

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 646 388**

21 Número de solicitud: 201630607

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

11.05.2016

43 Fecha de publicación de la solicitud:

13.12.2017

56 Se remite a la solicitud internacional:

PCT/ES2017/070298

71 Solicitantes:

**CONSEJO SUPERIOR DE INVESTIGACIONES
CIENTÍFICAS (CSIC) (50.0%)**

C/ Serrano, nº 117

28006 Madrid ES y

UNIVERSIDAD AUTÓNOMA DE MADRID (50.0%)

72 Inventor/es:

VINCENT, Olivier

74 Agente/Representante:

PONS ARIÑO, Ángel

54 Título: **MÉTODO DE DOBLE HÍBRIDO EN REVERSO PARA LA IDENTIFICACIÓN DE MUTACIONES
MISSENSE**

57 Resumen:

Método de doble híbrido en reverso para la identificación de mutaciones missense.

La presente invención se refiere a un método de doble híbrido en reverso útil para la identificación de mutaciones de tipo missense, es decir, aquellas mutaciones de cambio de sentido que impiden la unión entre dos proteínas interaccionantes. En la invención se describen además las construcciones génicas útiles en el método de la invención, así como la célula huésped que comprende dichas construcciones y que se utiliza en el método de la invención.

ES 2 646 388 A1

Método de doble híbrido en reverso para la identificación de mutaciones
missense

DESCRIPCIÓN

5

La presente invención se refiere a un método de doble híbrido en reverso útil para la identificación de mutaciones de tipo *missense*, es decir, aquéllas mutaciones de cambio de sentido que impiden la unión entre dos proteínas interaccionantes. La presente invención describe además las construcciones génicas útiles en el método de la invención, así como la célula huésped que comprende dichas construcciones y que se utiliza en el método de la invención.

10

ESTADO DE LA TÉCNICA

15

El sistema de doble híbrido (Fields S, Song O. Nature. 1989;340(6230):245-6) es una técnica de referencia para analizar las interacciones proteína-proteína. Esta técnica ha sido adaptada a cribados de alto rendimiento ("*high-throughput*") permitiendo caracterizar el interactoma de numerosos organismos, como por ejemplo el interactoma humano (Rolland T. *et al.*, Cell. 2014;159(5):1212-26). Actualmente, dos empresas, Clontech (USA) e Invitrogen (Life Technologies, Thermofisher, USA), comercializan kits del sistema de doble híbrido clásico.

20

25

El sistema de doble híbrido clásico ha derivado en el sistema de doble híbrido en reverso (White MA. Proc Natl Acad Sci U S A. 1996;93(19):10001-3). Esta metodología permite la selección de mutaciones que anulan una interacción definida entre un par de proteínas interaccionantes y, en consecuencia, permite también la identificación de los aminoácidos implicados en dicha interacción. La detección de las regiones, y más específicamente, de los aminoácidos concretos implicados en las interacciones proteína-proteína, es necesaria para definir la relevancia fisiológica y las bases moleculares de dichas interacciones. La técnica de doble híbrido en reverso se utiliza también para identificar moléculas que interfieren con dicha interacción proteína-proteína (Vidal M, Endoh H. Trends Biotechnol. 1999;17(9):374-81). Aunque el potencial del sistema de doble híbrido en reverso es evidente, dicho sistema no ha tenido tanto éxito como su predecesor. La razón principal de esta diferencia es que el sistema de doble híbrido en reverso genera un número demasiado alto de falsos positivos que corresponden a mutaciones que truncan la proteína, no a mutaciones de

30

35

cambio de sentido o de tipo *missense*, que son las que interesan al impedir la unión entre dos proteínas interaccionantes, de hecho, >97% de las mutaciones al azar generadas por PCR producen un truncamiento en la proteína.

5 Teniendo en cuenta el alto número de falsos positivos del sistema de doble híbrido en reverso se han propuesto diferentes mejoras para identificar únicamente las mutaciones *missense*, aunque ninguna de dichas mejoras ha conseguido resolver de forma definitiva el problema del alto número de falsos positivos detectados. La primera de dichas mejoras fue dirigida al uso de fusiones C-terminal a la proteína verde
10 fluorescente (GFP-“*Green Fluorescent Protein*”), para seleccionar los mutantes no truncados mediante análisis de la fluorescencia (Endoh H, *et al.*, *Methods Enzymol.* 2000;328:74-88). Sin embargo, la proteína GFP puede interferir en la interacción entre las proteínas a estudiar y además, el hecho de que la selección mediante fluorescencia se haga a posteriori, incrementa mucho la complejidad de la técnica.

15

Para evitar el uso de proteínas de tipo GFP y evitar la posible interferencia con la interacción analizada, se propuso utilizar fusiones a epítomos cortos reconocidos por anticuerpos (revisado en Bennett MA. *et al.* *Methods Mol Biol.* 2015;1278:433-46). De nuevo, este sistema es demasiado largo y complejo ya que requiere validar todos los
20 mutantes mediante la técnica de Western blot. Más recientemente, una fusión a GBD (*Gal4 Binding Domain*) ha sido utilizada en un sistema denominado “one-plus two-hybrid system” (Kim JY, *et al.*, *Mol Cell Proteomics.* 2007;6(10):1727-40; Kim JY *et al.*; *Methods Mol Biol.* 2012;812:209-23). En este último caso, el sistema utiliza dos marcadores o reporteros que son reconocidos por dos dominios de unión al ADN
25 distintos, al igual que en estudios previos (Jiang R, *et al.* *Genes Dev.* 1996;10(24):3105-15; Inouye C. *et al.* *Genetics.* 1997;147(2):479-92) en los cuales el sistema de doble híbrido fue adaptado para identificar mutaciones que anulan la interacción de una proteína con otra pero no con una tercera. En este sistema, la selección se hace en dos pasos diferentes pero con las mismas células transformadas
30 que comprenden y expresan los reporteros y las proteínas a estudiar. Sin embargo, al igual que la proteína GFP, los dominios de unión al ADN como GBD son dominios estructurales que pueden interferir con la interacción que se está analizando. Este mismo problema tampoco ha sido superado por el método descrito por el equipo de Solomon MJ, que utiliza fusiones que comprenden el gen que codifica la proteína a
35 analizar unida en su extremo C-terminal al gen *URA3* para detectar las mutaciones que truncan la proteína estudiada y en consecuencia no permiten la síntesis y

expresión de la proteína de fusión formada por la proteína a analizar y Ura3. Específicamente, se clona el gen *URA3* en fase con el gen que codifica la proteína a estudiar de tal forma que cuando dichos genes se expresan, se produce una proteína de fusión o proteína quimérica (proteína a estudiar-Ura3). La presencia de una mutación de truncamiento en la proteína estudiada impide la expresión y síntesis de la fusión de dicha proteína a Ura3, de tal forma que una levadura que comprenda el gen *URA3* endógeno inactivado no crecerá en un medio de cultivo en ausencia de uracilo (Burton JL. *et al.* EMBO J. 2011 May 4;30(9):1818-29; Lickfeld M. *et al.* Yeast. 2011 Jul;28(7):535-45).

Otro de los métodos alternativos descritos (Gray PN, *et al.* Mol Cell Proteomics. 2007 Mar;6(3):514-26; WO2006060595), y comercializados en forma de kit por Invitrogen, propuso la expresión de una proteína de fusión que comprende el gen mutado de una de las proteínas del par interaccionante a estudiar unido en el extremo C-terminal a un marcador de resistencia a antibióticos, por lo que era necesaria una doble selección en bacterias y levaduras dotando a la técnica de una gran complejidad metodológica. Dicha tecnología no consiguió superar las carencias antes mencionadas, ya que mediante dicha técnica se seguían detectando al menos un 50% de falsos positivos correspondientes a proteínas no mutadas.

En este sentido, sigue existiendo en el estado de la técnica la necesidad de sistemas de doble híbrido en reverso para la detección exclusiva de mutaciones *missense* que anulan la unión entre proteínas interaccionantes, donde dichos sistemas sean de una simplicidad técnica tal que permita su uso rutinario tanto en clínica como en investigación, y además, sin dar lugar a la detección de falsos positivos.

DESCRIPCIÓN DE LA INVENCION

La presente invención soluciona el problema de la técnica mencionado previamente mediante un nuevo sistema de doble híbrido en reverso capaz de detectar exclusivamente las mutaciones *missense* que impiden la unión entre un par de proteínas interaccionantes de las que se desee conocer las regiones concretas implicadas en su unión. El sistema de doble híbrido en reverso de la invención utiliza un plásmido que comprende la secuencia nucleotídica que codifica para un péptido capaz de interaccionar con una proteína heteróloga expresada desde una construcción génica que se ha integrado en el genoma de la célula utilizada en el

sistema de la invención, y donde la secuencia nucleotídica que codifica para dicho péptido está fusionada a la secuencia nucleotídica que codifica para una de las proteínas interaccionantes sometidas a estudio. La transformación de la célula utilizada en el método de doble híbrido en reverso de la invención con el plásmido que codifica para el péptido que se une a dicha proteína heteróloga expresada desde una construcción génica integrada en el genoma de dicha célula, junto con un sistema dual de genes reporteros de selección positiva y contra-selección, permite la identificación positiva e inequívoca de las mutaciones *missense* que impiden la interacción entre el par de proteínas a estudiar.

10

La utilización del péptido mencionado anteriormente en lugar de proteínas o dominios proteicos, tal y como se utiliza actualmente en el estado de la técnica para la detección de mutaciones *missense*, tiene la ventaja de reducir una posible interferencia de la fusión en los extremos C-terminal de las proteínas del par interaccionante a estudiar.

15

Adicionalmente, otra de las ventajas del método de doble híbrido en reverso de la invención es la identificación de los mutantes *missense* en una única etapa, mediante el uso de la técnica de PCR mutagénica unida a la técnica de recombinación *in vivo* ("*gap-repair*") y el sistema dual de reporteros que permite la selección simultánea de mutaciones que son *missense* y que provocan la pérdida de interacción entre el par de proteínas a estudiar.

20

Tal y como se muestran en los ejemplos incluidos en el presente documento, el sistema de doble híbrido en reverso de la invención para la identificación de mutaciones *missense* entre un par de proteínas interaccionantes sometidas a estudio, además de ser un sistema rápido y sencillo, es un sistema muy eficiente, llegando a identificar correctamente el 100% de las mutaciones *missense* obtenidas, sin dar lugar a falsos positivos.

30

Por tanto, la presente invención proporciona un método de doble híbrido en reverso capaz de identificar mutaciones *missense* que impiden la unión entre dos proteínas interaccionantes en una única etapa. La invención también presenta células, preferentemente cepas de levadura y varias construcciones génicas que son útiles para la identificación de las mutaciones que inhiben las interacciones moleculares entre las proteínas a estudiar mediante el método descrito aquí.

35

Método de la invención

5 En un primer aspecto, la presente invención se refiere a un método *in vitro* para identificar al menos una mutación en una proteína de referencia donde dicha mutación afecta a la capacidad de dicha proteína de referencia de unirse a otra proteína diana, donde dicho método comprende:

a) al menos una célula huésped que comprende:

10 i) Una primera secuencia nucleotídica que codifica para un gen reportero, donde dicha secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN,

15 ii) Una segunda secuencia nucleotídica que codifica para un segundo gen reportero, donde dicha segunda secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN, con la condición de que dicha secuencia de reconocimiento es diferente de la secuencia de reconocimiento de i), y

20 iii) Una tercera secuencia nucleotídica que codifica para una primera proteína de fusión que comprende el dominio de unión al ADN de ii) y una proteína heteróloga capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia de estudio, donde la secuencia nucleotídica que codifica para dicha primera proteína de fusión está operativamente unida a un promotor,

25 b) Pre-transformar la célula de la etapa a) con un plásmido que comprende la secuencia nucleotídica que codifica para una segunda proteína de fusión que comprende el dominio de unión al ADN de i) y la proteína diana, donde la secuencia nucleotídica que codifica para dicha segunda proteína de fusión está operativamente unida a un promotor,

30 c) Cultivar la célula pre-transformada de la etapa b) bajo condiciones que permitan exclusivamente el crecimiento de las células que hayan incorporado el plásmido de la etapa b),

35

- 5 d) Transformar la célula de la etapa c) con un vector linearizado y al menos un fragmento de ADN que previamente ha sido sometido a mutagénesis y que comprende la secuencia nucleotídica que codifica para la proteína de referencia con al menos una mutación, donde entre el vector linearizado y el fragmento de ADN sometido a mutagénesis se produce recombinación homóloga obteniéndose un plásmido que comprende la secuencia nucleotídica que codifica para una tercera proteína de fusión que comprende el dominio de transactivación de Gal4 y la proteína de referencia con al menos una mutación, donde la secuencia nucleotídica que codifica para dicha tercera proteína de fusión está operativamente unida a un promotor,
- 10
- e) Cultivar la célula de la etapa d) bajo condiciones tales que permitan exclusivamente el crecimiento de las células que presentan mutaciones *missense* que impiden la unión entre la proteína de referencia y la proteína diana,
- 15
- f) Comparar la secuencia de la proteína de referencia mutada con la secuencia de la proteína de referencia sin mutar (nativa o *wild-type*) e identificar la mutación que impide la unión de la proteína de referencia con la proteína diana.
- 20

A efectos de la presente invención, los términos proteína, o molécula de ADN “de referencia”, se refiere a aquélla molécula capaz de unirse o asociarse de forma transitoria o constitutiva con otra proteína, o molécula de ADN a la que a efectos de la presente invención se le denomina “diana”. En los ejemplos que se muestran en el presente documento para poner de manifiesto la validez del método de la invención se ha utilizado como proteína de referencia la proteína humana glucoquinasa (GK) de SEQ ID NO: 10, codificada por la secuencia nucleotídica SEQ ID NO: 9, y como proteína diana a la que se une la proteína de referencia, se ha utilizado la proteína reguladora de la glucoquinasa humana (GKRP) de SEQ ID NO: 2, codificada por la secuencia nucleotídica SEQ ID NO: 1. Mediante el método de la invención se puede utilizar cualquier par de proteínas interaccionantes de las que se quiera analizar la presencia de mutaciones *missense* que impidan su interacción.

25

30

35 A efectos de la presente invención, el término “dominio de unión al ADN” se entiende como un conjunto de amino ácidos que es capaz de dirigir la unión específica de un

polipéptido a una secuencia particular de ADN, por ejemplo a una secuencia de reconocimiento para un factor de transcripción que interactúa específicamente con una secuencia de ácido nucleico en el promotor de un gen. Alternativamente, el dominio de unión al ADN puede ser cualquier dominio proteico que interactúa específicamente con una secuencia que se puede producir de forma natural o se inserta artificialmente en el promotor de un gen reportero. El dominio de unión al ADN puede estar unido covalentemente a un dominio de transactivación, de manera que la unión de la proteína al ADN en un sitio localizado dentro del promotor de un gen reportero elegido, activa su transcripción.

10

A efectos de la presente invención, existe una gran variedad de dominios de unión al ADN y de dominios de transactivación que son adecuados para su uso en los diversos aspectos de la invención. En general, cualquier dominio de unión al ADN o cualquier dominio de transactivación de cualquier factor de transcripción puede ser utilizado en la presente invención. El dominio de unión al ADN y el dominio de transactivación pueden pertenecer o no a distintos factores de transcripción. Secuencias de reconocimiento útiles incluyen, sin ser limitativos, los sitios de unión de los factores de transcripción Gal4 y Ace1 de levadura y LexA de bacteria. Estos sitios de unión pueden utilizarse fácilmente con un promotor reprimido (por ejemplo, los promotores SPAL, SPEX y ESPACE combinan un promotor *SP013* con las secuencias de reconocimiento para Gal4, Ace1 y LexA, respectivamente). Otros factores de transcripción útiles incluyen la proteína Gcn4 de *S. cerevisiae* (véase, por ejemplo, Hope and Struhl, 1986, Cell 46: 885-894) y la proteína Adr1 de *S. cerevisiae* (véase, por ejemplo, Kumar *et al*, 1987, Cell 51: 941-951). La secuencia de reconocimiento debe incluir al menos un sitio de unión para el dominio de unión al ADN del factor de transcripción que se utiliza. El número de sitios de unión se puede ajustar para proporcionar mayor o menor sensibilidad al ensayo.

15

20

25

30

Por "secuencia de reconocimiento" se entiende un segmento de ADN que es necesario y suficiente para interactuar específicamente con un polipéptido dado, tal como por ejemplo, el dominio de unión al ADN de un factor de transcripción.

35

Por "unido operativamente" se entiende que un gen y una secuencia reguladora (s) (por ejemplo, un promotor) están conectados de tal manera que permiten la expresión génica cuando las moléculas apropiadas (por ejemplo, proteínas que incluyen dominios que activan la transcripción) se unen a las secuencias reguladoras.

Por “unido covalentemente” se entiende cuando dos moléculas, por ejemplo proteínas, están unidas directamente por uniones covalentes. Por ejemplo, proteínas o dominios proteicos unidos covalentemente se pueden localizar inmediatamente contiguos, o pueden localizarse separados por residuos de uno o más aminoácidos dentro de la misma proteína híbrida.

Tal y como se utiliza en la presente invención, el término “gen reportero” utilizado a lo largo del documento, se refiere a un gen cuya expresión puede analizarse como una medida de la capacidad de dos moléculas de ensayo para interactuar (es decir, como una medida de las interacciones proteína/proteína, proteína/ARN, ARN/ARN o proteína/ADN). Los genes reporteros descritos en este documento pueden localizarse en un plásmido o pueden integrarse en el genoma de una célula haploide o diploide. El gen reportero se une preferentemente de manera operativa a un promotor que tiene secuencias que dirigen la transcripción del gen reportero. El gen reportero está posicionado de tal manera que se expresa cuando un dominio de transactivación de un factor de transcripción se pone en estrecha proximidad con el gen (por ejemplo, mediante el uso de proteínas híbridas para reconstituir un factor de transcripción, o por unión covalente del dominio de transactivación a un dominio de unión al ADN). El gen reportero también puede estar unido operativamente a secuencias reguladoras que le hacen muy sensible a la presencia o ausencia de un factor de transcripción. Por ejemplo, en ausencia de un factor de transcripción específico, el gen reportero *URA3* unido operativamente a un promotor altamente sensible confiere un fenotipo $Ura^- Foa^r$ en la célula. En presencia de un factor de transcripción específico, esta construcción génica confiere un fenotipo $Ura^+ Foa^s$ en la célula. Métodos conocidos por el experto medio en el presente campo técnico pueden ser utilizados para conectar un gen reportero a un promotor y para introducir el gen reportero en una célula.

A efectos de la presente invención, un gen reportero útil tiene en su promotor una secuencia de reconocimiento para el dominio de unión al ADN de un factor de transcripción que puede ser reconstituido mediante la interacción de una proteína con un dominio de unión al ADN y otra proteína con un dominio de transactivación. Tales genes incluyen, sin limitación, *lacZ*, genes biosintéticos aminoacídicos tales como por ejemplo, los genes de levadura *LEU2*, *HIS3*, *LYS2*, *LYS5*, o *TRP1*, el gen *URA3*, genes biosintéticos de ácidos nucleicos, el gen que codifica para la cloranfenicol-transacetilasa bacteriana (*cat*), y el gen bacteriano *gus*. También se incluyen los

genes que codifican marcadores fluorescentes, tales como el gen de la proteína fluorescente verde (GFP). Ciertos genes reporteros son considerados como genes reporteros "seleccionables" o "contra-seleccionables" tal y como se describe a continuación.

5

En una realización más preferida, el método de doble híbrido en reverso según se describe en la presente invención se caracteriza por que la célula utilizada en dicho método comprende al menos dos genes reporteros, donde el primero de dichos genes es un gen de selección positiva y el segundo de dichos genes es un gen de contra-selección.

10

Por marcador "seleccionable" se entiende un gen que, cuando se expresa, confiere una ventaja de crecimiento sobre una célula que lo contiene. Ejemplos de marcadores seleccionables incluyen, sin limitación, *LEU2*, *TRP1*, e *HIS3*. Ciertos marcadores seleccionables descritos en este documento pueden usarse para promover el crecimiento de células que comprenden un plásmido que expresa este marcador seleccionable. En este caso, el promotor unido operativamente a este marcador seleccionable es el promotor natural para el marcador. Por otra parte, el marcador puede ser diseñado para ser unido operativamente a un promotor distinto de aquel al que está naturalmente ligado. Así, en el caso de los genes reporteros utilizados en esta invención, el marcador seleccionable está unido operativamente a un promotor con una secuencia de reconocimiento para un factor de transcripción que puede ser reconstituido mediante la interacción de una proteína con dominio de unión al ADN y otra proteína con dominio de transactivación.

15

20

25

Por marcador "contra-seleccionable" se entiende un gen que, cuando se expresa, impide el crecimiento de la célula que lo contiene. Ejemplos de marcadores contra-seleccionable incluyen *URA3*, *LYS2*, *GAL1*, *CYH2* y *CAN1*.

30

Por gen reportero "seleccionable" se entiende un gen reportero que, cuando se expresa bajo un determinado conjunto de condiciones, confiere una ventaja de crecimiento sobre las células que lo contienen. Ejemplos de genes reporteros seleccionables incluyen los marcadores *LEU2*, *TRP1*, e *HIS3*. En la presente invención, el gen reportero seleccionable preferido es el gen *HIS3*.

35

Por gen reportero "contra-seleccionable" se entiende un gen reportero que, cuando se expresa bajo un determinado conjunto de condiciones, impide el crecimiento de una célula que lo contiene. Ejemplos de genes reporteros contra-seleccionables incluyen los marcadores *URA3*, *LYS2*, *GAL1*, *CYH2* y *CAN1*. En la presente invención, el gen reportero contra-seleccionable preferido es el gen *URA3*.

Por gen reportero "detectable" se entiende un gen cuya expresión puede ser detectada en una célula por un medio distinto al de conferir una ventaja de crecimiento selectivo en una célula. Un ejemplo de un gen reportero detectable es el gen *lacZ*. Si se desea, un gen reportero detectable puede ser integrado en el genoma de una célula, preferentemente de una célula de levadura. Un gen reportero detectable se puede utilizar en la invención para medir la capacidad de dos moléculas de interactuar. Así, el promotor que está unido operativamente a un gen reportero detectable debe contener una secuencia de reconocimiento para un factor de transcripción que puede ser reconstituido mediante la interacción de una proteína con dominio de unión al ADN y otra proteína con dominio de transactivación.

Preferiblemente, cada uno de los genes reportero esta operativamente unido a un promotor que porta una secuencia represora que impide la transcripción en ausencia de un motivo de activación génica. Por lo tanto, el gen reportero debe colocarse de tal manera que su expresión sea altamente sensible a la presencia o ausencia de un factor de transcripción. Por ejemplo, se prefiere que cuando se utiliza el gen reportero que codifica para el alelo *URA3*, dicho alelo confiere un fenotipo $Ura^- Foa^+$ en ausencia del factor de transcripción, y un fenotipo $Ura^+ Foa^s$ en su presencia. Ciertos promotores, tales como el promotor *SP013*, contienen de forma natural una secuencia represora aguas arriba. Otros promotores pueden ser diseñados o modificados utilizando métodos convencionales de clonación, para que comprendan dichas secuencias. Cuando se utiliza un gen reportero de contra-selección, la expresión del gen puede ser detectada mediante la detección de la inhibición del crecimiento celular. Cuando se emplea más de un gen reportero, los genes reporteros pueden estar unidos operativamente a promotores que son idénticos entre sí sólo en sus secuencias de reconocimiento. Preferiblemente, el gen reportero es uno que permite la selección titulable; por lo tanto, el crecimiento celular se puede medir en un rango de condiciones (por ejemplo, concentraciones de 5-FoA).

35

Marcadores "*contra-seleccionables*": Mientras que los marcadores seleccionables se han utilizado para, bajo ciertas condiciones, promover el crecimiento de sólo aquellas células que expresan los marcadores seleccionables, el marcador *contra-seleccionable* se han utilizado, bajo ciertas condiciones, para promover el crecimiento de sólo aquellas células que no expresan el marcador *contra-seleccionable*. Los marcadores *contra-seleccionables* cuando están presentes en los plásmidos se pueden utilizar para seleccionar las células que han perdido el plásmido. Por ejemplo, la expresión del gen *URA3*, que codifica para la orotidina-5'-fosfato, es letal en presencia de un medio que contiene ácido 5-fluoro-orótico (5-FOA). Las células que expresan *URA3* también se pueden seleccionar positivamente haciéndolas crecer en un medio sin uracilo. Por lo tanto, dependiendo de las condiciones de crecimiento, el marcador *URA3* se puede utilizar ya sea para seleccionar de manera positiva o negativa. El gen *LYS2*, que codifica la α -aminoadipato reductasa, también se puede utilizar para *contra-selección*. Las células de levadura que expresan *LYS2* no crecen en un medio que contiene α -aminoadipato como fuente de nitrógeno primario. Del mismo modo, la expresión de *LYS5* en un medio que contiene α -aminoadipato es letal. Estos genes, que están implicados en la biosíntesis de lisina, se pueden seleccionar de una manera positiva en un medio libre de lisina. Otro marcador *contra-seleccionable* es el gen *CAN1* que codifica una permeasa de arginina. La expresión de este gen en ausencia de arginina y en presencia de canavanina es letal. Del mismo modo, la expresión del marcador de *contra-selección* *CYH2* es letal en presencia de cicloheximida. La expresión de un gen marcador *contra-seleccionable* se ha utilizado para identificar las mutaciones en el dominio de activación de receptor de estrógeno que inhiben su capacidad para activar la transcripción (Pierrat *et al*, 1992, Gene 119: 237-245).

