

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 606 706**

51 Int. Cl.:

**C12N 15/10** (2006.01)

**G01N 33/542** (2006.01)

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.01.2008 E 08000297 (5)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 2078750**

54 Título: **Un ensayo de doble híbrido fluorescente (F2H) para la visualización directa de interacciones de proteínas en células vivas**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**27.03.2017**

73 Titular/es:  
**LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN  
(100.0%)  
GESCHWISTER-SCHOLL-PLATZ 1  
80359 MÜNCHEN, DE**

72 Inventor/es:  
**ROTHBAUER, ULRICH;  
LEONHARDT, HEINRICH;  
ZOLGHADR, KOUROSH y  
MORTUSEWICZ, OLIVER**

74 Agente/Representante:  
**ARIAS SANZ, Juan**

ES 2 606 706 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

Un ensayo de doble híbrido fluorescente (F2H) para la visualización directa de interacciones de proteínas en células vivas

5 La presente invención se refiere a un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende: (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (ii) un (poli)péptido que se une específicamente a GFP; (b) expresar en la misma célula una segunda proteína de fusión que comprende (i) GFP; y (ii) un (poli)péptido señuelo; (c) expresar en la misma célula una tercera proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y (ii) un (poli)péptido presa; y (d) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación, en el que una colocalización de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del (poli)péptido señuelo y presa. La invención también se refiere a un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende: (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido fluorescente; (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (iii) un (poli)péptido señuelo; (b) expresar en la misma célula una segunda proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y (ii) un (poli)péptido presa y (c) detectar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula tras la excitación, en el que una colocalización de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del (poli)péptido señuelo y presa. Además, la presente invención se refiere a procedimientos para identificar un compuesto que modula la interacción de dos (poli)péptidos y procedimientos de determinación de la fuerza relativa de la interacción de dos proteínas con una tercera proteína.

En la presente memoria descriptiva, se citan varios documentos.

25 Después de la secuenciación del genoma humano, el siguiente reto ahora es analizar las redes de proteínas complejas subyacentes a las funciones celulares. En la última década, se ha desarrollado una amplia variedad de procedimientos para estudiar las interacciones proteína-proteína que varían desde enfoques bioquímicos hasta genéticos o basados en células. Los procedimientos bioquímicos, tales como la purificación por afinidad o la coimmunoprecipitación (Co-IP), permiten la detección de complejos de proteínas *in vitro*. Los procedimientos genéticos, tales como el sistema de doble híbrido de levaduras (Y2H), permiten el cribado ultrarrápido eficaz de interacciones dentro del entorno celular. El análisis de las interacciones proteína-proteína de mamífero en levadura puede verse afectado, sin embargo, por la ausencia o insuficiente conservación de los factores celulares que modulan las interacciones proteína-proteína, por ejemplo, a través de modificaciones postraduccionales (Parrish et al., 2006). Además, este procedimiento es laborioso y propenso a errores.

35 En los últimos años se han introducido nuevos procedimientos basados en fluorescencia para la visualización dentro de la célula de interacciones proteína-proteína. Dos técnicas establecidas, la transferencia de energía de resonancia de fluorescencia (FRET) (Miyawaki, 2003; Sekar y Periasamy, 2003) y la complementación de fluorescencia bimolecular (BiFC) (Kerppola, 2006), se basan en la expresión de proteínas marcadas con fluorescencia o fragmentos de las mismas. Sin embargo, FRET requiere una instrumentación costosa y conocimientos técnicos avanzados, mientras que BiFC se basa en la complementación irreversible y la maduración lenta de fluoróforos, lo que no permite la detección en tiempo real de las interacciones proteína-proteína (Kerppola, 2006).

Tumbar *et al.* 1999 describe una molécula de ácido nucleico que codifica una proteína de fusión que comprende una proteína fluorescente, una NLS y una proteína señuelo (el dominio ácido de activación de VP16), pero no divulga la expresión estable de la construcción en células.

45 El documento US 2006/051739 describe procedimientos de visualización e interferencia con la adhesión y segregación cromosómica de moléculas extracromosómicas, sin embargo, ni una proteína fluorescente ni una proteína señuelo está presente en las proteínas de fusión empleadas. También se divulgan construcciones de fusión que comprenden GFP o YFP, el represor de lac LacI y una señal de localización nuclear, pero la construcción de fusión carece de una proteína señuelo.

50 El documento WO 2005/030989 describe procedimientos de detección de proteínas usando aptámeros. Específicamente, el documento WO 2005/030989 divulga ácidos nucleicos que comprenden fusiones del epítipo FLAG con el represor de lac LacI o el represor de tet TetR. Se divulgan, además, fusiones que comprenden ácidos nucleicos que codifican FLAG, LacI, S35, HA, TetR y PML. No se describen ácidos nucleicos que codifican una proteína de fusión que comprende una proteína fluorescente y una proteína que se acumula en sitios distintos en el núcleo de la célula y una proteína señuelo.

El documento WO 02/086450 describe un dominio de dirección que se usa para inmovilizar el (poli)péptido señuelo en una localización subcelular. De forma importante, la proteína señuelo no está fusionada a una proteína fluorescente, pero la "fusión señuelo" consiste en un receptor de ligando de la fusión señuelo que puede ser

ecdisona, así como un compuesto adicional. Por lo tanto, en el documento WO 02/086450 la proteína señuelo no está marcada de forma detectable y se requiere una evaluación indirecta de la localización por medio de un marcador detectable fusionado a la proteína presa, lo que puede conducir a una frecuencia potenciada de resultados positivos falsos ya que la proteína presa detectable se puede agregar o unir a estructuras celulares a través de interacciones no específicas (incluyendo interacciones hidrófilas o hidrófobas) no proporcionadas por la proteína señuelo.

Todos estos procedimientos tienen inconvenientes inherentes y se combinan típicamente para obtener resultados más fiables.

Por lo tanto, existe la necesidad de procedimientos mejorados de detección de interacciones proteína-proteína.

La solución a dicho problema técnico se consigue proporcionando los modos de realización caracterizados en las reivindicaciones.

En consecuencia, en un primer aspecto, la presente invención se refiere a un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende: (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (ii) un (poli)péptido que se une específicamente a GFP; (b) expresar en la misma célula una segunda proteína de fusión que comprende (i) GFP; y (ii) un (poli)péptido señuelo; (c) expresar en la misma célula una tercera proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y (ii) un (poli)péptido presa; y (d) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación, en el que una colocalización de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del (poli)péptido señuelo y presa.

La expresión "interacciones proteína-proteína" se refiere a la interacción específica de dos o más compuestos proteicos, es decir, poli(péptidos) o proteínas. La interacción específica se caracteriza por una fuerza o afinidad de unión mínima. Las afinidades de unión para interacciones específicas alcanzan, en general, el intervalo desde pM hasta mM y también dependen en gran medida del entorno químico, por ejemplo, el valor de pH, la fuerza iónica, la presencia de cofactores etc. En el contexto de la presente invención, la expresión se refiere particularmente a interacciones proteína-proteína que se producen en condiciones fisiológicas, es decir, en una célula.

La expresión "expresar en una célula eucariota" se refiere a la transcripción y traducción de las proteínas de fusión de la invención usando elementos apropiados de control de la expresión que funcionan en la célula elegida. De esta manera, se pueden ensayar las propiedades de unión de proteínas de fusión individuales en sistemas de expresión celulares. Con este fin, una molécula de ácido nucleico que codifica una proteína de fusión se puede clonar en un vector de expresión adecuado, cuya composición depende, en general, del sistema de expresión. Para la presente invención, el sistema de expresión es eucariota, preferentemente de mamífero. Un vector de expresión de mamífero típico contiene un elemento promotor, que media el inicio de la transcripción del ARNm, la secuencia codificante de la proteína y señales requeridas para la terminación de la transcripción y la poliadenilación del transcrito. Elementos adicionales podrían incluir potenciadores, secuencias Kozak y secuencias intermedias flanqueadas por sitios donantes y aceptores para corte y empalme del ARN. Se puede conseguir una transcripción altamente eficaz con los promotores temprano y tardío de SV40, las repeticiones terminales largas (LTR) de retrovirus, por ejemplo, RSV, HTLV1, HIV1 y el promotor temprano de citomegalovirus (CMV). Sin embargo, también se pueden usar elementos celulares (por ejemplo, el promotor de la actina humana). Posibles ejemplos de elementos reguladores que garantizan el inicio de la transcripción comprenden el promotor de citomegalovirus (CMV), el promotor de RSV (virus del sarcoma de Rous), el promotor lacZ, el promotor gal10, el promotor del factor de elongación 1a humano, el potenciador de CMV, el promotor de la CaM-cinasa, el promotor poliédrico del virus de la poliodrosis nuclear múltiple de *Autographa californica* (AcMNPV) o el potenciador de SV40. Ejemplos de señales de terminación de la transcripción son el sitio poli-A de SV40 o el sitio poli-A de tk o las señales de poliadenilación de SV40, lacZ y poliédricas de AcMNPV, en dirección 3' del polinucleótido. Además, también se pueden incluir elementos tales como el origen de replicación, genes de resistencia a fármacos, reguladores (como parte de un promotor inducible) o sitios internos de entrada del ribosoma (IRES).

Los vectores de expresión adecuados para *Drosophila* son los que pertenecen al sistema pMT DES (Invitrogen) que usa el promotor de la metalotioneína (MT) de *Drosophila* (Bunch et al., 1988) o pAC5.1 que usa el promotor de la actina 5C de *Drosophila*. Un vector que usa el promotor USA inducible por GAL4 es pUAST. Los vectores de levadura son el vector pYEp (que usa un promotor Gal10), pYX142 (vector de una única copia) o pYX232 (plásmido 2 $\mu$  que usa el promotor de la triosafosfato isomerasa TPI) (ambos de Novagen).

Las células huésped de mamífero que se podrían usar incluyen, pero no se restringen a, células humanas Hela, 293, H9, SH-EP1 y Jurkat, células de ratón NIH3T3 y C2C12, células Cos 1, Cos 7 y CV1, células de codorniz QC1-3, células L de ratón, células de riñón de cría de hámster sirio (BHK) y células de ovario de hámster chino (CHO). De forma alternativa, las proteínas de fusión se pueden expresar en líneas celulares estables que contienen la construcción génica integrada en un cromosoma. La cotransfección con un marcador de selección tal como dhfr, gpt, neomicina, higromicina permite la identificación y aislamiento de las células transfectadas. La molécula de ácido

nucleico transfectada también se puede amplificar en la célula para que exprese grandes cantidades de la proteína de fusión codificada. El marcador DHFR (dihidrofolato reductasa) es útil para desarrollar líneas celulares que porten varios cientos o incluso varios miles de copias del gen de interés. Otro marcador de selección útil es la enzima glutamina sintasa (GS) (Bebbington et al., 1992; Murphy et al., 1991). Usando estos marcadores, las células de mamífero se cultivan en medio selectivo y se seleccionan las células con la resistencia más alta. Los medios de cultivo y condiciones apropiadas para las células huésped descritas anteriormente son conocidos en la técnica.

La expresión "proteína de fusión" se refiere a proteínas quiméricas que consisten en secuencias derivadas de al menos dos proteínas o (poli)péptidos diferentes. De acuerdo con el contenido de la presente invención, en las proteínas de fusión ejemplares, un (poli)péptido señuelo se fusiona a un (poli)péptido fluorescente y a un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula. De forma alternativa, un (poli)péptido presa se fusiona a un (poli)péptido fluorescente. La fusión se puede realizar por cualquier técnica conocida para los expertos en la técnica, siempre que dé como resultado la fusión en fase de las moléculas de ácido nucleico que codifican los componentes de las proteínas de fusión de la invención. La fusión de los componentes se puede lograr en cualquier orden. Convencionalmente, la generación de una proteína de fusión a partir de dos o más (poli)péptidos o dominios diferentes se basa en el "corte y empalme bilateral por extensión solapante" descrito en (Horton et al., 1989). Los fragmentos que codifican los (poli)péptidos individuales se generan en dos reacciones de PCR primarias por separado. Los cebadores internos para las reacciones de PCR primarias contienen una región complementaria significativa, de aproximadamente 20 pb, que permite la fusión de los dos fragmentos de dominio en la segunda PCR. De forma alternativa, las regiones codificantes se pueden fusionar haciendo uso de sitios de restricción que pueden ser de origen natural o se pueden introducir por tecnología de ADN recombinante.

El término "(poli)péptido", como se usa en el presente documento, describe un grupo de moléculas que comprende el grupo de péptidos, que consisten en un máximo de 30 aminoácidos, así como el grupo de polipéptidos, que consisten en más de 30 aminoácidos. Además, de acuerdo con la definición del término "(poli)péptido", describe fragmentos de proteínas. Los (poli)péptidos pueden formar, además, dímeros, trímeros y oligómeros superiores, es decir, que consisten en más de una molécula de (poli)péptido. Las moléculas de (poli)péptido que forman dichos dímeros, trímeros, etc. pueden ser idénticas o no idénticas. Las correspondientes estructuras de orden superior se denominan, en consecuencia, homo- o heterodímeros, homo- o heterotrímeros etc. Los términos "(poli)péptido" y "proteína" se refieren también a (poli)péptidos/proteínas modificadas de forma natural en las que la modificación se logra, por ejemplo, por glucosilación, acetilación, fosforilación y similares. Dichas modificaciones son bien conocidas en la técnica.

La expresión "(poli)péptido fluorescente" o "proteína fluorescente" se refiere a (poli)péptidos que emiten luz fluorescente tras su excitación a una longitud de onda específica. Se puede usar una variedad de proteínas fluorescentes en la presente invención. Un grupo de dichas proteínas fluorescentes incluye la proteína verde fluorescente aislada de *Aequorea victoria* (GFP), así como varias variantes de GFP, tales como la proteína cian fluorescente, la proteína azul fluorescente, la proteína amarilla fluorescente, etc. (Zhang et al., 2002; Zimmer, 2002). Típicamente, estas variantes comparten aproximadamente un 80 % o mayor identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 1 o la secuencia de ácido nucleico de la SEQ ID NO: 2, respectivamente. Los mutantes de GFP de desplazamiento de color tienen colores de emisión de azul a verde amarillento, brillo aumentado y fotoestabilidad (Tsien, 1998). Uno de dichos mutantes de GFP, denominado proteína amarilla fluorescente potenciada, muestra un máximo de emisión a 529 nm. También se han descrito variantes adicionales basadas en GFP que tienen espectros modificados de excitación y emisión (Tsien et al., solicitud de patente de EE. UU. 200201231 13A1), intensidad de fluorescencia potenciada y tolerancia térmica (Thastrup et al., solicitud de patente de EE. UU. 20020107362A1; Bjorn et al., solicitud de patente de EE. UU. 20020177189A1) y la formación de cromóforo a niveles reducidos de oxígeno (Fisher, patente de EE. UU. N.º 6.414.119).

Otro grupo de proteínas fluorescentes incluye las proteínas fluorescentes aisladas de antozoos incluyendo, sin limitación, la proteína roja fluorescente aislada de especies *Discosoma* de coral, DsRed (Matz et al., 1999), por ejemplo, la secuencia de aminoácidos de la SEQ ID NO: 3 o la secuencia de ácido nucleico de la SEQ ID NO: 4, respectivamente (véase, por ejemplo, el número de acceso AF168419 versión AF168492). DsRed y las otras proteínas fluorescentes de antozoos comparten solo aproximadamente un 26-30 % de identidad de secuencia de aminoácidos con la GFP natural de *Aequorea victoria*, aunque todos los motivos cruciales están conservados, lo que indica la formación de la estructura de barril beta de 11 hebras característica de GFP. La estructura cristalina de DsRed también se ha resuelto y muestra conservación de la estructura de barril beta de 11 hebras de GFP (MMDB Id: 5742).

Se han descrito varios mutantes de la proteína roja fluorescente DsRed de longitud de longitud de onda más larga y, de manera similar, se pueden emplear en la generación de las proteínas de fusión de la invención que comprenden (poli)péptidos fluorescentes. Por ejemplo, los mutantes DsRed recientemente descritos con espectros de emisión desplazados más hacia el rojo se pueden emplear en la práctica de la invención (Baird et al., 2000; Terskikh et al., 2000; Wiehler et al., 2001).

Versiones monoméricas de DsRed son, por ejemplo, mRFP (por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID NO: 5 o la secuencia de ácido nucleico de la SEQ ID NO: 6), mRFP1 (Campbell et al., 2002), mCherry,

mOrange o mPlum (Shaner et al., 2004) o TagRFP (Merzlyak et al., 2007).

Más recientemente, se describieron las GFP de los antozoos *Renilla reniformis* y *Renilla kollikeri* (Ward et al., solicitud de patente de EE. UU. 20030013849).

5 Un número cada vez mayor de otras proteínas fluorescentes de varias formas marinas de vida se ha descrito recientemente, y el Protein Data Bank actualmente enumera varias estructuras cristalinas de GFP y GFP mutante, así como las estructuras cristalinas de varios análogos de GFP. Se han descrito proteínas fluorescentes relacionadas con estructuras inferidas como similares a GFP a partir de corales, plumas de mar, ascidias y anémonas de mar, y se pueden usar en la generación de las proteínas de fusión de la invención que comprenden (poli)péptidos fluorescentes (para revisiones, véase (Zhang et al., 2002; Zimmer, 2002)).

10 También se ha informado de proteínas fluorescentes de *Anemonia majano*, *Zoanthus sp.*, *Discosoma striata*, *Discosoma sp.* y *Clavularia sp.* (Matz et al., 1999). Se ha informado de que una proteína fluorescente clonada a partir de la especie de coral pétreo, *Trachyphyllia geoffroyi*, emite luz verde, amarilla y roja, y convierte la emisión de luz verde a luz roja tras exposición a luz UV (Ando et al., 2002). Las proteínas fluorescentes recientemente descritas de anémonas de mar incluyen proteínas verdes y naranjas fluorescentes clonadas a partir de *Anemonia sulcata* (Wiedenmann et al., 2000), una proteína verde fluorescente potenciada de forma natural clonada a partir de los tentáculos de *Heteractis magnifica* (Tu et al., 2003), una cromoproteína púrpura en general no fluorescente que presenta fluorescencia roja débil clonada a partir de *Anemonia sulcata* y un mutante de la misma que presenta espectros de emisión de desplazamiento al rojo lejano (595 nm) (Lukyanov et al., 2000).

