



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 353 081**

51 Int. Cl.:
C12N 15/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07705420 .3**

96 Fecha de presentación : **09.01.2007**

97 Número de publicación de la solicitud: **1994149**

97 Fecha de publicación de la solicitud: **26.11.2008**

54 Título: **Nuevas proteínas fluorescentes y métodos para usarlas.**

30 Prioridad: **25.01.2006 US 761807 P**

45 Fecha de publicación de la mención BOPI:
25.02.2011

45 Fecha de la publicación del folleto de la patente:
25.02.2011

73 Titular/es: **EVROGEN IP**
16/10 Miklukho-Maklaya Street
Moscow 117997, RU

72 Inventor/es: **Lukyanov, Sergey A. y**
Chudakov, Dmitry M.

74 Agente: **García-Cabrerizo y del Santo, Pedro María**

ES 2 353 081 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

Nuevas proteínas fluorescentes y métodos para usarlas.

5 **Antecedentes de la invención****Campo de la invención**

Esta invención se refiere en general al campo de la biología y de la química. Más particularmente, la invención se refiere a proteínas fluorescentes.

Descripción de la técnica relacionada

Las proteínas fluorescentes incluyendo la Proteína Verde Fluorescente (GFP), sus mutantes y homólogos se conocen ampliamente hoy debido a su intensivo uso como marcadores fluorescentes *in vivo* en ciencias biomédicas como describen en detalle Lippincott-Schwartz y Patterson en el documento Science (2003) 300(5616): 87-91.

Las proteínas fluorescentes son proteínas que muestran fluorescencia después de irradiarlas con luz de la apropiada longitud de onda de excitación. La característica fluorescente de estas proteínas es una que surge de la interacción de dos o más restos de aminoácidos de la proteína, y no de un único resto de aminoácido.

La GFP de la hidromedusa *Aequorea aequorea* (sinónimo *A. victoria*), descrita por Johnson *et al.*, en el documento J Cell Comp Physiol. (1962), 60: 85-104, se descubrió como parte del sistema bioluminiscente de la medusa donde la GFP desempeñaba el papel de un emisor secundario que transforma la luz azul de la fotoproteína aequorina en luz verde. El ADNc que codifica la GFP de *A. victoria* fue clonado por Prasher *et al.* (Gene (1992), 111 (2): 229-33). La conclusión fue que este gen puede expresarse de forma heteróloga en prácticamente cualquier organismo debido a la actividad única de la GFP para formar un fluoróforo por sí misma (Chalfie *et al.*, Science 263 (1994), 802-805). Este descubrimiento abre amplias perspectivas para el uso de GFP en biología celular como marca fluorescente codificada genéticamente.

La GFP se aplicó para una amplia gama de aplicaciones incluyendo el estudio de la expresión génica y la localización de proteínas (Chalfie *et al.*, Science 263 (1994), 802-805, y Heim *et al.* en Proc. Nat. Acad. Sci. (1994), 91: 12501-12504), como herramienta para visualizar orgánulos subcelulares en células (Rizzuto *et al.*, Curr. Biology (1995), 5: 635-642), para la visualización del transporte de proteínas a lo largo de la ruta secretora (Kaether and Gerdes, FEBS Letters (1995), 369: 267-271).

Se está realizando gran cantidad de investigación para mejorar las propiedades de la GFP y para producir reactivos de GFP útiles y optimizados para diversos fines de investigación. Se han desarrollado nuevas versiones de la GFP, tales como un ADN de GFP "humanizado", cuyo producto proteico presenta una mayor síntesis en células de mamífero (Haas, *et al.*, Current Biology (1996), 6: 315-324; Yang, *et al.*, Nucleic Acids Research (1996), 24: 4592-4593). Una de dichas proteínas humanizadas es la "proteína verde fluorescente mejorada" (EGFP) variante mutante de la GFP que tiene dos sustituciones de aminoácidos: F64L y S65T (Heim *et al.*, Nature 373 (1995), 663-664). Otras mutaciones de la GFP dieron como resultado versiones que emiten luz azul, cian y amarilla-verde.

A pesar de la gran utilidad de la GFP, sin embargo, otras proteínas fluorescentes con propiedades similares a o diferentes de la GFP serían útiles en la técnica. En particular, los beneficios de las nuevas proteínas fluorescentes incluyen posibilidades basadas en nuevos espectros y mejor idoneidad para una excitación más amplia. En 1999, se clonaron homólogos de la GFP a partir de especies no bioluminiscentes de Antozoos (Matz *et al.*, Nature Biotechnol. (1999), 17: 969-973). Este descubrimiento demostró que estas proteínas no son un componente necesario de la maquinaria de la bioluminiscencia. Las proteínas similares a GFP obtenidas de antozoos mostraban una gran diversidad espectral incluyendo proteínas fluorescentes cian, verdes, amarillas, rojas y cromoproteínas (CP) no fluorescentes púrpura-azul (Matz *et al.*, Bioessays (2002), 24(10): 953-959). Seguidamente, el ADNc de proteínas similares a GFP se clonó de varias medusas hidroides y Copépodos (Shagin *et al.*, Mol Biol Evol. (2004), 21(5): 841-850). Las proteínas similares a GFP ya mostraban más de 120 homólogos de GFP fluorescentes y coloreados. La similitud de estas proteínas con la GFP varía entre el 80-90% y menos del 25% de identidad en la secuencia de aminoácidos.

Las estructuras cristalinas de la GFP de tipo silvestre y la GFP mutante S65T se han resuelto y muestran que la estructura terciaria de la GFP se parece a un barril (Ormo *et al.*, 1996, Science 273: 1392-1395; Yang, *et al.*, 1996, Nature Biotech 14: 1246-1251). El barril está constituido por láminas beta en una estructura anti-paralela compacta, dentro de la cual está contenida una hélice alfa que contiene el cromóforo. Se confirmó que el cromóforo está formado por ciclación oxidativa de tres restos de aminoácidos consecutivos, como se había inferido a partir de estudios bioquímicos anteriores (Cody *et al.*, Biochemistry (1993) 32, 1212-1218). Todas las proteínas similares a GFP ensayadas compartían el mismo plegamiento en barril beta que la GFP (Ormo *et al.* Science (1996) 273: 1392-1395; Wall *et al.* Nat Struct Biol (2000), 7: 1133-1138; Yarbrough *et al.* Proc Natl Acad Sci USA (2001) 98: 462-467; Prescott *et al.* Structure (Camb) (2003), 11: 275-284; Petersen *et al.* J Biol Chem (2003), 278: 44626-44631; Wilmann *et al.* J Biol Chem (2005), 280: 2401-2404; Remington *et al.* Biochemistry (2005), 44, 202-212; Quillin *et al.* Biochemistry (2005), 44: 5774-5787).

La utilidad de las proteínas fluorescentes como herramienta en biología molecular impulsó la búsqueda de otras proteínas fluorescentes con propiedades diferentes y mejoradas, en comparación con las proteínas fluorescentes conocidas. Por lo tanto, es un objeto proporcionar nuevas proteínas fluorescentes que muestren propiedades que no estén disponibles actualmente en las proteínas fluorescentes conocidas, así como ADN que las codifica.

5

Sumario de la invención

La presente invención proporciona moléculas de ácido nucleico aisladas que codifican una nueva proteína fluorescente de *Entacmaea quadricolor* (EqFP578) y mutantes funcionales de la misma, es decir proteínas relacionadas con EqFP578. En ciertas realizaciones, un ácido nucleico de la presente invención se aísla a partir de *Entacmaea quadricolor*. En otras realizaciones, un ácido nucleico de la presente invención está manipulado genéticamente.

10

En ciertas realizaciones, los ácidos nucleicos aislados de la presente invención codifican proteína fluorescente de *Entacmaea quadricolor* de tipo silvestre de SEC ID N° 02 (EqFP578). Una secuencia de ácido nucleico ejemplar se muestra en la SEC ID N° 01.

15

En otras realizaciones, se proporciona una molécula de ácido nucleico aislado, que tiene una secuencia que hibrida específicamente con partes preseleccionadas de o con toda la cadena complementaria de la SEC ID N° 01.

20

En ciertas realizaciones, los ácidos nucleicos de la presente invención codifican una proteína fluorescente que comprende una secuencia de aminoácidos que es sustancialmente igual que o idéntica a la secuencia de la proteína EqFP578 (SEC ID N° 02). Por consiguiente, también se contempla que las secuencias de ácido nucleico aisladas que codifican variantes alélicas naturales de la SEC ID N° 01 estén dentro del alcance de la presente invención.

25

En ciertas realizaciones, los ácidos nucleicos proporcionados codifican mutantes funcionales de EqFP578 que tienen propiedades bioquímicas y/o espectrales sustancialmente similares o alteradas en comparación con la EqFP578 de tipo silvestre.

30

En ciertas realizaciones, los ácidos nucleicos aislados de la presente invención codifican una proteína fluorescente mutante que comprende una secuencia de aminoácidos sustancialmente igual que la secuencia de EqFP578 (SEC ID N° 02) y que difiere de EqFP578 (SEC ID N° 02) en al menos una sustitución de aminoácidos.

35

En realizaciones preferidas, dicha sustitución de aminoácidos se selecciona entre el grupo constituido por R32G y S131P; dicha proteína fluorescente mutante tiene un plegamiento mejorado a 37°C que EqFP578. En realizaciones preferidas, dicho mutante comprende ambas sustituciones.

40

En realizaciones preferidas, dicho mutante comprende además otras mutaciones de plegamiento, por ejemplo, seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K. Estas mutaciones mejoran el plegamiento de la proteína y la tasa de maduración del cromóforo, dando como resultado la generación más rápida de una señal fluorescente *in vivo*.

45

En otras realizaciones preferidas, los ácidos nucleicos aislados de la presente invención codifican una proteína fluorescente manipulada que tiene partes N- y/o C-terminales modificadas fuera del dominio del cromóforo de EqFP578, en la que dicha proteína posee capacidad de agregación reducida en comparación con la EqFP578 de tipo silvestre. En una realización preferida, el extremo N modificado comprende una sustitución K6T y una sustitución R231S C-terminal. En algunas realizaciones, el extremo N modificado también comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY y MGED.

50

En ciertas realizaciones, los ácidos nucleicos aislados de la presente invención codifican una proteína fluorescente funcional manipulada que comprende al menos una sustitución seleccionada entre el grupo constituido por R155E, Q159D, S173N, F192V y F194Y, en la que dicha proteína fluorescente funcional tiene capacidad de oligomerización reducida en comparación con la EqFP578 de tipo silvestre. En una realización preferida, dicha proteína fluorescente funcional también comprende una sustitución N122R que reduce adicionalmente la tendencia de la proteína a oligomerizar. En una realización preferida, dicha proteína fluorescente funcional comprende todas las sustituciones indicadas y comprende además una o más sustituciones seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115L, R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K, donde esas sustituciones mejoran el plegamiento y la intensidad de la fluorescencia de la proteína *in vivo*.

55

60

En algunas realizaciones, los ácidos nucleicos aislados de la presente invención codifican una proteína fluorescente funcional manipulada que comprende al menos una sustitución seleccionada entre el grupo constituido por H197R, S158G, N143S, N143H, N143F y N143Y, donde dicha proteína fluorescente funcional tiene espectros de excitación y emisión alterados respecto a la proteína de tipo silvestre correspondiente.

65

Las moléculas de ácido nucleico ejemplares de la invención que codifican mutantes funcionales manipulados de EqFP578 incluyen las SEC ID N° 3, 5, 7, 9, 11, 13 y 15 o codifican las SEC ID N° 4, 6, 8, 10, 12, 14 y 16.

ES 2 353 081 T3

Las moléculas de ácido nucleico que difieren de las secuencias de ácido nucleico de la presente, debido a la degeneración del código genético, o hibridan con éstas, también están dentro del alcance de la presente invención.

5 En otras realizaciones más, se proporcionan vectores que comprenden un ácido nucleico de la presente invención. Además, la presente invención proporciona casetes de expresión que comprenden un ácido nucleico de la presente invención y elementos reguladores necesarios para la expresión del ácido nucleico en la célula huésped deseada. Adicionalmente, se proporcionan células huésped, líneas celulares estables, animales transgénicos y plantas transgénicas que comprenden ácidos nucleicos, vectores o casetes de expresión de la presente invención.

10 En otras realizaciones más, se proporcionan proteínas fluorescentes funcionales de la invención que son codificadas por los ácidos nucleicos indicados anteriormente.

En ciertas realizaciones, una proteína fluorescente funcional de la presente invención comprende una secuencia de aminoácidos que es sustancialmente igual que o idéntica a la secuencia de la proteína fluorescente de *Entacmaea* 15 *quadricolor* de tipo silvestre de la SEC ID N° 02 (EqFP578). Las proteínas de interés incluyen EqFP578 de tipo silvestre y mutantes de la misma, que tienen propiedades bioquímicas y/o espectrales sustancialmente similares o alteradas en comparación con la EqFP578 de tipo silvestre.

En realizaciones preferidas, dicha sustitución de aminoácidos se selecciona entre el grupo constituido por R32G y 20 S131P, y dicha proteína fluorescente funcional tiene un plegamiento mejorado a 37°C en comparación con EqFP578. En realizaciones preferidas, dicho mutante comprende ambas sustituciones.

En algunas realizaciones, dicho mutante comprende además otras mutaciones de plegamiento, por ejemplo, seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115L, R138L, 25 G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K. Estas mutaciones mejoran el plegamiento de las proteínas y la tasa de maduración del cromóforo, dando como resultado una generación más rápida de una señal fluorescente *in vivo*.

En otras realizaciones preferidas, la proteína fluorescente manipulada tiene parte(s) N- y/o C-terminal(es) modi- 30 ficada(s), en las que dicha proteína posee una capacidad de agregación reducida en comparación con la EqFP578 de tipo silvestre. En una realización preferida, el extremo N modificado comprende una sustitución K6T y una sustitución R231S C-terminal. En algunas realizaciones, el extremo N modificado también comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY y MGED (SEC ID N° 17 y 18, respectivamente).

35 En ciertas realizaciones, la proteína fluorescente funcional manipulada de la presente invención comprende al menos una sustitución seleccionada entre el grupo constituido por R155E, Q159D, S173N y F192V, donde dicha proteína fluorescente funcional tiene una capacidad de oligomerización reducida en comparación con EqFP578 de tipo silvestre. En una realización preferida, dicha proteína fluorescente funcional también comprende una sustitución N122R 40 que reduce adicionalmente la capacidad de oligomerización de la proteína. En una realización preferida, dicha proteína fluorescente funcional comprende todas las sustituciones indicadas y comprende además una o más sustituciones seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115L, R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K, donde dichas sustituciones mejoran el plegamiento de la proteína y la intensidad de la fluorescencia *in vivo*.

45 En algunas realizaciones, la proteína fluorescente funcional manipulada de la presente invención comprende al menos una sustitución seleccionada entre el grupo constituido por H197R, S158G, N143S, N143H, N143F y N143Y, donde dicha proteína fluorescente funcional tiene espectros de excitación y emisión alterados en comparación con la proteína de tipo silvestre.

50 Las proteínas fluorescentes ejemplares de interés incluyen las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.

Adicionalmente, se proporcionan kits que comprenden ácidos nucleicos o vectores o casetes de expresión que albergan dichos ácidos nucleicos, o proteínas de la presente invención.

55 Además, se proporcionan anticuerpos que se unen específicamente a las proteínas de la presente invención o a fragmentos de las mismas.

60 Breve descripción de los dibujos

Para que las características mencionadas anteriormente de la presente invención puedan entenderse con detalle, una descripción más particular de la invención, resumida brevemente anteriormente, puede haberse realizado en referencia a realizaciones, algunas de las cuales se ilustran en los dibujos adjuntos. Debe observarse, sin embargo, que los dibujos adjuntos ilustran solamente realizaciones típicas de esta invención y no debe considerarse, por lo tanto, que limitan su 65 alcance, para que la invención pueda admitir otras realizaciones igualmente eficaces.

La figura 1 muestra un alineamiento de múltiples secuencias de EqFP578 de tipo silvestre y mutantes de la misma.

ES 2 353 081 T3

La figura 2 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para la EqFP578 de tipo silvestre.

La figura 3 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para el mutante EqFP578m1.

La figura 4 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para el mutante M1-602.

La figura 5 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para el mutante M1-637.

La figura 6 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para el mutante M1-mono1.

La figura 7 ilustra los espectros fluorescentes de excitación (línea 1) y emisión (línea 2) para el mutante nrM181-5 (antes de la fotoconversión).

Descripción detallada

Como se ha resumido anteriormente, la presente invención se refiere a moléculas de ácido nucleico que codifican la proteína fluorescente de *Entacmaea quadricolor* EqFP578 y mutantes de la misma, así como proteínas codificadas por estos ácidos nucleicos. Las moléculas de ácido nucleico de interés se aíslan de *Entacmaea quadricolor* o se diseñan mediante manipulación genética. Las proteínas de interés incluyen proteínas fluorescentes que tienen composiciones de aminoácidos mostradas en las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16. También son de interés proteínas que son sustancialmente similares a, o mutantes de, las proteínas específicas señaladas anteriormente. También se proporcionan células huésped, líneas celulares estables y organismos transgénicos que comprenden las moléculas de ácido nucleico señaladas anteriormente. Además, se proporcionan anticuerpos específicos para las particulados de la invención.

Las composiciones de proteína y ácido nucleico objeto de la invención son útiles en una serie de diferentes aplicaciones y métodos, particularmente aplicaciones de marcado de células, orgánulos celulares o proteínas. Además, las composiciones de proteína y ácido nucleico objeto de la invención son útiles en los métodos de ensayo de actividades promotoras en diversas condiciones. Finalmente, se proporcionan kits para su uso en dichos métodos y aplicaciones.

Definiciones

Diversos términos que se refieren a las moléculas biológicas de la presente invención se usan anteriormente en este documento y también en toda la memoria descriptiva y las reivindicaciones.

Como se usa en este documento, la expresión “proteína fluorescente” significa una proteína que es fluorescente; por ejemplo, puede mostrar una fluorescencia baja, media o intensa después de la irradiación con luz de la longitud de onda de excitación apropiada. La característica fluorescente de la proteína fluorescente es una que surge a partir del cromóforo en la que el cromóforo es el resultado de la ciclación autocatalítica de dos o más restos de aminoácidos en la estructura principal polipeptídica. Como tal, las proteínas fluorescentes de la presente invención no incluyen proteínas que muestran fluorescencia solamente a partir de restos que actúan por sí mismos como fluorescentes intrínsecos, es decir, triptófano, tirosina y fenilalanina.

Como se usa en este documento, el término “GFP” se refiere a la proteína verde fluorescente de *Aequorea victoria*, incluyendo versiones de la técnica anterior de la GFP manipuladas para proporcionar mayor fluorescencia o ser fluorescentes a diferentes longitudes de onda. La secuencia de la GFP de *Aequorea victoria* ha sido descrita por Prasher *et al.* (1992, Gene 111: 229-33).

Como se usa en este documento, el término “EGFP” se refiere a una variante mutante de la GFP que tiene dos sustituciones de aminoácidos: F64L y S65T (Heim *et al.*, 1995, Nature 373: 663-664).