En otra realización preferida del método de la invención, este se caracteriza por que los genes reporteros de selección positiva se seleccionan del grupo que consiste en cualquiera de los siguientes marcadores seleccionables *HIS3*, *LEU2*, *URA3*, *ADE2*, *TRP1*, *LYS2* y *LYS5*. En otra realización más preferida aún, el gen reportero de selección positiva es preferentemente el gen *HIS3*.

En otra realización preferida, los genes reporteros de *contra-selección* se seleccionan del grupo que consiste en cualquiera de los marcadores *contra-seleccionables* *URA3*, *TRP1*, *LYS2*, *LYS5*, *CYH2*, *CAN1*, *GAL1* y *mazF*. En otra realización más preferida aún, el gen reportero de *contra-selección* es preferentemente el gen *URA3*.

En otra realización preferida, si se desea, la construcción génica con el gen reportero (por ejemplo, *SPAL10:URA3* o *(lexAop)4:HIS3*) se puede integrar en el genoma de una célula haploide o diploide. En una realización más preferida aún, se pueden
5 integrar en el genoma de una célula haploide o diploide, más de un gen reportero, preferiblemente al menos cuatro genes reporteros, al menos tres genes reporteros y más preferiblemente, al menos dos genes reporteros. Si se desea, una combinación de genes reporteros se puede integrar cromosómicamente en el genoma de la célula y otros genes reporteros pueden localizarse en un plásmido episomal y expresarse a
10 partir del mismo.

En una realización preferida del método de doble híbrido en reverso descrito en la presente invención, este se caracteriza por que la célula utilizada en dicho método comprende las secuencias de reconocimiento para proteínas que se unen al ADN
15 seleccionadas de entre cualquiera de los siguientes factores de transcripción: Gal4, LexA y Ace1.

A efectos de la presente invención, el método de doble híbrido en reverso se caracteriza por que la célula utilizada en dicho método comprende integrado en su
20 genoma al menos dos genes reporteros, preferentemente uno de selección positiva y otro de contra-selección. En otra realización más preferida aún, el gen reportero de selección positiva es *HIS3* y el gen reportero de contra-selección es *URA3*.

En otra realización preferida, la célula utilizada en el método de la invención también
25 se caracteriza por que comprende, preferentemente integrado en su genoma vía transformación con un plásmido integrativo, la secuencia nucleotídica que codifica para una primera proteína de fusión, donde dicha primera proteína de fusión comprende el mismo dominio de unión al ADN que el comprendido en el apartado ii), y una proteína heteróloga que se une al péptido utilizado en el método de la invención y
30 que está comprendido en un plásmido que comprende a su vez la secuencia nucleotídica que codifica para la proteína de referencia sometida a estudio. En una realización más preferida aún, el plásmido integrativo comprende las secuencias nucleotídicas de la construcción génica donde el promotor de *ADH1* esta operativamente unido a la secuencia nucleotídica que codifica para la fusión del
35 dominio de unión al ADN de LexA con la proteína heteróloga que se une al péptido de la invención. A efectos de la presente invención y a modo de ejemplo, sin querer ser

limitativo se ha utilizado como proteína heteróloga expresada desde una construcción génica integrada en el genoma de la célula de la invención, la proteína humana TSG101 (SEQ ID NO: 32, codificada por la secuencia nucleotídica SEQ ID NO: 31). Así, la célula de la invención expresa una primera proteína de fusión que comprende el dominio de unión al ADN de LexA y TSG101 (LexA-TSG101), siendo su secuencia nucleotídica la SEQ ID NO: 48 que codifica para la proteína de fusión de SEQ ID NO: 49.

La proteína TSG101 (del inglés Tumor Susceptibility Gene 101 pertenece a un grupo de enzimas aparentemente inactivas para conjugación con ubiquitina (ubiquitin-conjugating enzymes); en la presente invención TSG101 preferiblemente se refiere al Gene ID: 7251 de *Homo sapiens* aunque cualquier experto en el estado de la técnica entenderá que es posible el uso de otras secuencias procedentes de otras especies, preferiblemente mamíferos, más preferiblemente mamíferos primates, que son homólogas a la secuencia humana. Dicha proteína forma parte del complejo ESCRT-I que reconoce las proteínas ubiquitinadas en la ruta endocítica y está implicada en la formación del cuerpo multivesicular (MVB). TSG101 es a su vez capaz de interactuar con una pequeña secuencia de un péptido denominado P(S/T)TAP que está presente en la proteína Gag p6 del virus VIH (PNAS (2001) 98, 7724-9; Cell (2001) 107, 55-65). Se ha observado que una secuencia que comprende una triple repetición del péptido PTAP, que denominamos 3xPTAP, a lo largo del presente documento, presente de forma natural en una cepa de VIH aislada (GenBank: ACS76886.1) interactúa con TSG101 más fuertemente que una única repetición del péptido PTAP aislado. A efectos de la presente invención, para aumentar la fuerza de interacción de la proteína de fusión LexA-TSG101 (SEQ ID NO: 49) se ha utilizado esta triple repetición del péptido PTAP (3xPTAP) de SEQ ID NO: 12. Se podrían utilizar otras secuencias que comprendan el péptido P(S/T)AP y que interactúan con TSG101, u otras combinaciones de proteína y péptido que interactúan entre si. A modo de ejemplo, se podría utilizar la combinación del péptido YPX(L/I) y la proteína humana ALIX o cualquier homólogo de esta proteína en otros organismos que une este péptido (por ejemplo la proteína PalA del hongo filamentoso *Aspergillus nidulans*) (Mol Cell Biol (2003) 23, 1647-55).

El péptido 3xPTAP (SEQ ID NO: 12) une con gran afinidad la proteína humana TSG101 (SEQ ID NO: 32) en el sistema de doble híbrido en reverso utilizado en la presente invención. De esta manera, se ha llevado a cabo la fusión del péptido

3xPTAP en el extremo C-terminal de la proteína de referencia (que es sometida a mutagénesis), para identificar mutaciones que anulan la interacción con una proteína diana, sin afectar a la interacción mediada por el péptido 3xPTAP con la proteína TSG101, lo que permite descartar todos los falsos positivos que corresponden a mutaciones que truncan la proteína. La proteína de fusión con el péptido 3xPTAP en el extremo C-terminal de la proteína diana mutada, se encuentra fusionada al dominio de transactivación de Gal4 (GAD) en su extremo N-terminal, permitiendo la interacción en el sistema de doble-híbrido de la invención con la proteína de fusión LexA-TSG101 integrada en el cromosoma de la célula, y en consecuencia la activación del reportero (*lexAop*)4-*HIS3*. Este sistema permite una selección positiva e inequívoca de las mutaciones *missense* ya que todas las mutaciones que truncan, inestabilizan o impiden la entrada de la proteína en el núcleo bloquean su interacción con la proteína heteróloga TSG101 (SEQ ID NO: 32) y en consecuencia la activación del gen reportero *HIS3*, lo que se traduce en una falta de crecimiento en un medio sin histidina (ver **Figura 1**). La selección de los mutantes que han incorporado las mutaciones *missense*, se realiza en un solo paso, con un sistema doble de reporteros, tal y como hemos mencionado anteriormente, el primer reportero permite la selección de los mutantes que pierden la interacción, mientras que el segundo selecciona de forma simultánea los mutantes que no truncan la proteína.

En una realización más preferida aún, la célula utilizada en el método de la invención comprende, preferentemente integrado en su genoma los reporteros LexAop-*HIS3* (*(lexAop)4-HIS3*) y *UASGal-URA3* (*SPAL10::URA3*), que codifican para los marcadores de selección positiva y contra-selección *HIS3* y *URA3*, respectivamente. Adicionalmente, tal y como se ha mencionado anteriormente, la célula utilizada en el método descrito en la presente invención comprende la construcción génica donde el promotor de *ADH1* está operativamente unido a la secuencia nucleotídica que codifica para la fusión del dominio de unión al ADN de LexA con la proteína heteróloga, preferentemente dicha proteína heteróloga es la proteína humana TSG101, que se une al péptido de la invención. Así, la célula de la invención expresa la proteína de fusión LexA-TSG101 de SEQ ID NO: 49.

Los términos “híbrida” o “de fusión”, en relación a proteínas o ADN, hacen referencia a una quimera de al menos dos polipéptidos, o dos moléculas de ADN, unidas covalentemente.

Por “dominio de transactivación”, se entiende al conjunto de aminoácidos capaces de inducir la expresión de un gen de la región a cuyo promotor está unido.

5 A efectos de la presente invención se entiende por “péptido”, “etiqueta”, “epítipo”, “cola”, o “tallo”, funcional en el extremo C-terminal de la proteína de referencia, a un conjunto de amino ácidos ubicados en el extremo C-terminal de la proteína de referencia. A efectos de la presente invención se utiliza como péptido funcional en el extremo C-terminal de la proteína de referencia analizada a una secuencia peptídica de pequeño tamaño que interactúa con gran afinidad con la proteína heteróloga que
10 expresa la célula de la invención de la etapa a).

En una realización preferida, el péptido funcional en el extremo C-terminal de la proteína de referencia se refiere a la secuencia SEQ ID NO: 12 que se corresponde con tres repeticiones del péptido PTAP (3xPTAP). El péptido PTAP utilizado en la
15 presente invención se encuentra en la proteína Gag p6 del virus VIH, según se ha indicado anteriormente.

En una realización más preferida, la célula utilizada en el método de doble híbrido en reverso de la invención, comprende integrado en su genoma, las construcciones
20 génicas *SPAL10-URA3*, *(lexAop)4-HIS3* y *ADH1::LexA-TSG101*. Adicionalmente, también puede comprender la construcción génica que comprende el gen reportero detectable *GAL1-LacZ*, que, al igual que el gen reportero seleccionable *SPAL10-URA3*, está bajo el control de un promotor con la secuencia de reconocimiento *UASGal*. La presencia de dicho reportero detectable permite, si fuera necesario,
25 confirmar o validar los resultados obtenidos con el reportero *SPAL10-URA3* en el método de la invención, con ensayos de la actividad enzimática β -galactosidasa.

En otra realización preferida del método de la invención, el plásmido utilizado en la etapa b) comprende la secuencia nucleotídica que codifica para el dominio de unión al
30 ADN de Gal4 fusionada a la secuencia nucleotídica que codifica para la proteína diana a ensayar. A efectos de la presente invención y, a modo de ejemplo, se ha utilizado como proteína diana a ensayar la proteína humana GKRP de SEQ ID NO: 2, todo ello bajo el control de un promotor constitutivo, preferentemente y a modo de ejemplo, se ha utilizado el promotor de *ADH1*.

35

Según se describe en los ejemplos incluidos en el presente documento, la célula utilizada en el método de la invención se pre-transforma con dicho plásmido para asegurar la expresión de la proteína diana GKRP, de esta manera, se asegura que la célula expresa la proteína diana, y posteriormente tal y como se indica en la etapa c), se cultiva en condiciones que permiten exclusivamente el crecimiento de las células que han incorporado dicho plásmido.

Los medios de cultivo utilizados para el crecimiento de las células descritas en la presente invención y utilizados en el método aquí descrito, son medios de cultivo conocidos para el crecimiento preferentemente de células de levadura. Preferentemente, los medios de cultivo utilizados en la presente invención son el medio de cultivo mínimo (SD) y el medio de cultivo completo (YPAD). La preparación de dichos medios de cultivo se describe en *Methods in yeast genetics — A laboratory course manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 1990; pp 198. El medio de cultivo completo (YPAD) contiene extracto de levadura al 1% (p/v), bactopectona al 2% (p/v), 0.02% (p/v) de adenil(hemi)sulfato y como fuente de carbono, glucosa al 2%. El medio de cultivo mínimo (SD) se compone de base nitrogenada para levadura (YNB) sin aminoácidos al 0,17% (p/v), y sulfato amónico, (NH₄)₂SO₄, al 0,5% (p/v) suplementado con glucosa al 2% (p/v). Cuando no se especifica lo contrario, se añaden los requerimientos siguientes a una concentración final de 20 mg/l (histidina, triptófano, uracilo, arginina, metionina), 30 mg/l (isoleucina, lisina, tirosina), 50 mg/l (adenina, fenilalanina), 100 mg/l (leucina), 150 mg/l (valina) o 200 mg/l (treonina).

En otra realización preferida del método de la invención, la célula pre-transformada según se ha mencionado anteriormente, se transforma con un vector linearizado y un fragmento de ADN que comprende la secuencia nucleotídica que codifica para para la proteína de referencia que previamente ha sido sometida a un procedimiento de mutagénesis, *in vivo* o *in vitro*, y por lo tanto comprende al menos una mutación. Así, el cultivo de la célula pre-transformada con el vector linearizado y el fragmento de ADN previamente mencionado hace que entre ambas moléculas se produzca recombinación homóloga ("*gap-repair*") y se obtenga un plásmido que comprende el dominio de transactivación de Gal4 fusionado a la secuencia nucleotídica que codifica para la proteína de referencia, y estando la secuencia nucleotídica comprendida en dicho plásmido bajo el control de un promotor. A efectos de la presente invención y a

modo de ejemplo, sin ser limitativo, la proteína de referencia es la proteína GK de SEQ ID NO: 10, todo ello bajo el control del promotor constitutivo de ADH1.

5 En una realización preferida del método de la invención, las mutaciones en la secuencia nucleotídica que codifica para la proteína de referencia se obtienen utilizando preferentemente la técnica de PCR mutagénica y recombinación *in vivo* ("gap-repair"). A efectos de la presente invención, el método de mutagénesis mediante reacción en cadena de la polimerasa (PCR) proporciona un método conveniente para inducir mutagénesis de una secuencia elegida (Muhlrad *et al*, 1992, Yeast 8: 79-82), a
10 efectos de la presente invención en la secuencia de la proteína de referencia. En el método de PCR mutagénica y recombinación *in vivo* "gap-repair", el ADN que codifica la secuencia de la proteína que se muta se amplifica en una reacción de PCR en condiciones que favorecen la incorporación de nucleótidos incorrectos en la molécula de ADN. Tales condiciones incluyen niveles relativamente altos de manganeso y/o
15 una mezcla desigual de los diferentes nucleótidos. Los cebadores de PCR que se utilizan en el método de PCR mutagénica y recombinación *in vivo* "gap-repair" generan productos de PCR lineales que tienen en sus extremos secuencias homólogas a los extremos del plásmido linealizado, lo que favorece la recombinación entre ambos después de su co-transformación en levadura.

20

En una realización preferida del método de doble híbrido en reverso descrito en la presente invención se caracteriza por que las etapas b) y d) se pueden llevar a cabo simultáneamente. Así, las células de la invención se co-transformaron con el plásmido que codifica para la proteína diana y simultáneamente también con el plásmido
25 linearizado y los productos obtenidos de la PCR mutagénica.

A efectos de la presente invención, el término "mutado" o "mutación" se refiere a la secuencia modificada, bien mediante mutagénesis dirigida o mutagénesis al azar, respecto a la misma secuencia encontrada de manera natural en la naturaleza
30 (secuencia nativa o secuencia *wild-type*) y que no presenta mutaciones. Las secuencias mutadas incluyen aquellas secuencias que tienen mutaciones puntuales, inserciones, deleciones o reordenamientos.

A efectos de la presente invención, el término "promotor" se refiere a una mínima
35 secuencia suficiente para dirigir la transcripción; tales elementos pueden estar situados en los extremos 5' o 3' de la secuencia del gen nativo. Promotores

adecuados para la expresión de un gen reportero son aquellos que, cuando se enlaza con el gen reportero, puede dirigir la transcripción del mismo en presencia de moléculas apropiadas (es decir, proteínas que tienen dominios de transactivación). Un ejemplo de un promotor útil es el promotor de *ADH1* o el promotor de *SP013*. Otros promotores útiles incluyen aquellos promotores que contienen secuencias represoras aguas arriba (véase, por ejemplo, Vidal *et al*, 1995, Proc Natl Acad Sci EE.UU. 92:2370-2374), y que inhiben la expresión del gen reportero en ausencia de un activador transcripcional. La capacidad de un promotor para la transcripción de un gen reportero se puede medir con métodos convencionales de ensayos de expresión de genes (por ejemplo, la detección del producto del gen o su ARNm, o la detección del crecimiento celular bajo condiciones donde se requiere la expresión del gen reportero para el crecimiento de una célula).

Técnicas convencionales de biología molecular pueden ser usadas para construir derivados de promotores que incluyen una o más secuencias de reconocimiento para proteínas que se unen el ADN. Por ejemplo, el promotor de *SP013* puede diseñarse para incluir una o más copias del sitio de unión de Gal4. Los sitios de unión naturales de Gal4 han sido ampliamente caracterizados, permitiendo la creación de una secuencia sintética a la que se une Gal4 con una afinidad relativamente alta. Otras secuencias de reconocimiento útiles para proteínas que se unen al ADN incluyen los sitios de unión de LexA y Ace1. Además, donde se mide la capacidad de una proteína para unirse a una secuencia de ADN, la secuencia de reconocimiento de una proteína que se une al ADN puede ser de tipo nativo o *wild-type*, o puede tener cualquier diseño, tanto intencionalmente como al azar, para probar la capacidad de la secuencia de ADN para interactuar con esta proteína.

En una realización preferida del método de doble híbrido en reverso descrito en la presente invención, los promotores utilizados son preferentemente promotores constitutivos. En una realización más preferida aún, dichos promotores constitutivos se seleccionan de entre cualquiera de la lista que consiste en: *ADH1*, *PGK1*, *TEF1*, *TPL1*, *HXT7*, *TDH3* y *PYK1*.

En una realización preferida el método de doble híbrido en reverso de la invención se caracteriza por que el plásmido utilizado en la etapa d), antes de transformar la célula con él, es sometido a un procedimiento de mutagénesis, *in vivo* o *in vitro*, para inducir mutaciones al azar en la secuencia nucleotídica que comprende dicho plásmido y que

codifica para la proteína de referencia. En una realización más preferida aún, el procedimiento de mutagénesis se lleva a cabo mediante PCR mutagénica de la secuencia nucleotídica que codifica para la proteína de referencia y recombinación *in vivo* ("*gap-repair*") del producto obtenido en un plásmido. Preferiblemente, la etapa b) y d) del método de la invención, pueden realizarse por separado o simultáneamente, siendo preferido llevarlas a cabo por separado.

A efectos de la presente invención, en la PCR mutagénica utilizada para introducir mutaciones al azar en la secuencia nucleotídica que codifica para la proteína GK se pueden utilizar diferentes ADN polimerasas con diferentes tasas de mutación cada una de ellas, con la finalidad de comparar la idoneidad de cada polimerasa en el método de la invención. En los ejemplos mostrados en el presente documento, se han mostrado los resultados utilizando la ADN polimerasa Taq procedente de Takara o Roche, y la ADN polimerasa Mutazyme II procedente del kit de mutagénesis al azar Genemorph II de Agilent. A efectos de la presente invención es preferible usar una ADN polimerasa que incluya una única mutación en cada plásmido.

Otras técnicas conocidas para generar mutaciones al azar, y que pueden ser utilizadas en el método de la invención, pueden ser el uso de mutágenos físicos y/o químicos, cepas celulares mutadoras, preferentemente cepas bacterianas, tal como por ejemplo, la cepa de *Escherichia coli* mutadora descrita en Rasila TS. *et al.* Anal Biochem. 2009;388(1):71-80.

Técnicas convencionales de transformación celular, preferentemente de transformación en células de levadura pueden ser utilizadas a efectos de la presente invención. A modo de ejemplo y sin pretender limitar, electroporación, método de transformación en presencia de litio, método de transformación con esferoplastos, método de transformación con bolas de vidrio, etc.

En una realización preferida del método de la invención, a continuación en la etapa e) se cultiva la célula transformada de la etapa d) bajo condiciones que permiten el crecimiento de las células que presentan mutaciones *missense* que provocan la pérdida de interacción entre el par de proteínas a estudiar. El medio mínimo selectivo utilizado es un medio sin triptófano, leucina, adenina e histidina y con 0.1 % 5-FoA y de 1 a 5 mM 3-AT (SD-AHTL+5-FoA+3-AT). Este medio carece de adenina ya que la cepa utilizada es protótrofa para este requerimiento. La ausencia de triptófano y

leucina permite seleccionar los transformantes que han incorporado los dos plásmidos que expresan la proteína diana (marcador *TRP1*) y la proteína de referencia (marcador *LEU2*). La ausencia de histidina permite seleccionar los transformantes que activan el reportero (*(lexAop)4:HIS3*), y la presencia de 3-AT (3-aminotriazol), un inhibidor de *HIS3*, impide el crecimiento de los transformantes con un nivel basal de activación de este reportero. Finalmente, la presencia de 5-FoA impide el crecimiento de los transformantes que activan el reportero *SPAL10:URA3*. En consecuencia, el medio SD-AHTL+5-FoA+3-AT permite el crecimiento únicamente de los transformantes que han incorporado los dos plásmidos, el que expresan la proteína diana y el que expresa la proteína de referencia, y que por otra parte activan el reportero *(lexAop)4:HIS3*, pero no activan el reportero *SPAL10:URA3*. Dado que la interacción entre la proteína diana y la proteína de referencia activa el reportero *SPAL10:URA3*, los transformantes que no tienen mutaciones, o los que si las tienen pero que no bloquean la interacción entre las dos proteínas, activarán el reportero *SPAL10:URA3* y no podrán crecer en presencia de 5-FoA (**Figura 1a**). Por otra parte, la activación del reportero *(lexAop)4:HIS3* depende de la interacción entre TSG101 y el péptido 3xPTAP situado en el extremo C-terminal de la proteína de referencia. En consecuencia, los plásmidos con mutaciones en la secuencia codificante de la proteína de referencia que provocan su truncamiento, y en consecuencia eliminan el péptido 3xPTAP, así como los plásmidos recircularizados sin incorporar el producto de PCR mutagénico en el proceso de recombinación *in vivo* (o los que se habían quedado sin digerir) no activarán el reportero *(lexAop)4:HIS3* y no podrán crecer en ausencia de histidina (**Figura 1b**). En consecuencia, los únicos transformantes que podrán crecer en este medio son los que portan una mutación que bloquea la interacción entre la proteína diana y la proteína de referencia sin que esta mutación produzca un truncamiento en la proteína, es decir que sea una mutación *missense* (**Figura 1c**).

En una realización preferida, el método de doble híbrido en reverso se caracteriza por que la célula se selecciona de la lista que consiste en: célula de levadura, célula bacteriana y célula de mamífero. En una realización más preferida, la célula es preferentemente una célula de levadura. En una realización más preferida aún, la célula de levadura se selecciona del grupo que consiste en *Yarrowia lipolytica*, *Pichia pastoris* y *Saccharomyces cerevisiae*. Las células de la especie *S. cerevisiae* son particularmente útiles en el método de la invención. Las cepas utilizadas en la presente invención se mantienen en cultivos estándar mediante métodos estándar.

En otra realización preferida del método de la invención, en la etapa f) se compararán e identificarán las mutaciones *missense* presentes en los plásmidos que portan las células que han crecido en la etapa e) del método de la invención, respecto a la
5 secuencia nativa o *wild-type* de la proteína de referencia.

En una realización preferida, la identificación de las mutaciones *missense* obtenidas mediante el método de la invención se llevan a cabo mediante la extracción del plásmido que comprende la secuencia con la mutación *missense*, mediante técnicas
10 comúnmente conocidas en el presente campo técnico, y posteriormente se procede a la secuenciación de dicha secuencia mutada, directamente desde el ADN plasmídico, para determinar la mutación concreta comparándola con la secuencia nativa o *wild-type* de la proteína analizada.

15 Dada la eficiencia del sistema de doble híbrido en reverso descrito en la presente invención, para detectar el 100% de las mutaciones *missense* entre un par de proteínas interaccionantes de estudio, sin falsos positivos, sería posible aislar en masa, y de una sola vez el ADN de todas los clones positivos para luego amplificar la secuencia codificante de la proteína de referencia estudiada (GK en los ejemplos
20 mostrados) mediante PCR y secuenciar de una sola vez todos los mutantes obtenidos, con un sistema de secuenciación masiva (*Next Generation Sequencing*), de tipo Illumina. El análisis bioinformático del resultado permitiría generar una huella genética del sitio de interacción en la proteína del par interaccionante analizado y que ha sido sometida a mutagénesis.

25 Tal y como hemos mencionado previamente, la célula de levadura utilizada en la presente invención, ha integrado en su genoma un gen reportero contra-seleccionable, siendo preferido el gen reportero *URA3*, que está unido operativamente a un promotor que comprende una secuencia de reconocimiento para una proteína
30 que se une al ADN, siendo preferida la secuencia UASGal reconocida por el dominio de unión al ADN de Gal4. Además, la célula utilizada comprende también integrado en su genoma otro gen reportero seleccionable, siendo preferido el gen reportero *HIS3*, que está unido operativamente a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN, siendo preferida la secuencia
35 *lexAop* reconocida por el dominio de unión al ADN de LexA. Adicionalmente, la célula de la invención comprende integrado en su genoma la construcción génica

ADH1::LexA-TSG101 (SEQ ID NO: 48), que codifica para la proteína de fusión de SEQ ID NO: 49.

5 Cuando los tres genes mencionados anteriormente (es decir, un gen reportero contra-seleccionable, un gen reportero seleccionable, preferiblemente de selección positiva, y la proteína heteróloga a la que se une el péptido funcional del extremo C-terminal de la proteína de referencia sometida a mutagénesis y recombinación *in vivo*, que forma la proteína de fusión de SEQ ID NO: 49), se integran en el genoma de una célula, los promotores de los dos genes reporteros son distintos, específicamente en la región de 10 los sitios de reconocimiento de la unión ADN-proteína (**Figura 1**), mientras que el resto del promotor puede ser similar.

Célula huésped de la invención.

15 A efectos de la presente invención se utilizan los términos “célula”, “célula huésped”, “cepa”, indistintamente a lo largo del presente documento.