20 Adicionalmente, se ha descrito otra clase de proteínas relacionadas con GFP, que tienen propiedades de cromofóricas y fluorescentes. Uno de dichos grupos de proteínas derivadas de coral, las pocilloporinas, presentan una amplia gama de características espectrales y fluorescentes (Dove y Hoegh-Guldberg, 1999, solicitud PCT WO 00146233; (Dove et al., 2001)). Recientemente, se ha descrito la purificación y cristalización de la pocilloporina Rtms5 del coral *Montipora efflorescens* formador de arrecifes (Beddoe et al., 2003). Rtms5 es de color azul oscuro, aunque es débilmente fluorescente. Sin embargo, se ha informado de que Rtms5, así como otras cromoproteínas con homología de secuencia con Rtms5, se pueden interconvertir en una proteína rosada fluorescente a través de sustituciones de un único aminoácido (Beddoe et al., 2003; Bulina et al., 2002; Lukyanov et al., 2000).

25 También se conocen otras varias cromoproteínas derivadas de coral estrechamente relacionadas con las pocilloporinas (véase, por ejemplo, (Gurskaya et al., 2001; Lukyanov et al., 2000)). Ejemplos adicionales de proteínas fluorescentes son GFP de *Renilla reniformis*, MKO de *Fungia concinna*, Azami Green de *Galaxeidae* o cOFP de *Cerianthus*. Cualquiera de las proteínas fluorescentes o cromofóricas o fragmentos fluorescentes o cromofóricos de las mismas se pueden usar de acuerdo con el contenido de la presente invención. Los fragmentos de la proteína fluorescente o cromofórica son fragmentos preferentemente funcionales.

35 La acumulación de un (poli)péptido en sitios distintos en el núcleo de la célula puede estar causada por la interacción del (poli)péptido con las estructuras proteicas o no proteicas ya acumuladas en sitios distintos en el núcleo de la célula o por la acumulación en un sitio distinto que proporciona un entorno adecuado para la acumulación del (poli)péptido, preferentemente por unión directa o indirecta a dichos sitios.

40 "Colocalización de la emisión de fluorescencia en el núcleo" se refiere a la localización de dos emisiones de fluorescencia diferentes en el mismo sitio del núcleo de la célula. La colocalización se detecta en cuanto dos proteínas interactúan entre sí. La colocalización se detecta como la superposición espacial parcial o completa de la emisión de fluorescencia desde dos (poli)péptidos fluorescentes diferentes en el núcleo. La detección se puede lograr por el experimentador o por un software especializado conocido para los expertos en la técnica (por ejemplo, ImageJ co-localization plug-ins, <http://rsb.info.nih.gov/ij/>).

45 La presente invención se basa en el desarrollo de un novedoso ensayo de doble híbrido fluorescente (F2H) para la visualización directa de interacciones proteína-proteína en células eucariotas vivas. La lectura óptica sencilla de este ensayo F2H permite la observación de interacciones proteína-proteína en tiempo real y se puede emplear para exploraciones de alto rendimiento. El procedimiento de la invención se basa en la inmovilización de un (poli)péptido señuelo marcado con fluorescencia en una estructura nuclear distinta, que permite la detección de interacciones proteína-proteína como colocalización de un (poli)péptido presa marcado de forma diferente en esta estructura definida. El ensayo F2H de la invención se sometió a prueba en el ejemplo de líneas celulares con una integración estable de una matriz del operador de *lac* para inmovilizar un represor de *lac* fusionado a (poli)péptidos de interés marcados con fluorescencia ((poli)péptidos señuelo). Ya se han descrito líneas celulares que se pueden usarse fácilmente para seres humanos, ratón, hámster y *Drosophila* (Dietzel et al., 2004; Janicki et al., 2004; Robinett et al., 1996; Tsukamoto et al., 2000; Tumber et al., 1999; Vazquez et al., 2001). Para que sea independiente de las líneas celulares transgénicas específicas, este ensayo se podría modificar usando varias estructuras celulares como la lámina o los centrosomas como estructuras de anclaje para inmovilizar localmente (poli)péptidos señuelo.

55 Al igual que otros procedimientos genéticos de doble híbridos, el ensayo F2H de la invención puede proporcionar resultados positivos falsos o negativos falsos, que se tienen que controlar. Los (poli)péptidos presa que se unen a la matriz del operador de *lac* en ausencia de un (poli)péptido señuelo se pueden identificar por un cribado inicial. Con

este fin, la localización de los (poli)péptidos presa dentro del núcleo se determina por microscopía de fluorescencia en ausencia de una proteína señuelo respectiva. Posteriormente, se puede determinar una acumulación aleatoria de la proteína presa fluorescente en la matriz del operador de *lac* por una fluorescencia agrupada en esta estructura. Los (poli)péptidos identificados de esta manera se pueden usar únicamente como (poli)péptidos señuelo para evitar resultados positivos falsos. Se analizaron más de 20 interacciones proteína-proteína desde diferentes compartimentos subcelulares con el ensayo F2H de la presente invención y se obtuvieron resultados idénticos a los descritos anteriormente con otros procedimientos genéticos o bioquímicos. Las proteínas que se encuentran unidas por sí mismas a la matriz del operador de *lac* (tal como, por ejemplo, SUMO3 analizada en los ejemplos) se pueden usar únicamente como un (poli)péptido señuelo. Los resultados divulgados en los ejemplos muestran que el ensayo F2H de la presente invención es un procedimiento fiable y ampliamente aplicable para estudiar las interacciones proteína-proteína.

En algunos casos, las proteínas se pueden acumular en focos subnuclear y por tanto complicar el análisis F2H. Para evitar este problema, la matriz del operador de *lac* se podría visualizar e identificar con una tercera proteína de fusión fluorescente como CFP-LacI.

La presente invención se caracteriza, además, por que una construcción o proteína de fusión mínima que comprende un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula y un (poli)péptido que se une específicamente a GFP puede servir para el cribado ultrarrápido veloz y eficaz de las interacciones de las proteínas. Tras la coexpresión de una segunda proteína de fusión que comprende GFP y un (poli)péptido señuelo, se establece que la interacción de GFP con el (poli)péptido de unión a GFP unido al (poli)péptido acumulado en los sitios distintos en el núcleo de la célula constituye el complejo señuelo. Tras la coexpresión de una proteína de fusión adicional (tercera) que comprende un (poli)péptido fluorescente y un (poli)péptido presa, ambos (poli)péptidos fluorescentes se colocan tras la interacción de los (poli)péptidos señuelo y presa fusionados al mismo.

Hasta la fecha se han diseñado y establecido bibliotecas de proteínas fusionadas a GFP y de ácidos nucleicos que codifican estas últimas (Newman et al., 2006), lo que facilita en gran medida el cribado ultrarrápido de compañeros de interacción de una proteína.

En la solicitud de patente internacional WO 2007/068313, los autores de la presente invención divulgan un (poli)péptido que se une específicamente a GFP derivado de un dominio VHH de camello. En consecuencia, es preferente que la proteína que se une específicamente a GFP comprenda la secuencia de aminoácidos que se muestra en la SEQ ID NO: 7 o 9 o que está codificada por una secuencia de ácido nucleico de la SEQ ID NO: 8 o 10.

En resumen, este nuevo ensayo F2H permite la visualización directa de las interacciones proteína-proteína y debe perfectamente adecuarse para investigar los cambios dependientes del ciclo o la diferenciación celular, en tiempo real en células vivas. Una ventaja significativa del ensayo F2H sobre otras técnicas basadas en células es su simplicidad, que no requiere ni instrumentación costosa ni experiencia técnica avanzada. La lectura óptica simple del ensayo F2H ofrece adicionalmente la posibilidad de usar este ensayo en cribados ultrarrápidos automatizados para analizar sistemáticamente el interactoma de las proteínas en células vivas.

En comparación con el procedimiento descrito en (Miller et al., 2007) que se basa en estructuras víricas que se acumulan aleatoriamente en el citoplasma y que forman estructuras irregulares e imprevisibles, las estructuras nucleares usadas en la presente invención forman manchas definidas e identificables. Por el contrario, el procedimiento de Miller et al. tiene como objetivo y da como resultado, muchos agregados citoplásmicos mal definidos y grandes que son difíciles de distinguir de los agregados inespecíficos a menudo observados con proteínas de fusión sobreexpresadas de forma artificial o vesículas citoplásmicas. Usando estructuras inertes, tales como el operador de *lac*, es inmediatamente posible detectar la agregación no específica de un (poli)péptido fluorescente, por ejemplo, por la presencia de un número variable de agregados en las células. En el procedimiento de la presente invención, dependiendo de la ploidía de la célula, es detectable de una a dos manchas.

En un segundo aspecto, la presente invención se refiere a un procedimiento *in vitro* para identificar un compuesto que modula la interacción de dos (poli)péptidos (a) expresando en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula y (ii) un (poli)péptido que se une específicamente a GFP; (b) expresando en la misma célula una segunda proteína de fusión que comprende (i) GFP y (ii) un (poli)péptido señuelo; (c) expresando en la misma célula una tercera proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP y (ii) un (poli)péptido presa que se sabe o se sospecha que interacciona con el (poli)péptido señuelo; (d) poniendo en contacto la célula con un compuesto de prueba; y (e) detectando la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación; en el que un cambio en el grado de colocación de la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo que el compuesto puede modular la interacción del (poli)péptido señuelo y presa.

- 5 "Modulación de la interacción de dos (poli)péptidos" indica la capacidad de ciertos compuestos de influir en la interacción de dos (poli)péptidos o proteínas. La interacción se puede fortalecer o debilitarse hasta su completa anulación. La modulación en la forma de un debilitamiento puede tener lugar, por ejemplo, por competición de un compuesto de prueba con la proteína de interacción por el mismo sitio de unión o por unión a una proteína y, por lo tanto, la alteración de su estructura tridimensional, de manera que la interacción entre las dos proteínas ya no es posible debido a los cambios conformacionales en dicha proteína. Por otra parte, esto último también puede conducir a un aumento en la afinidad de unión hacia una proteína por la misma razón.
- 10 Un compuesto de prueba puede ser, pero no está restringido a, un compuesto que pertenece a las clases de, por ejemplo, ácidos nucleicos, (poli)péptidos, aptámeros peptídicos, aptámeros basados en ácidos nucleicos, moléculas pequeñas o anticuerpos o fragmentos de los mismos. El compuesto de prueba puede ser cualquier compuesto químico.
- 15 Los ácidos nucleicos pueden ser ADN, ARN o ribozimas. Los ácidos nucleicos se pueden sintetizar químicamente o se pueden producir junto con un promotor de expresión biológica *in vitro* o incluso *in vivo*.
- Los ácidos nucleicos se pueden sintetizar químicamente usando fosoramiditas de ribonucleósidos protegidas apropiadamente y un sintetizador convencional de ADN/ARN. Los proveedores de reactivos de síntesis de ARN son Proligo (Hamburgo, Alemania), Dharmacon Research (Lafayette, CO, EE. UU.), Pierce Chemical (parte de Perbio Science, Rockford, IL, EE. UU.), Glen Research (Sterling, VA, EE. UU.), ChemGenes (Ashland, MA, EE. UU.) y Cruachem (Glasgow, Reino Unido).
- 20 Una ribozima (de enzima de ácido ribonucleico, también llamada enzima de ARN o ARN catalítico) es una molécula de ARN que cataliza una reacción química. Muchas ribozimas naturales catalizan su propia escisión o la escisión de otros ARN, pero también se ha encontrado que catalizan la actividad aminotransferasa del ribosoma. Ejemplos de pequeños ARN de auto-escisión bien caracterizados son ribozimas en cabeza de martillo, de horquilla, del virus de la hepatitis delta y dependientes de plomo seleccionadas *in vitro*. La organización de estos pequeños catalizadores contrasta con la de ribozimas más grandes, tales como el intrón del grupo I.
- 25 Los aptámeros son moléculas de ácidos oligonucleicos o péptidos que se unen a una molécula diana específica. Los aptámeros se crean habitualmente seleccionándolos de una gran combinación de secuencias aleatorias, pero también existen aptámeros naturales en riboconmutadores. Los aptámeros se pueden usar tanto para investigación básica como con fines clínicos como fármacos macromoleculares. Los aptámeros se pueden combinar con ribozimas para auto-escisión en presencia de su molécula diana. Estas moléculas del compuesto tienen aplicaciones adicionales de investigación, industriales y clínicas. Más específicamente, los aptámeros se pueden clasificar como de aptámeros de ADN o ARN o aptámeros peptídicos. Mientras los primeros consisten en hebras (habitualmente cortas) de oligonucleótidos, los últimos consisten en un dominio peptídico variable corto, unido en ambos extremos a una estructura proteica.
- 30 Los aptámeros de ácido nucleico son especies de ácido nucleico que se han diseñado a través de rondas repetidas de selección *in vitro* o, de forma equivalente, SELEX (evolución sistemática de ligandos por enriquecimiento exponencial), para que se unan a varias dianas moleculares, tales como moléculas pequeñas, proteínas, ácidos nucleicos e incluso células, tejidos y organismos.
- 35 Los aptámeros peptídicos son (poli)péptidos que están diseñados para interferir con otras interacciones de las proteínas dentro de las células. Consisten en un bucle peptídico variable unido en ambos extremos a una estructura proteica. Esta doble restricción estructural aumenta enormemente la afinidad de unión del aptámero peptídico hasta niveles comparables a los de un anticuerpo (intervalo nanomolar). La longitud de bucle variable está comprendida, típicamente, de 10 a 20 aminoácidos, y la estructura puede ser cualquier proteína, que tenga buenas propiedades de solubilidad.
- 40 Actualmente, la proteína bacteriana tiorredoxina-A es la estructura proteica más frecuentemente usada, insertándose el bucle variable dentro del sitio activo de reducción, que es un bucle -Cys-Gly-Pro-Cys- en la proteína natural, pudiendo formar las dos cadenas laterales de cisteína un puente disulfuro. La selección del aptámero peptídico se puede hacer usando diferentes sistemas, pero el más frecuentemente usado es actualmente el sistema de doble híbrido de levaduras.
- 45 Un anticuerpo puede ser, por ejemplo, policlonal o monoclonal. El término "anticuerpo" también comprende derivados o fragmentos del mismo que aún retienen la especificidad de unión. Las técnicas para la producción de anticuerpos son bien conocidas en la técnica y se describen, por ejemplo, en Harlow y Lane "Antibodies, A Laboratory Manual", Cold Spring Harbor Laboratory Press, 1988 y Harlow y Lane "Using Antibodies: A Laboratory Manual" Cold Spring Harbor Laboratory Press, 1999.
- 50 Una molécula pequeña de acuerdo con la presente invención puede ser orgánica o inorgánica y tiene un peso molecular de un máximo de 2000 Dalton, preferentemente no más de 1000 Dalton, más preferentemente no más de 800 Dalton.
- 55 En un modo de realización preferente de este aspecto de la presente invención, el compuesto de prueba puede

debilitar la interacción de dos (poli)péptidos, en la que una disminución de la colocalización de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativa de que el compuesto puede debilitar la interacción del (poli)péptido señuelo y presa.

5 "Capaz de debilitar la interacción de dos (poli)péptidos", en el contexto de la presente invención, significa, en primer lugar, la influencia de un compuesto sobre la afinidad de unión de un (poli)péptido hacia otro (poli)péptido, que conduce a un debilitamiento o a la interrupción completa/anulación de la interacción o afinidad de unión. Como se describe anteriormente, esto puede tener lugar por cambios conformacionales inducidos en un (poli)péptido tras la unión del compuesto. En segundo lugar, el compuesto puede competir directamente con el (poli)péptido de interacción por el mismo sitio de unión o uno diferente adyacente a dicho sitio de unión en el otro (poli)péptido. Un debilitamiento o anulación completa de la interacción es detectable como una disminución o ausencia completa de colocalización de ambas señales fluorescentes.

Una célula de referencia, en el contexto de la presente invención, indica una célula que no se ha puesto en contacto con el compuesto de prueba. En consecuencia, la influencia del compuesto sobre el comportamiento de interacción del (poli)péptido señuelo y presa se puede comparar directamente.

En otro modo de realización preferente de este aspecto de la invención, el compuesto de prueba puede inducir o potenciar la interacción de dos (poli)péptidos, en la que un aumento en la colocalización de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede inducir o potenciar la interacción del (poli)péptido señuelo y presa.

"Capaz de inducir la interacción de dos (poli)péptidos" indica la influencia de un compuesto sobre la afinidad de unión de un (poli)péptido hacia otro (poli)péptido, que conduce a una afinidad de unión aumentada. Por ejemplo, si no se detecta o se detecta colocalización únicamente menor de las señales fluorescentes (poli)péptido señuelo y presa en la célula de referencia, un aumento en la colocalización indica una interacción reforzada.

25 En un aspecto diferente, la presente invención se refiere a un procedimiento *in vitro* para determinar la fuerza relativa de la interacción de dos proteínas con una tercera proteína, que comprende: (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (ii) un (poli)péptido que se une específicamente a GFP; (b) expresar en la misma célula una segunda proteína de fusión que comprende (i) GFP; y (ii) un (poli)péptido señuelo; (c) expresar en la misma célula una tercera proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y (ii) un primer (poli)péptido presa; (d) expresar en la misma célula una cuarta proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP y de la del (poli)péptido fluorescente de dicha tercera proteína de fusión; y (ii) un segundo (poli)péptido presa; y (e) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación, en el que una magnitud de colocalización de la emisión de fluorescencia de la segunda y la tercera proteína de fusión diferente en comparación con que de la segunda y la cuarta proteína de fusión en el núcleo de la célula es indicativa de una fuerza de unión diferente del primero y el segunda (poli)péptido presa al (poli)péptido señuelo.

Este y los siguientes dos aspectos de la invención se basan en dos construcciones básicas, una construcción presa y señuelo, y se puede aplicar en los casos en los que no están disponibles bibliotecas de ácidos nucleicos que codifican proteínas que comprenden GFP.

En un aspecto adicional, la presente invención se refiere a un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende: (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido fluorescente; (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (iii) un (poli)péptido señuelo; (b) expresar en la misma célula una segunda proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y (ii) un (poli)péptido presa y (c) detectar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula tras la excitación, en el que una colocalización de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del (poli)péptido señuelo y presa.

