

ctgagctacc agaccgcat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660

60 ctggagtccg tcagcgctg cagccacacc cacggcatgg acgagctgta ccgctga 717

ES 2 399 563 T3

<210> 22  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> Secuencia de aminoácidos de tagYFP1

<400> 22

10

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

15

Glu Leu Asp Gly Asp Val His Gly His Lys Phe Ser Val Arg Gly Glu  
 20 25 30

20

Gly Glu Gly Asp Ala Asp Tyr Gly Lys Leu Glu Ile Lys Phe Ile Cys  
 35 40 45

25

Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu  
 50 55 60

30

Thr Tyr Gly Val Gln Cys Phe Ala Arg Tyr Pro Lys His Met Lys Met  
 65 70 75 80

35

Asn Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Ile Gln Glu Arg  
 85 90 95

40

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

45

Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Lys  
 115 120 125

50

Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Ser  
 130 135 140

55

Phe Asn Ser His Asn Val Tyr Ile Thr Pro Asp Lys Ala Asn Asn Gly  
 145 150 155 160

60

Leu Glu Val Asn Phe Lys Thr Arg His Asn Ile Glu Gly Gly Gly Val  
 165 170 175

Gln Leu Ala Asp His Tyr Gln Thr Asn Val Pro Leu Gly Asp Gly Pro  
 180 185 190

ES 2 399 563 T3

Val Leu Ile Pro Ile Asn His Tyr Leu Ser Tyr Gln Thr Ala Ile Ser  
 195 200 205

5

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

10

Ser Ala Cys Ser His Thr His Gly Met Asp Glu Leu Tyr Arg  
 225 230 235

15

<210> 23  
 <211> 717  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <223> tagYFP2, secuencia de nucleótidos con el uso de codones humanizados

<400> 23

25

atgagcgggg gcgaggagct gttcgccggc atcgtgcccg tgctgatcga gctggacggc 60

gacgtgcacg gccacaagtt cagcgtgcgc ggcgagggcg agggcgacgc cgactacggc 120

30

aagctggaga tcaagttcat ctgcaccacc ggcaagctgc ccgtgccttg gccaccctg 180

gtgaccacce tcacctacgg cgtacagtgc ttcgcccgct accccaagca catgaagatg 240

35

aacgacttct tcaagagcgc catgcccagag ggctacatcc aggagcgcac catcctcttc 300

caagacgacg gcaagtataa gacccgcggc gaggtgaagt tcgagggcga caccctggtg 360

aaccgcatcg agctgaaggg caaggacttc aaggaggacg gcaacatcct gggccacaag 420

40

ctggagtaca gcttcaacag ccacaacgtc tacatcacc cgcacaaggc caacaacggc 480

ctggaggtga acttcaagac ccgccacaac atcgagggcg gcggcggtgca gctggccgac 540

45

cactaccaga ccaacgtgcc cctggggcgac ggccccgtgc tgatccccat caaccactac 600

ctgagctacc agaccgacat cagcaaggac cgcaacgagg cccgcgacca catggtgctc 660

50

ctggagtccg tcagcgcctg cagccacacc cacggcatgg acgagctgta ccgctga 717

55

<210> 24  
 <211> 238  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Secuencia de aminoácidos de tagYFP2