Tal y como se demuestra en el presente documento, los inventores han construido un conjunto de células o cepas de levadura que tienen las siguientes características:

- 20 i) Una primera secuencia nucleotídica que codifica para un gen reportero, donde dicha secuencia nucleotídica está operativamente unida un promotor que comprende una secuencia nucleotídica reconocida por una proteína que se une al ADN,
- 25 ii) Una segunda secuencia nucleotídica que codifica para un segundo gen reportero, donde dicha segunda secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN, con la condición de que dicha secuencia sea reconocida por un dominio de unión al ADN distinto del de i), y
- 30 iii) Una tercera secuencia nucleotídica que codifica para una primera proteína de fusión que comprende el dominio de unión al ADN de ii) y una proteína heteróloga capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia de estudio, donde la secuencia nucleotídica que codifica para dicha primera proteína de fusión está operativamente unida a un promotor.

35

En una realización preferida, la célula de la invención comprende dichas construcciones integradas en su genoma.

5 En otra realización más preferida, la célula de la invención se caracteriza por que la primera secuencia nucleotídica que comprende integrada en su genoma comprende un gen reportero que se selecciona de entre genes reporteros de selección positiva y/o, contra-selección, preferentemente un gen reportero de contra-selección, y más preferentemente seleccionado de entre cualquiera de la siguiente lista: *URA3*, *TRP1*, *LYS2*, *LYS5*, *CYH2*, *CAN1*, *GAL1*, *mazF*. En otra realización preferida, dicho gen
10 reportero está operativamente unido a una secuencia de reconocimiento para una proteína que se une al ADN, según se ha definido previamente de igual manera para el método de la invención. En una realización más preferida aún, la primera secuencia nucleotídica que comprende la célula de la invención se caracteriza por que comprende la construcción génica *UASGal-URA3 (=SPAL10::URA3)*, que expresa el
15 marcador de contra-selección *URA3*.

En otra realización preferida de la célula de la invención, ésta se caracteriza por que la segunda secuencia nucleotídica que comprende integrada en su genoma comprende un gen reportero que se selecciona de entre genes reporteros de selección positiva
20 y/o contra-selección, preferentemente un gen reportero de selección positiva, y más preferentemente seleccionado de entre cualquiera de la siguiente lista: *HIS3*, *LEU2*, *URA3*, *ADE2*, *TRP1*, *LYS2*, *LYS5*. En otra realización preferida, dicho gen reportero está operativamente unido a una secuencia de reconocimiento para una proteína que se une al ADN, con la condición de que dicha secuencia de reconocimiento es
25 diferente de la secuencia de reconocimiento comprendida en la primera secuencia nucleotídica que comprende la célula, es decir, diferente a la secuencia de reconocimiento de i). En una realización más preferida aún, la segunda secuencia nucleotídica que comprende la célula de la invención se caracteriza por que comprende la construcción génica *(lexAop)4-HIS3*, que expresa el reportero de
30 selección positiva *HIS3*.

Tal y como se ha descrito a lo largo del presente documento, los dominios de unión al ADN que pueden comprender las cepas de la invención se seleccionan de entre cualquiera de los siguientes factores de transcripción: Gal4, LexA y Ace1.

35

En otra realización preferida de la célula de la invención, ésta se caracteriza por que comprende además integrado en su genoma, una tercera secuencia nucleotídica que codifica para una proteína de fusión, denominada aquí primera proteína de fusión, que comprende el dominio de unión al ADN de LexA y la proteína heteróloga que se une al péptido funcional descrito y utilizado en la presente invención, todo ello bajo el control del promotor de *ADH1*. A efectos de la presente invención, y a modo de ejemplo, sin querer ser limitativo se ha utilizado como proteína heteróloga con la que se transforma la célula de la invención y que se integra en su genoma la proteína humana TSG101 (SEQ ID NO: 32, codificada por la secuencia nucleotídica SEQ ID NO: 31). Así, la célula de la invención expresa la proteína de fusión LexA-TSG101 de SEQ ID NO: 49, a partir de la secuencia nucleotídica SEQ ID NO: 48.

En una realización preferida, la construcción génica que codifica para la proteína de fusión de la célula de la invención aquí descrita, está unida operativamente a un promotor, preferentemente constitutivo y que se selecciona de entre cualquiera de los descritos anteriormente, siendo preferido el promotor de *ADH1*.

En otra realización preferida de la célula de la invención, ésta se caracteriza por que comprende integrado en su genoma, al menos dos genes reporteros, preferentemente uno de selección positiva y otro de contra-selección. En otra realización más preferida aún, el gen reportero de selección positiva es *HIS3* y el gen reportero de contra-selección es *URA3*.

En una realización más preferida, la célula de la invención, comprende integrado en su genoma, las construcciones génicas *SPAL10-URA3*, *(lexAop)4-HIS3* y *ADH1::LexA-TSG101*. Adicionalmente, también puede comprender la construcción génica que expresa para el gen reportero detectable *GAL1-LacZ*, que está bajo el control de un promotor con la secuencia de reconocimiento *UASGal*. La presencia de dicho reportero detectable permite, si fuera necesario, confirmar o validar los resultados obtenidos con el reportero *URA3* en el método de la invención, con ensayos de la actividad β -galactosidasa.

En una realización preferida, el método de doble híbrido en reverso se caracteriza por que la célula se selecciona de la lista que consiste en: célula de levadura, célula bacteriana y célula de mamífero. En una realización más preferida, la célula es preferentemente una célula de levadura. En una realización más preferida aún, la

célula de levadura se selecciona del grupo que consiste en *Yarrowia lipolytica*, *Pichia pastoris* y *Saccharomyces cerevisiae*. Las células de la especie *S. cerevisiae* son particularmente útiles en el método de la invención. Las cepas utilizadas en la presente invención se mantienen en cultivos estándar mediante métodos estándar.

5

En una realización más preferida aún, la célula de la invención es la célula OVY216. Dicha célula OVY216 se caracteriza por el genoma: *MATa ade2-101, his3-Δ200, leu2-3,112, trp1-901, gal4Δ, gal80Δ, LYS2:(lexAop)4-HIS3, SPAL10::URA3, GAL1-lacZ ADE2::LexA-TSG101*, que comprende integrado cromosómicamente las construcciones génicas *(lexAop)4-HIS3* y *UASGal-URA3 (=SPAL10::URA3)*, que expresan los reporteros *HIS3* y *URA3*, y la construcción génica *ADH1::LexA-TSG101*, que expresa la proteína de fusión LexA-TSG101 (SEQ ID NO: 49) que es la proteína que se une al péptido funcional 3xPTAP (SEQ ID NO: 12). Adicionalmente, tal y como se han indicado anteriormente, la cepa OVY216 de la presente invención también comprende la construcción que expresa el gen reportero detectable *GAL1-lacZ*. Este gen reportero, al igual que la construcción *SPAL10::URA3*, está bajo el control de *UASGal*. La presencia del reportero detectable *GAL1-lacZ* permite, si fuese necesario, confirmar o validar los resultados obtenidos con el reportero *URA3*, con ensayos de la actividad β-galactosidasa.

20

Construcciones génicas

Otro aspecto descrito en la presente invención se refiere a las construcciones génicas utilizadas para obtener la célula huésped descrita en la presente invención, así como para llevar a cabo el método de doble híbrido en reverso aquí descrito.

25

Una de las construcciones génicas aquí descritas, a partir de ahora la denominaremos primera construcción génica de la invención, comprende las secuencias nucleotídicas que codifican para:

30

- i) un promotor
- ii) un dominio de unión al ADN, y
- iii) una proteína heteróloga capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia.

En una realización preferida, la primera construcción génica de la invención se caracteriza por que comprende un promotor, preferentemente promotor constitutivo, que puede seleccionarse de entre cualquiera de los descritos previamente en la presente invención. Adicionalmente, dicha primera construcción génica comprende
5 también un dominio de unión al ADN, que se selecciona de entre cualquiera de los descritos a lo largo del presente documento. De la misma manera, la primera construcción génica de la invención, comprende también la secuencia nucleotídica que codifica para la proteína heteróloga que se une al péptido funcional descrito en la presente invención. En una realización preferida, dicha proteína heteróloga es la
10 proteína humana TSG101 (SEQ ID NO: 32, codificada por la secuencia nucleotídica SEQ ID NO: 31). La proteína TSG101 es capaz de interactuar con un pequeño péptido denominado PTAP que está presente en la proteína Gag p6 del virus VIH (PNAS (2001) 98, 7724-9; Cell (2001) 107, 55-65).

15 En una realización más preferida aún, la primera construcción génica de la invención es preferentemente el plásmido pRS402-LexA-TSG101, más preferentemente el plásmido pRS402-LexA-TSG101 que comprende la secuencia SEQ ID NO: 30.

En otra realización preferida, la primera construcción génica descrita en la presente
20 invención se encuentra integrada en el cromosoma de la célula huésped de la invención y codifica para la primera proteína de fusión descrita en la presente invención, la proteína de fusión LexA-TSG101 de SEQ ID NO: 49.

Otro aspecto descrito en la presente invención se refiere a una segunda construcción
25 génica, a partir de aquí la denominaremos segunda construcción génica de la invención, que comprende las secuencias nucleotídicas que codifican para:

- i) un promotor,
- ii) un dominio de unión al ADN, y
- iii) una proteína diana a la que se une la proteína de referencia.

30

En una realización preferida, la segunda construcción génica de la invención, comprende por tanto, un promotor, preferentemente constitutivo, pudiendo utilizarse cualquiera de los descritos en la presente invención, un dominio de unión al ADN, que se selecciona de entre cualquiera de las descritos a lo largo del presente documento,
35 y la secuencia de la proteína diana que se une a la proteína de referencia. A efectos

de la presente invención, la proteína diana ejemplificada es la proteína reguladora de la glucoquinasa humana de SEQ ID NO: 2 codificada por la SEQ ID NO: 1. En una realización más preferida aún, la tercera construcción génica de la invención se refiere al plásmido pGBKT7-GKRP, más preferentemente al plásmido pGBKT7-GKRP que comprende la SEQ ID NO: 6.

En otra realización preferida, la segunda construcción génica descrita en la presente invención codifica para la segunda proteína de fusión descrita en la presente invención, la proteína de fusión GBD-GKRP de SEQ ID NO: 51.

Otro aspecto descrito en la presente invención se refiere a una tercera construcción génica, a partir de aquí la denominaremos tercera construcción génica de la invención, que comprende las secuencias nucleotídicas que codifican para:

- i) un promotor
- ii) un dominio de transactivación, y
- iii) una proteína de referencia que además en su extremo carboxilo terminal comprende la secuencia que codifica para el péptido funcional 3xPTAP que se une específicamente a la proteína heteróloga codificada por la primera construcción génica de la invención.

En una realización preferida, la tercera construcción génica de la invención, comprende por tanto, un promotor, preferentemente constitutivo, pudiendo utilizarse cualquiera de los descritos en la presente invención, un dominio de transactivación, pudiendo utilizarse cualquiera de los descritos en la presente invención, y la secuencia de la proteína de referencia que se quiera estudiar, fusionada al péptido funcional 3xPTAP de SEQ ID NO: 12 en su extremo C-terminal. A efectos de la presente invención, la proteína de referencia ejemplificada es la proteína glucoquinasa humana de SEQ ID NO: 10 codificada por la SEQ ID NO: 9. En una realización más preferida aún, la segunda construcción génica de la invención se refiere al plásmido pACT2-GK-3xPTAP, más preferentemente al plásmido pACT2-GK-3xPTAP que comprende la SEQ ID NO: 14.

En otra realización preferida, la tercera construcción génica descrita en la presente invención codifica para la tercera proteína de fusión descrita en la presente invención, la proteína de fusión GAD-GK-3xPTAP de SEQ ID NO: 53.

A efectos de la presente invención se pueden utilizar otros plásmidos a los descritos aquí, con los mismos dominios de unión al ADN, tal como por ejemplo el plásmido pDESTTM32 (Invitrogen), o con otros (en sustitución de GBD y LexA) en combinación con una cepa con reporteros bajo el control de promotores con las secuencias de reconocimiento correspondientes, en sustitución de UASGal y lexAop, así como otros plásmidos con el mismo dominio de transactivación (GAD), tal como por ejemplo el plásmido pEXPTM-AD502 (Invitrogen), o con otro dominio de transactivación, tal como por ejemplo, VP16.

10

A efectos de la presente invención, también se podría utilizar como péptido funcional, en sustitución del péptido 3xPTAP utilizado en la presente invención, junto con su proteína interactora, TSG101, cualquier otra pareja, siempre que se cumpla la condición de que el péptido funcional debe ser un péptido pequeño, preferentemente menos de 30 aminoácidos y más preferentemente menos de 25 aminoácidos, que no interfiera con la interacción de las proteínas estudiadas, o que para ser detectado requiera de más etapas incrementando la complejidad de la técnica, así como otro promotor diferente a *ADH1* para expresar la fusión a TSG101 y otro marcador que *ADE2* para su integración genómica.

20

A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y figuras se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención.

25

BREVE DESCRIPCIÓN DE LAS FIGURAS

Figura 1. Representación esquemática de los genes reporteros de contra-selección *URA3* y selección positiva *HIS3*, bajo el control de promotores con secuencias de reconocimiento para distintos dominios de unión al ADN. (a) La ausencia de mutación en la proteína Y (ejemplificada en la proteína humana glucoquinasa) produce la activación de *URA3* y la falta de crecimiento en presencia de 5-FoA. (b) Una mutación de truncamiento en Y no permite la activación de *HIS3* y el crecimiento en ausencia de histidina. (c) Una mutación *missense* en Y produce la activación de *HIS3* pero no la de

35

URA3, lo que permite el crecimiento en un medio sin histidina y con 5-FoA. X: se refiere a la proteína reguladora de la glucoquinasa (GKRP).

Figura 2. Construcciones génicas utilizadas en el método de la invención. (a) Construcciones integradas en el genoma de la cepa de levadura OVY216. (b) Construcciones génicas utilizadas para la transformación de la cepa de la invención OVY216. X: se refiere a la proteína reguladora de la glucoquinasa humana (GKRP). Y: se refiere a la proteína humana glucoquinasa (GK).

Figura 3. Representación de la técnica de PCR mutagénica y recombinación *in vivo* (“*gap repair*”) para la introducción de mutaciones al azar en el gen Y (ejemplificada en el gen que codifica para la proteína humana glucoquinasa) comprendido en el plásmido pACT2-Y-3xPTAP de SEQ ID NO: 14.

Figura 4. Ensayo de la actividad β -galactosidasa en filtro para detectar la interacción en el sistema de doble-híbrido clásico de células OVY216 transformadas con el plásmido pACT2-GK-3xPTAP sin mutar o sometido a PCR mutagénica y con el plásmido pGBKT7-GKRP (panel izquierda) o con el plásmido pLexA(1-202)PL-TSG101 (panel derecha). La interacción entre las proteínas GK y GKRP (izquierda) o entre el péptido 3xPTAP y TSG101 (derecha) produce la activación del gen reportero *lacZ*, y en consecuencia la expresión de la β -galactosidasa y la hidrólisis del X-Gal en un compuesto de color azul. Un total de 7-8 células independientes transformadas han sido analizadas para cada interacción.

EJEMPLOS

25

A continuación se ilustrará la invención mediante unos ensayos realizados por los inventores, que pone de manifiesto la efectividad del producto de la invención.

30

En los ejemplos que se detallan a continuación se muestra como mediante el sistema de doble híbrido en reverso, utilizando la célula huésped descrita en la presente invención, así como las construcciones génicas específicas, diseñadas de acuerdo a los ejemplos mostrados, se han localizado mutaciones en la proteína glucoquinasa humana que bloquean su interacción con la proteína humana GKRP. Además, al ser conocida la estructura del complejo GK-GKRP (Choi JM. *et al.* Proc Natl Acad Sci U S A. 2013 Jun 18;110(25):10171-6), ha permitido validar los resultados obtenidos

35

mediante el método descrito en la presente invención, poniendo de manifiesto la consistencia entre los resultados aquí obtenidos con la estructura conocida del complejo.

5 **Ejemplo 1. Análisis de la validez del método de doble híbrido en reverso de la invención mediante el análisis de la interacción entre el par de proteínas GK y GKRP.**

1.1. Diseño y obtención de los plásmidos utilizados en el método de la invención.

10 Las construcción de los plásmidos utilizados en el método de invención ha sido realizada mediante técnicas estándar de ADN recombinante (Sambrook *et al.* Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

15 Para expresar una de las proteína de fusión descritas en el presente documento y que comprende una de las proteínas del par interaccionante a estudiar, específicamente para el presente ejemplo se ha seleccionado la proteína reguladora de la glucoquinasa humana (GKRP) de SEQ ID NO: 2, codificada por el gen de la glucoquinasa humana que comprende la SEQ ID NO: 1, se ha utilizado el plásmido
20 comercial pGBKT7 (Clontech) que comprende como marcador de selección *TRP1*. Dicho plásmido comercial pGBKT7 (SEQ ID NO: 3), comprende la secuencia nucleotídica que codifica para el dominio de unión al ADN del activador transcripcional Gal4 (GBD) (SEQ ID NO: 4), bajo el control del promotor de *ADH1* (SEQ ID NO: 5). De esta manera, para obtener el plásmido pGBKT7-GKRP (SEQ ID NO: 6), la
25 secuencia nucleotídica que codifica para la GKRP (SEQ ID NO: 1) se clona en fase, fusionada en el extremo N-terminal al GBD. Brevemente, mediante la técnica de PCR se clona en el plásmido pGBKT7 (Clontech) la secuencia codificante (SEQ ID NO: 1) de la proteína GKRP humana (SEQ ID NO: 2), incluyendo el codón de iniciación y el codón de parada, entre los sitios EcoR1 y BamH1, de tal forma que las secuencias
30 nucleotídicas que codifican para el dominio de unión al ADN de Gal4 (GBD) y la secuencia codificante (SEQ ID NO: 1) para la proteína GKRP (SEQ ID NO: 2) quedan en fase.

Para la amplificación de la secuencia codificante (SEQ ID NO: 1) de la proteína GKRP
35 (SEQ ID NO: 2) mediante PCR, se utilizó como molde el plásmido pFlag-ctc-hGKRP-

FlagC (Brocklehurst K, *et al.* Biochem. J. 2004;378(Pt2):693-7) y se diseñaron cebadores específicos capaces de hibridar con las regiones 5' y 3' del gen que codifica para la proteína GKRP (SEQ ID NO: 1) y con los sitios de las endonucleasas de restricción EcoR1 y BamH1 en los extremos. Los cebadores utilizados para dicho fin son los cebadores de SEQ ID NO: 7 (hGKRP-EcoR1) (5'-GGAATTCATGCCAGGCACAAAACGGTTTC-3') y de SEQ ID NO: 8 (hGKRP-BamH1) (5'-GGGATCCTACTGAACGTCAGGCTCTAGGATTC-3').

Por otro lado, para expresar la proteína de fusión que comprende la otra proteína del par interaccionante a estudiar, específicamente para el presente ejemplo, se trata de la proteína glucoquinasa humana (GK) de SEQ ID NO: 10, codificada por la secuencia nucleotídica SEQ ID NO: 9, y además las tres repeticiones del péptido PTAP (3xPTAP) (SEQ ID NO: 12: EPEPTAPPEPTAPPEPTAPPAE), codificado por la secuencia nucleotídica SEQ ID NO: 11, se ha utilizado el plásmido comercial pACT2 (Clontech) de SEQ ID NO: 13, que comprende como marcador de selección *LEU2*. Así, para la obtención del plásmido final pACT2-GK-3xPTAP (SEQ ID NO: 14), la secuencia codificante SEQ ID NO: 9 de la proteína GK (SEQ ID NO: 10) se fusiona en el extremo N-terminal al dominio de transactivación de Gal4 (GAD) (SEQ ID NO: 15), y en el extremo C-terminal a las tres repeticiones del péptido PTAP (SEQ ID NO: 11), todo ello bajo el control del promotor de *ADH1* (SEQ ID NO: 16). Brevemente, el plásmido comercial pACT2 (Clontech) de SEQ ID NO: 13 ha sido modificado mediante la inserción en el sitio Xho1 del *polylinker*, de la secuencia nucleotídica (SEQ ID NO: 11) que codifica para el péptido 3xPTAP de SEQ ID NO: 12 (EPEPTAPPEPTAPPEPTAPPAE). Esta secuencia nucleotídica (SEQ ID NO: 11) ha sido obtenida mediante el anillamiento de dos oligonucleótidos complementarios de SEQ ID NO: 17 (OV570) (5'-TCGAGAGCCAGAACCCACAGCACCGCCTGAGCCTACCGCCCCACCCGAACCGA CGGCGCCTCCAGCTGAGTAA-3') y de SEQ ID NO: 18 (OV571) (5'-TCGATTACTCAGCTGGAGGCGCCGTCGGTTCGGGTGGGGCGGTAGGCTCAGGC GGTGCTGTGGGTTCTGGCTC-3'). El oligonucleótido de doble cadena resultante presenta extremos protuberantes 5' (TCGA) que permiten su clonaje en el sitio Xho1 del plásmido pACT2. Así, el plásmido resultante tras el clonaje de dicha secuencia nucleotídica se denomina pACT2-3xPTAP (SEQ ID NO: 19) y presenta el mismo *polylinker* que el plásmido pACT2 original de SEQ ID NO: 11, pero limitado en cada extremo N-terminal y C-terminal con las secuencias que codifican para el dominio de activación de Gal4 (GAD) y para el péptido 3xPTAP, respectivamente, en fase uno

con el otro. A continuación, el plásmido pACT2-GK-3xPTAP de SEQ ID NO: 14 se obtiene mediante clonaje por PCR de la secuencia codificante de la GK humana (SEQ ID NO: 9), incluyendo el codón de iniciación, pero excluyendo el codón de parada, en el sitio BamH1 del *polylinker* del plásmido pACT2-3xPTAP de SEQ ID NO: 19, obtenido previamente, de modo que las secuencias nucleotídicas que codifican para el dominio de activación de Gal4 (GAD), para la proteína GK y para el péptido 3xPTAP, quedan en fase.

Para la amplificación por PCR de la secuencia nucleotídica SEQ ID NO: 9, que codifica para la proteína GK humana (SEQ ID NO: 10), se utilizó como molde el plásmido pGEX-5X-hGKi (Alvarez E, *et al.* J. Neurochem. 2002;80(1):45-53) y se diseñaron cebadores específicos capaces de hibridar con las regiones 5' y 3' del gen que codifica para la proteína GK y con los sitios de la endonucleasa de restricción BamH1 en cada extremo. Los cebadores utilizados para dicho fin son los cebadores de SEQ ID NO: 20 (OV500) (5'-GGGATCCTGATGCTGGACGACAGAGCCAGG-3') y de SEQ ID NO: 21 (OV562) (5'-GGGATCCACTGGCCCAGCATAACAGGCCTTC-3').

Por otro lado, para la expresión y síntesis de la proteína TSG101 humana (SEQ ID NO: 32) por parte de la célula OY216 descrita en la presente invención, se sintetizó el plásmido pRS402-LexA-TSG101 (SEQ ID NO: 30) en dos etapas. En una primera etapa, se construyó el plásmido pRS402-LexA de SEQ ID NO: 22. Para ello, se clonó la secuencia nucleotídica SEQ ID NO: 23 que codifica el dominio de unión al ADN del represor transcripcional LexA (aminoácidos 1-202) de SEQ ID NO: 24, junto con el promotor (SEQ ID NO: 25) y terminador (SEQ ID NO: 26) de *ADH1*, entre los sitios Kpn1 y Sac1 del plásmido integrativo pRS402 (Brachmann CB. *et al.* Yeast. 1998;14(2):115-32). Esta secuencia génica se amplificó mediante PCR utilizando como molde el plásmido pLexA(1-202)PL (Ruden DM, *et al.* Nature. 1991;350(6315):250-2) de SEQ ID NO: 27 y cebadores capaces de hibridar al inicio del promotor de *ADH1* (SEQ ID NO: 25) y al final del terminador de *ADH1* (SEQ ID NO: 26), y con los sitios de las endonucleasas de restricción Kpn1 y Sac1 en los extremos. Los cebadores utilizados para dicho fin fueron los cebadores de SEQ ID NO: 28 (OV502) (5'-GGGTACCTTTTGTGTTTCCGGGTGTAC-3') y SEQ ID NO: 29 (OV503) (5'-GGAGCTCGCATGCCGGTAGAGGTGTGGTC-3'). En la segunda etapa, se obtuvo el plásmido definitivo denominado pRS402-LexA-TSG101 de SEQ ID NO: 30 mediante PCR y clonaje de la secuencia nucleotídica SEQ ID NO: 31 que codifica para la proteína TSG101 (SEQ ID NO: 32) (incluyendo el codón de iniciación y el

codón de parada), en el sitio BamH1 del plásmido previamente construido pRS402-LexA (SEQ ID NO: 22). Como se mostrará a continuación, el plásmido pRS402-LexA-TSG101 (SEQ ID NO: 30) se integra en el cromosoma de la célula de levadura OVY216 descrita en la presente invención, específicamente en el locus del marcador *ADE2* de dicha célula.

Para la amplificación del gen *TSG101* (SEQ ID NO: 31) por PCR, se utilizó como molde una librería de ADNc humano (Clontech ref 638805) y cebadores capaces de hibridar con las regiones 5' y 3' de gen *TSG101* (SEQ ID NO: 31) y con sitios de la endonucleasa de restricción BamH1 en cada extremo. Los cebadores utilizados para dicho fin fueron los cebadores de secuencias SEQ ID NO: 34 (OV310) (5'-GGGATCCTCATGGCGGTGTCGGAGAGCCAGC-3') y SEQ ID NO: 35 (OV311) (5'-GGGATCCTCAGTAGAGGTCAGTACTGAGACCGGC-3').