El término “humanizado” se refiere a cambios realizados a la secuencia de ácido nucleico de la proteína fluorescente para optimizar el uso de codones para la expresión de la proteína en células de mamífero (Yang *et al.*, 1996, Nucleic Acids Research 24: 4592-4593).

Como se usa en este documento, el término “EqFP578” se refiere al ácido nucleico y a la proteína de la proteína fluorescente de *Entacmaea quadricolor* de tipo silvestre que tiene las secuencias de nucleótidos y de aminoácidos mostradas en las SEC ID N° 1 y 2, respectivamente.

Como se usa en este documento, el término “aislada” significa una molécula o una célula que está en un entorno diferente de aquel en el que la molécula o la célula se origina de forma natural.

Como se usan en este documento, los términos “mutante” o “derivados” se refieren a una proteína descrita en la presente invención, en la que uno o más aminoácidos se añaden y/o sustituyen y/o delecionan y/o insertan en el extremo N, y/o el extremo C, y/o dentro de las secuencias de aminoácidos nativas de las proteínas de la presente invención. Como se usa en este documento, el término “mutante” se refiere a una molécula de ácido nucleico que

ES 2 353 081 T3

codifica una proteína mutante. Además, el término “mutante” se refiere a cualquier versión más corta o más larga de la proteína o ácido nucleico de este documento.

5 Como se usa en este documento, “homólogo u homología” es un término usado en la técnica para describir el grado de relación de una secuencia de nucleótidos o péptidos con otra secuencia de nucleótidos o péptidos, que es determinado por el grado de identidad y/o similitud entre dichas secuencias comparadas.

10 Como se usa en este documento, una secuencia de aminoácidos o una secuencia de nucleótidos es “sustancialmente igual que” o “sustancialmente similar a” una secuencia de referencia si la secuencia de aminoácidos o la secuencia de nucleótidos tiene al menos el 85% de identidad en la secuencia con la secuencia de referencia en una ventana de comparación dada. Por lo tanto, secuencias sustancialmente similares incluyen aquellas que tienen, por ejemplo, al menos el 85% de identidad en la secuencia, al menos el 90% de identidad en la secuencia, al menos el 95% de identidad en la secuencia o al menos el 99% de identidad en la secuencia. Dos secuencias que son idénticas entre sí también son sustancialmente similares. Para los fines de esta invención, la longitud de secuencias de comparación de proteína fluorescente será generalmente de al menos 160 aminoácidos, preferentemente al menos 200 aminoácidos. Para los ácidos nucleicos, la longitud de las secuencias de comparación será generalmente de al menos 480 nucleótidos, preferentemente al menos 600 nucleótidos.

20 La identidad en la secuencia se calcula en base a una secuencia de referencia. En la técnica se conocen algoritmos para el análisis de secuencia, tales como BLAST, descrito en el documento Altschul *et al.*, J. Mol. Biol., 215, págs. 403-10 (1990). Para los fines de esta invención, comparaciones de secuencias de ácidos nucleicos o aminoácidos que se realizan con software Blast proporcionado por el *National Center for Biotechnology Information* (en la página <http://www.ncbi.nlm.nih.gov/blast>) usando un alineamiento con huecos con parámetros por defecto, pueden usarse para determinar el nivel de identidad y similitud entre secuencias de ácidos nucleicos y secuencias de aminoácidos.

25 Como se usa en este documento, la expresión “proteína fluorescente relacionada” se refiere a una proteína fluorescente que tiene una secuencia de aminoácidos sustancialmente igual en comparación con una proteína fluorescente de referencia. En general, una proteína fluorescente relacionada, en comparación con la secuencia de la proteína fluorescente de referencia, tiene una secuencia contigua de al menos aproximadamente 160 aminoácidos que comparte al menos el 85% de identidad en la secuencia con la proteína fluorescente de referencia.

30 Como se usa en este documento, la expresión “proteína relacionada con EqFP578” se refiere a la proteína EqFP578 de tipo silvestre de la SEC ID N° 2 y a mutantes funcionales de la misma, por ejemplo, que tienen secuencias de aminoácidos mostradas en las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16. Como se usa en este documento, la expresión “ácido nucleico relacionado con EqFP578” se refiere a un ácido nucleico que codifica una proteína relacionada con EqFP578 (por ejemplo, SEC ID N° 1, 3, 5, 7, 9, 11, 13 y 15). Como se usa en este documento, proteína relacionada con EqFP578 comprende una secuencia de aminoácidos que es sustancialmente igual que o idéntica a la secuencia de EqFP578 (SEC ID N° 02). Las expresiones “proteína relacionada con EqFP578” y “ácido nucleico relacionado con EqFP578” también se refieren a variantes más cortas o más largas de EqFP578 o sus mutantes y ácidos nucleicos que las codifican.

40 Como se usa en este documento, el término “funcional” implica que la secuencia de ácido nucleico o de aminoácidos es funcional para el ensayo o fin mencionado. El término “funcional” cuando se usa para describir proteínas fluorescentes significa que la proteína tiene espectros de excitación y emisión útiles (es decir, posee una fluorescencia detectable).

45 Como se usa en este documento, “propiedad bioquímica” se refiere al plegamiento y la tasa de maduración de la proteína, semivida antes de la degradación, capacidad de agregación, capacidad de oligomerización, estabilidad a y óptimo de pH o temperatura, y a otras propiedades similares.

50 Como se usa en este documento, “propiedad fluorescente” o “propiedad espectral” se refiere al coeficiente de extinción molar a una longitud de onda de excitación apropiada, la eficiencia cuántica de fluorescencia, la forma del espectro de excitación o el espectro de emisión, el máximo de longitud de onda de excitación y el máximo de longitud de onda de emisión, la proporción de amplitudes de excitación a dos longitudes de onda diferentes, la proporción de amplitudes de emisión a dos longitudes de onda diferentes, la vida útil en estado excitado, o la anisotropía de fluorescencia. Una diferencia medible en una cualquiera de estas propiedades entre la EqFP578 de tipo silvestre y la forma mutante es útil. Una diferencia medible puede determinarse como la cantidad de cualquier propiedad fluorescente cuantitativa, por ejemplo, la cantidad de fluorescencia a una longitud de onda particular, o la integral de la fluorescencia en el espectro de emisión.

60 Como se usa en este documento, “tasa de maduración” o “velocidad de maduración” se refiere a la tasa de formación de proteína fluorescente madura (es decir, proteína fluorescente capaz de producir fluorescencia) después de la traducción. La tasa de maduración puede caracterizarse por una semivida de maduración. Se ha descubierto que la maduración de una proteína fluorescente incluye dos etapas: (i) Plegamiento de la proteína que significa la formación de un barril beta de proteína con una hélice alfa central que contiene aminoácidos que formarán un cromóforo. Esta etapa se caracteriza habitualmente con una constante de tasa de aproximadamente $10^{(-2)}S^{(-1)}$ o semivida de varios segundos; (ii) Maduración del cromóforo, que es la ciclación y deshidratación de la estructura principal de proteínas. Esta fase se caracteriza habitualmente por una constante de tasa de aproximadamente $10^{(-4)}S^{(-1)}$

o semivida de aproximadamente varios minutos. Por lo tanto, esta etapa más lenta es la limitante en la maduración de proteínas verdes fluorescentes (Reid BG, Flynn GC. *Biochemistry*. 1997 V. 36(22), Págs. 6786-6791).

5 Como se usa en este documento, “agregación” se refiere a la tendencia o capacidad de una proteína expresada para formar precipitados insolubles (agregados). Hay que distinguir “agregación” de “oligomerización”. En particular, las mutaciones que reducen la agregación, por ejemplo, aumentan la solubilidad de la proteína, no reducen necesariamente la oligomerización (es decir, convierten tetrámeros en dímeros o monómeros o dímeros en monómeros).

10 Como se usa en este documento, “oligomerización” se refiere a la tendencia o capacidad de una proteína expresada para formar complejos (oligómeros) debido a la interacción específica de dos o más polipéptidos. Dicha interacción específica se produce en condiciones especificadas, por ejemplo, condiciones fisiológicas y es relativamente estable en estas condiciones. La referencia a una “capacidad” de las proteínas para oligomerizar indica que las proteínas pueden formar dímeros, trímeros, tetrámeros, o similares en condiciones especificadas. Generalmente, las proteínas fluorescentes tienen capacidad de oligomerizarse en condiciones fisiológicas aunque, como se describe en este documento, las
15 proteínas fluorescentes también pueden oligomerizar, por ejemplo, en condiciones de pH diferentes de las condiciones fisiológicas. Las condiciones en las que las proteínas fluorescentes oligomerizan o tienen tendencia a oligomerizar pueden determinarse usando métodos bien conocidos tales como filtración en gel u otra forma conocida en la técnica.

20 La expresión “unida de forma operativa” o “unida de forma funcional” o similares, cuando se usan para describir proteínas quiméricas, se refiere a secuencias de polipéptidos que están situadas en relación física y funcional entre sí. En una realización más preferida, las funciones de los componentes polipeptídicos de la molécula quimérica permanecen sin cambios en comparación con las actividades funcionales de las partes en aislamiento. Por ejemplo, una proteína fluorescente de la presente invención puede fusionarse a un socio de fusión de interés. En este caso, la molécula de fusión conserva su fluorescencia, y el polipéptido de interés conserva su actividad biológica original. En algunas reali-
25 zaciones de la presente invención, las actividades de la proteína fluorescente o la proteína de interés pueden reducirse en relación con sus actividades en aislamiento. Dichas fusiones también pueden ser útiles con la presente invención.

30 Como se usa en este documento, la expresión “hibrida específicamente” se refiere a la asociación entre dos moléculas de ácido nucleico de cadena sencilla de secuencia suficientemente complementaria para permitir dicha hibridación en condiciones predeterminadas usadas generalmente en la técnica (denominadas a veces “sustancialmente complementarias”).

35 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 cadena de clonación, cuya secuencia de nucleótidos es idéntica al ARNm y cuya secuencia se proporciona habitualmente en la lista de secuencias, así como su cadena complementaria, que se usa como plantilla para la transcripción. Como reconoce un experto en la materia, esto también incluye todas las secuencias de nucleótidos degeneradas que codifican la misma secuencia de aminoácidos. Las secuencias de nucleótidos que codifican un polipéptido incluyen secuencias que contienen intrones.

40 Como se usa en este documento, la numeración de los restos y sustituciones de aminoácidos corresponde a la numeración de restos de aminoácidos en la secuencia de la EqFP578 de tipo silvestre (SEC ID N° 2). Para las proteínas mutantes, la posición del resto o sustitución de aminoácido debe determinarse usando alineamiento de proteínas (figura 1).

45 *Moléculas de Ácido Nucleico*

50 La presente invención proporciona moléculas de ácido nucleico que codifican una proteína fluorescente EqFP578 que tiene una secuencia de aminoácidos SEC ID N° 2 y mutantes de la misma. Las moléculas de ácido nucleico que codifican variantes más cortas o más largas de la EqFP578 o sus mutantes también están dentro del alcance de la invención.

55 Las moléculas de ácido nucleico específicas de interés incluyen moléculas de ácido nucleico que codifican las siguientes proteínas fluorescentes: proteína fluorescente roja de *Entacmaea quadricolor* que tiene la secuencia de la SEC ID N° 2; y mutantes de la misma con propiedades útiles optimizadas: plegamiento de la proteína a 37°C, capacidad de agregación y/o oligomerización reducida, y características de espectros alteradas. Las secuencias de aminoácidos para estos mutantes se muestran en las SEC ID N° 4, 6, 8, 10, 12, 14 y 16. Los ácidos nucleicos mutantes específicos ejemplares de la invención se muestran en las SEC ID N° 3, 5, 7, 9, 11, 13 y 15.

60 Cada uno de estos tipos particulares de moléculas de ácido nucleico de interés se describe a continuación con más detalle en la parte experimental, más adelante.

65 Una molécula de ácido nucleico como se usa en este documento es una molécula de ADN, tal como una molécula de ADN genómico o una molécula de ADNc, o una molécula de ARN, tal como una molécula de ARNm. El término “ADNc” como se usa en este documento pretende incluir ácidos nucleicos que comparten la disposición de elementos de secuencia que se encuentra en especies de ARNm maduras nativas, donde los elementos de secuencia son exones y regiones no codificantes 5' y 3'.

ES 2 353 081 T3

Las moléculas de ácido nucleico que codifican las proteínas fluorescentes de la invención pueden sintetizarse a partir de nucleótido trifosfatos apropiados o aislarse a partir de fuentes biológicas. Ambos métodos utilizan protocolos bien conocidos en la técnica. Por ejemplo, la disponibilidad de la información de la secuencia de aminoácidos (por ejemplo SEC ID N° 2, 4, 6, 8, 10, 12, 14 ó 16) o información de la secuencia de ácidos nucleicos (por ejemplo, SEC ID N° 3, 5, 7, 9, 11, 13 ó 15) permite la preparación de moléculas de ácido nucleico aisladas de la invención mediante síntesis de oligonucleótidos. En el caso de la información de la secuencia de aminoácidos, pueden sintetizarse una serie de ácidos nucleicos que difieren entre sí debido a la degeneración del código. Los métodos para seleccionar variantes del uso de codones para huéspedes deseados se conocen bien en la técnica.

Pueden prepararse oligonucleótidos sintéticos mediante el método de fosforamidita, y las construcciones resultantes pueden purificarse de acuerdo con métodos bien conocidos en la técnica, tales como cromatografía líquida de alta resolución (HPLC) u otros métodos como se describen, por ejemplo, en el documento Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2ª Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, y según regulaciones descritas en, por ejemplo, el documento United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. Pueden sintetizarse moléculas de ADN de cadena doble largas de la presente invención mediante las siguientes fases: pueden sintetizarse varios segmentos más pequeños de complementariedad apropiada que comprenden extremos cohesivos apropiados para la unión de un segmento adyacente. Los segmentos adyacentes pueden unirse usando ADN ligasa o métodos basados en PCR.

Las moléculas de ácido nucleico que codifican las proteínas fluorescentes de la invención también pueden clonarse a partir de fuentes biológicas del phylum *Cnidaria*, preferentemente de la clase *Anthozoa*, más preferentemente de la subclase *Zoantharia*, más preferentemente de la orden *Actiniaria* y aún más preferentemente de la familia *Actiniidae*, por ejemplo, de *Entacmaea quadricolor*.

En ciertas realizaciones, una molécula de ácido nucleico de la invención es una molécula de ADN (o ADNc) que comprende un marco abierto de lectura que codifica la proteína fluorescente de *Entacmaea quadricolor* de la invención y es capaz, en condiciones apropiadas (por ejemplo, condiciones fisiológicas celulares), de expresarse como una proteína fluorescente de acuerdo con la invención. La invención también abarca ácidos nucleicos que son homólogos, sustancialmente iguales que, idénticos a, u obtenidos de los ácidos nucleicos que codifican proteínas de la presente invención. Los ácidos nucleicos objeto de la invención están presentes en un entorno diferente de su entorno 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 entorno de origen natural.

Los cambios o diferencias en la secuencia de nucleótidos entre secuencias de ácido nucleico estrechamente relacionadas pueden representar cambios de nucleótidos en la secuencia que surgen durante el transcurso de la replicación o duplicación normal en la naturaleza de la secuencia de ácido nucleico particular. Otros cambios pueden diseñarse específicamente e introducirse en la secuencia con fines específicos, tales como para cambiar un codón o secuencia de aminoácidos en una región reguladora del ácido nucleico. Dichos cambios específicos pueden realizarse *in vitro* usando diversas técnicas de mutagénesis o producirse en un organismo huésped situado en condiciones de selección particulares que inducen o seleccionan los cambios. Dichas variantes de secuencia generadas específicamente pueden denominarse “mutantes” o “derivados” de la secuencia original.

Un ácido nucleico que codifica un polipéptido relacionado con EqFP578 que es una secuencia de aminoácidos mutante, variante, derivado o alelo de la secuencia mostrada en la SEC ID N° 2 es proporcionado además por la presente invención. Un ácido nucleico que codifica dicho polipéptido puede mostrar más del 60% de homología con la secuencia codificante mostrada en la SEC ID N° 1, mayor de aproximadamente el 70% de homología, mayor de aproximadamente el 80% de homología, mayor de aproximadamente el 90% de homología o mayor de aproximadamente el 95% de homología.

Un ácido nucleico que codifica dicho polipéptido o fragmentos del mismo puede aislarse mediante cualquiera de una serie de métodos conocidos. Un fragmento de un ADNc de la presente invención puede usarse como sonda de hibridación contra una biblioteca de ADNc de un organismo diana usando condiciones de alta rigurosidad. La sonda puede ser un fragmento grande, o uno o más cebadores degenerados cortos. Los ácidos nucleicos que tienen similitud de secuencia se detectan mediante hibridación en condiciones de alta rigurosidad, por ejemplo 50°C o más (por ejemplo, 60°C o 65°C), formamida al 50%, 0,1 x SSC (cloruro sódico 15 mM/citrato sódico 1,5 mM), SDS al 0,1%. Los ácidos nucleicos que tienen una región de identidad sustancial con las secuencias proporcionadas, por ejemplo, variantes alélicas, versiones alteradas genéticamente del ácido nucleico, etc., se unen a las secuencias proporcionadas en condiciones de hibridación de alta rigurosidad. Mediante el uso de sondas, particularmente sondas marcadas de secuencias de ADN, pueden aislarse ácidos nucleicos relacionados.

Pueden generarse ácidos nucleicos mutantes o derivados en un ácido nucleico plantilla seleccionado entre los ácidos nucleicos descritos anteriormente modificando, delecionando o añadiendo uno o más nucleótidos a la secuencia plantilla, o una combinación de las mismas, para generar una variante del ácido nucleico plantilla. Las modificaciones, adiciones o deleciones pueden introducirse mediante cualquier método conocido en la técnica (véase por ejemplo los documentos Gustin *et al.*, Biotechniques (1993) 14: 22; Barany, Gene (1985) 37: 111-123; y Colicelli *et al.*, Mol. Gen. Genet. (1985) 199: 537-539, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, (1989), CSH Press, pp. 15.3-15408) incluyendo PCR propensa a error, transposiciones, mutagénesis dirigida a oligonucleótidos, PCR de ensamblaje, mutagénesis por PCR sexual, mutagénesis *in vivo*, mutagénesis por inserción de una casete, mutagénesis de

conjunto recursivo, mutagénesis de conjunto exponencial, mutagénesis dirigida, mutagénesis aleatoria, reensamblaje génico, mutagénesis saturada en sitio génico (GSSM), reensamblaje por ligamiento sintético (SLR), o una combinación de las mismas. Las modificaciones, adiciones, o deleciones también pueden introducirse mediante un método que comprende recombinación, recombinación de secuencias recursivas, mutagénesis de ADN modificado con fosfotioato, mutagénesis de plantilla que contiene uracilo, mutagénesis de cadena doble con huecos, mutagénesis de reparación de emparejamientos erróneos puntuales, mutagénesis de cepa huésped deficiente en reparación, mutagénesis química, mutagénesis radiogénica, mutagénesis por deleción, mutagénesis de selección-restricción, mutagénesis de purificación-restricción, síntesis génica artificial, mutagénesis de conjunto, creación de multímeros de ácido nucleico quiméricos y una combinación de las mismas. En algunas realizaciones, las proteínas fluorescentes codificadas por ácidos nucleicos mutantes o derivados tienen las mismas propiedades fluorescentes o bioquímicas que la proteína fluorescente de tipo silvestre. En otras realizaciones, los ácidos nucleicos mutantes o derivados codifican proteínas fluorescentes con propiedades alteradas.