60

<400> 24

ES 2 399 563 T3

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

**REIVINDICACIONES**

- 5 1. Ácido nucleico aislado, que comprende una secuencia de ácidos nucleicos que codifica una proteína fluorescente modificada genéticamente, cuya secuencia de aminoácidos tiene, como mínimo, el 80% de identidad de la secuencia con la secuencia de aminoácidos de la SEQ ID NO: 2, y cuya secuencia de aminoácidos difiere de la secuencia de aminoácidos de la SEQ ID NO: 2, como mínimo, en una sustitución de aminoácido F220L, y comprende una secuencia de aminoácidos que tiene, como mínimo, el 95% de identidad con la secuencia de aminoácidos seleccionada del grupo que comprende las SEQ ID 18, 20, 22, y 24.
- 10 2. Ácido nucleico, según la reivindicación 1, en el que la secuencia de aminoácidos de la proteína fluorescente modificada genéticamente comprende además una o más sustituciones de aminoácidos seleccionadas del grupo que comprende K3G, E6D, T9A, P58T, F99L, F99H, M128K, M128E, I136M, Y151H, N144S, K162E, K156M, T214A, G228C, G228S, y K238R.
- 15 3. Ácido nucleico, según la reivindicación 1, en el que la proteína fluorescente modificada genéticamente tiene una secuencia de aminoácidos que se selecciona del grupo que comprende las SEQ ID NO: 18, 20, 22, y 24.
4. Vector que comprende el ácido nucleico según cualquiera de las reivindicaciones 1 a 3.
- 20 5. Casete de expresión, que comprende:
- (a) una región de inicio de la transcripción que es funcional en un huésped de expresión;
- (b) el ácido nucleico según cualquiera de las reivindicaciones 1 a 3, y
- (c) una región de terminación de la transcripción funcional en dicho huésped de expresión.
- 25 6. Célula huésped aislada, o progenie de la misma, que comprende el casete de expresión, según la reivindicación 5, como parte de un elemento extracromosómico o incorporado en el genoma de una célula huésped como resultado de la introducción de dicho casete de expresión en dicha célula huésped.
- 30 7. Célula transgénica aislada, o progenie de la misma, que comprende el ácido nucleico según cualquiera de las reivindicaciones 1 a 3.
8. Proteína fluorescente modificada genéticamente, cuya secuencia de aminoácidos difiere de la SEQ ID NO: 2, como mínimo, en la sustitución de la secuencia de aminoácido F220L, y que tiene, como mínimo, el 95% de identidad con la secuencia de aminoácidos seleccionada del grupo que comprende las SEQ ID 18, 20, 22, y 24.
- 35 9. Proteína fluorescente modificada genéticamente, según la reivindicación 8, en la que la proteína fluorescente modificada genéticamente comprende una o más sustituciones de aminoácidos seleccionadas del grupo que comprende K3G, E6D, T9A, P58T, F99L, F99H, M128K, M128E, I136M, Y151H, N144S, K162E, K156M, T214A, G228C, G228S, y K238R.
- 40 10. Proteína fluorescente modificada genéticamente, según la reivindicación 8, en la que la proteína fluorescente modificada genéticamente tiene una secuencia de aminoácidos que se selecciona del grupo que comprende las SEQ ID NO: 18, 20, 22, y 24.
- 45 11. Kit que comprende, como mínimo, un ácido nucleico según la reivindicación 1.