En un aspecto diferente, la presente invención se refiere a un procedimiento *in vitro* para identificar un compuesto que modula la interacción de dos (poli)péptidos (a) expresando en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido fluorescente; (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (iii) un (poli)péptido señuelo; (b) expresando en la misma célula una segunda proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y (ii) un (poli)péptido presa que se sabe o se sospecha que interacciona con el (poli)péptido señuelo; (c) poniendo en contacto la célula con un compuesto de prueba; y (d) detectando la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula tras la excitación; en el que un cambio en el



grado de colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede modular la interacción del (poli)péptido señuelo y presa.

- 5 En un aspecto adicional, la presente invención se refiere a un procedimiento para determinar la fuerza relativa de la interacción de dos proteínas (usado de manera intercambiable con el término "(poli)péptidos") con una tercera proteína (o (poli)péptido), que comprende (a) expresar en una célula eucariota una primera proteína de fusión que comprende (i) un (poli)péptido fluorescente; (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (iii) un (poli)péptido señuelo; (b) expresar en la misma célula una  
10 segunda proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y (ii) un primer (poli)péptido presa; (c) expresar en la misma célula una tercera proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera y segunda proteína de fusión; y (ii) un segundo (poli)péptido presa; y (d) detectar la emisión de  
15 fluorescencia de las partes fluorescentes de la primera, la segunda y la tercera proteína de fusión en la célula tras la excitación, en el que una magnitud de colocalización de la emisión de fluorescencia de la primera y la segunda proteína de fusión diferente en comparación con la de la primera y la tercera proteína de fusión en el núcleo de la célula es indicativa de una fuerza de unión diferente de la primera y la segunda proteína presa al (poli)péptido señuelo.
- 20 En un modo de realización preferente de los aspectos de la presente invención que se refiere a procedimientos para detectar interacciones proteína-proteína, la detección se emplea para investigar la dependencia de las interacciones proteína-proteína de los procesos celulares, en el que el procedimiento comprende además (d1) supervisar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula en el curso de uno o más procesos en la célula; en el que un cambio en el grado de colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en la célula es indicativo de una  
25 dependencia de la interacción de uno o más procesos celulares.

"Dependencia de los procesos celulares" indica la posibilidad de que el (poli)péptido señuelo o presa o ambos alteren su comportamiento de interacción debido a modificaciones en el curso de los procesos celulares, tales como el ciclo celular. Las proteínas que desempeñan un papel en el ciclo celular experimentan varias modificaciones tales como fosforilación y desfosforilación, que podrían alterar su afinidad de unión hacia otras proteínas. Dependiendo del punto temporal o el período durante un proceso celular cuando se logra esta modificación, un (poli)péptido podría perder o ganar afinidad de unión a uno o más (poli)péptidos diferentes.

En un modo de realización más preferente, el proceso celular es el ciclo celular, la secreción, la traslocación o la transducción de señales.

- 35 En otro modo de realización preferente, la detección se emplea para determinar la fuerza de una interacción proteína-proteína, en el que el procedimiento comprende además (d2) en caso de que se detecte una colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión, la extinción selectiva de la fluorescencia de dicha segunda proteína de fusión y el control del restablecimiento de la colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en el tiempo, en el que el tiempo necesario para establecer colocalización es indicativo de la fuerza de la interacción proteína-proteína.

- Este modo de realización de la presente invención, en parte, utiliza el procedimiento FRAP (recuperación de fluorescencia después de fotoblanqueado) bien establecido. Este procedimiento indica una técnica óptica que puede cuantificar la difusión y la movilidad de sondas marcadas con fluorescencia. Esta técnica proporciona una gran  
45 utilidad en estudios biológicos de unión de proteínas y se usa comúnmente junto con proteínas fluorescentes (FP), en la que la proteína estudiada está fusionada con una FP. Cuando se excita por una longitud de onda específica de la luz (típicamente, un haz de láser), la proteína emitirá fluorescencia. Cuando la proteína que se está estudiando se produce con la FP, entonces la fluorescencia se puede rastrear. Después de la fotodestrucción de la FP (típicamente con un pulso de láser fuerte/intenso), la cinética de la recuperación de fluorescencia en la zona blanqueada proporciona información acerca de la fuerza de las interacciones de las proteínas, la continuidad del orgánulo y el tráfico de proteínas que evitar o retarda el intercambio de FP blanqueadas y sin blanquear. Esta observación se ha explotado más recientemente para investigar la unión de proteínas.

En otro modo de realización preferente de la presente invención, los componentes (i), (ii) y/o (iii) de dicha primera proteína de fusión y/o los componentes (i) y (ii) de dicha segunda proteína de fusión y/o los componentes (i) y (ii) de dicha tercera proteína de fusión están conectados a través de un engarce.

- 55 El término "engarce" se refiere a la conexión entre los componentes de las proteínas de fusión de la invención. Un engarce puede ser un enlace peptídico o un tramo de aminoácidos que comprende al menos un resto de aminoácido que puede estar dispuesto entre los componentes de las proteínas de fusión en cualquier orden. Dicho engarce puede ser útil en algunos casos, por ejemplo, para mejorar el plegamiento por separado de los dominios individuales o para modular la estabilidad de la proteína de fusión. Además, dichos restos de engarce pueden contener señales

para el transporte, secuencias de reconocimiento de proteasa o señales para modificación secundaria. Los restos de aminoácido que forman el engarce pueden estar estructurados o no estructurados. Preferentemente, el engarce puede ser tan corto como de 1 residuo de aminoácido o de un máximo 2, 3, 4, 5, 10, 20 o 50 restos. En casos particulares, el engarce puede incluso implicar un máximo de 100 o 150 restos.

5 En otro modo de realización preferente de la presente invención, el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona directamente con estructuras proteicas o no proteicas acumuladas en sitios distintos en el núcleo de la célula.

En un modo de realización preferente diferentes de la presente invención, el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona indirectamente con estructuras proteicas o no proteicas acumuladas en sitios distintos en el núcleo de la célula.

10 La interacción indirecta de nuevo se puede producir a través de moléculas proteicas o no proteicas tales como (poli)péptidos o ácidos nucleicos.

En un modo de realización preferente diferente, cualquiera de los (poli)péptidos presa comprende una señal de localización nuclear.

15 Una señal de localización nuclear dirige un producto de expresión al núcleo de la célula. Un ejemplo de una NLS es la secuencia peptídica PKKKRKV (señal de localización nuclear (NLS) del antígeno T grande de SV40 (Kalderon et al., 1984)), que puede dirigir proteínas heterólogas al núcleo. Más NLS son, por ejemplo, KR [PAATKKAGQA]KKKK, la NLS de la nucleoplasmina como un prototipo de una señal bipartita ubicua o KIPK, la NLS del represor Mata2 de la transcripción de levaduras. Muchas otras secuencias de señales de localización nuclear son conocidos para los expertos en la técnica y se describen en la literatura.

20 Una NLS tiene que estar presente si el (poli)péptido presa fusionado al (poli)péptido fluorescente no se puede trasladar por sí mismo al núcleo. En general, (poli)péptidos de un tamaño de un máximo de 60 kDa se pueden trasladar al núcleo. Además, la composición de aminoácidos del (poli)péptido desempeña un papel, es decir, varios aminoácidos básicos consecutivos en un (poli)péptido pueden promover la traslocación del (poli)péptido al núcleo. Para garantizar que el (poli)péptido presa se trasloca y se acumula en el núcleo, es preferente que esté presente una NLS. Lo mismo es cierto para (poli)péptidos que se acumulan en sitios distintos del núcleo, pero que únicamente se trasladan al núcleo si está presente una NLS. Un ejemplo de dicho (poli)péptido es el represor de lac que por sí mismo no comprende una NLS.

25 En un modo de realización preferente diferente, la expresión en la célula eucariota se logra transfectando las moléculas de ácido nucleico que codifican dicha primera y segunda proteína de fusión en uno o más vectores.

La transfección es la introducción de moléculas de ácido nucleico en células eucariotas. Los procedimientos comúnmente usados comprenden, pero no se restringen a, transfección usando cloruro de calcio, JetPEI™ (PolyPlus) o lipofectin™ (Invitrogen).

30 El vector que comprende la molécula de ácido nucleico que codifica dicha primera y/o segunda y/o tercera y/o cuarta proteína de fusión es un vector de expresión eucariota, preferentemente un vector de expresión de mamífero. La incorporación de la molécula de ácido nucleico en un vector ofrece la posibilidad de introducir la molécula de ácido nucleico de manera eficaz en las células y, preferentemente, el ADN de una célula huésped. La célula huésped puede ser una única célula tal como una célula de una línea celular. Dicha medida hace posible expresar, si se eligen los vectores de expresión, la molécula de ácido nucleico respectiva en la célula huésped. Por lo tanto, la incorporación de la molécula de ácido nucleico en un vector de expresión abre el camino a un nivel elevado de forma permanente del (poli)péptido o proteína codificada en cualquier célula o un subconjunto de células seleccionadas de la célula huésped.

35 En otro modo de realización preferente, el (poli)péptido acumulado en sitios distintos del núcleo de la célula se ha introducido en la célula. En un modo de realización preferente diferente, en caso de que dicho (poli)péptido interaccione con estructuras proteicas o no proteicas acumuladas en sitios distintos en el núcleo, dichas estructuras proteicas o no proteicas se pueden introducir en la célula.

40 De manera similar al vector o vectores que comprenden las moléculas de ácido nucleico que codifican las proteínas de fusión de la presente invención, también las moléculas de ácido nucleico que codifican las estructuras proteicas o no proteicas se pueden introducir en un vector. La molécula de ácido nucleico se integra preferentemente de manera estable en los cromosomas de la célula, lo que conduce a la generación de una línea celular estable.

45 En un modo de realización preferente diferente, los sitios distintos de la célula forman estructuras inertes.

En un modo de realización más preferente, la estructura inerte es la lámina nuclear o manchas nucleares.

En un modo de realización preferente diferente, el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona con cuerpos de PML (también denominados cuerpos

nucleares de PML o PML NB).

Los cuerpos de PML se han asociado con muchas funciones nucleares, incluyendo la transcripción, la reparación del ADN, la defensa vírica, el estrés, la regulación del ciclo celular, la proteólisis y la apoptosis. La célula de mamífero promedio contiene 10-30 cuerpos nucleares de PML. Los cuerpos de PML se definen por la presencia de la proteína PML, identificado por primera vez por su fusión con el receptor alfa de ácido retinoico en la traslocación cromosómica t(15,17) (Borden, 2002; Mutilar et al., 2000; Moller et al., 2003) asociada con la leucemia promielocítica aguda (APL). La proteína PML es esencial para la formación de PML NB y, cuando está ausente, o sus dedos RING (que comprenden dedos de C<sub>3</sub>HC<sub>4</sub> zinc como motivo estructural) están mutados, los cuerpos de PML se interrumpen.

5 En un modo de realización preferente diferente, el (poli)péptido que se acumula en sitios distintos del núcleo o las estructuras proteicas o no proteicas acumuladas en sitios distintos de la célula es heterólogo.

"Heterólogo", como se usa en la presente invención, indica el origen de una estructura proteica o no proteica que es diferente de la de la célula en la que se expresa, es decir, una especie diferente.

15 En otro modo de realización preferente, el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona con el ADN.

En un modo de realización más preferente, el ADN es el operador de *lac*.

20 En un modo de realización incluso más preferente, el operador de *lac* está presente en el núcleo en varias copias. El número de copias depende del experimento individual, es decir, del tipo de célula huésped usada o el (poli)péptido señuelo y/o el (poli)péptido presa. El número de copias comúnmente aplicado alcanza de al menos 150 copias de un plásmido que comprende 256 copias del operador de *lac* (es decir, 38400 copias del operador de *lac*) hasta aproximadamente 2000 copias de dicho plásmido (es decir, 512000 copias del operador de *lac*). En caso de que se necesite una mayor sensibilidad, el número de copias puede ser incluso mayor y llegar a un máximo de 3000 copias. De forma alternativa, se puede utilizar cualquier clase de plásmido que comprenda un número de copias diferente para obtener el número de copias deseado del operador de *lac*. El operador de *lac* puede estar dispuesto en tándem con o sin extensiones de ácido nucleico separando cada elemento. Los elementos se pueden disponer de forma antiparalela o paralela. Las líneas celulares que tienen varias copias del operador de *lac* integradas de forma estable en el núcleo son conocidas en la técnica (Janicki et al., 2004; Tsukamoto et al., 2000).

25

En otro modo de realización preferente, dicha primera proteína de fusión comprende Lacl.

30 El represor de *lac* Lacl es un tetrámero de subunidades idénticas. Cada subunidad contiene un motivo de hélice-giro-hélice (HTH) que se puede unir al ADN. El sitio del operador en el que se une el represor es una secuencia de ADN con simetría de repeticiones invertidas. Los dos semisitios de ADN del operador se unen a dos de las subunidades del represor tetramérico.

35 Tanto el operador de *lac* como Lacl forman parte del operón *lac* bacteriano, una unidad de regulación del metabolismo de lactosa de bacterias. Si falta lactosa en el medio de cultivo, el represor se une muy fuertemente al operador de *lac* localizado en dirección 3 del promotor.

La presente invención hace uso de la interacción entre Lacl y el operador de *lac* para formar una estructura de anclaje no proteica en el núcleo con la que el procedimiento de la invención se puede llevar a cabo convenientemente. En general, la interacción de Lacl con el operador de *lac* se detecta en un sitio distinto en el núcleo.

40 Una ventaja importante de este modo de realización de la presente invención, en comparación con los procedimientos previamente conocidos, en particular el de (Miller et al., 2007), es que el número de copias del operador de *lac* que forma una matriz del operador de *lac* se puede variar para adaptar el sistema al nivel de expresión los (poli)péptidos tanto señuelo como presa. Por ejemplo, si la expresión de uno o ambos (poli)péptidos se encuentra demasiado alta en la célula, por lo tanto, por ejemplo, interfiriendo con los procesos celulares o interrumpiendo la señal fluorescente, se puede tomar otra célula que tenga un número de copias inferior del operador de *lac*, posibilitando, por lo tanto, una valoración de la detección de la interacción.

45

En un modo de realización más preferente, dicha segunda proteína de fusión comprende GFP.

En otro modo de realización preferente, la detección se lleva a cabo usando un microscopio de fluorescencia.

50 Un microscopio de fluorescencia es un microscopio óptico usado para estudiar las propiedades de sustancias orgánicas o inorgánicas que usan los fenómenos de fluorescencia y fosforescencia en lugar de, o además de, la reflexión y la absorción. La muestra se ilumina con luz de una longitud de onda (o longitudes de onda) específica que se absorbe por los fluoróforos, provocando que emitan longitudes de onda más largas de luz (de un color diferente al de la luz absorbida). La luz de iluminación se separa de la fluorescencia emitida mucho más débil a través del uso de un filtro de emisión. Los componentes típicos de un microscopio de fluorescencia son la fuente de

luz (lámpara de descarga de arco de xenón o de mercurio), el filtro de excitación, el espejo dicróico (o divisor de haz dicromático) y el filtro de emisión. Los filtros y el espejo dicróico se eligen para acoplar las características espectrales de excitación y emisión del fluoróforo usado para marcar la muestra. La mayoría de los microscopios de fluorescencia en uso son los microscopios de epifluorescencia (es decir: la excitación y la observación de la fluorescencia son desde arriba (epi) de la muestra). Estos microscopios se han convertido en una parte importante en el campo de la biología, abriendo las puertas para diseños de microscopios más avanzados, tales como el microscopio de barrido de láser confocal (CLSM) y el microscopio de fluorescencia de reflexión interna total (TIRF). Dichas tecnologías son bien conocidas para los expertos en la técnica.

En otro modo de realización preferente, la célula eucariota es una célula viva.

- 10 En otro modo de realización preferente, el procedimiento de la invención comprende, además, la localización de los (poli)péptidos señuelo y/o presa en la célula.

Este modo de realización sirve para confirmar que la colocalización tiene lugar, efectivamente, en el núcleo y para descartar la interacción no específica de los (poli)péptidos señuelo y presa. La localización se detecta por el experimentador o bien por un programa informático especializado que puede distinguir diferentes compartimentos celulares y es bien conocido para el experto en la técnica (por ejemplo ImageJ (versión 1.38, <http://rsb.info.nih.gov/ij/>)).

También se describe en el presente documento una molécula de ácido nucleico que codifica una proteína de fusión que comprende (i) un (poli)péptido fluorescente; (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula y (iii) un (poli)péptido señuelo.

- 20 También se describe en el presente documento una molécula de ácido nucleico que codifica una proteína de fusión que comprende (i) un (poli)péptido que se une al operador de lac y (ii) un (poli)péptido que se une específicamente a GFP, un vector que comprende dicho ácido nucleico y una proteína codificada por dicho ácido nucleico.

Una "molécula de ácido nucleico" incluye ADN, tal como ADNc o ADN genómico, y ARN, tanto la hebra con sentido como la hebra antisentido. Se incluyen, además, moléculas que imitan ácidos nucleicos conocidas en la técnica, tales como derivados sintéticos o semisintéticos de ADN o ARN y polímeros mixtos. Dichas moléculas que imitan ácidos nucleicos o derivados de ácidos nucleicos incluyen ácido nucleico de fosforotioato, ácido nucleico de fosforamidoato, ácido 2'-O-metoxietil ribonucleico, ácido morfolino nucleico, ácido hexitol nucleico (HNA) y ácido nucleico bloqueado (LNA) (véase (Braasch y Corey, 2001)). El LNA es un derivado de ARN en el que el anillo de ribosa está limitado por un enlace metileno entre el oxígeno 2' y el carbono 4'. Pueden contener bases nucleotídicas no naturales o derivadas adicionales, como se apreciará fácilmente por los expertos en la técnica.

Se describe en el presente documento, además, que la proteína de fusión comprende LacI.

También se describe en el presente documento una proteína codificada por la molécula de ácido nucleico de la invención. Las características de una proteína o (poli)péptido se definen en otra parte en la presente solicitud.

Se describe, además, que la proteína comprende LacI.