Además, también se proporcionan 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 por otros codones que codifican los mismos aminoácidos. En particular, se generan variantes degeneradas de los ácidos nucleicos para aumentar su expresión en una célula huésped. En esta realización, los codones del ácido nucleico que no son preferidos o son menos preferidos en genes en la célula huésped se sustituyen por los codones sobre-representados en secuencias codificantes en genes en la célula huésped, donde dichos codones sustituidos codifican el mismo aminoácido. Las versiones humanizadas de los ácidos nucleicos de la presente invención son de particular interés. Como se usa en este documento, el término “humanizado” se refiere a cambios realizados en la secuencia de ácido nucleico para optimizar los codones para la expresión de la proteína en células de mamífero (ser humano) (Yang *et al.*, *Nucleic Acids Research* (1996) 24: 4592-4593). Véase también la Patente de Estados Unidos N° 5.795.737 que describe la humanización de proteínas. Ejemplos de variantes degeneradas de interés se describen con más detalle en la parte experimental, más adelante.

Los ácidos nucleicos que codifican variantes más cortas o más largas de la EqFP578 o mutantes de la misma también están dentro del alcance de la invención. Como se usan en este documento, estas variantes de proteína comprenden secuencias de aminoácidos de proteína relacionada con EqFP578 con el extremo C, N, o ambos modificados. En las variantes más largas, el extremo C o N de la proteína puede comprender restos de aminoácidos adicionales. En las variantes más cortas uno o más (habitualmente hasta 8, más habitualmente hasta 7 y preferentemente hasta 5) restos de aminoácidos deben eliminarse de la secuencia o sustituirse por cualesquiera otros restos de aminoácidos. Dichas modificaciones no alteran sustancialmente las propiedades fluorescentes de las proteínas, pero pueden facilitar el plegamiento de las proteínas en células huésped, reducir la capacidad de agregación o modular otras propiedades bioquímicas de las proteínas, por ejemplo, la semivida antes de la degradación. En algunas realizaciones, estas modificaciones no modifican propiedades bioquímicas de la proteína. Todos los tipos de modificaciones y mutaciones indicadas anteriormente se realizan a nivel del ácido nucleico.

Las moléculas de ácido nucleico de la invención pueden codificar todas o una parte de las proteínas objeto de la invención. Pueden obtenerse fragmentos de cadena doble o sencilla a partir de la secuencia de ADN sintetizando químicamente oligonucleótidos de acuerdo con métodos convencionales, mediante digestión con enzimas de restricción, mediante amplificación por PCR, etc. En su mayor parte, los fragmentos de ADN tendrán al menos aproximadamente 15 nucleótidos de longitud, habitualmente al menos aproximadamente 18 nucleótidos de longitud o aproximadamente 25 nucleótidos de longitud, y pueden tener al menos aproximadamente 50 nucleótidos de longitud. En algunas realizaciones, las moléculas de ácido nucleico objeto de la invención pueden tener aproximadamente 100, aproximadamente 200, aproximadamente 300, aproximadamente 400, aproximadamente 500, aproximadamente 600, aproximadamente 700 nucleótidos o más de longitud. Los ácidos nucleicos objeto de la invención pueden codificar fragmentos de las proteínas objeto de la invención o las proteínas de longitud completa; por ejemplo, los ácidos nucleicos objeto de la invención pueden codificar polipéptidos de aproximadamente 25 aminoácidos, aproximadamente 50, aproximadamente 75, aproximadamente 100, aproximadamente 125, aproximadamente 150, o aproximadamente 200 aminoácidos hasta la proteína de longitud completa.

Los ácidos nucleicos objeto de la invención pueden aislarse y obtenerse en forma sustancialmente purificada. Forma sustancialmente purificada significa que los ácidos nucleicos son al menos aproximadamente el 50% puros, habitualmente al menos aproximadamente el 90% puros y son típicamente “recombinantes”, es decir, están flanqueados por uno o más nucleótidos con los que normalmente no están asociados en un cromosoma de origen natural en su organismo huésped natural.

También se proporcionan ácidos nucleicos que codifican proteínas de fusión que comprenden una proteína de la presente invención, o fragmentos de la misma que se describen con más detalle a continuación.

También se proporcionan construcciones de vector y otras de ácido nucleico que comprenden los ácidos nucleicos objeto de la invención. Los vectores adecuados incluyen vectores virales y no virales, plásmidos, cósmidos, fagos, etc., preferentemente plásmidos, y se usan para clonar, amplificar, expresar, transferir etc., la secuencia de ácido nucleico de la presente invención en el huésped apropiado. La elección del vector apropiado está dentro de las competencias técnicas, y muchos de dichos vectores están disponibles en el mercado. Para preparar las construcciones, el ácido nucleico parcial o de longitud completa se inserta en un vector típicamente por medio de unión con ADN ligasa a un sitio de enzima de restricción escindido en el vector. Como alternativa, la secuencia de nucleótidos deseada

puede insertarse mediante recombinación homóloga *in vivo*, típicamente uniendo regiones de homología al vector en los flancos de la secuencia de nucleótidos deseada. Se añaden regiones de homología mediante ligamiento de oligonucleótidos, o mediante reacción en cadena de la polimerasa usando cebadores que comprenden tanto la región de homología como una parte de la secuencia de nucleótidos deseada, por ejemplo.

5 También se proporcionan casetes o sistemas de expresión usados, entre otras cosas, para la producción de las proteínas cromogénicas o fluorescentes objeto de la invención o proteínas de fusión de las mismas o para replicación de las moléculas de ácido nucleico objeto de la invención. La casete de expresión puede existir como un elemento extracromosómico o puede estar integrada en el genoma de la célula como resultado de la introducción de dicha casete
10 de expresión en la célula. Para la expresión, el producto génico codificado por el ácido nucleico de la invención se expresa en cualquier sistema de expresión conveniente, incluyendo, por ejemplo, sistemas bacterianos, de levadura, de insecto, de anfibio, o de mamífero. En el vector de expresión, un ácido nucleico objeto de la invención está unido de forma operativa a una secuencia reguladora que puede incluir promotores, potenciadores, terminadores, operadores, represores e inductores. Los métodos para preparar casetes o sistemas de expresión capaces de expresar el producto deseado son conocidos por un experto en la materia.

Las líneas celulares, que expresan de forma estable las proteínas de la presente invención, pueden seleccionarse mediante los métodos conocidos en la técnica (por ejemplo, la co-transfección con un marcador seleccionable tal como dhfr, gpt, neomicina, higromicina permite la identificación y el aislamiento de las células transfectadas que contienen
20 el gen integrado en un genoma).

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

Cuando cualquiera de las células huésped señaladas anteriormente, u otras células u organismos huésped apropiados se usan para replicar y/o expresar los ácidos nucleicos de la invención, el ácido nucleico replicado, proteína expresada o polipéptido resultante, está dentro del alcance de la invención como producto de la célula u organismo huésped. El producto puede recuperarse mediante medios apropiados conocidos en la técnica.

También se proporcionan pequeños fragmentos de ADN de los ácidos nucleicos objeto de la invención, que son útiles como cebadores para PCR, sondas de cribado de hibridación, etc. Los fragmentos de ADN más grandes son útiles para la producción del polipéptido codificado, como se ha descrito anteriormente. Sin embargo, para su uso
35 en reacciones de amplificación geométrica, tales como PCR geométrica, se usará un par de pequeños fragmentos de ADN, es decir, cebadores. La composición exacta de las secuencias de cebador no es crítica para la invención, pero para la mayoría de las aplicaciones, los cebadores hibridarán con la secuencia objeto de la invención en condiciones rigurosas, como se conoce en la técnica. Es preferible seleccionar un par de cebadores que generarán un producto de amplificación de al menos aproximadamente 50 nucleótidos, preferentemente al menos aproximadamente 100
40 nucleótidos y puede extenderse hasta la secuencia completa del ácido nucleico. Generalmente se conocen algoritmos para la selección de secuencias de cebadores, y están disponibles en paquetes de software comerciales. Los cebadores de amplificación hibridan con cadenas complementarias de ADN e iniciarán el cebado uno hacia el otro.

Las moléculas de ácido nucleico de la presente invención también pueden usarse para identificar la expresión de un gen en una muestra biológica. La manera en la que se sondan células en busca de la presencia de secuencias de nucleótidos particulares, tales como ADN o ARN genómico, está bien establecida en la técnica. En resumen, se aísla ADN o ARNm de una muestra de células. El ARNm puede amplificarse mediante RT-PCR, usando transcriptasa inversa para formar una cadena de ADN complementaria, seguida de amplificación por reacción en cadena de la polimerasa usando cebadores específicos para las secuencias de ADN objeto de la invención. Como alternativa, la muestra
50 de ARNm se separa mediante electroforesis en gel, se transfiere a un soporte adecuado, por ejemplo, nitrocelulosa, nylon, etc., y a continuación se sondea con un fragmento del ADN objeto de la invención como sonda. Otras técnicas, tales como ensayos de ligamiento de oligonucleótidos, hibridaciones *in situ*, e hibridación con sondas de ADN dispuestos en series en un chip sólido también pueden usarse. La detección de ARNm que hibrida con la secuencia objeto de la invención es indicativa de expresión génica en la muestra.

55 Proteínas

La presente invención también proporciona proteínas fluorescentes, derivados y mutantes de las mismas incluyendo
60 proteínas de longitud completa, así como partes o fragmentos de las mismas. También se proporcionan variantes de la proteína de *Entacmaea quadricolor* de origen natural (SEC ID N° 2), donde dichas variantes son homólogas o sustancialmente iguales que la proteína de origen natural, y mutantes de las proteínas de origen natural, como se describe con más detalle a continuación.

Las proteínas objeto de la invención son fluorescentes, es decir, poseen fluorescencia detectable. En muchas realizaciones, las proteínas objeto de la invención poseen fluorescencia en el rojo o rojo lejano, es decir, tienen un máximo de absorbancia que varía entre aproximadamente 450 nm y 700 nm, habitualmente entre aproximadamente 470 nm y 650 nm y más habitualmente entre aproximadamente 500 y 600 nm, por ejemplo, entre 550 y 595 nm; mientras

ES 2 353 081 T3

que los espectros de emisión de las proteínas objeto de la invención varían entre aproximadamente 530 y 700 nm, habitualmente entre aproximadamente 550 nm y 670 nm y más habitualmente entre aproximadamente 560 y 650 nm, por ejemplo, entre 574 y 637 nm. En otras realizaciones, las proteínas objeto de la invención tienen un máximo de absorbancia que varía entre aproximadamente 350 y 500 nm, habitualmente entre aproximadamente 370 nm y 450 nm y más habitualmente entre aproximadamente 390 y 420 nm, por ejemplo, en 400 nm; mientras que los espectros de emisión de las proteínas objeto de la invención varían típicamente entre aproximadamente 400 y 530 nm, habitualmente entre aproximadamente 420 nm y 510 nm, por ejemplo aproximadamente 470 nm.

Las proteínas objeto de la invención generalmente tienen un coeficiente de extinción máximo que varía entre aproximadamente 30.000 y 150.000 y habitualmente entre aproximadamente 60.000 y 120.000, por ejemplo, 90.000 y 120.000.

Las proteínas objeto de la invención típicamente varían en longitud entre aproximadamente 150 y 300 aminoácidos y habitualmente entre aproximadamente 200 y 300 restos de aminoácidos, y generalmente tienen un peso molecular que varía entre aproximadamente 15 y 35 kDa, habitualmente entre aproximadamente 17,5 y 32,5 kDa.

En ciertas realizaciones, las proteínas objeto de la invención son brillantes, donde brillantes significa que las proteínas fluorescentes pueden detectarse mediante métodos comunes (por ejemplo, cribado visual, espectrofotometría, espectrofluorometría, microscopía fluorescente, mediante máquinas de FACS, etc.). El brillo de fluorescencia de una proteína fluorescente particular se determina mediante su rendimiento cuántico multiplicado por el coeficiente de extinción máximo y dividido por 1000. En muchas realizaciones, las proteínas objeto de la invención tienen un brillo de fluorescencia desde aproximadamente 10 a 90, habitualmente de aproximadamente 40 a 80, y más habitualmente de aproximadamente 50 a 75.

En algunas realizaciones, las proteínas objeto de la invención son fotoactivables, fotoinactivables o fotodesplazables, es decir, cambian propiedades fluorescentes bajo irradiación de luz de una cierta longitud de onda e intensidad. Por ejemplo, pueden poseer fluorescencia azul (con emisión a aproximadamente 440-500 nm) antes de la irradiación y fluorescencia roja (con emisión a aproximadamente 540-650 nm) después de la irradiación con luz UV-violeta (por ejemplo aproximadamente 400 nm). En algunas realizaciones dichas proteínas pueden poseer fluorescencia roja antes de la irradiación que se reduce después de la irradiación con luz (por ejemplo luz UV o azul). En otra realización, dichas proteínas pueden ser no fluorescentes (cromogénicas) antes pero fluorescentes después de la irradiación con luz UV o azul.

En ciertas realizaciones, las proteínas objeto de la invención se pliegan rápidamente después de la expresión en la célula huésped. Por se pliegan rápidamente, se entiende que las proteínas alcanzan su estructura terciaria que da origen a su cualidad fluorescente en un corto periodo de tiempo. En estas realizaciones, el tiempo de plegamiento de la mitad de las proteínas generalmente no supera aproximadamente 48 horas, habitualmente no supera aproximadamente 24 horas y más habitualmente no supera aproximadamente 12 horas (por ejemplo el tiempo de plegamiento de la mitad de las proteínas puede ser de 3 horas).

Las proteínas específicas de interés incluyen la proteína roja fluorescente EqFP578 de *Entacmaea quadricolor* del phylum *Cnidaria*, preferentemente de la clase *Anthozoa*, más preferentemente de la subclase *Zoantharia*, más preferentemente del orden *Actiniaria* y aún más preferentemente de la familia *Actiniidae* (SEC ID N° 2); y mutantes funcionales de la misma.

Los mutantes pueden conservar propiedades biológicas de las proteínas de tipo silvestre (por ejemplo, de origen natural), o pueden tener propiedades biológicas que difieren de las proteínas de tipo silvestre. La expresión "propiedad biológica" de las proteínas de la presente invención se refiere a, aunque sin limitación, propiedades espectrales, tales como máximo de absorbancia, máximo de emisión, coeficiente de extinción máximo, brillo (por ejemplo, en comparación con la proteína de referencia), y similares; propiedades bioquímicas, tales como estabilidad *in vivo* y/o *in vitro* (por ejemplo, semivida); velocidad de maduración, tendencia a la agregación y tendencia a la oligomerización y otras propiedades semejantes (en comparación con la proteína de referencia). Las mutaciones incluyen cambios de aminoácidos individuales, deleciones o inserciones de uno o más aminoácidos, truncamientos o extensiones N-terminales, truncamientos o extensiones C-terminales y similares.

Los mutantes pueden generarse usando técnicas convencionales de biología molecular como se describen en detalle en la sección "Moléculas de ácido nucleico" anteriormente. Dada la guía proporcionada en los Ejemplos, y usando técnicas convencionales, los expertos en la materia pueden generar fácilmente una amplia variedad de mutantes adicionales y ensayar si una propiedad biológica (por ejemplo, bioquímica, espectral, etc.) ha sido alterada. Por ejemplo, la intensidad de fluorescencia puede medirse usando un espectrofotómetro a diversas longitudes de onda de excitación.

Las proteínas de interés también pueden modificarse usando técnicas convencionales que incluyen editado del ARN, modificaciones químicas, modificaciones postraduccionales y postranscripcionales y similares. Por ejemplo, pueden generarse derivados de las proteínas de interés mediante procesos tales como fosforilación alterada, o glucosilación, o acetilación, o lipidación, o mediante diferentes tipos de maduración por escisión y similares.

ES 2 353 081 T3

En ciertas realizaciones, dicho mutante comprende al menos una sustitución de aminoácidos que mejoró el plegamiento de la proteína a 37°C, donde dicha sustitución se selecciona entre el grupo constituido por R32G y S131P. En ciertas realizaciones, dicho mutante comprende ambas sustituciones.

5 En ciertas realizaciones, dicho mutante comprende además una o más sustituciones seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, 1115L1 R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K, donde dichas sustituciones de plegamiento potencian el plegamiento de la proteína y la tasa de maduración del cromóforo a 37°C.

10 En ciertas realizaciones, dicho mutante tiene partes N- y/o C-terminales modificadas y comprende al menos una sustitución de entre K6T y R231S, o ambas, donde dichas sustituciones reducen la capacidad de agregación del mutante en comparación con la proteína de tipo silvestre correspondiente.

15 En algunas realizaciones, dicho mutante es una variante más larga de EqFP578, que comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY (SEC ID N° 17) o MGED (SEC ID N° 18), donde dicha secuencia de aminoácidos reduce la capacidad de agregación del mutante.

20 En ciertas realizaciones, el mutante fluorescente funcional de interés comprende al menos una sustitución seleccionada entre el grupo constituido por R155E, Q159D, S173N, F192V y F194Y, donde dicho mutante tiene una capacidad de oligomerización reducida en comparación con la EqFP578 de tipo silvestre. En una realización preferida, dicha proteína fluorescente funcional también comprende una sustitución N122R que reduce adicionalmente la capacidad de oligomerización de la proteína. En una realización preferida, dicho mutante comprende todas las sustituciones indicadas y comprende además una o más sustituciones seleccionadas entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, 1115L1 R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, 25 H193Y, M216V, K220R y R231K, donde esas sustituciones potencian el plegamiento de la proteína y la intensidad de fluorescencia de la proteína *in vivo*.

30 En algunas realizaciones, la proteína fluorescente funcional manipulada de la presente invención comprende al menos una sustitución seleccionada entre el grupo constituido por H197R, S158G, N143S, N143H, N143F o N143Y, donde dicha proteína fluorescente funcional tiene espectros de excitación y emisión alterados con comparación con la proteína de tipo silvestre correspondiente.

35 Los mutantes específicos de interés incluyen proteínas fluorescentes funcionales manipuladas que comprenden composiciones de aminoácidos de las SEC ID N° 4, 6, 8, 10, 12, 14 y 16 y se describen con más detalle en la parte experimental, a continuación.

40 También se proporcionan proteínas que son sustancialmente iguales que las proteínas específicas proporcionadas, donde sustancialmente iguales significa que la proteína tiene una identidad en la secuencia de aminoácidos con la secuencia de la proteína de tipo silvestre de al menos aproximadamente el 85% de identidad en la secuencia, habitualmente al menos aproximadamente el 90% y más habitualmente al menos aproximadamente el 95%, (por ejemplo 95%; 96%; 97%; 98%; 99% o 100% de identidad en la secuencia).

