

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 985 624**

51 Int. Cl.:

G01N 33/50 (2006.01)

C12Q 1/6897 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.12.2019** **PCT/EP2019/086750**

87 Fecha y número de publicación internacional: **25.06.2020** **WO20128015**

96 Fecha de presentación y número de la solicitud europea: **20.12.2019** **E 19832117 (6)**

97 Fecha y número de publicación de la concesión europea: **19.06.2024** **EP 3899535**

54 Título: **Ensayo de detección**

30 Prioridad:

20.12.2018 GB 201820863

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

06.11.2024

73 Titular/es:

THE UNIVERSITY OF BATH (100.0%)

Claverton Down

Bath Somerset BA2 7AY, GB

72 Inventor/es:

MASON, JODY MICHAEL y

KAD, NEIL MARK

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 985 624 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

Ensayo de detección

5 Campo de la invención

La presente invención se refiere a un método para detectar un antagonista de una proteína de unión al ADN, tal como un factor de transcripción, como se define en las reivindicaciones. Los métodos de la invención encuentran aplicación, por ejemplo, en la identificación de antagonistas que puedan ser útiles en el tratamiento de cánceres.

10 Antecedentes

Los factores de transcripción suelen estar mal regulados en los cánceres humanos y atacarlos puede ser muy eficaz en el tratamiento de neoplasias malignas concretas, por ejemplo, los receptores de hormonas nucleares (Soucek *et al.*, 2013). La diversidad y potencia de los factores de transcripción como impulsores de la transformación celular justifica una búsqueda continua como dianas terapéuticas para el descubrimiento de fármacos.

Un ejemplo de un factor de transcripción cuya alteración está alterada en numerosos cánceres humanos es la Proteína Activadora-1 (AP-1). La AP-1 consiste en un heterodímero formado por un miembro de la familia Fos y un miembro de la familia Jun y se une a variaciones del sitio de unión consenso del ADN conocido como elemento de respuesta TPA (TRE): TGA[G/C]TCA (Ozanne *et al.* 2007). La familia Fos contiene cuatro proteínas (c-Fos, FosB, Fra1 y Fra2), mientras que la familia Jun contiene tres proteínas (c-Jun, JunB y JunD). La AP-1 es miembro de la familia de reguladores de la transcripción de cremallera de leucina básica(bZIP). La dimerización, típicamente por miembros de las familias Jun y Fos, es dirigida por el motivo de cremallera de leucina/hélice superenrollada y es necesario para la actividad funcional de la proteína. A continuación, los restos clave dentro de la región básica pueden interactuar con bases específicas dentro del TRE y, por lo tanto, pueden regular la expresión de una serie de genes implicados en la diferenciación, la proliferación y la apoptosis. Los niveles elevados de AP-1 conducen a una mayor transactivación de la expresión del gen diana, lo que conduce al crecimiento y la transformación maligna en muchos tipos de células.

Los factores de transcripción tales como AP-1 generalmente siguen siendo dianas farmacológicas muy estimulantes cuando se utilizan enfoques dirigidos por moléculas pequeñas debido a la falta de bolsas tratables y a la falta de interacciones proteína-proteína. Por lo general, los equipos han intentado apuntar a los factores de transcripción indirectamente, por ejemplo, promover la selección de ellos para la proteólisis o inhibir los factores de señalización hacia 5' (McCormick, 2016).

Otro abordaje para atacar los factores de transcripción es identificar pequeños antagonistas peptídicos. Estos péptidos son más grandes que una molécula pequeña típica y tienen una superficie más grande capaz de formar muchos más puntos de contacto con alta especificidad para alterar mejor las interacciones proteína-proteína (PPI). Muchos abordajes de diseño racionales, abordajes de detección aleatoria y sistemas de selección dan como resultado la identificación con éxito de compuestos capaces de unirse a dianas proteicas determinadas.

Un ejemplo de un método que tiene como objetivo identificar inhibidores que alteran las PPI se describe en Miranda *et al.* (2011). Este método utiliza un sistema bacteriano de dos híbridos inversos que permite la identificación de proteínas asociadas que interactúan. Los compuestos que interrumpen la interacción entre dos proteínas específicas, expresados como fusiones híbridas de un complejo represor quimérico (434 y P22), previenen la expresión de genes informadores. De esta forma, se informa que son identificables inhibidores que interrumpen la interacción de las dos proteínas diana.

Otro ejemplo de un método que busca identificar inhibidores de los PPI se describe en Mason *et al.* 2006. Este método identificó inhibidores potenciales de c-Jun utilizando un ensayo de complementación de fragmentos de proteínas (PCA) que implican la secuencia de cFos como estructura de diseño y con semialeatorización de restos clave implicado en la heterodimerización con cJun. Con este método, se identificó el inhibidor de PPI FosW. FosW es un péptido de 37 restos basado en el dominio cremallera de leucina cFos que se ha demostrado que se une a c-Jun con una afinidad elevada (Mason *et al.* 2006). El uso del abordaje PCA para identificar PPI también se describe en Pelletier *et al.* (1998).

Sin embargo, lo que es mucho más difícil de garantizar es que la unión a dicha diana dará como resultado la eliminación de la función de la proteína diana. Se han diseñado métodos que identifican inhibidores que alteran los PPI, tal como el descrito en Miranda *et al.* (2011) y Mason *et al.* (2006) para identificar agentes de unión a una diana. No revelan si el inhibidor identificado también es capaz de eliminar la función de las proteínas. Dicho de otro modo, no revelan si el inhibidor es un antagonista funcionalmente activo de la proteína. Hasta que el inhibidor se pruebe a nivel de proteína, no hay forma de saber si los inhibidores identificados se traducirán en antagonistas funcionales de la función proteica. Hay muchos casos en los que la formación de una PPI no ha garantizado la pérdida de función. Esto también es aún más relevante para usar como objetivos factores de transcripción unidos al ADN, ya que estos complejos suelen ser muy estables. Por ejemplo, Olive *et al.* (1997) describen la generación de una construcción inhibidora denominada 4H-Fos que era capaz de inhibir la unión de la proteína Fos a Jun, pero un experimento posterior de desplazamiento en gel mostró que esta construcción no podía inhibir la actividad de unión al ADN de la proteína. Esto demuestra que

los inhibidores potenciales identificados utilizando estos métodos que detectan inhibidores de PPI no necesariamente se traducen en antagonistas funcionales.

Varios documentos de la técnica anterior describen métodos para analizar PPI y/o la actividad de proteínas de unión a ADN específicas e implican el uso de un gen informador unido operativamente a un promotor, donde el promotor contiene un sitio de unión para la proteína de unión al ADN. Por ejemplo:

el documento WO 2018/075486 A1 desvela el uso de una proteína de fusión que comprende una proteína de unión a ligando y una proteína de unión a ADN en un método para detectar un ligando en una célula o una mezcla de reacción, donde la proteína de unión al ADN es capaz de unirse a un promotor unido operativamente a un gen informador.

El documento FR 2861742 desvela métodos que seleccionan compuestos que se unen al represor micobacteriano EthR y evitan que EthR reprima la expresión del gen *ethA*. Este documento desvela un método que implica una construcción donde la región promotora y operadora del gen *ethA* clonan en 5' de un gen informador.

El documento WO 99/13069 A1 desvela un sistema para analizar PPI utilizando proteínas de fusión "cebo" y "presa". La proteína cebo incluye un dominio de unión al ADN y la proteína presa incluye un dominio de represión de la transcripción y el sistema se describe como un ensayo para compuestos que interfieren con la interacción de las proteínas de fusión cebo y presa. Este documento desvela el uso de genes informadores que tienen promotores con sitios de unión para el dominio de unión al ADN de la proteína de fusión cebo.

El documento WO 01/59450 A2 desvela un método para detectar un compuesto por su efecto sobre un proceso celular proporcionando un primer polinucleótido que codifica una proteína con dedos de cinc operativamente unida a un elemento de control de la transcripción y un segundo polinucleótido que codifica un informador. El informador puede estar unido operativamente a elementos de control de la transcripción que están modulados por la proteína dedos de cinc.

El documento WO 99/57535 A2 desvela líneas celulares recombinantes y métodos de detección útiles para identificar agentes que inducen apoptosis en las células diana. Estas líneas celulares y métodos pueden comprender plásmidos que expresan promotores unidos operativamente a genes informadores.

El documento WO 2008/110784 A1 desvela un método que implica determinar si un agente de prueba modula al menos una actividad o función de *M. tuberculosis* Rv3574, un miembro de la familia TetR de represores de la transcripción. Este método puede implicar medir la expresión de un gen informador unido operativamente a una secuencia de ADN que está unida por Rv3574.

Stanley Sabrina Y R: "An Investigation of Bacteriophage Anti-CRISPR and Anti-CRISPR Associated Proteins", Tesis de la Universidad de Toronto, 1 de noviembre de 2018 (01/11/2018), desvela un ensayo indicador de transcripción en el que una proteína de unión al ADN (Cascade) está dirigida un casete de expresión de gen informador (gen *phzM*) en presencia de un anti-CRISPR de interés. El anti-CRISPR se une al complejo Cascade, previene la unión al ADN, de modo que el acceso de la ARN polimerasa al promotor *phzM* no se ve impedido y los cultivos aparecen de color verde.

Por lo tanto, sigue existiendo la necesidad de métodos que sean capaces de identificar de forma fiable antagonistas funcionalmente activos de los factores de transcripción, en particular los factores de transcripción que se sabe que su regulación está alterada en los cánceres humanos.

La presente invención se ha ideado a la luz de las consideraciones anteriores.

Divulgación de la invención

Los presentes inventores han reconocido que existe la necesidad de un método de detección que sea capaz de identificar antagonistas funcionalmente activos de proteínas de unión a ADN, tales como factores de transcripción. Los antagonistas funcionalmente activos son compuestos que pueden unirse a la proteína de unión al ADN y antagonizar su actividad de unión al ADN. Como se ha descrito en la sección de antecedentes anterior, en los cánceres humanos, la regulación de los factores de transcripción normalmente está alterada y, por lo tanto, se espera que el uso de este método de detección para identificar antagonistas funcionalmente activos sea útil para identificar inhibidores que puedan tener uso terapéutico en el tratamiento de enfermedades como el cáncer.

El método de detección de la presente invención hace uso de un casete de expresión de gen informador que codifica un producto de expresión de gen informador, tal como una proteína que proporciona una lectura fenotípica (también denominada "proteína informadora"). Los presentes inventores reconocieron que al diseñar el casete de expresión de gen informador para que contenga un sitio de unión para una proteína de unión al ADN, es posible hacer uso de la interacción entre la proteína de unión al ADN y el sitio de unión para inhibir la expresión del producto de expresión de gen informador. Sin desear quedar ligados a teoría alguna, se cree que la unión de la proteína de unión al ADN al sitio de unión inhibe la transcripción del casete de expresión de gen informador, inhibiendo así la expresión del producto de expresión de gen informador. También se cree que la inhibición de la expresión del producto informador es

completamente independiente de si la proteína de unión al ADN está clasificada como un activador de la transcripción o un represor de la transcripción (por ejemplo, si normalmente funciona para activar o reprimir la transcripción cuando se une a secuencias promotoras o potenciadoras que contienen sus sitios de unión en el genoma de una célula).

Los presentes inventores reconocieron que estos componentes podrían usarse en un método para identificar antagonistas "funcionalmente activos" de la proteína de unión al ADN. Si un compuesto de prueba es capaz de unirse e inhibir la actividad de unión al ADN de la proteína de unión al ADN, entonces reducirá la capacidad de la proteína de unión al ADN para inhibir la expresión del producto de expresión de gen informador. Por lo tanto, se ideó un método en el que la proteína de unión al ADN, el compuesto de prueba y el casete de expresión de gen informador estaban todos presentes. Si el compuesto de prueba es un antagonista funcionalmente activo, habrá una mayor expresión del producto de expresión de gen informador en presencia del compuesto de prueba con respecto a la expresión del producto de expresión en ausencia del compuesto de prueba. Se puede determinar un aumento de la expresión en presencia del compuesto de prueba comparando (i) la expresión del producto de expresión de gen informador en situaciones en las que la proteína de unión al ADN y el casete de expresión de gen informador están presentes sin el compuesto de prueba, contra (ii) la expresión del producto de expresión de gen informador en situaciones en las que la célula comprende la proteína de unión al ADN, el casete de expresión y el compuesto de prueba. Si se observa que la expresión del producto de expresión de gen informador es mayor en (ii) que en (i), aumenta la expresión del producto de expresión de gen informador en presencia del compuesto de prueba. De esta forma, se puede determinar si un determinado compuesto de prueba es un antagonista funcionalmente activo de la proteína de unión al ADN.

Por lo tanto, en un aspecto, la presente invención proporciona un método para detectar un antagonista de una proteína de unión al ADN, comprendiendo el método:

i) proporcionar una célula, en donde la célula comprende un compuesto de prueba, una proteína de unión al ADN y un casete de expresión de gen informador que codifica un producto de expresión de gen informador, en donde el casete de expresión de gen informador comprende al menos un sitio de unión para la proteína de unión a ADN de manera que la unión de la proteína de unión a ADN al sitio de unión inhibe la expresión del producto de expresión de gen informador; y

ii) determinar la expresión del producto de expresión de gen informador en presencia del compuesto de prueba;

en donde un aumento en la expresión del producto de expresión de gen informador en presencia del compuesto de prueba indica que el compuesto de prueba es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN, y

en donde algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador.

Es importante destacar que los presentes inventores han demostrado que este método de detección puede distinguir entre 1) compuestos de prueba que son capaces de unirse a la proteína de unión al ADN pero no interrumpen la actividad de unión al ADN de la proteína de unión al ADN y 2) aquellos inhibidores que son capaces tanto de unirse como de interrumpir la actividad de unión al ADN de la proteína de unión al ADN. Sólo los de la segunda categoría son antagonistas funcionalmente activos. Como se describe en el presente documento, sólo aquellos compuestos de prueba que son capaces de unirse y disociar el complejo unido al ADN, o de unirse e impedir la unión al ADN de la proteína de unión al ADN, dan como resultado el aumento en la expresión del producto de expresión de gen informador del método actual. Para disipar cualquier duda, la referencia en el presente documento a un compuesto de prueba que es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN (un "antagonista funcionalmente activo") pretende abarcar tanto inhibidores que se unen y disocian el complejo unido al ADN como inhibidores que se unen e impiden la unión al ADN de la proteína de unión al ADN.

Por lo tanto, de manera ventajosa, el método de detección permite la identificación de antagonistas funcionalmente activos de las proteínas de unión al ADN. Se espera que sea más probable que dichos antagonistas funcionalmente activos representen compuestos que serán útiles para inhibir la proteína de unión al ADN en un entorno terapéutico, por ejemplo, en el tratamiento del cáncer, que aquellos compuestos de prueba que se unen pero no alteran la función de la proteína de unión al ADN. La capacidad de detectar y desarrollar únicamente aquellos compuestos de prueba que inhiben la actividad de unión al ADN tiene el potencial de ahorrar una inversión considerable de tiempo y dinero.

Asimismo, el método indica que el antagonista identificado es específico de la proteína de unión al ADN diana, ya que se espera que sólo la disociación de la proteína de unión al ADN diana dé como resultado un aumento de la expresión. Por ejemplo, los compuestos de prueba que inhiben otras proteínas de unión al ADN que generalmente están implicadas en la expresión del producto de expresión de gen informador (por ejemplo, componentes del complejo de iniciación) no darán como resultado un aumento de la expresión y, por lo tanto, no serán identificados mediante el presente método de detección.

Una ventaja adicional del método es que la lectura es un aumento de la expresión. Tener un aumento en la expresión como lectura es ventajoso, ya que es menos probable que se produzcan falsos positivos que pueden ocurrir cuando

la lectura es una disminución en la expresión. Por lo tanto, el método de detección descrito en el presente documento produce resultados con un alto grado de confianza, reduciendo la probabilidad de necesitar pruebas de detección adicionales para confirmar el resultado.

- 5 En otro aspecto, la presente invención proporciona un método sin células para detectar un antagonista de una proteína de unión al ADN, comprendiendo el método:

- 10 i) poner en contacto un compuesto de prueba con una proteína de unión a ADN y un casete de expresión de gen informador que codifica un producto de expresión de gen informador, en donde el casete de expresión de gen informador comprende al menos un sitio de unión para la proteína de unión a ADN de manera que la unión de la proteína de unión a ADN al sitio de unión inhibe la expresión del producto de expresión de gen informador; y
- 15 ii) determinar la expresión del producto de expresión de gen informador; en donde un aumento en la expresión del producto de expresión de gen informador en presencia del compuesto de prueba indica que el compuesto de prueba es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN, y en donde el método se lleva a cabo fuera de una célula en un sistema *in vitro* que comprende los componentes necesarios para la expresión del producto de expresión de gen informador y
- 20 en donde algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador.

Otro aspecto de la presente divulgación se refiere a antagonistas identificados mediante los métodos de detección de la presente invención.

- 25 Algunos aspectos particulares de la invención se tratarán ahora en más detalle.

Producto de expresión de gen informador

- 30 El producto de expresión de gen informador utilizado en el presente documento puede ser un compuesto peptídico o una molécula de ARN (tal como microARN, ARNip o una ribozima). Los métodos para medir la expresión de proteínas son bien conocidos en la técnica e incluyen transferencia Western, inmunohistoquímica, ensayo del informador del gen de la luciferasa, ensayos colorimétricos tales como el ensayo BCA o el ensayo de Bradford, espectroscopia de UV, así como métodos que implican observar la lectura fenotípica de la proteína, como se describe con más detalle a
- 35 continuación. Los métodos para medir la expresión de una molécula de ARN también son bien conocidos en la técnica e incluyen PCR cuantitativa (PCRc), análisis transcriptómicos, espectroscopia UV y análisis de microfluidos. Preferentemente, el producto de expresión es una proteína.

- 40 En realizaciones preferidas, el producto de expresión de gen informador es una proteína que proporciona una lectura fenotípica (también denominada "proteína informadora"). Una proteína informadora que proporciona una lectura fenotípica permite ventajosamente una detección sencilla y rápida de los compuestos de prueba.

- 45 Ejemplos de proteínas informadoras incluyen proteínas de supervivencia celular, proteínas de reproducción celular, proteínas fluorescentes, proteínas de bioluminiscencia, enzimas que actúan sobre un sustrato para producir una señal colorimétrica, proteínas cinasas, proteasas, factores de transcripción y proteínas reguladoras tal como la ubiquitina. El uso de proteínas informadoras adecuadas en ensayos para determinar los PPI se describe, por ejemplo, en Wehr y Rossner (2016).

- 50 En algunas realizaciones, la proteína informadora es una proteína de supervivencia celular o una proteína de reproducción celular. Una proteína de supervivencia celular es una proteína esencial para la supervivencia celular, de modo que la supervivencia depende de la presencia o actividad de la proteína de supervivencia celular. Una proteína de reproducción celular es una proteína que es esencial para la reproducción de la célula, de modo que la proliferación (división) celular depende de la actividad de la proteína de reproducción celular. La esencialidad de la proteína de supervivencia o reproducción celular puede depender de determinadas condiciones, por ejemplo, la presencia de
- 55 ciertos factores, tal como un compuesto citotóxico, en el medio celular.

- Si la proteína informadora es una proteína de supervivencia celular, la inhibición de la expresión de la proteína de supervivencia celular dará como resultado la muerte celular. Por lo tanto, en métodos descritos en el presente documento en donde la proteína informadora es una proteína de supervivencia celular, la unión de la proteína de unión al ADN al sitio de unión en ausencia de un compuesto de prueba dará como resultado la muerte celular. La muerte celular se puede determinar mediante una de varias técnicas conocidas por el experto en la técnica, por ejemplo, la observación de cambios morfológicos tales como ampollas citoplasmáticas, reducción del volumen celular, fragmentación internucleosómica y condensación de cromatina. La escisión del ADN típica del proceso apoptótico se puede demostrar mediante ensayos TUNEL y de escalera de ADN. En estas situaciones, cuando se añade un
- 60 compuesto de prueba que es un antagonista funcionalmente activo de la proteína de unión al ADN, dará como resultado la supervivencia celular y, por lo tanto, dicho método utiliza la supervivencia celular como indicador de que
- 65

el compuesto de prueba es un antagonista funcionalmente activo. El uso de una proteína de supervivencia celular como proteína informadora puede ser ventajoso ya que proporciona una lectura binaria simple, es decir, la célula está viva o muerta.

- 5 Si la proteína informadora es una proteína de reproducción celular, la inhibición de la expresión de la proteína de reproducción celular dará como resultado que la célula sea incapaz de proliferar y, por tanto, incapaz de formar descendencia. Por lo tanto, en los métodos descritos en el presente documento en donde la proteína informadora es una proteína de reproducción celular, la unión de la proteína de unión al ADN al sitio de unión en ausencia de un compuesto de prueba inhibirá la proliferación celular. La proliferación celular se puede determinar mediante una de
- 10 varias técnicas conocidas por el experto en la técnica, por ejemplo, mediante el recuento de células individuales, focos o colonias, midiendo la actividad metabólica utilizando colorantes tales como MTT y WST-1, utilizando análogos nucleosídicos tales como bromodesoxiuridina (BrdU) y midiendo la incorporación de este análogo en las células, tiñendo células en división usando reactivos tales como éster de succinimidilo de diacetato de carboxifluoresceína y detectando marcadores de proliferación como PCNA, poisomerasa IIB o fosfohistona H3. La inhibición de la
- 15 proliferación celular también puede provocar la muerte celular, que se puede medir como se ha descrito anteriormente. En estas situaciones, cuando se añade un compuesto de prueba que es un antagonista funcionalmente activo de la proteína de unión al ADN, esto restaurará la proliferación celular y, por lo tanto, dicho método utiliza la proliferación celular como indicador de que el compuesto de prueba es un antagonista funcionalmente activo.
- 20 Ejemplos de proteínas de supervivencia celular incluyen enzimas que participan en la síntesis de compuestos necesarios para la supervivencia celular y proteínas que son capaces de inhibir la acción de un agente tóxico, tal como un antibiótico. Ejemplos de proteínas de reproducción celular incluyen enzimas necesarias para la reproducción celular.
- 25 Ejemplos de enzimas que participan en la síntesis de compuestos necesarios para la supervivencia o reproducción celular se exponen en la **Tabla 1**. Por lo tanto, en algunas realizaciones, la proteína de supervivencia celular o proteína de reproducción celular es una enzima seleccionada de la primera columna **Tabla 1**.

Tabla 1 - Ejemplos de enzimas implicadas en la síntesis de compuestos necesarios para la supervivencia o reproducción celular

Enzima	Compuestos/condiciones capaces de inhibir la función enzimática
Dihidrofolato reductasa (DHFR)	metotrexato o trimetoprima, cultivados sin nucleósidos
Timidina cinasa	ganciclovir, hipoxantina/aminopterina/timidina (HAT)
timidilato sintasa	2-fluorodesoxiuridina
Xantina-guanina fosforribosil	ácido micofenólico con xantina limitante
Asparagina sintetasa	B-aspartil hidroxamato o albizina puromicina
Citosina metiltransferasa	5-azacitidina (5-aza-CR) y 5-aza-2'-desoxicitidina
O6-alquilguanina alquiltransferasa	N-metil-N-nitrosourea
Glicinamida ribonucleótido transformilasa	dideazatetrahidrofolato, cultivados sin purina
Glicinamida ribonucleótido sintetasa	cultivados sin purina
Fosforribosil-aminoimidazol sintetasa	cultivados sin purina
Formilglicinamida ribótido amidotransferasa	L-azaserina, 6-diazo-5-oxo-L-nor-leucina, cultivados sin purina
Fosforribosil-aminoimidazol carboxilasa	cultivados sin purina
Fosforribosil-aminoimidazol carboxamida formiltransferasa	cultivados sin purina
Ácido graso sintasa	cerulenina
IMP deshidrogenasa	ácido micofenólico
histidinol deshidrogenasa	cultivados sin histidina

Por ejemplo, la dihidrofolato reductasa (DHFR) cataliza la reducción del dihidrofolato a tetrahidrofolato, para su uso en la transferencia de unidades de un carbono necesarias para la biosíntesis de serina, metionina, purinas, pantotenato y timidilato. En ausencia de la función DHFR, la síntesis *de novo* de precursores de nucleósidos (hipoxantina y timidina) se inhibe. Por lo tanto, si las células se cultivan en ausencia de una DHFR funcional y en ausencia de nucleósidos (por ejemplo, todos los nucleósidos, o al menos los nucleósidos de purina), las células morirán. Se puede controlar *in vivo* la reconstitución de la actividad enzimática mediante la supervivencia celular en células DHFR negativas cultivadas en ausencia de nucleósidos.

- 40 Ejemplos de proteínas que son capaces de inhibir la acción de un agente tóxico incluyen enzimas que son capaces

de metabolizar un agente tóxico, por ejemplo, a un agente menos tóxico y proteínas resistentes a los antibióticos, por ejemplo, proteínas que se unen e inhiben antibióticos. Ejemplos de estos se exponen en la **Tabla 2**. Por lo tanto, en algunas realizaciones, la proteína de supervivencia celular es una proteína seleccionada de la primera columna en la **Tabla 2**.

5

Tabla 2 - Ejemplos de proteínas que son capaces de inhibir la acción de un agente tóxico

Proteína de supervivencia celular	Agente tóxico/antibiótico
beta-lactamasa	antibióticos β -lactámicos tales como penicilinas, cefalosporinas, cefamicinas
cloranfenicol acetil transferasa	cloranfenicol
Puromicina-N-acetiltransferasa	puromicina
Aminoglucósido fosfotransferasa	neomicina, G418, gentamicina
Higromicina B fosfotransferasa	higromicina B
Proteína de unión a blebomicina	Blebomicina
Adenosina desaminasa	Xil-A o adenosina, alanosina y 2'-desoxicoformicina

Por ejemplo, la higromicina B es un aminociclitol que inhibe la síntesis de proteínas al alterar la translocación y promover la lectura errónea. La enzima de *E. coli* higromicina-B-fosfotransferasa desintoxica las células fosforilando la higromicina-B. Cuando se expresa en células de mamíferos, la higromicina-B-fosfotransferasa puede conferir resistencia a la higromicina-B (Gritz y Davies, 1983).

10

Como ejemplo adicional, la adenosina desaminasa (ADA) cataliza la conversión irreversible de nucleósidos de adenina citotóxicos a su respectiva conversión de nucleósidos de adenina citotóxicos a sus respectivos análogos de inosina no tóxicos. La ADA sólo se convierte en una proteína de supervivencia celular cuando se añaden concentraciones citotóxicas de adenosina. Al añadir concentraciones citotóxicas de adenosina o análogos de adenosina citotóxicos tales como 9-b-D-xilofuranosiladenina a las células, se requiere ADA para que el crecimiento celular desintoxique el agente citotóxico. Un método de ejemplo que utiliza ADA como proteína informadora se describe en Kaufman. *et al.* 1986.