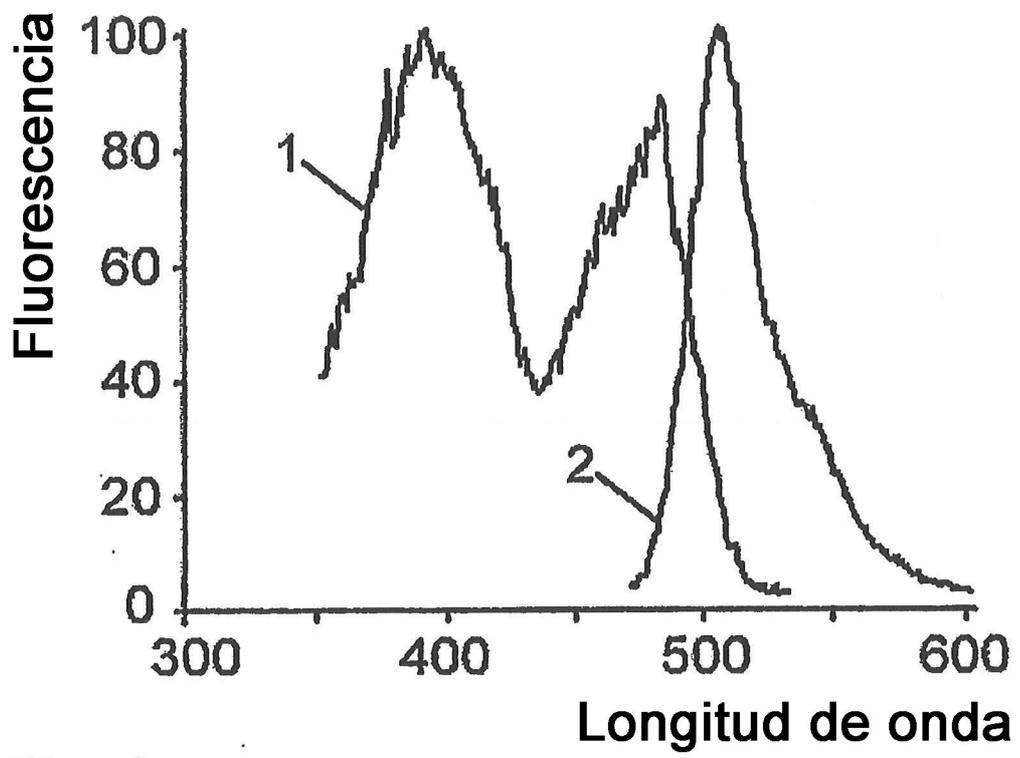
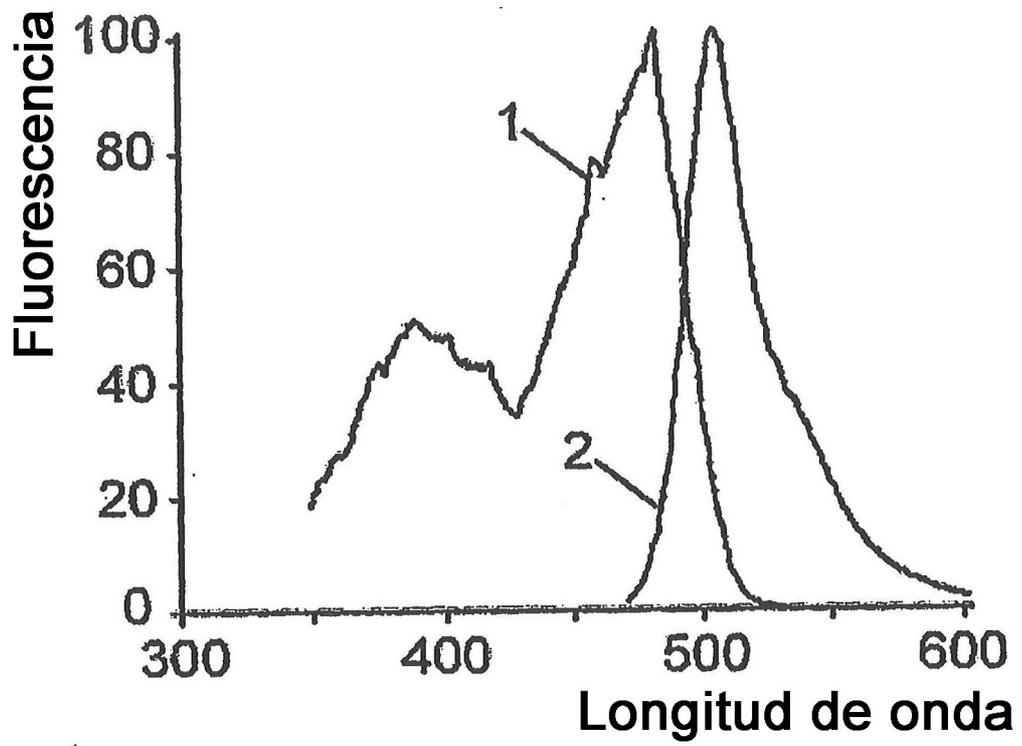