- 35 En un aspecto alternativo, la presente invención se refiere a un vector que comprende la molécula de ácido nucleico caracterizada en las reivindicaciones adjuntas. Las propiedades de un vector se han descrito en otra parte en la presente solicitud. Además de las características ya descritas, el vector de la presente invención puede ser un vector procariota. Los vectores procariotas y sus propiedades son bien conocidos en la técnica. Preferentemente, el vector es un plásmido, cósmido, virus, bacteriófago u otro vector usado convencionalmente, por ejemplo, en ingeniería genética.

La molécula de ácido nucleico se puede insertar en varios vectores disponibles comercialmente. Los ejemplos no limitantes incluyen vectores compatibles con una expresión en células de mamífero como pREP (Invitrogen), pcDNA3 (Invitrogen), pCEP4 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2neo, pBPV-1, pdBPVMMTneo, pRSVgpt, pRSVneo, pSV2-dhfr, pZD35, pLXIN, pSIR (Clontech), pIRES-EGFP (Clontech), pEAK-10 (Edge Biosystems), pTriEx-Hygro (Novagen), pCINeo (Promega), el vector de expresión de ADNc pcDV1 de Okayama-Berg (Pharmacia), pRc/CMV, pcDNA1, pSPORT1 (GIBCO BRL), pGEMHE (Promega) o pSVL y pMSG (Pharmacia, Uppsala, Suecia).

Para las técnicas de modificación de vectores, véase Sambrook y Russel (2001), loc. cit. En general, los vectores pueden contener uno o más orígenes de replicación (ori) y sistemas de herencia para la clonación o expresión, uno o más marcadores para la selección en el huésped, por ejemplo, resistencia a antibióticos, y uno o más casetes de expresión.

Las moléculas de ácido nucleico insertadas en el vector se pueden sintetizar, por ejemplo, por procedimientos estándar, o se pueden aislar de fuentes naturales. El ligamiento de las secuencias codificantes a elementos reguladores de la transcripción y/o a otras secuencias codificantes de aminoácidos se puede llevar a cabo usando procedimientos establecidos.

De acuerdo con la presente invención, el vector comprende, además, una molécula de ácido nucleico que codifica una proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en la proteína codificada por la molécula de ácido nucleico descrita anteriormente y (ii) un (poli)péptido presa.

- 5 En un aspecto diferente, la presente invención se refiere a una célula eucariota transfectada con el vector de la invención.

Es preferente que la célula eucariota sea una célula de mamífero.

- 10 En otro aspecto de la invención, una célula eucariota como se caracteriza en las reivindicaciones adjuntas se transfecta con un vector que comprende la molécula de ácido nucleico descrita anteriormente y con un vector que comprende una molécula de ácido nucleico que codifica una proteína de fusión que comprende (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en la proteína descrita anteriormente y (ii) un (poli)péptido presa.

Además, se describe en el presente documento una célula eucariota que tiene varias copias de operador de *lac* integradas de manera estable en su ADN y que expresan de forma estable la proteína de la invención.

- 15 En otro aspecto, la invención se refiere a una célula eucariota que tiene varias copias del operador de *lac* integradas de manera estable en su ADN y que expresan de forma estable una proteína que comprende (i) un (poli)péptido que se une al operador de *lac* y (ii) un (poli)péptido que se une específicamente a GFP.

- 20 En un modo de realización preferente, dicho (poli)péptido que se une específicamente a GFP comprende un (poli)péptido que tiene la secuencia de aminoácidos de las SEQ ID NO: 7 o 9, o está codificada por una secuencia de ácido nucleico de la SEQ ID NO: 8 o 10.

- 25 También se describe en el presente documento una molécula de ácido nucleico que codifica una proteína de fusión que comprende (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y (ii) un (poli)péptido que se une específicamente a GFP; así como la proteína codificada por dicha molécula de ácido nucleico, un vector que comprende dicha molécula de ácido nucleico y una célula eucariota transfectada con el vector que comprende dicha molécula de ácido nucleico. Los modos de realización preferentes y las definiciones que se describen para los aspectos de la presente invención referidos a un sistema F2H que comprende una construcción señuelo y al menos una construcción presa son igualmente aplicables a los aspectos anteriores que utilizan el (poli)péptido específico para GFP. Todos los modos de realización referidos a la primera, la segunda o la tercera proteína de fusión de los tres primeros aspectos de la invención se aplican igualmente a las cuatro proteínas de fusión de los aspectos que utilizan el (poli)péptido específico para GFP.
- 30

Las figuras muestran:

### Figura 1

- 35 Boceto esquemático del ensayo de doble híbrido fluorescente (F2H). (a) Esquema de los vectores de expresión pF2H-presa y pF2H-señuelo que codifican proteínas presa y señuelo marcados con fluorescencia usados para el ensayo F2H. (b) El dominio *Lacl* de la proteína señuelo media la unión a la matriz del operador de *lac* integrado cromosómicamente, que es visible como una mancha fluorescente en los núcleos de células transfectadas. Si la presa marcada de forma diferencial interacciona con el señuelo, se enriquece en la misma mancha, lo que da como resultado la colocalización de las señales fluorescentes en el operador de *lac* (visibles como manchas amarillas en la imagen superpuesta). (c) Si la presa no interacciona con la proteína señuelo, permanece disperso en el núcleo y la matriz del operador de *lac* se visualiza únicamente por la proteína señuelo (mancha roja). FP1 y FP2 se refieren a dos proteínas fluorescentes distinguibles, por ejemplo, GFP o YFP y mCherry o mRFP.
- 40

### Figura 2

- 45 Interacción específica de la ADN ligasa III con XRCC1 revelada por F2H. (a) Se cotransfectaron células BHK transgénicas que contenían una matriz del operador de *lac* cromosómico con construcciones XRCC1-*Lacl*-RFP y de ADN ligasa III o ADN ligasa I marcadas con GFP. La parte del represor de *lac* de la proteína de fusión XRCC1-*Lacl*-RFP media la unión a la matriz del operador de *lac* (visible por microscopía de fluorescencia como una mancha roja). La ADN ligasa III se recluta a la matriz del operador de *lac* a través de la interacción con XRCC1. Obsérvese que la ADN ligasa I altamente homóloga no se acumula en la matriz del operador de *lac*, lo que indica que no interacciona con XRCC1. Barras de escala 5  $\mu$ m. (b) Comparación de los resultados de F2H y de experimentos de coimmunoprecipitación (Co-IP). Se realizaron Co-IP con células HEK 293T que coexpresaban RFP-XRCC1 y GFP-ligasa III o GFP-ligasa I, respectivamente. Para la identificación de interacciones, se usaron el dominio BRCT marcado con GFP de la ADN ligasa III y una construcción de delección que carece del dominio BRCT. Las inmunoprecipitaciones se realizaron con un GFP-nanotrap (Rothbauer et al., 2007) (como se muestra anteriormente (Mortusewicz et al., 2006)). Las proteínas de fusión precipitadas se detectaron entonces con anticuerpos específicos contra RFP y GFP en las bandas Western. Se coprecipitó RFP-XRCC1 con GFP-ligasa III, pero no con GFP-ligasa I. También se coprecipitó RFP-XRCC1 con GFP-BRCT de ligasa III pero no con GFP- $\Delta$ BRCT de N-ligasa III. Para la
- 50
- 55

comparación de los resultados de F2H se alienaron las bandas de entrada (izquierda) y unidas (derecha) de Co-PI con las señales correspondientes del ensayo F2H. La mancha de LacI de la construcción señuelo XRCC1-LacI-RFP mostrada en rojo y la fracción unida se alinearon con la señal respectiva de las construcciones de presas marcadas con GFP. Las imágenes de células enteras de los respectivos experimentos F2H se muestran en (a) y la figura 6.

### 5 Figura 3

Análisis F2H de la interacción independiente del ciclo celular de Dnmt1 con PCNA. (a) Boceto esquemático de Dnmt1 de ratón de longitud completa y proteínas de fusión. PBD, dominio de unión a PCNA; NLS, secuencia de localización nuclear; TS, secuencia de dirección; ZnF, región de unión a Zn<sup>2+</sup>; BAH 1 y 2, dos dominios de homología adyacentes a bromo. (b) Esquema de las posibilidades de unión de las proteínas de fusión a la matriz del operador de *lac* (*lac op*) y en la horquilla de replicación. (c) Células BHK transgénicas que contienen una matriz del operador de *lac* cromosómico se cotransfectaron con las construcciones PBD-LacI-YFP y RFP-PCNA. RFP-PCNA muestra la distribución dependiente del ciclo celular característica (disperso en células que no están en fase S (fila superior) y patrones focales en fase S (fila inferior)). La parte del represor de *lac* de la proteína de fusión PBD-LacI-YFP media la unión a la matriz del operador de *lac* (visible como una mancha verde y destacado por puntas de flecha) y el PBD media la unión a PCNA en los sitios de replicación (patrón focal en fase S). Obsérvese que RFP-PCNA se localiza en la matriz del operador de *lac* en células en fase S y que no están en fase S, lo que indica una interacción del PBD de Dnmt1 con PCNA en todo el ciclo celular e independiente de la maquinaria de replicación. (d) Células BHK se transfectaron con vectores de expresión para  $\Delta$ PBD-LacI-YFP y RFP-PCNA. Como anteriormente, RFP-PCNA muestra una distribución dispersa fuera de la fase S (fila superior) y se redistribuye en los sitios de replicación en la fase S (fila inferior). La proteína de fusión  $\Delta$ PBD-LacI-YFP se une a la matriz del operador de *lac* (mancha verde marcada por la punta de flecha), pero no se une a sitios de replicación en fase S, ya que carece del PBD. De forma importante, en estas células RFP-PCNA (presa) no se localiza en la matriz del operador de *lac* (marcado por puntas de flecha), lo que indica que la unión depende de la presencia del PBD, que está ausente en  $\Delta$ PBD-LacI-YFP (señuelo). Barras de escala 5  $\mu$ m.

### 25 Figura 4

Análisis de las interacciones relacionadas con la enfermedad de Huntington por F2H. Interacciones indicadas entre (a) SUMO3 y HZFH y (b) HZFH y vimentina reveladas por F2H. (c) El análisis F2H muestra ausencia de interacción entre SUMO3 y vimentina como se describe previamente (Goehler et al., 2004). En (b) el núcleo se perfila mediante una línea discontinua y en (c) la matriz del operador de *lac* está indicada (puntas de flecha). Barras de escala 5  $\mu$ m.

### 30 Figura 5

Análisis de las interacciones proteína-proteína mitocondriales y el efecto de una mutación asociada con el síndrome de Mohr-Tranebjaerg. (a) Vista general esquemática del complejo hexamérico DDP1-TIMM13 en el espacio intermembrana (IMS) de las mitocondrias. (b + c) Células BHK que expresan la proteína señuelo mCherry-LacI-TIMM13 junto con GFP-DDP1 (b) o el mutante de pérdida de función GFP-DDP1<sup>C66W</sup> (c). Mientras que la fusión natural funcional GFP-DDP1 muestra interacción con TIMM13 revelada por colocalización de las señales fluorescentes en la matriz del operador de *lac* (b), el mutante GFP-DDP1<sup>C66W</sup> se distribuye de forma dispersa por todo el núcleo, lo que indica que no hay interacción (c). Barras de escala 5  $\mu$ m.

### Figura 6

Interacción mediada por BRCT de ADN-ligasa III con XRCC1 revelada por el ensayo F2H. Células BHK transgénicas que contenían una matriz del operador de *lac* se cotransfectaron con XRCC1-LacI-RFP y varias construcciones de ADN-ligasa III marcadas con GFP. La parte del represor de *lac* de la proteína de fusión XRCC1-LacI-RFP media la unión a la matriz del operador de *lac* (visible como una mancha roja). El dominio BRCT es necesario y suficiente para dirigir la ADN ligasa III a la matriz del operador de *lac* a través de la interacción con XRCC1. Obsérvese que la ADN ligasa I altamente homóloga no se acumula en la matriz del operador de *lac*, lo que indica que no interacciona con XRCC1. Barras de escala 5  $\mu$ m.

### Figura 7

El ensayo F2H revela la interacción de XRCC1 con PCNA, PARP-1 y PARP-2. Células BHK que contenían una matriz del operador de *lac* se transfectaron con vectores de expresión para XRCC1-LacI-RFP y GFP-PARP-1, GFP-PARP-2 o GFP-PCNA. La parte del represor de *lac* de la proteína de fusión XRCC1-LacI-RFP media la unión a la matriz del operador de *lac* (visible como una mancha roja). GFP-PARP-1, GFP-PARP-2 y GFP-PCNA están dirigidos a la matriz del operador de *lac*, lo que indica una interacción con XRCC1. Barra de escala 5  $\mu$ m.

### Figura 8

El ensayo F2H revela la interacción mediada por PBD de la ADN-ligasa I con PCNA. Células U2OS transgénicas que contenían una matriz del operador de *lac* se cotransfectaron con NLS-PCNA-LacI-RFP y varias construcciones de ADN-ligasa I marcadas con GFP. La parte del represor de *lac* de la proteína de fusión NLS-PCNA-LacI-RFP media la unión a la matriz del operador de *lac* (visible como una mancha roja). El PBD es necesario y suficiente para dirigir la

ADN ligasa I a la matriz del operador de *lac* a través de la interacción con PCNA. Barra de escala 5  $\mu$ m.

### Figura 9

5 Interacción de diferentes proteínas de replicación y reparación con PCNA revelada por el ensayo F2H. Células BHK transgénicas que contenían una matriz del operador de *lac* que expresaban NLS-PCNA-Lacl-RFP y varias proteínas de replicación y reparación marcadas por GFP. Todas las proteínas sometidas a prueba interaccionan con PCNA. Barra de escala 5  $\mu$ m.

Los ejemplos ilustran la invención.

### Ejemplo 1: Materiales y procedimientos

#### Construcciones de expresión

10 La secuencia codificante de Lacl se amplificó por PCR a partir del vector de expresión p3'SS EGFP-Lacl (Robinett et al., 1996) usando los siguientes cebadores: cebador directo 5'-TCT AGA AAG CTT TCC ATG GTG AAA CCA GTA-3' y cebador inverso 5'-CCA TGC CCG GGA CAG TCG TCG TCG GGA AAC-3' (sitios de restricción en cursiva). Este fragmento de PCR se digirió con HindIII y XmaI y se clonó en los mismos sitios de dos vectores de expresión Dnmt1-YFP (MTNY.2 y PBHD-YFP) (Easwaran et al., 2004) generando PBD-Lacl-YFP y  $\Delta$ PBD-Lacl-YFP. Las construcciones NLS-PCNA-Lacl-RFP y XRCC1-Lacl-RFP se generaron mediante amplificación por PCR del ADNc de PCNA y XRCC1 usando los siguientes cebadores (sitios de restricción en cursiva):

PCNA directo	5'- CCCCCTCGAGATGTTTCGAGGCGCGC -3'
PCNA inverso	5'- GGGGAAGCTTGGAGATCCTTCTTCATCCTC- 3'
XRCC1 directo	5'- CCCAGATCTATGCCGAGATCCGC -3'
XRCC1 inverso	5'- GGGGAATTTCGGGGCTTGCGGCACCAC -3'

Posteriormente, los fragmentos de PCR se clonaron en un vector de expresión Lacl-RFP usando los sitios XhoI/HindIII para NLS-PCNA-Lacl-RFP y los sitios BglII/EcoRI para el vector de expresión XRCC1-Lacl-RFP.

20 Todas las otras construcciones F2H se generaron por amplificación por PCR de ADNc codificantes y posterior ligamiento en los sitios AsiSI y NotI de los vectores de expresión de señuelo y presa descritos en la figura 1a. Se usaron los siguientes cebadores con el sitio de restricción indicado en cursiva:

DDP1 directo	5'- GGGGAATTTCGGGGCTTGCGGCACCAC -3'
DDP1 inverso	5'-CCCCGCGGCCGCTCAGTCAGAAAGGCTTTCTG-3'
TIMM13 directo	5'-CCCCGCGATCGCGAGGGCGGCTTCGGCTCC-3'
TIMM13 inverso	5'-CCCCGCGATCGCGAGGGCGGCTTCGGCTCC-3'
HZFH directo	5'-GGGGGCGATCGCCACGCCGCTTCC-3'
HZFH inverso	5'-CCCCGCGGCCGCTTAGTCGTCTATACAGATCACCTCC-3'
SUMO3 directo	5'-CCCCGCGATCGCGCCGACGAAAAGCCCAAG-3'
SUMO3 inverso	5'-CCCCGCGGCCGCTCAGTAGACACCTCCCG-3'
Vim directo	5'-GGGGTGTACAGCGATCGCATGTCGACCCACGCGT-3'
Vim inverso	5'-CCCCGAATTTCGGGCCGCTTATTCAAGGTCATCGTGATGCT-3'

25 Las construcciones de expresión en mamíferos que codifican fusiones de traducción de DNMT1 humano, ADN-ligasa I, ADN-ligasa III, p21, FEN I, la subunidad p66 de la polimerasa  $\delta$ , PARP-1, PARP-2 y PCNA se describieron previamente (Cazzalini et al., 2003; Maeda et al., 2006; Meder et al., 2005; Mortusewicz et al., 2005; Schermelleh et al., 2005; Sporbert et al., 2005). Las construcciones de delección y los dominios aislados de ADN-ligasa I y III se describieron en Mortusewicz et al. (Mortusewicz et al., 2006). Las inmunoprecipitaciones se realizaron con un GFP-nanotrap (Rothbauer et al., 2007) como se describe anteriormente (Mortusewicz et al., 2006). Todas las construcciones de fusión se sometieron a prueba para la expresión y localización correctas.

#### Cultivo celular y transfección

Células BHK transgénicas (clon n.º 2) y células U2OS (clon 2-6-3) que contenían repeticiones del operador de *lac* se cultivaron en condiciones selectivas en DMEM suplementado con suero bovino fetal al 10 % y 150 µg/ml de higromicina B (PAA Laboratories) como se describe (Janicki et al., 2004; Tsukamoto et al., 2000). Para la microscopía, las células se cultivaron hasta un 50-70 % de confluencia, en cubreobjetos de vidrio de 18 x 18 o en portaobjetos µ-Slide (ibidi, Munich, Alemania) y después se cotransfectaron con las construcciones de expresión indicadas usando el reactivo de transfección de Polyplus jetPEI™ (BIOMOL GmbH, Hamburgo, Alemania) de acuerdo con las instrucciones del fabricante. Después de 6-10 horas, el medio de transfección se cambió a medio de cultivo recién preparado y las células se incubaron entonces durante otras 12-24 horas antes de la microscopía de células vivas o la fijación con formaldehído al 3,7 % en PBS durante 10 min a temperatura ambiente. Las células fijadas se permeabilizaron con Triton X-100 al 0,2 % en PBS durante 3 min, se tiñeron con contraste con DAPI y se montaron en Vectashield (Vector Laboratories, CA, EE. UU.).