15

20

La bleomicina, un miembro de la familia de antibióticos leomicina/fileomicina, es tóxico para las bacterias, hongos, plantas y células de mamíferos. La expresión de la proteína de unión a bleomicina confiere resistencia al unirse y secuestrar el fármaco y así prevenir su asociación e hidrólisis del ADN.

25

Se conocen métodos que utilizan proteínas de supervivencia celular como proteínas informadoras en la detección de inhibidores que alteran las PPI. Véase, por ejemplo, Park *et al.* (2007), que describe métodos que implican beta-lactamasa en una estrategia de fragmentación y complementación.

30

En algunas realizaciones, la proteína de supervivencia celular es una dihidrofolato reductasa (DHFR). La DHFR puede ser DHFR murina, que puede ser la proteína identificada por el número de acceso de UniProt P00375-1 (versión 3, última modificación el 23 de enero de 2007). Por ejemplo, la DHFR murina puede tener una secuencia de aminoácidos que posee al menos un 80 %, al menos un 85 % o al menos un 90 % idéntico a la secuencia establecida en la **SEQ ID NO: 1**. En realizaciones particularmente preferidas, la DHFR murina tiene una secuencia de aminoácidos que posee al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o un 100 % idéntica a la secuencia expuesta en la **SEQ ID NO: 2**.

35

40

La identidad de secuencia puede medirse mediante cualquier algoritmo de alineación adecuado, incluidos, entre otros, el algoritmo BLAST (consulte, por ejemplo, la herramienta de alineación BLAST disponible en blast.ncbi.nlm.nih.gov/Blast.cgi, opcionalmente con la configuración predeterminada), el algoritmo Needleman-Wunsch (véase, por ejemplo, el alineador EMBOSS Needle disponible en https://www.ebi.ac.uk/Tools/psa/emboss_needle/, opcionalmente con la configuración predeterminada) o el algoritmo Smith-Waterman (véase, por ejemplo, el alineador EMBOSS Water disponible en https://www.ebi.ac.uk/Tools/psa/emboss_water/, opcionalmente con la configuración predeterminada). El alineamiento óptimo se puede evaluar utilizando cualquier parámetro adecuado de un algoritmo elegido, incluidos los parámetros predeterminados.

45

50

La DHFR puede ser DHFR humana, que puede ser la proteína identificada por el número de acceso de UniProt P00374-1 (versión 2, última modificación el 23 de enero de 2007). Por ejemplo, la DHFR humana puede tener una secuencia de aminoácidos que es al menos un 80 %, al menos un 85 % o al menos un 90 % idéntico a la secuencia establecida en la **SEQ ID NO: 3**.

En algunas realizaciones, la célula puede depender condicionalmente de la actividad de la proteína de supervivencia celular o de la proteína de reproducción celular para su supervivencia o reproducción, respectivamente. En algunos

casos, la célula puede no contener una proteína de supervivencia celular endógena o una proteína de reproducción celular y, por lo tanto, requiere la adición de una proteína exógena para proliferar. En otros casos, la célula puede contener una proteína de supervivencia celular endógena o una proteína de reproducción celular endógena que es necesaria para la supervivencia o proliferación celular, respectivamente, y la función de esta proteína endógena puede inhibirse o eliminarse en determinadas condiciones (también denominadas "condiciones de selección"). Por lo tanto, la célula puede hacer uso de la proteína endógena para su supervivencia hasta que se activen las condiciones de selección, momento en el que la actividad de la proteína de supervivencia celular (también denominada "proteína de supervivencia celular exógena") se vuelve esencial para la supervivencia de la célula. Esto es ventajoso ya que permite que las células sobrevivan hasta que el método de detección esté listo para ejecutarse. Puede ser posible obtener estas condiciones de selección usando, por ejemplo, un agente de selección, donde el agente de selección es un compuesto que inhibe la actividad de la proteína endógena pero no inhibe la actividad de la proteína de supervivencia celular. Por ejemplo, la proteína endógena de supervivencia celular o la proteína de reproducción celular puede ser una de las proteínas expuestas en la primera columna de la **Tabla 1**, arriba y puede inhibirse utilizando las condiciones de selección expuestas en la segunda columna.

El término "endógeno" en el contexto de proteínas de supervivencia celular y proteínas de reproducción celular pretende significar una proteína que se origina en la célula en la que se realiza el método de selección. El término "exógeno" en el contexto de proteínas de supervivencia celular y proteínas de reproducción celular pretende significar una proteína que tiene actividad equivalente a la proteína endógena de manera que puede compensar una deficiencia en la función de la proteína de supervivencia celular endógena, pero es resistente a las condiciones de selección, por ejemplo, la presencia de un compuesto particular que inhibe la función de la proteína endógena, de modo que la supervivencia o proliferación de la célula depende de la actividad de la proteína exógena en estas condiciones de selección. La proteína exógena y endógena normalmente tendrán características similares, pero no secuencias de aminoácidos idénticas. Por ejemplo, la proteína exógena puede ser al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % idéntica a la proteína endógena y la proteína exógena puede contener una o más modificaciones en su secuencia de aminoácidos en comparación con la secuencia de aminoácidos de la proteína de supervivencia celular endógena. La proteína exógena y la proteína endógena pueden ser ortólogas, es decir, genes de diferentes especies que descienden de una secuencia ancestral común. Por ejemplo, la proteína de supervivencia celular endógena o la proteína de reproducción celular puede ser una versión bacteriana de la proteína de supervivencia celular o la proteína de reproducción celular expuesta anteriormente, por ejemplo en la **Tabla 1** o la **Tabla 2**, y la proteína exógena puede ser una proteína ortóloga de una especie de mamífero, por ejemplo, murina o humana. De forma alternativa o adicional, la proteína exógena puede contener una o más mutaciones en su secuencia de aminoácidos que la hacen resistente a las condiciones de selección que inhiben la función de la proteína endógena.

Por ejemplo, cuando la célula es una célula bacteriana, la proteína endógena puede ser una proteína de supervivencia de células bacterianas y la proteína de supervivencia de células exógenas puede ser una proteína de supervivencia de células eucariotas ortóloga, tal como una proteína de supervivencia celular de mamíferos, por ejemplo, proteína de supervivencia celular humana o de ratón. Después se puede utilizar un inhibidor bacteriano específico como agente de selección para inhibir la proteína de supervivencia de las células bacterianas sin afectar a la función de la proteína de supervivencia de las células eucariotas.

En un ejemplo más específico, la proteína de supervivencia de las células bacterianas puede ser DHFR de *E. coli*, que puede ser la proteína identificada por el número de acceso de UniProt P0ABQ4-1 (versión 1, modificada por última vez el 21 de julio de 1986) y la proteína de supervivencia de células eucariotas puede ser DHFR de ratón o humana, como se ha expuesto anteriormente. La DHFR bacteriana puede inhibirse específicamente utilizando compuestos tales como trimetoprima (TMP), que hace que las células dependan de la actividad de la DHFR exógena, por ejemplo, DHFR murina o humana, para su supervivencia.

Por lo tanto, las células bacterianas pueden cultivarse en un medio, tal como un medio de caldo líquido rico, hasta que el método de detección esté listo para realizar. En este punto, las células pueden hacer uso de la proteína endógena para sobrevivir y/o proliferar. Cuando el método de detección esté listo para ser realizado, las células pueden cultivarse en un medio que carezca de nucleósidos tales como purinas y se añade un agente de selección, tal como trimetoprima (TMP) que inhibe la DHFR bacteriana. Una vez añadido el agente de selección, las células dependen condicionalmente de la actividad de la proteína de supervivencia celular exógena, tal como DHFR de mamíferos, para su supervivencia. Por lo tanto, la supervivencia celular dependerá de la actividad de la proteína informadora celular y un aumento en la supervivencia celular indicará que el compuesto de prueba es un antagonista funcionalmente activo. Un experto en la técnica sería capaz de seleccionar un tipo y una cantidad apropiados de agente de selección para usar de manera que la supervivencia celular dependa de la actividad de la proteína informadora celular. Por ejemplo, cuando se usa TMP para inhibir la DHFR bacteriana, la concentración puede estar entre 4 y 20 μM . Si las células pueden sobrevivir en presencia de TMP, las colonias bacterianas normalmente se observan entre 2 días y 2 semanas de incubación a 37 °C.

En otro ejemplo, cuando la célula es una célula de mamífero, la proteína de supervivencia celular endógena puede ser una DHFR de mamífero. Los métodos para utilizar la detección de PPI utilizando una DHFR de mamífero como proteína de supervivencia celular se describen en Remy *et al.* (2007). Brevemente, el principio del ensayo de supervivencia de DHFR en células de mamífero es que las células que carecen de actividad de DHFR endógena,

pueden rescatarse mediante la expresión simultánea de DHFR complementaria en medios empobrecidos en nucleósidos. El ensayo podría realizarse en células negativas para DHFR o la selección se puede lograr en células positivas para DHFR usando una DHFR exógena como proteína de supervivencia celular, en donde la DHFR exógena contiene una o más mutaciones que hacen que la DHFR sea resistente a un agente de selección, tal como el fármaco antifolato metotrexato (MTX). Cuando las células se cultivan en ausencia de nucleótidos con selección de resistencia a MTX, sólo sobrevivirán las células que puedan hacer uso de la DHFR exógena.

Un ejemplo de una mutación en una DHFR de mamífero que hace que la DHFR de mamífero sea resistente al MTX es la mutación F31S, en donde la numeración de los restos es de acuerdo con la DHFR murina expuesta en la **SEQ ID NO: 1**. Por lo tanto, la proteína de supervivencia celular puede ser una DHFR murina que tiene una secuencia de aminoácidos que es al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % idéntica a la secuencia expuesta en la **SEQ ID NO: 1**, en donde la DHFR murina comprende además una serina (S) en la posición 31 y en donde la numeración de restos es según la DHFR murina expuesta en la **SEQ ID NO: 1**. La proteína de supervivencia celular puede ser una DHFR murina que tiene una secuencia de aminoácidos que es al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, un 99 % o un 100 % idéntica a la secuencia expuesta en la **SEQ ID NO: 2**, en donde la DHFR murina comprende además una serina (S) en la posición 31.

Tal método puede comprender cultivar las células de mamífero que comprenden la proteína de supervivencia celular exógena en condiciones en las que la célula depende de la actividad de la proteína informadora de células exógenas para su supervivencia. Por ejemplo, la proteína de supervivencia celular exógena puede ser DHFR murina que ha sido modificada para ser resistente al fármaco antifolato metotrexato (MTX) y la célula de mamífero puede cultivarse en ausencia de nucleósidos y en presencia de MTX. Por lo tanto, la supervivencia celular dependerá de la actividad de la proteína informadora celular y un aumento en la supervivencia celular indicará que el compuesto de prueba es un antagonista funcionalmente activo.

Como ejemplo adicional de una proteína informadora que proporciona un fenotipo observable, la proteína informadora puede ser una proteína informadora fluorescente. En estos casos, la unión de la proteína de unión al ADN al sitio de unión en ausencia de un compuesto de prueba inhibirá la señal fluorescente. Cuando se añade un compuesto de prueba que es un antagonista funcionalmente activo de la proteína de unión al ADN, esto dará como resultado un aumento de la señal fluorescente y, por lo tanto, dicho método utiliza la fluorescencia como indicador de que el compuesto de prueba es un antagonista funcionalmente activo. Las células que expresan fluorescencia podrían clasificarse (por ejemplo, mediante clasificación de células activadas por fluorescencia, FACS) para clasificar las células según la fluorescencia y, por lo tanto, el o los compuestos de prueba más eficaces, donde la célula con el nivel más alto de fluorescencia indica el compuesto de prueba más eficaz.

Por lo tanto, en algunas realizaciones, la proteína informadora es una proteína informadora fluorescente, tales como la proteína verde fluorescente (GFP), proteína amarilla fluorescente (YFP), mNeonGreen, proteína fluorescente mCherry o Kusabira-Green (mKG).

En algunas realizaciones, la proteína informadora es una proteína bioluminiscente, tal como una enzima luciferasa. Estas proteínas funcionan de manera similar a las proteínas fluorescentes, excepto que en lugar de requerir una fuente de luz externa, requieren la adición de luciferina. Las células que expresan bioluminiscencia se pueden clasificar, por ejemplo, mediante FACS, para clasificar las células de una manera similar a la descrita anteriormente para la fluorescencia.

En otro ejemplo de una proteína informadora que proporciona un fenotipo observable, la proteína informadora puede ser una enzima que actúa sobre un sustrato para producir una señal colorimétrica. En estos casos, la unión de la proteína de unión al ADN al sitio de unión en ausencia de un compuesto de prueba inhibirá la señal colorimétrica. Cuando se añade un compuesto de prueba que es un antagonista funcionalmente activo de la proteína de unión al ADN, esto dará como resultado un aumento en la señal colorimétrica y, por lo tanto, dicho método utiliza la señal colorimétrica como indicador de que el compuesto de prueba es un antagonista funcionalmente activo.

Por lo tanto, en algunas realizaciones, la proteína informadora es una enzima que actúa sobre un sustrato para producir una señal colorimétrica. Por ejemplo, la enzima puede ser peroxidasa de rábano picante o beta-galactosidasa.

Otro ejemplo de proteína informadora es una proteína cinasa, tal como la cinasa de adhesión focal (FAK). FAK es una tirosina cinasa que se compone de dominios distintos que están fosforilados. La fosforilación puede detectarse mediante, por ejemplo, lisado e inmunotransferencia del lisado celular. Se describe un método para sondear las interacciones proteína-proteína utilizando FAK, por ejemplo, en Ma *et al.* (2014).

Por lo tanto, en algunas realizaciones, la proteína informadora es una proteína cinasa, tal como FAK.

Otro ejemplo de proteína informadora es una proteasa, tal como la proteasa del virus del grabado del tabaco (TEV). El TEV es una cisteína proteasa viral altamente específica y se puede aplicar para analizar las PPI utilizando un abordaje modular de varias proteínas informadoras, incluidas las fluorescentes y luminiscentes "silenciosas" que

requieren proteólisis para activarse.

Por lo tanto, en algunas realizaciones, la proteína informadora es una proteasa, tales como TEV, usada en combinación con una proteína informadora fluorescente o luminiscente silenciosa que requiere proteólisis para activarse. Se describe un método para monitorizar las PPI usando TEV, por ejemplo, en Wehr *et al.* (2006).

En algunas realizaciones, la proteína informadora es un factor de transcripción, tal como un activador de la transcripción. Ejemplos de activadores de la transcripción incluyen GAL4, que es bien conocido por su uso en sistemas "de dos híbridos" para estudiar las PPI. Véase, por ejemplo, Young, 1998. Un activador de la transcripción se une a una secuencia de ADN provocando la activación de un gen informador en 3'. Por ejemplo, GAL4 se une al UAS e impulsa la transcripción del gen informador en 3'. El gen informador en 3' puede codificar cualquiera de las proteínas informadoras descritas anteriormente, por ejemplo, puede codificar una proteína de supervivencia celular o puede codificar una proteína fluorescente. Por lo tanto, la expresión del activador de la transcripción se puede medir indirectamente midiendo la expresión de la proteína codificada por el gen informador en 3'.

En algunas realizaciones, la proteína informadora no es una proteína informadora dividida. Las proteínas informadoras divididas están formadas por una proteína informadora funcional que se ha dividido en dos o más fragmentos inactivos, es decir, los fragmentos inactivos no proporcionan una lectura fenotípica a menos que se vuelvan a ensamblar. Por lo tanto, en algunas realizaciones, la proteína informadora es capaz de proporcionar una lectura fenotípica sin requerir reensamblaje con otra proteína o péptido.

Casetes de expresión

En la presente memoria descriptiva, la expresión "casete de expresión" pretende significar una secuencia de polinucleótidos de ADN que es capaz de efectuar la transcripción de un producto de expresión. El casete de expresión puede derivar de un gen eucariota o de un gen procariota. Un gen eucariota normalmente comprende, de 5' a 3', un promotor, una región no traducida en 5' (UTR), un marco de lectura abierto formado por exones e intrones, una UTR en 3' y puede comprender además uno o más potenciadores y/o silenciadores. Es bien sabido que los promotores son regiones del ADN responsables del inicio de la transcripción. Es bien sabido que los potenciadores son regiones del ADN que pueden unirse mediante proteínas activadoras para aumentar la probabilidad de que progrese la transcripción. Es bien sabido que los silenciadores son regiones del ADN que pueden unirse a proteínas represoras para disminuir la probabilidad de que progrese la transcripción.

Durante la transcripción en células eucariotas, el gen eucariota normalmente se transcribe primero en pre-ARNm en el núcleo de las células, que contiene las UTR 5' y 3', y los exones e intrones que conforman el marco de lectura abierto. Después de esto, el pre-ARNm se procesa en ARNm, que implica la adición de un capuchón en 5' al comienzo del ARN, la adición de una cola de poli-A al final del ARN y la eliminación de intrones. Después, el ARNm maduro final puede salir del núcleo y traducirse en una proteína.

Un gen procariota tiene una estructura similar, excepto que no contiene intrones dentro del marco de lectura abierto. Esto significa que el transcrito del ARN de un gen procariota está listo para actuar como un ARNm maduro y no requiere el procesamiento que se presenta en las células eucariotas. La transcripción del ARNm de un operón suele estar controlada por un represor que se une a un segmento de ADN conocido como operador. Por ejemplo, el operón Lac codifica una proteína represora, que está bajo regulación alostérica. En la célula procariota, la proteína represora normalmente se une al operador, que impide la transcripción del marco de lectura abierto. Sin embargo, cuando el represor se une a la molécula efectora lactosa o al análogo estructural isopropil β -D-1-tiogalactopiranosido (IPTG), el represor no se unirá al operador, lo que permite que se produzca la transcripción. De esta forma, el inicio de la transcripción depende de la disponibilidad de lactosa o IPTG dentro de la célula procariota.

Se pretende que una "secuencia codificante" signifique una porción de la secuencia de ADN de un gen que codifica el producto de expresión. Cuando el producto de expresión es una proteína, esta secuencia puede denominarse "secuencia codificante de proteínas". La secuencia codificante de proteínas normalmente comienza en el extremo 5' con un codón de iniciación y termina en el extremo 3' con un codón de terminación. Asimismo, la secuencia codificante de la proteína normalmente es la secuencia del exón(es) del gen que en un gen está flanqueado por las UTR 5' y 3'. Un ejemplo de una secuencia codificante de proteína se expone en la **SEQ ID NO: 4**.

Habitualmente, el casete de expresión comprende un promotor unido operativamente a una secuencia codificante de proteína. La expresión "unido operativamente" utilizada en el presente documento incluye la situación en la que una secuencia codificante y un promotor seleccionados están unidos de manera covalente de tal manera que colocan la expresión de la secuencia codificante de la proteína bajo la influencia o el control del promotor. Por tanto, un promotor está unido operativamente a la secuencia codificante de la proteína si el promotor es capaz de efectuar la transcripción de la secuencia codificante de la proteína. Cuando sea apropiado, el transcrito resultante puede traducirse en una proteína deseada. En algunas realizaciones, el casete de expresión puede comprender además componentes adicionales de un gen eucariota o procariota, tales como uno o más seleccionados entre una lista que consiste en: un intrón, un potenciador, un silenciador, una UTR en 5', una UTR en 3' y un regulador.

Se puede utilizar cualquier promotor adecuado conocido en la técnica en el casete de expresión siempre que funcione en el tipo de célula que se utiliza. Por ejemplo, cuando la célula es una célula bacteriana, la expresión puede estar bajo el control del operón lac. En tales casos, la célula también puede contener una proteína represora de lac, mediante lo cual la expresión puede controlarse mediante la introducción de isopropil β -D-1-tiogalactopiranosido (IPTG). El

- 5 promotor puede ser endógeno a la célula en la que se lleva a cabo el método. Cuando se utilizan múltiples casetes de expresión, cada secuencia codificante puede estar unida operativamente de forma independiente a su propio promotor. Como alternativa, la secuencia codificante para uno o más de los casetes de expresión puede estar unida operativamente al mismo promotor.
- 10 Como ya se ha descrito en el presente documento, el producto de expresión de gen informador está codificado por un casete de expresión de gen informador. La proteína de unión al ADN también puede estar codificada por un casete de expresión, denominado en el presente documento "casete de expresión de proteína de unión a ADN". Donde la proteína de unión al ADN es un complejo proteico, tal como un heterodímero, cada componente del complejo puede estar codificado en el mismo casete de expresión o en casetes de expresión separados. El compuesto de prueba
- 15 también puede ser un péptido o polipéptido que se expresa intracelularmente a partir de un casete de expresión, denominado en el presente documento "casete de expresión del compuesto de prueba".

Los casetes de expresión descritos en el presente documento pueden ser parte de uno o más vectores de expresión. Un "vector de expresión" como se utiliza en el presente documento es una molécula de ADN utilizada para la expresión de material genético extraño en una célula. Se puede utilizar cualquier vector adecuado conocido en la técnica. Los

- 20 vectores adecuados incluyen plásmidos, vectores binarios, vectores víricos y cromosomas artificiales (por ejemplo, cromosomas artificiales de levadura). Como alternativa, los casetes de expresión descritos en el presente documento pueden incorporarse al genoma de la célula.
- 25 Los métodos descritos en el presente documento pueden comprender administrar uno o más casetes de expresión descritos en el presente documento a la célula. Por ejemplo, el método puede comprender administrar un casete de expresión de gen informador, casete de expresión de la proteína de unión al ADN y/o un casete de expresión del compuesto de prueba a la célula, opcionalmente cuando el(los) casete(s) de expresión son parte de uno o más vectores de expresión. Las técnicas de biología molecular adecuadas para administrar casetes de expresión y producir
- 30 proteínas tales como la proteína de unión al ADN y la proteína informadora descritas en el presente documento en células son bien conocidas en la técnica, tales como las expuestas en Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Nueva York: Cold Spring Harbor Press, 1989.

Casete de expresión de gen informador y sitio(s) de unión

Como se ha descrito anteriormente, el casete de expresión de gen informador comprende al menos un sitio de unión de modo que la unión de la proteína de unión a ADN al sitio de unión sea capaz de inhibir la expresión del producto de expresión de gen informador. Preferentemente, el producto de expresión comprende una pluralidad de tales sitios de unión.

- 40 El al menos un sitio de unión puede estar ubicado en cualquier lugar del casete de expresión de gen informador, de modo que proporciona la unión de la proteína de unión al ADN al menos a un sitio de unión es capaz de inhibir la expresión del producto de expresión de gen informador. Por ejemplo, el(los) sitio(s) de unión pueden estar ubicados en un promotor, secuencia codificante de proteínas, potenciador, silenciador, UTR en 5', UTR en 3', regulador, exón y/o intrón. La unión de la proteína de unión al ADN a los sitios de unión puede inhibir la expresión del producto de expresión de gen informador en menos del 50 %, menos del 40 %, menos del 30 %, menos del 20 %, menos del 10 % o menos del 5 % de la expresión del producto de expresión de gen informador cuando la célula comprende el casete de expresión de gen informador sin la proteína de unión al ADN.

- 50 Por lo tanto, en algunas realizaciones, el casete de expresión de gen informador comprende al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14 o al menos 15 sitios de unión. Preferentemente, el casete de expresión de gen informador comprende al menos 2, más preferentemente al menos 5, incluso más preferentemente al menos 10, aún más preferentemente al menos 12, aún más preferentemente al menos 15 sitios de unión. En algunas realizaciones, el
- 55 casete de expresión de gen informador comprende: entre 1 y 20, entre 1 y 18, entre 1 y 15, entre 1 y 10, entre 1 y 5, entre 2 y 20, entre 2 y 18, entre 2 y 15, entre 2 y 10, entre 2 y 5, entre 5 y 18, entre 5 y 10, entre 10 y 18 o entre 12 y 16 sitios de unión. En algunas realizaciones, el casete de expresión de gen informador comprende hasta 5, hasta 10, hasta 15, hasta 18, hasta 20 sitios de unión. En una realización ilustrada, el casete de expresión de gen informador comprende 15 sitios de unión.

- 60 Algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador, por ejemplo, en la secuencia codificante del casete de expresión de gen informador. Preferentemente, el casete de expresión de gen informador comprende una pluralidad de sitios de unión que están ubicados en la secuencia transcrita o secuencia codificante. Sin desear quedar ligados a teoría alguna, se cree que una pluralidad de
- 65 sitios de unión ubicados en la región transcrita o secuencia codificante aumentarán la probabilidad de que la unión de la proteína de unión al ADN a los sitios de unión inhiba eficazmente la expresión del producto de expresión de gen

informador.

En realizaciones en las que el producto de expresión de gen informador es una proteína informadora, es preferible que el producto de expresión sea funcional para determinar si la expresión de la proteína informadora aumenta en presencia del compuesto de prueba del método de detección. En realizaciones preferidas, la presencia del sitio o sitios de unión en el casete de expresión de gen informador no afecta sustancialmente a la función de la proteína informadora. Por ejemplo, la proteína informadora puede retener al menos un 50 %, al menos un 70 %, al menos un 90 % o al menos un 95 % de la función de una proteína informadora original, en donde la proteína informadora original está codificada por un casete de expresión de gen informador original que corresponde al casete de expresión de gen informador pero que no comprende el(los) sitio(s) de unión.

Para preservar la actividad de la proteína informadora, al menos algunos del o los sitios de unión, preferentemente la mayoría del o los sitios de unión pueden introducirse en la secuencia codificante de proteínas del casete de expresión de gen informador como mutaciones silenciosas, semiconservadoras y/o conservadoras. La secuencia codificante de la proteína está formada por una serie de codones, cada uno de los cuales codifica un aminoácido específico o una señal de terminación cuando la secuencia codificante de la proteína se transcribe y traduce. Las mutaciones silenciosas son mutaciones en un codón de la secuencia codificante de la proteína que no afectan al resto de aminoácido resultante del codón. Por ejemplo, el codón GCA codifica el aminoácido Alanina (A). La mutación del codón GCA a GCG se consideraría una mutación silenciosa ya que el codón GCG todavía codifica el aminoácido Alanina (A).

Una mutación conservadora o semiconservadora es un cambio en un codón determinado que conduce a la sustitución de un aminoácido por uno bioquímicamente similar, por ejemplo, tal como se expone en la siguiente tabla.