Fig. 1



**Fig. 2**

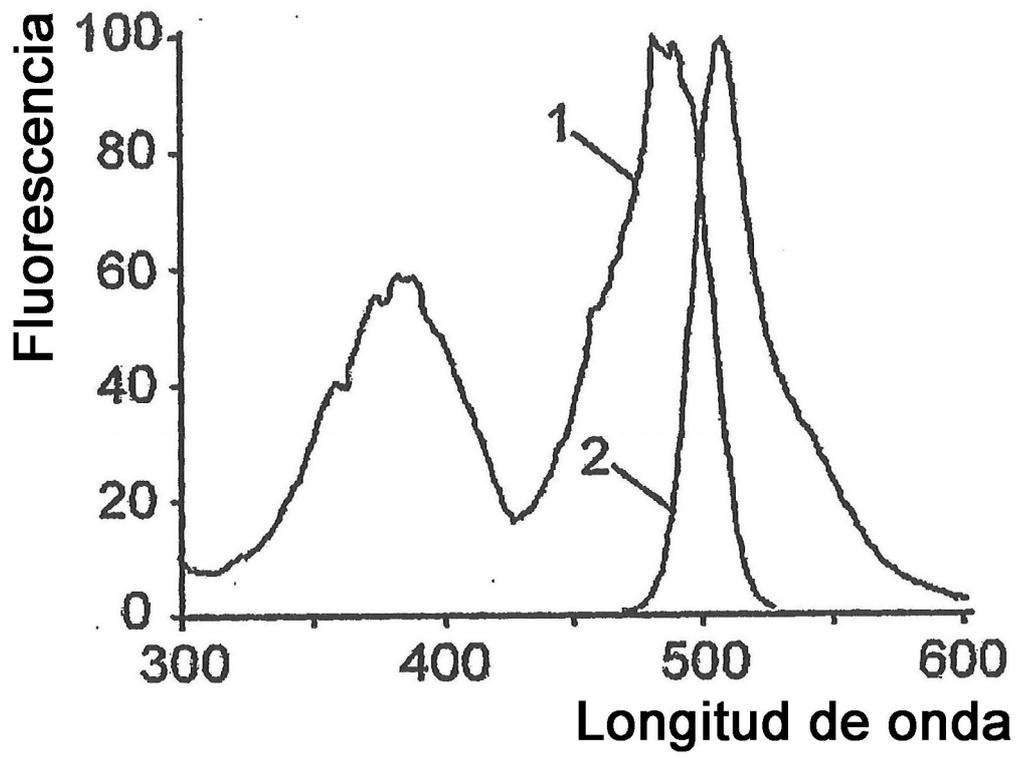


Fig. 3