Por otra parte, el plásmido pLexA(1-202)PL-TSG101 (SEQ ID NO: 33) utilizado en el sistema de doble híbrido clásico, que se utiliza para validar el método de la invención y que se describe a continuación, se obtuvo de la misma manera que el plásmido pRS402-LexA-TSG101 (SEQ ID NO: 30) mencionado previamente, pero utilizando el plásmido pLexA(1-202)PL (SEQ ID NO: 27), en vez del plásmido pRS402-LexA (SEQ ID NO: 22), como vector receptor.

1.2. Diseño y obtención de la cepa OVY216 de la invención

La cepa de levadura OVY216 de la presente invención comprende integrado cromosómicamente las construcciones génicas *(lexA_o)4p-HIS3* y *UASGal-URA3* (=SPAL10::URA3), que expresan los genes reporteros *HIS3* (utilizado en la presente invención como gen reportero de selección positiva) y *URA3* (utilizado en la presente invención como gen reportero de contra-selección), y la construcción génica *ADH1::LexA-TSG101*, que expresa la proteína de fusión LexA-TSG101 (SEQ ID NO: 49), que se une al péptido 3xPTAP (SEQ ID NO: 12) expresado por el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14).

La cepa de levadura OVY216 de la presente invención se ha obtenido mediante el cruce de la cepa comercial procedente de Invitrogen MAV203 (*MAT α* , *leu2-3,112*, *trp1-901*, *his3- Δ 200*, *gal4 Δ* , *gal80 Δ* , *can1R*, *cyh2R*, *LYS2::GAL1-HIS3*, *GAL1-lacZ*, *SPAL10::URA3*) que comprende el gen reportero *URA3* bajo el control de UASGal, y

de la cepa comercial procedente de Invitrogen L40-ura (*MATa*, *leu2-3,112*, *trp1-901*, *his3-Δ200*, *ade2-101*, *gal80Δ*, *LYS2:(lexAop)4-HIS3*, *ura3:(lexAop)8-lacZ*), que comprende el gen reportero *HIS3* bajo el control de LexAop.

5 Los medios de cultivo de levadura utilizados en la presente invención para el crecimiento de las cepas son el medio de cultivo mínimo (SD) y el medio de cultivo completo (YPAD). La preparación de dichos medios de cultivo se describe en *Methods in yeast genetics — A laboratory course manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. 1990; pp 198.

10

Así, para obtener la cepa OVY216 se mezclaron cantidades equivalentes de cada cepa L40-ura y MAV203, en una placa de medio de cultivo completo (YPAD) y se incubaron a 30°C durante 5 horas. La obtención de organismos diploides (*ade2-101/ade2-101*, *his3-Δ200/his3-Δ200*, *leu2-3,112/leu2-3,112*, *trp1-901/trp1-901*,
15 *gal4Δ/GAL4*, *gal80Δ/gal80Δ*, *LYS2::GAL1-HIS3/LYS2:(lexAop)4-HIS3*, *ura3:(lexAop)8-lacZ/SPAL10::URA3*, *GAL1-lacZ/-*) se comprobó por observación al microscopio y varios de ellos fueron recogidos con un micromanipulador (Modelo MSM system SINGER) y depositados en otra placa de cultivo en presencia de medio de cultivo completo YPAD. Trascorridos dos días de cultivo a 30°C, los diploides fueron
20 transferidos a una placa de medio de pre-esporulación (0.8% extracto de levadura, 0.3% bactopectona, 10% glucosa, 2% agar) e incubados a 30°C durante 24 horas, para luego ser transferidos a una placa de medio de esporulación (1% acetato potásico, 0.1% extracto de levadura, 0.05% glucosa, 2% agar) e incubados a 30°C durante 3 días. Posteriormente, se preparó una suspensión de cada uno de los
25 diploides y se comprobó la obtención de tétradas por observación al microscopio. Para la disección de tétradas, se incubó una de las suspensiones de los individuos diploides en presencia de beta-glucuronidasa durante 15 minutos a 25°C y las esporas fueron separadas con el micromanipulador MSM system SINGER y cultivadas en una placa de medio completo YPAD.

30

De dicho cultivo se obtuvieron 70 segregantes haploides viables que se caracterizaron mediante análisis genotípico/fenotípico. Para la caracterización fenotípica se seleccionaron aquellos segregantes que no pueden utilizar la galactosa como fuente de carbono (*gal4Δ*) y que requieren adenina en el medio de cultivo (*ade2-101*). Estos
35 cultivos se llevaron a cabo a 30°C en placas que comprendían medio de cultivo completo con un 2% de galactosa como fuente de carbono o medio mínimo SD sin

adenina. Por otra parte, el análisis genotípico se realizó mediante técnicas de PCR, seleccionando aquellos segregantes que comprenden las construcciones génicas *SPAL10-URA3*, *(lexAop)4-HIS3* y *GAL1-lacZ*. Para la detección de las construcciones génicas mencionadas se amplificó vía PCR un fragmento de 300-400 pb del genoma de los segregantes utilizando como cebadores los descritos a continuación:

Construcción génica	Cebador directo (5'-3')	Cebador inverso (5'-3')
<i>SPAL10-URA3</i>	SEQ ID NO: 36 (OV713) (GCGAGGCATATTTATGG TGAAGG)	SEQ ID NO: 37 (OV714) (CATTTCCGTGCAAAGGTA CTAAC)
<i>(lexAop)4-HIS3</i>	SEQ ID NO: 38 (OV731) (CTGTATATAAAACCAGT GGTTATATGTAC)	SEQ ID NO: 39 (OV732) (TCGAGTGCTCTATCGCTA GGG)
<i>GAL1-lacZ</i>	SEQ ID NO: 40 (OV715) (CCATAGGATGATAATGC GATTAG)	SEQ ID NO: 41 (OV741) (CGCTTCTGGTGCCGGAAA CC)

De los 70 segregantes haploides obtenidos del cruce de las cepas de levadura L40-ura (Invitrogen) y MAV203 (Invitrogen), se ha identificado un único segregante denominado en la presente invención cepa OVY211, que presenta el genotipo: *MATa ade2-101, his3-Δ200, leu2-3,112, trp1-901, gal4Δ, gal80Δ, LYS2:(lexAop)4-HIS3, SPAL10::URA3, GAL1-lacZ*.

En el locus *ADE2* de esta cepa OVY211 se ha integrado el plásmido pRS402-LexA-Tsg101 de SEQ ID NO: 30 (obtenido previamente según se ha descrito) que codifica para la proteína de fusión LexA-Tsg101 (SEQ ID NO: 49), dando lugar a la cepa de levadura de la invención, OVY216 (*MATa ade2-101, his3-Δ200, leu2-3,112, trp1-901, gal4Δ, gal80Δ, LYS2:(lexAop)4-HIS3, SPAL10::URA3, GAL1-lacZ ADE2::LexA-TSG101*) que comprende integrado cromosómicamente las construcciones génicas *(lexAop)4-HIS3* y *UASGal-URA3 (=SPAL10::URA3)*, que expresan los genes reporteros *HIS3* (selección positiva) y *URA3* (contra-selección), y la construcción génica *ADH1::LexA-TSG101*, que expresa la proteína de fusión LexA-TSG101 (SEQ ID NO: 49) que es la proteína que se une al péptido funcional 3xPTAP (SEQ ID NO: 12). (**Figura 2a**). Brevemente, se transformó la cepa de levadura OVY211 con el plásmido pRS402-LexA-TSG101 (SEQ ID NO: 30) linearizado con la endonucleasa de

restricción Stu1 que reconoce un único sitio en el marcador *ADE2* de dicho plásmido. La selección de los transformantes en los cuales dicho plásmido se ha integrado mediante recombinación homóloga en el locus del marcador *ADE2* se llevó a cabo en una placa de medio mínimo SD sin adenina. La correcta integración del plásmido se confirmó mediante amplificación por PCR de un fragmento de 2.3Kb con los cebadores de SEQ ID NO: 42 (OV508) (5'-CAGATTGTAAGAGAGTGCACC-3') y SEQ ID NO: 43 (OV747) (5'-ATTCCTTGCTTCTTGTACTGG-3'). Por lo tanto, la cepa de levadura de la invención, OVY216, tal y como hemos indicado anteriormente comprende integrado cromosómicamente las construcciones génicas *LexAop-HIS3* y *UASGal-URA3* (=SPAL10::URA3), que expresan los reporteros *HIS3* y *URA3*, y la construcción génica *ADH1::LexA-TSG101*, que expresa la proteína de fusión LexA-TSG101 (SEQ ID NO: 49) que es la proteína que se une al péptido funcional 3xPTAP (SEQ ID NO: 12) incluido en el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14).

Adicionalmente, la cepa OVY216 de la presente invención comprende la construcción que expresa el gen reportero inducible *GAL1-lacZ*. Este gen reportero, al igual que la construcción *SPAL10::URA3*, está bajo el control de *UASGal*. La presencia del gen reportero inducible *GAL1-lacZ* permite, si fuese necesario, confirmar o validar los resultados obtenidos con el gen reportero *URA3*, mediante ensayos de actividad β -galactosidasa.

1.3. Transformación de la cepa OVY216 con el plásmido pGBKT7-GKRP de SEQ ID NO: 6

Tras la obtención de la cepa de la invención OVY216 según se ha descrito previamente en 1.2, ésta se transforma con el plásmido pGBKT7-GKRP (SEQ ID NO: 6) para que exprese una de las proteínas del par interaccionante que se estudia en el presente ejemplo, la proteína GKRP (SEQ ID NO: 2). Simultáneamente a la transformación de la célula OVY216 con el plásmido pGBKT7-GKRP de SEQ ID NO: 6 se podría llevar a cabo la transformación de dicha célula, para obtener mediante recombinación *in vivo* "gap repair", el segundo plásmido pACT2-GK-3xPTAP, pero para asegurar la expresión de la proteína de fusión GBD-GKRP (SEQ ID NO: 51), por parte de la célula OVY216, se prefiere llevar a cabo en primer lugar la transformación de dicha célula con el plásmido pGBKT7-GKRP (SEQ ID NO: 6), y una vez que la célula comprende dicho plásmido, posteriormente llevar a cabo la segunda transformación y obtener el plásmido pACT2-GK-3xPTAP recombinado. De esta

manera, el plásmido pGBKT7-GKRP (SEQ ID NO: 6) expresa la proteína de fusión GBD-GKRP (SEQ ID NO: 51) a un nivel suficiente como para activar la expresión del gen reportero de selección positiva URA3 (mediante su interacción con la proteína de fusión GAD-GK-3xPTAP (SEQ ID NO: 53)) expresada por el plásmido pACT2-GK-3xPTAP de SEQ ID NO: 14, e impedir el crecimiento, en presencia de 5-FoA, de los clones que comprenden la proteína GK que no presentan mutaciones *missense*.

El protocolo de transformación de la cepa de levadura OVY216 con el plásmido pGBKT7-GKRP (SEQ ID NO: 6) se llevó a cabo mediante la metodología estándar con acetato de litio. Brevemente, la célula de levadura OVY216 se mantiene en cultivo en medio completo YPAD hasta que llega a la fase logarítmica de crecimiento. A continuación, se centrifuga el cultivo durante 5 min a 2000 rpm y se desecha el sobrenadante. Se resuspende el pellet donde se encuentra la célula OVY216 de la invención en 1 ml de agua y se transfiere a un tubo eppendorf para volver a centrifugarlo durante 5 min a 2000 rpm y volver a desechar el sobrenadante. Posteriormente, se resuspende el pellet de la célula OVY216 en la solución TELiAc (100 µl/5 ml cultivo) que comprende el tampón TE con 0.1 M de LiAc. A continuación, se mezcla en un eppendorf 50 µl de suspensión de la célula OVY216 de la invención con 2 µl (20 µg) de ADN *carrier* y 100 ng del plásmido pGBKT7-GKRP (SEQ ID NO: 6). Adicionalmente, se añaden 250µl de la solución TELiPEG que comprende TE con 0.1 M LiAc y 40% PEG 3350 y se vortea durante 15 seg. Posteriormente, se incuba dicha mezcla durante 30 min a 30°C y transcurrido dicho tiempo se vuelve a incubar durante 15 min a 42 °C. A continuación, se centrifuga 1 min a 13000 rpm y se elimina todo el sobrenadante. Finalmente, se resuspende el pellet en 50 µl de agua y se siembra en una placa con medio de cultivo mínimo SD sin triptófano (SD-T) dado que el marcador de selección del plásmido pGBKT7-GKRP es *TRP1*. Las cepas transformadas que han incorporado el plásmido pGBKT7-GKRP (SEQ ID NO: 6) son visibles a cabo de 3-4 días de cultivo a 30°C.

1.4. Generación y selección simultánea de mutaciones missense mediante PCR-mutagénica y recombinación in vivo ("gap-repair") en la secuencia nucleotídica que codifica para la proteína GK comprendida en el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14).

Para generar mutaciones al azar en la secuencia nucleotídica SEQ ID NO: 9 que codifica para la proteína GK (SEQ ID NO: 10) comprendida en el plásmido pACT2-GK-

3xPTAP (SEQ ID NO: 14), se ha utilizado la técnica de PCR mutagénica y recombinación *in vivo* “*gap-repair*”. La generación de mutaciones mediante dicha técnica se basa en la utilización de condiciones de PCR que favorecen la introducción al azar de mutaciones en el producto de PCR obtenido, haciendo uso de cebadores
5 que permiten la amplificación de la secuencia nucleotídica que codifica para la GK, junto con las secuencias flanqueantes presentes en el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14). La técnica de PCR mutagénica y recombinación *in vivo* se basa en la recombinación, después de transformación, entre un plásmido linearizado y un producto de PCR obtenido en condiciones que favorecen la introducción al azar de
10 mutaciones. La homología entre los extremos del producto de PCR y los extremos del plásmido pACT2 digerido con BamH1, permiten la recombinación homóloga entre ambos (“*gap repair*”) después de co-transformación en la célula de levadura (**Figura 3**). De esta forma, los transformantes obtenidos comprenden un plásmido, a priori, idéntico a pACT2-GK-3xPTAP, pero con mutaciones introducidas al azar en la
15 secuencia nucleotídica que codifica para la proteína GK.

Los cebadores utilizados en la PCR mutagénica han sido los cebadores de SEQ ID NO: 44 (OV621) (5'-CACTGTACCTGGTTGGACGG-3') y de SEQ ID NO: 45 (OV622) (5'-CTATAGATCAGAGGTTACATGGC-3'), que hibridan 213bp corriente
20 arriba y 174bp corriente abajo de la secuencia nucleotídica (SEQ ID NO: 9) que codifica para la proteína GK (SEQ ID NO: 10) en el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14).

En el presente ejemplo se han llevado a cabo dos PCR mutagénicas para introducir
25 mutaciones al azar en la secuencia nucleotídica que codifica para la proteína GK, siendo la única diferencia entre dichas PCRs mutagénicas la enzima ADN polimerasa utilizada en cada una de ellas. Las enzimas ADN polimerasas presentan tasas de mutación diferentes entre ellas, de ahí que se ha querido comparar dos de dichas enzimas para conocer cuál es la más idónea para el método de la invención. En el
30 presente ejemplo se han probado la ADN polimerasa Taq procedente de Takara o Roche, y la ADN polimerasa Mutazyme II procedente del kit de mutagénesis al azar Genemorph II de Agilent.

Las condiciones para las PCR mutagénicas en función de la ADN polimerasa utilizada
35 se describen a continuación:

- ADN polimerasa Mutazyme II (kit de mutagénesis al azar Genemorph II de Agilent). Se utilizan 2.5 µg de ADN molde (pACT2-GK-3xPTAP de SEQ ID NO: 14) y se amplifica durante 20 ciclos para minimizar el número de mutaciones por Kb. Se siguieron las recomendaciones indicadas por el fabricante. Brevemente, la mezcla de reacción (50µl total) comprende: 33.1 µl H₂O + 5 µl buffer 10x + 1 µl dNTPs 10 mM + 0.25 µl de los cebadores de SEQ ID NO: 44 y SEQ ID NO: 45 (100 µM) + 9.4 µl pACT2-GK-3xPTAP (2.5 µg) + 1 µl Mutazyme II (ADN pol). Los ciclos de la PCR fueron: (1) 95 °C 2 min; (2) 95 °C 30 sec + 55 °C 30 sec + 72 °C 1 min/Kb (repetir 20 veces el paso (2)) y (3) 72 °C 10 min.
- ADN polimerasa Taq (Takara Ref. R001 o Roche Ref 11647679001). Se han utilizado condiciones estándar para minimizar el número de mutaciones por Kb. La mezcla de reacción (50 µl total) comprende: 42 µl H₂O + 5 µl buffer + 1 µl dNTPs 10 mM + 0.25 µl de los cebadores de SEQ ID NO: 44 y SEQ ID NO: 45 (100 µM) + 1 µl pACT2-GK-3xPTAP (265 ng) + 0.5 µl Taq (ADN pol). Los ciclos de la PCR fueron (1) 94 °C 2 min; (2) 94 °C 30 sec + 55 °C 30 sec + 72 °C 1 min/Kb (repetir 30 veces el paso (2)) y (3) 72 °C 7 min.

Después de cada reacción de PCR mutagénica, se añadió 1 µl de la enzima *dpn1* (Roche) y se incubó la mezcla durante 1 h a 37°C para eliminar el ADN molde. Posteriormente se precipitó el producto de PCR con acetato de sodio y etanol, para después resuspenderlo en 20 µl H₂O.

Para la técnica de la PCR mutagénica y recombinación *in vivo* se lineariza el plásmido pACT2 (SEQ ID NO: 13) (Clontech) en el *polylinker* con el enzima de restricción BamH1 (**Figura 3**). Para dicha linearización del plásmido pACT2 en el *polylinker* se podrían utilizar otras enzimas de restricción tales como Nco1, Sma1, EcoR1, Sac1 o Xho1. La mezcla de digestión (100 µl total) comprende: 80 µl H₂O + 10 µl buffer 10x + 7.4 µl pACT2 (1.74 µg) + 2.5 µl (25u) BamH1 (Roche). Se incubó la mezcla durante 2h a 37°C. Posteriormente se precipitó el ADN con acetato de sodio y etanol, para después resuspenderlo en 20 µl H₂O.

A continuación, se lleva a cabo la transformación de la célula OVY216, previamente pre-transformada con el plásmido pGBKT7-GKRP de SEQ ID NO: 6, con el plásmido

pACT2 linearizado con BamH1 y los diferentes productos obtenidos en las PCR mutagénicas, según se ha descrito anteriormente.

5 La generación de mutaciones en el plásmido pACT2-GK-3xPTAP (SEQ ID NO: 14) mediante la técnica de PCR mutagénica y recombinación *in vivo*, así como la selección de las mutaciones “*missense*” en la proteína GK que producen una pérdida de interacción con la proteína GKRP, se llevan a cabo en un solo paso mediante la utilización combinada de genes reporteros, específicamente en la presente invención mediante los genes reporteros *SPAL10-URA3* e *(lexAop)4-HIS3*, y el medio selectivo
10 SD-AHTL+5FoA+3-AT, que permiten exclusivamente el crecimiento de los transformantes en los que la mutación en el gen de la GK bloquea su interacción con GKRP, pero no trunca la proteína (mutación *missense*). El compuesto 3-AT (3-aminotriazol) es un inhibidor competitivo del producto del gen *HIS3* que se utiliza en el presente ejemplo para evitar que la actividad basal del reportero *LexA(op)-HIS3* sea
15 suficiente para permitir el crecimiento de la célula en ausencia de histidina. Se ha utilizado una concentración del compuesto 3-AT de 1 mM pero se ha observado que utilizando dicha concentración existen células OVY216 transformadas que son capaces de crecer en ausencia de histidina debido a la actividad basal del reportero *LexA(op)-HIS3*, aunque las colonias obtenidas son mucho más pequeñas que las
20 obtenidas con los transformantes que comprenden la secuencia que codifica para la proteína GK con mutaciones *missense*. Para evitar este inconveniente se ha probado una concentración de 3-AT de 5mM. Esta mayor concentración es capaz de inhibir completamente el crecimiento de estos transformantes.

25 Brevemente, la célula de levadura de la invención OVY216, previamente transformada con el plásmido pGBKT7-GKRP (SEQ ID NO: 6), se transforma ahora con el plásmido pACT2 linearizado con BamH1 junto con los productos de la PCR mutagénica. Para ello, dicha célula se inocula en 5ml de medio de cultivo mínimo sin triptófano (SD-T) y se mantiene en cultivo hasta que su crecimiento llegue a la fase logarítmica. Este
30 cultivo asegura el crecimiento de las células que comprenden el plásmido pGBKT7-GKRP y en consecuencia, expresarán la proteína de fusión GBD-GKRP (SEQ ID NO: 51). A continuación, se centrifuga dicho cultivo y se resuspende el pellet de las células en 15 ml de medio completo (YPAD) para mejorar la eficiencia de la transformación. Se mantiene dicho cultivo hasta que llegue de nuevo a fase logarítmica (alrededor de
35 3-4h). A continuación, se sigue el protocolo de transformación con acetato de litio según se ha explicado previamente, utilizando como mezcla de transformación: 50µl

de suspensión de la célula de levadura en TELiAc + 2 μ l (20 μ g) de ADN carrier + 2 μ l (175 ng) de pACT2 linearizado con BamH1 + 2 μ l (Taq) o 4 μ l (Mutazyme II) de PCR mutagénico. Se utiliza una concentración alta del producto de PCR mutagénico respecto al plásmido linearizado para favorecer su integración mediante recombinación o “*gap repair*” (**Figura 3**). Para incrementar aún más la eficiencia del procedimiento de transformación y asegurar la expresión de los marcadores de selección *URA3* e *HIS3* al final del protocolo de transformación, se resuspende el pellet de las células de levadura transformadas en 5ml de medio completo YPAD y se cultiva a 30°C durante 150 min. Posteriormente, se centrifuga durante 1 min a 13000 rpm y se elimina todo el sobrenadante. A continuación, se resuspende el pellet de levadura en 50 μ l de agua y se siembra el 1% de la mezcla de transformación en una placa control con medio mínimo no selectivo sin triptófano, leucina, ni adenina (SD-TLA) para calcular el número total de transformantes que contienen los dos plásmidos pGBKT7-GKRP (SEQ ID NO: 6) y pACT2-GK-3xPTAP obtenido mediante recombinación (“*gap repair*”), incluyendo los transformantes que no llevan mutaciones o los que llevan mutaciones sin el efecto deseado. Este medio selectivo carece de adenina ya que la cepa es protótrofa para este requerimiento y carece de triptófano y leucina dado que los marcadores de los dos plásmidos son *TRP1* y *LEU2*. Se ha obtenido una media de 500 transformantes en esta placa control, lo que indica que el número total de transformantes rastreados es de 50000, al haberse sembrado el 1% de la mezcla de transformación.

Para seleccionar las células transformadas que expresan la secuencia mutante de la proteína GK que bloquea su unión a la proteína GKRP (mutaciones *missense*) (**Figura 1**), se siembra el resto de la mezcla de transformación (99%) en una placa de cultivo con medio mínimo selectivo sin triptófano, leucina, adenina e histidina y con 0.1% 5-FoA y 1mM 3-AT (SD-AHTL+5-FoA+3-AT). En dicho medio selectivo solo crecerán aquellas células OVY216 transformadas según se describe en el presente ejemplo, que presenten mutaciones *missense* en la secuencia que codifica para la proteína GK y que, por tanto, impidan su unión a la proteína GKRP. Por el contrario, las células OVY216 transformadas con el plásmido pACT2 que ha quedado sin digerir o que se ha recircularizado, así como las células transformadas con el plásmido pACT2-GK-3xPTAP no mutado (SEQ ID NO: 14) o con mutaciones que producen un truncamiento en la proteína GK, no son capaces de crecer en dicho medio selectivo SD-AHTL+5-FoA+3-AT. Adicionalmente, también se ha observado que con 1mM 3-AT en el medio de cultivo selectivo, el bloqueo del crecimiento de los transformantes que no activan el

reportero *HIS3* no es absoluto, existiendo una pequeña actividad basal del reportero (*lexAop*)4-*HIS3*, aunque las colonias obtenidas son mucho más pequeñas que las obtenidas con los transformantes que comprenden la secuencia que codifica para la proteína GK con mutaciones *missense*. Para evitar este inconveniente se ha probado una concentración de 3-AT de 5mM. Esta mayor concentración es capaz de inhibir completamente el crecimiento de estos transformantes.

Es interesante mencionar que el número de transformantes obtenidos en el medio selectivo SD-AHTL+5-FoA+3-AT varía según la ADN polimerasa utilizada para introducir mutaciones al azar en la secuencia nucleotídica que codifica para la proteína GK en la PCR mutagénica. Así, cuando se utiliza la ADN polimerasa Mutazyme II, el número de transformantes en el medio selectivo es el 1.25% del total de los transformantes rastreados. Sin embargo, con la ADN polimerasa Taq, se han obtenido 10 veces menos clones positivos (0.125% del total de transformantes) que los obtenidos con Mutazyme II. Esta diferencia se debe probablemente al hecho de que con la ADN polimerasa Taq se han utilizado condiciones estándar que no favorecen la aparición de mutaciones, con el objetivo de limitar el número de mutaciones por plásmido.

1.5. Análisis de las células OVY216 transformadas según se ha descrito previamente que han crecido en el medio selectivo SD-AHTL+5-FoA+3-AT y purificación de los plásmidos que comprenden las mutaciones *missense* en la secuencia que codifica para la proteína GK humana.