Hidrófobo (apolar)	Alquilo	G A V L I M P
	Aromático	F Y W
Hidrófilo (polar)	Neutro	S T C Q N
	Ácido	E D
	Básico	K H R

Por ejemplo, un cambio en un codón determinado que reemplaza un aminoácido hidrófobo por otro aminoácido hidrófobo o un aminoácido hidrófilo por otro aminoácido hidrófilo, puede considerarse una mutación semiconservadora. Por ejemplo, un cambio en un codón determinado que reemplaza una serina (S) por ácido aspártico (D) puede considerarse una mutación semiconservadora. Un cambio en un codón determinado que reemplaza un alquilaminoácido por otro alquilaminoácido o un aminoácido aromático por otro aminoácido aromático, o un aminoácido neutro por otro aminoácido neutro, o un aminoácido ácido por otro aminoácido ácido o un aminoácido básico por otro aminoácido básico, puede considerarse una mutación conservadora. Por ejemplo, un cambio en un codón determinado que reemplaza un aminoácido hidrófilo neutro por otro aminoácido hidrófilo neutro (por ejemplo, treonina (T) a glutamina (Q)) puede considerarse una mutación conservadora.

Por lo tanto, la proteína informadora puede tener una secuencia de aminoácidos que tiene al menos un 80 %, al menos un 85 %, al menos un 90 % o al menos un 95 % idéntico a una proteína informadora original, en donde la proteína informadora original está codificada por un casete de expresión de gen informador original que corresponde al casete de expresión de gen informador pero que no comprende el(los) sitio(s) de unión.

En algunas realizaciones, la mayoría de las diferencias en la secuencia de aminoácidos de la proteína informadora y la secuencia de aminoácidos de la proteína informadora original son sustituciones conservadoras y/o semiconservadoras. En estos casos, se espera que la proteína informadora tenga sustancialmente la misma función que la proteína informadora original.

La ubicación del(los) sitio(s) de unión en el casete de expresión de gen informadora se puede seleccionar para evitar afectar a la función de la proteína informadora. Por ejemplo, el(los) sitio(s) de unión pueden estar ubicados en una posición en la secuencia codificante de la proteína que no codifica un resto que forma parte, o está muy cerca del centro catalítico (sitio activo) de la proteína informadora, o forma parte, o está muy cerca de un resto implicado en la unión del cofactor (por ejemplo, NADH, NDDPH). Proximidad estrecha puede significar que el resto está alejado menos de 15 Å, más preferentemente menos de 10 Å; incluso más preferentemente menos de 5 Å de un resto que forma parte del centro catalítico y/o está implicado en la unión del cofactor. De forma alternativa o adicional, proximidad estrecha puede significar que el resto está alejado menos de 5 restos, más preferentemente menos de 4 restos, incluso más preferentemente menos de 3 restos, incluso más preferentemente menos de 2 restos de un resto que forma parte del centro catalítico y/o está implicado en la unión del cofactor, cuando se evalúa en una secuencia lineal de aminoácidos.

Se espera que los cambios fuera del centro catalítico de la proteína informadora minimicen las alteraciones funcionales. De forma alternativa o adicional, el(los) sitio(s) de unión pueden estar ubicados en una posición en la

secuencia codificante de la proteína que codifica un resto expuesto al disolvente en la proteína informadora. Se espera que los cambios realizados en las regiones expuestas al disolvente de la proteína informadora minimicen las perturbaciones estructurales y, por lo tanto, minimicen la perturbación de la función general.

- 5 Se conocen métodos para identificar las regiones expuestas al disolvente de la proteína informadora. Por ejemplo, es posible tomar los archivos de coordenadas para la proteína informadora, por ejemplo, un archivo de banco de datos de proteínas (PDB) y utilizar un programa que calcula el área de superficie accesible (ASA) que informa al usuario qué tan expuestos/enterrados están los restos dentro de una estructura. Se puede encontrar un programa ASA ilustrativo en <http://cib.cf.ocha.ac.jp/bitool/ASA/>. Se puede usar un valor de corte ilustrativo de 20 Å², de manera que los restos inferiores a este se consideran enterrados y los superiores se consideran expuestos. De esta forma, se pueden identificar las ubicaciones de los restos expuestos al disolvente y modificar los codones en consecuencia.

En algunas realizaciones, el casete de expresión de gen informador codifica una proteína informadora que es una proteína de fusión, en donde la proteína de fusión comprende dos o más de las proteínas de supervivencia celular, proteínas de reproducción celular, proteínas fluorescentes, proteínas de bioluminiscencia, enzimas que actúan sobre un sustrato para producir una señal colorimétrica, proteínas cinasas, proteasas, factores de transcripción y proteínas reguladoras que se describen en el presente documento. Por ejemplo, el casete de expresión de gen informador puede codificar una proteína de fusión que comprende una proteína de supervivencia celular como se describe en el presente documento y una proteína de fluorescencia como se describe en el presente documento. En un ejemplo tal, el(los) sitio(s) de unión pueden estar ubicados en la parte del casete de expresión de gen informador (por ejemplo, la secuencia codificante) que codifica la proteína de supervivencia celular. Por lo tanto, este casete de expresión de gen informador ilustrativo proporcionaría dos lecturas de eficacia, a saber, supervivencia celular y fluorescencia. En un ejemplo particular, la proteína de fusión puede comprender una DHFR como proteína de supervivencia celular y mNeonGreen como proteína de fluorescencia.

25 Proteína de unión al ADN

La proteína de unión al ADN utilizada en los métodos de la invención puede ser o puede contener un factor de transcripción que se une a un sitio de unión (también conocido como sitio de reconocimiento) de una manera específica de secuencia, o un fragmento de unión al ADN del mismo. Se espera que cualquier factor de transcripción, o fragmento de unión al ADN del mismo, que sea capaz de unirse a un sitio de unión de una manera específica de secuencia se puede utilizar con los métodos descritos en el presente documento. En realizaciones particulares, el factor de transcripción es un factor de transcripción eucariota, tal como un factor de transcripción humano. Como se describe en la sección Antecedentes, se sabe que numerosos factores de transcripción tienen su regulación alterada en enfermedades tales como el cáncer. En algunas realizaciones, se sabe o se sospecha que la regulación del factor de transcripción humano está alterada en enfermedades tales como el cáncer. En consecuencia, cabe esperar que la identificación de antagonistas funcionalmente activos de factores de transcripción, o fragmentos de unión al ADN de los mismos, sea útil en la identificación de antagonistas que puedan ser útiles en el tratamiento del cáncer.

40 La proteína de unión al ADN puede ser un complejo proteico que comprende una o más cadenas polipeptídicas. Por ejemplo, puede ser un complejo proteico dimérico, que puede ser un homodímero o un heterodímero.

La proteína de unión al ADN puede ser cualquiera de los factores de transcripción humanos descritos en Vaquerizas *et al.* (2009) (por ejemplo, cualquiera de los enumerados en la información complementaria S3), o un fragmento de unión al ADN de los mismos. Por ejemplo, la proteína de unión al ADN puede ser un miembro de la familia de dedos de zinc C2H2, la familia del homeodominio o la familia hélice-bucle-hélice o un fragmento de unión al ADN de la misma.

En algunas realizaciones, la proteína de unión al ADN es una cremallera de leucina básica (bZIP), factor de transcripción hélice-bucle básico (bHLH) o cremallera de leucina bHLH (bHLH-Zip) o un fragmento de unión al ADN del mismo. Los factores de transcripción bHLH y bHLH-Zip son proteínas exclusivamente eucariotas que se unen a ADN bicatenario específico de secuencia como homodímeros o heterodímeros para activar o reprimir la transcripción génica. Los factores de transcripción bZIP forman una de las familias más grandes de factores de transcripción en células eucariotas y contienen una región básica que contacta las bases del ADN para unirse a su sitio de unión al ADN. Los factores de transcripción bZIP también contienen repeticiones de leucina a través de las cuales las proteínas se dimerizan para formar una hélice superenrollada, que es necesaria para la actividad funcional de la proteína. Además de las proteínas humanas, ciertas proteínas víricas, tales como BZLF1, forman parte de la familia bZIP. En Vinson *et al.* (2002), Newman y Keating (2003) and Rodríguez-Martínez *et al.* (2017) se proporcionan detalles de los factores de transcripción bZIP, bHLH and bHLH-ZIP y sus secuencias consenso. Habitualmente, el fragmento de unión al ADN de estos factores de transcripción será (o comprenderá) las porciones básicas de los factores de transcripción que interactúan físicamente con el ADN.

El factor de transcripción bHLH ilustrativo incluye ATOH1, AhR, AHRR, ARNT, ASCL1, BHLH2, BHLH3, BHLH9, ARNTL, ARNTL2, CLOCK, EPAS1, FIGLA, HAND1, HAND2, HES5, HES6, HEY1, HEY2, HEYL, HES1, HIF1A, HIF3A, ID1, ID2, ID3, ID4, LYL1, MESP2, MXD4, MYCL1, MYCN, MyoD, Miogenina, MYF5, MYF6, Neurogenin1, Neurogenin2, Neurogenin3, NeuroD1, NeuroD2, NPAS1, NPAS2, NPAS3, OLIG1, OLIG2, Phox, Scleraxis, SIM1, SIM2, TAL1, TAL2, Twist y USF1. Los factores de transcripción bHLH-ZIP ilustrativos incluyen AP-4, Max, MXD1,

MXD3, MITF, MNT, MLX, MLXIPL, MXI1, Myc, SREBP1 y SREBP2. En realizaciones particulares, el factor de transcripción bHLH-ZIP utilizado puede ser c-Myc o Max, o un heterodímero entre c-Myc y Max (c-Myc-Max). Los factores de transcripción bHLH y bHLH-ZIP normalmente se unen a una secuencia consenso llamada caja E, que puede tener la secuencia CANNTG (siendo "N" cualquier nucleótido) y en casos particulares tiene la secuencia CACGTG. La proteína de unión a ADN puede ser o puede comprender cualquiera de estos factores de transcripción bHLH o bHLH-Zip y el casete de expresión de gen informador comprende al menos una caja E como sitio de unión, en donde la caja E puede tener la secuencia CANNTG, por ejemplo, CACGTG.

Subfamilias ilustrativas del factor de transcripción bZIP humano, las secuencias de nucleótidos de sus sitios de unión y ejemplos de proteínas de estas subfamilias se exponen en la siguiente tabla. La proteína de unión al ADN puede ser, o puede comprender, cualquiera de estas proteínas bZIP humanas o un fragmento de unión a ADN de las mismas, y al menos un sitio de unión puede ser la secuencia de nucleótidos del sitio de unión expuesto en la misma fila que la proteína bZIP en la siguiente tabla. Por ejemplo, la proteína de unión al ADN puede ser una proteína de la subfamilia Fos/Jun (por ejemplo, cJun) y al menos un sitio de unión puede tener la secuencia de nucleótidos TGACTCA o TGAGTCA.

Subfamilia bZIP humana	Proteína bZIP ilustrativa	Secuencia(s) de nucleótidos del sitio de unión	Nombre del sitio de unión
PAP	PAP1, YAP1, YAP2, YAP3, YAP4, YAP5, YAP6, YAP7, Cap1	TTACGTAA	PAP/CREB-2/PAR
CREB-2	AFT4, mATFP4, ApCREB-2, hCREB2, acr1		
PAR	DBP, VBP/TEF, HLF, CES2, TEF		
C/EBP	C/EBP α , C/EBP β , C/EBP δ , C/EBP ϵ , C/EBP γ , CRP1, CRP2, CRP3, Ig/EBP, lap, DDIT3	ATTGCGCAAT	CCAAT
Fos/Jun	cFos, FRA1 (FosL1), FRA2 (FosL2), cJun, JUNB, JUND, GCN4, BATF, BATF2, BATF3	TGACTCA o TGAGTCA	Elemento de respuesta a TPA (TRE)
CREB	CREB1, ATF1, ATF2, ATF3, ARF5, ATFa, BBF-2, CREB3L1	TGACGTCA	elemento de respuesta a AMPc (CRE)
Maf	MafA, MafB, BACH1, BACH2	TGCTGA(G/C)TCAGCA y TGCTGAG (C/C) GTCAG CA	Elemento de reconocimiento de Maf (MARE)

AP-1 es un dímero, normalmente un heterodímero, que está compuesto por proteínas pertenecientes a la subfamilia Fos/Jun (por ejemplo, cFos, FRA1, FRA2, cJun, JUNB, JUND, GCN4, BATF, BATF2, BATF3).

Además de los factores de transcripción bZIP humanos, ciertas proteínas víricas que se unen al ADN también pertenecen a la familia bZIP. Esto incluye el transactivador bZIP del virus de Epstein-Barr, BZLF1. BZLF1 puede unirse al sitio de unión a TRE (TGACTCA o TGAGTCA) o al sitio de unión CCAAT (ATTGCGCAAT). La proteína de unión al ADN puede ser, o puede comprender, BZLF1 o un fragmento de unión a ADN del mismo, y al menos un sitio de unión puede ser un sitio de unión a TRE o un sitio de unión a CCAAT.

La proteína de unión al ADN puede ser alternativamente un factor de transcripción o un fragmento de unión al ADN del mismo que no forma parte de las familias bZIP, bHLH o bHLH-ZIP. Ejemplos de factores de transcripción eucariotas adecuados adicionales que pueden ser, comprender o derivar de la proteína de unión al ADN se exponen en la siguiente tabla, junto con las secuencias de nucleótidos de sus sitios de unión al ADN y los nombres de estos sitios de unión. La proteína de unión al ADN puede ser, o puede comprender, cualquiera de estos factores de transcripción eucariotas (por ejemplo, humanos) o un fragmento de unión al ADN de los mismos, y al menos un sitio de unión puede ser la secuencia de nucleótidos del sitio de unión expuesto en la misma fila que el factor de transcripción en la tabla siguiente.

Factor(es) de transcripción eucariotas	Nombre del sitio de unión	Secuencia(s) de nucleótidos del sitio de unión
Factor de unión a la caja CAAT *	Caja CAAT	GGCCAATCT
Factor de respuesta sérica *	Caja CArG	CC(A/T) ₆ GG
Proteínas Snail (por ejemplo, SNAI1) *	Caja E2	CAGGTG y CACCTG
Runx2*	Caja HY	TG(A/T)GGG
Factores de transcripción de la caja T *	Caja T	TCACACCT

ES 2 985 624 T3

ARN polimerasa en eucariotas *	Caja TATA	TATAAA
Proteínas RFX (por ejemplo, RFX1) *	Caja X	GTTGGCATGGCAAC
Proteína de unión a la caja Y *	Caja Y	(A/G)CTAACC(A/G)(A/G)(C/T)

(continuación)

Factor(es) de transcripción eucariotas	Nombre del sitio de unión	Secuencia(s) de nucleótidos del sitio de unión
Proteínas de unión a elementos respondedores a etileno	Caja ATA	AAATAT
AtSR1 (genes respondedor a señales de <i>Arabidopsis thaliana</i>)	Caja CGCG	(A/C/G)CGCG(C/G/T)
Proteínas similares a las de unión a elementos respondedores a la deshidratación (DREB)	Caja DREB	TACCGACAT
Proteína Fur	Caja Fur	GATAATGATAATCATTATC
EmBP1	Caja G	GCCACGTGGC
Proteínas similares a EREBP	Caja GCC	AGCCGCC
Proteína KAP-2	Caja H	ACACCA
factor de unión a la caja de prolamina de cebada (caja P)	Caja de prolamina	TGTAAAG
Proteínas aleurona	Caja de pirimidina	CCTTTT
U2 snRNP	Caja TACTAAC	ATTTACTAAC
* Factores de transcripción eucariotas que también son factores de transcripción humanos.		

Por lo tanto, en algunas realizaciones,

- 5 a) el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**;
- b) el al menos un sitio de unión es un elemento de respuesta a la caja E que tiene la secuencia de nucleótidos **CACGTG (SEQ ID NO: 7)** o **CACATG (SEQ ID NO: 8)**;
- 10 c) el al menos un sitio de unión es un sitio de unión CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**;
- d) el al menos un sitio de unión es un elemento de respuesta a AMPc (CRE) que tiene la secuencia de nucleótidos **TGACGTCA (SEQ ID NO: 10)**;
- 15 e) el al menos un sitio de unión es un elemento de reconocimiento de Maf (MARE) que tiene la secuencia de nucleótidos **TGCTGA^g/cTCAGCA (SEQ ID NO: 32)** o **TGCTGA^{gc}/ccTCAGCA (SEQ ID NO: 33)**; o
- 20 f) el al menos un sitio de unión es un sitio de unión a PAP/CREB-2/PAR que tiene la secuencia de nucleótidos **TTACGTAA (SEQ ID NO: 34)**.

En realizaciones más específicas,

- 25 a) el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**;
- b) el al menos un sitio de unión es un elemento de respuesta a la caja E que tiene la secuencia de nucleótidos **CACGTG (SEQ ID NO: 7)** o **CACATG (SEQ ID NO: 8)**; o
- 30 c) el al menos un sitio de unión es un sitio de unión CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**.

35 En realizaciones aún más específicas, el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**.

En realizaciones particulares,

- 40 a) la proteína de unión al ADN es AP-1 o un miembro de la subfamilia Fos/Jun de factores de transcripción (tal como c-Jun), o un fragmento de unión al ADN de la misma y al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**;
- b) la proteína de unión al ADN es un factor de transcripción bHLH, tal como c-Myc o Max, o un fragmento de unión al ADN del mismo y al menos un sitio de unión es un elemento de respuesta a la caja E que tiene la secuencia de nucleótidos **CACGTG (SEQ ID NO: 7)** o **CACATG (SEQ ID NO: 8)**;
- 45

c) la proteína de unión al ADN es un miembro de la subfamilia C/EBP de factores de transcripción (tal como C/EBP alfa), o un fragmento de unión al ADN de la misma, y al menos un sitio de unión es un sitio de unión CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**;

d) la proteína de unión al ADN es BZLF1, o un fragmento de unión al ADN de la misma, y el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**, o un sitio de unión a CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**;

e) la proteína de unión al ADN es un miembro de la subfamilia CREB de factores de transcripción (tal como CRE), o un fragmento de unión al ADN de la misma, y el al menos un sitio de unión es un elemento de respuesta al AMPc (CRE) que tiene la secuencia de nucleótidos **TGACGTCA (SEQ ID NO: 10)**;

f) la proteína de unión al ADN es un factor de transcripción Maf, o un fragmento de unión a ADN del mismo, y el al menos un sitio de unión es un elemento de reconocimiento de Maf (MARE) que tiene la secuencia de nucleótidos **TGCTGA^{GC}/cTCAGCA (SEQ ID NO: 32)** o **TGCTGA^{GC}/ccTCAGCA (SEQ ID NO: 33)**;

g) la proteína de unión al ADN es un miembro de la subfamilia de factores de transcripción poli(ADP-ribosa) (PAR), o un fragmento de unión al ADN de la misma, y al menos un sitio de unión es un sitio de unión a PAP/CREB-2/PAR que tiene la secuencia de nucleótidos **TTACGTAA (SEQ ID NO: 34)**; o

h) la proteína de unión al ADN es un miembro de la subfamilia CREB-2 de factores de transcripción, o un fragmento de unión al ADN de la misma, y al menos un sitio de unión es un sitio de unión a PAP/CREB-2/PAR que tiene la secuencia de nucleótidos **TTACGTAA (SEQ ID NO: 34)**.

En realizaciones más específicas,

a) la proteína de unión al ADN es AP-1 o un miembro de la subfamilia Fos/Jun de factores de transcripción (tal como c-Jun), o un fragmento de unión al ADN de la misma y al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**;

b) la proteína de unión al ADN es un factor de transcripción bHLH, tal como c-Myc o Max, o un fragmento de unión al ADN del mismo y al menos un sitio de unión es un elemento de respuesta a la caja E que tiene la secuencia de nucleótidos **CACGTG (SEQ ID NO: 7)** o **CACATG (SEQ ID NO: 8)**;

c) la proteína de unión al ADN es un miembro de la subfamilia C/EBP de factores de transcripción (tal como C/EBP alfa), o un fragmento de unión al ADN de la misma, y al menos un sitio de unión es un sitio de unión CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**; o

d) la proteína de unión al ADN es BZLF1, o un fragmento de unión al ADN de la misma, y el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**, o un sitio de unión a CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**.

En realizaciones particulares, la proteína de unión al ADN es AP-1, una proteína de la subfamilia Fos/Jun (por ejemplo, c-Jun), o un fragmento de unión al ADN de la misma, y al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**. Un ejemplo de un fragmento de unión al ADN de AP-1 es el motivo básico de cualquiera de las proteínas que son miembros de la subfamilia Fos/Jun, por ejemplo, un fragmento que contiene el motivo básico de c-Jun, como se usa en los ejemplos. El motivo básico de c-Jun puede ser la secuencia expuesta en las posiciones 252-279 del acceso UniProt P05412, versión 2. Como se describe en otra parte, los factores de transcripción bZIP se unen como un homodímero o heterodímero y, por lo tanto, el factor de transcripción bZIP puede contener un dominio de dimerización, tal como un dominio de cremallera de leucina, además de su dominio de unión al ADN.

Como se describe en los ejemplos, se generó un casete de expresión de gen informador que codifica DHFR murina como proteína informadora, donde el casete de expresión de gen informador contenía 15 TRE en su secuencia codificante de proteínas. Esta secuencia codificante de proteínas ilustrada de este casete de expresión de gen informador tiene la secuencia expuesta en la **SEQ ID NO: 4**.

Por lo tanto, en algunas realizaciones, el casete de expresión de gen informador comprende una secuencia codificante de proteínas que es al menos un 90 %, al menos un 95 %, al menos un 98 % o un 100 % idéntico a la secuencia expuesta en la **SEQ ID NO: 4** y la proteína de unión a ADN es AP-1 o un fragmento de unión a ADN de la misma.

Células

El método para detectar un antagonista funcionalmente activo de la invención funciona en células vivas aisladas, es decir, los métodos se realizan *in cellulo* a menos que el contexto indique claramente lo contrario. El término "*in cellulo*" tiene como objetivo abarcar experimentos que tienen lugar con células y pueden ser en células cultivadas o en células o tejidos que han sido obtenidos de un organismo. Los métodos de la invención no se ponen en práctica en el cuerpo humano o animal.

Puede usarse cualquier célula adecuada para la expresión de productos de expresión para el método de selección descrito en el presente documento. La célula puede ser procariota o eucariota. Normalmente, las células son células aisladas.

La célula utilizada en el método de detección puede ser una célula bacteriana. En algunas realizaciones, la célula bacteriana es una célula de *Escherichia coli*, por ejemplo, BL21 (DE3), XL-1, RV308 o DH5alpha. Los métodos de detección en los que la célula es una célula bacteriana pueden implicar el cultivo de la célula bacteriana en medios adecuados. Tales técnicas son bien conocidas por los expertos en la materia.

Como alternativa, la célula es una célula eucariota, tal como una célula de levadura, una célula vegetal, una célula de insecto o una célula de mamífero. En algunas realizaciones, la célula es una célula de mamífero, por ejemplo una célula humana. Las células de mamífero, especialmente células humanas, pueden ser células somáticas. Los métodos de detección en donde la célula es una célula eucariota pueden implicar el cultivo o la fermentación de la célula eucariota. El cultivo o la fermentación puede realizarse en un biorreactor provisto de un suministro apropiado de nutrientes, aire/oxígeno y/o factores de crecimiento. Los expertos en la materia conocen bien las técnicas de cultivo, fermentación y separación.

Como se ha descrito anteriormente, se espera que los compuestos de prueba que sean antagonistas funcionalmente activos de las proteínas de unión al ADN, en particular, los factores de transcripción cuya regulación se sabe que está alterada en los cánceres humanos, tendrán uso terapéutico en el tratamiento de enfermedades tales como cáncer. En algunas realizaciones en donde se utiliza una célula de mamífero, la proteína de unión al ADN es producida de forma natural por la célula de mamífero, es decir, la proteína de unión al ADN no se administra a la célula, por ejemplo, en forma de un vector de expresión de proteína de unión al ADN. En particular, la célula de mamífero puede ser una célula humana aislada de un paciente con cáncer en donde se sospecha o se sabe que la regulación de la proteína de unión al ADN está alterada en la célula. Se cree que esto sea ventajoso, ya que el método probará si el compuesto de prueba puede antagonizar funcionalmente la expresión natural de la proteína de unión al ADN y, por lo tanto, se espera que los resultados obtenidos reflejen lo que puede ocurrir *in vivo*.

Los métodos en donde la célula es una célula humana tienen la ventaja adicional de que, además de detectar antagonistas funcionalmente activos de la proteína de unión al ADN, el método perfilará simultáneamente el compuesto de prueba para obtener otras propiedades deseables que conduzcan al desarrollo de fármacos. Por ejemplo, se puede utilizar para determinar si el compuesto de prueba es tóxico, si es selectivo para la proteína de unión al ADN en cuestión y si el compuesto de prueba es estable en células humanas. Esto se compara favorablemente con métodos conocidos para identificar inhibidores de PPI que funcionan como compuestos terapéuticos en células humanas, donde una primera etapa sería identificar un inhibidor de la PPI, la segunda etapa para confirmar que el inhibidor de la PPI elimina la función de la proteína y luego la tercera etapa para comprobar que funciona en las células humanas. Por lo tanto, la presente invención permite ventajosamente combinar todas estas etapas individuales en una etapa de detección intracelular en células humanas.

Por ejemplo, como se describe en la sección de proteína informadora anterior, cuando la célula es una célula de mamífero, la proteína informadora puede ser DHFR de mamífero, por ejemplo, DHFR murino que ha sido modificada de tal manera que se vuelve resistente al fármaco antifolato metotrexato (MTX), por ejemplo como lo describen Remy *et al.* (2007). De esta forma, la supervivencia celular se puede utilizar como lectura para determinar si el compuesto de prueba es un antagonista funcionalmente activo de la proteína de unión al ADN en cuestión en la célula humana.

Compuesto de prueba

Los compuestos de prueba para su uso con el método de detección de la invención no están particularmente limitados. En algunas realizaciones, el compuesto de prueba es peptídico. "Peptídico", como se usa en el presente documento, incluye compuestos que están compuestos o comprenden una cadena lineal de aminoácidos unidos por enlaces peptídicos e incluyen péptidos y polipéptidos. En la presente memoria descriptiva, el término "péptido" pretende significar moléculas que consisten en entre 2 y 50 aminoácidos y el término "polipéptido" pretende significar moléculas que están formadas por más de 50 aminoácidos. En otras realizaciones, el compuesto de prueba es una molécula pequeña, sintético o de origen natural. Una molécula pequeña es un compuesto (normalmente un compuesto orgánico) que tiene un peso molecular de 500 dalton o menos.