### Microscopía

Se analizaron células vivas o fijadas que expresaban proteínas fluorescentes usando un microscopio confocal Leica TCS SP2 AOBS equipado con un objetivo de inmersión en aceite 63x/1,4 NA Plan-Apochromat. Los fluoróforos se excitaban con un láser de diodo de 405 nm, un láser de argón de 488 nm y uno de 514 nm y un láser de estado sólido bombeado por diodo (DPSS) de 561 nm. Los apilamientos de imágenes confocales de células vivas o fijadas se registraron típicamente con un tamaño de fotograma de 512 x 512 píxeles, un tamaño de píxel de 50-100 nm, un tamaño z de pasadas de 250 nm y el orificio abierto a 1 unidad de Airy. Se generó una proyección de máxima intensidad de varias semisecciones z con ImageJ (versión 1.38, <http://rsb.info.nih.gov/ij/>).

### Ejemplo 2: procedimiento para detectar la interacción de proteínas

Para visualizar las interacciones proteína-proteína en células vivas en tiempo real, se desarrolló un ensayo de doble híbrido de fluorescencia (F2H). El fundamento para el ensayo F2H se basa en el hecho de que las proteínas están transitando libremente en la célula a menos que las interacciones con otros componentes celulares las inmovilicen en estructuras específicas (Phair y Misteli, 2000). Se usó una línea celular BHK y U2OS descritas previamente que albergan ambas una integración estable de aproximadamente 200-1000 copias de un plásmido que porta 256 copias de la secuencia del operador de *lac* (Janicki et al., 2004; Tsukamoto et al., 2000). Se generó una construcción de expresión que codificaba una proteína señuelo fluorescente que consiste en una proteína fluorescente (FP), el represor de *lac* (LacI) y la proteína X que se va a someter a prueba para determinar las interacciones (señuelo), que da como resultado la proteína de triple fusión *FP-LacI-X* (figura 1a) o *X-LacI-FP*. Esta proteína de fusión se une a la matriz del operador, que entonces se hace visible debido al enriquecimiento focal de la señal de FP. Una segunda proteína de fusión marcada de forma diferente (FP-Y, presa) puede interactuar con la proteína señuelo X, lo que conduce a la colocalización de las señales de PF (figura 1b) o puede no interactuar, lo que da como resultado una distribución dispersa de la fluorescencia presa (figura 1c).

### Ejemplo 3: Visualización de interacciones entre proteínas de reparación del ADN

Para someter a prueba el ensayo F2H, se analizó la interacción descrita previamente entre las dos proteínas de reparación del ADN, la ADN ligasa III y XRCC1 (Caldecott et al., 1994; Wei et al., 1995) y los resultados se compararon con los datos obtenidos a partir de ensayos desplegables. Se ha demostrado previamente que esta interacción está mediada por el dominio BRCT de la ADN ligasa III que lo dirige a sitios de reparación del ADN (Mortusewicz et al., 2006). Se generó una proteína de fusión señuelo que consiste en XRCC1 seguido de LacI y la proteína roja fluorescente RFP monomérica (mRFP). Como era de esperar, esta proteína de fusión se localizaba en la matriz del operador de *lac* en células BHK transfectadas de forma transitoria (figura 2a). Ambos, la ADN ligasa III de longitud completa marcada con GFP, así como el dominio BRCT marcado con GFP aislado se colocalizan con XRCC1 en la matriz del operador de *lac*, mientras que una proteína de fusión a la que le falta el dominio BRCT muestra una distribución dispersa. De forma destacable, la ADN ligasa I altamente homóloga, que cataliza la misma reacción que la ADN ligasa III, no se une a XRCC1 (figura 2a y figura 6). Una comparación directa de los datos de F2H con datos obtenidos a partir de experimentos de Co-IP revela que estos dos procedimientos dieron resultados similares (figura 2b). Además, también se pudo observar la interacción descrita recientemente de XRCC1 con PCNA (Fan et al., 2004) y los dos PARP dependientes de daños del ADN, PARP-1 y PARP-2 (Masson et al., 1998; Schreiber et al., 2002) (figura 7). Estos resultados demuestran que el ensayo F2H es muy adecuado para estudiar las interacciones proteína-proteína en células vivas.

### Ejemplo 4: Análisis de la dependencia del ciclo celular de las interacciones proteína-proteína

Un reto en el análisis de las interacciones proteína-proteína es controlar los cambios transitorios provocados por, por ejemplo, la progresión del ciclo celular u otros estímulos externos. Se analizó la interacción descrita previamente entre la ADN metiltransferasa 1 (Dnmt1) y PCNA, que está mediada por el dominio de unión a PCNA (PBD) y dirige Dnmt1 a sitios de replicación del ADN en fase S (Chuang et al., 1997; Easwaran et al., 2004). Estos hallazgos plantearon la cuestión de si esta interacción se produce únicamente en la fase S en focos de replicación o en todo el ciclo celular. Se generaron dos proteínas señuelo que comprenden partes de Dnmt1 fusionadas a LacI e YFP. Un señuelo (PBD-LacI-YFP) comprende los aa 118-427 de Dnmt1 incluyendo el PBD, mientras que el segundo señuelo ( $\Delta$ PBD-LacI-YFP) carece del PBD y comprende los aa 629-1089 de Dnmt1 (figura 3a). Como proteína presa, se usó



RFP-PCNA que además marca sitios de replicación del ADN, lo que permite la identificación de células en fase S (Easwaran et al., 2005; Sporbert et al., 2005). Las posibilidades de unión de estas proteínas de fusión en la matriz del operador de *lac* y la horquilla de replicación se resumen en la figura 3b.

5 Fuera de la fase S, la parte Lacl de las proteínas señuelo se une únicamente a la matriz del operador de *lac* integrado cromosómicamente, que - dependiendo de la ploidía de la célula - se hace visible como una o dos manchas fluorescentes en el núcleo. La interacción de RFP-PCNA con la parte PBD de la proteína señuelo da como resultado la colocalización de las señales fluorescentes en la matriz del operador de *lac* (panel superior de la figura 3c), mientras que la delección del PBD en la proteína señuelo conduce a una distribución dispersa de RFP-PCNA en células que no está en la fase S (panel superior de la figura 3d). Esto ilustra claramente que la interacción dependiente de PBD de Dnmt1 con PCNA también se produce fuera de la fase S.

10 En células en fase S, RFP-PCNA se localiza en sitios de replicación del ADN en curso y, además, se recluta a la matriz del operador de *lac* por la proteína señuelo PBD-Lacl-YFP (panel inferior de la figura 3c). En contraste, cuando RFP-PCNA se coexpresa junto con una proteína señuelo que carece de un PBD funcional ( $\Delta$ PBD-Lacl-YFP), RFP-PCNA se enriquece de forma exclusiva en los sitios de replicación del ADN y no en la matriz del operador de *lac*, destacado por  $\Delta$ PBD-Lacl-YFP (panel inferior de la figura 3d).

15 Estos resultados muestran claramente que la localización de RFP-PCNA (presa) en la matriz del operador de *lac* depende de la presencia del PBD en la construcción señuelo y que esta interacción no está limitada a la fase S.

A continuación, se analizó la interacción de otras proteínas que contienen PBD con PCNA. Se generó una proteína de fusión señuelo que comprende PCNA fusionado con una NLS adicional, seguida de Lacl y RFP (NLS-PCNA-Lacl-RFP). Cuando se coexpresaba con GFP-ligasa I, ambas proteínas de fusión se localizaban en la matriz del operador de *lac*, lo que indica interacción entre PCNA y la ADN ligasa I. La delección del PBD conduce a una distribución dispersa de ADN ligasa I, mientras que el PBD de la ADN ligasa I en solitario era suficiente para la unión a PCNA en la matriz del operador de *lac* (figura 8). Esto está de acuerdo con estudios previos que muestran que el PBD de la ADN ligasa I es necesario y suficiente para dirigirse a los sitios de replicación y reparación del ADN (Cardoso et al., 1997; Montecucco et al., 1995; Mortusewicz et al., 2006). De forma destacable, usando el ensayo F2H se pudo demostrar que la ADN ligasa I, así como el PBD aislado, se pueden unir a PCNA también fuera de la fase S. Asimismo, se pudo demostrar la unión de varias proteínas de replicación y reparación adicionales como FEN1, p21 y la subunidad p66 de la polimerasa  $\delta$  a PCNA en células que no están en la fase S (figura 9). Tomados conjuntamente, se pudo demostrar que la interacción entre proteínas de replicación y PCNA no está limitada a la fase S, sino que también se produce en células que no están en la fase S y fuera de la maquinaria de replicación. Esto ilustra que el ensayo F2H ofrece el potencial único para analizar cambios específicos del ciclo celular en las interacciones proteína-proteína en células vivas.

#### Ejemplo 5: Detección de interacciones entre proteínas relacionadas con la enfermedad de Huntington

35 Para investigar si el ensayo F2H también puede detectar interacciones proteína-proteína que tienen lugar en otros compartimentos celulares, se sometió a prueba el ensayo F2H con interacciones de proteínas identificadas en el contexto de la enfermedad de Huntington por ensayos de doble híbrido de levaduras (Y2H) (Goehler et al., 2004). Se analizó la interacción de una proteína citoplásmica (vimentina) y dos proteínas nucleares (HZFH y SUMO3). La vimentina se ha descrito como un componente del citoesqueleto y participa en procesos de transporte, mientras que HZFH y SUMO3 están implicadas en la regulación de la transcripción y el mantenimiento del ADN (Goehler et al., 2004). Estas proteínas se fusionan con una mCherry-Lacl-NLS roja fluorescente o con NLS-GFP para generar conjuntos de proteínas señuelo y presa. Células BHK que portan una matriz del operador de *lac* se transfectaron con todas las posibles combinaciones de construcciones de expresión y se sometieron a análisis microscópico. Se pudo detectar una interacción entre vimentina y HZFH independiente de si estas dos proteínas se usaban como señuelo o presa (figura 4 y datos no mostrados). También se pudo detectar la interacción indicada entre SUMO3 y HZFH, mientras que vimentina y SUMO3 no interactuaban, como se describe previamente (figura 4) (Goehler et al., 2004). Estos resultados muestran que las interacciones de las proteínas nucleares y citoplásmicas se pueden estudiar con el ensayo F2H.

#### Ejemplo 6: Detección de interacciones entre proteínas mitocondriales

50 A continuación, se investigó si el ensayo F2H también es adecuado para detectar interacciones proteína-proteína que se producen en otros orgánulos celulares. Con este fin, se analizó la interacción entre dos proteínas mitocondriales, DDP1 (péptido 1 de distonía por hipoacusia) y TIMM13. Ambas proteínas están codificadas en el núcleo y se importan al espacio intermembrana mitocondrial (IMS), formando un complejo hexamérico (figura 5a). Dentro del IMS, el complejo DDP1-TIMM13 facilita la importación de proteínas hidrófobas de la maquinaria de importación mitocondrial a la membrana interna mitocondrial (Rothbauer et al., 2001). Una mutación del gen *DDP1* se asoció con el síndrome de Mohr-Tranebjaerg, que es un trastorno progresivo y neurodegenerativo (Tranebjaerg et al., 1995). Se sabe que esta mutación sin sentido C66W provoca un fenotipo completo y afecta al motivo Cys(4) altamente conservado de DDP1. Estudios previos han demostrado que este intercambio de aminoácidos anula la interacción entre DDP1 y TIMM13 en el IMS (Hofmann et al., 2002).

5 Usando una proteína de fusión señuelo roja fluorescente que comprende LacI-NLS-TIMM13 y proteínas presa DDP1 naturales (GFP-DDP1) o mutantes (GFP-DDP1<sup>C66W</sup>) marcadas con GFP, se analizó esta interacción de proteínas mitocondriales específica con el ensayo F2H. Se descubrió que GFP-DDP1 se colocaliza con TIMM13 en la matriz del operador de *lac* (figura 5b), mientras que GFP-DDP1<sup>C66W</sup> estaba distribuida uniformemente (figura 5c). Estos resultados demuestran que el ensayo F2H también es adecuado para el análisis de interacciones proteína-proteína que se producen fuera del núcleo y la caracterización de mutaciones puntuales relacionadas con enfermedades que interrumpen estas interacciones.

## REFERENCIAS

- Ando, R., H. Hama, M. Yamamoto-Hino, H. Mizuno, y A. Miyawaki. 2002. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc Natl Acad Sci U S A.* 99:12651-6.
- 5 Baird, G.S., D.A. Zacharias, y R.Y. Tsien. 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A.* 97:11984-9.
- Bebbington, C.R., G. Renner, S. Thomson, D. King, D. Abrams, y G.T. Yarranton. 1992. High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology (N Y).* 10:169-75.
- 10 Beddoe, T., M. Ling, S. Dove, O. Hoegh-Guldberg, R.J. Devenish, M. Prescott, y J. Rossjohn. 2003. The production, purification and crystallization of a pocilloporin pigment from a reef-forming coral. *Acta Crystallogr D Biol Crystallogr.* 59:597-9.
- Borden, K.L. 2002. Pondering the promyelocytic leukemia protein (PML) puzzle: possible functions for PML nuclear bodies. *Mol Cell Biol.* 22:5259-69.
- 15 Braasch, D.A., y D.R. Corey. 2001. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem Biol.* 8:1-7.
- Bulina, M.E., D.M. Chudakov, N.N. Mudrik, y K.A. Lukyanov. 2002. Interconversion of Anthozoa GFP-like fluorescent and non-fluorescent proteins by mutagenesis. *BMC Biochem.* 3:7.
- Bunch, T.A., Y. Grinblat, y L.S. Goldstein. 1988. Characterization and use of the Drosophila metallothionein promoter in cultured Drosophila melanogaster cells. *Nucleic Acids Res.* 16:1043-61.
- 20 Caldecott, K.W., C.K. McKeown, J.D. Tucker, S. Ljungquist, y L.H. Thompson. 1994. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol Cell Biol.* 14:68-76.
- Campbell, R.E., O. Tour, A.E. Palmer, P.A. Steinbach, G.S. Baird, D.A. Zacharias, y R.Y. Tsien. 2002. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A.* 99:7877-82.
- 25 Cardoso, M.C., C. Joseph, H.P. Rahn, R. Reusch, B. Nadal-Ginard, y H. Leonhardt. 1997. Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication *in vivo*. *J Cell Biol.* 139:579-87.
- Cazzalini, O., P. Perucca, F. Riva, L.A. Stivala, L. Bianchi, V. Vannini, B. Ducommun, y E. Prosperi. 2003. p21CDKN1A does not interfere with loading of PCNA at DNA replication sites, but inhibits subsequent binding of DNA polymerase delta at the G1/S phase transition. *Cell Cycle.* 2:596-603.
- 30 Chuang, L.S., H.I. Ian, T.W. Koh, H.H. Ng, G. Xu, y B.F. Li. 1997. Human DNA- (cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science.* 277:1996-2000.
- Dietzel, S., K. Zolghadr, C. Hepperger, y A.S. Belmont. 2004. Differential large- scale chromatin compaction and intranuclear positioning of transcribed versus non-transcribed transgene arrays containing {beta}-globin regulatory sequences. *J Cell Sci.* 117:4603-4614.
- 35 Dove, S.G., O. Hoegh-Guldberg, y S. Ranganathan. 2001. Major colour patterns of reef-building corals are due to a family of GFP-like proteins. *Coral Reefs.* 19:197-204.
- Easwaran, H.P., H. Leonhardt, y M.C. Cardoso. 2005. Cell Cycle Markers for Live Cell Analyses. *Cell Cycle.* 4.
- Easwaran, H.P., L. Schermelleh, H. Leonhardt, y M.C. Cardoso. 2004. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* 5:1181-6.
- 40 Fan, J., M. Otterlei, H.K. Wong, A.E. Tomkinson, y D.M. Wilson, 3°. 2004. XRCC1 co-localizes and physically interacts with PCNA. *Nucleic Acids Res.* 32:2193-201.
- Goehler, H., M. Lalowski, U. Stelzl, S. Waelter, M. Stroedicke, U. Worm, A. Droege, K.S. Lindenberg, M. Knoblich, C. Haenig, M. Herbst, J. Suopanki, E. Scherzinger, C. Abraham, B. Bauer, R. Hasenbank, A. Fritzsche, A.H. Ludewig, K. Bussow, S.H. Coleman, C.A. Gutekunst, B.G. Landwehrmeyer, H. Lehrach, y E.E. Wanker. 2004. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell.* 15:853-65.
- 45 Gurskaya, N.G., A.F. Fradkov, A. Tersikh, M.V. Matz, Y.A. Labas, V.I. Martynov, Y.G. Yanushevich, K.A. Lukyanov, y S.A. Lukyanov. 2001. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett.* 507:16-20.
- Hofmann, S., U. Rothbauer, N. Muhlenbein, W. Neupert, K.D. Gerbitz, M. Brunner, y M.F. Bauer. 2002. The C66W mutation in the deafness dystonia peptide 1 (DDP1) affects the formation of functional DDP1.TIM13 complexes in the mitochondrial intermembrane space. *J Biol Chem.* 277:23287-93.