Se ha extraído el ADN plasmídico de 19 transformantes obtenidos en el medio selectivo SD-AHTL+5-FoA+3-AT. En estos 19 transformantes (10 obtenidos con la DNA polimerasa mutazyme II y 9 obtenidos con la DNA polimerasa Taq), que presentan las mutaciones en la secuencia nucleotídica que codifica para la proteína GK, obtenidas mediante la técnica de PCR mutagénica y recombinación *in vivo* "gap-repair", y seleccionadas mediante el método de doble híbrido en reverso descrito en la presente invención, se ha aislado el plásmido pACT2-GK-3xPTAP mutado que comprendían dichos transformantes.

Brevemente, dado que *LEU2* es el marcador para seleccionar este plásmido en levaduras, cada transformante se cultiva en 5mL de medio mínimo sin leucina (SD-L) hasta que llega a la fase estacionaria de su crecimiento. A continuación, se aísla el

ADN plasmídico mediante el kit comercial “*High Pure Plasmid Isolation Kit*” (Roche), que se ha adaptado para la extracción de plásmidos en levadura, ya que dicho kit es específico para purificación de plásmidos en bacterias. Se resuspendió el pellet de los transformantes de levaduras seleccionados en el medio selectivo SD-AHTL+5-FoA+3-AT en 250 µl de la solución 1 del kit “*High Pure Plasmid Isolation Kit*” (Roche) y se añadieron 10µl de zimoliasa 50 mg/ml que se incubó durante 30 min a 37 °C para digerir la pared celular de la levadura. Posteriormente, se añadió la solución 2 de dicho kit y se siguieron las instrucciones del fabricante hasta la obtención del ADN plasmídico purificado. El ADN plasmídico purificado de cada transformante comprende una mezcla del plásmido pGBKT7-GKRP (que comprende un marcador de resistencia a kanamicina) y del plásmido pACT2-GK-3xPTAP con la secuencia génica GK mutada (que comprende un marcador de resistencia a ampicilina). Posteriormente, de dicha mezcla se aísla el plásmido con la secuencia nucleotídica que codifica para la GK mutada. Para ello, se transforma mediante la metodología estándar del cloruro de rubidio la cepa de *E. coli* DH5α con 10µl de la mezcla del ADN plasmídico extraído tal y como se ha mencionado anteriormente. Después de un choque térmico de 2 min a 37°C, se cultivan dichas células bacterianas transformadas en una placa con medio de cultivo LB en presencia de 50µg/ml ampicilina para seleccionar únicamente aquéllos transformantes bacterianos que habían incorporado el plásmido pACT2-GK-3xPTAP que presentaba las mutaciones *missense*.

A continuación, se purificó el plásmido pACT2-GK-3xPTAP que comprende las mutaciones *missense* con el kit “*High Pure Plasmid Isolation Kit*” (Roche). Para ello, se cultivó un transformante bacteriano de cada placa de transformación en 5mL de medio LB con ampicilina y se purificó el plásmido siguiendo las instrucciones del fabricante. Finalmente, se comprobó con el enzima de restricción BamH1 que el plásmido pACT2-GK-3xPTAP mutado presenta el mismo patrón de restricción que el plásmido parental pACT2-GK-3xPTAP (SEQ ID NO: 14) no mutado. En el presente ejemplo, los 19 plásmidos purificados que presentan la secuencia nucleotídica GK mutada mediante la técnica de PCR mutagénica y recombinación *in vivo* “*gap-repair*”, liberaron 2 fragmentos de 8.2 Kb y 1.4 Kb cada uno, al igual que el plásmido pACT2-GK-3xPTAP parental (SEQ ID NO: 14) que no presentaba las mutaciones.

1.6. Validación del método de doble híbrido en reverso de la invención con el sistema de doble híbrido clásico.

Mediante el sistema de doble híbrido clásico (ensayo de actividad β -galactosidasa en filtro) se ha confirmado que los 19 plásmidos pACT2-GK-3xPTAP que comprenden la secuencia de la proteína GK mutada, no interaccionan con el plásmido pGBKT7-GKRP, es decir existe una pérdida de la interacción entre las proteínas GK-GKRP, esto indica que la mutación en la secuencia nucleotídica que codifica para la proteína GK es una mutación *missense* que bloquea su interacción con la proteína GKRP. Pero, por otro lado, mediante el sistema de doble híbrido clásico se ha confirmado que los plásmidos pACT2-GK-3xPTAP que comprenden la secuencia de la proteína GK mutada, si interaccionan con el plásmido pLexA(1-202)PL-TSG101 de SEQ ID NO: 33, a través de la unión a TSG101-3xPTAP, lo que demuestra que estas mutaciones son *missense* y no producen un truncamiento de la proteína.

Para validar los resultados obtenidos mediante el sistema de doble híbrido en reverso descrito en la presente invención, ejemplificado en la detección de mutaciones *missense* que bloquean la interacción entre las proteína GK codificada por el plásmido pACT2-GK-3xPTAP y la proteína GKPR codificada por el plásmido pGBKT7-GKRP, mediante el sistema de doble híbrido clásico, se ha utilizado la cepa de levadura Y187 (Clontech) que comprende el reportero *lacZ* bajo el control de UASGal. Por otro lado, para validar la unión entre la proteína TSG101 codificada por el plásmido pLexA(1-202)PL-TSG101 (SEQ ID NO: 33), y el péptido 3xPTAP codificado por el plásmido pACT2-GK-3xPTAP, que comprende la secuencia GK mutada, se ha utilizado la cepa de levadura CTY10-5d (*Cellular interactions in development: a practical approach* ed. D.A. Hartley Oxford: Oxford University Press 153-17) que comprende el reportero *lacZ* bajo el control de *lexAop*. En ambos casos, se han realizado ensayos de la actividad β -galactosidasa en filtro para cuantificar la activación de los reporteros.

Brevemente, para llevar a cabo dicho análisis de interacción, se transformaron las cepas de levadura mencionadas, Y187 (Clontech) y CTY10-5d, siguiendo el protocolo con acetato de litio descrito previamente, pero utilizando las mezclas siguientes:

- 50 μ l de una suspensión de la cepa de levadura Y187 (Clontech) en TELiAc + 2 μ l (20 μ g) de ADN carrier + 1 μ l (100 ng) del plásmido pGBKT7-GKRP (SEQ ID NO: 6) + 1 μ l (100 ng) del plásmido pACT2-GK-3xPTAP mutante aislado según se ha descrito previamente. Como controles negativos y positivos, hemos co-transformado también pGBKT7-GKRP + pACT2, así como pGBKT7-GKRP + pACT2-GK-3xPTAP (original, no mutado). La mezcla de transformación se sembró en una placa de medio mínimo sin triptófano,

leucina y uracilo (SD-TLU) ya que la cepa Y187 es capaz de sintetizar uracilo y los marcadores de los dos plásmidos son *TRP1* y *LEU2*.

- 50 µl de suspensión de la cepa de levadura CTY10-5d en TELiAc + 2 µl (20 µg) de ADN carrier + 1 µl (100ng) del plásmido pLexA(1-202)PL-TSG101 + 1 µl (100 ng) del plásmido pACT2-GK-3xPTAP mutante. Como controles negativos y positivos, hemos co-transformado también pLexA(1-202)PL-TSG101 + pACT2, así como pLexA(1-202)PL-TSG101 + pACT2-GK-3xPTAP (original, no mutado). La mezcla de transformación se sembró en una placa de medio mínimo sin histidina, leucina y uracilo (SD-HLU) ya que la cepa CTY10-5d es capaz de sintetizar uracilo y los marcadores de los dos plásmidos son *HIS3* y *LEU2*.

Después de 3-4 días de cultivo a una temperatura de 30°C, replicamos de 8-10 transformantes de cada mezcla de transformación a placas con los medios de cultivo SD-TLU o SD-HLU, y se mantuvieron en cultivo durante 2 días a 30 °C. Transcurrido dicho tiempo se llevó a cabo el ensayo de actividad β-galactosidasa en filtro (Breedon L, Nasmyth K. Cold Spring Harb Symp Quant Biol. 1985;50:643-50) para cuantificar la activación de los reporteros. Para ello se colocó un filtro de nitrocelulosa encima de la placa de cultivo para transferir, mediante aplicación suave, la levadura al filtro. Se incubó el filtro a -80 °C durante 1 h para permeabilizar las células y posteriormente se colocó el filtro encima de papel Whatman 3MM mojado en tampón Z (Na₂HPO₄ 60 mM, NaH₂PO₄ 40 mM, KCl 10 mM, MgSO₄ 1 mM, β-mercaptoetanol 38 mM) con 0.1% X-Gal y se incubó durante 1 h a 30 °C. La interacción entre las proteínas produce la activación del gen reportero *lacZ*, y en consecuencia la expresión de la β-galactosidasa y la hidrólisis del X-Gal en un compuesto de color azul.

Los resultados obtenidos indican que en todos los casos, es decir con los 19 plásmidos pACT2-GK-3xPTAP que comprenden la secuencia codificante de la proteína GK mutada, no se detecta actividad β-galactosidasa en la cepa Y187 co-transformada con cada uno de estos plásmidos y con pGBKT7-GKRP, lo que confirma el hecho de que las proteínas GKRP y GK mutadas no interaccionan en el sistema de doble-híbrido clásico y no activan el reportero que comprende *lacZ* bajo el control de UASGal. Este resultado indica que la selección de mutaciones en GK que bloquean su interacción con GKRP ha sido eficiente al 100% con este método. Por otra parte, los resultados obtenidos indican que en todos los casos, se detecta actividad β-galactosidasa en la cepa CTY10-5d co-transformada con cada uno de estos

plásmidos y pLexA(1-202)PL-TSG101, lo que confirma el hecho de que las proteínas TSG101 y GK mutadas siguen interaccionando mediante la unión entre TSG101 y el péptido 3xPTAP fusionado al extremo C-terminal de GK y activan el reportero que comprende *lacZ* bajo el control de *lexAop*. Este resultado demuestra que la selección de mutaciones *missense* en GK que no truncan la proteína y en consecuencia no eliminan el péptido 3xPTAP ha sido eficiente al 100% con este método. En conjunto, estos resultados muestran que la selección de mutaciones *missense* en GK que bloquean su interacción con GKRP ha sido eficiente al 100% con este método. En la **Figura 4**, se presentan los ensayos de la actividad β -galactosidasa en filtro para 10 de los mutantes seleccionados.

Ejemplo 2. Identificación de las mutaciones *missense* presentes en la secuencia que codifica para la proteína GK y que impiden su unión a la proteína GKRP.

Las mutaciones *missense* en los 19 plásmidos mutados aislados y validados con el sistema de doble-híbrido clásico según se ha descrito anteriormente, han sido identificadas mediante secuenciación del inserto que comprende el gen de la GK en cada plásmido. Los oligonucleótidos utilizados para la secuenciación y la identificación de las mutaciones *missense* en la secuencia nucleotídica del gen que codifica para la proteína GK fueron los siguientes: SEQ ID NO: 46 (OV284) (5'-CGATGATGAAGATACCCACC-3'), que hibrida 67bp corriente arriba del *polylinker* donde la secuencia nucleotídica del gen GK está clonada y SEQ ID NO: 47 (OV285) (5'-GAGATGGTGCACGATGCACAG-3'), que hibrida 120bp corriente abajo del *polylinker* donde la secuencia nucleotídica del gen GK está clonada.

Por otro lado, también se han identificado mediante secuenciación las mutaciones presentes en 5 transformantes adicionales, diferentes a los 19 anteriores y procedentes de la transformación con el producto de PCR obtenido con la ADN polimerasa Taq. Para estos 5 transformantes adicionales, el plásmido analizado no ha sido extraído ni validado previamente en el sistema de doble-híbrido clásico. El inserto que comprende el gen de la GK mutado ha sido amplificado directamente desde el ADN plasmídico extraído de los 5 transformantes adicionales y secuenciado. Teniendo en cuenta la alta eficiencia del sistema de doble híbrido en reverso descrito en la presente invención (100% de positivos detectados-100% mutaciones de tipo *missense* con los 19 transformantes), el hecho de poder amplificar y secuenciar directamente el ADN plasmídico, sin tener que aislar el plásmido y validarlo permite ahorrar tiempo. La

extracción del plásmido mutado y su validación con el sistema de doble-híbrido clásico puede llevarse a cabo posteriormente, una vez que la mutación haya sido identificada, lo que evita validar mutantes repetidos o ya conocidos.

5 La identificación de las mutaciones en los 5 transformantes adicionales ha sido realizada de la forma siguiente: Extracción del ADN plasmídico de los transformantes de levadura seleccionados mediante el sistema de doble híbrido en reverso de la invención (ver apartado 1.5) y utilización de este ADN como molde para amplificar el gen de la GK mediante PCR con los oligonucleótidos utilizados para la secuenciación (SEQ ID NO: 46 (OV284) y SEQ ID NO: 47 (OV285)). Posteriormente se secuencia el producto de PCR amplificado con estos mismos oligonucleótidos para identificar la ó las mutaciones presentes en el inserto.

15 La secuenciación del total de los 24 clones analizados (19+5) junto con los ensayos de validación con el sistema de doble híbrido clásico ha demostrado que todos ellos comprenden mutaciones *missense* que provocan una pérdida de interacción entre la GK y la GKRP. Esto demuestra la eficacia del método ya que el 100% de los clones analizados, un total de 24, comprenden mutaciones *missense* en el gen de la GK que bloquean su interacción con la proteína GKRP.

20 Los resultados obtenidos con la secuenciación de los 24 mutantes analizados ha permitido la identificación de 21 mutaciones *missense* responsables de la pérdida de la interacción de la proteína GK con la proteína GKRP (Tabla 1). Varias de estas mutaciones están repetidas y algunos clones que comprendían varias mutaciones no han sido caracterizados.

Tabla 1. *Mutaciones missense identificadas en la proteína GK que impiden la interacción con la proteína GKRP (substituciones aa).*

Substitución	Repetición
A201V	1
T60A	2
G72W	1
G72E	1
S64P	4

Substitución	Repetición
L75P	4
T209P	1
C220R	1
C233Y	1
C252Y	3
G407D	1
L306R	1

La secuenciación de los 24 mutantes muestra que la frecuencia de mutaciones *missense* en cada clon es distinta en función de la ADN polimerasa utilizada. La situación óptima sería obtener una única mutación *missense* en la secuencia codificante de la GK (1.5 Kb). En el caso de que sean varias, es necesario analizar en el sistema de doble híbrido clásico el efecto de cada mutación de forma aislada para saber cuál es la responsable de la pérdida de interacción. En este caso, los plásmidos que comprendían una sola mutación se han obtenido mediante mutagénesis dirigida.

5

Utilizando la ADN polimerasa Mutazyme II, se han obtenido cerca de 2 mutaciones *missense* por Kb de media. En el caso de la ADN polimerasa Taq, se han obtenido una media de 1 mutación *missense* por Kb (la mitad que con Mutazyme II). Esta diferencia es consistente con el hecho de que se han obtenido 10 veces menos transformantes en el medio selectivo SD-AHTL+5-FoA+3-AT cuando se ha utilizado la ADN polimerasa Taq que cuando se ha utilizado la ADN polimerasa Mutazyme II. Aunque el resultado con Taq es mejor en cuanto a frecuencia de mutaciones, esta polimerasa presenta un espectro mutacional con una fuerte tendencia para los cambios AT a GC, lo que hace que muchas de las mutaciones obtenidas se repitan. Así, tal y como muestran los resultados obtenidos, la ADN polimerasa Taq da lugar a un 70% de mutaciones *missense* repetidas, mientras que la ADN polimerasa Mutazyme II solo da lugar a un 40% de mutaciones repetidas.

10

15

20

25

Una vez identificadas las mutaciones *missense* que impiden la unión entre las proteínas GK y GKRP, se procedió a comparar dichas mutaciones con la estructura ya conocida del complejo GK-GKPR (Choi JM *et al.* Proc Natl Acad Sci U S A. 2013 Jun 18;110(25):10171-6), poniéndose de manifiesto la buena correlación existente entre

los resultados obtenidos en la identificación de mutaciones *missense* mediante el método de doble híbrido en reverso de la invención y la estructura del complejo, ya que:

- 5 – los residuos aminoacídicos Ala201 y Thr60 se localizan junto con el bolsillo hidrofóbico de la GK que media la interacción con GKRP (Proc Natl Acad Sci U S A. 2006 Sep 5;103(36):13368-73),
- 10 – el residuo aminoacídico Gly72 está próximo a los residuos aminoacídicos Ala201 y Thr60, y el puente de hidrogeno que forma con el residuo aminoacídico Tyr215 parece estabilizar la conformación súper-abierta de la GK que interacciona con GKRP (J Biol Chem. 2006 Dec 29;281(52):40201-7),
- 15 – las sustituciones a prolina de los residuos S64P, L75P y T209P, pueden tener un efecto indirecto en la rotura de la interacción (vía Thr60, Gly72 y Ala201), ya que el residuo Ser64 se localiza en la lámina beta adyacente a Thr60 mientras que Leu75 y Thr209 se localizan en láminas beta adyacentes a los motivos conteniendo Gly72 y Ala201, y debido a la interrupción de estas láminas beta por los residuos de prolina,
- 20 – los residuos de cisteína: Cys220, Cys233 y Cys252, forman parte de un anillo de 5 residuos de cisteína, próximo al bolsillo hidrofóbico y probablemente implicados en la formación de puentes disulfuro y en el mantenimiento de la conformación de la GK (Arch Biochem Biophys. 2000 Mar 15;375(2):251-60). Además, el residuo aminoacídico Gly407 se localiza junto al anillo de residuos de cisteína.

Ejemplo 3. Comparación entre sistema de doble híbrido en reverso de la invención que comprende el gen reportero de contra-selección *URA3* respecto al mismo sistema de doble híbrido en reverso pero utilizando como sistema de contra-selección el gen represor TetR.

Para poner de manifiesto que los reporteros de selección utilizados en el sistema de doble híbrido de la invención presentan ventajas adicionales sorprendentes respecto de otros reporteros conocidos en el estado de la técnica, se ha probado un sistema alternativo a la contra-selección con el gen reportero *URA3* y 5-FoA, basado en el gen represor TetR (Shih HM *et al.* Proc Natl Acad Sci U S A. 1996;93:13896-901). En este sistema de doble híbrido en reverso alternativo, la unión de las proteínas interaccionantes a estudiar activa la transcripción de TetR que reprime la transcripción

del reportero *ADE2*. En consecuencia, siguiendo con el ejemplo mostrado en la presente invención, una mutación *missense* en la secuencia codificante de la proteína GK que bloquea su interacción con GKRP activa el reportero *ADE2*, lo que permite el crecimiento de la cepa en ausencia de adenina. Para probar este sistema, se ha integrado la proteína de fusión *ADH1::LexA-TSG101* en el locus *URA3* de la cepa LY26 (Thomas LR *et al.* J Biol Chem. 2002;277:34343-8) para obtener la cepa OVY158 (*MATalpha can1 his3 leu2 met15 trp1 ura3 gal4::hisG gal80::hisG LYS2::LexA(op)-HIS3 TetO-ADE2 ho::KanMX::GAL1-TetR URA3::LexA-TSG101*).

10 En este caso, la pérdida de interacción entre las proteínas GK y GKRP se selecciona mediante crecimiento en un medio selectivo con ausencia de adenina (medio de cultivo mínimo (SD) con arginina y metionina como únicos requerimientos y 1 mM 3-AT) en vez de crecimiento en presencia de 5-FoA. Se ha observado que este sistema de doble híbrido en reverso utilizando como método de contra-selección TetR en lugar de *URA3* y 5-FoA no es tan eficiente como el sistema de doble híbrido en reverso con 5-FoA, ya que el crecimiento de los clones no mutados no está completamente bloqueado apareciendo en los cultivos colonias más pequeñas. Para intentar solucionar este problema se ha utilizado sacarosa como fuente de carbono, en lugar de glucosa, intentando así eliminar el efecto negativo de la glucosa sobre el promotor *GAL1p* y, en consecuencia la expresión del reportero TetR, se ha resuelto sólo parcialmente este problema, pero este sistema es menos eficiente que el sistema que utiliza *URA3* y 5-FoA, ya que mediante el uso de TetR se obtienen un 15% de falsos positivos, mientras que, como hemos mencionado anteriormente, mediante el sistema de doble híbrido en reverso utilizando *URA3*, *HIS3* y 5-FoA, descrito en la presente invención no se obtienen falsos positivos.

REIVINDICACIONES

1. Método *in vitro* para identificar al menos una mutación *missense* en una proteína de referencia donde dicha mutación afecta a la capacidad de unión de dicha proteína de referencia con otra proteína diana, donde dicho método comprende:
- 5
- a) al menos una célula huésped que comprende integrado en su genoma:
- i) Una primera secuencia nucleotídica que codifica para un gen reportero, donde dicha secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia nucleotídica reconocida por una proteína que se une al ADN,
- 10
- ii) Una segunda secuencia nucleotídica que codifica para un segundo gen reportero, donde dicha segunda secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia nucleotídica reconocida por una proteína que se une al ADN, con la condición de que dicha secuencia nucleotídica sea reconocida por un dominio de unión al ADN distinto del de i), y
- 15
- iii) Una tercera secuencia nucleotídica que codifica para una primera proteína de fusión que comprende el dominio de unión al ADN de ii) y una proteína heteróloga, capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia de estudio, donde la secuencia nucleotídica que codifica para dicha primera proteína de fusión está operativamente unida a un promotor,
- 20
- 25
- b) Pre-transformar la célula de la etapa a) con un plásmido que comprende la secuencia nucleotídica que codifica para una segunda proteína de fusión que comprende el dominio de unión al ADN de i) y la proteína diana, donde la secuencia nucleotídica que codifica para dicha segunda proteína de fusión está operativamente unida a un promotor,
- 30

- c) Cultivar la célula de la etapa b) bajo condiciones que permitan exclusivamente el crecimiento de las células que hayan incorporado el plásmido de la etapa b),
- 5 d) Transformar la célula de la etapa c) con un vector linearizado y al menos un fragmento de ADN que previamente ha sido sometido a mutagénesis y que comprende la secuencia nucleotídica que codifica para la proteína de referencia con al menos una mutación, donde entre el vector linearizado y el fragmento de ADN sometido a mutagénesis se produce recombinación homóloga obteniéndose un plásmido que
- 10 comprende la secuencia nucleotídica que codifica para una tercera proteína de fusión que comprende el dominio de transactivación de Gal4 y la proteína de referencia con al menos una mutación, donde la secuencia nucleotídica que codifica para dicha tercera proteína de fusión está operativamente unida a un promotor,
- 15 e) Cultivar la célula de la etapa d) bajo condiciones tales que permitan exclusivamente el crecimiento de las células que presentan mutaciones *missense* que impiden la unión entre la proteína de referencia y la proteína diana,
- 20 f) Comparar la secuencia de la proteína de referencia mutada con la secuencia de la proteína de referencia *wild-type* e identificar la mutación *missense* que impide la unión de la proteína de referencia con la proteína diana.
- 25 2. Método según la reivindicación 1 donde el procedimiento de mutagénesis al que ha sido sometido el fragmento de ADN de la etapa d) es un procedimiento de mutagénesis *in vitro* o *in vivo*.
3. Método según la reivindicación 1 donde el plásmido de la etapa d) se obtiene mediante recombinación *in vivo* "gap-repair".
4. Método según cualquiera de las reivindicaciones 1 a 3 donde el gen reportero se selecciona de entre genes reporteros de selección positiva y genes reporteros de contra-selección.
- 30 5. Método según la reivindicación 4 donde los genes reporteros de selección positiva se seleccionan del grupo que consiste en: *HIS3*, *LEU2*, *URA3*, *ADE2*,

TRP1, *LYS2* y *LYS5*; y los genes reporteros de contra-selección se seleccionan del grupo que consiste en: *URA3*, *TRP1*, *LYS2*, *LYS5*, *CYH2*, *CAN1*, *GAL1* y *mazF*.

- 5 6. Método según la reivindicación 5 donde el gen reportero de selección positiva es el gen *HIS3*, y el gen reportero de contra-selección es el gen *URA3*
7. Método según cualquiera de las reivindicaciones 1 a 6 donde las secuencias de reconocimiento para dominios de unión al ADN comprenden los sitios de unión de una proteína seleccionada de entre cualquiera de las siguientes: Gal4, LexA y Ace1.
- 10 8. Método según cualquiera de las reivindicaciones 1 a 7 donde la secuencia nucleotídica que codifica para un dominio de transactivación comprende los dominios de activación transcripcional de una proteína seleccionada de entre cualquiera de las siguientes: Gal4, VP16 y Ace1.
- 15 9. Método según cualquiera de las reivindicaciones 1 a 8 donde los promotores son promotores constitutivos.
10. Método según la reivindicación 9 donde los promotores constitutivos se seleccionan de entre cualquiera de la lista que consiste en: *ADH1*, *PGK1*, *TEF1*, *TPI1*, *HXT7*, *TDH3* y *PYK1*.
- 20 11. Método según cualquiera de las reivindicaciones 1 a 10 donde la proteína heteróloga es la proteína de SEQ ID NO: 32, y el péptido funcional es el péptido de SEQ ID NO: 12.
12. Método según cualquiera de las reivindicaciones 1 a 11 donde la primera proteína de fusión es la proteína de SEQ ID NO: 49.
- 25 13. Método según cualquiera de las reivindicaciones 1 a 12 donde la segunda proteína de fusión es la proteína de SEQ ID NO: 51.
14. Método según cualquiera de las reivindicaciones 1 a 13 donde la célula se selecciona del grupo que consiste en: célula de levadura, célula bacteriana y célula de mamífero.
- 30 15. Método según la reivindicación 14 donde la célula de levadura se selecciona del grupo que consiste en *Yarrowia lipolytica*, *Pichia pastoris* y *Scacharomyces*

cerevisiae, preferentemente la célula de levadura pertenece a la especie *S. cerevisiae*.