En algunas realizaciones, el compuesto de prueba es un péptido mimético. Las expresiones "péptido mimético", "peptidomimético" y "análogo peptídico" se usan indistintamente y se refieren a un compuesto químico que no está compuesto enteramente por aminoácidos pero que tiene sustancialmente las mismas características que un compuesto peptídico que está compuesto enteramente por aminoácidos. Un péptido mimético puede ser peptídico, en

el sentido de que es una molécula quimérica que está formada tanto por aminoácidos peptídicos naturales como por análogos no naturales de aminoácidos. Como alternativa, un péptido mimético puede no ser peptídico, en cuanto a que está compuesto enteramente por análogos de aminoácidos sintéticos o no naturales. Los péptidos miméticos pueden clasificarse como se expone en Pelay-Gimeno *et al.* (2015). Brevemente, los miméticos de "clase A" corresponden a compuestos peptídicos que están formados principalmente por aminoácidos con alteraciones menores de la cadena lateral o del esqueleto; los miméticos de "clase B" corresponden a compuestos peptídicos con diversas alteraciones de la cadena principal y lateral; los miméticos de "clase C" corresponden a estructuras similares a moléculas pequeñas que proyectan sustituyentes en analogía con las cadenas laterales peptídicas; y los miméticos de "clase D" corresponden a moléculas que imitan el modo de acción de un péptido sin un enlace directo a sus cadenas laterales.

En algunas realizaciones, el compuesto de prueba es un compuesto de prueba peptídico que se expresa intracelularmente a partir de una secuencia de nucleótidos. Por ejemplo, la secuencia de nucleótidos puede ser un casete de expresión (también denominado "casete de expresión del compuesto de prueba", que puede estar contenido en un vector presente en la célula o puede incorporarse al genoma de la célula como se ha descrito anteriormente.

Se espera que el método de detección de la invención sea útil con bibliotecas peptídicas codificadas genéticamente. Se conocen bibliotecas peptídicas codificadas genéticamente y se han utilizado en métodos de detección para identificar inhibidores de proteínas de unión al ADN. Véase, por ejemplo, Mern *et al.* (2010). Brevemente, dichas bibliotecas se forman a partir de bibliotecas de casetes de expresión de compuestos de prueba, cada uno de los cuales codifica y es capaz de dirigir la expresión de un compuesto de prueba peptídico diferente. Al transformar la biblioteca en células que contienen la proteína de unión al ADN y el casete de expresión de gen informador, es posible determinar si un miembro determinado de la biblioteca puede actuar como un antagonista funcionalmente activo de la proteína de unión al ADN. Tales bibliotecas peptídicas codificadas genéticamente se pueden usar con el método de la presente invención para seleccionar rápidamente múltiples compuestos de prueba diferentes al mismo tiempo.

Así, en algunas realizaciones, la célula utilizada en el método se obtuvo de un conjunto de células que se transformaron con una biblioteca genéticamente codificada de compuestos de prueba peptídicos, de modo que la célula exprese el compuesto peptídico de prueba intracelularmente.

Los presentes inventores también han reconocido que el método de selección se puede utilizar con compuestos de prueba que se añaden extracelularmente. Por ejemplo, las células que contienen la proteína de unión al ADN y el casete de expresión de gen informador se pueden cultivar y sembrar en placas de microtitulación (por ejemplo, placas de 1536 pocillos) y se pueden cribar bibliotecas de compuestos de prueba mediante adición directa a cada pocillo. La adición de las bibliotecas de compuestos de prueba a los pocillos puede ocurrir antes o después de la adición de las células. Este método se puede utilizar para detectar rápidamente múltiples compuestos de prueba diferentes y tiene la ventaja adicional de permitir al usuario alejarse de las bibliotecas de péptidos estándar, por ejemplo, permitiendo al usuario crear perfiles de péptidos restringidos por hélice, peptidomiméticos, aminoácidos no naturales o incluso bibliotecas de moléculas pequeñas. Los compuestos de prueba que se añaden extracelularmente deben poder cruzar la membrana celular (y la pared celular, si está presente) para ingresar a la célula y examinarse para determinar si son antagonistas funcionalmente activos usando los métodos de la invención. Esto significa que el método de adición del compuesto de prueba extracelular permite al usuario perfilar la penetrancia celular concomitantemente con el antagonismo funcional de la proteína de unión al ADN, ya que un aumento en la expresión del producto de expresión de gen informador indicará que el compuesto de prueba es capaz de ingresar a la célula y capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN. Los compuestos utilizados en un entorno terapéutico en seres humanos deberán entrar en las células para tener un efecto terapéutico. Por lo tanto, sin desear quedar ligados a teoría alguna, se espera que aquellos compuestos de prueba añadidos extracelularmente que dan como resultado un aumento en la expresión del producto de expresión de gen informador usando los métodos descritos en el presente documento representen buenos candidatos para su uso como agentes terapéuticos potenciales. Asimismo, dado que la penetrancia celular y el antagonismo funcional se determinan concomitantemente, esto se compara favorablemente con los métodos que requieren ensayos separados para probar la penetrancia celular y el antagonismo funcional *in cellulo*.

En algunas realizaciones, la transcripción de la proteína de unión al ADN y el producto de expresión de gen informador en la célula puede iniciarse primero (por ejemplo, mediante la activación del promotor o promotores que están unidos operativamente al casete de expresión de la proteína de unión al ADN y al casete de expresión de gen informador), seguido de la administración o expresión del compuesto de prueba en la célula. Esto es, en estas realizaciones, la proteína de unión al ADN se transcribe, se traduce y se le permite unirse a su(s) sitio(s) de unión al ADN para formar un complejo unido al ADN en el casete de expresión de gen informador antes de que el compuesto de prueba esté presente en la célula. Sin desear quedar ligados a teoría alguna, se cree que en estas realizaciones, un aumento en la expresión del producto de expresión de gen informador después de añadir el compuesto de prueba proporciona una indicación de que el compuesto de prueba es capaz de unirse y disociar el complejo unido al ADN.

Por lo tanto, en algunas realizaciones, el método comprende administrar el compuesto de prueba extracelularmente para obtener una célula que comprende el compuesto de prueba. Por ejemplo, el compuesto de prueba se puede añadir a los medios de cultivo en los que se está cultivando la célula. En realizaciones en las que el compuesto de

prueba se administra extracelularmente, un aumento en la expresión del producto de expresión de gen informador indica que el compuesto de prueba es capaz de entrar en la célula así como de inhibir la actividad de unión al ADN de la proteína de unión al ADN.

- 5 En algunas realizaciones, el compuesto de prueba es uno que previamente se ha identificado como capaz de interactuar con la proteína de unión al ADN, o se sospecha que puede inhibir la actividad de la proteína de unión al ADN. Por ejemplo, se puede sospechar que el compuesto de prueba es un inhibidor basado en un ensayo de PCA y luego se puede usar el método descrito en el presente documento para confirmar que el inhibidor es un antagonista funcionalmente activo de la proteína de unión al ADN. Como se describe en los ejemplos, un péptido ("FosW")
10 identificado mediante un ensayo de PCA se clasificó como un antagonista funcionalmente activo según el método descrito en el presente documento.

Los restos presentes en la superficie de una proteína que son responsables de las PPI están asociados con motivos de estructura secundaria de proteínas, tales como una alfa-hélice, láminas beta y giros beta. Cabe destacar que se cree que las alfa-hélices comprenden aproximadamente el 60 % de todas las estructuras secundarias en los complejos proteicos (Jochim y Arora, 2010). Adicionalmente, se ha demostrado que las alfa-hélices median en una gran cantidad de interfaces de PPI clave desde el punto de vista terapéutico, de las cuales el 60 % se une a una cara de la hélice (Raj *et al.*, 2013). Las alfa-hélices contienen un enlace de hidrógeno entre el grupo carbonilo (C=O) de un aminoácido determinado y el grupo amino (NH) de un aminoácido a tres o cuatro restos de distancia.

Se ha informado que la restricción de péptidos en una conformación helicoidal confiere beneficios que incluyen mejora de la resistencia a las proteasas, la estabilidad en las células, aumenta la captación celular, mejora las propiedades biofísicas y se anticipa que se unirán a sus dianas con mayor potencia en comparación con las secuencias peptídicas de tipo salvaje (Azzarito *et al.* 2013). Como resultado, los péptidos que contienen alfa-hélices restringidas (también denominados "péptidos restringidos por hélice") han sido de gran interés para identificar inhibidores de PPI (Robertson y Spring, 2018).

Por lo tanto, en algunas realizaciones, el compuesto de prueba peptídico comprende un péptido restringido por hélice.

30 La expresión "péptido restringido por hélice" pretende significar un péptido que tiene al menos una modificación química que da como resultado un entrecruzamiento intramolecular entre dos aminoácidos para producir una alfa-hélice estabilizada. En general, el entrecruzamiento se extiende a lo largo de una o dos vueltas helicoidales (es decir, aproximadamente 3-3,6 o aproximadamente 7 aminoácidos). En consecuencia, los aminoácidos situados en *i* y uno de: *i*+3, *i*+4 e *i*+7 son candidatos ideales para el entrecruzamiento. Por lo tanto, por ejemplo, cuando un péptido tiene la secuencia... X1, X2, X3, X4, X5, X6, X7, X8, X9 ... , y el aminoácido X se selecciona independientemente para cada posición, los entrecruzamientos entre X1 y X4, o entre X1 y X5, o entre X1 y X8 son útiles al igual que los entrecruzamientos entre X2 y X5, o entre X2 y X6, o entre X2 y X9, etc. El uso de múltiples entrecruzamientos (por ejemplo, 2, 3, 4 o más) también se contempla.

40 La modificación química incluye una modificación química para incorporar una atadura molecular, tal como una fibra de hidrocarbano y una modificación química para promover la formación de un puente disulfuro. El entrecruzamiento puede ser un enlace iónico, covalente o de hidrógeno que une los dos restos, preferentemente el entrecruzamiento es un enlace covalente.

45 La presencia de una alfa-hélice estabilizada se puede determinar utilizando métodos tales como la espectroscopia de dicroísmo circular para una alfa-hélice, por ejemplo tal como se describe en Jo *et al.* (2012). Se puede utilizar dicroísmo circular para medir un aumento de helicidad, es decir, de lineal a cíclico. En situaciones en las que el entrecruzamiento se produce mediante la formación de un puente disulfuro entre dos grupos tiol, tal como entre dos restos de cisteína, la presencia de una alfa-hélice estabilizada también se puede determinar mediante un ensayo que determina si los tioles en la muestra están libres o conjugados. Por ejemplo, los tioles libres se pueden analizar mediante reacción con el reactivo de Ellman (ácido 5,5'-ditiobis(2-nitrobenzoico; DNTB) (Sigma)) y monitorizando la absorbancia a 412 nm.

Los métodos para inducir entrecruzamientos entre aminoácidos son bien conocidos e incluyen métodos que inducen entrecruzamientos entre la cadena principal del péptido, por ejemplo, entre el grupo carbonilo y el grupo amino como en las alfa-hélices naturales, así como entre las cadenas laterales de los péptidos. Los métodos incluyen la formación de enlaces disulfuro (por ejemplo, como se describe en Leduc *et al.* (2003)), sustitutos de enlaces de hidrógeno (por ejemplo, como se describe en Wang *et al.* (2005)), metátesis de cierre de anillo (por ejemplo, como se describe en Walensky *et al.* (2004)), alquilación de cisteína usando derivados de α -haloacetamida (por ejemplo, como se describe en Woolley (2005)) o haluros de biarilo (por ejemplo, como se describe en Muppidi *et al.* (2011)), formación de anillos lactámicos (por ejemplo, como se describe en Fujimoto *et al.* (2008)), enlace de hidracina (por ejemplo, como se describe en Cabezas y Satterthwait (1999)), enlace oxima (por ejemplo, como se describe en Haney *et al.* (2011)), quelación de metales (por ejemplo, como se describe en Ruan *et al.* (1990)), y química de "clic" (por ejemplo, como se describe en Holland-Nell y Meldal (2011)).

65 En algunas realizaciones, el entrecruzamiento se introduce entre los aminoácidos en el compuesto de prueba peptídico para producir un péptido restringido por hélice antes de administrar el compuesto de prueba a la célula, por ejemplo,

administrando el compuesto de prueba extracelularmente.

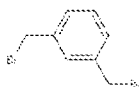
Los presentes inventores también han hecho el sorprendente descubrimiento de que es posible introducir intracelularmente el entrecruzamiento intramolecular en el compuesto de prueba. Por lo tanto, se podría utilizar un método en el que el compuesto de prueba peptídico se entrecruza durante la etapa de selección intracelular para detectar directamente péptidos restringidos por hélice dentro de la célula. Dado que el péptido restringido por hélice está presente dentro de la célula, las células se pueden usar inmediatamente para la detección posterior para determinar si el compuesto de prueba es capaz de alterar las PPI y/o si el compuesto de prueba es un antagonista funcionalmente activo usando el método de detección descrito en el presente documento. Asimismo, este método es aplicable para polipéptidos que contienen el péptido restringido por hélice, permitiendo que se críbe el péptido restringido por hélice para determinar si puede alterar las PPI en el contexto del polipéptido.

Por lo tanto, en algunas realizaciones del método de detección descrito en el presente documento donde el compuesto de prueba comprende un péptido, el método comprende además administrar un agente de entrecruzamiento en la célula, en donde el agente de entrecruzamiento modifica químicamente el péptido para introducir un entrecruzamiento entre dos restos de aminoácidos para producir una alfa-hélice estabilizada, produciendo así el compuesto de prueba que comprende el péptido restringido por hélice. El compuesto de prueba puede expresarse intracelularmente a partir de un casete de expresión del compuesto de prueba.

En algunas realizaciones, el entrecruzamiento se forma entre aminoácidos en las posiciones i e $i+3$, i e $i+4$ o i e $i+7$ en la secuencia de aminoácidos del péptido. En algunas realizaciones, el entrecruzamiento se produce entre restos de cisteína (C) ubicados en estas posiciones. En otras realizaciones, el entrecruzamiento se produce entre los restos de lisina (K) y ácido aspártico (D) en estas posiciones. Preferentemente, el entrecruzamiento se forma entre los aminoácidos en las posiciones i e $i+4$.

En algunas realizaciones, el método comprende determinar la expresión del producto de expresión de gen informador tanto antes como después de la adición del agente de entrecruzamiento. De esta forma, se puede determinar si el péptido o polipéptido es capaz de actuar como antagonista funcionalmente activo tanto antes como después del entrecruzamiento, proporcionando por lo tanto una indicación del efecto funcional que tiene la restricción de la alfa-hélice en el péptido.

En realizaciones preferidas, el péptido comprende una cisteína (C) en las posiciones i e $i+4$ en su secuencia de aminoácidos. Como se describe en Jo *et al.* (2012), la introducción de restos de cisteína en las posiciones i e $i+4$ es útil porque este espaciado acerca dos restos de tioéter cuando están en la alfa-hélice. Agentes de entrecruzamiento adecuados para estabilizar la alfa-hélice dentro del péptido que contiene una cisteína (C) en la posición i e $i+4$ se describen en Jo *et al.* (2012). Por ejemplo, el agente de entrecruzamiento podría ser un agente de entrecruzamiento seleccionado entre el grupo que consiste en un bromuro de alquilo, un yoduro de alquilo, un bromuro de bencilo, un bromuro de alilo, una maleimida y un difluorobenceno electrófilo. En realizaciones preferidas, el agente de entrecruzamiento es un bromuro de bencilo a base de *m*-xileno, a base de *o*-xileno o a base de *p*-xileno, más preferentemente un bromuro de bencilo a base de *m*-xileno. En realizaciones particularmente preferidas, el agente de entrecruzamiento es 1,3-dibromometilbenceno (DBMB) que tiene la siguiente fórmula química:



En algunas realizaciones, el péptido comprende una lisina (K) y ácido aspártico (D) en las posiciones i e $i+4$ en su secuencia de aminoácidos. Esto es, la posición i es una lisina (K) y la posición $i+4$ es un ácido aspártico (D), o la posición i es un ácido aspártico (D) y la posición $i+4$ es una lisina (K). Se describen métodos para llevar a cabo la lactamización K-D, por ejemplo, in de Araujo *et al.* (2014).

El método puede comprender añadir el agente de entrecruzamiento a un pH de entre 7,5 y 8,5, preferentemente un pH de 8,0. Esto se puede lograr utilizando varios tampones, como es bien conocido en la técnica. El método puede comprender adicionalmente tratar las células con tris(2-carboxietil)fosfina (TCEP), lo que puede ayudar a dirigir la bialquilación específica. En métodos ilustrativos concretos, el agente de entrecruzamiento DBMB se puede añadir al compuesto de prueba que comprende un péptido restringido por hélice con TCEP y bicarbonato de amonio, y se puede hacer reaccionar a pH 8,0 y temperatura ambiente durante 4 a 5 horas en la oscuridad.

Método sin células

En otro aspecto, la presente invención proporciona un método sin células para detectar un antagonista de una proteína de unión al ADN, comprendiendo el método:

- i) poner en contacto un compuesto de prueba con una proteína de unión a ADN y un casete de expresión de gen

informador que codifica un producto de expresión de gen informador, en donde el casete de expresión de gen informador comprende al menos un sitio de unión para la proteína de unión a ADN de manera que la unión de la proteína de unión a ADN al sitio de unión inhibe la expresión del producto de expresión de gen informador; y

ii) determinar la expresión del producto de expresión de gen informador;

en donde un aumento en la expresión del producto de expresión de gen informador en presencia del compuesto de prueba indica que el compuesto de prueba es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN, en donde el método se lleva a cabo fuera de una célula en un sistema *in vitro* que comprende los componentes necesarios para la expresión del producto de expresión de gen informador y en donde algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador.

Estos métodos se llevan a cabo utilizando sistemas de expresión *in vitro* que comprenden los componentes necesarios para la expresión del informador. Aunque se describe como "sin células", puede haber presentes células. Sin embargo, la expresión del informador no tiene lugar dentro de las células. Estos métodos normalmente se llevan a cabo *in vitro* y en ausencia de células. Estos sistemas de expresión contienen los componentes necesarios para sintetizar proteínas a partir del ADN *in vitro*, es decir, contienen enzimas, incluida la ARN polimerasa, ribosomas, ARNt, aminoácidos, factores de iniciación, elongación y terminación, etc. y sólo requieren la adición de ADN molde. Se pueden usar kits de transcripción-traducción *in vitro* disponibles comercialmente. Un ejemplo de un kit de transcripción-traducción *in vitro* disponible comercialmente es el kit de síntesis de proteínas *in vitro* PURExpress® disponible en New England Biolabs (número de catálogo E6800).

En tales métodos sin células, la proteína informadora puede ser cualquier proteína que proporcione un fenotipo observable, por ejemplo una proteína informadora fluorescente o una proteína que proporciona una señal colorimétrica. Anteriormente se han descrito más detalles sobre proteínas informadoras adecuadas.

Como alternativa, la proteína informadora podría ser DHFR y podría controlarse NADPH para determinar la expresión de proteínas. DHFR es una enzima que reduce el ácido dihidrofólico a ácido tetrahidrofólico, utilizando NADPH como donante de electrones, lo que significa que a medida que se produce ácido tetrahidrofólico, el NADPH se oxida a NADP+. La oxidación de NADPH a NADP+ va acompañada de una disminución de la absorbancia a 340 nM (A340), que puede controlarse mediante espectrofotometría. Por lo tanto, cuando la proteína informadora es DHFR, un aumento en la expresión de proteínas puede revelarse mediante una disminución en la absorbancia a 340 nM.

Antagonistas, células, kits y bibliotecas

En algunas realizaciones, los métodos de detección descritos en el presente documento comprenden además aislar el compuesto de prueba que se ha indicado como un antagonista funcionalmente activo como una proteína de unión a ADN. Como se ha indicado anteriormente, los antagonistas funcionalmente activos de las proteínas de unión al ADN proporcionados por los métodos de la presente invención pueden ser útiles para inhibir la proteína de unión al ADN en un entorno terapéutico. Por lo tanto, los antagonistas funcionalmente activos pueden tener utilidad en el tratamiento *per se* como productos farmacéuticos o pueden ser compuestos líderes valiosos para modificación y mejora. Por tanto, los aspectos de la presente divulgación descritos anteriormente pueden comprender además la etapa de formular el agente identificado mediante el cribado con un excipiente farmacéuticamente aceptable. Las composiciones farmacéuticas pueden formularse y estar destinadas a la administración mediante cualquier número de vías que incluyen, sin limitación, las vías oral, intravenosa, intramuscular, intraarticular, intraarterial, intramedular, intratecal, intraventricular, transdérmica, subcutánea, intraperitoneal, intranasal, entérica, tópica, sublingual o rectal.

La invención incluye la combinación de los aspectos y rasgos preferidos descritos, salvo donde dicha combinación sea claramente inadmisibles o se evite expresamente.

Las características desveladas en la descripción anterior o en las reivindicaciones a continuación o en los dibujos adjuntos, expresadas en sus formas específicas o en términos de un medio para realizar la función desvelada o un método o proceso para obtener los resultados desvelados, según corresponda, pueden, por separado o en cualquier combinación de tales características, utilizarse para realizar la invención en diversas formas de la misma.

Para evitar cualquier duda, cualesquier explicaciones teóricas proporcionadas en el presente documento se proporcionan con el fin de mejorar la comprensión del lector. Los inventores no desean quedar ligados a ninguna de estas explicaciones teóricas.

Cualesquier encabezados de sección utilizados en el presente documento tienen fines solamente organizativos y no deben considerarse limitantes de la materia objeto descrita.

A lo largo de la presente memoria descriptiva, incluyendo las reivindicaciones a continuación, a menos que el contexto

requiera otra cosa, la palabra "comprenden" e "incluyen" y variaciones como "comprende", "que comprende" y "que incluye" se entenderá que implican la inclusión de un entero o medida o grupo de enteros o de medidas establecidos, pero no la exclusión de ningún otro entero o medida o grupo de enteros o de medidas.

- 5 Cabe señalar que, como se usa en la memoria descriptiva y en las reivindicaciones adjuntas, las formas singulares "un", "uno/a" y "el" o "la", incluyen las referencias en plural, a menos que el contexto dicte claramente lo contrario. En el presente documento, los intervalos se pueden expresar como desde "aproximadamente" un valor concreto y/o hasta "aproximadamente" otro valor concreto. Cuando se expresa un intervalo de este tipo, otra realización incluye desde un valor particular y/o hasta el otro valor particular. De manera similar, cuando los valores se expresan como aproximaciones, mediante el uso del antecedente "aproximadamente", se entenderá que el valor particular constituye otra realización. El término "aproximadamente" en relación con un valor numérico es opcional y significa, por ejemplo, +/- 10 %.

Sumario de las figuras

- 15 Las realizaciones y experimentos que ilustran los principios de la invención se analizarán ahora con referencia a las figuras adjuntas en las que:

20 **Figura 1. Ensayo de supervivencia en bloque de transcripción (TBS) para obtener inhibidores de AP-1 funcionalmente activos.** Esquema que ilustra los principios del ensayo TBS que utiliza mDHFR como casete de expresión de gen informador y elementos de respuesta de TPA (TRE) como sitios de unión al ADN. **A)** Se generó un gen mDHFR que contiene quince TRE introducidos en su región transcrita, unido operativamente a un promotor (por ejemplo, el operón lac). **B)** cJun básico puede formar homodímeros unidos al ADN que se unen a los TRE y previenen la transcripción de mDHFR, inhibiendo la formación de colonias en condiciones selectivas donde la expresión de mDHFR es necesaria para la supervivencia celular. **C)** Los péptidos que se unen a la región de hélice superenrollada de cJun básico (por ejemplo, cFos) pero no disocian el complejo unido al ADN no son "antagonistas funcionales" y no rescatarán la transcripción del gen mDHFR y, por lo tanto, no rescatarán la formación de colonias en condiciones selectivas. **D)** La expresión de un inhibidor funcionalmente activo da como resultado la disociación de cJun básico de los sitios TRE en el gen mDHFR, lo que conduce a la restauración de la transcripción-traducción de mDHFR y la formación de colonias.

Figura 2. Ensayos de informador de Luciferasa.
Ensayo informador del gen de luciferasa impulsado por AP-1 sin transfección (control), vector ficticio transfectado (hélice ficticia), c-jun ácido, c-Fos ácido y péptido derivado de biblioteca (FosW ácido). Las proteínas AP-1 ácidas reducen la actividad de la luciferasa impulsada por AP-1 donde el péptido inhibidor de la **Figura 1D** tiene una reducción espectacular en la señal. El asterisco (*) indica los resultados obtenidos cuando se usó el inhibidor peptídico de la **Figura 1D**. Estos datos sugieren que los péptidos que pueden disociar las proteínas AP-1 unidas al ADN en el ensayo de TBS bacteriano producen un fuerte efecto fenotípico en las células eucariotas.

40 **Figura 3. Cuantificación de los resultados del ensayo TBS utilizando TRE DHFR.**
Esta figura proporciona la cuantificación de la formación de colonias en un ensayo de TBS que utiliza el gen de DHFR diseñado para contener 15 sitios de unión a TRE (TRE mDHFR) como informador y AP-1 como proteína de unión al ADN. Las células bacterianas transfectadas con células bacterianas transfectadas con TRE mDHFR, la parte de la cremallera de leucina de cJun (cJun LZ) y la parte de la cremallera de leucina de cFos (cFos LZ) dieron como resultado un gran número (>300) de colonias bacterianas supervivientes. Las células transfectadas con TRE mDHFR, cJun que contiene el dominio básico de unión al ADN (cJun bZIP) y cFos LZ dio como resultado un número muy bajo de colonias (<20). Esto demuestra que el dominio básico de unión al ADN de cJun es capaz de unirse a los sitios TRE en TRE mDHFR e inhibir la expresión de la proteína DHFR. Las células transfectadas con TRE mDHFR, cJun bZIP y un péptido (FosW ácido) que es capaz de disociar la proteína cJun bZIP de los sitios TRE en TRE mDHFR dan como resultado un aumento sustancial de colonias bacterianas (>200).

Ejemplos

EJEMPLO 1 - Desarrollo de un abordaje generalizado para derivar inhibidores peptídicos funcionalmente activos de la actividad del factor de transcripción

Muchos abordajes de diseño racionales, abordajes de detección aleatoria y sistemas de selección dan como resultado la identificación con éxito de compuestos capaces de unirse a dianas proteicas determinadas. Sin embargo, lo que es mucho más difícil de asegurar, es que la unión a dicha diana dará como resultado la ablación de la función de la proteína diana. Hay muchos casos en los que la formación de una interacción proteína-proteína (PPI) no ha garantizado la pérdida de función. Para abordar este importante obstáculo en la selección y el diseño de antagonistas y acelerar el diseño de antagonistas funcionalmente activos, los inventores de han inspirado en el sistema de unión al ADN del factor de transcripción y se ha invertido su papel en la transcripción.

Introducción de sitios de unión al ADN en el gen DHFR

Puede resultar difícil predecir si un compuesto derivado para unirse a una proteína determinada antagonizará su función. Para abordarlo se ha tomado el gen correspondiente a la enzima esencial, dihidrofolato reductasa (DHFR) e introdujeron 15 elementos de respuesta a TPA (TRE) en el gen. Esto se ha logrado utilizando una combinación de mutaciones tanto silenciosas como conservadas, de modo que se conserve la actividad de la enzima.