- Horton, R.M., H.D. Hunt, S.N. Ho, J.K. Pullen, y L.R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap *extension*. *Gene*. 77:61-8.
- 5 Janicki, S.M., T. Tsukamoto, S.E. Salghetti, W.P. Tansey, R. Sachidanandam, K.V. Prasanth, T. Ried, Y. Shav-Tal, E. Bertrand, R.H. Singer, y D.L. Spector. 2004. From silencing to gene expression: real-time analysis in single cells. *Cell*. 116:683-98.
- Kalderon, D., W.D. Richardson, A.F. Markham, y A.E. Smith. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature*. 311:33-8.
- Kerppola, T.K. 2006. Visualization of molecular interactions by fluorescence complementation. *Nat Rev Mol Cell Biol*. 7:449-56.
- 10 Lukyanov, K.A., A.F. Fradkov, N.G. Gurskaya, M.V. Matz, Y.A. Labas, A.P. Savitsky, M.L. Markelov, A.G. Zaraisky, X. Zhao, Y. Fang, W. Tan, y S.A. Lukyanov. 2000. Natural animal coloration can Be determined by a nonfluorescent green fluorescent protein homolog. *J Biol Chem*. 275:25879- 82.
- Maeda, Y., T.C. Hunter, D.E. Loudy, V. Dave, V. Schreiber, y J.A. Whitsett. 2006. PARP-2 interacts with TTF-1 and regulates expression of surfactant protein- B. *J Biol Chem*. 281:9600-6.
- 15 Masson, M., C. Niedergang, V. Schreiber, S. Muller, J. Menissier-de Murcia, y G. de Murcia. 1998. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol Cell Biol*. 18:3563-71.
- Matz, M.V., A.F. Fradkov, Y.A. Labas, A.P. Savitsky, A.G. Zaraisky, M.L. Markelov, y S.A. Lukyanov. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol*. 17:969-73.
- 20 Maul, G.G., D. Negorev, P. Bell, y A.M. Ishov. 2000. Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. *J Struct Biol*. 129:278-87.
- Meder, V.S., M. Boeglin, G. de Murcia, y V. Schreiber. 2005. PARP-1 and PARP-2 interact with nucleophosmin/B23 and accumulate in transcriptionally active nucleoli. *J Cell Sci*. 118:211-22.
- 25 Merzlyak, E.M., J. Goedhart, D. Shcherbo, M.E. Bulina, A.S. Shcheglov, A.F. Fradkov, A. Gaintzeva, K.A. Lukyanov, S. Lukyanov, T.W. Gadella, y D.M. Chudakov. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods*. 4:555-7.
- Miller, C.L., M.M. Arnold, T.J. Broering, C. Eichwald, J. Kim, J.B. Dinoso, y M.L. Nibert. 2007. Virus-derived platforms for visualizing protein associations inside cells. *Mol Cell Proteomics*. 6:1027-38.
- Miyawaki, A. 2003. Visualization of the spatial and temporal dynamics of intracellular signaling. *Dev Cell*. 4:295-305.
- 30 Moller, A., H. Sirma, T.G. Hofmann, S. Rueffer, E. Klimczak, W. Droge, H. Will, y M.L. Schmitz. 2003. PML is required for homeodomain-interacting protein kinase 2 (HIPK2)-mediated p53 phosphorylation and cell cycle arrest but is dispensable for the formation of HIPK domains. *Cancer Res*. 63:4310-4.
- Montecucco, A., E. Savini, F. Weighardt, R. Rossi, G. Ciarrocchi, A. Villa, y G. Biamonti. 1995. The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. *Embo J*. 14:5379-86.
- 35 Mortusewicz, O., U. Rothbauer, M.C. Cardoso, y H. Leonhardt. 2006. Differential recruitment of DNA Ligase I and III to DNA repair sites. *Nucleic Acids Res*. 34:3523-32.
- Mortusewicz, O., L. Schermelleh, J. Walter, M.C. Cardoso, y H. Leonhardt. 2005. Recruitment of DNA methyltransferase I to DNA repair sites. *Proc Natl Acad Sci U S A*.
- 40 Murphy, G., M.I. Cockett, R.V. Ward, y A.J. Docherty. 1991. Matrix metalloproteinase degradation of elastin, type IV collagen and proteoglycan. A quantitative comparison of the activities of 95 kDa and 72 kDa gelatinases, stromelysins-1 and -2 and punctuated metalloproteinase (PUMP). *Biochem J*. 277 (Pt 1):277-9.
- Newman, J.R., S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, y J.S. Weissman. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 441:840-6.
- 45 Parrish, J.R., K.D. Gulyas, y R.L. Finley, Jr. 2006. Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol*. 17:387-93.
- Phair, R.D., y T. Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature*. 404:604-9.
- Robinett, C.C., A. Straight, G. Li, C. Wilhelm, G. Sudlow, A. Murray, y A.S. Belmont. 1996. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell*

*Biol.* 135:1685-700.

Rothbauer, U., S. Hofmann, N. Muhlenbein, S.A. Paschen, K.D. Gerbitz, W. Neupert, M. Brunner, y M.F. Bauer. 2001. Role of the deafness dystonia peptide 1 (DDP1) in import of human Tim23 into the inner membrane of mitochondria. *J Biol Chem.* 276:37327-34.

- 5 Rothbauer, U., K. Zolghadr, S. Muyldermans, A. Schepers, M.C. Cardoso, y H. Leonhardt. 2007. A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics.*

Schermelleh, L., F. Spada, H.P. Easwaran, K. Zolghadr, J.B. Margot, M.C. Cardoso, y H. Leonhardt. 2005. Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods.* 2:751-756.

- 10 Schreiber, V., J.C. Ame, P. Dolle, I. Schultz, B. Rinaldi, V. Fraulob, J. Menissier-de Murcia, y G. de Murcia. 2002. Poly(ADP-ribose) polymerase-2 (PARP-2) is required for efficient base excision DNA repair in association with PARP-1 and XRCC1. *J Biol Chem.* 277:23028-36.

Sekar, R.B., and A. Periasamy. 2003. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol.* 160:629-33.

- 15 Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N. Giepmans, A.E. Palmer, y R.Y. Tsien. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol.* 22:1567-72.

Sporbert, A., P. Domaing, H. Leonhardt, y M.C. Cardoso. 2005. PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res.* 33:3521-8.

- 20 Terskikh, A., A. Fradkov, G. Ermakova, A. Zaisky, P. Tan, A.V. Kajava, X. Zhao, S. Lukyanov, M. Matz, S. Kim, I. Weissman, y P. Siebert. 2000. "Fluorescent timer": protein that changes color with time. *Science.* 290:1585- 8.

Tranebjaerg, L., C. Schwartz, H. Eriksen, S. Andreasson, V. Ponjavic, A. Dahl, R.E. Stevenson, M. May, F. Arena, D. Barker, et al. 1995. A new X linked recessive deafness syndrome with blindness, dystonia, fractures, and mental deficiency is linked to Xq22. *J Med Genet.* 32:257-63.

Tsien, R.Y. 1998. The green fluorescent protein. *Annu Rev Biochem.* 67:509-44.

- 25 Tsukamoto, T., N. Hashiguchi, S.M. Janicki, T. Tumber, A.S. Belmont, and D.L. Spector. 2000. Visualization of gene activity in living cells. *Nat Cell Biol.* 2:871-8.

Tu, H., Q. Xiong, S. Zhen, X. Zhong, L. Peng, H. Chen, X. Jiang, W. Liu, W. Yang, J. Wei, M. Dong, W. Wu, y A. Xu. 2003. A naturally enhanced green fluorescent protein from magnificent sea anemone (*Heteractis magnifica*) and its functional analysis. *Biochem Biophys Res Commun.* 301:879-85.

- 30 Tumber, T., G. Sudlow, y A.S. Belmont. 1999. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol.* 145:1341-54.

Vazquez, J., A.S. Belmont, y J.W. Sedat. 2001. Multiple regimes of constrained chromosome motion are regulated in the interphase *Drosophila* nucleus. *Curr Biol.* 11:1227-39.

- 35 Wei, Y.F., P. Robins, K. Carter, K. Caldecott, D.J. Pappin, G.L. Yu, R.P. Wang, B.K. Shell, R.A. Nash, P. Schar, et al. 1995. Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol Cell Biol.* 15:3206-16.

Wiedenmann, J., C. Elke, K.D. Spindler, y W. Funke. 2000. Cracks in the beta-can: fluorescent proteins from *Anemonia sulcata* (Anthozoa, Actinaria). *Proc Natl Acad Sci U S A.* 97:14091-6.

- 40 Wiehler, J., J. von Hummel, y B. Steipe. 2001. Mutants of *Discosoma* red fluorescent protein with a GFP-like Chromophore. *FEBS Lett.* 487:384-9.

Zhang, J., R.E. Campbell, A.Y. Ting, y R.Y. Tsien. 2002. Creating new fluorescent probes for cell biology. *Nat Rev Mol Cell Biol.* 3:906-18.

Zimmer, M. 2002. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem Rev.* 102:759-81.

45

# ES 2 606 706 T3

## LISTADO DE SECUENCIAS

<110> Ludwig-Maximilians-Universität München

<120> Un ensayo de doble híbrido fluorescente (F2H) para la visualización directa de interacciones de proteínas en células vivas

5 <130> N3214 EP

<160> 10

<170> PatentIn versión 3.3

<210> 1

<211> 238

10 <212> PRT

<213> *Aequorea victoria*

<400> 1

```

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

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

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

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

Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln
65          70          75          80

His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
85          90          95

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

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

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

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

Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val
165         170         175
    
```

ES 2 606 706 T3

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

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

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

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

<210> 2

<211> 717

<212> ADN

5 <213> Aequorea victoria

<400> 2

atgagtaaag gagaagaact tttcactgga gttgtcccaa ttcttggtga attagatggt 60  
 gatgttaatg ggcacaaatt ttctgtcagt ggagaggggtg aaggatgatgc aacatacggga 120  
 aaacttacc ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 180  
 gtcactactt tctcttatgg tgttcaatgc ttttcaagat acccagatca tatgaaacag 240  
 catgactttt tcaagagtgc catgcccga ggttatgtac aggaaagaac tatatttttc 300  
 aaagatgacg ggaactacaa gacacgtgct gaagtcaagt ttgaaggtga tacccttggt 360  
 aatagaatcg agttaaagg tattgatttt aaagaagatg gaaacattct tggacacaaa 420  
 ttggaataca actataactc acacaatgta tacatcatgg cagacaaaaca aaagaatgga 480  
 atcaaagtta acttcaaaat tagacacaac attgaagatg gaagcgttca actagcagac 540  
 cattatcaac aaaatactcc aattggcgat ggccctgtcc ttttaccaga caaccattac 600  
 ctgtccacac aatctgcctt ttcgaaagat cccaacgaaa agagagacca catggtcctt 660  
 cttgagtttg taacagctgc tgggattaca catggcatgg atgaactata caaataa 717

<210> 3

10 <211> 225

<212> PRT

<213> Discosoma

<400> 3

ES 2 606 706 T3

Met Arg Ser Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys Val  
1 5 10 15

Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu  
20 25 30

Gly Glu Gly Arg Pro Tyr Glu Gly His Asn Thr Val Lys Leu Lys Val  
35 40 45

Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Gln  
50 55 60

Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His Pro Ala Asp Ile Pro  
65 70 75 80

Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val  
85 90 95

Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser  
100 105 110

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

Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu  
130 135 140

Ala Ser Thr Glu Arg Leu Tyr Pro Arg Asp Gly Val Leu Lys Gly Glu  
145 150 155 160

Ile His Lys Ala Leu Lys Leu Lys Asp Gly Gly His Tyr Leu Val Glu  
165 170 175

Phe Lys Ser Ile Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Tyr  
180 185 190

Tyr Tyr Val Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr  
195 200 205

Thr Ile Val Glu Gln Tyr Glu Arg Thr Glu Gly Arg His His Leu Phe  
210 215 220

Leu

225

5 <210> 4

<211> 678

<212> ADN

<213> Discosoma



ES 2 606 706 T3

<400> 4

atgaggtctt ccaagaatgt tatcaaggag ttcattgaggt ttaaggttcg catggaagga 60  
acgggtcaatg ggcacgagtt tgaaatagaa ggcgaaggag aggggaggcc atacgaaggc 120  
cacaataaccg taaagcttaa ggtaaccaag gggggacctt tgccatttgc ttgggatatt 180  
ttgtcaccac aatttcagta tggaagcaag gtatatgtca agcacctgc cgacatacca 240  
gactataaaa agctgtcatt tcctgaagga tttaaattggg aaagggtcat gaactttgaa 300  
gacgggtggcg tcgttactgt aaccaggat tccagtttgc aggatggctg tttcatctac 360  
aagggtcaagt tcattggcgt gaactttcct tccgatggac ctggttatgca aaagaagaca 420  
atgggctggg aagccagcac tgagcgtttg taccctcgtg atggcgtgtt gaaaggagag 480  
attcataagg ctctgaagct gaaagacggt ggtcattacc tagttgaatt caaaagtatt 540  
tacatggcaa agaagcctgt gcagctacca ggggtactact atggtgactc caaactggat 600  
ataacaagcc acaacgaaga ctatacaatc gttgagcagt atgaaagaac cgagggacgc 660  
caccatctgt tcctttag 678

5 <210> 5

<211> 229

<212> PRT

<213> secuencia artificial

10 <220>

<221> fuente

<223> /nota = "Descripción de secuencia artificial: mRFP1"

<400> 5

15

ES 2 606 706 T3

Met Ala Ser Ser Glu Asp Val Ile Lys Glu Phe Met Arg Phe Lys Val  
1 5 10 15

Arg Met Glu Gly Ser Val Asn Gly His Glu Phe Glu Ile Glu Gly Glu  
20 25 30

Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr Ala Lys Leu Lys Val  
35 40 45

Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp Ile Leu Ser Pro Gln  
50 55 60

Phe Gln Tyr Gly Ser Lys Ala Tyr Val Lys His Pro Ala Asp Ile Pro  
65 70 75 80

Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe Lys Trp Glu Arg Val  
85 90 95

Met Asn Phe Glu Asp Gly Gly Val Val Thr Val Thr Gln Asp Ser Ser  
100 105 110

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

Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys Thr Met Gly Trp Glu  
130 135 140

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

Ile Lys Met Arg Leu Lys Leu Lys Asp Gly Gly His Tyr Asp Ala Glu  
165 170 175

Val Lys Thr Thr Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Ala  
180 185 190

Tyr Lys Thr Asp Ile Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr  
195 200 205

Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly Arg His Ser Thr Gly  
210 215 220

Ala Gly Leu Tyr Lys  
225

<210> 6

<211> 690

5 <212> ADN

<213> secuencia artificial

# ES 2 606 706 T3

<220>

<221> fuente

<223> /nota = "Descripción de secuencia artificial: mRFP1"

5

<400> 6

```
atggcctcct ccgaggacgt catcaaggag ttcattgcgt tcaaggtgcg catggagggc      60
tccgtgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc      120
accagaccg ccaagctgaa ggtgaccaag ggcggccccc tgccttcgc ctgggacatc      180
ctgtcccctc agttccagta cggctccaag gcctacgtga agcaccocgc cgacatcccc      240
gactacttga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag      300
gacggcggcg tggtgaccgt gaccaggac tcctccctgc aggacggcga gttcatctac      360
aaggtgaagc tgcgcgccac caacttcccc tccgacggcc ccgtaatgca gaagaagacc      420
atgggctggg aggctccac cgagcggatg taccocgagg acggcgcctt gaagggcgag      480
atcaagatga ggctgaagct gaaggacggc ggccactacg acgccgaggt caagaccacc      540
tacatggcca agaagcccgt gcagctgccc ggcgcctaca agaccgacat caagctggac      600
atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgcgc cgagggccgc      660
cactccaccg gcgcccgcct gtacaagtaa                                     690
```

<210> 7

<211> 117

10 <212> PRT

<213> secuencia artificial

<220>

<221> fuente

15 <223> /nota = "Descripción de secuencia artificial: fragmento VHH de cromocuerpo fluorescente contra GFP"

<400> 7

ES 2 606 706 T3

Met Ala Asp Val Gln Leu Val Glu Ser Gly Gly Ala Leu Val Gln Pro  
1 5 10 15

Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Pro Val Asn  
20 25 30

Arg Tyr Ser Met Arg Trp Tyr Arg Gln Ala Pro Gly Lys Glu Arg Glu  
35 40 45

Trp Val Ala Gly Met Ser Ser Ala Gly Asp Arg Ser Ser Tyr Glu Asp  
50 55 60

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ala Arg Asn Thr  
65 70 75 80

Val Tyr Leu Gln Met Asn Ser Leu Lys Pro Glu Asp Thr Ala Val Tyr  
85 90 95

Tyr Cys Asn Val Asn Val Gly Phe Glu Tyr Trp Gly Gln Gly Thr Gln  
100 105 110

Val Thr Val Ser Ser  
115

<210> 8

<211> 351

<212> ADN

5 <213> secuencia artificial

<220>

<221> fuente

10 <223> /nota = "Descripción de secuencia artificial: secuencia que codifica el fragmento VHH de cromocuerpo fluorescente de la SEQ ID NO: 7"

<400> 8

```
atggccgatg tgcagctggg ggagtctggg ggagccttgg tgcagccggg ggggtctctg 60
agactctcct gtgcagcctc tggattcccc gtcaatcgct atagtatgag gtggtaccgc 120
caggctccag ggaaggagcg cgagtgggtc gcggtatga gtagtgctgg tgatcgttca 180
agttatgaag actccgtgaa gggccgattc accatctcca gagacgacgc caggaatagc 240
gtgtatctgc aatgaacag cctgaaacct gaggacacgg ccgtgtatta ctgtaatgtc 300
aatgtgggct ttgagtactg gggccagggg acccaggtca ccgtctcctc a 351
```

<210> 9

15 <211> 117

ES 2 606 706 T3

<212> PRT

<213> secuencia artificial

<220>

5 <221> fuente

<223> /nota = "Descripción de secuencia artificial: fragmento VHH de cromocuerpo fluorescente contra GFP (mutante C92S)"

<400> 9

Met Ala Asp Val Gln Leu Val Glu Ser Gly Gly Ala Leu Val Gln Pro  
1 5 10 15

Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Pro Val Asn  
20 25 30

Arg Tyr Ser Met Arg Trp Tyr Arg Gln Ala Pro Gly Lys Glu Arg Glu  
35 40 45

Trp Val Ala Gly Met Ser Ser Ala Gly Asp Arg Ser Ser Tyr Glu Asp  
50 55 60

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ala Arg Asn Thr  
65 70 75 80