16. Método según cualquiera de las reivindicaciones 1 a 15 donde la célula es OVY216.
- 5 17. Método según cualquiera de las reivindicaciones 1 a 16 caracterizado por que las células que presentan al menos una mutación *missense*, crecen en un medio de cultivo que comprende 5-FoA y 3-AT, y carece de triptófano, leucina, adenina e histidina.
- 10 18. Método según cualquiera de las reivindicaciones 1 a 17 donde la identificación de las mutaciones *missense*, se lleva a cabo mediante técnicas de amplificación y secuenciación.
19. Célula huésped que comprende integrado en su genoma:
- 15 i) Una primera secuencia nucleotídica que codifica para un gen reportero, donde dicha secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN,
- 20 ii) Una segunda secuencia nucleotídica que codifica para un segundo gen reportero, donde dicha segunda secuencia nucleotídica está operativamente unida a un promotor que comprende una secuencia de reconocimiento para una proteína que se une al ADN, con la condición de que dicha secuencia de reconocimiento sea diferente de la secuencia de reconocimiento de i), y
- 25 iii) Una tercera secuencia nucleotídica que codifica para una primera proteína de fusión que comprende el dominio de unión al ADN de ii) y una proteína heteróloga capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia, donde la secuencia nucleotídica que codifica para dicha primera proteína de fusión está operativamente unida a un promotor.
- 30 20. Célula huésped según la reivindicación 17, donde el gen reportero se selecciona de la lista que consiste en genes reporteros de selección positiva y genes reporteros de contra-selección.

21. Célula huésped según la reivindicación 20 donde los genes reporteros de selección positiva se seleccionan del grupo que consiste en: *HIS3*, *LEU2*, *URA3*, *ADE2*, *TRP1*, *LYS2*, *LYS5*; y los genes reporteros de contra-selección se seleccionan del grupo que consiste en: *URA3*, *TRP1*, *LYS2*, *LYS5*, *CYH2*, *CAN1*, *GAL1*, *mazF*.
22. Célula huésped según la reivindicación 21 donde el gen reportero de selección positiva es el gen *HIS3*, y el gen reportero de contra-selección es el gen *URA3*
23. Célula huésped según cualquiera de las reivindicaciones 19 a 22 donde las secuencias de reconocimiento para dominios de unión al ADN comprenden los sitios de unión de una proteína seleccionada de entre cualquiera de las siguientes: Gal4, LexA y Ace1.
24. Célula huésped según cualquiera de las reivindicaciones 19 a 23 donde la secuencia nucleotídica que codifica para un dominio de transactivación comprende los dominios de activación transcripcional de una proteína seleccionada de entre cualquiera de las siguientes: Gal4, VP16 y Ace1.
25. Célula huésped según cualquiera de las reivindicaciones 19 a 24 donde los promotores son promotores constitutivos.
26. Célula huésped según la reivindicación 25 donde los promotores constitutivos se seleccionan de entre cualquiera de la lista que consiste en: *ADH1*, *PGK1*, *TEF1*, *TPI1*, *HXT7*, *TDH3* y *PYK1*.
27. Célula huésped según cualquiera de las reivindicaciones 19 a 26 donde la proteína heteróloga es la proteína de SEQ ID NO: 32, y el péptido funcional es el péptido de SEQ ID NO: 12.
28. Célula huésped según cualquiera de las reivindicaciones 19 a 27 donde la primera proteína de fusión es la proteína de SEQ ID NO: 49.
29. Célula huésped según cualquiera de las reivindicaciones 19 a 28 donde la célula se selecciona del grupo que consiste en: célula de levadura, célula bacteriana y célula de mamífero.
30. Célula huésped según la reivindicación 29 donde la célula de levadura se selecciona del grupo que consiste en: *Yarrowia lipolytica*, *Pichia pastoris* y *Saccharomyces cerevisiae*, preferentemente la célula de levadura pertenece a

la especie *S. cerevisiae*.

31. Célula huésped según cualquiera de las reivindicaciones 19 a 30 donde la célula es la célula de levadura OY216.

5 32. Construcción génica que comprende las secuencias nucleotídicas que codifican para:

i) un promotor

ii) un dominio de unión al ADN, y

iii) la proteína heteróloga capaz de unirse a un péptido funcional localizado en el extremo C-terminal de la proteína de referencia.

10 33. Construcción génica según la reivindicación 26 donde el promotor es un promotor constitutivo, preferentemente el promotor de *ADH1*, el dominio de unión al ADN es preferentemente el dominio de unión de la proteína LexA, y la proteína heteróloga es preferentemente la proteína humana TSG101 que comprende la SEQ ID NO: 32.

15 34. Construcción génica según cualquiera de las reivindicaciones 32 a 33 caracterizada por que es un plásmido, preferentemente el plásmido pRS402-LexA-Tsg101 que comprende la secuencia SEQ ID NO: 30.

35. Construcción génica que comprende las secuencias nucleotídicas que codifican para:

20 i) un promotor

ii) un dominio de transactivación y

iii) una proteína de referencia que en su extremo carboxilo terminal comprende la secuencia que codifica para el péptido funcional 3xPTAP de SEQ ID NO: 12.

25 36. Construcción génica según la reivindicación 35 donde el promotor es un promotor constitutivo, preferentemente el promotor de *ADH1*, el dominio de transactivación es preferentemente el dominio de transactivación del activador transcripcional Gal4, y la proteína de referencia que comprende en su extremo carboxilo terminal la secuencia que codifica para el péptido funcional 3xPTAP

de SEQ ID NO: 12.

37. Construcción génica según las reivindicaciones 35 a 36 caracterizada por que es un plásmido, preferentemente el plásmido pACT2-GK-3xPTAP que comprende la secuencia SEQ ID NO: 14.
- 5 38. Construcción génica que comprende las secuencias nucleotídicas que codifican para:
- i) un promotor
 - ii) un dominio de unión al ADN, y
 - iii) una proteína diana a la que se une la proteína de referencia.
- 10 39. Construcción génica según la reivindicación 38 donde el promotor es un promotor constitutivo, preferentemente el promotor es el promotor de *ADH1*, el dominio de unión al ADN es preferentemente el del activador transcripcional Gal4, y la proteína diana.
- 15 40. Construcción génica según cualquiera de las reivindicaciones 38 a 39 caracterizada por que es un plásmido, preferentemente el plásmido PGBKT7-GKRP que comprende la secuencia SEQ ID NO: 6.

Figura 1a

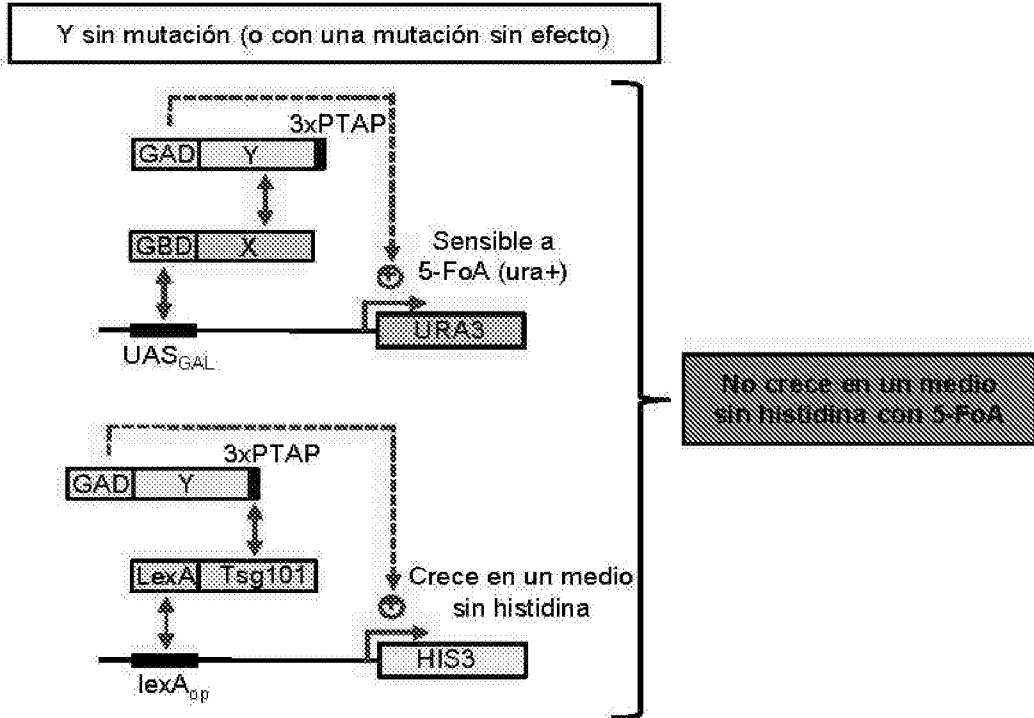


Figura 1b

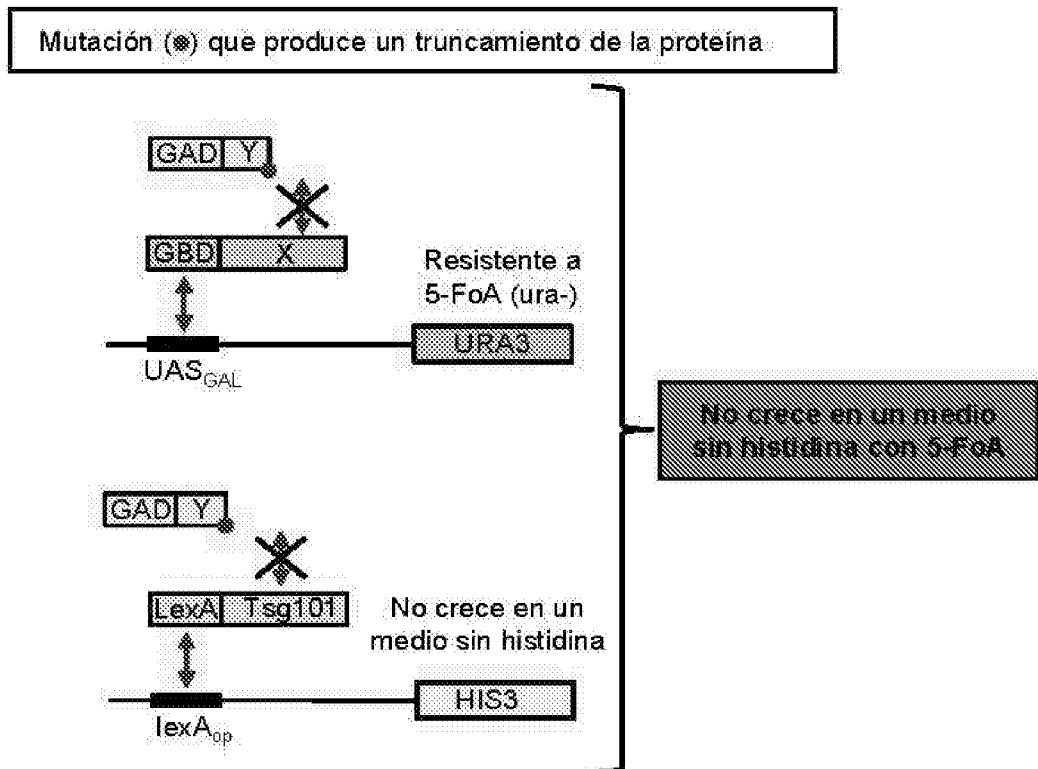


Figura 1c

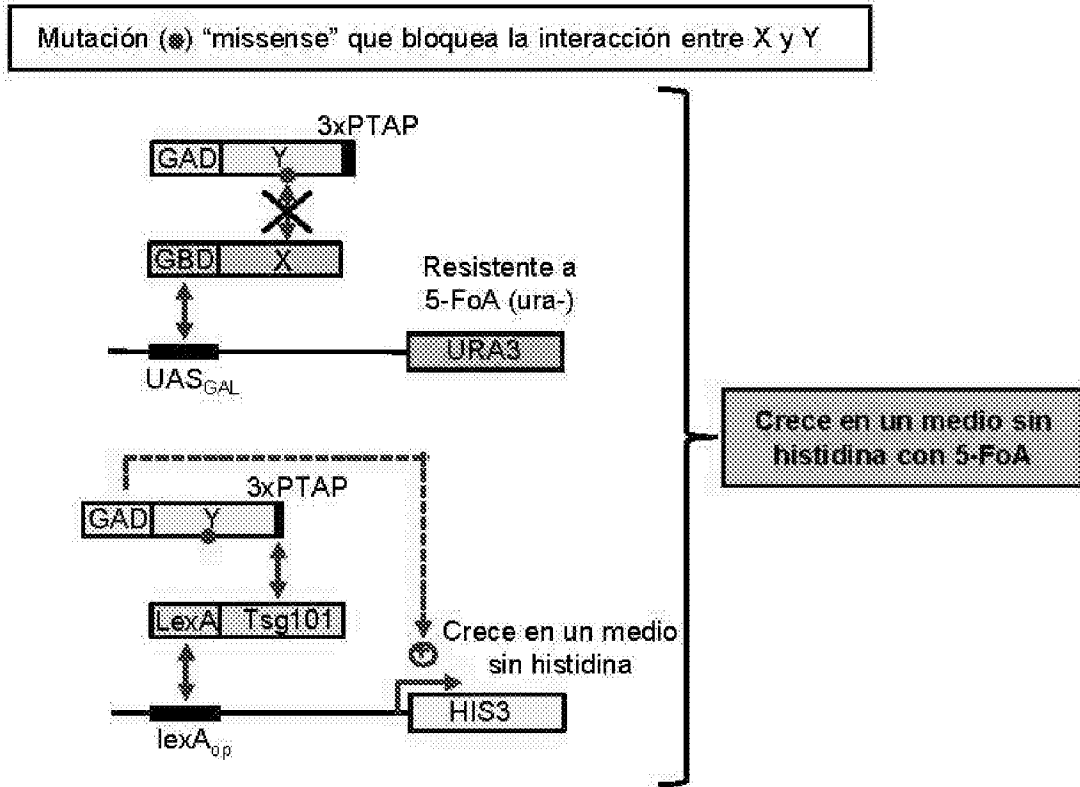


Figura 2a

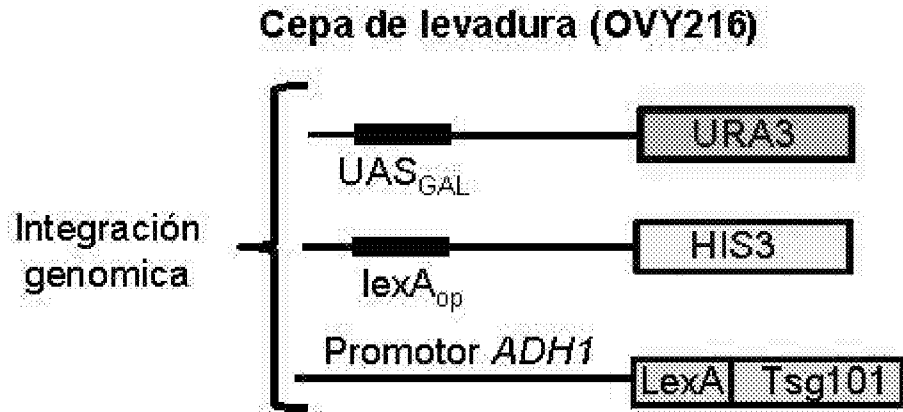


Figura 2b

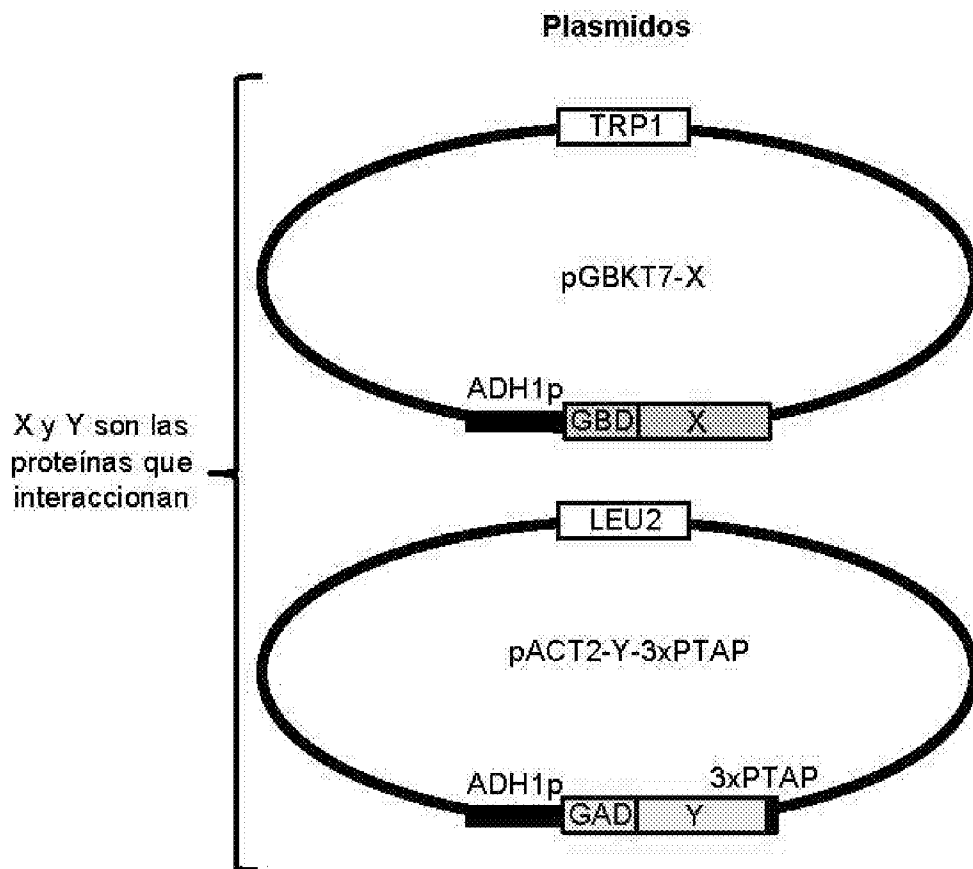


Figura 3

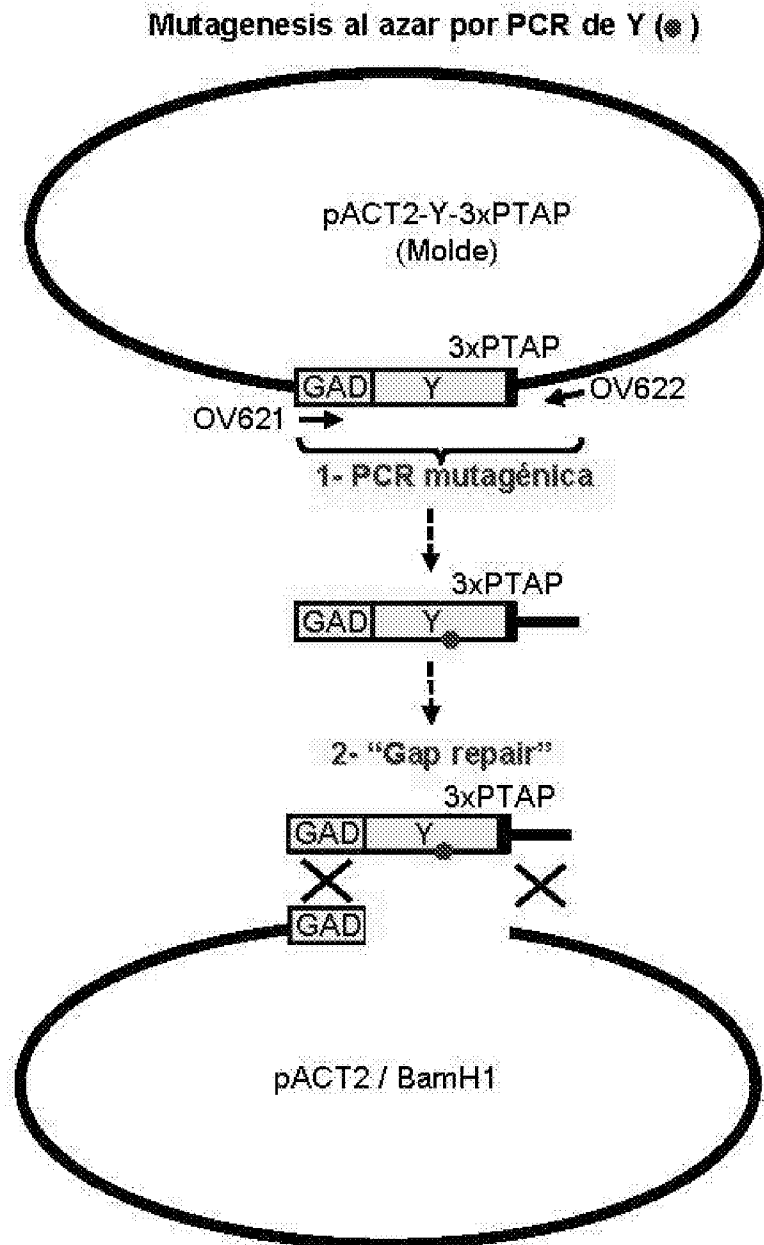
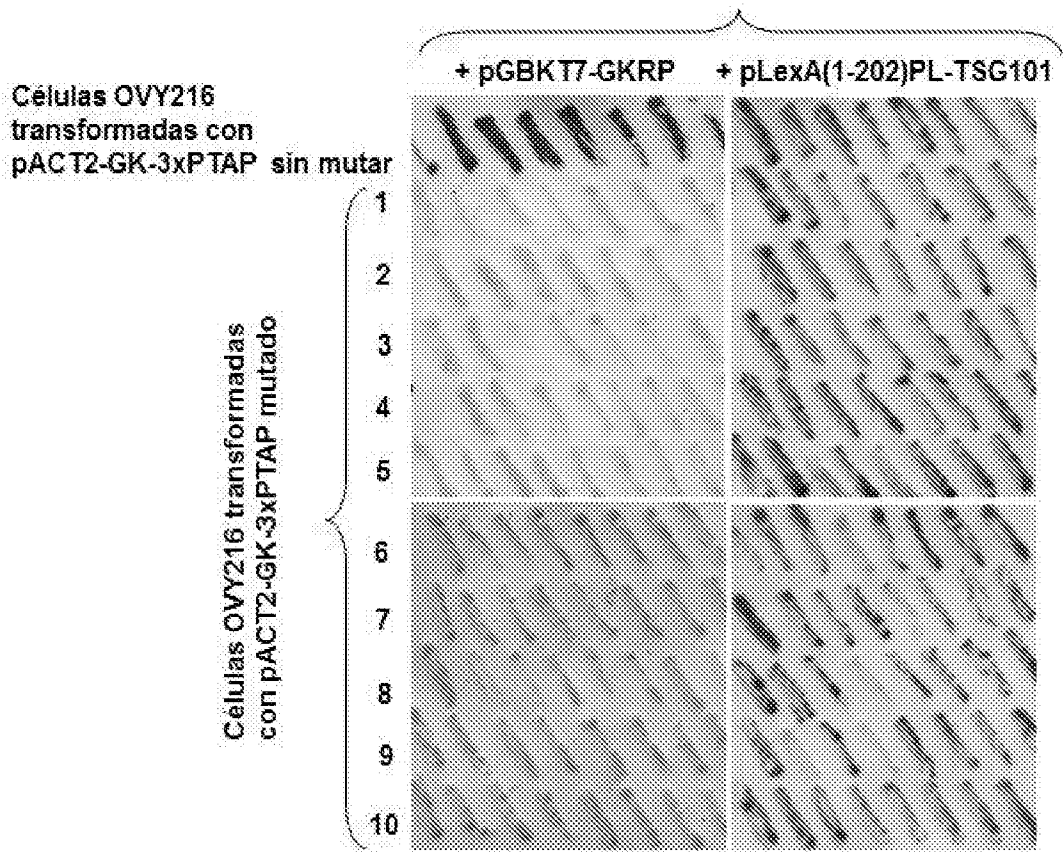


Figura 4



ES 2 646 388 A1

LISTADO DE SECUENCIAS

<110> Consejo Superior de Investigaciones Científicas (CSIC)
 <120> Método de doble híbrido en reverso para la identificación de mutaciones missense
 <130> 1641.1111
 <160> 53
 <170> PatentIn version 3.5
 <210> 1
 <211> 1878
 <212> DNA
 <213> Homo sapiens

 <220>
 <221> CDS
 <222> (1)..(1878)