Todos los cambios se han realizado en regiones de la molécula expuestas al disolvente para minimizar las perturbaciones estructurales, con varios cambios propuestos eliminados mediante una inspección minuciosa del área de superficie accesible (ASA) dentro del archivo pdb (PDBid = 2FZJ (Cody *et al.* (2006))). Esto se hizo introduciendo el archivo pdb en la calculadora ASA en <http://cib.cf.ocha.ac.jp/bitool/ASA/>. Se utilizó un valor de corte de 20: los restos que tenían un valor de ASA inferior a este se consideraron enterrados y no modificados; se consideran expuestos los restos que tuvieron un valor de ASA mayor a este.

No se han realizado cambios en los restos considerados importantes para la catálisis o la unión de NADPH. Se conocen métodos para identificar las regiones expuestas al disolvente de la proteína informadora. Por ejemplo, es posible tomar los archivos de coordenadas para la proteína informadora, por ejemplo, un archivo de banco de datos de proteínas (PDB) y utilizar un programa que calcula el área de superficie accesible (ASA) que informa al usuario qué tan expuestos/enterrados están los restos dentro de una estructura. Se puede encontrar un programa ASA ilustrativo en <http://cib.cf.ocha.ac.jp/bitool/ASA/>. Se puede utilizar un valor de corte ilustrativo de 20, de manera que los restos inferiores a este se consideran enterrados y los superiores se consideran expuestos. De esta forma, se pueden identificar las ubicaciones de los restos expuestos al disolvente y modificar los codones en consecuencia.

A continuación se muestra la secuencia del gen mDHFR. (**SEQ ID NO: 11**) con mutaciones de ADN en negrita y subrayadas y cambios dentro de la secuencia de proteínas traducida (**SEQ ID NO: 31**) mostrado. En cursiva y negrita se muestran los sitios NheI y HindIII utilizados para subclonar el gen en el vector pES300d. Las mutaciones se realizaron mediante inspección de las secuencias consenso deseadas (TGACTCA o TGAGTCA) y los tres marcos y los cambios correspondientes en la secuencia de aminoácidos al realizar los cambios necesarios de un solo par de bases. Por ejemplo, cualquiera de las dos secuencias deseadas anteriores se puede colocar en cualquiera de los tres marcos de lectura y se pueden proporcionar la secuencia de aminoácidos correspondiente y las variaciones toleradas:

- i) Cuadro 1: TGA CTC Axx 1=terminación 2=LV 3=I/M/T/N/K/S/R
- ii) Cuadro 2: xTG ACT CAx 1=LMV 2=TS 3=HQ
- iii) Cuadro 3: xxT GAC TCA 1=FSYCLPHRITNVADG 2=DE 3=S

Esto da lugar a que se identifiquen una serie de codones para la mutación silenciosa y, en consecuencia, una serie de opciones para mutaciones conservadas o semiconservadas que permitirían la introducción de TRE en el gen mDHFR:

i) Sin opciones

ii) LSH, LSQ, LTH, LTQ, MSH, MSQ, MTH, MTQ, VSH, VSQ, VTH, VTQ

iii) ADS, AES, CDS, CES, DDS, DES, FDS, FES, GDS, GES, HDS, HES, IDS, IES, LDS, LES, NDS, NES, PDS, PES, RDS, RES, SDS, SES, TDS, TES, VDS, VES, YDS, YES

A partir de esto, los inventores pudieron implementar los siguientes cambios en el gen mDHFR para generar una perturbación mínima en la secuencia general. Cuando fue posible, las mutaciones fueron silenciosas o conservadoras. Todas las mutaciones también se colocaron en sitios expuestos al disolvente y lejos del centro catalítico (E116) y lejos de los restos necesarios para la unión de NADPH/sustrato (A10/R71). Esto tuvo como resultado en la introducción de 15 TRE en el gen mDHFR:

- | | | | |
|-----|---------|--------------|-----------------------------|
| 1. | VSQ | (silencioso) | = <u>GTG AGT CAG</u> |
| 2. | NEF→NES | (F32S) | = <u>AAT GAG TCA</u> |
| 3. | MTT→MTQ | (T40Q) | = <u>ATG ACT CAG</u> |
| 4. | TSS→TDS | (S42D) | = <u>ACT GAC TCA</u> |
| 5. | VEG→VES | (G46S) | = <u>GTT GAG TCA</u> |
| 6. | PEK→PES | (K64S) | = <u>CCT GAG TCA</u> |
| 7. | LSR→LSQ | (R78Q) | = <u>CTG AGT CAA</u> |
| 8. | IEQ→IES | (Q103S) | = <u>ATT GAG TCA</u> |
| 9. | VDM→VDS | (M112S) | = <u>GTT GAC TCA</u> |
| 10. | MNQ→MTQ | (N127T) | = <u>ATG ACT CAA</u> |
| 11. | VTR→VTQ | (R138Q) | = <u>GTG ACT CAG</u> |
| 12. | FES | (silencioso) | = <u>TTT GAG TCA</u> |
| 13. | IDL→IDS | (L154S) | = <u>ATT GAC TCA</u> |
| 14. | PEY→PES | (Y163S) | = <u>CCT GAG TCA</u> |
| 15. | LSE→LSQ | (E169Q) | = <u>CTG AGT CAG</u> |

Este proceso de diseño dio lugar a la siguiente secuencia:

```

A  S  V  R  P  L  N  C  I  V  A  V  S  Q  N  M  G
GCT AGC GTT CGA CCA TTG AAC TGC ATC GTC ACC GTC AGT CAG AAT ATG GGG
I  G  K  R  G  D  L  F  W  P  F  L  R  N  E  S  K
ATT GGC AAG AAC GGA GAC CTA CCC TGG OCT CCG CTC AGG AAT GAG TCA ARG
Y  F  Q  R  M  T  Q  T  D  S  V  E  S  R  Q  N  L
TAC TTC CAA AGA ATG ACT CAG ACT GAC TCA GTT GAG TCA AAA CAG AAT CTG
V  I  M  G  R  E  F  W  F  S  I  F  E  S  N  R  P
GTG ATT ATG GGT AGG AAA ACC TGG TTC TCC ATT CCT GAG TCA AAT CGA CCT
L  F  D  R  I  N  I  V  L  S  Q  E  L  R  E  P  P
TTA AAG GAC AGA ATT AAT ATA GTT CTG AGT CAA GAA CTC AAA GAA CCA CCA
R  G  A  H  F  L  A  E  S  L  D  D  A  L  R  L  I
CGA CGA GCT CAT TTT CTT GGC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT
E  S  F  E  L  A  S  K  V  D  S  V  W  I  V  G  G
GAG TCA CCG GAA TTG GCG AGC AAA GTT GAC TCA GTT TGG ATC GTC GGA GGC
S  S  V  Y  Q  E  A  M  T  Q  F  G  H  L  R  L  F
AGT TCT GTT TAC CAG GAA CCC ATG ACT CAA CCA GCG CAC CTT AGA CTC TTT
V  T  Q  I  M  Q  E  F  E  S  D  T  F  F  F  E  I
GTC ACT CAG ATC ATG CAG GAA TTT GAG TCA GAC ACG TTT TTC CCA GAA ATT
D  S  G  K  Y  K  L  L  P  E  S  P  G  V  L  S  Q
GAC TCA GGG AAA TAT AAA CTT CTC CCT GAG TCA CCA GCG GTC CTG AGT CAG
V  Q  E  E  K  G  I  E  F  E  F  E  V  Y  E  K  K
GTC CAG GAG GAA AAA GGC ATC AAG TAT AAG TTT GAA GTC TAC GAG AAG AAA
D  *  A  *
GAC TAA GCT TAA

```

- 5 Los inventores introdujeron 15 TRE mediante mutaciones silenciosas y conservadas en posiciones expuestas a disolventes dentro del gen que codifica la enzima esencial dihidrofolato reductasa (DHFR; **figura 1A**). Demostraron que estos cambios dan como resultado una enzima funcional. En condiciones selectivas, la introducción de AP-1 previene la expresión de DHFR uniéndose a sitios TRE dentro del gen, bloqueando la transcripción e impidiendo la formación de colonias en condiciones selectivas. Por el contrario, las versiones atenuadas de AP-1 que carecen de una región básica de unión al ADN no logran prevenir la formación de colonias. De manera significativa, la introducción de péptidos que pueden unirse a AP-1 y antagonizar la función se puede utilizar para revertir este efecto, dando lugar a la formación de colonias bacterianas. Esto representa un potente abordaje para la selección de péptidos funcionalmente activos que pueden unirse a los componentes de AP-1 y anular su función, lo que lleva a una rápida selección de bibliotecas e identificación de secuencias terapéuticamente interesantes.

Prueba de la funcionalidad de la proteína DHFR

El sistema de selección se basa en el hecho de que la DHFR bacteriana puede inhibirse específicamente utilizando trimetoprima, haciendo que las células dependan de la actividad de DHFR murina (mDHFR) para su supervivencia. La primera prueba del sistema fue establecer que la proteína mDHFR se repliega y está activa. Se utilizó el análisis SDS-PAGE para confirmar que la proteína se expresa altamente tras la adición de IPTG. Más evidencia de que la proteína se expresa, se pliega y es funcionalmente activo se verificó mediante la transformación de células bacterianas y se confirmó mediante la presencia de múltiples colonias en medios mínimos que contienen trimetoprima (datos no mostrados).

Establecimiento del ensayo TBS

A continuación, fue necesario establecer que la introducción de un componente AP-1 (en este caso cJun básico) daría como resultado la unión a los 15 TRE introducidos dentro del gen mDHFR y, por lo tanto, el fracaso de la transcripción del gen.

Se utilizaron tres plásmidos para el ensayo TBS. Estos son i) p300-mDHFR (Cm; **SEQ ID NO: 42**) para expresar la secuencia de consenso 15x que contiene mDHFR, que está bajo el control del operón lac; ii) p230d-cJun-básico (Amp; **SEQ ID NO: 43**) que también está bajo el control del operón lac; iii) pREP4 (Kan; **SEQ ID NO: 44**) para expresar el represor lac.

Las células se cultivaron en condiciones no selectivas (es decir, agar LB/LB) que contenían Cm/Amp/Kan hasta el momento del ensayo. Durante la selección de TBS, las células se cultivan en medio mínimo M9 (agar o caldo) en presencia de Cm/Amp/Kan, así como Tmp (para inhibir las copias bacterianas de DHFR) e IPTG (para inducir la expresión de las proteínas mDHFR y bZIP). Durante la selección del ensayo, se usa medio en IPTG para servir como control negativo para garantizar que la supervivencia celular esté impulsada exclusivamente por la pérdida en la interacción entre la proteína diana bZIP y las secuencias consenso ubicadas dentro del gen mDHFR. Las placas de agar M9 con Tmp en ausencia de IPTG no forman colonias (datos no mostrados). El uso de IPTG para expresar

mDHFR que comprende los 15 TRE confiere supervivencia en medio mínimo M9 que contiene Tmp para inhibir la DHFR bacteriana (datos no mostrados).

Como era de esperar, la sobreexpresión de cJun-básico en el segundo plásmido resultó en una pérdida completa de colonias bacterianas en medios mínimos (datos no mostrados). Sin desear quedar ligados a teoría alguna, se cree que esto funciona porque AP-1 se une a los múltiples TRE que se encuentran dentro del gen mDHFR y, por lo tanto, funciona de manera opuesta a su función natural. Más bien funciona bloqueando la transcripción e impidiendo que la maquinaria se mueva a lo largo del ADN. (**Figura 1B**). Como control, se probó una versión de cJun que contiene la cremallera de leucina, pero sin la región básica de unión al ADN (**SEQ ID NO: 45**). Como se esperaba, esta versión no impidió la formación de colonias bacterianas en medio mínimo (datos no mostrados).

Por último, los inventores añadieron compuestos que se sabe que se unen a la hélice superenrollada de cJun para establecer si también son péptidos funcionalmente activos, es decir, capaces de secuestrar cJun básico como un heterodímero no funcional y, por lo tanto, prevenir la unión al ADN y la transcripción de rescate. El complejo homodímero c-Jun/ADN forma una interacción muy estrecha, por lo tanto, se requiere un inhibidor genuino de c-Jun no solo para unirse al homodímero c-Jun en solución sino también para disociar el complejo unido al ADN para prevenir la actividad transcripcional de AP-1. Los péptidos que se unen al dominio de hélice superenrollada (el motivo de dimerización del homodímero c-Jun) no necesariamente disocian el complejo unido al ADN (**Figura 1C**). Por lo tanto, aquellos que no lo hacen no rescatarán el crecimiento bacteriano de las células tratadas con TMP/IPTG (datos no mostrados). Sin embargo, los péptidos que conducen a la disociación del homodímero c-Jun del ADN permitirán la transcripción del gen mDHFR y rescatarán las células bacterianas TMP/IPTG y crecerán. Esto se demuestra mediante el uso del péptido FosW que se sabe que se une a cJun con alta afinidad ($K_D = 90$ nM; Mason *et al.* 2006). La expresión del péptido FosW dio como resultado la formación de colonias (datos no mostrados), demostrando que el péptido FosW era capaz de rescatar la transcripción del gen mDHFR y por tanto es un antagonista funcionalmente activo (**Figura 1D**).

Este ensayo permite la selección de péptidos específicos capaces no solo de unirse a la región de hélice superenrollada que se predice que será necesaria para impulsar la dimerización de AP-1, pero lo más importante es que esa función desactiva la actividad de unión al ADN de la proteína. Por lo tanto, los péptidos deben ser funcionalmente activos para ser seleccionados y pueden aislarse fácilmente de aglutinantes que no son útiles utilizando la supervivencia celular como marcador de éxito.

Actividad in vitro en células de mamífero

Las proteínas AP-1 de tipo salvaje tienen una gran cantidad de aminoácidos básicos en el extremo N de la proteína que permiten la unión al ADN. Se pueden generar versiones ácidas de las proteínas AP-1 con una larga serie de aminoácidos cargados negativamente (en lugar de los positivos) que pueden servir como imitadores del ADN. Por lo tanto, las variantes del péptido AP-1 ácido pueden formar heterodímeros con proteínas AP-1 de tipo salvaje impulsadas tanto por el motivo de dimerización como por la interacción de carga negativa/positiva, lo que da como resultado su secuestro del ADN.

Para probar si es probable que los péptidos identificados como antagonistas funcionales en el ensayo TBS también tengan la misma actividad en células eucariotas, se llevó a cabo un ensayo informador del gen de luciferasa impulsado por AP-1. Este ensayo se llevó a cabo sin transfección (control), vector ficticio transfectado, c-jun ácido, c-Fos ácido y tratado con péptido derivado de biblioteca (FosW). Los resultados se muestran en la **Figura 2**. Las proteínas AP-1 ácidas reducen la actividad luciferasa dirigida por AP-1 donde el inhibidor peptídico del Ejemplo 1 (FosW) tiene una reducción espectacular en la señal. Estos datos sugieren que los péptidos que pueden disociar las proteínas AP-1 unidas al ADN en el ensayo de TBS bacteriano producen un fuerte efecto fenotípico en las células eucariotas.

Análisis

Existen muchos ensayos para derivar PPI de alta afinidad utilizando abordajes basados en bibliotecas, pero muy pocos que garanticen la pérdida de función dentro de la proteína diana. Usando AP-1 como sistema ilustrativo, los inventores han desarrollado un ensayo para obtener secuencias funcionalmente activas capaces de detener la actividad del factor de transcripción. Utilizando la enzima esencial mDHFR han demostrado que **i)** la actividad enzimática se conserva tras la introducción de 15 TRE en el gen **ii)** en condiciones selectivas, la actividad se pierde cuando se introduce cJun básico **iii)** la región básica dentro de cJun básico es un requisito absoluto para esta pérdida de actividad de mDHFR y **iv)** los péptidos derivados para unirse a cJun se pueden separar en aquellos que dan como resultado la pérdida de la actividad de unión al ADN de AP-1 (y por lo tanto de su función) y los que no. Por lo tanto, este ensayo utiliza la supervivencia celular como marcador para permitir la detección rápida de bibliotecas de péptidos y, en consecuencia, la derivación de antagonistas funcionalmente activos de la función del factor de transcripción.

EJEMPLO 2 - Creación de biblioteca

Ensayo TBS: construcción de biblioteca genéticamente codificada:

Como se ha descrito anteriormente, se utilizaron tres plásmidos para el ensayo TBS. Estos son i) p300-mDHFR (Cm) para expresar la secuencia de consenso 15x que contiene mDHFR, que está bajo el control del operón lac; ii) p230d-cJun básico (Amp) que también está bajo el control del operón lac; iii) pREP4 (Kan) para expresar el represor lac.

5 Las bibliotecas codificadas genéticamente se crean mediante PCR de extensión superpuesta, subclonadas en el vector p410d (Tet) y sembradas. Cada colonia representa entonces un miembro de la biblioteca. Normalmente los inventores recogieron 2-5 veces el tamaño de la biblioteca en números de colonias para ganar aproximadamente el 95 % de cobertura total. El tamaño máximo de biblioteca que se puede filtrar utilizando este abordaje es 10^6 . Una vez que la biblioteca está completa, se agrupan las colonias y se realiza una minipreparación del ADN. Finalmente, la biblioteca de plásmidos se transforma en células que contienen p300/p230/pREP4. Durante la selección de una sola etapa, las células se siembran en agar LB (para demostrar una transformación con éxito), agar M9 sin IPTG (como control negativo donde no se expresa bZIP o mDHFR) y finalmente en agar M9 que contiene Cm/Amp/Kan/Tet/Tmp/IPTG para dirigir la producción de bZIP/mDHFR/Biblioteca de modo que solo se restablezca la viabilidad celular. Si un miembro de la biblioteca determinado puede evitar que el objetivo bZIP interactúe con las secuencias afines dentro del gen mDHFR. A continuación, las colonias supervivientes se pueden agrupar, cultivar y diluir en serie en cultivos líquidos en condiciones selectivas (medio mínimo M9 con 1 µg/ml de trimetoprima). El crecimiento más rápido y, por tanto, los socios que interactuaban con mayor afinidad dominaban el grupo. Se secuenciaron conjuntos de bibliotecas y colonias de clones individuales para verificar la llegada a una secuencia. Para evaluar la calidad de la biblioteca, los inventores secuenciaron grupos y clones individuales para encontrar distribuciones aproximadamente iguales de aminoácidos variados. Las colonias agrupadas excedieron el tamaño de la biblioteca entre 5 y 10 veces. Usando métodos de ligación más recientes (Topo/Gibson/Gateway), puede ser posible pasar a TBS directamente desde la ligación, dando la importante ventaja de poder examinar bibliotecas más grandes (posiblemente hasta 10^{10} o 10^{11}), sin embargo, será necesario implementar procesos (por ejemplo, secuenciación de próxima generación) para garantizar que el tamaño y la calidad de la biblioteca estén completamente representados antes de la transformación en el ensayo TBS.

Otra posibilidad es utilizar pET24a como alternativa al vector pREP4 utilizado para expresar el represor lac. Esto permitiría la expresión tanto del represor lac como del miembro/antagonista de la biblioteca en un único plásmido, es decir, evitando la necesidad de otro antibiótico.

Ensayo TBS - Adición de compuestos extracelulares:

Para bibliotecas extracelulares, las células que contienen el plásmido p300-mDHFR se cultivan en presencia de los plásmidos p230d-bZIP y pREP4 en condiciones no selectivas (agar LB/medio). Una vez listo para el ensayo, se pueden colocar las noches en cada pocillo de las placas de microtitulación (96, 384, 1536) a $A_{600} = 0,05$ y bibliotecas compuestas analizadas mediante adición directa a cada pocillo. Las placas se incuban a 37 °C y se agitan y los compuestos se identifican con éxito mediante el seguimiento de la señal de absorbancia a 600 nm. Este método de **adición de compuestos extracelulares** tiene la ventaja de permitir al usuario alejarse de las bibliotecas de péptidos estándar (por ejemplo, se pueden crear perfiles de péptidos restringidos por hélice, peptidomiméticos, aminoácidos no naturales, etc., o incluso bibliotecas de moléculas pequeñas) y, lo que es más importante, permite al usuario perfilar la penetrancia celular concomitantemente con el antagonismo funcional del bZIP diana. Una vez más, todas las proteínas están bajo el control de un promotor lac y la expresión se indujo con isopropil β-D-1-tiogalactopiranosido (IPTG).

Selección de los péptidos ganadores

Brevemente, durante el TBS, solo los péptidos (intracelulares) o compuestos (extracelulares) que pueden interactuar con el bZIP diana y disociarlo de las secuencias consenso dentro del gen mDHFR darán como resultado la formación de colonias/crecimiento celular en placas con medio mínimo M9/medios con 1 µg/ml. trimetoprima para inhibir la DHFR bacteriana.

EJEMPLO 3 - Experimento adicional de ensayo de TBS

Se llevó a cabo un experimento adicional para establecer el ensayo TBS como un ensayo adecuado para identificar antagonistas funcionales.

Este experimento utilizó plásmidos similares a los establecidos en el Ejemplo 1, con la excepción de que se usó pET24a para expresar tanto el represor lac como el compuesto de prueba usando un único plásmido. Específicamente, se utilizaron tres plásmidos: i) p300d que expresa el gen mDHFR modificado para incluir 15 TRE (TRE mDHFR, generado como se describe en el Ejemplo 1); ii) p230d que expresa la parte de la cremallera de leucina de cJun (cJun LZ) o cJun con la región básica y la cremallera de leucina (cJun bZip); iii) pET24a que expresa el represor lac y la parte de cremallera de leucina de cFos (cFos LZ) o el péptido FosW ácido (FosW ácido).

Las placas se generaron como se expone en la siguiente tabla:

Placa	p300d	p230d	pET24a
1	TRE mDHFR	cJun LZ	cFos LZ
2	TRE mDHFR	cJun bZIP	cFos LZ
3	TRE mDHFR	cJun bZIP	FosW ácido

Se contó el número de colonias producidas en cada placa y esta cuantificación se ilustra en la **figura 3**. El cJun LZ expresado en la placa de control 1 es incapaz de unirse e inhibir la expresión de TRE DHFR, por lo tanto, se produce un gran número (>300) de colonias bacterianas. cJun con el dominio básico de unión al ADN (cJun bZip) es capaz de unirse e inhibir la expresión de DHFR y, por lo tanto, cuando se expresa en la placa 2, resultó en una reducción sustancial en el número de colonias bacterianas (<20) en comparación con la placa 1. La coexpresión con un péptido que se une a la región en hélice superenrollada de cJun (cFos LZ) no disocia el complejo unido al ADN y, por lo tanto, no es capaz de rescatar la expresión de la proteína DHFR en el ensayo de TBS. Sin embargo, la coexpresión con un péptido (FosW ácido) que es capaz de disociar el factor de transcripción AP-1 unido al ADN de sus sitios de unión en el gen TRE DHFR da como resultado un aumento sustancial en el número de colonias en comparación con la placa 2

En general, estos resultados proporcionan evidencia adicional de que el ensayo TBS se puede utilizar para identificar antagonistas que pueden inhibir la actividad de unión al ADN de un factor de transcripción como AP-1.

EJEMPLO 4 - introducción de los sitios de unión de CRE, CCAAT y caja E en el gen DHFR

Se diseñaron construcciones mediante las cuales se insertaron los sitios de unión a CRE, CCAAT y caja E, respectivamente, en el gen de la DHFR. Estas construcciones se pueden probar en el ensayo TBS como se describe en los Ejemplos 1 y 3.

Inserción del sitio de unión de CRE en el gen de DHFR

CRE generalmente se define como **TGACGTCA** (SEQ ID NO: 10). Las mutaciones en el gen de DHFR se pueden realizar mediante la inspección de la secuencia consenso deseada y los tres marcos y los cambios correspondientes en la secuencia de aminoácidos al realizar los cambios necesarios de un solo par de bases. El CRE tiene 8 pb, por lo que puede abarcar cuatro codones. Por ejemplo, la secuencia definida anteriormente se puede colocar en uno cualquiera de los tres marcos de lectura y se pueden proporcionar la secuencia de aminoácidos correspondiente y las variaciones toleradas:

A: Cuadro 1: TGA CGT CAX 1=terminación 2:R 3:H/Q
 B: Cuadro 2: xTG ACG TCA 1:LMV 2:T 3:S
 C: Cuadro 3: xxT GAC GTC Axx 1:FLIVSPTAYHND CRG 2:D 3:V 4:IMTNKSR

A partir de esto es posible implementar cambios en el gen mDHFR para provocar una perturbación mínima en la secuencia general. Las mutaciones deben ubicarse en sitios expuestos a disolventes y lejos del centro catalítico y, cuando sea posible, las mutaciones deben ser silenciosas o conservadoras.

A continuación se muestra un ejemplo de un gen mDHFR que se modifica para contener sitios de unión a CRE:

```

ATGCTTTCGACCATTTGAACCTCCATCGCTGCGCGCTGTGCCCCAAATATGGCGATTGGCAGAAACGGAGACCTAC
CCTGGGCTTCGGCTCAGGAAACGAGTTCAAGTACTTCCBAAGAATGACGTCAACCTCTTCACTGGGAAGGTAA
ACAGAAATCTGGTGGTTATGGCTAGGAAAACCTGGTTCTCCATTCTCTGAGAAGAAATCGACCTTTAAAGGAC
AGAATTAATATATGTGACGTCAAGAGAACTCAAAGAAACCAACGAGGAGCTCATTTTCTTGGCCAAAAGTT
TGGATGATGCCCTTAAGACTTATTTGAACAAACCGGAATTGACGTCAAAAGTAGACATGGTTTGGATCTCTCG
AGGCACTTCTGTTTTACAGGAAGCCATGAATCAACAGGCCACCTTAGACTTTTTGTGACGTCAATCATG
CAGGAATTTGAAAGTCACACGTTTTTCCAGAAAATTGATTTGGGGAAATATAAACTCTCTCCAGAAATACC
CAGGCGGTGACGTCAGAGCTCTACGAGCAAAAAGGCATCAAGTATAACTTTGAAGTCTACGAGAAAGAA
CTAAGCTTAA

```

Los restos de nucleótidos en **subrayado en negrita** indican sitios de unión a CRE de consenso.

Los restos de nucleótidos en minúscula y cursiva corresponden a los sitios de enzimas de restricción para Ascl y HindIII en los extremos 5' y 3' de la secuencia, respectivamente.