45 Las proteínas de la presente invención están presentes en un entorno no de origen natural, por ejemplo, están separadas de su entorno de origen natural o son recombinantes. Las proteínas de la presente invención pueden estar presentes en forma aislada, por lo que se entiende que la proteína está sustancialmente libre de otras proteínas y otras moléculas biológicas de origen natural, tales como oligosacáridos, ácidos nucleicos y fragmentos de los mismos, y similares, donde la expresión “sustancialmente libre” en este caso significa que menos del 70%, habitualmente menos del 60% y más habitualmente menos del 50% de la composición que contiene la proteína aislada es alguna otra molécula biológica de origen natural. En ciertas realizaciones, las proteínas están presentes en forma sustancialmente 50 purificada, donde por “forma sustancialmente purificada” se entiende al menos el 95%, habitualmente al menos el 97% y más habitualmente al menos el 99% pura.

55 También se proporcionan fragmentos de la proteína de origen natural así como de las proteínas mutante y derivada descritas anteriormente. Los fragmentos biológicamente activos y/o fragmentos correspondientes a dominios funcionales, y similares son de particular interés. Los fragmentos de interés son polipéptidos que tienen típicamente al menos aproximadamente 30 aminoácidos de longitud, habitualmente al menos aproximadamente 50 aminoácidos de longitud, preferentemente de al menos aproximadamente 75 ó 100 aminoácidos de longitud y pueden tener hasta 300 aminoácidos de longitud o más, pero habitualmente no superarán aproximadamente 250 aminoácidos de longitud, donde el fragmento tendrá un tramo de aminoácidos que es idéntico a la proteína objeto de la invención de al menos 60 aproximadamente 25 aminoácidos, y habitualmente al menos aproximadamente 45 aminoácidos, y en muchas realizaciones al menos aproximadamente 50 aminoácidos de longitud. En algunas realizaciones, los polipéptidos objeto de la invención tienen aproximadamente 25 aminoácidos, aproximadamente 50, aproximadamente 75, aproximadamente 100, aproximadamente 125, aproximadamente 150, aproximadamente 200, o aproximadamente 250 aminoácidos de longitud, hasta la longitud completa de la proteína. En algunas realizaciones, un fragmento de proteína conserva todas 65 o sustancialmente todas las propiedades específicas de la proteína de tipo silvestre o mutantes específicos de la misma.

Las proteínas y polipéptidos objeto de la invención pueden obtenerse a partir de fuentes de origen natural o producirse de forma sintética. Por ejemplo, las proteínas de tipo silvestre pueden obtenerse a partir de fuentes biológicas

que expresan las proteínas, por ejemplo, *Entacmaea quadricolor*, tal como la específica enumerada anteriormente. Las proteínas objeto de la invención también pueden obtenerse de medios sintéticos, por ejemplo, expresando una secuencia codificante de ácido nucleico recombinante que codifica la proteína de interés en un huésped adecuado, como se ha descrito anteriormente. Pueden emplearse cualesquiera procedimientos de purificación de proteínas adecuados, donde metodologías de purificación de proteínas adecuadas se describen en el documento Guide to Protein Purification, (Deuthser ed., Academic Press, 1990). Por ejemplo, puede prepararse un lisado a partir de la fuente original y purificarse usando HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía por afinidad y similares.

También se proporcionan proteínas de fusión que comprenden una proteína de la presente invención, o fragmentos de la misma, fusionado, por ejemplo, a una secuencia de degradación, una secuencia de localización subcelular (por ejemplo, señal de localización nuclear, señal de reconocimiento de peroxisomas, secuencia de reconocimiento del aparato de Golgi, secuencia de reconocimiento de mitocondrias, etc.), un péptido señal, o cualquier proteína o polipéptido de interés. Las proteínas de fusión pueden comprender por ejemplo, una proteína fluorescente del polipéptido de la presente invención y un segundo polipéptido ("el socio de fusión") fusionados en marco en el extremo N- y/o extremo C- de la proteína fluorescente. Los socios de fusión incluyen, aunque sin limitación, polipéptidos que pueden unirse a anticuerpos específicos para el socio de fusión (por ejemplo, marcas epitópicas), 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. En dichas proteínas de fusión, el socio de fusión generalmente no está asociado de forma natural a la parte de proteína fluorescente de la proteína de fusión, y típicamente no es una proteína fluorescente de *Entacmaea quadricolor* de la presente invención o derivado/fragmento de la misma; es decir, no se encuentra en la especie *Entacmaea quadricolor*.

También se proporcionan anticuerpos que se unen específicamente a las proteínas fluorescentes de la presente invención. Los anticuerpos adecuados pueden producirse usando las técnicas conocidas en la técnica. Por ejemplo, pueden obtenerse anticuerpos policlonales como se describe en el documento (Harlow and Lane Antibodies: A Laboratory Manual, (1988) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York) y pueden obtenerse anticuerpos monoclonales como se describe en el documento (Goding Monoclonal Antibodies: Principles and Practice: Production and Application of Monoclonal Antibodies in Cell Biology, Biochemistry and Immunology; 3ª edición, (1996) Academic Press). Los anticuerpos quiméricos incluyendo anticuerpos humanizados así como anticuerpos de cadena sencilla y fragmentos de anticuerpo tales como Fv, F(ab')₂ y Fab también son de interés.

Transformantes

Los ácidos nucleicos de la presente invención pueden usarse para generar transformantes incluyendo organismos transgénicos o modificaciones génicas específicas de sitio en líneas celulares. Las células transgénicas de la presente invención incluyen uno o más ácidos nucleicos de acuerdo con la presente invención presentes como un transgén. Para los fines de la invención, puede usarse cualquier célula huésped adecuada incluyendo células huésped procariontas (por ejemplo, *Escherichia coli*, *Streptomyces* sp., *Bacillus subtilis*, *Lactobacillus acidophilus*, etc.) o eucariotas. Los organismos transgénicos de la presente invención pueden ser procariontas o un organismo eucariota incluyendo bacterias, cianobacterias, hongos, plantas y animales, en las que una o más de las células del organismo contienen ácido nucleico heterólogo de la presente invención introducido por medio de intervención humana, tal como mediante técnicas transgénicas bien conocidas en la técnica.

El ácido nucleico aislado de la presente invención puede introducirse en el huésped mediante métodos conocidos en la técnica, por ejemplo infección, transfección, transformación o transconjugación. Técnicas para transferir las moléculas de ácido nucleico (es decir, ADN) a dichos organismos se conocen ampliamente y se proporcionan en referencias tales como Sambrook *et al.* (Molecular Cloning: A Laboratory Manual, 3ª Ed., (2001) Cold Spring Harbor Press, Cold Spring Harbor, NY).

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

En otra realización, el organismo transgénico puede ser un hongo, por ejemplo una levadura. La levadura se usa ampliamente como vehículo para expresión génica heteróloga (por ejemplo véase el documento Goodey *et al* Yeast biotechnology, D R Berry *et al*, eds, (1987) Allen and Unwin, Londres, páginas 401-429) y por King *et al* Molecular and Cell Biology of Yeasts, E F Walton and G T Yarronton, eds, Blackie, Glasgow (1989) págs. 107-133). Varios tipos de vectores de levadura están disponibles, incluyendo vectores integradores, que requieren recombinación con el genoma del huésped para su mantenimiento, y para replicar de forma autónoma vectores plasmídicos.

Otro organismo huésped es un animal. Los animales transgénicos pueden obtenerse mediante técnicas transgénicas bien conocidas en la técnica y proporcionadas en referencias tales como Pinkert, Transgenic Animal Technology: a Laboratory Handbook, 2ª edición (2003) San Diego: Academic Press; Gersenstein y Vintersten, Manipulating the Mouse Embryo: A Laboratory Manual, 3ª ed, (2002) Nagy A. (Ed), Cold Spring Harbor Laboratory; Blau *et al.*, Laboratory Animal Medicine, 2ª Ed., (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 by Alexandra L. Joyner (Ed.)

Oxford University Press; 2ª edición (2000). Por ejemplo, pueden obtenerse animales transgénicos mediante recombinación homóloga, donde el locus endógeno está alterado. Como alternativa, una construcción de ácido nucleico se integra aleatoriamente en el genoma. Los vectores para integración estable incluyen plásmidos, retrovirus y otros virus animales, YAC (*chromosomas artificiales de levadura*), y similares.

El ácido nucleico puede introducirse en la célula, directa o indirectamente mediante introducción en un precursor de la célula, mediante manipulación genética deliberada, tal como mediante microinyección o mediante infección con un virus recombinante o con un vector viral recombinante y similares. La expresión manipulación genética no incluye el cruzamiento convencional, o fertilización *in vitro*, sino que en su lugar se refiere a la introducción de una molécula de ácido nucleico recombinante. Esta molécula de ácido nucleico puede estar integrada dentro de un cromosoma, o puede ser ADN que se replica de forma extracromosómica.

Las construcciones de ADN para recombinación homóloga comprenderán al menos una parte de un ácido nucleico de la presente invención, donde el gen tiene la(s) modificación(es) genética(s) deseada(s), e incluye regiones de homología con el locus diana. No es necesario que las construcciones de ADN para integración aleatoria incluyan regiones de homología para mediar en la recombinación. Convenientemente, pueden incluirse marcadores para selección positiva y negativa. Los métodos para generar células que tienen modificaciones génicas fijadas como objetivo mediante recombinación homóloga se conocen en la técnica. Para diversas técnicas para transfectar células de mamífero, véase el documento Keown *et al.*, *Meth. Enzymol.* (1990) 185: 527-537.

Para células madre embrionarias (ES), puede emplearse una línea celular ES, o pueden obtenerse células embrionarias nuevas de un huésped, tal como un ratón, rata, cobaya, etc. Dichas células se cultivan en una capa alimentadora de fibroblastos apropiada o se cultivan en presencia del factor inhibidor de leucemia (LIF). Las ES o células embrionarias transformadas pueden usarse para producir animales transgénicos usando la técnica apropiada descrita en la técnica.

Los animales transgénicos pueden ser cualesquiera animales no humanos incluyendo un mamífero no humano (por ejemplo ratón, rata), un ave o un anfibio, etc., y usarse en estudios funcionales, cribado de fármacos y similares. Los ejemplos representativos del uso de animales transgénicos incluyen los que se describen más adelante.

También pueden producirse plantas transgénicas. Los métodos para preparar células y plantas transgénicas se describen en las Patentes de Estados Unidos N° 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.

Métodos para producir plantas transgénicas también se revisan en el documento *Plant Biochemistry and Molecular Biology* (eds. Lea and Leegood, John Wiley & Sons) (1993) págs. 275-295 y en el documento *Plant Biotechnology and Transgenic Plants* (eds. Oksman-Caldentey and Barz), (2002) 719 p.

Por ejemplo, pueden usarse explantes embriogénicos que comprenden células somáticas para la preparación del huésped transgénico. Después de la recogida de células o tejido, se introduce ADN exógeno de interés en las células vegetales, donde una serie de diferentes técnicas están disponibles para dicha introducción. Con protoplastos aislados, surge la oportunidad de introducción mediante protocolos de transferencia génica mediada por ADN, incluyendo 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 electroporación de los protoplastos en presencia de ADN desnudo que comprende la secuencia exógena de interés. Los protoplastos que han captado con éxito el ADN exógeno se seleccionan a continuación, se cultivan en un callo, y finalmente en una planta transgénica a través del contacto con las cantidades y proporciones apropiadas de factores estimuladores, tales como auxinas y citoquininas.

Pueden usarse otros métodos adecuados para producir plantas, tales como el enfoque con “pistola génica” o transformación mediada por *Agrobacterium* disponibles para los expertos en la materia.

Métodos de Uso

Las proteínas fluorescentes de la presente invención (así como otros componentes de la presente invención descritos anteriormente) son útiles en una serie de diferentes aplicaciones. Por ejemplo, pueden usarse en los métodos para marcar, analizar o detectar una molécula biológica, célula u orgánulo celular. Los usos representativos para cada uno de estos tipos de proteínas se describirán a continuación, donde los usos descritos en este documento son simplemente ejemplares y no pretenden limitar en modo alguno el uso de las proteínas de la presente invención a los descritos.

En una realización preferida que se refiere al método para marcar una molécula biológica, célula u orgánulo celular, las proteínas objeto de la invención son útiles como marcas *in vivo* (o moléculas informadoras) en ensayos en biología celular y molecular. Los ensayos de interés incluyen, aunque no se limitan a, ensayos para expresión génica, localización y co-localización de proteínas, interacciones proteína-proteína, interacciones proteína-ácido nucleico, interacciones ácido nucleico-ácido nucleico, localización e interacciones de células y orgánulos celulares, etc. Las proteínas fluorescentes de la presente invención son útiles como marcas de biomoléculas, o marcas de orgánulos celulares en células vivas y fijadas; como marcadores en la fusión de células u orgánulos celulares, como marcadores de la

integridad de una célula u orgánulo celular, como marcadores de transfección (por ejemplo, como marcas para la selección de células transfectadas que contienen un vector de expresión que codifica al menos una proteína fluorescente de la invención), como sondas en tiempo real que funcionan a concentraciones cercanas a la fisiológica, etc.

5 Además, las proteínas objeto de la invención pueden usarse en un método para analizar la expresión génica (por ejemplo, actividad promotora). En otras palabras, son útiles para identificar y/o medir la expresión 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 de acuerdo con la presente invención donde dicha molécula de ácido nucleico está unida de forma operativa a y bajo el control de una
10 secuencia de control de la expresión que modera la expresión de dicha proteína o polipéptido de interés; ii) expresión de dicho ácido nucleico en condiciones adecuadas; y iii) detección de la emisión de fluorescencia de la proteína fluorescente como medio para medir la expresión de la proteína de interés.

15 En particular, las proteínas objeto de la invención son útiles para identificar y/o medir la expresión y/o localización de la 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 de acuerdo con la presente invención donde dicha molécula de ácido nucleico está fusionada a una secuencia que codifica una proteína o polipéptido de interés y está unida de forma operativa a y bajo el control de una secuencia de control de la expresión que modera la expresión de dicha proteína o polipéptido de interés; ii) cultivar la célula
20 en condiciones adecuadas para la expresión de la proteína de interés; y iii) detectar la emisión de fluorescencia de la proteína fluorescente como un medio para medir la expresión/localización de la proteína interés.

Las aplicaciones de interés incluyen el uso de las proteínas objeto de la invención en métodos de transferencia de energía por resonancia de fluorescencia (FRET). En estos métodos, las proteínas objeto de la invención sirven como donadores y/o aceptores en combinación con una segunda proteína fluorescente o colorante con espectros de excitación/emisión apropiados; otros colorantes fluorescentes tales como coumarina y sus derivados, 7-amino-4-metilcoumarina y aminocoumarina; colorantes Bodipy; azul cascada; o fluoresceína y sus derivados, tales como isotiocianato de fluoresceína y verde Oregon; colorantes de rodamina tales como rojo Texas, tetrametilrodamina, eosinas y eritrosinas; colorantes de cianina tales como Cy3 y Cy5; quelatos macrocíclicos de iones de lantánidos, tales como colorante Quantum Dye; y colorantes quimioluminiscentes tales como luciferasas, incluyendo 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.

35 Los ejemplos específicos de cuando pueden usarse ensayos FRET que emplean las proteínas fluorescentes objeto de la invención incluyen, aunque sin limitación, la detección de interacciones proteína-proteína, tales como en un sistema di-híbrido de mamífero, dimerización del factor de transcripción, multimerización de las proteínas de membrana, formación del complejo multiproteico; como biosensor para una serie de diferentes sucesos, cuando un péptido o proteína se une covalentemente a una combinación fluorescente de FRET que incluye las proteínas fluorescentes objeto de la invención, y el péptido o proteína enlazador es, por ejemplo, un sustrato específico de proteasa para la escisión mediada por caspasa, un péptido que sufre un cambio conformacional después de recibir una señal que
40 aumenta o disminuye la FRET, tales como un dominio regulador de PKA (sensor de AMPc), un sitio de fosforilación (por ejemplo, donde hay un sitio de fosforilación en el péptido o el péptido tiene especificidad de unión por un dominio fosforilado/desfosforilado de otra proteína), o el péptido tiene un dominio de unión a Ca^{2+} . Además, las aplicaciones de transferencia de energía por resonancia de fluorescencia o FRET en las que las proteínas de la presente invención son útiles incluyen, aunque sin limitación, las descritas 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.

50 Las proteínas fluorescentes de la presente invención son útiles en un método para detectar los efectos de una sustancia de ensayo sobre la regulación de la expresión y/o translocación de una o más proteínas de interés en una célula. Como alternativa, son útiles en un método para detectar 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 ensayo. 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 ensayo. Dichos métodos pueden realizarse en presencia y en ausencia de una
55 sustancia de ensayo, cuyo efecto sobre el proceso se va a medir.

Las proteínas fluorescentes de la presente invención también son útiles en aplicaciones que implican el cribado automatizado de series de células que expresan grupos informadores fluorescentes mediante el uso de análisis de formación de imágenes al microscopio y electrónicos. Puede usarse el cribado para el descubrimiento de fármacos y en el campo de la genómica funcional donde las proteínas objeto de la invención se usan como marcadores de células completas para detectar cambios en la reorganización y migración multicelular, por ejemplo en la formación de túbulos multicelulares (formación de vasos sanguíneos) mediante células endoteliales, migración de células mediante el sistema de inserción Fluoroblok (Becton Dickinson Co.), cicatrización de heridas, o crecimiento de neuritas. También puede emplearse cribado donde las proteínas de la presente invención se usan como marcadores fusionados a péptidos (tales como secuencias fijadoras de objetivos) o proteínas que detectan cambios en la ubicación intracelular como
65 indicador para la actividad celular, por ejemplo en la transducción de señales, tales como translocación de quinasa y factor de transcripción después de estímulos. Los ejemplos incluyen proteína quinasa C, proteína quinasa A, factor de transcripción NFκB, y NFAT; proteínas del ciclo celular, tales como ciclina A, ciclina B1 y ciclina E; escisión de pro-

teasa con el posterior movimiento de sustrato escindido; fosfolípidos con marcadores para estructuras intracelulares tales como el retículo endoplasmático, el aparato de Golgi, mitocondrias, peroxisomas, núcleo, nucleolos, membrana plasmática, histonas, endosomas, lisosomas o microtúbulos.

5 Las proteínas de la presente invención también pueden usarse en cribado de alto contenido para detectar la co-localización de otras proteínas de fusión fluorescentes con marcadores de localización como indicadores de movimientos de proteínas/péptidos fluorescentes intracelulares o como marcadores en solitario. Los ejemplos de aplicaciones que implican el cribado automatizado de series de células en las que las proteínas fluorescentes objeto de la invención son útiles incluyen la Patente de Estados Unidos N° 5.989.835; así como los documentos WO 0017624; WO 00/26408; 10 WO 00/17643; y WO 00/03246.