 <400> 1
 atg cca ggc aca aaa cgg ttt caa cat gtc att gag acc ccg gag cct 48
 Met Pro Gly Thr Lys Arg Phe Gln His Val Ile Glu Thr Pro Glu Pro
 1 5 10 15

 ggc aag tgg gag ttg tct ggg tac gag gca gct gtg cca atc acg gag 96
 Gly Lys Trp Glu Leu Ser Gly Tyr Glu Ala Ala Val Pro Ile Thr Glu
 20 25 30

 aag tca aac cca ctg acc cag gat cta gac aaa gca gat gct gag aac 144
 Lys Ser Asn Pro Leu Thr Gln Asp Leu Asp Lys Ala Asp Ala Glu Asn
 35 40 45

 att gtt cga ctg cta ggg caa tgt gat gct gag atc ttc cag gag gag 192
 Ile Val Arg Leu Leu Gly Gln Cys Asp Ala Glu Ile Phe Gln Glu Glu
 50 55 60

 ggg caa gcc ctg tcc aca tac cag aga ctc tac agc gaa tcc att ctg 240
 Gly Gln Ala Leu Ser Thr Tyr Gln Arg Leu Tyr Ser Glu Ser Ile Leu
 65 70 75 80

 acc acc atg gta cag gtg gct ggg aaa gtt cag gaa gtg ctg aag gag 288
 Thr Thr Met Val Gln Val Ala Gly Lys Val Gln Glu Val Leu Lys Glu
 85 90 95

 cca gat ggg ggg ctg gtt gtg ctg agt gga ggg ggc acc tct ggc cgg 336
 Pro Asp Gly Gly Leu Val Val Leu Ser Ser Gly Gly Thr Ser Gly Arg
 100 105 110

 atg gca ttc ctc atg tcg gtg tcc ttt aat cag ctg atg aaa ggt ctg 384
 Met Ala Phe Leu Met Ser Val Ser Phe Asn Gln Leu Met Lys Gly Leu
 115 120 125

 gga cag aaa cct ctt tac acc tac ctc att gca ggt ggt gac agg tct 432
 Gly Gln Lys Pro Leu Tyr Thr Tyr Leu Ile Ala Gly Gly Asp Arg Ser
 130 135 140

 gtg gtg gcc tct agg gag ggg aca gaa gat agt gcc ttg cac ggg att 480
 Val Val Ala Ser Arg Glu Gly Thr Glu Asp Ser Ala Leu His Gly Ile
 145 150 155 160

 gag gaa ctg aag aag gtg gct gcc ggg aag aag aga gtg att gtc att 528
 Glu Glu Leu Lys Lys Val Ala Ala Gly Lys Lys Arg Val Ile Val Ile
 165 170 175

 ggc att tct gtg gga ctc tct gct ccc ttt gtg gca ggc cag atg gac 576
 1

ES 2 646 388 A1

Gly	Ile	Ser	Val	Gly	Leu	Ser	Ala	Pro	Phe	Val	Ala	Gly	Gln	Met	Asp		
			180					185					190				
tgc	tgc	atg	aac	aac	aca	gct	gtc	ttc	ttg	cca	gtc	ctg	gtt	ggc	ttc	624	
Cys	Cys	Met	Asn	Asn	Thr	Ala	Val	Phe	Leu	Pro	Val	Leu	Val	Gly	Phe		
		195				200						205					
aat	cca	gtg	agc	atg	gcc	aga	aat	gac	ccc	att	gaa	gac	tgg	agt	tca	672	
Asn	Pro	Val	Ser	Met	Ala	Arg	Asn	Asp	Pro	Ile	Glu	Asp	Trp	Ser	Ser		
	210					215					220						
aca	ttc	cga	caa	gta	gca	gag	cgg	atg	cag	aaa	atg	cag	gag	aaa	cag	720	
Thr	Phe	Arg	Gln	Val	Ala	Glu	Arg	Met	Gln	Lys	Met	Gln	Glu	Lys	Gln		
	225				230					235					240		
aaa	gct	ttt	gtg	ctc	aat	cct	gcc	atc	ggg	ccc	gag	ggt	ctc	agc	ggc	768	
Lys	Ala	Phe	Val	Leu	Asn	Pro	Ala	Ile	Gly	Pro	Glu	Gly	Leu	Ser	Gly		
				245					250					255			
tcc	tcc	cgg	atg	aaa	ggt	gga	agt	gcc	acc	aag	att	ctg	ctg	gaa	acc	816	
Ser	Ser	Arg	Met	Lys	Gly	Gly	Ser	Ala	Thr	Lys	Ile	Leu	Leu	Glu	Thr		
			260					265					270				
ctg	tta	tta	gca	gcc	cat	aag	act	gtg	gac	cag	ggc	att	gca	gca	tct	864	
Leu	Leu	Leu	Ala	Ala	His	Lys	Thr	Val	Asp	Gln	Gly	Ile	Ala	Ala	Ser		
		275					280					285					
caa	aga	tgc	ctc	ctg	gaa	atc	ttg	cgg	aca	ttt	gag	cga	gct	cat	cag	912	
Gln	Arg	Cys	Leu	Leu	Glu	Ile	Leu	Arg	Thr	Phe	Glu	Arg	Ala	His	Gln		
	290					295					300						
gtg	acc	tac	agc	caa	agc	ccc	aag	att	gcc	acc	ctg	atg	aag	agt	gtc	960	
Val	Thr	Tyr	Ser	Gln	Ser	Pro	Lys	Ile	Ala	Thr	Leu	Met	Lys	Ser	Val		
	305				310				315						320		
agc	acc	agt	ctg	gag	aag	aaa	ggc	cac	gtg	tac	ctg	gtt	ggc	tgg	cag	1008	
Ser	Thr	Ser	Leu	Glu	Lys	Lys	Gly	His	Val	Tyr	Leu	Val	Gly	Trp	Gln		
				325					330					335			
acc	ctg	ggt	atc	att	gcc	atc	atg	gat	gga	gta	gag	tgc	atc	cac	acc	1056	
Thr	Leu	Gly	Ile	Ile	Ala	Ile	Met	Asp	Gly	Val	Glu	Cys	Ile	His	Thr		
			340					345					350				
ttt	ggt	gct	gat	ttc	cga	gat	gtc	cgt	ggc	ttt	ctc	att	ggt	gat	cac	1104	
Phe	Gly	Ala	Asp	Phe	Arg	Asp	Val	Arg	Gly	Phe	Leu	Ile	Gly	Asp	His		
		355					360					365					
agt	gac	atg	ttt	aac	cag	aag	gct	gag	ctc	acc	aac	cag	ggt	ccc	cag	1152	
Ser	Asp	Met	Phe	Asn	Gln	Lys	Ala	Glu	Leu	Thr	Asn	Gln	Gly	Pro	Gln		
	370					375					380						
ttc	acc	ttc	tcc	cag	gag	gac	ttc	ccg	act	tcc	atc	ctt	ccc	tct	ctc	1200	
Phe	Thr	Phe	Ser	Gln	Glu	Asp	Phe	Pro	Thr	Ser	Ile	Leu	Pro	Ser	Leu		
				390						395					400		
acg	gaa	atc	gat	act	gtg	gtc	ttc	att	ttc	acc	ctg	gat	gac	aac	ctc	1248	
Thr	Glu	Ile	Asp	Thr	Val	Val	Phe	Ile	Phe	Thr	Leu	Asp	Asp	Asn	Leu		
				405				410						415			
acg	gag	gtg	cag	act	ata	gtg	gag	cag	gtg	aaa	gag	aag	acc	aac	cac	1296	
Thr	Glu	Val	Gln	Thr	Ile	Val	Glu	Gln	Val	Lys	Glu	Lys	Thr	Asn	His		
			420				425						430				
atc	cag	gcc	ctg	gca	cac	agc	acc	gtg	ggt	cag	acc	ttg	ccg	atc	cct	1344	
Ile	Gln	Ala	Leu	Ala	His	Ser	Thr	Val	Gly	Gln	Thr	Leu	Pro	Ile	Pro		
		435					440					445					
ctg	aag	aag	ctc	ttt	ccc	tcc	atc	atc	agc	atc	aca	tgg	cca	ctg	ctt	1392	
Leu	Lys	Lys	Leu	Phe	Pro	Ser	Ile	Ile	Ser	Ile	Thr	Trp	Pro	Leu	Leu		

ES 2 646 388 A1

450	455	460	
ttc ttt gaa tat gaa ggg aac ttc atc cag aag ttc cag cgt gag cta			1440
Phe Phe Glu Tyr Glu Gly Asn Phe Ile Gln Lys Phe Gln Arg Glu Leu			480
465	470	475	
agc acc aaa tgg gtg ctg aat aca gtg agt aca ggt gct cat gtg ctt			1488
Ser Thr Lys Trp Val Leu Asn Thr Val Ser Thr Gly Ala His Val Leu			495
485	490		
ctt ggt aag atc cta caa aac cac atg ttg gac ctt cgg att agc aac			1536
Leu Gly Lys Ile Leu Gln Asn His Met Leu Asp Leu Arg Ile Ser Asn			510
500	505	510	
tcc aag ctc ttc tgg cgg gcg ctg gcc atg ctg cag cgg ttc tct gga			1584
Ser Lys Leu Phe Trp Arg Ala Leu Ala Met Leu Gln Arg Phe Ser Gly			520
515	520	525	
cag tcc aag gct cga tgc atc gag agc ctc ctc cga gcg atc cac ttt			1632
Gln Ser Lys Ala Arg Cys Ile Glu Ser Leu Leu Arg Ala Ile His Phe			530
530	535	540	
ccc cag cca ctg tca gat gat att cgg gct gct ccc atc tcc tgc cgt			1680
Pro Gln Pro Leu Ser Asp Asp Ile Arg Ala Ala Pro Ile Ser Cys Arg			550
545	550	555	
gtc cag gtt gca cat gag aag gaa cag gtg ata ccc atc gcc ttg ctg			1728
Val Gln Val Ala His Glu Lys Glu Gln Val Ile Pro Ile Ala Leu Leu			565
565	570	575	
agc ctc cta ttc cgg tgc tcg atc act gag gct cag gca cac ctg gct			1776
Ser Leu Leu Phe Arg Cys Ser Ile Thr Glu Ala Gln Ala His Leu Ala			580
580	585	590	
gca gct cct tct gtc tgt gag gct gtc agg agt gct ctt gct ggg cca			1824
Ala Ala Pro Ser Val Cys Glu Ala Val Arg Ser Ala Leu Ala Gly Pro			595
595	600	605	
ggt cag aag cgc act gcg gac ccc ctc gag atc cta gag cct gac gtt			1872
Gly Gln Lys Arg Thr Ala Asp Pro Leu Glu Ile Leu Glu Pro Asp Val			610
610	615	620	
cag tga			1878
Gln			625
625			

<210> 2
 <211> 625
 <212> PRT
 <213> Homo sapiens

<400> 2

Met	Pro	Gly	Thr	Lys	Arg	Phe	Gln	His	Val	Ile	Glu	Thr	Pro	Glu	Pro
1				5					10					15	
Gly	Lys	Trp	Glu	Leu	Ser	Gly	Tyr	Glu	Ala	Ala	Val	Pro	Ile	Thr	Glu
			20					25					30		
Lys	Ser	Asn	Pro	Leu	Thr	Gln	Asp	Leu	Asp	Lys	Ala	Asp	Ala	Glu	Asn
		35					40					45			
Ile	Val	Arg	Leu	Leu	Gly	Gln	Cys	Asp	Ala	Glu	Ile	Phe	Gln	Glu	Glu
	50					55					60				

ES 2 646 388 A1

Gly 65 Gln Ala Leu Ser Thr 70 Tyr Gln Arg Leu Tyr 75 Ser Glu Ser Ile Leu 80
 Thr Thr Met Val Gln 85 Val Ala Gly Lys Val 90 Gln Glu Val Leu Lys 95 Glu
 Pro Asp Gly Gly 100 Leu Val Val Leu Ser 105 Gly Gly Gly Thr Ser Gly Arg 110
 Met Ala Phe 115 Leu Met Ser Val Ser 120 Phe Asn Gln Leu Met 125 Lys Gly Leu
 Gly Gln Lys Pro Leu Tyr Thr 135 Tyr Leu Ile Ala Gly 140 Gly Asp Arg Ser
 Val 145 Val Ala Ser Arg Glu 150 Gly Thr Glu Asp Ser 155 Ala Leu His Gly Ile 160
 Glu Glu Leu Lys Lys 165 Val Ala Ala Gly Lys 170 Lys Arg Val Ile Val 175 Ile
 Gly Ile Ser Val 180 Gly Leu Ser Ala Pro 185 Phe Val Ala Gly Gln 190 Met Asp
 Cys Cys Met 195 Asn Asn Thr Ala Val 200 Phe Leu Pro Val Leu 205 Val Gly Phe
 Asn Pro 210 Val Ser Met Ala Arg 215 Asn Asp Pro Ile Glu 220 Asp Trp Ser Ser
 Thr Phe Arg Gln Val Ala 230 Glu Arg Met Gln Lys 235 Met Gln Glu Lys Gln 240
 Lys Ala Phe Val Leu 245 Asn Pro Ala Ile Gly 250 Pro Glu Gly Leu Ser Gly 255
 Ser Ser Arg Met 260 Lys Gly Gly Ser Ala 265 Thr Lys Ile Leu Leu 270 Glu Thr
 Leu Leu Leu 275 Ala Ala His Lys Thr Val Asp Gln Gly Ile 285 Ala Ala Ser
 Gln Arg 290 Cys Leu Leu Glu Ile 295 Leu Arg Thr Phe Glu 300 Arg Ala His Gln
 Val Thr Tyr Ser Gln Ser 310 Pro Lys Ile Ala Thr 315 Leu Met Lys Ser Val 320
 Ser Thr Ser Leu Glu 325 Lys Lys Gly His Val 330 Tyr Leu Val Gly Trp 335 Gln

ES 2 646 388 A1

Thr Leu Gly Ile Ile Ala Ile Met Asp Gly Val Glu Cys Ile His Thr
 340 345 350

Phe Gly Ala Asp Phe Arg Asp Val Arg Gly Phe Leu Ile Gly Asp His
 355 360 365

Ser Asp Met Phe Asn Gln Lys Ala Glu Leu Thr Asn Gln Gly Pro Gln
 370 375 380

Phe Thr Phe Ser Gln Glu Asp Phe Pro Thr Ser Ile Leu Pro Ser Leu
 385 390 395 400

Thr Glu Ile Asp Thr Val Val Phe Ile Phe Thr Leu Asp Asp Asn Leu
 405 410 415

Thr Glu Val Gln Thr Ile Val Glu Gln Val Lys Glu Lys Thr Asn His
 420 425 430

Ile Gln Ala Leu Ala His Ser Thr Val Gly Gln Thr Leu Pro Ile Pro
 435 440 445

Leu Lys Lys Leu Phe Pro Ser Ile Ile Ser Ile Thr Trp Pro Leu Leu
 450 455 460

Phe Phe Glu Tyr Glu Gly Asn Phe Ile Gln Lys Phe Gln Arg Glu Leu
 465 470 475 480

Ser Thr Lys Trp Val Leu Asn Thr Val Ser Thr Gly Ala His Val Leu
 485 490 495

Leu Gly Lys Ile Leu Gln Asn His Met Leu Asp Leu Arg Ile Ser Asn
 500 505 510

Ser Lys Leu Phe Trp Arg Ala Leu Ala Met Leu Gln Arg Phe Ser Gly
 515 520 525

Gln Ser Lys Ala Arg Cys Ile Glu Ser Leu Leu Arg Ala Ile His Phe
 530 535 540

Pro Gln Pro Leu Ser Asp Asp Ile Arg Ala Ala Pro Ile Ser Cys Arg
 545 550 555 560

Val Gln Val Ala His Glu Lys Glu Gln Val Ile Pro Ile Ala Leu Leu
 565 570 575

Ser Leu Leu Phe Arg Cys Ser Ile Thr Glu Ala Gln Ala His Leu Ala
 580 585 590

Ala Ala Pro Ser Val Cys Glu Ala Val Arg Ser Ala Leu Ala Gly Pro
 595 600 605

Gly Gln Lys Arg Thr Ala Asp Pro Leu Glu Ile Leu Glu Pro Asp Val

ES 2 646 388 A1

610

615

620

Gln
625

<210> 3
<211> 7304
<212> DNA
<213> Artificial Sequence

<220>
<223> plásmido comercial pGBKT7

<400> 3
 cggtgcgggc ctcttcgcta ttacgccaga tccttttggt gtttccgggt gtacaatatg 60
 gacttcctct tttctggcaa ccaaaccat acatcgggat tcctataata ccttcgttgg 120
 tctccctaac atgtaggtgg cggaggggag atatacaata gaacagatac cagacaagac 180
 ataatgggct aaacaagact acaccaatta cactgcctca ttgatggtgg tacataacga 240
 actaatactg tagccctaga cttgatagcc atcatcatat cgaagtttca ctaccctttt 300
 tccatttgcc atctattgaa gtaataatag gcgcatgcaa cttcttttct ttttttttct 360
 tttctctctc ccccgttggt gtctcaccat atccgcaatg acaaaaaaaaa tgatggaaga 420
 cactaaagga aaaaattaac gacaaagaca gcaccaacag atgtcgtttgt tccagagctg 480
 atgaggggta tctcgaagca cacgaaactt tttccttctc tcattcacgc acactactct 540
 ctaatgagca acggtatacg gccttccttc cagttacttg aatttgaaat aaaaaaagtt 600
 tgctgtcttg ctatcaagta taaatagacc tgcaattatt aatcttttgt ttcctcgtca 660
 ttgttctcgt tccctttctt ccttgtttct ttttctgcac aatatttcaa gctataccaa 720
 gcatacaatc aactccaagc ttgaagcaag cctcctgaaa gatgaagcta ctgtcttcta 780
 tcgaacaagc atgcatatt tgccgactta aaaagctcaa gtgctccaaa gaaaaaccga 840
 agtgcgcaa gtgtctgaag aacaactggg agtgtcgcta ctctcccaa accaaaaggt 900
 ctccgctgac tagggcacat ctgacagaag tggaatcaag gctagaaaga ctggaacagc 960
 tatttctact gatttttctc cgagaagacc ttgacatgat tttgaaaatg gattctttac 1020
 aggatataaa agcattgtta acaggattat ttgtacaaga taatgtgaat aaagatgccg 1080
 tcacagatag attggcttca gtggagactg atatgcctct aacattgaga cagcatagaa 1140
 taagtgcgac atcatcatcg gaagagagta gtaacaaagg tcaaagacag ttgactgtat 1200
 cgccggaatt tgtaatacga ctactatag ggcgagccgc catcatggag gagcagaagc 1260
 tgatctcaga ggaggacctg catatggcca tggaggccga attcccgggg atccgctgac 1320
 ctgcagcggc cgcataacta gcataacccc ttggggcctc taaacggggtc ttgagggggt 1380
 ttttgcgcgc ttgcagcaa gctaattccg ggcgaatttc ttatgattta tgatttttat 1440
 tattaataaa gttataaaaa aaataagtgt atacaaattt taaagtgact cttaggtttt 1500
 aaaacgaaaa ttcttattct tgagtaactc tttcctgtag gtcaggttgc tttctcaggt 1560
 atagcatgag gtcgctctta ttgaccacac ctctaccggc atgcaagctt ggcgtaatca 1620

ES 2 646 388 A1

tggtcatagc	tgtttctgt	gtgaaattgt	tatccgctca	caattccaca	caacatacga	1680
gccggaagca	taaagtgtaa	agcctggggg	gcctaatagag	tgaggtaact	cacattaatt	1740
gcgttgcgct	cactgcccgc	tttccagtcg	ggaaacctgt	cgtgccagct	gcattaatga	1800
atcggccaac	gcgcggggag	aggcggtttg	cgtattgggc	gctcttccgc	ttcctcgctc	1860
actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	tatcagctca	ctcaaaggcg	1920
gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	agaacatgtg	agcaaaaggc	1980
cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	2040
ccccctgacg	agcatcacia	aaatcgacgc	tcaagtcaga	ggtaggcgaaa	cccgcagga	2100
ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	tgttccgacc	2160
ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	2220
agctcacgct	gtaggtatct	cagttcgggtg	taggtcgttc	gctccaagct	gggctgtgtg	2280
cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	2340
aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	2400
gcgaggtatg	taggcgggtc	tacagagttc	ttgaagtggg	ggcctaacta	cggtacact	2460
agaaggacag	tatttgggat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	2520
ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtgggtttttt	tgtttgcaag	2580
cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	2640
tctgacgctc	agtggaacga	aaactcacgt	taagggatth	tggtcatgag	attatcaaaa	2700
aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaaataaat	ctaaagtata	2760
tatgagtaac	ctgaggctat	ggcagggcct	gccgccccga	cgttggctgc	gagccctggg	2820
ccttcacccg	aacttggggg	gtgggggtggg	gaaaaggaag	aaacgcgggc	gtattggccc	2880
caatgggggtc	tcgggtgggt	atcgacagag	tgccagccct	gggaccgaac	cccgcgttta	2940
tgaacaaacg	acccaacacc	gtgcgtttta	ttctgtcttt	ttattgccgt	catagcgcgg	3000
gttcttccg	gtattgtctc	cttccgtggt	tcagttagcc	tccccctagg	gtgggcgaag	3060
aactccagca	tgagatcccc	gcgctggagg	atcatccagc	cggcgtcccc	gaaaacgatt	3120
ccgaagccca	acctttcata	gaaggcggcg	gtggaatcga	aatctcgtga	tggcaggttg	3180
ggcgtcgctt	ggtcggttca	tttcgaacc	cagagtcccg	ctcagaagaa	ctcgtcaaga	3240
aggcgataga	aggcgatgcy	ctgcgaatcg	ggagcggcga	taccgtaaag	cacgaggaag	3300
cggtcagccc	attcgccgcc	aagctcttca	gcaatatcac	gggtagccaa	cgctatgtcc	3360
tgatagcggg	ccgccacacc	cagccggcca	cagtcgatga	atccagaaaa	gcggccattt	3420
tccaccatga	tattcggcaa	gcaggcatcg	tcatgggtca	cgacgagatc	ctcgcgctcg	3480
ggcatgctcg	ccttgagcct	ggcgaacagt	tcggctggcg	cgagcccctg	atgctcttcg	3540
tccagatcat	cctgatcgac	aagaccggct	tccatccgag	tacgtgctcg	ctcgatgcga	3600
tgtttcgctt	ggtaggtcga	tgggcaggta	gccggatcaa	gcgtatgcag	ccgccgcatt	3660

ES 2 646 388 A1

gcatcagcca tgatggatac tttctcggca ggagcaaggt gagatgacag gagatcctgc 3720
cccggcactt cgccaatag cagccagtcc cttcccgctt cagtgacaac gtcgagcaca 3780
gctgcgcaag gaacgcccgt cgtggccagc cacgatagcc gcgctgcctc gtcttgagct 3840
tcattcaggg caccggacag gtcggtcttg acaaaaagaa cggggcgccc ctgctgctgac 3900
agccggaaca cggcggcatc agagcagccg attgtctggt gtgcccagtc atagccgaat 3960
agcctctcca cccaagcggc cggagaacct gcgtgcaatc catcttgttc aatcatactc 4020
ttcctttttc aatattattg aagcatttat cagggttatt gtctcatgag cggatacata 4080
tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc cggaaaagtg 4140
ccacctgaac gaagcatctg tgcttcattt tgtagaacia aaatgcaacg cgagagcgct 4200
aatTTTTcaa acaagaatc tgagctgcat ttttacagaa cagaaatgca acgCGaaagc 4260
gctattttac caacgaagaa tctgtgcttc atttttgtaa acaaaaaatg caacgcgaga 4320
gCGctaattt ttcaaacaaa gaatctgagc tgcattttta cagaacagaa atgcaacgcg 4380
agagcgctat tttaccaaca aagaatctat acttcttttt tgttctacia aaatgcatcc 4440
cgagagcgct atttttctaa caaagcatct tagattactt tttttctcct ttgtgcgctc 4500
tataatgcag tctcttgata actttttgca ctgtaggtcc gtttaaggta gaagaaggct 4560
actttgggtg ctattttctc ttccataaaa aaagcctgac tccacttccc gcgtttactg 4620
attactagcg aagctgcggg tgcatttttt caagataaag gcatccccga ttatattcta 4680
taccgatgtg gattgcgcat actttgtgaa cagaaagtga tagcgttgat gattcttcat 4740
tggtcagaaa attatgaacg gtttcttcta ttttgtctct atatactacg tataggaaat 4800
gtttacattt tcgtattggt ttcgattcac tctatgaata gttcttacta caattttttt 4860
gtctaaagag taatactaga gataaacata aaaaatgtag aggtcgagtt tagatgcaag 4920
ttcaaggagc gaaagggtga tgggtagggt atatagggat atagcacaga gatatatagc 4980
aaagagatac ttttgagcaa tgtttgtgga agcggatttc gcaatatttt agtagctcgt 5040
tacagtccgg tgcgTTTTg gTTTTTgaa agtgcgtctt cagagcgctt ttggttttca 5100
aaagcgctct gaagttccta tactttctag agaataggaa cttcggaata ggaacttcaa 5160
agcgtttccg aaaacgagcg cttccgaaaa tgcaacgcga gctgCGcaca tacagctcac 5220
tgttcacgtc gcacctatat ctgCGtggtg cctgtatata tatatacatg agaagaacgg 5280
catagtgcgt gtttatgctt aaatgcgtac ttatatgcgt ctatttatgt aggatgaaag 5340
gtagtctagt acctcctgtg atattatccc attccatgcg gggatcgtg tgcttccttc 5400
agcactacc tttagctggt ctatatgctg cactcctca attggattag tctcatcctt 5460
caatgctatc atttccttg atattggatc atattaagaa accattatta tcatgacatt 5520
aacctataaa aataggcgta tcacgaggcc ctttcgtctc gcgCGtttcg gtgatgacgg 5580
tgaaaacctc tgacacatgc agctccccga gacggtcaca gcttgtctgt aagcggatgc 5640
cgggagcaga caagcccgtc agggcgctc agcgggtggt ggcgggtgct ggggctggct 5700
taactatgcg gcatcagagc agattgtact gagagtgcac catagatcaa cgacattact 5760