La secuencia de aminoácidos resultante se muestra a continuación:

```

M  V  R  P  L  R  C  I  V  A  V  S  Q  W  M  G  F  G  K  N
G  D  L  E  W  D  P  L  R  N  E  P  K  Y  P  Q  R  M  T  W
T  S  S  V  E  G  K  Q  N  L  V  I  N  G  R  K  T  W  P  S
I  P  E  K  R  R  P  L  K  D  R  I  N  I  V  E  D  R  E  L
K  E  P  P  R  G  A  H  F  L  A  K  S  L  D  D  A  L  R  L
I  E  Q  P  E  L  E  S  S  V  D  M  V  W  I  V  G  G  S  S
V  I  Q  E  A  H  N  Q  P  G  S  L  R  L  P  V  T  S  I  M
Q  E  P  E  Q  D  T  S  F  P  E  I  D  L  G  K  Y  K  L  L
P  S  Y  P  Q  V  E  S  E  V  Q  E  E  K  G  I  R  Y  K  P
S  V  Y  S  K  K  D
  
```

- 5 Los restos de aminoácidos en *cursiva* son restos expuestos a disolventes. Los demás restos se clasifican como restos enterrados.

Los restos de aminoácidos en **subrayado en negrita** son restos que han sido alterados como resultado de la inserción de CRE en la secuencia de nucleótidos.

10

A continuación se proporciona un resumen de los cambios de aminoácidos:

1. MTT→MTS (T40S) = **ATG ACG TCA** ASA en la posición = 36
2. VLS→VTS (L76T) = **GTG ACG TCA** ASA en la posición = 21
3. LAS→LTS (A107T) = **TTG ACG TCA** ASA en la posición = 37
4. VTR→VTS (R138S) = **GTG ACG TCA** ASA en la posición = 57
5. VLS→VTS (L167T) = **GTG ACG TCA** ASA en la posición = 99

Inserción del sitio de unión a CCAAT en el gen de DHFR

15

CCAAT generalmente se define como **ATTGCGCAAT** (SEQ ID NO: 9). Las mutaciones en el gen de DHFR se pueden realizar mediante la inspección de la secuencia consenso deseada y los tres marcos y los cambios correspondientes en la secuencia de aminoácidos al realizar los cambios necesarios de un solo par de bases. El CCAAT tiene 10 pb y, por lo tanto, puede abarcar cinco codones. Por ejemplo, la secuencia definida anteriormente se puede colocar en uno cualquiera de los tres marcos de lectura y se pueden proporcionar la secuencia de aminoácidos correspondiente y las variaciones toleradas:

20

A: Cuadro 1:	<u>ATT GCG CAA</u> Txx	1:1	2:A	3:Q	4:FLSYCW*
B: Cuadro 2:	xAT TGC GCA ATx	1:YHND	2:C	3:A	4:IM
C: Cuadro 3:	xxA TTG CGC AAT	1:LIVSPTAQKERG*	2:L	3:R	4:N

25

A partir de esto es posible implementar cambios en el gen mDHFR para provocar una perturbación mínima en la secuencia general. Las mutaciones deben ubicarse en sitios expuestos a disolventes y lejos del centro catalítico y, cuando sea posible, las mutaciones deben ser silenciosas o conservadoras.

A continuación se muestra un ejemplo de un gen mDHFR que se modifica para contener CCAAT:

30

```

ATGGTTCGACCATTGAACTGCATCGTCGCCGTGTCCCAAATATGGGGATTGGCAAGAACGGAGACCTAC
CCTGGCCTCCATTGCGCAATGAGTTCAAGTACTTCCAAAGAATGACCACAACCTCTTCAGTGGAAGGTAA
ACAGAATCTGGTGATTATGGGTAGGAAAACCTGGTTCTCCATTCTGAGAAGAATCGACCATTGCGCAAT
AGAATTAATATAGTTCTCAGTAGAGAATTGCGCAATCCACCACGAGGAGCTCATTTTATTGCGCAATCCT
TGGATGATGCATTGCGCAATTATTGAACAACCGGAATTGGCGAGCAAAGTAGACATGGTTTGGATCGTCGG
AGGCAGTTCTGTTTACCAGGAAGCCATGAATCAACCAGGCCACCTTAGACTCTTTGTGACAAGGATCATG
CAGGAATTTGAAAGTGACACGTTTTTTCCAGAAATTGATTTGGGGAAATATAAACTTCTCCAGAATACC
CAGGCGTCCTCTCTGAATTGCGCAATGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGA
CTAAGCTTAA
  
```

Los restos de nucleótidos en **subrayado en negrita** indican LOS sitios de unión a CCAAT consenso.

Los restos de nucleótidos en minúscula y cursiva corresponden a los sitios de enzimas de restricción para Ascl y HindIII en los extremos 5' y 3' de la secuencia, respectivamente.

La secuencia de aminoácidos resultante se muestra a continuación:

```

M  V  R  P  L  N  C  I  V  A  V  S  Q  N  N  G  I  G  K  N
G  D  L  F  W  E  E  L  N  E  E  E  K  K  E  Q  R  M  T  P
N  S  S  F  S  G  N  Q  S  L  V  I  N  S  N  K  T  W  F  N
I  F  E  K  S  R  E  L  R  N  R  I  N  I  V  L  S  E  E  L
R  N  E  E  R  G  A  E  F  I  A  Q  S  L  D  D  A  L  E  N
I  E  Q  P  E  L  A  S  E  V  D  N  V  W  I  V  G  S  S  S
V  Y  Q  E  A  N  N  Q  E  G  S  L  E  L  F  V  F  R  I  N
Q  E  F  E  S  D  T  E  F  P  E  I  D  L  G  K  Y  K  L  L
F  E  Y  E  G  F  L  S  E  V  Q  E  E  K  Q  I  N  Y  E  E
S  V  Y  S  K  K  D
  
```

Los restos de aminoácidos en *cursiva* son restos expuestos a disolventes. Los demás restos se clasifican como restos enterrados.

Los restos de aminoácidos en **subrayado en negrita** son restos que han sido alterados como resultado de la inserción de CCAAT en la secuencia de nucleótidos.

A continuación se proporciona un resumen de los cambios de aminoácidos:

1. PLRN (silencioso)
2. PLKD→PLRN (K69R, D70N) = ASA en las posiciones = 90, 97
3. ELKE→ELRN (K81R, E82N) = ASA en las posiciones = 175,131
4. LAKS→IAQS (L90I, K92Q) = ASA en las posiciones = 47,139
5. ALRL→ALRN (L100N) = ASA en la posición = 46

Se podría insertar un sitio CCAAT adicional para realizar la siguiente mutación:

6. EVQE→ELRN (V170L, Q171R, E172N)

Inserción del sitio de unión a la caja E en el gen de DHFR

En el contexto de c-Myc, la caja E normalmente se define como CACGTG (SEQ ID NO: 7) o CACATG (SEQ ID NO: 8). Las mutaciones en el gen de DHFR se pueden realizar mediante la inspección de la secuencia consenso deseada y los tres marcos y los cambios correspondientes en la secuencia de aminoácidos al realizar los cambios necesarios de un solo par de bases. Por ejemplo, las secuencias definidas anteriormente se pueden colocar en uno cualquiera de los tres marcos de lectura y se pueden proporcionar la secuencia de aminoácidos correspondiente y las variaciones toleradas:

- | | | | | |
|--------------|------------------------|--------------------|-----|-------------------|
| A: Cuadro 1: | <u>CAC GTG</u> xxx = | 1:H | 2:V | 3: cualquier cosa |
| B: Cuadro 2: | x <u>CA CGT</u> Gxx = | 1:S/P/T/A | 2:R | 3:V/A/D/E/G |
| C: Cuadro 3: | xx <u>C ACG TG</u> x = | 1:FLIVSPTAYHNDCRSG | 2:T | 3:C/W/* |
| A: Cuadro 1: | <u>CAC ATG</u> xxx = | 1:H | 2:M | 3: cualquier cosa |
| B: Cuadro 2: | x <u>CA CAT</u> Gxx = | 1:S/P/T/A | 2:H | 3:V/A/D/E/G |
| C: Cuadro 3: | xx <u>C ACA TG</u> x = | 1:FLIVSPTAYHNDCRSG | 2:T | 3:C/W/* |

A partir de esto es posible implementar cambios en el gen mDHFR para provocar una perturbación mínima en la secuencia general. Las mutaciones deben ubicarse en sitios expuestos a disolventes y lejos del centro catalítico y, cuando sea posible, las mutaciones deben ser silenciosas o conservadoras.

A continuación se muestra un ejemplo de un gen mDHFR que se modifica para contener cajas E:

ATGGTTCGACCATTTGAACTGCATCGTCGCCGTGTCCCAAAATATGGGGATTGGCAAGAACGGAGACCTAC
CCTGGCCTCCGCTCAGGAACGAGTTCAAGTACTTCCAAAGAATGACCACAACCTCTTCAGTGGAAGGTAA
ACAGAATCTGGTGATTATGGGTAGGCGG**CACGTG**GTTCTCCATTCTGAGAAGAATCGACCTTTAAAGGAC
AGAATTAATATAGTTCTCT**CACGTG**AACTCAAAGAACCAC**CACGTG**GAGCT**CACGTG**CTTGCCAAATCAC
TGGATGATGCATTAAGACTTATTGAACAACCGGAATTGGCGT**CACGTG**TAGACATGGTTTGGATCGTCGG
AGGCAGTTCTGTTTACCAGGAAGCCATGAATCAACCAGGC**CACGTG**GAGACTCTTTGTGA**CACGTG**TCATG
CAGGAATTTGAAAGTGACACGTTTTTCCAGAAATTGATTTGGGGAAATATAAACTTCTCCAGAATACC
CAGGCGTCTCT**CACGTG**TCCAGGAGGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGA
CTAAGCTTAA

Los restos de nucleótidos en **subrayado en negrita** indican los sitios de unión a la caja E consenso.

- 5 Los restos de nucleótidos en minúscula y cursiva corresponden a los sitios de enzimas de restricción para Ascl y HindIII en los extremos 5' y 3' de la secuencia, respectivamente.

La secuencia de aminoácidos resultante se muestra a continuación:

M	V	R	P	L	N	C	I	V	A	V	S	Q	N	M	G	I	G	K	N
G	D	L	P	W	P	P	L	R	N	E	F	K	Y	F	Q	R	M	T	T
T	S	S	V	E	G	K	Q	N	L	V	I	M	G	R	R	T	W	F	S
I	P	E	K	N	R	P	L	K	D	R	I	N	I	V	L	S	R	E	L
K	E	P	P	R	G	A	H	V	L	A	K	S	L	D	D	A	L	R	L
I	E	Q	P	E	L	A	S	R	V	D	M	V	W	I	V	G	G	S	S
V	Y	Q	E	A	M	N	Q	P	G	H	V	R	L	F	V	T	R	V	M
Q	E	F	E	S	D	T	F	F	P	E	I	D	L	G	K	Y	K	L	L
P	E	Y	P	G	V	L	S	R	V	Q	E	E	K	G	I	K	Y	K	F
10	E	V	Y	E	K	K	D												

Los restos de aminoácidos en *cursiva* son restos expuestos a disolventes. Los demás restos se clasifican como restos enterrados.

- 15 Los restos de aminoácidos en **subrayado en negrita** son restos que han sido alterados como resultado de la inserción de la caja E en la secuencia de nucleótidos.

A continuación se proporciona un resumen de los cambios de aminoácidos:

- | | | | | | |
|----|---------|--------------|-----------------------------|--------------------------|------------|
| 1. | KTW→RTW | (K56R) | = CGC ACG TGG | ASA en la posición = 143 | (expuesto) |
| 2. | SRE | (silencioso) | = TCA CGT GAA | ASA en la posición = N/D | |
| 3. | PRG | (silencioso) | = CCA CGT GGA | ASA en la posición = N/D | |
| 4. | HFL→HVL | (F89V) | = CAC GTG CTT | ASA en la posición = 71 | (expuesto) |
| 5. | SKV→SRV | (K109R) | = ACA CGT GTA | ASA en la posición = 109 | (expuesto) |
| 6. | HLR→HVR | (L132V) | = CAC GTG AGA | ASA en la posición = 1,6 | |
| 7. | TRI→TRV | (I139V) | = ACA CGT GTC | ASA en la posición = 1,6 | |
| 8. | SEV→SRV | (E151R) | = ACA CGT GTC | ASA en la posición = 141 | (expuesto) |

Los cambios 6 y 7 se ubican en los restos que se clasifican como enterrados. En consecuencia, se podrían hacer construcciones que contengan los 8 sitios de caja E, uno al que le falta el sitio "6", uno al que le falta el sitio "7" y otro al que le faltan los sitios "6" y "7" para determinar si la mutación en estos sitios "enterrados" afecta a la función de la proteína DHFR resultante.

EJEMPLO 5 - Ampliación del ensayo TBS para su uso con dianas de factores de transcripción adicionales

Se llevaron a cabo experimentos para establecer el ensayo TBS para su uso en la identificación de inhibidores funcionales de factores de transcripción distintos de AP-1. Se generaron los siguientes genes de DHFR modificados:

Mutante de mDHFR TBS	Secuencia de nucleótidos del sitio de unión	Número total de sitios de unión	Secuencia de nucleótidos del mutante de mDHFR:	Notas
TRE mDHFR	TGA(C/G)TCA	15	SEQ ID NO: 4	Véase el Ejemplo 1 para detalles de la construcción
CCAAT mDHFR	ATTGCGCAAT	6	SEQ ID NO: 38	Véase el Ejemplo 4 para detalles de la construcción
CAJA E mDHFR	CAC(G/A)TG	8	SEQ ID NO: 40	Véase el Ejemplo 4 para detalles de la construcción
CRE mDHFR	TGACGTCA	5	SEQ ID NO: 36	Véase el Ejemplo 4 para detalles de la construcción

Como se establece en el Ejemplo 1, la expresión de mDHFR TRE fue capaz de restaurar colonias bacterianas en presencia de TMP, lo que indica que la proteína producida por este mutante de mDHFR era funcional. Se llevaron a cabo experimentos similares para determinar si los otros mutantes de mDHFR (mDHFR CCAAT, mDHFR CAJA E, mDHFR CRE) también estaban activos. Estos experimentos revelaron que la expresión del mutante de mDHFR en presencia de TMP dio como resultado un mayor número de colonias en comparación con las placas donde estaba presente TMP sin que se expresara el mutante de mDHFR (datos no mostrados). Estos experimentos confirman que las proteínas DHFR producidas por los mutantes mDHFR eran funcionales y capaces de conferir supervivencia en presencia de TMP.

También se llevaron a cabo experimentos para determinar si el ensayo TBS podría usarse para identificar antagonistas funcionales de C/EBP alfa y BZLF1. C/EBP alfa es un factor de transcripción bZip que se une a los sitios CCAAT y la regulación positiva en las células humanas se asocia con el crecimiento del cáncer colorrectal, metástasis, e indica un mal resultado de supervivencia. BZLF1 es un factor de transcripción bZip que se une a los sitios TRE y CCAAT y está asociado con el linfoma de Burkitt, la enfermedad de Hodgkin, el linfoma no Hodgkin, el carcinoma nasofaríngeo y linfomas.

Para C/EBP alfa, se generaron plásmidos p230d que codificaban C/EBP alfa con la región básica de unión al ADN (C/EBPa incluida la región básica) o sin la región básica (C/EBPa menos la región básica). Luego, las células se sembraron en placas con p230d que expresaba cualquiera de estas construcciones C/EBP alfa junto con p300d que expresaba mDHFR CCAAT y un plásmido pET24a vacío. El experimento se llevó a cabo utilizando TMP 8 μ M y en ausencia y presencia de IPTG 1 mM (para inducir la expresión de las proteínas C/EBP alfa y DHFR).

Las células se transfectaron primero con CCAAT mDHFR, un plásmido que codifica C/EBPa incluida la región básica y un plásmido pET24a vacío en presencia de trimetoprima (TMP) 8 μ M pero sin IPTG. TMP inhibe la DHFR bacteriana, lo que resulta en que no haya colonias bacterianas (datos no mostrados). La adición de IPTG 1 mM induce la transcripción de las proteínas mDHFR y C/EBPa y dio como resultado la observación de un número muy bajo de colonias bacterianas (aproximadamente 8; datos no mostrados). Este resultado sugiere que el dominio de unión al ADN de C/EBPa fue capaz de unirse a los sitios de unión de CCAAT e inhibir la expresión de mDHFR, como se esperaba en base a resultados anteriores usando el ensayo TBS (véanse los Ejemplos 1 y 3). La expresión de mDHFR y C/EBPa menos la región básica da como resultado un gran número (>100) de colonias bacterianas. Esto sugiere que la ausencia de un dominio de unión al ADN impide que la proteína C/EBPa se una a los sitios CCAAT en CCAAT mDHFR, permitiendo la expresión de la proteína mDHFR y permitiendo la supervivencia de las células bacterianas.

El experimento reveló que la coexpresión de mDHFR CCAAT con la proteína alfa C/EBP que contiene la región básica dio como resultado un número sustancialmente menor de colonias en comparación con la placa que contenía células que expresan mDHFR CCAAT y C/EBP alfa que carecían de la región básica. Este resultado indica que la región básica de unión al ADN de la proteína alfa C/EBP fue capaz de unirse a los sitios de unión de CCAAT en mDHFR CCAAT e inhibir la producción de mDHFR, lo que sugiere que estas construcciones podrían usarse en el ensayo de TBS para identificar compuestos que sean capaces de inhibir la actividad de unión al ADN de la proteína alfa C/EBP.

Se llevó a cabo un experimento similar para evaluar si las construcciones descritas anteriormente podrían usarse para identificar compuestos capaces de inhibir la actividad de unión al ADN de BZLF1. Se generaron plásmidos p230d que codificaban BZLF1 con la región básica de unión al ADN (BZLF1 incluida la región básica) o sin la región básica (BZLF1 menos región básica). Luego, las células se sembraron con p230d que expresaba cualquiera de estas construcciones BZLF1 junto con p300d que expresaba mDHFR TRE y un plásmido pET24a vacío. El experimento se llevó a cabo utilizando TMP 8 μ M y en ausencia y presencia de IPTG 1 mM (para inducir la expresión de las proteínas BZLF1 y DHFR).

Los resultados demostraron que la coexpresión de mDHFR TRE con la proteína BZLF1 que contiene la región básica dio como resultado un número sustancialmente menor de colonias en comparación con la placa que contenía mDHFR TRE y BZLF1 que carecía de la región básica (datos no mostrados). Este resultado indica que la región básica de unión al ADN de la proteína BZLF1 fue capaz de unirse a los sitios de unión de TRE en mDHFR TRE e inhibir la producción de mDHFR, lo que sugiere que estas construcciones podrían usarse en el ensayo de TBS para identificar compuestos que sean capaces de inhibir la actividad de unión al ADN de la proteína BZLF1.

Resumiendo, esos resultados proporcionan evidencia de que el ensayo TBS se puede utilizar para identificar compuestos capaces de inhibir la actividad de unión al ADN de factores de transcripción distintos de AP-1, tales como los factores de transcripción C/EBP alfa y BZLF1. Esto demuestra además que el ensayo TBS se puede utilizar como método general para identificar antagonistas funcionales de proteínas de unión a ADN.

Bibliografía

Se citan anteriormente varias publicaciones con el fin de describir y divulgar más completamente la invención y el estado de la técnica al que pertenece la invención. Las citas completas para estas referencias se proporcionan a continuación.

Cabezas, E.; Satterthwait, A. C. (1999) *J. Am. Chem. Soc.*, 121, 3862.

Cody, V. *et al.* (2006) New insights into DHFR interactions: Analysis of *Pneumocystis carinii* and mouse DHFR complexes with NADPH and two highly potent 5-(omega-carboxy(alkyloxy) trimethoprim derivatives reveals conformational correlations with activity and novel parallel ring stacking interactions. *Proteins* 65(4): 959-969

de Araujo A.D. *et al.* (2014) Comparative α -helicity of cyclic pentapeptides in water. *Angew Chem Int Ed Engl.* 53(27):6965-9

Fujimoto, K. *et al.* (2008) Development of a series of cross-linking agents that effectively stabilize alpha-helical structures in various short peptides. *Chemistry* 14(3):857-63.

Gritz L y Davies J. (1983) Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene.* 25(2-3):179-88.

Haney, C.M. *et al.* (2011) Promoting peptide α -helix formation with dynamic covalent oxime side-chain cross-links. *Chem Commun (Camb).* 47(39):10915-7.

Holland-Nell, K.; Meldal, M. Maintaining biological activity by using triazoles as disulfide bond mimetics. *Angew Chem Int Ed Engl.* 50(22):5204-6.

Heinis, C. *et al.* (2009) Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol.* 5(7): 502-507

Jo, H. *et al.* (2012) Development of α -helical calpain probes by mimicking a natural protein-protein interaction. *J Am Chem Soc.* 134(42):17704-17713

Jochim, A.L.; Arora, P.S. (2010) Systematic analysis of helical protein interfaces reveals targets for synthetic inhibitors. *ACS Chem. Biol.*, 5, 919-923.

Kaufman *et al.* (1986) Selection and amplification of heterologous genes encoding adenosine deaminase in mammalian cells. *Proc Natl Acad Sci USA.* 83(10): 3136-3140.

Leduc, A.M. *et al.* (2003) Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor-coactivator interactions. *Proc Natl Acad Sci USA.* 100(20):11273-8

Ma. Y. *et al.* (2014) Split focal adhesion kinase for probing protein-protein interactions. *Biochemical Engineering Journal.* 90: 272-278

Mason, J. *et al.* (2006) Semirational design of Jun-Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc Natl Acad Sci USA.* 103(24): 8989-94.

McCormick, F. (2015) KRAS as a Therapeutic Target. *Clin Cancer Res.* 21(8): 1797-1801.

Mern, D.S. *et al.* (2010) Inhibition of Id proteins by a peptide aptamer induces cell-cycle arrest and apoptosis in ovarian cancer cells. *Br J Cancer.* 103(8): 1237-1244.

Miranda, E. *et al.* (2011) Deciphering interactions used by the influenza virus NS1 protein to silence the host antiviral sensor protein RIG-I using a bacterial reverse two-hybrid system. *Mol. Biosyst.*, 7, 1042-1045

Muppidi, A. *et al.* (2011) Achieving cell penetration with distance-matching cysteine cross-linkers: a facile route to cell-permeable peptide dual inhibitors of Mdm2/Mdmx. *Chem Commun (Camb)*. 47(33):9396-8. Newman & Keating (2003) Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*. 300(5628):2097-101

Olive, M. *et al.* (1997) A Dominant Negative to Activation Protein-1 (AP1) That Abolishes DNA Binding and Inhibits Oncogenesis. *J Biol Chem*. 272(30): 18586-94.

Park, J.H. (2007) Bacterial beta-lactamase fragmentation complementation strategy can be used as a method for identifying interacting protein pairs. *Journal of Microbiology and Biotechnology*. 17 (10): 1607-15.

Pelay-Gimeno, M. *et al.* (2015) Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angew Chem Int Ed Engl*. 54(31):8896-927

Pelletier J.N., Campbell-Valois F.X., Michnick S.W. (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A*. 95(21): 12141-6.

Soucek, L. *et al.* (2013) Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes. Dev*. 27(5): 504-13.

Raj, M. *et al.* (2013) Plucking the high hanging fruit: A systematic approach for targeting protein-protein interactions. *Bioorg. Med. Chem*. 21, 4051-4057.

Remy, I. *et al.* (2007) Detection of protein-protein interactions using a simple survival protein-fragment complementation assay based on the enzyme dihydrofolate reductase. *Nat Protoc*. 2(9): 2120-5. Robertson and Spring (2018) Using Peptidomimetics and Constrained Peptides as Valuable Tools for Inhibiting Protein-Protein Interactions. *Molecules* 23(4): 959.

Rodriguez-Martinez *et al.* (2017). Combinatorial bZIP dimers display complex DNA-binding specificity landscapes. *Elife*. 6 e19272

Ruan, F. *et al.* (1990) Metal ion-enhanced helicity in synthetic peptides containing unnatural, metal-ligating residues *J. Am. Chem. Soc.*, 112 (25): 9403-9404

Vaquerizas, J.M. *et al.* (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 10(4): 252-63

Vinson, C. *et al.* (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol*. 22(18):6321-35.

Walensky, L.D. *et al.* (2004) Activation of apoptosis *in vivo* by a hydrocarbon-stapled BH3 helix. *Science*. 305(5689):1466-70.

Wang, D. *et al.* (2005) Enhanced metabolic stability and protein-binding properties of artificial alpha helices derived from a hydrogen-bond surrogate: application to Bcl-xL. *Angew Chem Int Ed Engl*. 44(40):6525-9.

Wehr, M.C. y Rossner, M.J. (2016) Split protein biosensor assays in molecular pharmacological studies. *Drug Discov Today*. 21(3):415-29.

Wehr, M.C. *et al.* (2006) Monitoring regulated protein-protein interactions using split TEV. *Nat Methods*. 3(12):985-93.

Woolley, G.A. (2005) Photocontrolling peptide alpha helices. *Acc Chem Res*; 38(6):486-93.

Young, K.H. (1998) Yeast two-hybrid: so many interactions, (in) so little time... *Biol Reprod*. 58(2):302-11.