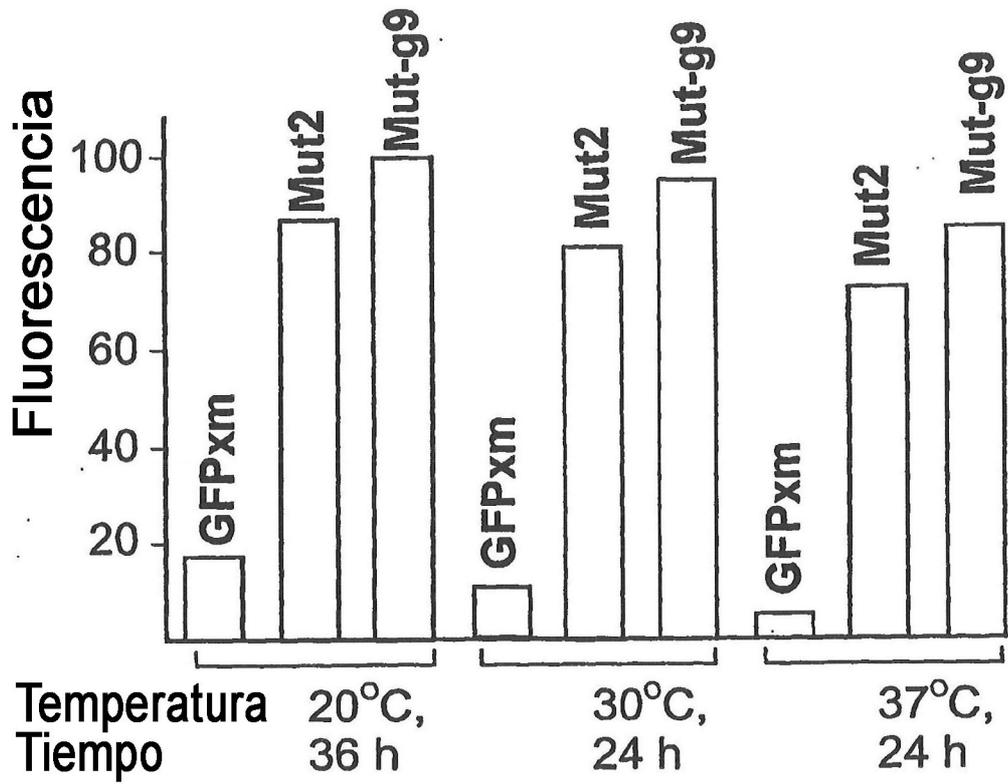


Fig. 4

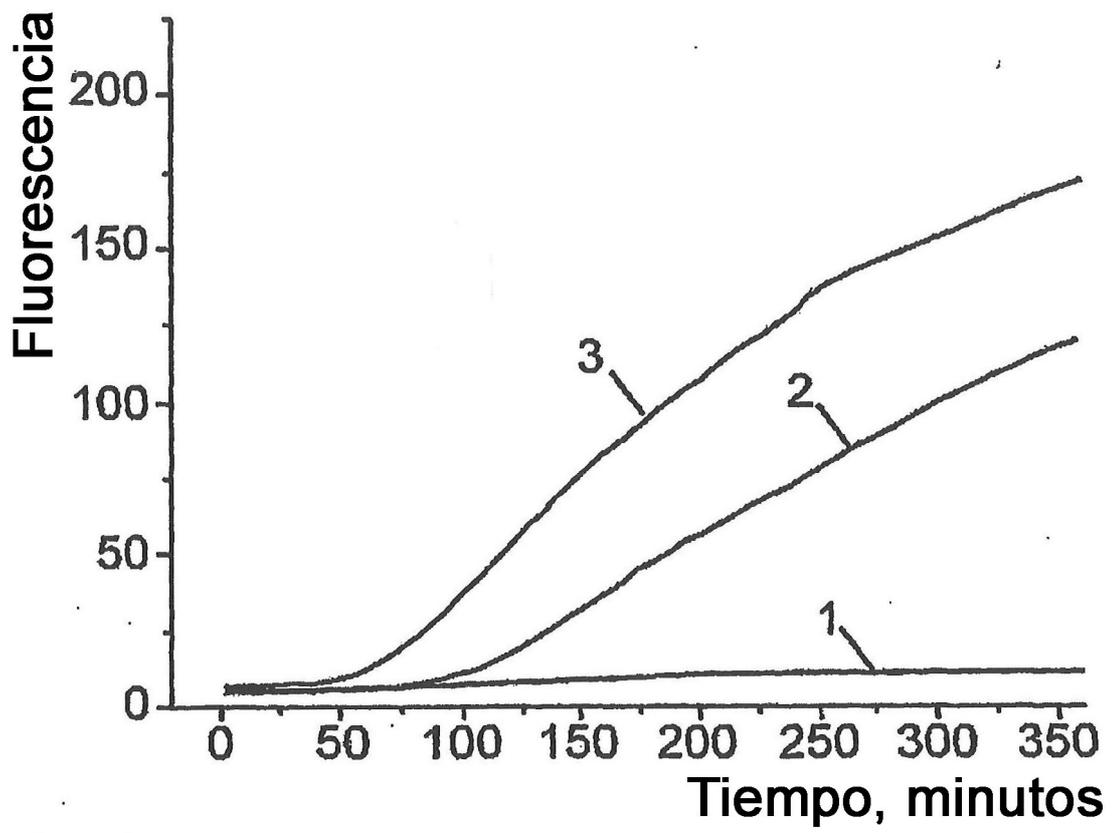