Las proteínas fluorescentes de la presente invención también son útiles en ensayos de cribado de alto rendimiento. Las proteínas fluorescentes objeto de la invención son proteínas estables con semividas de más de 24 horas. También se proporcionan versiones desestabilizadas de las proteínas fluorescentes objeto de la invención con semividas reducidas 15 que pueden usarse como informadores de transcripción para el descubrimiento de fármacos. Por ejemplo, una proteína de acuerdo con la presente invención puede fusionarse con una supuesta secuencia señal proteolítica obtenida de una proteína con una semivida más corta, tal como una secuencia PEST del gen de ornitina descarboxilasa de ratón, una secuencia de destrucción de ciclina B1 de ratón o ubiquitina, etc. Para una descripción de proteínas desestabilizadas y vectores que pueden emplearse para producirlas, véase por ejemplo, la Patente de Estados Unidos N° 6.130.313. Los 20 promotores en rutas de transducción de señales pueden detectarse usando versiones desestabilizadas de las proteínas fluorescentes objeto de la invención para el cribado de fármacos tales como, por ejemplo, API, NFAT, NFkB, Smad, STAT, p53, E2F, Rb, myc, CRE, ER, GR y TRE, y similares.

Las proteínas objeto de la invención pueden usarse como segundos detectores de mensajeros fusionando las proteínas objeto de la invención a dominios específicos tales como el dominio de unión a PKCgamma Ca, el dominio de 25 unión a PKCgamma DAG, el dominio SH2 o el dominio SH3, etc.

Las formas secretadas de las proteínas objeto de la invención, que a su vez pueden usarse en una serie de diversas aplicaciones pueden prepararse fusionando secuencia líder secretadas a las proteínas objeto de la invención. 30

Las proteínas objeto de la invención también son útiles en aplicaciones de clasificación de células activadas por fluorescencia (FACS). En dichas aplicaciones, la proteína fluorescente objeto de la invención se usa como marca para marcar una población de células y la población de células marcada resultante se clasifica a continuación con un dispositivo de clasificación de células activadas por fluorescencia, como se conoce en la técnica. Los métodos de FACS 35 se describen en las Patentes de Estados Unidos N° 5.968.738 y 5.804.387.

Las proteínas objeto de la invención también son útiles como marcas *in vivo* en animales transgénicos. Por ejemplo, la expresión de la proteína objeto de la invención puede estar impulsada por promotores específicos del tejido, donde dichos métodos son útiles en investigación para terapia génica, tal como ensayar la eficacia de la expresión transgénica, 40 entre otras aplicaciones. Una aplicación representativa de proteínas fluorescentes en animales transgénicos que ilustra dichas aplicaciones se encuentra en el documento WO 00/02997.

Aplicaciones adicionales de las proteínas de la presente invención incluyen el uso como marcadores después de la inyección en células o animales y en el calibrado para mediciones cuantitativas; como marcadores o informadores en 45 dispositivos biosensores de oxígeno para monitorizar la viabilidad celular; como marcadores o marcas para animales, mascotas, juguetes, alimentos, y similares.

Las proteínas fluorescentes objeto de la invención también son útiles en ensayos de escisión con proteasa. Por ejemplo, pueden desarrollarse ensayos de fluorescencia activada por escisión usando las proteínas objeto de la invención, manipuladas para incluir una secuencia de escisión específica de proteasa. Después de la escisión de la proteína fluorescente mediante una proteasa activada, la fluorescencia cambiaría debido a cambios conformacionales de la proteína objeto de la invención. Las aplicaciones anteriores podrían desarrollarse en ensayos para una serie de diferentes 50 tipos de proteasas, tales como caspasas y otras.

Las proteínas objeto de la invención también pueden usarse en ensayos para determinar la composición de fosfolípidos en membranas biológicas. Por ejemplo, proteínas de fusión de las proteínas objeto de la invención (o cualquier otro tipo de modificación covalente o no covalente de las proteínas objeto de la invención) que permita la unión a fosfolípidos específicos para localizar/visualizar patrones de distribución de fosfolípidos en membranas biológicas, mientras se permite la co-localización de proteínas de membrana en "microdominios" (*rafts*) de fosfolípidos específi- 60 cos, puede realizarse con las proteínas objeto de la invención.

Las proteínas fluorescentes objeto de la invención también son útiles como biosensores, fuentes de proteínas fluorescentes permutadas de forma circular y biosensores de las mismas. Los métodos de preparación y uso de proteínas fluorescentes permutadas de forma circular se describen en los documentos Nagai *et al.*, Proc Natl Acad Sci U S A, 2001, V. 98(6), págs. 3197-3202, Nagai *et al.*, Proc Natl Acad Sci USA, 2004, V. 101(29), págs. 10554-10559, Filippin *et al.*, J Biol Chem., 2003, V. 278(40), págs. 39224-34, y en las Patentes de Estados Unidos N° 6.469.154 y 6.699.687. Los biosensores pueden usarse en células procariotas y eucariotas, tales como indicadores de iones Ca²⁺, un indicador de pH, un indicador de fosforilación, otros indicadores de actividad enzimática, o como un indicador de iones, tales 65

ES 2 353 081 T3

como magnesio, sodio, potasio, cloro, haluros, etc. Los métodos de uso de proteínas fluorescentes como biosensores también incluyen los descritos en las Patentes de Estados Unidos N° 5.972.638, 5.824.485 y 5.650.135 (así como las referencias mencionadas en ellas).

- 5 Los anticuerpos de la presente invención, descritos anteriormente, también son útiles en una serie de aplicaciones, incluyendo la diferenciación de las proteínas objeto de la invención de otras proteínas fluorescentes.

Kits

10

La presente invención también proporciona kits para su uso en la puesta en práctica de una o más de las aplicaciones descritas anteriormente. En realizaciones preferidas pueden usarse kits para marcar una molécula biológica. Los kits incluyen típicamente la proteína de la invención como tal, o un ácido nucleico que la codifica preferentemente con los elementos para expresar las proteínas objeto de la invención, por ejemplo, una construcción tal como un vector que comprende un ácido nucleico que codifica la proteína objeto de la invención.

Los siguientes ejemplos se ofrecen a modo de ilustración y no a modo de limitación.

20 Ejemplos

Ejemplo 1

Clonación de ADNc que codifica una proteína roja fluorescente de Entacmaea quadricolor

25

Para buscar proteínas fluorescentes de *Entacmaea quadricolor*, se usó una estrategia basada en cribar una biblioteca de ADNc de expresión en *E. coli*. Un pequeño fragmento (aproximadamente 1 mm de longitud) de tentáculo de *Entacmaea quadricolor* (*Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria; Nynantheae; Actiniidae; Entacmaea*) que poseía fluorescencia roja brillante se usó para la preparación de ARN total. El ARN total se aisló mediante un kit *NucleoSpin RNA II kit* (Clontech). La muestra de ADNc amplificado se preparó usando un kit de amplificación de ADNc SMART (Clontech) y se clonó en el vector PCR-Script (Stratagene). Aproximadamente 5×10^4 clones recombinantes se cribaron visualmente usando un estereomicroscopio fluorescente. Como resultado, se identificó una nueva proteína roja fluorescente llamada EqFP578 (o eqFP578, SEC ID N° 1 y 2). La proteína fluorescente comparte aproximadamente el 76% de identidad en la secuencia de aminoácidos con otra proteína fluorescente de *Entacmaea quadricolor*, eqFP611 (GenBank Id AAN05449) y el 64% de identidad en la secuencia de aminoácidos con la proteína roja no fluorescente asCP562 de *Anemonia sulcata* (GenBank Id AAG41206).

Ejemplo 2

40

Caracterización de EqFP578 (SEC ID N° 1, 2)

Las secuencias codificantes de ácido nucleico de EqFP578 se obtuvieron como se ha descrito anteriormente en el Ejemplo 1 y se clonaron en un vector de expresión pQE30 (Qiagen), de modo que la proteína recombinante contenía una marca de seis histidinas en su extremo N. Después de la expresión en *E. coli*, la proteína se purificó mediante una resina de afinidad por metal TALON (Clontech) y se caracterizó.

La proteína tenía picos espectrales de excitación-emisión a 552 y 578 nm, respectivamente (figura 2). La proteína purificada posee un coeficiente de extinción molar de $102.000 \text{ M}^{-1} \text{ cm}^{-1}$ y un rendimiento cuántico de fluorescencia de 0,54. Para la determinación del coeficiente de extinción molar, se estimó la concentración de cromóforo maduro. La proteína se sometió a desnaturalización alcalina con un volumen igual de NaOH 2 M. En estas condiciones, los cromóforos similares a DsRed (incluyendo el cromóforo de EqFP578) se convierten en el cromóforo similar a GFP, que absorbe a 446 nm con un coeficiente de extinción molar de $44.000 \text{ M}^{-1} \text{ cm}^{-1}$ (Ward, W. W., *Bioluminescence and Chemiluminescence* (1981), Academic Press, 235-242). Se midieron los espectros de absorción para EqFP578 nativa y sometida a desnaturalización alcalina. El coeficiente de extinción molar para la proteína en estado nativo se estimó en base a la absorción de la proteína desnaturalizada. Para la determinación del rendimiento cuántico, la fluorescencia de EqFP578 se comparó con DsRed2 de igual absorción con un rendimiento cuántico de 0,55.

Los resultados de un ensayo de filtración en gel indicaban que EqFP578 es una proteína dimerica. Se cargaron muestras de proteína purificadas (~1 mg/ml) en una columna Sephadex-100 (0,7 x 60 cm) y se eluyeron con una solución de tampón fosfato 50 mM (pH 7,0) y NaCl 100 mM. Se usaron EGFP, HcRed1 y DsRed2 (Clontech) como patrones de monómero, dímero y tetramero, respectivamente. Además, se ha indicado una tendencia a formar tetrameros cuando se usó EqFP578 purificada recombinante para la filtración en gel a una concentración de aproximadamente 10 mg/ml.

65

Ejemplo 3

Preparación de mutantes de EqFP578

5 En cada caso, el kit *Diversity PCR Random Mutagenesis kit* (CLONTECH) se usó para una mutagénesis aleatoria, en condiciones óptimas para 5-6 mutaciones por cada 1000 pb.

10 La secuencia codificante de ácido nucleico de la EqFP578 de tipo silvestre se obtuvo como se ha descrito anteriormente en el Ejemplo 1. Para potenciar la expresión en células de mamífero, realizamos la siguiente estrategia: 1. se usó mutagénesis aleatoria para encontrar una proteína con una mayor velocidad de maduración y plegamiento; 2. la secuencia de la proteína se “humanizó”; 3. se realizó una mutagénesis aleatoria adicional para aumentar adicionalmente la velocidad de maduración y el brillo de la proteína en *E. coli* a 37°C. Como resultado, se obtuvo el mutante de EqFP578 llamado EqFP578m1 (SEC ID N° 3, 4) con el uso de codón optimizado en mamífero y que comprendía seis sustituciones de aminoácidos en comparación con la EqFP578 de tipo silvestre: R32G, T68A, L79F, L110F, S131P y L138R.

15 La EqFP578m1 recombinante se preparó, purificó y caracterizó como se ha descrito en el Ejemplo 2 para la EqFP578 de tipo silvestre. Las propiedades de fluorescencia de este mutante eran similares, pero no idénticas a la EqFP578 de tipo silvestre. Tiene picos espectrales de excitación-emisión en 553 y 574 nm, respectivamente (figura 3). La proteína purificada posee un coeficiente de extinción molar de $104.000 \text{ M}^{-1}\text{cm}^{-1}$ y un rendimiento cuántico de fluorescencia de 0,65. Al igual que la EqFP578 de tipo silvestre, EqFP578m1 es un dímero.

20 Aunque EqFP578m1 no demuestra ninguna agregación cuando se expresa *in vivo*, demostró formar agregados *in vitro*, de acuerdo con los datos de electroforesis en gel. Esta propiedad puede potencialmente, restringir las aplicaciones de proteínas para la generación de líneas celulares estables y animales transgénicos. Para reducir la capacidad de agregación de la proteína, el ácido nucleico que las codifica se sometió a modificaciones adicionales que aumentan la carga negativa local (es decir, concentración de restos de aminoácidos cargados negativamente) en los extremos C y N de la proteína. En particular, los aminoácidos en los extremos N y C se sustituyeron para reducir la carga positiva local, por ejemplo, se introdujeron sustituciones K6T y R231S mediante mutagénesis dirigida. Además, una secuencia de aminoácidos adicional (MGEY) se añadió al extremo N de la proteína. La mutagénesis aleatoria del ácido nucleico obtenido dio como resultado una mutación de plegamiento K188R adicional que proporciona una generación más rápida de una señal fluorescente *in vivo*, como se muestra tras la expresión de la proteína en la cepa XL-1Blue de *E. coli*. La secuencia de nucleótidos y aminoácidos del mutante no agregante llamado M1-NA se muestran en las SEC ID N° 5, 6, respectivamente.

35 Se preparó, purificó y caracterizó M1-NA recombinante como se ha descrito en el Ejemplo 2 para la EqFP578 de tipo silvestre. De acuerdo con los datos de filtración en gel, M1-NA es una proteína dimerica a una concentración de 1 mg/ml. Las propiedades fluorescentes de M1-NA son las mismas que las de EqFP578m1. M1-NA es una proteína fluorescente roja brillante, caracterizada con espectros de fluorescencia de excitación/emisión con máximos a 553/574 nm. El rendimiento cuántico de fluorescencia de M1-A = 0,67, el coeficiente de extinción molar (a 553 nm) = $92.000 \text{ M}^{-1}\text{cm}^{-1}$. M1-NA no forma agregados *in vitro*, de acuerdo con los datos de electroforesis en gel. Tampoco forma agregados detectables algunos después de la expresión en células eucariotas, como se mostró para las líneas celulares HeLa y 293T.

45 Una variante de EqFP578m1 (M1-602) con espectros de excitación/emisión desplazados a rojo también se produjo mediante mutagénesis aleatoria. En comparación con EqFP578m1, comprende una sustitución N143S, que proporciona un desplazamiento a rojo de los espectros de excitación y emisión de fluorescencia. La mutagénesis aleatoria produjo además las sustituciones F110L, I115L, R138L, G152S y F174L que potencian el plegamiento de la proteína y la maduración del cromóforo, dando como resultado la generación más rápida de una señal fluorescente *in vivo* como se muestra después de la expresión en la cepa XL-1Blue de *E. coli*. Así como M1-NA, los extremos N y C de M1-602 se modificaron con las sustituciones K6T y R231S y la adición de aminoácidos N-terminales (MGED). Las secuencias de nucleótidos y de aminoácidos de la proteína M1-602 se muestran en las SEC ID N° 7, 8, respectivamente.

50 Se preparó, purificó y caracterizó M1-602 recombinante como se ha descrito en el Ejemplo 2 para la EqFP578 de tipo silvestre. Presentaba máximos de los espectros de excitación/emisión a 574/602 nm (véase la figura 4 para los espectros). Rendimiento cuántico de fluorescencia de M1-602 = 0,35, coeficiente de extinción molar (a 574 nm) = $74.400 \text{ M}^{-1}\text{cm}^{-1}$. De acuerdo con los datos de filtración en gel, M1-602 es una proteína dimerica a una concentración de 1 mg/ml.

60 Un mutante con un desplazamiento a rojo adicional en los espectros de excitación/emisión llamado M1-637 se preparó en base al ácido nucleico que codifica la proteína M1-602 mediante mutagénesis dirigida de los restos H197 y G158. En comparación con M1-602, M1-637 comprende las sustituciones de aminoácidos H197R y G158S, que proporcionan un desplazamiento al rojo lejano de los espectros de fluorescencia. Las secuencias de nucleótidos y de aminoácidos de la proteína M1-602 se muestran en las SEC ID N° 9, 10, respectivamente.

65 Se preparó, purificó y caracterizó M1-637 recombinante como se describe en el Ejemplo 2 para la EqFP578 de tipo silvestre. M1-637 es una proteína fluorescente rojo lejano, caracterizada por espectros de excitación/emisión con máximos a 591/637 nm (véase la figura 5 para los espectros). De acuerdo con los datos de filtración en gel, M1-637 es una proteína dimerica a una concentración de 1 mg/ml.

ES 2 353 081 T3

Debido a sus características superiores, EqFP578 y sus mutantes descritos anteriormente representan una base atractiva para la generación de una proteína fluorescente rojo brillante monomérica de propiedades ventajosas. La disponibilidad de la estructura cristalina para eqFP611 (Petersen *et al.*, 2003, J Biol Chem v. 278: págs. 44626-44631), el homólogo cercano de eqFP578, nos permitió determinar restos de aminoácidos, responsables de la dimerización y la tetramerización débil de eqFP578 y mutantes de la misma. La mutagénesis dirigida simultánea de varios restos de aminoácidos clave, responsables de la dimerización a través de la llamada interfaz “segunda” o “hidrófila”, es decir, R155E, Q159D, S173N, F192V y F194Y, se ha empleado en el ácido nucleico que codifica la proteína EqFP578m1. Simultáneamente, para adaptar el plegamiento y la maduración de la proteína monomérica, se realizó mutagénesis aleatoria para la secuencia completa de la proteína, mientras que las posiciones clave de la interfaz se fijaron con cebadores. Para inhibir la dimerización secundaria a través de la “primera” interfaz (“hidrófoba”), también se introdujo la sustitución N122R, que se había descrito que altera eficazmente la tetramerización de eqFP611 (Wiedenmann *et al.*, 2005, J Biomed Opt v. 10: p. 14003).

Las variantes fluorescentes rojas con mala maduración obtenidas después de la primera ronda de mutagénesis se comprobaron mediante filtración en gel y demostraron ser monoméricas, de acuerdo con los datos de filtración en gel. Esto indicaba que las mutaciones introducidas fueron suficientes para alterar la dimerización. A continuación, se realizaron siete rondas adicionales de mutagénesis aleatoria para optimizar el plegamiento y la maduración de proteínas a 37°C. Después de cada ronda, las bibliotecas generadas se cribaron en un sistema de expresión en *E. coli*, usando un estereomicroscopio fluorescente Olympus SZX-12, juego de filtros TRITC. De 10 a 20 clones más brillantes se seleccionaron y se comprobaron mediante secuenciación. Solamente se usaron para el trabajo adicional aquellas variantes de proteínas, que contenían todas las sustituciones clave en la “segunda” interfaz. De 5 a 10 de las variantes seleccionadas se compararon con respecto a rendimiento cuántico de fluorescencia, coeficiente de extinción molar y fotoestabilidad. La completitud de la maduración del cromóforo rojo se comprobó mediante mediciones del espectro de absorbancia. La variante final, M1-mono1, era una proteína fluorescente roja monomérica con máximos de excitación/emisión a 555/584 nm (Figura 6). Las secuencias de nucleótidos y de aminoácidos de la proteína M1-mono1 se muestran en las SEC ID N° 11, 12, respectivamente.

Además, se produjeron dos proteínas cian sobre la base de M1-mono1. Un ácido nucleico que codifica M1-mono1 se sometió a mutagénesis dirigida y una posterior aleatoria. Los plásmidos resultantes que codifican proteínas fluorescentes mutantes se transfectaron en *E. coli* y las variantes cian más brillantes se seleccionaron. La primera, nrM181-5, contiene las sustituciones N143H y R67K, que dan como resultado fluorescencia cian del cromóforo de la proteína y la capacidad de fotoconversión de la proteína a la forma fluorescente roja. En comparación con M1-mono1, nrM181-también contiene las sustituciones R42K, C172A, V216M y R220K que potencian el plegamiento de la proteína y la maduración del cromóforo. Las secuencias de nucleótidos y de aminoácidos de la proteína nrM181-5 se muestran en las SEC ID N° 13, 14, respectivamente.