For standard molecular biology techniques, véase Sambrook, J., Russel, D.W. *Molecular Cloning, A Laboratory Manual*. 3ª ed. 2001, Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press

Anexo de secuencias

Secuencia de aminoácidos de la dihidrofolato reductasa murina de tipo salvaje (SEQ ID NO: 1)

MVRPLNCIVAVSQNMIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTFWFSIPEKNRPLKD
 RINIVLSRELKEPPRGHAFLAKSLDDALRLIEQPELASKVDMVWIVGGSSVYQEAMNQPGHLRLFVTRIM
 QEFESDTFFPEIDLKGYKLLPEYPGVLSEVQEEKGIKYKFEVYEKKD

Secuencia de aminoácidos de la dihidrofolato reductasa murina diseñada para contener sitios TRE (SEQ ID NO: 2)

MVRPLNCIVAVSQNMIGKNGDLPWPPLRNEFKYFQRMQTQDSVESKQNLVIMGRKTFWFSIPESNRPLKD
 RINIVLSQELKEPPRGHAFLAKSLDDALRLIESPELASKVDSVWIVGGSSVYQEAMNQPGHLRLFVTQIM
 QEFESDTFFPEIDSGKYKLLPESPGVLSQVQEEKGIKYKFEVYEKKD

Secuencia de aminoácidos de la dihidrofolato reductasa humana de tipo salvaje (SEQ ID NO: 3)

MVGSINNCIVAVSQNMIGKNGDLPWPPLRNEFRYFQRMTTTSSVEGKQNLVIMGKKTWFSIPEKNRPLKG
 RINIVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMVWIVGGSSVYKEAMNHPGHLKLFVTRIM
 QDFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIKYKFEVYEKND

Secuencia de ácido nucleico para la secuencia codificante de proteínas de la dihidrofolato reductasa murina diseñada para contener sitios TRE (SEQ ID NO: 4)

ATGGTTCGACCATTGAACTGCATCGTCGCCGTGAGTCAGAATATGGGGATTGGCAAGAACGGAGACCTACCTGGCC
 TCCGCTCAGGAATGAGTCAAAGTACTTCCAAAGAATGACTCAGACTGACTCAGTTGAGTCAAACAGAATCTGGTGA
 TTATGGGTAGGAAAACCTGGTTCTCCATTCTGAGTCAAATCGACCTTTAAAGGACAGAATTAATATAGTTCTGAGT
 CAAGAAGTCAAAGAACCACACGAGGAGCTCATTTTCTTGCCAAAAGTTTGGATGATGCCTTAAGACTTATTGAGTC
 ACCGGAATTGGCGAGCAAAGTTGACTCAGTTTGGATCGTCGGAGGCAGTTCTGTTTACCAGGAAGCCATGACTCAAC
 CAGGCCACCTTAGACTCTTTGTGACTCAGATCATGCAGGAATTTGAGTCAGACACGTTTTTCCAGAAATTGACTCA
 GGGAAATATAAACTTCTCCCTGAGTCACCAGGCGTCTGAGTCAGGTCCAGGAGGAAAAAGGCATCAAGTATAAGTT
 TGAAGTCTACGAGAAGAAAGACTAA

Secuencias de ácidos nucleicos de elementos de respuesta a TPA (TRE)

TGACTCA (SEQ ID NO: 5)
 TGAGTCA (SEQ ID NO: 6)

Secuencias de ácidos nucleicos de elementos de respuesta a la caja E

CACGTG (SEQ ID NO: 7)
 CACATG (SEQ ID NO: 8)

Secuencia de ácido nucleico del elemento de respuesta a proteínas C/EBP

ATTGCGCAAT (SEQ ID NO: 9)

Secuencia de ácidos nucleicos del elemento de respuesta al AMPc (CRE)

TGACGTCA (SEQ ID NO: 10)

Secuencias de ácidos nucleicos de elementos de reconocimiento Maf (MARE)

TGCTGA⁹/cTCAGCA (SEQ ID NO: 32)
 TGCTGA^{9c}/CGTCAGCA (SEQ ID NO: 33)

Secuencia de ácidos nucleicos del sitio de unión PAP/CREB-2/PAR

TTACGTAA (SEQ ID NO: 34)

Secuencia de ácido nucleico del polinucleótido que codifica la dihidrofolato reductasa murina diseñada que incluye sitios de enzimas de restricción (SEQ ID NO: 11)

GCTAGCGTTTCGACCATTGAACTGCATCGTCGCCGTGAGTCAGAATATGGGGATTGGCAAGAACGGAGACCTACCCTG
GCCTCCGCTCAGGAATGAGTCAAAGTACTTCCAAAGAATGACTCAGACTGACTCAGTTGAGTCAAAACAGAATCTGG
TGATTATGGGTAGGAAAACCTGGTTCTCCATTCTGAGTCAAATCGACCTTTAAAGGACAGAATTAATATAGTTCTG
AGTCAAGAACTCAAAGAACCACCACGAGGAGCTCATTTTCTGCCAAAAGTTTGATGATGCCTTAAGACTTATTGA
GTCACCGGAATTGGCGAGCAAAGTTGACTCAGTTTGGATCGTCGGAGGCAGTTCTGTTTACCAGGAAGCCATGACTC
AACCAGGCCACCTTAGACTCTTTGTGACTCAGATCATGCAGGAATTTGAGTCAGACACGTTTTTCCCAGAAATTGAC
TCAGGGAAATATAAACTTCTCCCTGAGTCACCGGCGTCTGAGTCAGGTCCAGGAGGAAAAAGGCATCAAGTATAA
GTTTGAAGTCTACGAGAAGAAAGACTAA

Secuencias de ácidos nucleicos de marcos de lectura de ejemplo

- 5 Ejemplo de marco de lectura 1: TGA CTC Axx (SEQ ID NO: 12)
Ejemplo de marco de lectura 2: xTG ACT CAX (SEQ ID NO: 13)
Ejemplo de marco de lectura 3: xxT GAC TCA (SEQ ID NO: 14)

Secuencia de aminoácidos del marco de lectura de ejemplo 3 (SEQ ID NO: 15)

10 FSYCLPHRITNVADG

Secuencia de aminoácidos de tripletes de codones de ejemplo que contienen TRE

- 15 GTGAGTCAG (SEQ ID NO: 16)
AATGAGTCA (SEQ ID NO: 17)
ATGACTCAG (SEQ ID NO: 18)
ACTGACTCA (SEQ ID NO: 19)
GTTGAGTCA (SEQ ID NO: 20)
20 CCTGAGTCA (SEQ ID NO: 21)
CTGAGTCAA (SEQ ID NO: 22)
ATTGAGTCA (SEQ ID NO: 23)
GTTGACTCA (SEQ ID NO: 24)
ATGACTCAA (SEQ ID NO: 25)
25 GTGACTCAG (SEQ ID NO: 26)
TTTGAGTCA (SEQ ID NO: 27)
ATTGACTCA (SEQ ID NO: 28)
CCTGAGTCA (SEQ ID NO: 29)
CTGAGTCAG (SEQ ID NO: 30)
30

Secuencia de aminoácidos de la dihidrofolato reductasa murina diseñada utilizada durante el proceso de diseño (SEQ ID NO: 31)

* = codón de terminación

- 35 ASVRPLNCIVAVSQNMGIGKNGDLPWPPLRNESKYFQRMQTQDSVESKQNLVIMGRKTFWFSIPESNRPLK
DRINIVLSQELKEPPRGAFHFLAKSLDDALRLIESPELASKVDSVWIVGGSSVYQEAQTQPGHLRLFTVQI
MQEFESDTFFPEIDSGKYKLLPESPGVLSQVQEEKGIKYKFEVYEKKD*A*

Secuencia de nucleótidos de un gen de dihidrofolato reductasa murino ilustrativo diseñado para incluir sitios de unión a CRE (SEQ ID NO: 36)

- 40 ATGGTTTCGACCATTGAACTGCATCGTCGCCGTGTCCTCCAAATATGGGGATTGGCAAGAACGGAGACCTAC
CCTGGCCTCCGCTCAGGAACGAGTTCAAGTACTTCCAAAGAATGACGTCAACCTCTTCAGTGGAAGGTAA
ACAGAATCTGGTGATTATGGGTAGGAAAACCTGGTTCTCCATTCTGAGAAGAATCGACCTTTAAAGGAC
AGAATTAATATAGTGACGTCAAGAGAACTCAAAGAACCACCACGAGGAGCTCATTTTCTTGCCAAAAGTT
TGGATGATGCCTTAAGACTTATTGAACAACCGGAATTGACGTCAAAAGTAGACATGGTTTGGATCGTCGG
AGGCAGTTCTGTTTACCAGGAAGCCATGAATCAACCAGGCCACCTTAGACTCTTTGTGACGTCAATCATG
CAGGAATTTGAAAGTGACACGTTTTTCCCAGAAATTGATTTGGGGAAATATAAACTTCTCCCAGAAATACC
CAGGCGTGACGTGAGAGGTCCAGGAGGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGA
CTAAGCTTAA

Secuencia de aminoácidos de una dihidrofolato reductasa murina ilustrativa diseñada para incluir sitios de unión a CRE (SEQ ID NO: 37)

45

MVRPLNCIVAVSQNMGIKNGDLFPWPLRNEFKYFQRMSTSSVEGKQNLVIMGRKTWFSIPEKNRPLKDRINIVTS
RELKEPPRGHFLAKSLDDALRLIEQPELTSKVDWVWIVGGSSVYQEAMNQPGHLRLFVTSIMQEFESDTFFPEIDL
GKYKLLPEYPGVTSEVQEEKGIKYKFEVYEKKD

Secuencia de nucleótidos de un gen de dihidrofolato reductasa murino ilustrativo diseñado para incluir sitios de unión a CCAAT (SEQ ID NO: 38)

5

ATGGTTCGACCATTGAACTGCATCGTCGCCGTGTCCCAAATATGGGGATTGGCAAGAACGGAGACCTACCCCTGGCC
TCCATTGCGCAATGAGTTCAGTACTTCCAAAGAATGACCACAACCTCTTCAGTGAAGGTAAACAGAATCTGGTGA
TTATGGGTAGGAAAACCTGGTTCTCCATTCTGAGAAGAATCGACCATTGCGCAATAGAATTAATATAGTTCTCAGT
AGAGAATTGCGCAATCCACCACGAGGAGCTCATTTTATTGCGCAATCCTTGGATGATGCATTGCGCAATATTGAACA
ACCGGAATTGGCGAGCAAAGTAGACATGGTTTGGATCGTCGGAGGCAGTTCTGTTTACCAGGAAGCCATGAATCAAC
CAGGCCACCTTAGACTCTTTGTGACAAGGATCATGCAGGAATTTGAAAGTGACACGTTTTTCCCAGAAATTGATTTG
GGGAAATATAAACTTCTCCAGAATACCCAGGCGTCTCTCTGAATTGCGCAATGAAAAAGGCATCAAGTATAAGTT
TGAACTCTACGAGAAGAAAGACTAAGCTTAA

Secuencia de aminoácidos de una dihidrofolato reductasa murina ilustrativa diseñada para incluir sitios de unión a CCAAT (SEQ ID NO: 39)

10

MVRPLNCIVAVSQNMGIKNGDLFPWPLRNEFKYFQRMSTSSVEGKQNLVIMGRKTWFSIPEKNRPLRNRINIVLS
RELNRNPPRGHFLIAQSLDDALRNIEQPELASKVDWVWIVGGSSVYQEAMNQPGHLRLFVTRIMQEFESDTFFPEIDL
GKYKLLPEYPGVLSVQEEKGIKYKFEVYEKKD

Secuencia de nucleótidos de un gen de dihidrofolato reductasa murino ilustrativo diseñado para incluir cajas E (SEQ ID NO: 40)

15

ATGGTTCGACCATTGAACTGCATCGTCGCCGTGTCCCAAATATGGGGATTGGCAAGAACGGAGACCTAC
CCTGGCCTCCGCTCAGGAACGAGTTCAGTACTTCCAAAGAATGACCACAACCTCTTCAGTGAAGGTAA
ACAGAATCTGGTGATTATGGGTAGGCGCACGTGGTTCTCCATTCTGAGAAGAATCGACCTTTAAAGGAC
AGAAATTAATATAGTTCTCTCACGTGAACCTCAAAGAACCACCACGTGGAGCTCACGTGCTTGCCAAATCAC
TGGATGATGCATTAAGACTTATTGAACAACCGGAATTGGCGTCACGTGTAGACATGGTTTGGATCGTCGG
AGGCAGTTCTGTTTACCAGGAAGCCATGAATCAACCAGGCCACGTGAGACTCTTTGTGACACGTGTCATG
CAGGAATTTGAAAGTGACACGTTTTTCCCAGAAATTGATTTGGGGAAATATAAACTTCTCCCAGAATACC
CAGGCGTCTCTCACGTGTCCAGGAGGAAAAAGGCATCAAGTATAAGTTTGAAGTCTACGAGAAGAAAGA
CTAAGCTTAA

20 Secuencia de aminoácidos de una dihidrofolato reductasa murina ilustrativa diseñada para incluir cajas E (SEQ ID NO: 41)

MVRPLNCIVAVSQNMGIKNGDLFPWPLRNEFKYFQRMSTSSVEGKQNLVIMGRRTWFSIPEKNRPLKDRINIVLS
RELKEPPRGHVLAKSLDDALRLIEQPELASRVDWVWIVGGSSVYQEAMNQPGHVRLFVTRVMQEFESDTFFPEIDL
GKYKLLPEYPGVLSRVQEEKGIKYKFEVYEKKD

25 Secuencia de nucleótidos del plásmido p300-mDHFR usado en los ejemplos (SEQ ID NO: 42)

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACA
 ATTTACACAGAATTCATTAAAGAGGAGAAATTAAGCATGCACCATCACCATCACCATgctagcgttcgaccattga
 actgcatcgtcgccgtgagtcagaatatggggattggcaagaacggagacctaccctggcctccgctcaggaatgag
 tcaaagtacttccaaagaatgactcagactgactcagttgagtcaaaacagaatctggtgattatgggtaggaaaac
 ctggttctccattcctgagtc aaatcgacctttaaaggacagaattaatatagttctgagtcagaactcaaagaac
 caccacgaggagctcattttcttgccaaaagtttgatgatgccttaagacttattgagtcaccggaattggcgagc
 aaagttgactcagtttgatcgctcgaggcagttctgtttaccaggaagccatgactcaaccaggccaccttagact
 ctttgtgactcagatcatgcaggaatttgagtcagacacgtttttccagaaattgactcagggaatatataaacttc
 tccctgagtcaccaggcgtcctgagtcaggtccaggaggaaaaggtcatcaagtataagtttgaagtctacgagaag
 aaagactaagcttAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTT
 GTTCAGAACGCTCGGTTGCCCGGGCGCTTTTATTGTTGAGAATCCAAGCTAGTTTGGGAGGTTCCAACCTTTCAC
 CATAATGAAATAAGATCACTACCGGGCGTATTTTTTGTAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGA
 GAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAGAACATTTTGAGGCATTTTCAGTCAG
 TTGCTCAATGTACCTATAACCAGACCGTTGAGCTGGATATTACGGCCTTTTTTAAAGACCGTAAAGAAAAATAAGCAC
 AAGTTTTATCCGGCCTTATTTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAGTTCGGTATGGCAATGAAAGA
 CGGTGAGCTGGTGATATGGGATAGTGTTCACCCCTTGTACACCGCTTTTCCATGAGCAAACCTGAAACGTTTTCATCGC
 TCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAAAC
 CTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCGATTT
 TGATTTAAACGTGGCCAATATGGACAACCTTCTCGCCCCGTTTTCACCATGGGCAAATATTATACGCAAGGCGACA
 AGGTGCTGATGCCGCTGGCGATTGAGGTTTCATCATGCCGTTTGTGATGGCTTCCATGTGGGCGAATGCTTAATGAA
 TTACAACAGTACTGCGATGAGTGGCAGGGCGGGCGTAATTTTTTAAAGCAGTTATTGGTGCCCTTAAACGCCTGG
 GGTAAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCTTTTCGTTTTATCTGT
 TGTTTGTGGTGAAACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGAGCTGCCTCGCGCTTTCGGTGATGACGGT
 GAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCG
 TCAGGGCGCGTTCAGCGGTGTTGGCGGGTGTGCGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGATA
 CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCG
 TAAGGAGAAAAATACCGCATCAGGCGCTCTTCCGCTTCTCTCGCTCACTGACTCGCTGCGCTCGGTCTCGGCTGCGG
 CGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAGAACATGTG
 AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTG
 ACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCC
 CCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGG
 AAGCGTGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTG
 TGCACGAACCCCCCGTTACGCCCCGACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACAC
 GACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
 GAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCG
 GAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAG
 ATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAA
 CTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT
 TTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
 GCGATCTGTCTATTTCTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGATATAACTACGATACGGGAGGGCTTACC
 ATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAG
 CCGCGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTGAGATCCAGTTCGATGTAACCCACTCGTGACCC
 CAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAA
 AGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTGAAGCATTTATCAGGGT
 TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCCG
 AAAAGTGCCACCTGACGTCTAAGAAACCATTTATTCATGACATTAACTATAAAAAATAGGCGTATCACGAGGCCCT
 TTCGTCTTAC

Secuencia de nucleótidos del plásmido p230d-cJun básico usado en los ejemplos (SEQ ID NO: 43)

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACA
 ATTTACACAGAATTCATTAAAGAGGAGAAATTAAGCATGCGCATTAAGCCGAACGCAAACGGATGCGCAACCGCA
 TCGCAGCCTCCAAGTGCCGCAAACGCAAATTGGAGCGCATCGCCCGCTTGAAGAAAAGGTGAAAACCTGAAAGCA
 CAGAACTATGAGCTGGCTCCACCGCCAACATGTTGCGCGAACAGGTGGCCAGCTCGGCGCGCCTCATCACCATCA
 CCATCACTGATAAAGCGCGCCTTGATAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCT
 CAGAACTCCATCTGGATTTGTTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTTGGTGAGAATCCAGGCGAGATT
 TTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAATCACTGGATATACACCGTTGATATATCCCAATGGCATCGTA
 AAGAACATTTTGAGGCATTTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTT
 TTAAAGACCGTAAGAAAAATAAGCACAGTTTATCCGGCCTTTATTCACATTCCTGCCCGCTGATGAATGCTCA
 TCCGGAATTTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTACCCCTTGTTACACCGTTTCC
 ATGAGCAAACCTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCG
 CAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGC
 CAATCCCTGGGTGAGTTTCACCAAGTTTGAATTAAACGTGGCCAATATGGACAACCTCTTCGCCCCCGTTTTACCA
 TGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTGAGGTTTCATCATGCCGTTTGTGATGGC
 TTCCATGTCGGCAGAAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGG
 CAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCG
 AAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGAGC
 TGCCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTA
 AGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCC
 AGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACCTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATA
 TGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCGCTTCTCGCTCACTGAC
 TCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
 AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG
 CGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
 GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGG
 ATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT

AGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTAT
CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
GGTGGTTTTTTTTTTTGTTCGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC
GGGGTCTGACGCTCAGTGGAAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
AGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAA
TGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTA
GATAACTACGATACGGGAGGGCTTACCATCTGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCACC GGCTC
CAGATTTATCAGCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATC
CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGC
TACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTCAGCTCCGGTCCCAACGATCAAGGCGAGTTA
CATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGTTTATGGCAGCACTGCATAATTCTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGAC
TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGG
ATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATTCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTAC
CAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA
TACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGATTTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCAT
TATTATCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCCTTTTCGTCTTCAC

Secuencia de nucleótidos de pREP4 que expresa el represor lac usado en los ejemplos (SEQ ID NO: 44)

AAGCTTCACGCTGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAG
AAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCA
GGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCCAG
CTGGGGCGCCCTCTGTTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGG
CGCAGGGGATCAAGATCTGATCAAGAGACAGGATGACGGTCTGTTTCGCATGCTTGAACAAGATGGATTGCACGCAGG
TTCTCCGGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCG
TGTTCCGGCTGTGACGCGAGGGGCGCCCGGTTCTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAG
GACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGC
GGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCATCTCACCTTGCTCCTGCCGAGAAAG
TATCCATCATGGCTGATGCAATGCCGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAA
CATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGG
GCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCTGACCCATGGCG
ATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCG
GACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCT
CGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGG
GACTCTGGGGTTGCAAAATGACCGACCAAGCGACGCCCCAACCTGCCATCACGAGATTTGATTCACCGCCGCCCTTCT
ATGAAAGGTTGGGCTTCGGAATCGTTTCCGGGACGCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAG
TTCTTCGCCACCCCGGGCTCGATCCCCTCGCGAGTTGGTTTCAGCTGCTGCCGTGAGGCTGGACGACCTCGCGGAGTT
CTACCGGCACTGCAAAATCCGTCGGCATCCAGGAAACAGCAGCGGCTATCCGCGCATCCATGCCCCGAACTGCAGG

AGTGGGGAGGCACGATGGCCGCTTTGGTCGACAATTCGCGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCT
 TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGCGGTTTGCCTATTGG
 GCGCCAGGGTGGTTTTCTTTTCCACAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCTGGCCCTGAGAGAGT
 TGCAGCAAGCGGTCCACGCTGTTTGGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACA
 TGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAAATGGCGC
 GCATTGCGCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG
 GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATA
 TTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTGCTGGTGAC
 CCAATGCGACCAGATGCTCCACGCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGG
 TCAGAGACATCAAGAAATAACGCCGGAACATTAGTGAGGAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGG
 ATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCAGCGCGCTTC
 GTTCTACCATCGACACCACCAGCTGGCACCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
 GCGTGACAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGGCCGCCAGTTGTTGTGCCACGCGGTT
 GGAATGTAATTCAGTCCGCCATCGCCGCTTCCACTTTTTCCCGGCTTTTCGAGAAACGTGGCTGGCCTGGTTCA
 CCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTACC
 ACCCTGAATTGACTCTCTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCAACGTA
 AATGCATGCCGCTTCGCCCTTCGCGCGCAATTGTGACCCCTGTCCCTCCTGTTTACGCTACTGACGGGGTGGTGCGTA
 ACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCA
 GTGAAGTGCTTCATGTGGCAGGAGAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCG
 CTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTTCCACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGAT
 TTCTTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAGCCGTTTTTCCATAGGCTCCGC
 CCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGC
 GTTTCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCTCCTGCTTTCCGTTTACCGGTGTCATTCCGCTGTTATGGCC
 GCGTTTGTCTCATTCCACGCTGACACTCAGTTCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCC
 CCCGTTCACTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAGACATGCAAAAGCACC
 ACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGG
 ACAAGTTTTGGTGAAGTGCCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAA
 ACCGCCCTGCAAGGGCGTTTTTTTCTGTTTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATC
 TTATTAATCAGATAAAATATTTCTAGATTTTCACTGCAATTTATCTCTTCAATGTAGCACCTGAAGTCAGCCCCATA
 CGATATAAGTTGTTAATTCTCATGTTTGACAGCTTATCATCGAT

Secuencia de nucleótidos del plásmido cJun de control que carece de la región básica de unión al ADN utilizada en los ejemplos (SEQ ID NO: 45)

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACA
 ATTTACACAGAATTCATTAAAGAGGAGAAATTAAGCATGCACCATCACCATCACCATGCTAGCATCGCCCGGCTGG
 AGGAAAAAGTGAAGACCTTGAAGGCCAGAACTATGAGCTGGCGTCCACGGCCAAACATGCTCCGGGAACAGGTGGCA
 CAGCTTGGCGCGCCTTAAGGTAGCTCTAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACC
 TCAGAACTCCATCTGGATTTGTTTCAAGACGCTCGGTTGCCGCCGGGCGTTTTTTTATTGGTGAGAATCCAGGCGAGAT
 TTTTCAAGAGCTAAGGAAGCTAAATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGT
 AAAGAACATTTTGAAGGATTTTCACTGAGTTGCTCAATGTACCTATAACCAGACCGTTTCACTGGATATTACGGCCTT
 TTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTACATTTCTGCGCGCCTGATGAATGCTT
 AATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAAC
 GCCTGGGGTAATGACTCTTAGCTTGAGGCATCAAATAAACGAAAGGCTCAGTCGAAAGACTGGGCCCTTCGTTTTT

ATCTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCCTCTAGAGCTGCCTCGCGCGTTTCGGTGAT
 GACGGTGAAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACA
 AGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAG
 TGTATACTGGCTTAACCTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACA
 GATGCGTAAGGAGAAAAATACCGCATCAGGCGCTCTTCGGCTTCTCGCTCACTGACTCGCTCGCGCTCGGTCTGTTCCG
 CTGCGGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAA
 CATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCC
 CCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACAGGCG
 TTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCCTTTCTCCC
 TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGG
 GCTGTGTGCACGAACCCCCGTTACGCCCCACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTA
 AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
 GTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
 CCTTCGGA AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAG
 CAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA
 CGAAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA AAAAT
 GAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCT
 ATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
 CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTACCGGCTCCAGATTTATCAGCAATAAACCC
 AGCCAGCCGGAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGG
 GAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTGGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACG
 CTCGTGCTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCA
 AAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGCAAGTAAGTTGGCCGAGTGTATCACTCATGCTTATG
 GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
 ATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCTCAATACGGGATAATACCGCGCCACATAGCA
 GAACTTTAAAAAGTGCTCATCATTTGGA AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
 AGTTCGATGTAACCCACTCGTGACCCCACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAA
 AACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCTCTTTTTC
 AATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAAACAA
 ATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCTAAGAAACCATTATTATCATGACATTAACTTA
 TAAAAATAGGCGTATCACGAGGCCCTTTTCGTCTTAC

Secuencia de nucleótidos del plásmido pET24a que contiene FosW (SEQ ID NO: 46)

ATCCGATATAGTTCCCTTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGTTATGCTAGTTAT
 TGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGT
 GTGCTCGAGTGCGGCCGAAGCTTGTGACGGAGCTCGAATTCGGATCCTTAAGGCGCGCCAGTTTCTCCAGCTGT
 TTCTGGAGGTCTTCGATCTCTTTGCGCAAGGCATAGTTGCGTCTTCCAGCTGTTCAATCTCGGCCTGCAGTTCATC
 GAGTTCCCTGTTTCGAGCTCTTCGGCCTCTTTTTTCCAATCTTCGTTCTCGCGGGCCAGTTCTTCGGCCCCGCTGTTCCA
 GGCTAGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTATCCGCTCACAATT
 CCCCTATAGTGAGTCGTATTAATTTTCGCGGGATCGAGATCTCGATCCTCTACGCCGGACGCATCGTGGCCGGCATCA
 CCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGG
 CTCATGAGCGCTTGTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCA
 TGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATA

AGGGAGAGCGTTCGAGATCCCGGACACCATCGAATGGCGCAAAACCTTTTCGCGGTATGGCATGATAGCGCCCGGAAGA
 GAGTCAATTTCAGGGTGGTGAATGTGAAACCACTAACGTTATACGATGTGCGCAGATATGCCGGTGTCTCTTATCAGA
 CCGTTTTCCCGCGTGGTGAACCAGGCCAGCCGCTTTCTGCGAAAACGCGGGAAAAAGTGGAGCGCGCATGGCGGAG
 CTGAATTACATTCCCAACCGCGTGGCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAG
 TCTGGCCCTGCACGCGCCGTGCAAAATTGTGCGGGCGATTAAATCTGCGCGCGATCAACTGGGTGCCAGCGTGGTGG
 TGTGATGGTAGAACGAAGCGCGTGAAGCCTGTAAAGCGGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGG
 CTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGCGCTIATT
 TCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGC
 ATCTGGTCGCATTGGGTCAACAGCAAATCGCCCTGTAGCGGGCCCATTAAGTTCTGTCTCGCGCGCTCTGCGTCTG
 GCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTC
 CGGTTTTCAACAAACCATGCAAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGG
 CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGAT
 ACCGAAGACAGCTCATGTTATATCCGCGCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACAGCGT
 GGACCGCTTGTGCAACTCTCTCAGGGCCAGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAA
 AAACCAACCTGGCGCCCAATACGCAAAACCGCTCTCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCAGCAG
 GTTTCGCCACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCGGGATCTC
 GACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGATGACATATCGTCGCGCGCACTT
 ATGACTGTCTTCTTTATCATGCAACTCTGACACAGGTGCCGGCAGCGCTCTGGGTCAATTTTCGGCGAGGACCGCTT
 TCGCTGGAGCGCAGATGATCGCCCTGTGCTTGGCGGTATTGGAATCTTGACGCGCCCTCGCTCAAGCCTTCCTCA
 CTGGTCCCGCCACCAACGTTTCGCGGAGAAGCAGGCCATTATCGCGGCGATGGCGGCCCCACGGGTGCGCATGATC
 GTGCTCCTGTGTTGAGGACCCGCTAGGCTGGCGGGTTGCCCTTACTGGTTAGCAGAATGAATCACCAGATACGCGA
 GCGAAGGTGAAGCGACTGCTGCTGCANACGTCTGGACCTGAGCAACACATGAATGGTCTTCGGTTTCGGTGTCTT
 CGTAAGTCTGGAACGCGGAGTCAAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGCTA
 CCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGAT
 CCATACCGCCAGTTGTTTACCTCACAACGTTCCAGTAACCGGGCATGTTTATCATCAGTAACCCGTATCGTGAGCA
 TCCTCTCTCGTTTCATCGGTATCATTAACCCCATGAACAGAAATCCCCCTTACACGAGGATCAGTGACCAAACAG
 GAAAAACCGCCCTTAACATGGCCCGCTTTATCAGAACCCAGACATTAACGCTTCTGGAGAACTCAACGAGCTGGA
 CGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCAGCTGATGAGCTTTACCGCAGCTGCCTCGCGCGTT
 TCGGTGATGACGGTGAACACCTCTGACACATCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGG
 AGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGCGGGCGCAGCCATGACCCAGTCACGTAGCGA
 TAGCGGAGTGATACTGGCTTAACATATGCGGCATCAGAGCAGATTGTAAGAGAGTGACCATATATGCGGTGTGAA
 ATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTC
 GGTGTTTCGGTTCGCGGAGCGGTATCAGCTCACTCAAGGCGGTAAACGGTTATCCACAGAAATCAGGGGATTAACG
 CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCCGCTTGTGCGGTTTTTCCAT
 AGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
 ATACCAAGCGCTTTCCCGCTGGAACCTCCCTCTGCGCTCTCTGTTCCGACCTCCCGCTTACCGGATACCTGTCCG
 CCTTTCTCCCTTCGGGAAGCGTGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTTCGGTGTAGGTGCTTCGC
 TCCAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCTTATCCGGTAACATCGTCTTGAGTC
 CAACCGGTAAAGACACGACTTATCGCCACTGCGAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGG
 GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTG
 AAGCCAGTTACCTTCGGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACACCGCTGGTAGCGGTGGTTTTTT
 TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACG
 CTCAGTCGAACCAAACTCACGTAAAGCGATTTTGGCTCATGAACAATAAACTCTCTGCTTACATAAACAGTAATAC

AAGGGGTGTTATGAGCCATATTCAACGGGAACGTCTTGCTCTAGGCCGCGATTAAATTC AACATGGATGCTGATT
TATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGAT
GCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTG
GCTGACGGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCCTG
CGATCCCCGGGAAAACAGCATTCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCA
GTGTTCTGCGCCGGTTGCATTGATTCTGTTTGTAAATTGTCCTTTTAACAGCGATCGCGTATTTCTGCTCGCTCA
GGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC
AAGTCTGGAAGAAATGCATAAACTTTTGCCATTCTCACCAGGATTCAGTCGTCACATCATGGTGATTTCTCACTTGAT
AACCTTATTTTGGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGA
TCTTGCCATCCTATGGAATGCCTCGGTGAGTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTG
ATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTCTAAGAATTAATTATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGAAATTG
TAAACGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTAAACCAATAGGCCGAAATC
GGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGAACAAGAGTCCACT
ATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCAC
CCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGAGCCCCGATTTAGAGCT
TGACGGGGAAAGCCGCGCAACGTGGCGAGAAAGGAAGGAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAG
GTGAGCGGTACGCTGCGCGTAACCAACACACCCGCGCGCTTAATGCGCGCTACAGGGCGCTCCCATTCGCCA

Secuencia de nucleótidos del plásmido pET24a que contiene cFos (SEQ ID NO: 47)

ATCCGGATATAGTTCTCTCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCAAGGGGTTATGCTAGTTAT
TGCTCAGCGGTGGCAGCAGCCAACCTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGT
GTGCTCGAGTGCGGCCCAAGCTTGTGACGAGCTCGAATTCGGATCCTTAAGGCGCGCCGAGTTTCTCTCTCC
TTCAGCAGGTTGGCAATTTCCGTCTGCAGGGCGTACTTCTCATCTTCCAGTTGGTCTGTCTCCGCTTGGAGTGTATC
AGTGCTAGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATATTTCTAGAGGGGAATTGTTATCCGCTCACAA
TTCCTCTATAGTAGTGTATTAATTTTCGCGGATCGAGATCTCGATCCTCTACGCCGACGCATCGTGGCCGGCAT
CACCGGCGCCACAGGTGCGGTTGCTGGCGCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCG
GGCTCATGAGCGCTTGTTCGGCGTGGGTATGGTGGCAGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTG
CATGCACCATTCCTTGGCGGCGGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCA
TAAGGGAGAGCGTCGAGATCCCGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCCCCGGA
GAGAGTCAATTCAGGGTGGTGAATGTGAAACAGTAACGTTATACGATGTGCGAGAGTATGCCGGTGTCTCTTATCA
GACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAAGCGCGATGGCGG
AGCTGAATTACATTCCCAACCGCTGGCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCC
AGTCTGGCCCTGCACGCGCGTTCGCAAATGTGCGGCGGATTAAATCTCGCGCGGATCAACTGGGTGCCAGCGTGGT
GGTGTGATGGTAGAACGAAGCGCGCTCGAAGCCTGTAAAGCGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTG
GGCTGATCATTAACATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTA
TTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGA
GCATCTGGTCGATTGGGTCAACAGCAAATCGCGCTGTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTC
TGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATG
TCCGGTTTCAACAAACCATGCAAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGAT
GGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACG
ATACCGAAGACAGCTCATGTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACAGC
GTGGACCGCTTGTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCGCTCTCACTGGTGAAAAG

AAAAACCACCCTGGCGCCCAATACGCAACCGCCTCTCCCCGCGC3TTGGCCGATTCAATTAATGCAGCTGGCAGCAG
 AGGTTTCCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAAT3TAAGTTAGCTCACTCATTAGGCACCGGGATC
 TCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCG3TGGGCGCGGGGCATGACTATCGTCGCCGCAC
 TTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCAATTTTCGGCGAGGACCGC
 TTTCGCTGGAGCGCAGATGATCGGCCTGTGCTTTCGGGTATTCGGAATCTTGACGCGCCTCGCTCAAGCCTTCGT
 CACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCCCACGGGTGCGCATGA
 TCGTGCTCCTGTGCTTGGAGACCCGGCTAGGCTGGCGGGGTTCCTTACTGGTTAGCAGAATGAATCACCGATACGC
 GAGCGAACGTGAAGCGACTGCTGCTGCAAAACGCTCTGCGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGT
 TTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCATTAAT3TTCGGATCTGCATCGCAGGATGCTGCTGGC
 TACCTGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCTGAGTGATTTTCTCTGGTCCCGCCGC
 ATCCATACCGCCAGTTGTTTACCCTCACAAACGTTCCAGTAACCGGGCATGTTCAATCATCAGTAACCCGTATCGTGAG
 CATCCTCTCTCGTTTCATCGGTATCATTACCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAAC
 AGGAAAAAACCGCCCTTAACATGGCCCGCTTATCAGAAGCCAGACATTAACGCTTCTGGAGAACTCAACGAGCTG
 GACGCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGCG
 TTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAGCGGATGCCG
 GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTCGGCGGT3TCGGGGCGCAGCCATGACCCAGTCACGTAGC
 GATAGCGGAGTGATATACTGGCTTAACATATGCGGCATCAGAGCAGATTGACTGAGAGTGACCATATATGCGGTGTG
 AAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCTCTTCGGCTTCCTCGCTCACTGACTCGCTGCGC
 TCGGTCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG3TAATACGGTTATCCACAGAATCAGGGGATAA
 CGCAGGAAAGAATATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTC
 ATAGGCTCCGCCCCCTGACGAGCATCACAAAATCAGCGCTCAA3TCAGAGETGGCGAAACCCGACAGGACTATAA
 AGATACCAAGCGCTTTCGCCCTGGAAGCTCCCTCGTGCGCTCTCCT3TTCGACCCCTGCCGCTTACCGGATACCTGTC
 CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGATCTCAGTTTCGGTGTAGGTCGTTTC
 GCTCCAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAG
 TCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT3GTAAACAGATTAGCAGAGCGAGGTATGTAGG
 CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC
 TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG3CAAAACAAACACCGCTGGTAGCGGTGGTTTT
 TTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGA
 CGCTCAGTGAACGAAACTCACGTTAAGGGATTTTGGTCATGAACAATAAACTGTCTGCTTACATAAACAGTAAT
 ACAAGGGGTGTTATGAGCCATATTCACGGGAAACGCTTGCTCTAGGCCCGCATTAATTCACATGGATGCTGA
 TTTATATGGGTATAAATGGGCTCGCGATAATGTGCGGAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCG
 ATGCGCCAGAGTTGTTTCTGAAACATGGCAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAAC
 TGGCTGACGGAATTTATGCCTCTTCGACCATCAAGCATTTTATCCGTAATCCTGATGATGCATGGTTACTCACCAC
 TCGCATCCCCGGGAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAAGTGAAAATATTGTTGATGCGCTGG
 CAGTGTTCTGCGCCGGTTCATTTCGATTTCCTGTTGTAATTGTCCTTTTACAGCGATCGCGTATTTCTGCTCGCT
 CAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGA
 ACAAGTCTGGAAGAAATGCATAAATTTTGCCATTCTCACCGGATTGATCCTCACTCATGGTGATTTCTCACTTG
 ATAACCTTATTTTTCAGCAGGGGAAATTAATAGGTTGTAATGATGTTGGACGAGTCGGAATCGCAGACCGATACCAG
 GATCTTGCCATCCTATGGAACGCTCGGTGAGTTTCTCCTTCATTACAGAAACGGCTTTTCAAAAATATGGTAT
 TGATAATCCTGATATGAATAAATGCAGTTTCATTTGATGCTCGATGAGTTTTCTAAGAATTAATTCATGAGCGGA
 TACATATTTGAATGATTTAGAAAAATAACAAATAGGGGTTCGCGCGACATTTCCCCGAAAAGTGCCACCTGAAAT
 TGTAACGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTAAACCAATAGGCCGAAA
 TCGGCAAAATCCCTTATAAATCAAAAGATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCA
 CTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATC
 ACCCTAATCAAGTTTTTTGGGGTTCAGGTGCGGTAAAGCACTAAATCGGAACCTAAAGGAGCCCCGATTAGAG
 CTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGGAAGGAGCGGGCTAGGGCGCTGGCA
 AGTGTAGCGGTACGCTGCGCGTAACCCACACCCCGCGCTTAATGCGCGCTACAGGGCGCGTCCCATTCGCC

A

ES 2 985 624 T3

Caja CAAT	GGCCAATCT	(SEQ ID NO: 48)
Caja CArG	CC(A/T ₆ GG	(SEQ ID NO: 49)
Caja E2	CAGGTG y CACCTG	(SEQ ID NO: 50 y 51)
Caja HY	TG(A/T)GGG	(SEQ ID NO: 52)
Caja T	TCACACCT	(SEQ ID NO: 53)
Caja TATA	TATAAA	(SEQ ID NO: 54)
Caja X	GTTGGCATGGCAAC	(SEQ ID NO: 55)
Caja Y	(A/G) CTAACC (A/G) (A/G) (C/T)	(SEQ ID NO: 56)
Caja ATA	AAATAT	(SEQ ID NO: 57)
Caja CGCG	(A/C/G) CGCG (C/G/T)	(SEQ ID NO: 58)
Caja DREB	TACCGACAT	(SEQ ID NO: 59)
Caja Fur	GATAATGATAATCATTATC	(SEQ ID NO: 60)
Caja G	GCCACGTGGC	(SEQ ID NO: 61)
Caja GCC	AGCCGCC	(SEQ ID NO: 62)
Caja H	ACACCA	(SEQ ID NO: 63)
Caja de prolamina	TGTAAAG	(SEQ ID NO: 64)
Caja de pirimidina	CCTTTT	(SEQ ID NO: 65)
Caja TACTAAC	ATTACTAAC	(SEQ ID NO: 66)

REIVINDICACIONES

1. Un método para detectar un antagonista de una proteína de unión al ADN, comprendiendo el método:

- 5 i) proporcionar una célula, en donde la célula comprende un compuesto de prueba, una proteína de unión al ADN y un casete de expresión de gen informador que codifica un producto de expresión de gen informador, en donde el casete de expresión de gen informador comprende al menos un sitio de unión para la proteína de unión a ADN de manera que la unión de la proteína de unión a ADN al sitio de unión inhibe la expresión del producto de expresión de gen informador; y
- 10 ii) determinar la expresión del producto de expresión de gen informador en presencia del compuesto de prueba; en donde un aumento en la expresión del producto de expresión de gen informador en presencia del compuesto de prueba indica que el compuesto de prueba es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN, y
- 15 en donde algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador.

2. El método de la reivindicación 1, en donde el producto de expresión de gen informador es una proteína informadora, opcionalmente en donde la proteína informadora es una proteína de supervivencia celular, una proteína de reproducción celular una proteína fluorescente, una proteína bioluminiscente, una proteasa, una enzima que actúa sobre un sustrato para producir una señal colorimétrica, una proteína cinasa, un activador de la transcripción o una proteína reguladora tal como ubiquitina.

20

3. El método de la reivindicación 2, en donde la proteína informadora es una proteína de supervivencia celular, opcionalmente en donde la proteína de supervivencia celular es una proteína de supervivencia celular exógena que es capaz de compensar una deficiencia en una proteína de supervivencia celular endógena; y

25 en donde el método se realiza en condiciones de selección tales que la supervivencia de la célula depende de la actividad de la proteína de supervivencia celular exógena.

4. El método de la reivindicación 3, en donde la proteína de supervivencia celular es dihidrofolato reductasa (DHFR), opcionalmente, en donde la DHFR tiene una secuencia de aminoácidos que es al menos un 80 % idéntica a la secuencia expuesta en la SEQ ID NO: 1.

30

5. El método de una cualquiera de las reivindicaciones anteriores, en donde:

- 35 i) el casete de expresión de gen informador comprende entre 1 y 5, entre 1 y 10, entre 1 y 15, entre 1 y 20, entre 5 y 10, entre 5 y 15, entre 5 y 20, entre 10 y 15, entre 10 y 20, entre 10 y 18 o entre 12 y 16 sitios de unión; y/o
- ii) algunos o todos los sitios de unión están ubicados en la secuencia codificante de proteínas del casete de expresión de gen informador.

40 6. El método de una cualquiera de las reivindicaciones anteriores, en donde la proteína de unión a ADN es un factor de transcripción o un fragmento de unión al ADN del mismo, opcionalmente en donde la proteína de unión al ADN es un factor de transcripción eucariota o un fragmento de unión a ADN del mismo.

45 7. El método de una cualquiera de las reivindicaciones anteriores, en donde la proteína de unión al ADN es un factor de transcripción de cremallera de leucina básica (bZIP), un factor de transcripción básico hélice-bucle hélice (bHLH), un factor de transcripción de cremallera de leucina bHLH (bHLH-ZIP), o un fragmento de unión al ADN del mismo, y opcionalmente en donde:

- 50 a) el al menos un sitio de unión es un elemento de respuesta a TPA (TRE) que tiene la secuencia de nucleótidos **TGACTCA (SEQ ID NO: 5)** o **TGAGTCA (SEQ ID NO: 6)**;
- b) el al menos un sitio de unión es un elemento de respuesta a la caja E que tiene la secuencia de nucleótidos **CACGTG (SEQ ID NO: 7)** o **CACATG (SEQ ID NO: 8)**;
- c) el al menos un sitio de unión es un sitio de unión CCAAT que tiene la secuencia de nucleótidos **ATTGCGCAAT (SEQ ID NO: 9)**;
- 55 d) el al menos un sitio de unión es un elemento de respuesta a AMPc (CRE) que tiene la secuencia de nucleótidos **TGACGTCA (SEQ ID NO: 10)**;
- e) el al menos un sitio de unión es un elemento de reconocimiento de Maf (MARE) que tiene la secuencia de nucleótidos **TGCTGA^{gc}/cTCAGCA (SEQ ID NO: 32)** o **TGCTGA^{gc}/ccTCAGCA (SEQ ID NO: 33)**; o
- 60 f) el al menos un sitio de unión es un sitio de unión a PAP/CREB-2/PAR que tiene la secuencia de nucleótidos **TTACGTAA (SEQ ID NO: 34)**.

8. El método de una cualquiera de las reivindicaciones anteriores, en donde:

- 65 i) la célula es una célula bacteriana, opcionalmente una célula de *Escherichia coli*; o
- ii) la célula es una célula eucariota, opcionalmente una célula de mamífero.

9. El método de la reivindicación 8 ii), en donde la célula eucariota se aisló de un paciente humano y en donde la proteína de unión al ADN es producida naturalmente por la célula, y opcionalmente en donde se sospecha o se sabe que la regulación de la proteína de unión al ADN está alterada en la célula.

5 10. El método de una cualquiera de las reivindicaciones 1 a 8, en donde el método comprende administrar un casete de expresión de la proteína de unión al ADN que codifica la proteína de unión al ADN para proporcionar la célula que comprende la proteína de unión al ADN.

10 11. El método de una cualquiera de las reivindicaciones anteriores, en donde el compuesto de prueba es un compuesto de prueba peptídico.

12. El método de la reivindicación 11, en donde el compuesto de prueba peptídico se expresa intracelularmente a partir de un casete de expresión del compuesto de prueba, opcionalmente, en donde el método comprende proporcionar el casete de expresión del compuesto de prueba a la célula.

13. El método de la reivindicación 11 o la reivindicación 12, en donde el método comprende administrar un agente de entrecruzamiento en la célula para introducir un entrecruzamiento entre dos restos de aminoácidos en una alfa-hélice del compuesto de prueba peptídico para producir un compuesto peptídico restringido por hélice.

14. El método de una cualquiera de las reivindicaciones 1 a 11, en donde el método comprende administrar el compuesto de prueba extracelularmente para proporcionar la célula que comprende el compuesto de prueba, opcionalmente en donde el compuesto de prueba es un compuesto de prueba peptídico, en donde el compuesto de prueba peptídico comprende un péptido restringido por hélice, y en donde el péptido restringido por hélice comprende un entrecruzamiento entre dos restos de aminoácidos.

15. Un método sin células para detectar un antagonista de una proteína de unión al ADN, comprendiendo el método:

i) poner en contacto un compuesto de prueba con una proteína de unión a ADN y un casete de expresión de gen informador que codifica un producto de expresión de gen informador, en donde el casete de expresión de gen informador comprende al menos un sitio de unión para la proteína de unión a ADN de manera que la unión de la proteína de unión a ADN al sitio de unión inhibe la expresión del producto de expresión de gen informador; y

ii) determinar la expresión del producto de expresión de gen informador, en donde un aumento en la expresión del producto de expresión de gen informador en presencia del compuesto de prueba indica que el compuesto de prueba es capaz de inhibir la actividad de unión al ADN de la proteína de unión al ADN, en donde el método se lleva a cabo fuera de una célula en un sistema *in vitro* que comprende los componentes necesarios para la expresión del producto de expresión de gen informador y

en donde algunos o todos los sitios de unión están ubicados en la secuencia transcrita del casete de expresión de gen informador.

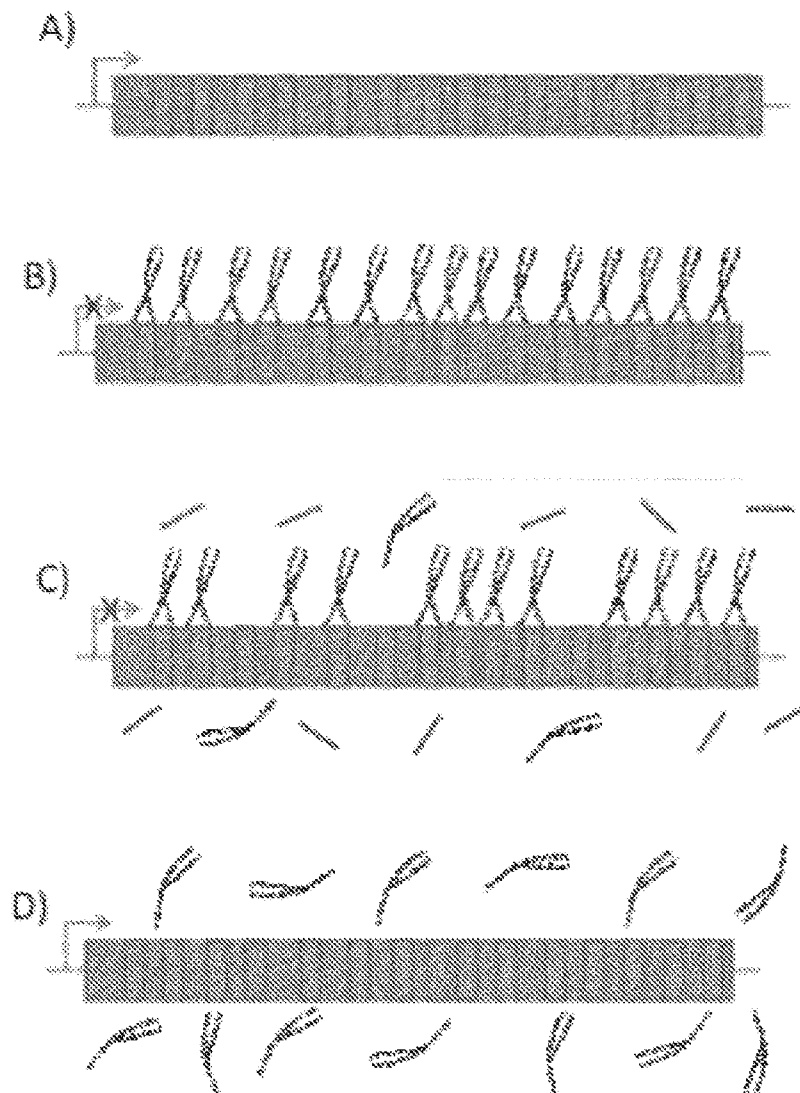


Figura 1

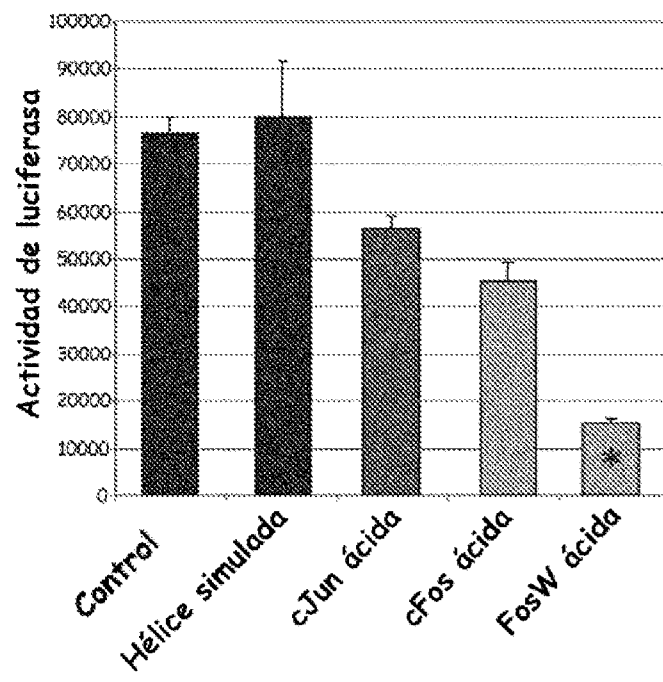


Figura 2

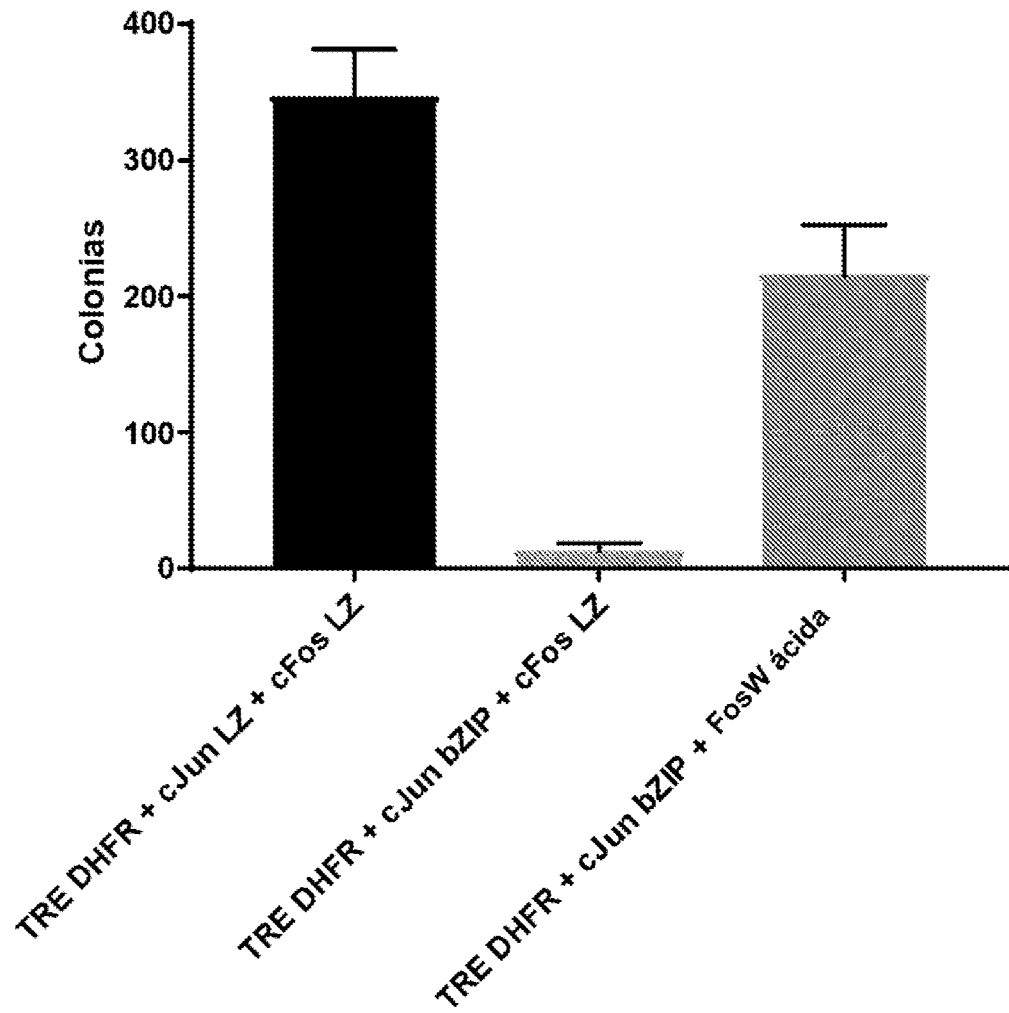


Figura 3