Fig. 5

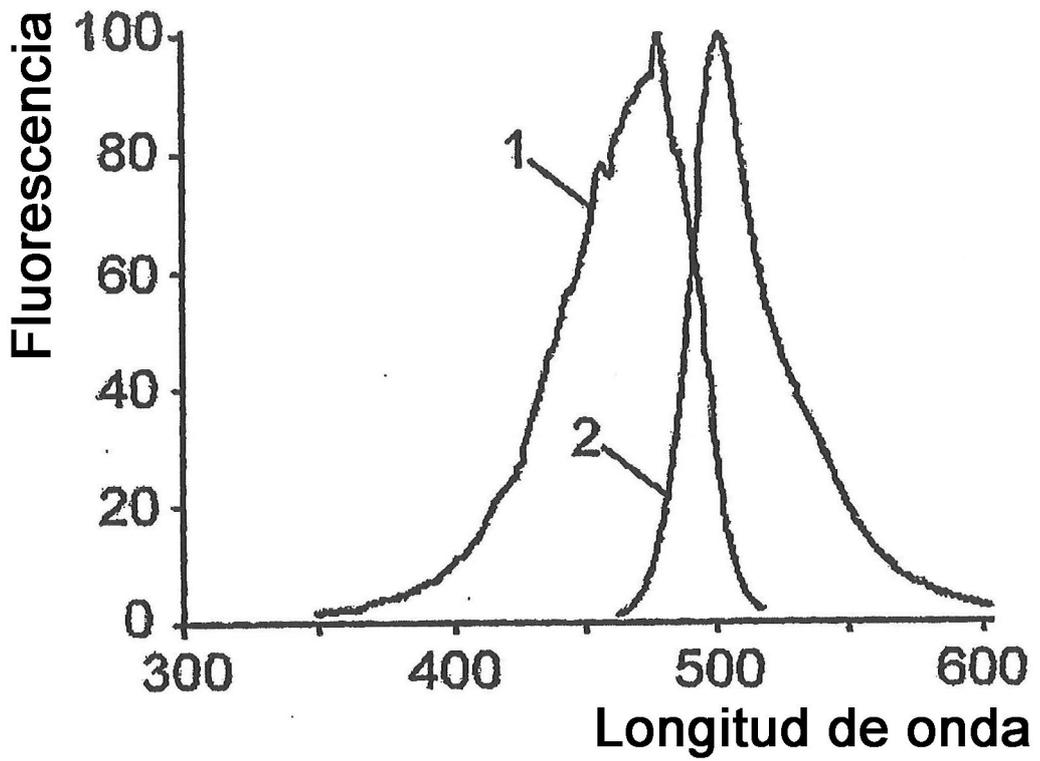


Fig. 6