Se preparó, purificó y caracterizó nrM181-5 recombinante como se ha descrito en el Ejemplo 2 para la EqFP578 de tipo silvestre. nrM181-5 es una proteína fluorescente cian, caracterizada por espectros de excitación/emisión con máximos a 400/470 nm (véase la figura 7 para los espectros). De acuerdo con los datos de filtración en gel, nrM181-5 es una proteína monomérica a una concentración de 1 mg/ml. La proteína es capaz de fotoconversión a la forma fluorescente roja en respuesta a irradiación con luz violeta.

La segunda proteína Cyan-2-1 comprende las sustituciones N143F y H197Y, que dan como resultado fluorescencia cian del cromóforo de la proteína. Cyan-2-1 también comprende las sustituciones E36G y F53V que potencian el plegamiento de la proteína y la maduración del cromóforo. Las secuencias de nucleótidos y de aminoácidos de la proteína Cyan-2-1 se muestran en las SEC ID N° 15, 16, respectivamente.

Se preparó, purificó y caracterizó Cyan-2-1 recombinante como se ha descrito en el Ejemplo 2 para la EqFP578 de tipo silvestre. Cyan-2-1 es una proteína fluorescente cian, caracterizada por espectros de excitación/emisión con máximos a 400/470 nm. De acuerdo con los datos de filtración en gel, Cyan-2-1 es una proteína monomérica a una concentración de 1 mg/ml. La proteína no es capaz de fotoconversión a la forma fluorescente roja en respuesta a la irradiación con luz violeta.

55 Ejemplo 4

Preparación de anticuerpo policlonal

Las regiones codificantes de ácidos nucleicos de EqFP578m1 preparadas como se ha descrito anteriormente en el Ejemplo 3 se clonaron en el vector de expresión pQE30 (Qiagen), de modo que la proteína recombinante contenía una marca de seis histidinas como su extremo N. Después de la expresión en *E. coli*, la proteína se purificó mediante resina de afinidad por metal TALON (Clontech) en condiciones de desnaturalización. Se inmunizaron y se reforzaron conejos cuatro veces a intervalos mensuales con polipéptidos recombinantes emulsionados en adyuvante completo de Freund. Diez u 11 días después de cada refuerzo se extrajo sangre de los animales. Se ensayó anticuerpo antisuero policlonal en proteína recombinante mediante ELISA y mediante inmunotransferencia de Western.

ES 2 353 081 T3

Ejemplo 5

Marcado de células y orgánulos de mamífero usando mutantes de EqFP578

5 Para el marcado fluorescente de células eucariotas, las secuencias codificantes de EqFP578m1, M1-NA, M1-602, M1-637, M1-mono1, nrM181-5 y Cyan-2-1 preparadas como se ha descrito anteriormente en el Ejemplo 3 se clonaron en el vector pEGFP-N1 (CLONTECH) entre los sitios de restricción AgeI y BglIII (en lugar de la región codificante EGFP). Las líneas celulares HeLa, 293T y Phoenix se transfectaron transitoriamente con estos vectores usando el reactivo LipofectAMINE (Invitrogen) y se ensayaron 20 h después de la transfección. Un microscopio de fluorescencia Olympus CK40 equipado con una cámara CCD (DP-50, Olympus) se usó para generar imágenes de las células. La expresión de cada proteína en todas las líneas celulares ensayadas dio como resultado señales fluorescentes brillantes sin agregación visible. La fluorescencia era claramente detectable dentro de las 20 horas después de la transfección.

15 Para ensayar las proteínas en fusiones con señales de localización subcelular, la secuencia que reconoce a la mitocondria (MTS) de la subunidad VIII de la citocromo c oxidasa humana se clonó en los vectores N1 obtenidos como se ha descrito anteriormente. La transfección de células HeLa dio como resultado la translocación eficaz de las proteínas a la mitocondria de células huésped. En cada caso, la fluorescencia brillante era claramente detectable dentro de 24 horas después de la transfección. No se observó agregación de proteínas visible.

Ejemplo 7

Marcado de proteínas usando mutantes de EqFP578

25 Las secuencias codificantes de EqFP578m1, M1-NA, M1-602, M1-637, M1-mono1, nrM181-5 y Cyan-2-1 preparadas como se ha descrito anteriormente en el Ejemplo 3 se clonaron en el vector pEGFP-C1 (CLONTECH) en lugar de la región codificante de EGFP. Las secuencias codificantes de socios de fusión (beta-actina citoplásmica humana, fibrilarina y proteína BID) se unieron de forma operativa con las secuencias de mutantes de EqFP578 indicadas anteriormente mediante la clonación las secuencias del socio de fusión en múltiples sitios de clonación de los vectores C correspondientes en marco con las secuencias codificantes de proteínas fluorescentes. Los vectores se transfectaron a células HeLa. En cada caso, la distribución de la fluorescencia reflejaba la distribución esperada del socio de fusión en una célula. La fluorescencia brillante era claramente detectable dentro de 24 horas después de la transfección.

35 Sin embargo, en el caso de EqFP578m1, M1-NA, M1-602, M1-637, se observó la distribución anormal de las señales fluorescentes en células huésped cuando se usaba alfa tubulina como socio de fusión. Por otro lado, las fusiones de M1-mono1, nrM181-5 y Cyan-2-1 con alfa tubulina mostraban una distribución esperada de señales fluorescentes que sugería la naturaleza monomérica de estas proteínas.

40 Una realización de la presente invención se refiere a un ácido nucleico aislado, que comprende una secuencia de ácido nucleico que codifica una proteína fluorescente funcional relacionada con EqFP578, donde dicha proteína es sustancialmente igual que o idéntica a la proteína fluorescente EqFP578 de tipo silvestre de la SEC ID N° 2.

45 En una realización preferida del ácido nucleico aislado, la proteína fluorescente funcional relacionada con EqFP578 es una proteína fluorescente funcional manipulada genéticamente cuya secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por R32G y S131P, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una mayor velocidad de maduración a una temperatura de 37°C en comparación con EqFP578.

50 En una realización preferida adicional del ácido nucleico aislado, la secuencia de aminoácidos comprende al menos una sustitución de aminoácidos adicional seleccionada entre el grupo constituido por E36G, K42R, F53V, K67R, T6BA, L79F, I93V, F110L, N112D, I115L, R138L, G152S, H157R, Y169H, H1711, C172A, F174L, K188R, H193Y, M216V, K220R, y R231K, donde dicha proteína fluorescente funcional manipulada genéticamente tiene un plegamiento potenciado a una temperatura de 37°C en comparación con EqFP578.

55 En una realización preferida adicional más del ácido nucleico aislado, la proteína fluorescente funcional relacionada con EqFP578 es una proteína fluorescente funcional manipulada genéticamente cuya secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por K6T y R231S, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de agregación reducida en comparación con EqFP578.

60 En otra realización preferida más del ácido nucleico aislado, la secuencia de aminoácidos comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY y MGED, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de agregación reducida en comparación con EqFP578.

65 En una realización preferida adicional del ácido nucleico aislado, la proteína fluorescente funcional relacionada con EqFP578 es una proteína fluorescente funcional manipulada genéticamente cuya secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos

ES 2 353 081 T3

seleccionada entre el grupo constituido por R155E, Q159D, F192V, S173N y F194Y, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de oligomerización reducida en comparación con EqFP578.

5 En una realización preferida adicional del ácido nucleico aislado, la secuencia de aminoácidos comprende una sustitución N122R, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de oligomerización reducida en comparación con EqFP57B.

10 En una realización preferida adicional del ácido nucleico aislado, la secuencia de aminoácidos comprende además al menos una sustitución de plegamiento seleccionada entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, 193V, F110L, N112D, 1115L, R138L, G152S, H157R, Y169H, H1711, C172A, F174L, K188R, H193Y, M216V, K220R y R231K.

15 En otra realización preferida más del ácido nucleico aislado, la proteína fluorescente funcional relacionada con EqFP578 es una proteína fluorescente funcional manipulada genéticamente cuya secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por H197R, S158G, N1435, N143H, N143F y N143Y, donde dicha proteína fluorescente funcional manipulada genéticamente tiene un espectro de excitación y uno de emisión diferentes de los espectros de la EqFP578.

20 En una realización preferida adicional del ácido nucleico aislado, la proteína fluorescente funcional relacionada con EqFP578 comprende una secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.

25 Otro aspecto de la presente invención se refiere a un vector que comprende el ácido nucleico.

Otro aspecto más de la presente invención se refiere a una casete de expresión que comprende (a) una región de inicio de la transcripción funcional en un huésped de expresión; (b) el ácido nucleico; y (c) una región de terminación de la transcripción funcional en dicho huésped de expresión.

30 Otro aspecto más de la presente invención se refiere a una célula huésped o proteína de la misma, que comprende la casete de expresión como parte de un elemento extracromosómico o integrada en el genoma de una célula huésped como resultado de la introducción de dicha casete de expresión en dicha célula huésped.

35 Un aspecto adicional de la presente invención se refiere a una célula transgénica, o progenie de la misma, que comprende el ácido nucleico.

40 Otro aspecto más de la presente invención se refiere a una proteína fluorescente funcional relacionada con EqFP578 aislada que es codificada por el ácido nucleico, donde dicha proteína es sustancialmente igual que o idéntica a la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2.

45 En una realización de la proteína fluorescente funcional relacionada con EqFP578 aislada, la proteína está manipulada genéticamente y su secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por R32G y S131P, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una mayor velocidad de maduración a una temperatura de 37°C en comparación con Eq FP578.

50 En otra realización de la proteína fluorescente funcional relacionada con EqFP578 manipulada genéticamente, su secuencia de aminoácidos comprende al menos una sustitución de aminoácidos adicional seleccionada entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, 93V, F110L, N112D, 1115L, R138L, G152S, H157R, Y169H, H171, C172A, F174L, K188R, H193Y, M216V, K220R y R231K, donde dicha proteína fluorescente funcional manipulada genéticamente tiene un plegamiento potenciado a una temperatura de 37°C en comparación con EqFP578.

55 En otra realización más de la proteína fluorescente funcional relacionada con EqFP578 aislada, la proteína está manipulada genéticamente y su secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por K6T y R231S, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de agregación reducida en comparación con EqFP578.

60 En una realización adicional de la proteína fluorescente funcional relacionada con EqFP578 manipulada genéticamente, su secuencia de aminoácidos comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY y MGED, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de agregación reducida en comparación con EqFP578.

65 En otra realización más de la proteína fluorescente funcional relacionada con EqFP578 aislada, la proteína está manipulada genéticamente y su secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por R155E, Q159D, F192V, S173N, y F194Y, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de oligomerización reducida en comparación con EqFP578.

ES 2 353 081 T3

En otra realización más de la proteína fluorescente funcional relacionada con EqFP578 manipulada genéticamente, su secuencia de aminoácidos comprende una sustitución N122R, donde dicha proteína fluorescente funcional manipulada genéticamente tiene una capacidad de oligomerización reducida en comparación con EqFP578.

5 En una realización adicional de la proteína fluorescente funcional relacionada con EqFP578 manipulada genéticamente, su secuencia de aminoácidos comprende además al menos una sustitución de plegamiento seleccionada entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115L, R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K.

10 En una realización adicional más de la proteína fluorescente funcional relacionada con EqFP578 aislada, la proteína está manipulada genéticamente y su secuencia de aminoácidos difiere de la proteína fluorescente de tipo silvestre EqFP578 de la SEC ID N° 2 en al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por H197R, S158G, N143S, N143H, N143F o N143Y, donde dicha proteína fluorescente funcional manipulada genéticamente tiene un espectro de excitación y uno de emisión diferentes de los espectros de la EqFP578.

15 En otra realización de la proteína fluorescente funcional relacionada con EqFP578, ésta comprende una secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

- 5 1. Un ácido nucleico aislado que comprende una secuencia de aminoácidos que codifica una proteína fluorescente, donde dicha proteína tiene al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16 en una ventana de comparación de al menos 200 aminoácidos.
- 10 2. El ácido nucleico de acuerdo con la reivindicación 1, en el que dicha proteína tiene al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos de la SEC ID N° 2 en una ventana de comparación de al menos 200 aminoácidos y en la que la proteína fluorescente comprende al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por R32G y S131P, y en la que la proteína fluorescente opcionalmente comprende además una sustitución seleccionada entre el grupo constituido por E36G, K42R, F53V, K67R, T68A, L79F, I93V, F110L, N112D, I115L, R138L, G152S, H157R, Y169H, H171I, C172A, F174L, K188R, H193Y, M216V, K220R y R231K.
- 15 3. El ácido nucleico de acuerdo con la reivindicación 1, en el que dicha proteína tiene al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos de la SEC ID N° 2 en una ventana de comparación de al menos 200 aminoácidos y en la que la proteína fluorescente comprende una secuencia de aminoácidos N-terminal adicional seleccionada entre el grupo constituido por MGEY y MGED.
- 20 4. El ácido nucleico de acuerdo con la reivindicación 1, en el que dicha proteína tiene al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos de la SEC ID N° 2 en una ventana de comparación de al menos 200 aminoácidos y en la que la proteína fluorescente comprende al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por R155E, Q159D, F192V, S173N, F194Y y N122R y tiene una capacidad de oligomerización reducida en comparación con EqFP578.
- 25 5. El ácido nucleico de acuerdo con la reivindicación 1, en el que dicha proteína tiene al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos de la SEC ID N° 2 en una ventana de comparación de al menos 200 aminoácidos y en la que la proteína fluorescente comprende al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por H197R, S158G, N143S, N143H, N143F y N143Y, en la que dicha proteína fluorescente tiene un espectro de excitación y uno de emisión diferentes de los espectros de EqFP578.
- 30 6. Un ácido nucleico de acuerdo con la reivindicación 1, en el que la proteína fluorescente comprende una secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.
- 35 7. Un vector que comprende el ácido nucleico de acuerdo con la reivindicación 1.
- 40 8. Una casete de expresión que comprende (a) una región de inicio de la transcripción funcional en un huésped de expresión; (b) el ácido nucleico de acuerdo con la reivindicación 1; y (c) una región de terminación de la transcripción funcional en dicho huésped de expresión.
- 45 9. Un huésped o progenie del mismo, que comprende la casete de expresión de acuerdo con la reivindicación 8 como parte de un elemento extracromosómico, o integrada en el genoma de una célula huésped como resultado de la introducción de dicha casete de expresión en dicha célula huésped.
- 50 10. Un proceso para la fabricación de una célula transgénica que comprende introducir el ácido nucleico de acuerdo con la reivindicación 1 en una célula.
- 55 11. Una proteína fluorescente aislada que es codificada por el ácido nucleico de acuerdo con las reivindicaciones 1-6.
- 60 12. Una proteína fluorescente aislada que comprende una secuencia de aminoácidos con al menos el 90% de identidad en la secuencia con la secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.
- 65 13. Una proteína fluorescente de acuerdo con la reivindicación 12, que comprende una secuencia de aminoácidos seleccionada entre el grupo constituido por las SEC ID N° 2, 4, 6, 8, 10, 12, 14 y 16.
14. El ácido nucleico de acuerdo con la reivindicación 1, en el que la proteína fluorescente comprende al menos una sustitución de aminoácidos seleccionada entre el grupo constituido por K6T y R231S.

```

          *           20           *           40
SEC ID N° 2 (EqFP578) : ...MSELIKENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:40
SEC ID N° 4 (EqFP578m1) : ...MSELIKENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:40
SEC ID N° 6 (M1-NA) : MGEYSELITENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:43
SEC ID N° 8 (M1-602) : MGEDSELITENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:43
SEC ID N° 10 (M1-637) : MGEDSELITENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:43
SEC ID N° 12 (M1-mono1) : ...MSELIKENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:40
SEC ID N° 14 (nrM181-5) : ...MSELIKENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:40
SEC ID N° 16 (Cyan2-1) : ...MSELIKENMHHMKLYMEGTVNNHFFKCTSEGERKPYEGTQT:40

          *           60           *           80
SEC ID N° 2 (EqFP578) : MKIKVVEGGPLEAFDILATSFMYGSKTFINHTQGIPLDFKQS:83
SEC ID N° 4 (EqFP578m1) : MKIKVVEGGPLEAFDILATSFMYGSKAFINHTQGIPLDFKQS:83
SEC ID N° 6 (M1-NA) : MKIKVVEGGPLEAFDILATSFMYGSKAFINHTQGIPLDFKQS:86
SEC ID N° 8 (M1-602) : MKIKVVEGGPLEAFDILATSFMYGSKAFINHTQGIPLDFKQS:86
SEC ID N° 10 (M1-637) : MKIKVVEGGPLEAFDILATSFMYGSKAFINHTQGIPLDFKQS:86
SEC ID N° 12 (M1-mono1) : MKIKVVEGGPLEAFDILATSFMYGSRTFINHTQGIPLDFKQS:83
SEC ID N° 14 (nrM181-5) : MKIKVVEGGPLEAFDILATSFMYGSKTFINHTQGIPLDFKQS:83
SEC ID N° 16 (Cyan2-1) : MKIKVVEGGPLEAFDILATSFMYGSRTFINHTQGIPLDFKQS:83

          *           100          *           120
SEC ID N° 2 (EqFP578) : FPEGFTWERITTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:126
SEC ID N° 4 (EqFP578m1) : FPEGFTWERITTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:126
SEC ID N° 6 (M1-NA) : FPEGFTWERITTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:129
SEC ID N° 8 (M1-602) : FPEGFTWERITTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:129
SEC ID N° 10 (M1-637) : FPEGFTWERITTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:129
SEC ID N° 12 (M1-mono1) : FPEGFTWERVTTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:126
SEC ID N° 14 (nrM181-5) : FPEGFTWERVTTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:126
SEC ID N° 16 (Cyan2-1) : FPEGFTWERVTTYEDGGVLTATQDTSFQNGCIIYNVKINGVNF:126

          *           140          *           160           *
SEC ID N° 2 (EqFP578) : PSNGFVMQKKTIGWEANTEMLYPADSGLRCHSQMALKLVGGGY:169
SEC ID N° 4 (EqFP578m1) : PSNGFVMQKKTIGWEANTEMLYPADSGLRCHSQMALKLVGGGY:169
SEC ID N° 6 (M1-NA) : PSNGFVMQKKTIGWEANTEMLYPADSGLRCHSQMALKLVGGGY:172
SEC ID N° 8 (M1-602) : PSNGFVMQKKTIGWEANTEMLYPADSGLRCHSQMALKLVGGGY:172
SEC ID N° 10 (M1-637) : PSNGFVMQKKTIGWEANTEMLYPADSGLRCHSQMALKLVGGGY:172
SEC ID N° 12 (M1-mono1) : PSNGFVMQKKTIGWEANTEMLYPADSGLEGRSDMALKLVGGGH:169
SEC ID N° 14 (nrM181-5) : PSNGFVMQKKTIGWEANTEMLYPADSGLEGRSDMALKLVGGGH:169
SEC ID N° 16 (Cyan2-1) : PSNGFVMQKKTIGWEANTEMLYPADSGLEGRSDMALKLVGGGH:169

          180          *           200           *
SEC ID N° 2 (EqFP578) : LHCSLKTYYRSKKPAKNLKMPCGFHFDVDRLERIKEADKETYVE:212
SEC ID N° 4 (EqFP578m1) : LHCSLKTYYRSKKPAKNLKMPCGFHFDVDRLERIKEADKETYVE:212
SEC ID N° 6 (M1-NA) : LHCSLKTYYRSKKPAKNLRMPGFHFDVDRLERIKEADKETYVE:215
SEC ID N° 8 (M1-602) : LHCSLKTYYRSKKPAKNLKMPCGFHFDVDRLERIKEADKETYVE:215
SEC ID N° 10 (M1-637) : LHCSLKTYYRSKKPAKNLKMPCGFHFDVDRLERIKEADKETYVE:215
SEC ID N° 12 (M1-mono1) : LIANEKTTYRSKKPAKNLKMPCGVYVDFRLERIKEADKETYVE:212
SEC ID N° 14 (nrM181-5) : LIANEKTTYRSKKPAKNLKMPCGVYVDFRLERIKEADKETYVE:212
SEC ID N° 16 (Cyan2-1) : LIANEKTTYRSKKPAKNLKMPCGVYVDFRLERIKEADKETYVE:212

          220          *
SEC ID N° 2 (EqFP578) : QHEMAVAKYCDLPSKLGHR...:231
SEC ID N° 4 (EqFP578m1) : QHEMAVAKYCDLPSKLGHR...:231
SEC ID N° 6 (M1-NA) : QHEMAVAKYCDLPSKLGHS...:234
SEC ID N° 8 (M1-602) : QHEMAVAKYCDLPSKLGHS...:234
SEC ID N° 10 (M1-637) : QHEMAVAKYCDLPSKLGHS...:234
SEC ID N° 12 (M1-mono1) : QHEVAVARYCDLPSKLGHR...:231
SEC ID N° 14 (nrM181-5) : QHEMAVAKYCDLPSKLGHRKLN:233
SEC ID N° 16 (Cyan2-1) : QHEVAVARYCDLPSKLGHRKLN:233