Val Tyr Leu Gln Met Asn Ser Leu Lys Pro Glu Asp Thr Ala Val Tyr  
85 90 95

Tyr Ser Asn Val Asn Val Gly Phe Glu Tyr Trp Gly Gln Gly Thr Gln  
100 105 110

Val Thr Val Ser Ser  
115

10

<210> 10

<211> 351

<212> ADN

<213> secuencia artificial

15

<220>

<221> fuente

<223> /nota = "Descripción de secuencia artificial: secuencia que codifica el fragmento VHH de cromocuerpo fluorescente de la SEQ ID NO: 9"

20

<400> 10

## ES 2 606 706 T3

```
atggccgatg tgcagctggt ggagtctggg ggagccttgg tgcagccggg ggggtctctg      60
agactctcct gtgcagcctc tggattcccc gtcaatcgct atagtatgag gtggtaccgc      120
caggctccag ggaaggagcg cgagtgggtc gcgggtatga gtagtgctgg tgatcgttca      180
agttatgaag actccgtgaa gggccgattc accatctcca gagacgacgc caggaatagc      240
gtgtatctgc aaatgaacag cctgaaacct gaggacacgg ccgtgtatta ctcgaatgtc      300
aatgtgggct ttgagtactg gggccagggg acccaggtca ccgtctcctc a                351
```

**REIVINDICACIONES**

1. Un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende:
  - (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - 5 (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - (ii) un (poli)péptido que se une específicamente a la proteína verde fluorescente (GFP)
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende
    - (i) GFP; y
    - (ii) un (poli)péptido señuelo
  - 10 (c) expresar en la misma célula una tercera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y
    - (ii) un (poli)péptido presa
  - (d) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación,
  - 15 en el que una colocación de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del dicho (poli)péptido señuelo y presa.
2. Un procedimiento *in vitro* para identificar un compuesto que modula la interacción de dos (poli)péptidos
  - (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - 20 (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - (ii) un (poli)péptido que se une específicamente a la proteína verde fluorescente (GFP)
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende
    - (i) GFP; y
    - (ii) un (poli)péptido señuelo
  - 25 (c) expresar en la misma célula una tercera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y
    - (ii) un (poli)péptido presa que se sabe o se sospecha que interacciona con el (poli)péptido señuelo
  - (d) poner en contacto la célula con un compuesto de prueba; y
  - (e) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación;
  - 30 en el que un cambio en el grado de colocación de la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede modular la interacción del (poli)péptido señuelo y presa.
- 35 3. Un procedimiento *in vitro* para determinar la fuerza relativa de la interacción de dos proteínas con una tercera proteína, que comprende:
  - (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - (i) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - 40 (ii) un (poli)péptido que se une específicamente a la proteína verde fluorescente (GFP)
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende

- (i) GFP; y
- (ii) un (poli)péptido señuelo
- (c) expresar en la misma célula una tercera proteína de fusión que comprende
  - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP; y
  - (ii) un primer (poli)péptido presa
- (d) expresar en la misma célula una cuarta proteína de fusión que comprende
  - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la de GFP y de la del (poli)péptido fluorescente de dicha tercera proteína de fusión; y
  - (ii) un segundo (poli)péptido presa
- (e) detectar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula tras la excitación,

en el que una magnitud de colocalización de la emisión de fluorescencia de la segunda y la tercera proteína de fusión diferente en comparación con la de la segunda y la cuarta proteína de fusión en el núcleo de la célula es indicativa de una fuerza de unión diferente del primero y el segundo (poli)péptido presa con el (poli)péptido señuelo.

4. Un procedimiento *in vitro* para detectar interacciones proteína-proteína, que comprende:
- (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente
    - (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - (iii) un (poli)péptido señuelo;
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y
    - (ii) un (poli)péptido presa;
  - (c) detectar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula tras la excitación,

en el que una colocalización de la emisión de fluorescencia de ambas proteínas de fusión en el núcleo de la célula es indicativa de una interacción del dicho (poli)péptido señuelo y presa.

5. Un procedimiento *in vitro* para identificar un compuesto que modula la interacción de dos (poli)péptidos
- (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente
    - (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - (iii) un (poli)péptido señuelo
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y
    - (ii) un (poli)péptido presa que se sabe o se sospecha que interacciona con el (poli)péptido señuelo
  - (c) poner en contacto la célula con un compuesto de prueba; y
  - (d) detectar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula tras la excitación;



en el que un cambio en el grado de colocación de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede modular la interacción del (poli)péptido señuelo y presa.

- 5 6. Un procedimiento *in vitro* para determinar la fuerza relativa de la interacción de dos proteínas con una tercera proteína, que comprende
- (a) expresar en una célula eucariota una primera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente
    - 10 (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
    - (iii) un (poli)péptido señuelo
  - (b) expresar en la misma célula una segunda proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera proteína de fusión; y
    - 15 (ii) un primer (poli)péptido presa
  - (c) expresar en la misma célula una tercera proteína de fusión que comprende
    - (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en dicha primera y segunda proteína de fusión; y
    - (ii) un segundo (poli)péptido presa
  - 20 (d) detectar la emisión de fluorescencia de las partes fluorescentes de la primera, la segunda y la tercera proteína de fusión en la célula tras la excitación,

en el que una magnitud de colocación de la emisión de fluorescencia de la primera y la segunda proteína de fusión diferente en comparación con la de la primera y la tercera proteína de fusión en el núcleo de la célula es indicativa de una fuerza de unión diferente de la primera y la segunda proteína presa con el (poli)péptido señuelo.

- 25 7. El procedimiento de la reivindicación 2 o 5, en el que una disminución de la colocación de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede debilitar la interacción del (poli)péptido señuelo y presa.
- 30 8. El procedimiento de la reivindicación 2 o 5, en el que un aumento en la colocación de las partes fluorescentes de ambas proteínas de fusión en el núcleo de la célula en comparación con la observada en el núcleo de una célula de referencia que no está en contacto con el compuesto de prueba es indicativo de que el compuesto puede inducir o potenciar la interacción del (poli)péptido señuelo y presa.
- 35 9. El procedimiento de la reivindicación 4, en el que la detección se emplea para investigar la dependencia de las interacciones proteína-proteína de los procesos celulares y en el que el procedimiento comprende además (d1) supervisar la emisión de fluorescencia de las partes fluorescentes de la primera y la segunda proteína de fusión en la célula en el curso de uno o más procesos en la célula; en el que un cambio en el grado de colocación de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en la célula es indicativo de una dependencia de la interacción de uno o más procesos celulares.
- 40 10. El procedimiento de la reivindicación 1, en el que la detección se emplea para investigar la dependencia de las interacciones proteína-proteína de los procesos celulares y en el que el procedimiento comprende además (d1) supervisar la emisión de fluorescencia de las partes fluorescentes de la segunda y la tercera proteína de fusión en la célula en el curso de uno o más procesos en la célula; en el que un cambio en el grado de colocación de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en la célula es indicativo de una dependencia de la interacción de uno o más procesos celulares.
- 45 11. El procedimiento de la reivindicación 9 o 10, en el que el proceso celular es el ciclo celular, la secreción, la traslocación o la transducción de señales.
- 50 12. El procedimiento de la reivindicación 4, en el que la detección se emplea para determinar la fuerza de una interacción proteína-proteína y en el que el procedimiento comprende además, en caso de que se detecte una colocación de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión, (d2) la extinción selectiva de la fluorescencia de dicha segunda proteína de fusión y el control del restablecimiento de la

colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en el tiempo, en el que el tiempo necesario para establecer colocalización es indicativo de la fuerza de la interacción proteína-proteína.

- 5 13. El procedimiento de la reivindicación 1, en el que la detección se emplea para determinar la fuerza de una interacción proteína-proteína y en el que el procedimiento comprende además, en caso de que se detecte una colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión, (d2) la extinción selectiva de la fluorescencia de dicha tercera proteína de fusión y el control del restablecimiento de la colocalización de la emisión de fluorescencia de las partes fluorescentes de ambas proteínas de fusión en el tiempo, en el que el tiempo necesario para establecer colocalización es indicativo de la fuerza de la interacción proteína-proteína.
- 10 14. El procedimiento de una cualquiera de las reivindicaciones 1 a 13, en el que el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona directamente con estructuras proteicas o no proteicas acumuladas en sitios distintos en el núcleo de la célula.
- 15 15. El procedimiento de una cualquiera de las reivindicaciones 1 a 13, en el que el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona indirectamente con estructuras proteicas o no proteicas acumuladas en sitios distintos en el núcleo de la célula.
- 20 16. El procedimiento de una cualquiera de las reivindicaciones 1 a 13, en el que los componentes (i), (ii) y/o (iii) de dicha primera proteína de fusión y/o los componentes (i) y (ii) de dicha segunda proteína de fusión y/o los componentes (i) y (ii) de dicha tercera proteína de fusión y/o los componentes (i) y (ii) de dicha cuarta proteína de fusión están conectados a través de un engarce.
17. El procedimiento de una cualquiera de las reivindicaciones 1 a 16, en el que cualquiera de los (poli)péptidos presa comprende una señal de localización nuclear.
18. El procedimiento de una cualquiera de las reivindicaciones 1 a 3, en el que dicha segunda y/o tercera y/o cuarta proteína de fusión comprende una señal de localización nuclear.
- 25 19. El procedimiento de una cualquiera de las reivindicaciones 1 a 18, en el que la expresión en la célula eucariota se logra transfectando las moléculas de ácido nucleico que codifican dicha primera y segunda proteína de fusión en uno o más vectores.
20. El procedimiento de una cualquiera de las reivindicaciones 1 a 19, en el que el (poli)péptido acumulado en sitios distintos del núcleo de la célula se ha introducido en la célula.
- 30 21. El procedimiento de una cualquiera de las reivindicaciones 1 a 20, en el que los sitios distintos de la célula forman estructuras inertes.
22. El procedimiento de la reivindicación 21, en el que la estructura inerte es la lámina nuclear o especie nuclear.
23. El procedimiento de una cualquiera de las reivindicaciones 1 a 20, en el que el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona con cuerpos de PML.
- 35 24. El procedimiento de una cualquiera de las reivindicaciones 1 a 23, en el que el (poli)péptido acumulado en sitios distintos del núcleo es heterólogo a la célula.
25. El procedimiento de una cualquiera de las reivindicaciones 1 a 21 y 24, en el que el (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula, interacciona con ADN.
26. El procedimiento de la reivindicación 25, en el que el ADN es el operador de lac.
- 40 27. El procedimiento de la reivindicación 26, en el que el operador de lac está presente en el núcleo en varias copias.
28. El procedimiento de una cualquiera de las reivindicaciones 1 a 27, en el que dicha primera proteína de fusión comprende LacI.
- 45 29. El procedimiento de una cualquiera de las reivindicaciones 1 a 28, en el que la detección se lleva a cabo usando un microscopio de fluorescencia.
30. El procedimiento de una cualquiera de las reivindicaciones 1 a 29, en el que la célula eucariota es una célula viva.
31. El procedimiento de una cualquiera de las reivindicaciones 1 a 30, que comprende, además, la localización de los (poli)péptidos señuelo y/o presa en la célula.
- 50 32. Un vector para su uso en la expresión de la primera y la segunda proteína de fusión de acuerdo con el

procedimiento de una cualquiera de las reivindicaciones 4 a 6 en una célula eucariota, que comprende

- 5 A. una molécula de ácido nucleico que codifica la proteína de fusión que comprende
- (i) un (poli)péptido fluorescente
  - (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
  - (iii) un (poli)péptido señuelo; y
- 10 B. una molécula de ácido nucleico que codifica la proteína de fusión que comprende
- (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en la proteína de fusión codificada por la molécula de ácido nucleico de acuerdo con A; y
  - (ii) un (poli)péptido presa.
33. Una célula eucariota transfectada con el vector de la reivindicación 32.
34. Una célula eucariota en la que se expresa la primera y la segunda proteína de fusión de acuerdo con el procedimiento de una cualquiera de las reivindicaciones 4 a 6, en la que la célula se transfecta con
- 15 A. un vector que comprende una molécula de ácido nucleico que codifica la proteína de fusión que comprende
- (i) un (poli)péptido fluorescente
  - (ii) un (poli)péptido que, cuando se expresa en una célula, se acumula en sitios distintos en el núcleo de la célula; y
  - (iii) un (poli)péptido señuelo; y
- 20 B. un vector que comprende una molécula de ácido nucleico que codifica la proteína de fusión que comprende
- (i) un (poli)péptido fluorescente, cuya longitud de onda de excitación y/o emisión difiere de la del (poli)péptido fluorescente comprendido en la proteína de fusión codificada por la molécula de ácido nucleico de acuerdo con A; y
  - (ii) un (poli)péptido presa.
- 25 35. Una célula eucariota que tiene varias copias del operador de lac integradas de manera estable en su ADN y que expresan de forma estable una proteína que comprende (i) un (poli)péptido que se une al operador de lac y (ii) un (poli)péptido que se une específicamente a la proteína verde fluorescente (GFP).
- 30 36. El procedimiento de una cualquiera de las reivindicaciones 1 a 3 y 7 a 31 o la célula eucariota de la reivindicación 35, en el que dicho (poli)péptido que se une específicamente a la proteína verde fluorescente (GFP) comprende un (poli)péptido que tiene la secuencia de aminoácidos de la SEQ ID NO: 7 o 9.