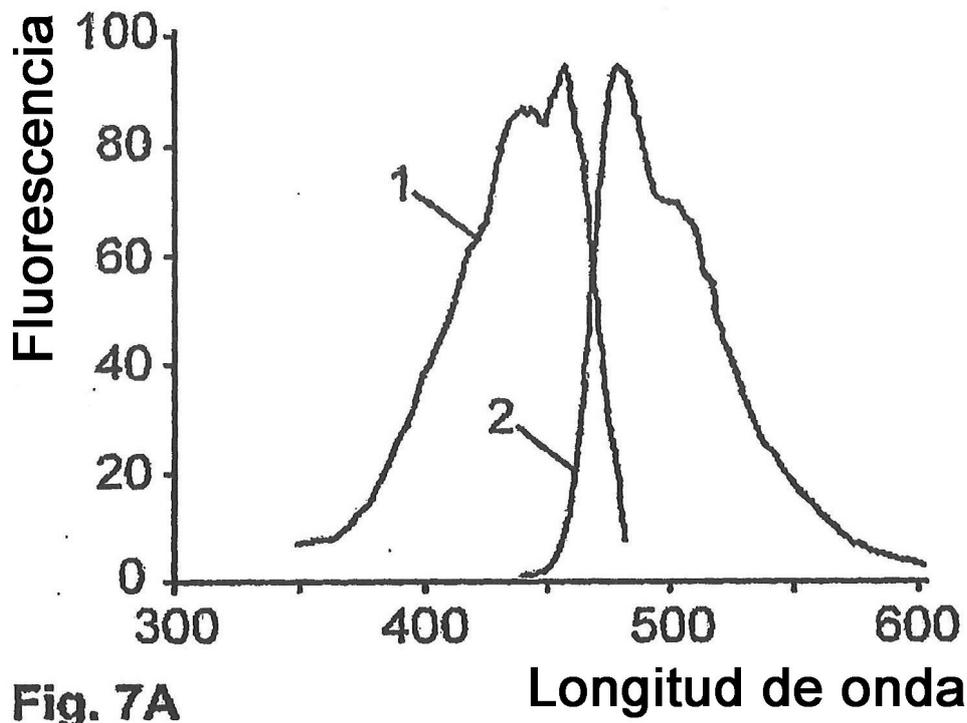


Fig. 7A

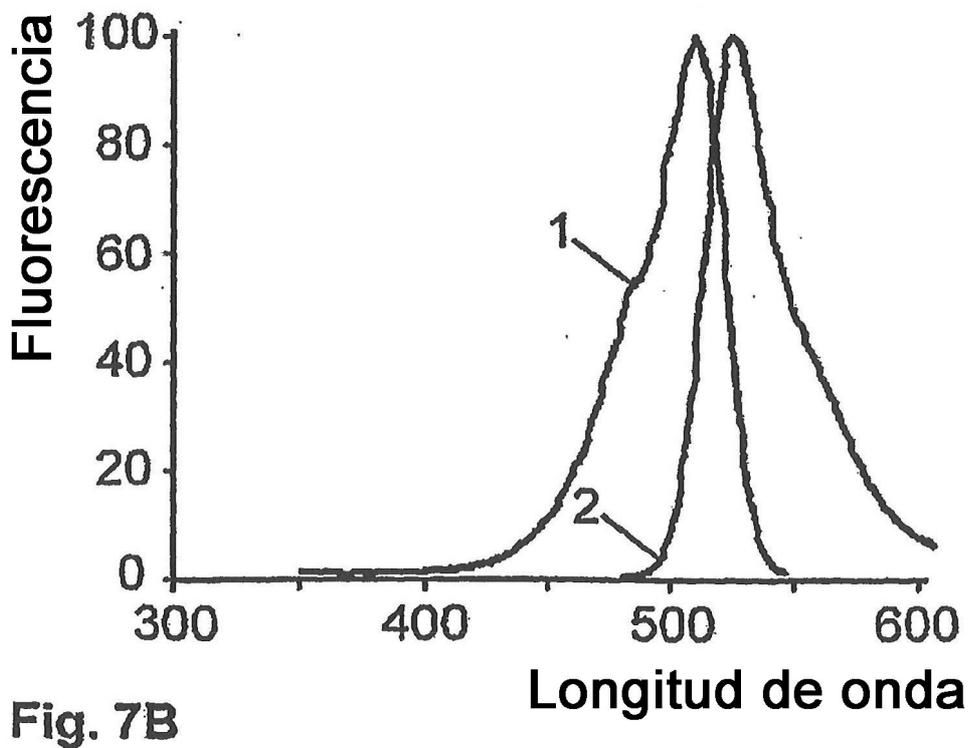


Fig. 7B