```

Fig. 1

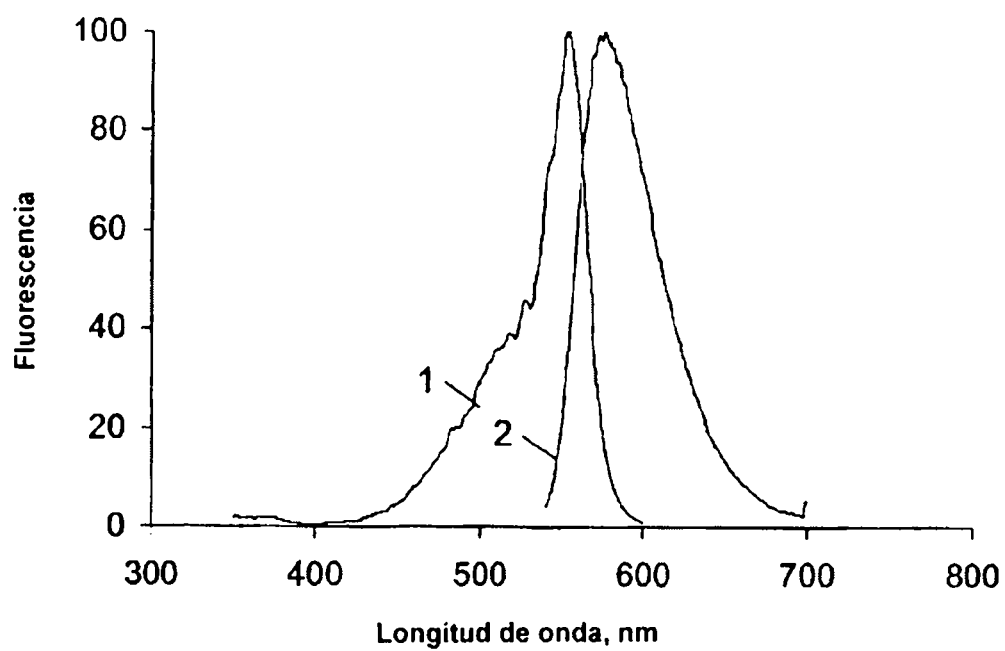


Fig. 2

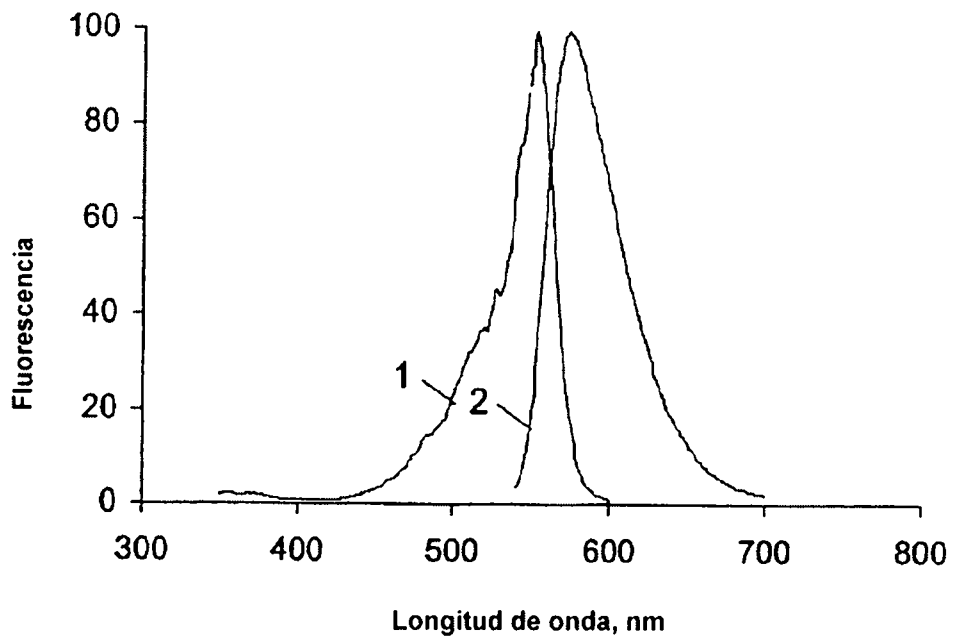


Fig. 3

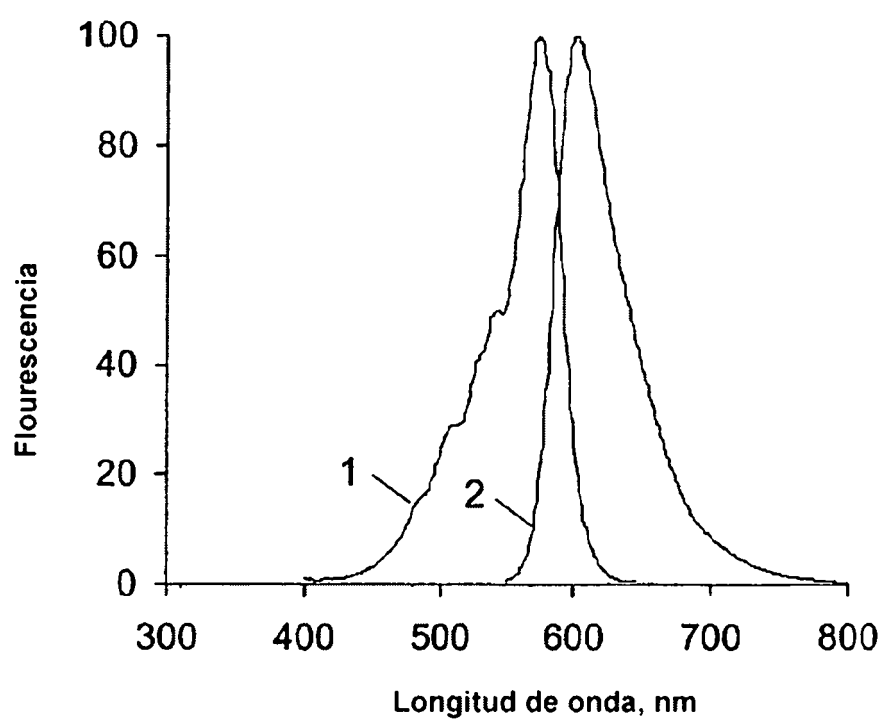


Fig. 4

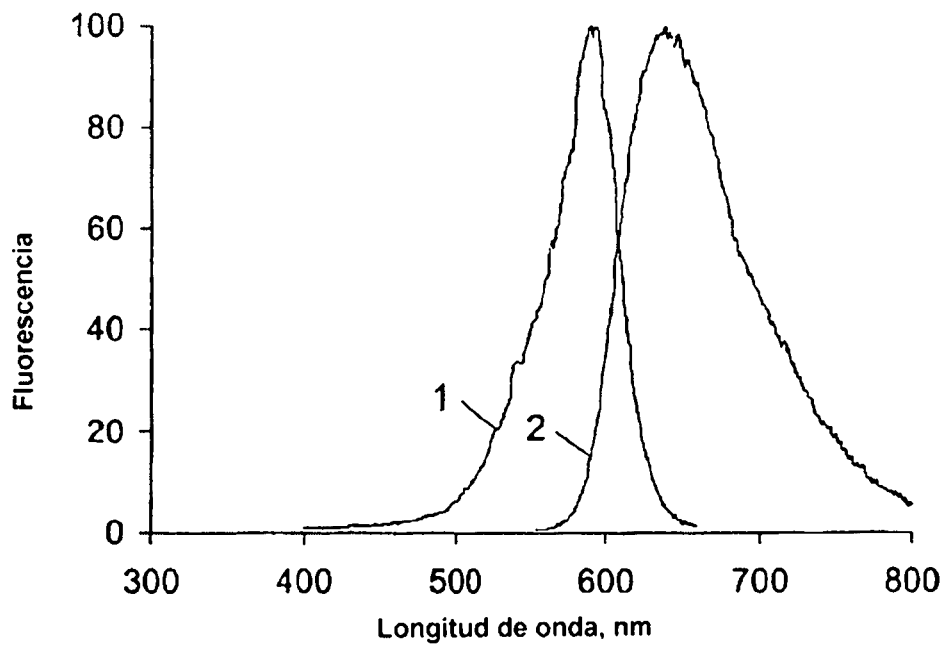


Fig. 5

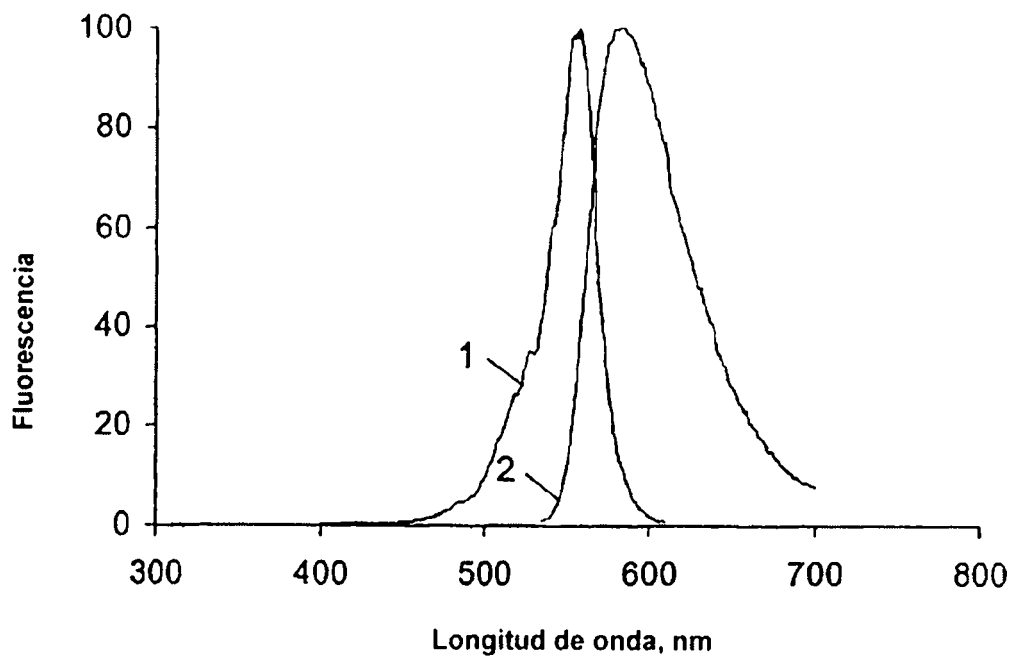


Fig. 6

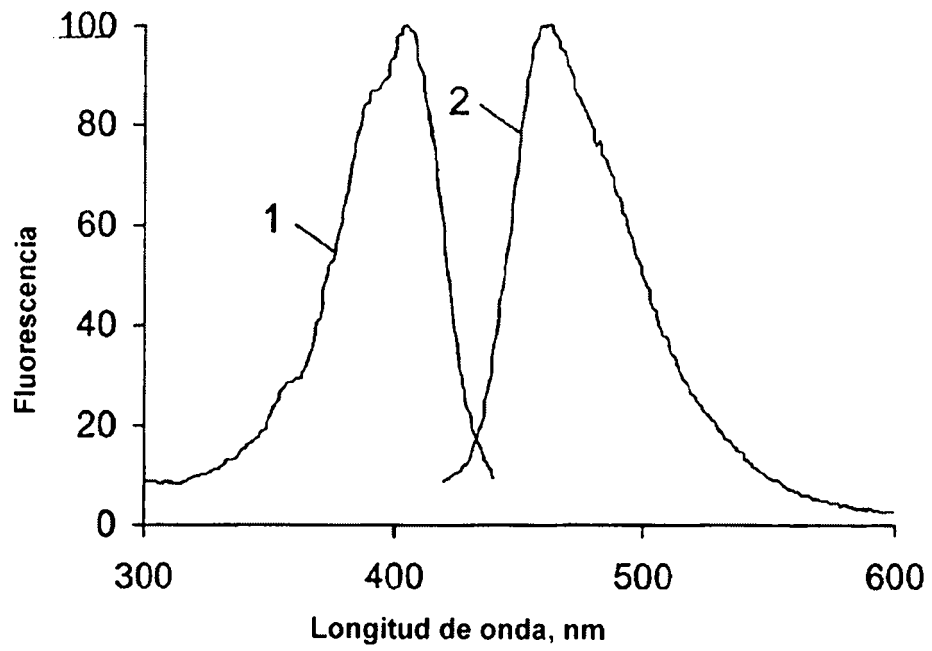


Fig. 7

ES 2 353 081 T3

LISTA DE SECUENCIAS

<110> Evrogen, IP

5 <120> NUEVAS PROTEÍNAS FLUORESCENTES Y MÉTODOS PARA USARLAS

<130> 13893R-EP

10 <140> EP 07705420.3

<141> 25-08-2008

15 <160> 18

<170> PatentIn version 3.1

<210> 1

20 <211> 696

<212> ADN

<213> *Entacmaea quadricolor*

25 <400> 1

```
    atgagtgaac tgatcaaaga aaacatgcac atgaagctgt acatggaagg tacggtcaat    60
30    aaccaccatt tcaaatgcac gcttgaagg gaacgcaagc catacgaagg aactcaaacc    120
    atgaagatca aagtcgtcga gggaggacct ctgccatttg cctttgacat tcttgcgaca    180
    agcttcatgt acggcagcaa gacttttacc aatcacactc aaggcattcc tgacttgttt    240
35    aagcagtcac ttctgaagg ttttacttgg gaaagaatta caacatacga ggatggggga    300
    gtccttactg ctacgcaaga caccagcctt caaaatggct gtatcattta caacgtcaag    360
40    atcaatggag taaactttcc ttcaaatggt tctgtgatgc aaaagaagac tcttggttgg    420
    gaggccaata cagagatgct gtatccagca gatggtggtc tgaggggaca ctcccaaatg    480
    gcactaaagc ttgttggagg tggctatctg cattgctctt tcaaaacaac ttacaggtca    540
45    aaaaagcctg cgaagaatth gaagatgccc ggtttccatt ttgttgatca ccgtttagaa    600
    aggataaagg aggctgacaa agaaacgtac gtggaacaac acgaaatggc agttgccaag    660
    tactgtgacc ttccatccaa actgggacat cgttaa    696
```

50

<210> 2

<211> 231

55 <212> PRT

<213> *Entacmaea quadricolor*

60

65

ES 2 353 081 T3

<400> 2

5 Met Ser Glu Leu Ile Lys Glu Asn Met His Met Lys Leu Tyr Met Glu
1 5 10 15

10 Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu Gly Glu Arg
20 25 30

15 Lys Pro Tyr Glu Gly Thr Gln Thr Met Lys Ile Lys Val Val Glu Gly
35 40 45

20 Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser Phe Met Tyr
50 55 60

25 Gly Ser Lys Thr Phe Ile Asn His Thr Gln Gly Ile Pro Asp Leu Phe
65 70 75 80

30 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Ile Thr Thr Tyr
85 90 95

35 Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser Leu Gln Asn
100 105 110

40 Gly Cys Ile Ile Tyr Asn Val Lys Ile Asn Gly Val Asn Phe Pro Ser
115 120 125

45 Asn Gly Ser Val Met Gln Lys Lys Thr Leu Gly Trp Glu Ala Asn Thr
130 135 140

50 Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Arg Gly His Ser Gln Met
145 150 155 160

55 Ala Leu Lys Leu Val Gly Gly Gly Tyr Leu His Cys Ser Phe Lys Thr
165 170 175

60 Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys Met Pro Gly Phe
180 185 190

65 His Phe Val Asp His Arg Leu Glu Arg Ile Lys Glu Ala Asp Lys Glu
195 200 205

70 Thr Tyr Val Glu Gln His Glu Met Ala Val Ala Lys Tyr Cys Asp Leu
210 215 220

75 Pro Ser Lys Leu Gly His Arg
225 230

<210> 3

<211> 696

<212> ADN

<213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de eqFP578m1

5 <400> 3

```

    atgagcgagc tgatcaagga gaacatgcac atgaagctgt acatggaggg caccgtgaac    60
10  aaccaccact tcaagtgcac atccgagggc gaaggcaagc cctacgaggg caccagacc    120
    atgaagatca aggtggtcga gggcggccct ctccccctcg ccttcgacat cctggctacc    180
    agcttcatgt acggcagcaa agccttcatc aaccacacc agggcatccc cgacttcttt    240
15  aagcagtcct tccctgaggg cttcacatgg gagagaatca ccacatacga agacgggggc    300
    gtgctgaccg ctaccagga caccagcttc cagaacggct gcatcatcta caacgtcaag    360
    atcaacgggg tgaacttccc atccaacggc cctgtgatgc agaagaaaac acgcggctgg    420
20  gaggccaaca ccgagatgct gtaccccgct gacggcggcc tgagaggcca cagccagatg    480
    gccctgaagc tcgtgggcgg gggctacctg cactgctcct tcaagaccac atacagatcc    540
25  aagaaacccg ctaagaacct caagatgcc ggcttccact tcgtggacca cagactggaa    600
    agaatcaagg aggccgacaa agagacctac gtcgagcagc acgagatggc tgtggccaag    660
    tactgcgacc tccctagcaa actggggcac agatga                                696