Figura 1

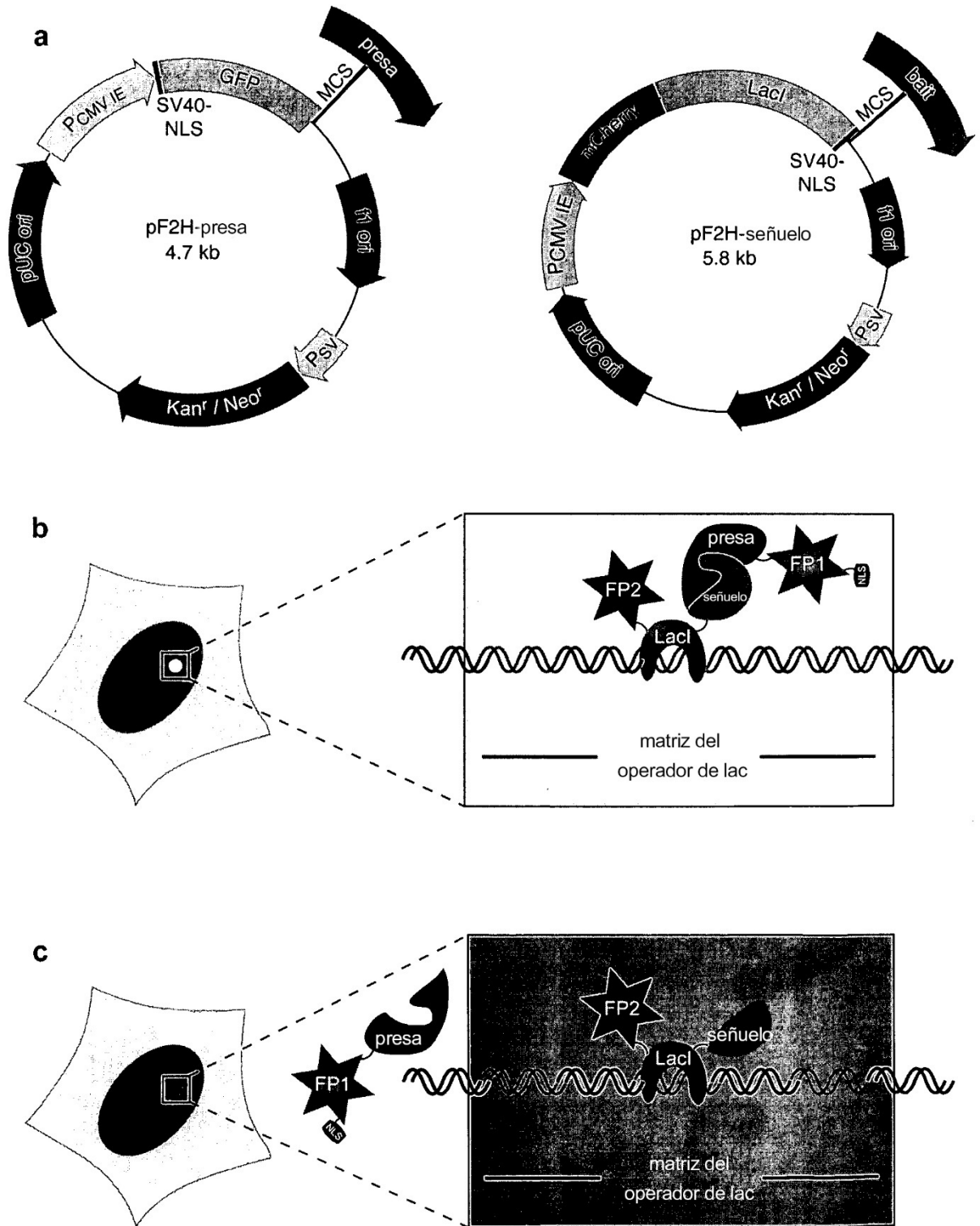


Figura 2

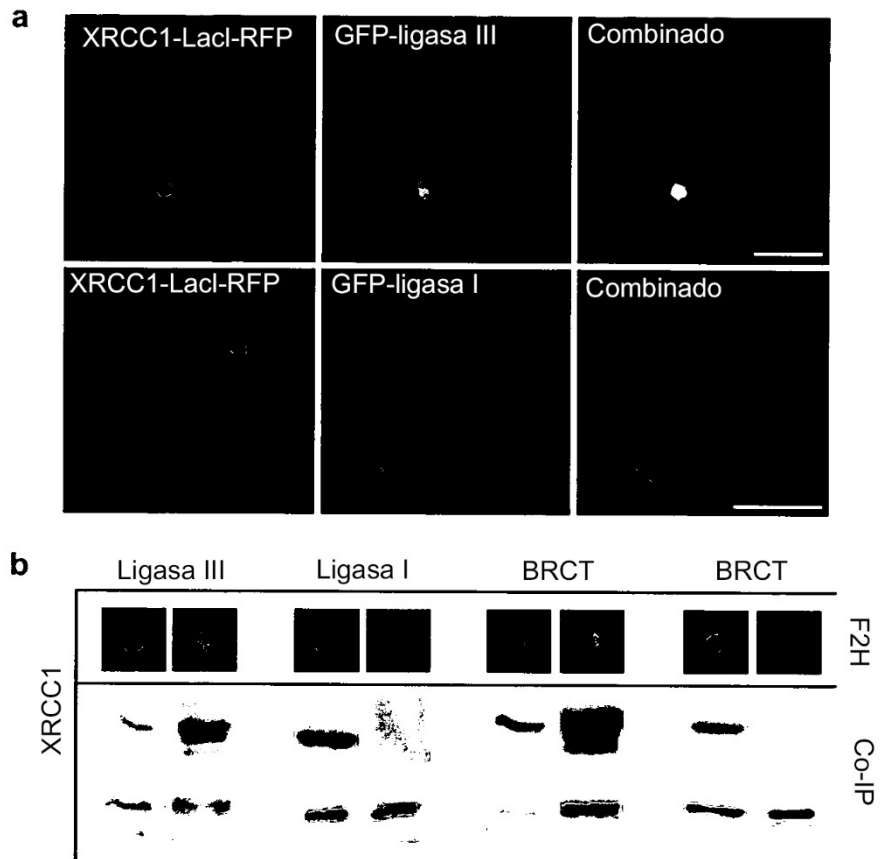


Figura 3

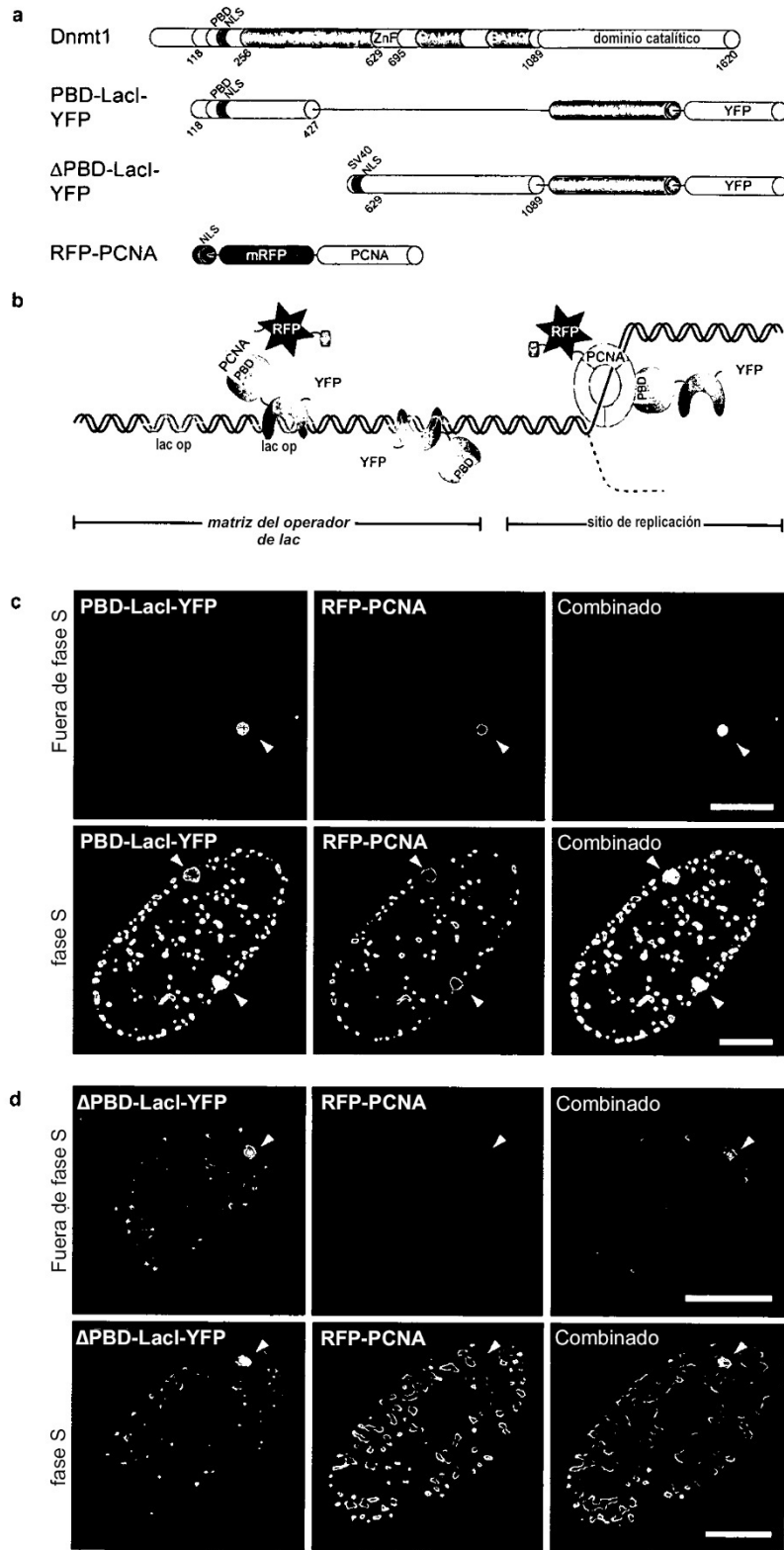


Figura 4

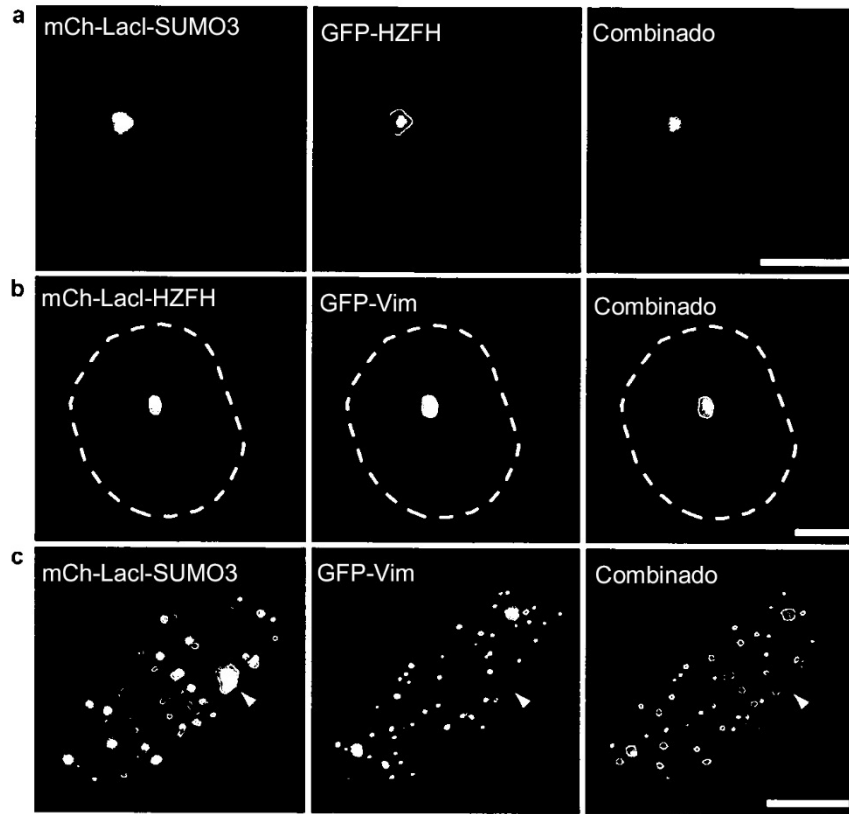


Figura 5

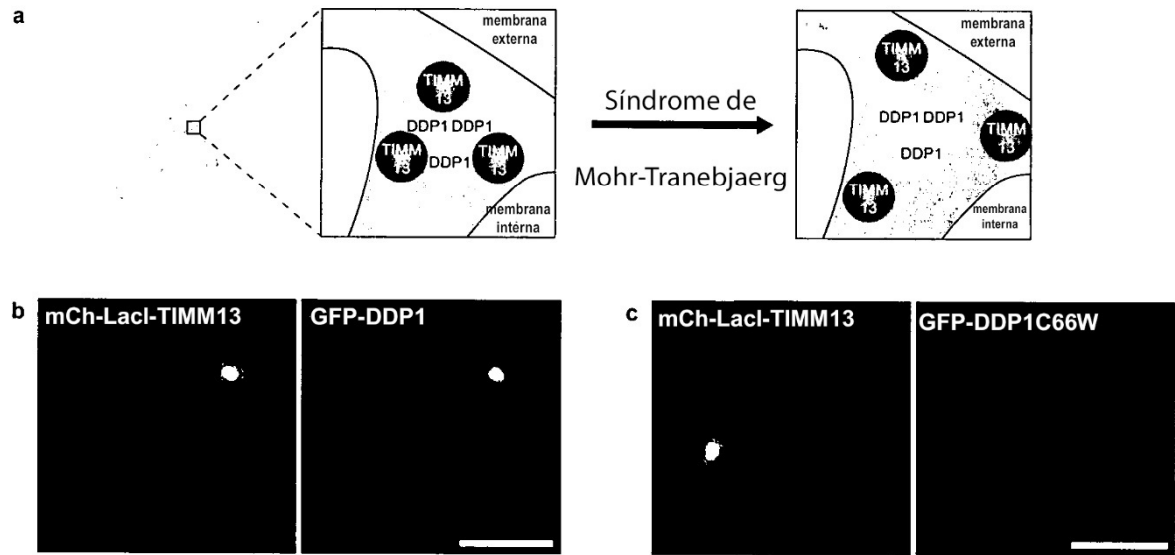




Figura 6

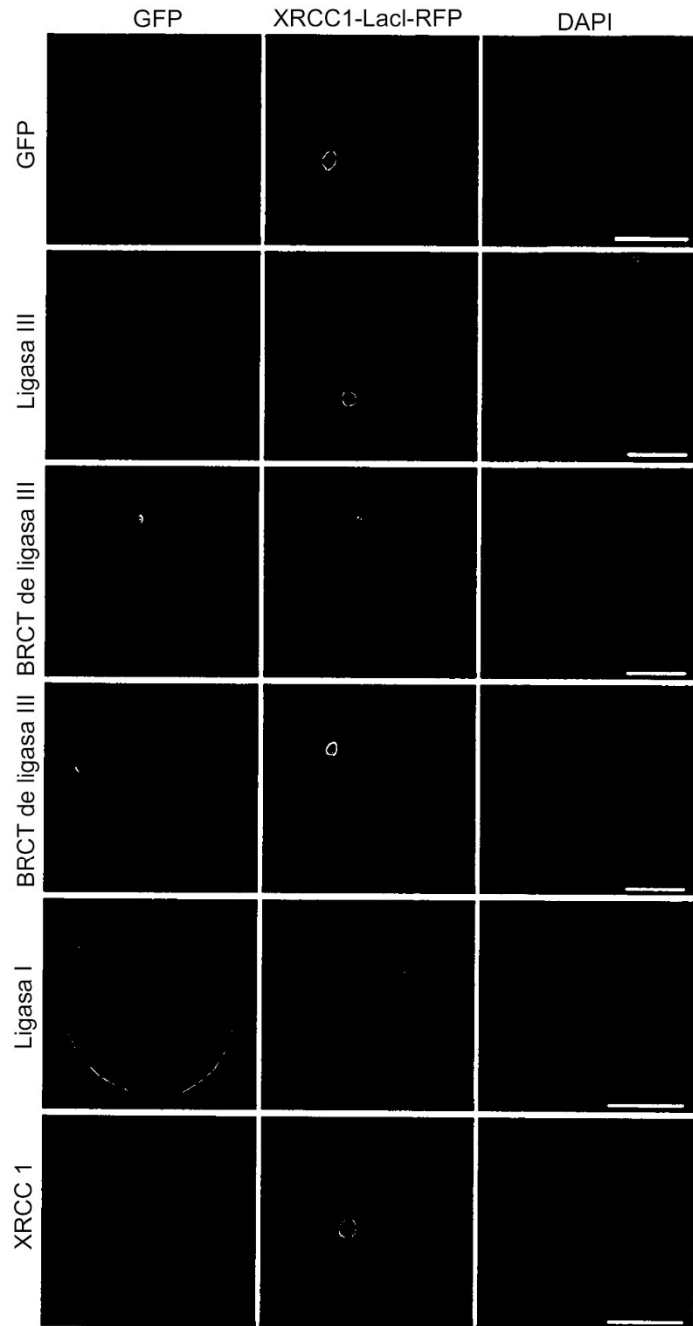


Figura 7

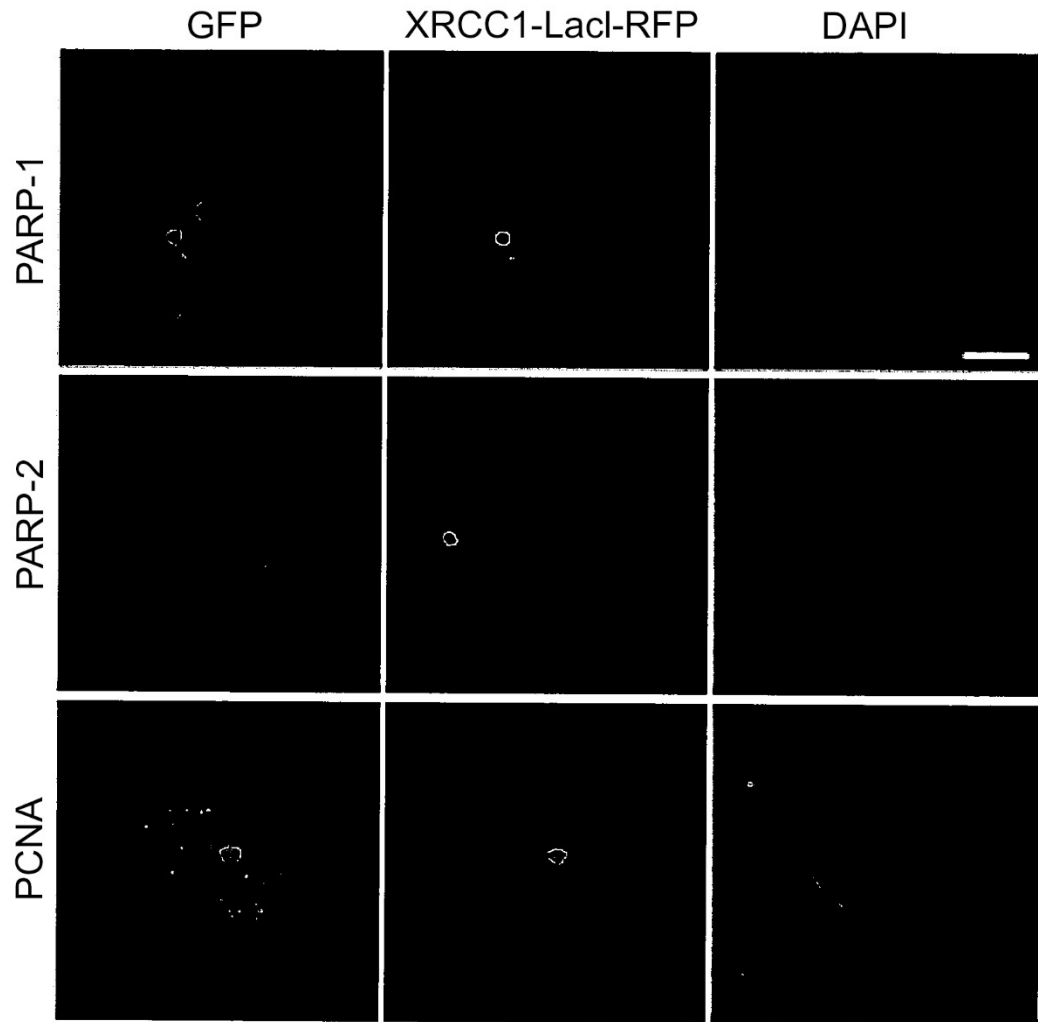


Figura 8

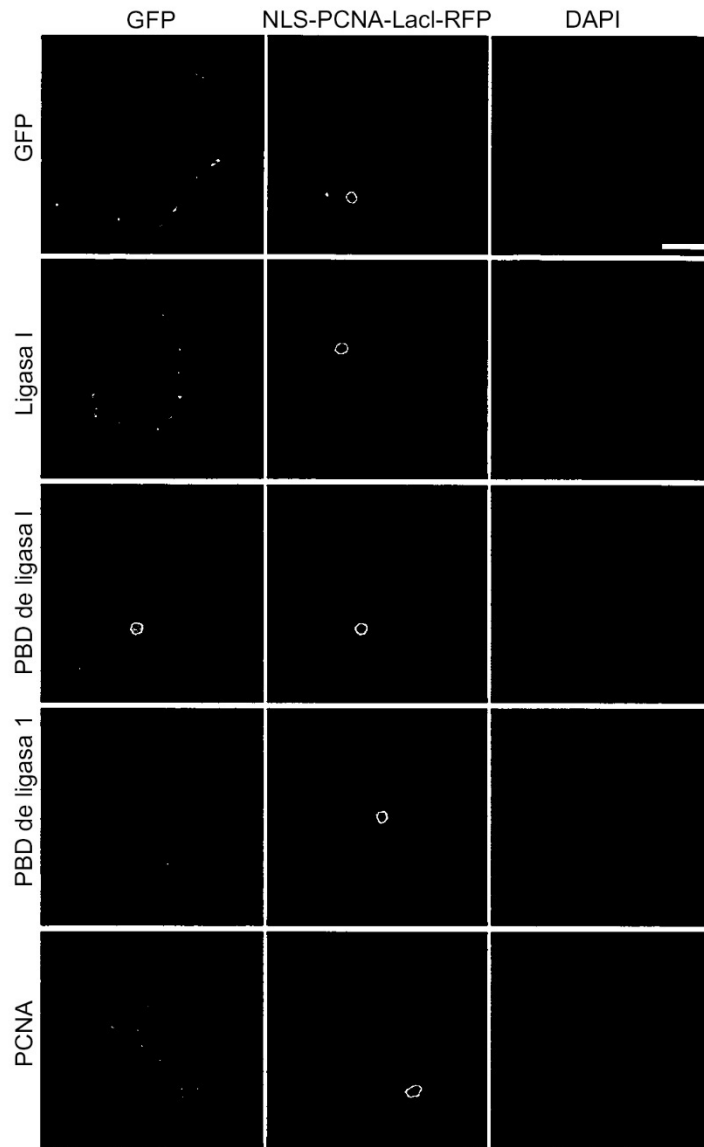


Figura 9