```

30

<210> 4

<211> 231

35 <212> PRT

<213> Secuencia artificial

<220>

40 <223> secuencia de aminoácidos de eqFP578m1

<400> 4

```

45  Met Ser Glu Leu Ile Lys Glu Asn Met His Met Lys Leu Tyr Met Glu
    1           5           10           15
50  Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu Gly Glu Gly
    20           25           30
55  Lys Pro Tyr Glu Gly Thr Gln Thr Met Lys Ile Lys Val Val Glu Gly
    35           40           45

```

55

60

65

ES 2 353 081 T3

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

<210> 5

<211> 705

<212> ADN

50 <213> Secuencia artificial

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante M1-NA

55

60

65

ES 2 353 081 T3

<400> 5

5	atgggtgagt atagcgagct gatcaccgag aacatgcaca tgaagctgta catggagggc	60
	accgtgaaca accaccactt caagtgacaca tccgagggcg aaggcaagcc ctacgagggc	120
10	acccagacca tgaagatcaa ggtggtcgag ggcggccctc tccccttcgc cttcgacatc	180
	ctggctacca gcttcatgta cggcagcaaa gccttcatca accacaccca gggcatcccc	240
	gacttcttta agcagtcctt ccctgagggc ttcacatggg agagaatcac cacatacgaa	300
15	gacgggggcg tgctgactgc taccagggac accagcttcc agaacggctg catcatctac	360
	aacgtcaaga tcaacggggt gaacttccca tccaacggcc ctgtgatgca gaagaaaaca	420
20	cgcggctggg aggccaacac cgagatgctg taccctgctg acggcggcct gagaggccac	480
	agccagatgg ccctgaagct cgtgggcggg ggctacctgc actgctcctt caagaccaca	540
	tacagatcca agaaaccgc taagaacctc aggatgcccg gcttccactt cgtggaccac	600
25	agactggaaa gaatcaagga agccgacaaa gagacctacg tcgagcagca cgagatggct	660
	gtggccaagt actgcgacct ccctagcaaa ctggggcaca gctga	705

30

<210> 6

<211> 234

<212> PRT

35

<213> Secuencia artificial

<220>

<223> secuencia de aminoácidos de la proteína fluorescente mutante M1-NA

40

45

50

55

60

65

ES 2 353 081 T3

<400> 6

5 Met Gly Glu Tyr Ser Glu Leu Ile Thr Glu Asn Met His Met Lys Leu
1 5 10 15

10 Tyr Met Glu Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu
20 25

15 Gly Glu Gly Lys Pro Tyr Glu Gly Thr Gln Thr Met Lys Ile Lys Val
30 35

20 Val Glu Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser
50 55 60

25 Phe Met Tyr Gly Ser Lys Ala Phe Ile Asn His Thr Gln Gly Ile Pro
65 70 75 80

30 Asp Phe Phe Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Ile
85 90 95

35 Thr Thr Tyr Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser
100 105 110

40 Phe Gln Asn Gly Cys Ile Ile Tyr Asn Val Lys Ile Asn Gly Val Asn
115 120 125

45 Phe Pro Ser Asn Gly Pro Val Met Gln Lys Lys Thr Arg Gly Trp Glu
130 135 140

50 Ala Asn Thr Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Arg Gly His
145 150 155 160

55 Ser Gln Met Ala Leu Lys Leu Val Gly Gly Gly Tyr Leu His Cys Ser
165 170 175

60 Phe Lys Thr Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Arg Met
180 185 190

65 Pro Gly Phe His Phe Val Asp His Arg Leu Glu Arg Ile Lys Glu Ala
195 200 205

70 Asp Lys Glu Thr Tyr Val Glu Gln His Glu Met Ala Val Ala Lys Tyr
210 215 220

75 Cys Asp Leu Pro Ser Lys Leu Gly His Ser
225 230

<210> 7

<211> 705

65 <212> ADN

<213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante M1-602

5 <400> 7

```
atgggtgagg atagcgagct gatcaccgag aacatgcaca tgaaactgta catggagggc 60
accgtgaaca accaccactt caagtgcaca tccgagggcg aaggcaagcc ctacgagggc 120
10 acccagacca tgaagatcaa ggtggtcgag ggcggccctc tccccttcgc cttcgacatc 180
ctggctacca gcttcatgta cggcagcaaa gccttcatca accacacca gggcatcccc 240
15 gacttcttta agcagtcctt ccctgagggc ttcacatggg agagaatcac cacatacgaa 300
gacggggggc tgctgaccgc taccaggac accagcctc agaacggctg cctcatctac 360
aacgtcaaga tcaacggggt gaacttccca tccaacggcc ctgtgatgca gaagaaaaca 420
20 ctcggctggg aggccagcac cgagatgctg taccgcctg acagcggcct gagaggccat 480
ggccagatgg ccctgaagct cgtgggcggg ggctacctgc actgctccct caagaccaca 540
25 tacagatcca agaaacccgc taagaacctc aagatgcccg gcttccactt cgtggaccac 600
agactggaaa gaatcaagga ggccgacaaa gagacctacg tcgagcagca cgagatggct 660
gtggccaagt actgcgacct ccctagcaaa ctggggcaca gctga 705
```

30

<210> 8

<211> 234

35 <212> PRT

<213> Secuencia artificial

<220>

40 <223> secuencia de aminoácidos de la proteína fluorescente mutante M1-602

45

50

55

60

65

ES 2 353 081 T3

<400> 8

5 Met Gly Glu Asp Ser Glu Leu Ile Thr Glu Asn Met His Met Lys Leu
1 5 10

10 Tyr Met Glu Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu
20 25 30

15 Gly Glu Gly Lys Pro Tyr Glu Gly Thr Gln Thr Met Lys Ile Lys Val
35 40 45

20 val Glu Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser
50 55 60

25 Phe Met Tyr Gly Ser Lys Ala Phe Ile Asn His Thr Gln Gly Ile Pro
65 70 75 80

30 Asp Phe Phe Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Ile
85 90 95

35 Thr Thr Tyr Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser
100 105 110

40 Leu Gln Asn Gly Cys Leu Ile Tyr Asn Val Lys Ile Asn Gly Val Asn
115 120 125

45 Phe Pro Ser Asn Gly Pro Val Met Gln Lys Lys Thr Leu Gly Trp Glu
130 135 140

50 Ala Ser Thr Glu Met Leu Tyr Pro Ala Asp Ser Gly Leu Arg Gly His
145 150 155 160

55 Gly Gln Met Ala Leu Lys Leu Val Gly Gly Gly Tyr Leu His Cys Ser
165 170 175

60 Leu Lys Thr Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys Met
180 185 190

65 Pro Gly Phe His Phe Val Asp His Arg Leu Glu Arg Ile Lys Glu Ala
195 200 205

70 Asp Lys Glu Thr Tyr Val Glu Gln His Glu Met Ala Val Ala Lys Tyr
210 215 220

75 Cys Asp Leu Pro Ser Lys Leu Gly His Ser
225 230

<210> 9

<211> 705

65 <212> ADN

<213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante M1-637

5 <400> 9

```
atgggtgagg atagcgagct gatcaccgag aacatgcaca tgaaactgta catggagggc 60
accgtgaaca accaccactt caagtgcaca tccgagggcg aaggcaagcc ctacgagggc 120
accagacca tgaagatcaa ggtggtcgag ggcggccctc tccccttcgc cttcgacatc 180
ctggctacca gcttcatgta cggcagcaaa gccttcatca accacacca gggcatcccc 240
gacttcttta agcagtcctt ccctgagggc ttcacatggg agagaatcac cacatacgaa 300
gacgggggcg tgctgaccgc taccaggac accagcctcc agaacggctg cctcatctac 360
aacgtcaaga tcaacggggt gaacttccca tccaacggcc ctgtgatgca gaagaaaaca 420
ctcggctggg aggccagcac cgagatgctg taccctgctg acagcggcct gagaggccat 480
agccagatgg ccctgaagct cgtgggcggg ggctacctgc actgctccct caagaccaca 540
tacagatcca agaaacccgc taagaacctc aagatgcccg gcttccactt cgtggacagg 600
agactggaaa gaatcaagga ggccgacaaa gagacctacg tcgagcagca cgagatggct 660
gtggccaagt actgcgacct ccctagcaaa ctggggcaca gctga 705
```

30

<210> 10

<211> 234

35 <212> PRT

<213> Secuencia artificial

<220>

40 <223> secuencia de aminoácidos de la proteína fluorescente mutante M1-637

45

50

55

60

65

ES 2 353 081 T3

<400> 10

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

<210> 11

<211> 693

<212> ADN

65 <213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante M1-mono1

5 <400> 11

```
    atgagcgagc tgattaagga gaacatgcac atgaagctgt acatggaggg caccgtgaac    60
    aaccaccact tcaagtgcac atccgagggc gaaggcaagc cctacgaggg caccagacc    120
10  atgagaatca aggtggtcga gggcggcctt ctccccttcg ccttcgacat cctggctacc    180
    agcttcatgt acggcagcag aaccttcac aaccacacc agggcatccc cgacttcttt    240
15  aagcagtcct tccctgaggg cttcacatgg gagagagtca ccacatacga agacgggggc    300
    gtgctgaccg ctaccagga caccagcctc caggacggct gcctcatcta caacgtcaag    360
    atcagagggg tgaacttccc atccaacggc cctgtgatgc agaagaaaac actcggctgg    420
20  gaggccaaca ccgagatgct gtaccccgtc gacggcggcc tgggaaggcag aagcgacatg    480
    gccctgaagc tcgtgggagg gggccacctg atctgcaact tcaagaccac atacagatcc    540
    aagaaacccg ctaagaacct caagatgccg ggcgttact atgtggacca cagactggaa    600
25  agaatcaagg aggccgacaa agagacctac gtcgagcagc acgaggtggc tgtggccaga    660
    tactgagacc tccctagcaa actggggcac aag                                693
```

30

<210> 12

<211> 231

<212> PRT

35 <213> Secuencia artificial

<220>

<223> secuencia de aminoácidos de la proteína fluorescente mutante M1-mono1

40

45

50

55

60

65

ES 2 353 081 T3

<400> 12

5 Met Ser Glu Leu Ile Lys Glu Asn Met His Met Lys Leu Tyr Met Glu
1 5 10 15

10 Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu Gly Glu Gly
20 25 30

15 Lys Pro Tyr Glu Gly Thr Gln Thr Met Arg Ile Lys Val Val Glu Gly
35 40 45

20 Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser Phe Met Tyr
50 55 60

25 Gly Ser Arg Thr Phe Ile Asn His Thr Gln Gly Ile Pro Asp Phe Phe
65 70 75 80

30 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Val Thr Thr Tyr
85 90 95

35 Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser Leu Gln Asp
100 105 110

40 Gly Cys Leu Ile Tyr Asn Val Lys Ile Arg Gly Val Asn Phe Pro Ser
115 120 125

45 Asn Gly Pro Val Met Gln Lys Lys Thr Leu Gly Trp Glu Ala Asn Thr
130 135 140

50 Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Glu Gly Arg Ser Asp Met
145 150 155 160

55 Ala Leu Lys Leu Val Gly Gly Gly His Leu Ile Cys Asn Phe Lys Thr
165 170 175

60 Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys Met Pro Gly Val
180 185 190

65 Tyr Tyr Val Asp His Arg Leu Glu Arg Ile Lys Glu Ala Asp Lys Glu
195 200 205

70 Thr Tyr Val Glu Gln His Glu Val Ala Val Ala Arg Tyr Cys Asp Leu
210 215 220

75 Pro Ser Lys Leu Gly His Lys
225 230

<210> 13

<211> 702

<212> ADN

65 <213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante nrM181-5

5 <400> 13

```

    atgagcgagc tgattaagga gaacatgcac atgaagctgt acatggaggg caccgtgaac    60
10    aaccaccact tcaagtgcac atccgagggc gaaggcaagc cctacgaagg caccagacc    120
    atgaaaatca aggtggtcga gggcggccct ctccccttcg ccttcgacat cctggctacc    180
    agcttcatgt acggcagcaa aaccttcac aaccacaccc agggcatccc cgacttcttt    240
15    aagcagtcct tccctgaggg cttcacatgg gagagagtca ccacatacga agacgggggc    300
    gtgctgaccg ctaccagga caccagcctc caggacggct gcctcatcta caacgtcaag    360
    atcagagggg tgaacttccc atccaacggc cctgtgatgc agaagaaac actcggtgg    420
20    gaggcccaca ccgagatgct gtaccccgct gacggcggcc tggaaggcag aagcgacatg    480

    gccctgaagc tcgtgggcgg gggccacctg atcgccaact tcaagacca atacagatcc    540
25    aagaaacccg ctaagaacct caagatgccc ggcgtctact atgtggacca cagactggaa    600
    agaatcaagg aggccgacaa agagacctac gtcgagcagc acgagatggc tgtggccaaa    660
30    tactgcgacc tcctagcaa actggggcac aagcttaatt ag                                702
```

<210> 14

35 <211> 233

<212> PRT

<213> Secuencia artificial

40 <220>

<223> secuencia de aminoácidos de la proteína fluorescente mutante nrM181-5

45

50

55

60

65

ES 2 353 081 T3

<400> 14

5 Met Ser Glu Leu Ile Lys Glu Asn Met His Met Lys Leu Tyr Met Glu
1 5 10

10 Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu Gly Glu Gly
20 25 30

15 Lys Pro Tyr Glu Gly Thr Gln Thr Met Lys Ile Lys Val Val Glu Gly
35 40 45

20 Gly Pro Leu Pro Phe Ala Phe Asp Ile Leu Ala Thr Ser Phe Met Tyr
50 55 60

25 Gly Ser Lys Thr Phe Ile Asn His Thr Gln Gly Ile Pro Asp Phe Phe
65 70 75 80

30 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Val Thr Thr Tyr
85 90 95

35 Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser Leu Gln Asp
100 105 110

40 Gly Cys Leu Ile Tyr Asn Val Lys Ile Arg Gly Val Asn Phe Pro Ser
115 120 125

45 Asn Gly Pro Val Met Gln Lys Lys Thr Leu Gly Trp Glu Ala His Thr
130 135 140

50 Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Glu Gly Arg Ser Asp Met
145 150 155 160

55 Ala Leu Lys Leu Val Gly Gly Gly His Leu Ile Ala Asn Phe Lys Thr
165 170 175

60 Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys Met Pro Gly Val
180 185 190

65 Tyr Tyr Val Asp His Arg Leu Glu Arg Ile Lys Glu Ala Asp Lys Glu
195 200 205

70 Thr Tyr Val Glu Gln His Glu Met Ala Val Ala Lys Tyr Cys Asp Leu
210 215 220

75 Pro Ser Lys Leu Gly His Lys Leu Asn
225 230

<210> 15

<211> 702

65 <212> ADN

<213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> secuencia de nucleótidos de la proteína fluorescente mutante Cyan-2-1

5 <400> 15

```
    atgagcgagc tgattaagga gaacatgcac atgaagctgt acatggaggg caccgtgaac    60
    aaccaccact tcaagtgcac atccgagggc gaaggcaagc cctacggagg cacccagacc    120
10  atgagaatca aggtggtcga gggcgccctt ctcccgtcg ccttcgacat cctggctacc    180
    agcttcatgt acggcagcag aaccttcatt aaccacacc agggcatccc cgacttcttt    240
15  aagcagtcct tcctgaggg cttcacatgg gagagagtca ccacatacga agacgggggc    300
    gtgctgaccg ctaccagga caccagcctc caggacggct gcctcatcta caacgtcaag    360
    atcagagggg tgaacttccc atccaacggc cctgtgatgc agaagaaaac actcggctgg    420
20  gaggccttca ccgagatgct gtaccccgtt gacggcggcc tggaaggcag aagcgacatg    480
    gccctgaagc tcgtgggagg gggccacctg atctgcaact tcaagaccac atacagatcc    540
    aagaaacccg ctaagaacct caagatgccc ggcgtctact atgtggacta cagactggaa    600
25  agaatcaagg aggccgacaa agagacctac gtcgagcagc acgaggtggc tgtggccaga    660
    tactgagacc tcctagcaa actggggcac aagcttaatt ag                                702
```

30

<210> 16

<211> 233

<212> PRT

35

<213> Secuencia artificial

<220>

40

<223> secuencia de aminoácidos de la proteína fluorescente mutante Cyan-2-1

45

50

55

60

65

ES 2 353 081 T3

<400> 16

5 Met Ser Glu Leu Ile Lys Glu Asn Met His Met Lys Leu Tyr Met Glu
1 5 10

10 Gly Thr Val Asn Asn His His Phe Lys Cys Thr Ser Glu Gly Glu Gly
20 25 30

15 Lys Pro Tyr Gly Gly Thr Gln Thr Met Arg Ile Lys Val Val Glu Gly
35 40 45

20 Gly Pro Leu Pro Val Ala Phe Asp Ile Leu Ala Thr Ser Phe Met Tyr
50 55 60

25 Gly Ser Arg Thr Phe Ile Asn His Thr Gln Gly Ile Pro Asp Phe Phe
65 70 75 80

30 Lys Gln Ser Phe Pro Glu Gly Phe Thr Trp Glu Arg Val Thr Thr Tyr
85 90 95

35 Glu Asp Gly Gly Val Leu Thr Ala Thr Gln Asp Thr Ser Leu Gln Asp
100 105 110

40 Gly Cys Leu Ile Tyr Asn Val Lys Ile Arg Gly Val Asn Phe Pro Ser
115 120 125

45 Asn Gly Pro Val Met Gln Lys Lys Thr Leu Gly Trp Glu Ala Phe Thr
130 135 140

50 Glu Met Leu Tyr Pro Ala Asp Gly Gly Leu Glu Gly Arg Ser Asp Met
145 150 155 160

55 Ala Leu Lys Leu Val Gly Gly Gly His Leu Ile Cys Asn Phe Lys Thr
165 170 175

60 Thr Tyr Arg Ser Lys Lys Pro Ala Lys Asn Leu Lys Met Pro Gly Val
180 185 190

65 Tyr Tyr Val Asp Tyr Arg Leu Glu Arg Ile Lys Glu Ala Asp Lys Glu
195 200 205

70 Thr Tyr Val Glu Gln His Glu Val Ala Val Ala Arg Tyr Cys Asp Leu
210 215 220

75 Pro Ser Lys Leu Gly His Lys Leu Asn
225 230

<210> 17

<211> 4

65 <212> PRT

<213> Secuencia artificial

ES 2 353 081 T3

<220>

<223> composición de aminoácidos para el extremo N de proteínas fluorescentes no agregantes

5 <400> 17

Met Gly Glu Tyr
1

10

<210> 18

<211> 4

<212> PRT

15

<213> Secuencia artificial

<220>

20 <223> composición de aminoácidos para el extremo N de proteínas fluorescentes no agregantes

<400> 18

25

Met Gly Glu Asp
1

30

35

40

45

50

55

60

65