ES 2 646 388 A1

atatatataa tataggaagc atttaataga acagcatcgt aatatatgtg tacttttgcag 5820
 ttatgacgcc agatggcagt agtggaagat attctttatt gaaaaatagc ttgtcacctt 5880
 acgtacaatc ttgatccgga gcttttcttt ttttgccgat taagaattaa ttcggtcgaa 5940
 aaaagaaaag gagagggcca agagggaggg cattggtgac tattgagcac gtgagtatac 6000
 gtgattaagc acacaaaggc agcttgagat atgtctgtta ttaatttcac aggtagtctt 6060
 ggtccattgg tgaaagtttg cggcttgcag agcacagagg cgcagaatg tgctctagat 6120
 tccgatgctg acttgctggg tattatatgt gtgcccaata gaaagagaac aattgacccg 6180
 gttattgcaa ggaaaatttc aagtcttgta aaagcatata aaaatagttc aggcactccg 6240
 aaatacttgg ttggcgtggt tcgtaatcaa cctaaggagg atgttttggc tctggtcaat 6300
 gattacggca ttgatatcgt ccaactgcat ggagatgagt cgtggcaaga ataccaagag 6360
 ttctcgggtt tgccagttat taaaagactc gtatttccaa aagactgcaa catactactc 6420
 agtgcagctt cacagaaacc tcattcgttt attcccttgt ttgattcaga agcaggtggg 6480
 acaggtgaac ttttggattg gaactcgatt tctgactggg ttggaaggca agagagcccc 6540
 gaaagcttac attttatggt agctggtgga ctgacgccag aaaatgttgg tgatgcgctt 6600
 agattaaatg gcgttattgg tgttgatgta agcggagggtg tggagacaaa tgggtgtaaaa 6660
 gactctaaca aatagcaaa tttcgtcaaa aatgctaaga aataggttat tactgagtag 6720
 tatttattta agtattgttt gtgcacttgc cgatctatgc ggtgtgaaat accgcacaga 6780
 tgcgtaagga gaaaataccg catcaggaaa ttgtaaactg taatattttg ttaaaattcg 6840
 cgtaaattt ttgttaaact agctcatttt ttaaccaata ggccgaaatc ggcaaaatcc 6900
 cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt tggaacaaga 6960
 gtccactatt aaagaactg gactccaacg tcaaagggcg aaaaaccgtc tatcagggcg 7020
 atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg tgccgtaaag 7080
 cactaaatcg gaaccctaaa gggagcccc gatttagagc ttgacgggga aagccggcga 7140
 acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg ctggcaagtg 7200
 tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg ctacagggcg 7260
 cgtccattcg ccattcaggc tgcgcaactg ttgggaaggg cgat 7304

<210> 4
 <211> 440
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Dominio de unión a Gal4 (GBD-Gal4 Binding Domain)

<400> 4
 atgaagctac tgtcttctat cgaacaagca tgcgatattt gccgacttaa aaagccaagt 60
 gtcctaaaga aaaaccgaag tgcgccaagt gtctgaagaa caactgggag tgctcgctact 120
 ctcccaaac caaaaggctt ccgctgacta gggcacatct gacagaagtg gaatcaaggc 180

ES 2 646 388 A1

tagaaagact ggaacagcta tttctactga tttttcctcg agaagacctt gacatgattt 240
 tgaaaatgga ttctttacag gatataaaag cattgttaac aggattattt gtacaagata 300
 atgtgaataa agatgccgtc acagatagat tggcttcagt ggagactgat atgcctctaa 360
 cattgagaca gcatagaata agtgcgacat catcatcgga agagagtagt aacaaaggtc 420
 aaagacagtt gactgtatcg 440

<210> 5
 <211> 705
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Promotor ADH1

<400> 5
 atccttttgt tgtttccggg tgtacaatat ggacttcctc ttttctggca accaaaccca 60
 tacatcggga ttctataat accttcgttg gtctccctaa catgtaggtg gcggagggga 120
 gatatacaat agaacagata ccagacaaga cataatgggc taaacaagac tacaccaatt 180
 aactgcctc attgatggtg gtacataacg aactaatact gtagccctag acttgatagc 240
 catcatcata tcgaagtttc actacccttt ttccatttgc catctattga agtaataata 300
 ggcgcatgca acttcttttc ttttttttc ttttctctct cccccgttgt tgtctcacca 360
 tatccgcaat gacaaaaaaaa atgatggaag acactaaagg aaaaaattaa cgacaaagac 420
 agcaccaaca gatgtcgttg ttccagagct gatgaggggt atctcgaagc acacgaaact 480
 ttttcttcc ttcatcagc cacactactc tctaatagagc aacgggtatac ggccttcctt 540
 ccagttactt gaatttgaaa taaaaaaagt ttgctgtctt gctatcaagt ataaatagac 600
 ctgcaattat taatcttttg tttcctcgtc attgttctcg ttccctttct tccttgtttc 660
 ttttctgca caatatttca agctatacca agcatacaat caact 705

<210> 6
 <211> 9177
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> plásmido pGBKT7-GKRP

<400> 6
 cggtgcgggc ctcttcgcta ttacgccaga tccttttggt gtttccgggt gtacaatatg 60
 gacttcctc tttctggcaa ccaaaccat acatcgggat tcctataata ctttcgttgg 120
 tctccctaac atgtaggtgg cggagggggag atatacaata gaacagatac cagacaagac 180
 ataatgggct aaacaagact acaccaatta cactgcctca ttgatggtgg tacataacga 240
 actaatactg tagccctaga cttgatagcc atcatcatat cgaagtttca ctaccctttt 300
 tccatttgcc atctattgaa gtaataatag gcgcatgcaa cttcttttct tttttttct 360
 tttctctctc cccccgttgt gtctcacat atccgcaatg aaaaaaaaaa tgatggaaga 420
 cactaaagga aaaaattaac gacaaagaca gcaccaacag atgtcgttgt tccagagctg 480

ES 2 646 388 A1

atgaggggta tctcgaagca cacgaaactt tttccttctt tcattcacgc aactactctt 540
ctaattgagca acgggtatacg gccttccttc cagttacttg aatttgaaat aaaaaaagtt 600
tgctgtcctg ctatcaagta taaatagacc tgcaattatt aatcttttgt ttcctcgtca 660
ttgttctcgt tccctttctt ccttgtttct ttttctgcac aatatttcaa gctataccaa 720
gcatacaatc aactccaagc ttgaagcaag cctcctgaaa gatgaagcta ctgtcttcta 780
tcgaacaagc atgcatgatt tgccgactta aaaagctcaa gtgctccaaa gaaaaaccga 840
agtgcgccaa gtgtctgaag aacaactggg agtgtcgcta ctctcccaa accaaaaggt 900
ctccgctgac tagggcacat ctgacagaag tggaatcaag gctagaaaga ctggaacagc 960
tatttctact gatttttctt cgagaagacc ttgacatgat tttgaaaatg gattctttac 1020
aggatataaa agcattgtta acaggattat ttgtacaaga taatgtgaat aaagatgccg 1080
tcacagatag attggcttca gtggagactg atatgcctct aacattgaga cagcatagaa 1140
taagtgcgac atcatcatcg gaagagagta gtaacaaagg tcaaagacag ttgactgtat 1200
cgccggaatt tgtaatacga ctactatag ggcgagccgc catcatggag gagcagaagc 1260
tgatctcaga ggaggacctg catatggcca tggaggccga attcatgcca ggcacaaaac 1320
ggtttcaaca tgcattgag accccggagc ctggcaagtg ggagtgtctt gggtagcagg 1380
cagctgtgcc aatcacggag aagtcaaacc cactgacca ggatctagac aaagcagatg 1440
ctgagaacat tgttcgactg ctagggcaat gtgatgctga gatcttccag gaggaggggc 1500
aagccctgtc cacataccag agactctaca gcgaatccat tctgaccacc atggtacagg 1560
tggctgggaa agttcaggaa gtgctgaagg agccagatgg ggggctgggt gtgctgagtg 1620
gagggggcac ctctggccgg atggcattcc tcatgtcggg gtcctttaat cagctgatga 1680
aaggtctggg acagaaacct ctttacacct acctcattgc aggtgggtgac aggtctgtgg 1740
tggcctctag ggaggggaca gaagatagtg ccttgcacgg gattgaggaa ctgaagaagg 1800
tggctgccgg gaagaagaga gtgattgtca ttggcatttc tgtgggactc tctgctcctt 1860
ttgtggcagg ccagatggac tgctgcatga acaacacagc tgtcttcttg ccagtcctgg 1920
ttggcttcaa tccagtgagc atggccagaa atgaccccat tgaagactgg agttcaacat 1980
tccgacaagt agcagagcgg atgcagaaa tgcaggagaa acagaaagct tttgtgctca 2040
atcctgccat cgggcccag ggtctcagcg gtcctcccg gatgaaaggt ggaagtgcc 2100
ccaagattct gctggaaacc ctgttattag cagcccataa gactgtggac cagggcattg 2160
cagcatctca aagatgcctc ctggaaatct tgcggacatt tgagcgagct catcaggtga 2220
cctacagcca aagcccaag attgccacc tgatgaagag tgtcagcacc agtctggaga 2280
agaaaggcca cgtgtacctg gttggctggc agaccctggg tatcattgcc atcatggatg 2340
gagtagagtg catccacacc tttgggtgctg atttccgaga tgtccgtggc tttctcattg 2400
gtgatcacag tgacatgttt aaccagaagg ctgagctcac caaccagggt cccagttca 2460
ccttctcca ggaggacttc ccgacttcca tccttccctc tctcacggaa atcgatactg 2520

ES 2 646 388 A1

tggtcttcat	ttcacctg	gatgacaacc	tcacggaggt	gcagactata	gtggagcagg	2580
tgaagagaa	gaccaaccac	atccaggccc	tggcacacag	caccgtgggt	cagaccttgc	2640
cgatccctct	gaagaagctc	tttcctcca	tcatcagcat	cacatggcca	ctgcttttct	2700
ttgaatatga	aggaacttc	atccagaagt	tccagcgtga	gctaagcacc	aaatgggtgc	2760
tgaatacagt	gagtacaggt	gctcatgtgc	ttcttggtaa	gacacctaaa	aaccacatgt	2820
tggaccttcg	gattagcaac	tccaagctct	tctggcgggc	gctggccatg	ctgcagcggg	2880
tctctggaca	gtccaaggct	cgatgcatcg	agagcctcct	ccgagcgatc	cactttcccc	2940
agccactgtc	agatgatatt	cgggctgctc	ccatctcctg	ccgtgtccag	gttgcacatg	3000
agaaggaaca	ggtgataccc	atcgccttgc	tgagcctcct	attccgggtgc	tcgatcactg	3060
aggctcaggc	acacctggct	gcagctcctt	ctgtctgtga	ggctgtcagg	agtgtctttg	3120
ctgggccagg	tcagaagcgc	actgcggacc	ccctcgagat	cctagagcct	gacgttcagt	3180
aggatccgtc	gacctgcagc	ggccgcataa	ctagcataac	cccttggggc	ctctaaacgg	3240
gtcttgaggg	gttttttgcg	cgcttgagc	caagctaatt	ccgggcgaat	ttcttatgat	3300
ttatgatttt	tattattaa	taagttataa	aaaaaataag	tgtatacaaa	ttttaagtgc	3360
actcttaggt	tttaaacga	aaattcttat	tcttgagtaa	ctctttcctg	taggtcaggt	3420
tgctttctca	ggtatagcat	gaggctgctc	ttattgacca	cacctctacc	ggcatgcaag	3480
cttggcgtaa	tcatggcat	agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	3540
acacaacata	cgagccggaa	gcataaagtgc	taaagcctgg	ggtgcctaata	gagtgaggta	3600
actcacatta	attgcgttgc	gctcactgcc	cgctttccag	tcgggaaacc	tgctgtgcca	3660
gctgcattaa	tgaatcggcc	aacgcgcggg	gagaggcggg	ttgctgattg	ggcgtcttcc	3720
cgcttcctcg	ctcactgact	cgctgcgctc	ggctgctcgg	ctgcggcgag	cggtatcagc	3780
tcactcaaag	gcggtataac	ggttatccac	agaatcaggg	gataacgcag	gaaagaacat	3840
gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgttgc	tggcgttttt	3900
ccataggctc	cgccccctg	acgagcatca	caaaaatcga	cgctcaagtc	agaggtggcg	3960
aaaccgaca	ggactataaa	gataccaggc	gtttccccct	ggaagctccc	tcgtgcgctc	4020
tcctgttccg	accctgccgc	ttaccggata	cctgtccgcc	tttctccctt	cggaagcgt	4080
ggcgttttct	catagctcac	gctgtaggta	tctcagttcg	gtgtaggctg	ttcgtccaa	4140
gctgggctgt	gtgcacgaac	ccccgttca	gcccgaccgc	tgcgcttat	ccgtaacta	4200
tcgtcttgag	tccaaccgg	taagacacga	cttatcgcca	ctggcagcag	ccactggtaa	4260
caggattagc	agagcgaggt	atgtaggcgg	tgctacagag	ttcttgaagt	ggtggcctaa	4320
ctacggctac	actagaagga	cagtatttgg	tatctgcgct	ctgctgaagc	cagttacctt	4380
cgaaaaaga	gttggtagct	cttgatccgg	caaaaaacc	accgctggta	gcggtggttt	4440
ttttgttgc	aagcagcaga	ttacgcgcag	aaaaaaagga	tctcaagaag	atcctttgat	4500
cttttctacg	gggtctgacg	ctcagtgga	cgaaaactca	cgtaaggga	ttttggtcat	4560
gagattatca	aaaaggatct	tcacctagat	ccttttaaat	taaaaatgaa	gttttaaatc	4620

ES 2 646 388 A1

aatctaaagt atatatgagt aacctgaggc tatggcaggg cctgccgccc cgacgttggc 4680
tgcgagccct gggccttcac ccgaacttgg ggggtgggggt ggggaaaagg aagaaacgcg 4740
ggcgtattgg cccaatggg gtctcgggtg ggtatcgaca gagtgccagc cctgggaccg 4800
aaccgcgct ttatgaaca acgaccaac accgtgctt ttattctgtc tttttattgc 4860
cgtcatagcg cgggttcctt ccggtattgt ctcttccgt gtttcagtta gcctccccct 4920
aggggtgggcg aagaactcca gcatgagatc cccgctgctg aggatcatcc agccggcgctc 4980
ccggaaaacg attccgaagc ccaacctttc atagaaggcg gcggtggaat cgaaatctcg 5040
tgatggcagg ttgggcgtcg cttggctcgg tcatctcgaa cccagagtc ccgctcagaa 5100
gaactcgtca agaaggcgat agaaggcgat gcgctgcgaa tcgggagcgg cgataccgta 5160
aagcacgagg aagcggctag cccttcgcc gccaaactct tcagcaatat cacgggtagc 5220
caacgctatg tcctgatagc ggtccgccac acccagccgg ccacagtcga tgaatccaga 5280
aaagcggcca tttccacca tgatattcgg caagcaggca tcgtcatggg tcacgacgag 5340
atcctcgccg tcgggcatgc tcgccttgag cctggcgaac agttcggctg gcgagagccc 5400
ctgatgctct tcgtccagat catcctgatc gacaagaccg gcttccatcc gagtacgtgc 5460
tcgctcgatg cgatgtttcg cttgggtggtc gaatgggcag gtagccggat caagcgtatg 5520
cagccgccgc attgcatcag ccatgatgga tactttctcg gcaggagcaa ggtgagatga 5580
caggagatcc tgccccggca cttcgcccaa tagcagccag tcccttcccg cttcagtgac 5640
aacgtcgagc acagctgcgc aaggaacgcc cgtcgtggcc agccacgata gccgcgctgc 5700
ctcgtcttgc agttcattca gggcaccgga caggtcggctc ttgacaaaaa gaaccgggcg 5760
cccctgcgct gacagccgga acacggcggc atcagagcag ccgattgtct gttgtgccca 5820
gtcatagccg aatagcctct ccaccaagc ggccggagaa cctgcgtgca atccatcttg 5880
ttcaatcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt attgtctcat 5940
gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt 6000
tccccgaaaa gtgccacctg aacgaagcat ctgtgcttca tttttagtaa caaaaatgca 6060
acgcgagagc gctaattttt caaacaaga atctgagctg cttttttaca gaacagaaat 6120
gcaacgcgaa agcgtatatt taccaacgaa gaatctgtgc ttcatttttg taaaacaaaa 6180
atgcaacgcg agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca 6240
gaaatgcaac gcgagagcgc tttttacca acaagaatc tatacttctt ttttgttcta 6300
caaaaatgca tcccagagagc gctatTTTTT taacaaagca tcttagatta ctttttttct 6360
cctttgtgcg ctctataatg cagtctcttg ataactTTTT gcaactgtagg tccgtaaggg 6420
ttagaagaag gctacttttg tgtctatTTT ctcttcata aaaaaagcct gactccactt 6480
cccgcgttta ctgattacta gcgaagctgc ggggtgcattt tttcaagata aaggcatccc 6540
cgattatatt ctataccgat gtggattgcg catactttgt gaacagaaag tgatagcgtt 6600
gatgattctt cattggctag aaaattatga acggtttctt ctattttgtc tctatatact 6660

ES 2 646 388 A1

acgtatagga aatgtttaca ttttcgtatt gttttcgatt cactctatga atagttctta 6720
 ctacaatfff tttgtctaaa gagtaatact agagataaac ataaaaaatg tagagggtcga 6780
 gtttagatgc aagttcaagg agcgaaggt ggatgggtag gttatatagg gatatagcac 6840
 agagatatat agcaaagaga tacttttgag caatgtttgt ggaagcggta ttcgcaatat 6900
 ttttagtagct cgttacagtc cgggtcgttt ttggtttttt gaaagtgcgt cttcagagcg 6960
 cttttggttt tcaaaagcgc tctgaagttc ctatactttc tagagaatag gaacttcgga 7020
 ataggaactt caaagcgttt ccgaaaacga gcgcttccga aaatgcaacg cgagctgcgc 7080
 acatacagct cactgttcac gtcgcaccta tatctgcgtg ttgcctgtat atatatatac 7140
 atgagaagaa cggcatagtg cgtgtttatg cttaaagtcg tacttatatg cgtctattta 7200
 tgtaggatga aaggtagtct agtacctcct gtgatattat cccattccat gcgggggtatc 7260
 gtatgcttcc ttcagcacta cccttttagct gttctatatg ctgccactcc tcaattggat 7320
 tagtctcatc cttcaatgct atcatttcct ttgatattgg atcatattaa gaaaccatta 7380
 ttatcatgac attaacctat aaaaataggc gtatcacgag gccctttcgt ctcgcgcggt 7440
 tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggagacggtc acagcttgtc 7500
 tgtaagcggg tgccgggagc agacaagccc gtcagggcgc gtcagcgggt gttggcgggt 7560
 gtcggggctg gcttaactat gcggcatcag agcagattgt actgagagtg caccatagat 7620
 caacgacatt actatatata taatatagga agcatttaat agaacagcat cgtaatatat 7680
 gtgtactttg cagttatgac gccagatggc agtagtgga gatattcttt attgaaaaat 7740
 agcttgtcac cttacgtaca atcttgatcc ggagcttttc tttttttgcc gattaagaat 7800
 taattcggtc gaaaaaagaa aaggagaggg ccaagagggg gggcattggt gactattgag 7860
 cacgtgagta tacgtgatta agcacacaaa ggcagcttg agtatgtctg ttattaatft 7920
 cacaggtagt tctggtccat tggtgaaagt ttgctgctg cagagcacag aggccgcaga 7980
 atgtgctcta gattccgatg ctgacttgct gggattata tgtgtgcca atagaaagag 8040
 aacaattgac ccggttattg caaggaaaat ttcaagtctt gtaaaagcat ataaaaatag 8100
 ttcaggcact ccgaaatact tggttggcgt gtttcgtaat caacctaagg aggatgtttt 8160
 ggctctggtc aatgattacg gcattgatat cgtccaactg catggagatg agtcgtggca 8220
 agaataccaa gagttcctcg gtttgccagt tattaaga ctcgtatttc caaaagactg 8280
 caacatacta ctcagtgcag cttcacagaa acctattcg tttattccct tgtttgattc 8340
 agaagcaggt gggacaggtg aacttttgga ttggaactcg atttctgact gggttggaag 8400
 gcaagagagc cccgaaagct tacatftttat gttagctggt ggactgacgc cagaaaatgt 8460
 tggatgatgc cttagattaa atggcgttat tgggtttgat gtaagcggag gtgtggagac 8520
 aaatggtgta aaagactcta acaaaatagc aaatttcgtc aaaaatgcta agaaataggt 8580
 tattactgag tagtattttat ttaagtattg tttgtgcact tgccgatcta tgcgggtgta 8640
 aataccgcac agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgttaatatt 8700
 ttgttaaaat tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa 8760

ES 2 646 388 A1

atcggcaaaa tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca 8820
 gtttgaaca agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc 8880
 gtctatcagg gcgatggccc actacgtgaa ccatcacctt aatcaagttt tttggggctcg 8940
 aggtgccgta aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgg 9000
 ggaaagccgg cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg 9060
 gcgctggcaa gtgtagcggc cacgctgctc gtaaccacca cacccgccgc gcttaatgcg 9120
 ccgctacagg gcgctccat tcgccattca ggctgctgcaa ctggtgggaa gggcgat 9177

<210> 7
 <211> 29
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Cebador hGKRP-EcoR1

<400> 7
 ggaattcatg ccaggcacia aacggtttc 29

<210> 8
 <211> 33
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Cebador hGKRP-BamH1

<400> 8
 gggatcctac tgaacgtcag gctctaggat ttc 33

<210> 9
 <211> 1398
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)..(1398)

<400> 9
 atg ctg gac gac aga gcc agg atg gag gcc gcc aag aag gag aag gta 48
 Met Leu Asp Asp Arg Ala Arg Met Glu Ala Ala Lys Lys Glu Lys Val
 1 5 10 15

gag cag atc ctg gca gag ttc cag ctg cag gag gag gac ctg aag aag 96
 Glu Gln Ile Leu Ala Glu Phe Gln Leu Gln Glu Glu Asp Leu Lys Lys
 20 25 30

gtg atg aga cgg atg cag aag gag atg gac cgc ggc ctg agg ctg gag 144
 Val Met Arg Arg Met Gln Lys Glu Met Asp Arg Gly Leu Arg Leu Glu
 35 40 45

acc cat gaa gag gcc agt gtg aag atg ctg ccc acc tac gtg cgc tcc 192
 Thr His Glu Glu Ala Ser Val Lys Met Leu Pro Thr Tyr Val Arg Ser
 50 55 60

acc cca gaa ggc tca gaa gtc ggg gac ttc ctc tcc ctg gac ctg ggt 240
 Thr Pro Glu Gly Ser Glu Val Gly Asp Phe Leu Ser Leu Asp Leu Gly
 15

ES 2 646 388 A1

65	70					75					80					
ggc Gly	act Thr	aac Asn	ttc Phe	agg Arg 85	gtg Val	atg Met	ctg Leu	gtg Val	aag Lys 90	gtg Val	gga Gly	gaa Glu	ggt Gly	gag Glu 95	gag Glu	288
ggg Gly	cag Gln	tgg Trp	agc Ser 100	gtg Val	aag Lys	acc Thr	aaa Lys	cac His 105	cag Gln	atg Met	tac Tyr	tcc Ser	atc Ile 110	ccc Pro	gag Glu	336
gac Asp	gcc Ala	atg Met 115	acc Thr	ggc Gly	act Thr	gct Ala	gag Glu 120	atg Met	ctc Leu	ttc Phe	gac Asp	tac Tyr 125	atc Ile	tct Ser	gag Glu	384
tgc Cys	atc Ile 130	tcc Ser	gac Asp	ttc Phe	ctg Leu	gac Asp 135	aag Lys	cat His	cag Gln	atg Met	aaa Lys 140	cac His	aag Lys	aag Lys	ctg Leu	432
ccc Pro 145	ctg Leu	ggc Gly	ttc Phe	acc Thr	ttc Phe 150	tcc Ser	ttt Phe	cct Pro	gtg Val	agg Arg 155	cac His	gaa Glu	gac Asp	atc Ile	gat Asp 160	480
aag Lys	ggc Gly	atc Ile	ctt Leu	ctc Leu 165	aac Asn	tgg Trp	acc Thr	aag Lys	ggc Gly 170	ttc Phe	aag Lys	gcc Ala	tca Ser	gga Gly 175	gca Ala	528
gaa Glu	ggg Gly	aac Asn	aat Asn 180	gtc Val	gtg Val	ggg Gly	ctt Leu	ctg Leu 185	cga Arg	gac Asp	gct Ala	atc Ile	aaa Lys 190	cgg Arg	aga Arg	576
ggg Gly	gac Asp	ttt Phe 195	gaa Glu	atg Met	gat Asp	gtg Val	gtg Val 200	gca Ala	atg Met	gtg Val	aat Asn	gac Asp 205	acg Thr	gtg Val	gcc Ala	624
acg Thr	atg Met 210	atc Ile	tcc Ser	tgc Cys	tac Tyr	tac Tyr 215	gaa Glu	gac Asp	cat His	cag Gln	tgc Cys 220	gag Glu	gtc Val	ggc Gly	atg Met	672
atc Ile 225	gtg Val	ggc Gly	acg Thr	ggc Gly	tgc Cys 230	aat Asn	gcc Ala	tgc Cys	tac Tyr	atg Met 235	gag Glu	gag Glu	atg Met	cag Gln	aat Asn 240	720
gtg Val	gag Glu	ctg Leu	gtg Val	gag Glu 245	ggg Gly	gac Asp	gag Glu	ggc Gly	cgc Arg 250	atg Met	tgc Cys	gtc Val	aat Asn	acc Thr 255	gag Glu	768
tgg Trp	ggc Gly	gcc Ala	ttc Phe 260	ggg Gly	gac Asp	tcc Ser	ggc Gly	gag Glu 265	ctg Leu	gac Asp	gag Glu	ttc Phe	ctg Leu 270	ctg Leu	gag Glu	816
tat Tyr	gac Asp	cgc Arg 275	ctg Leu	gtg Val	gac Asp	gag Glu	agc Ser 280	tct Ser	gca Ala	aac Asn	ccc Pro	ggt Gly 285	cag Gln	cag Gln	ctg Leu	864
tat Tyr	gag Glu 290	aag Lys	ctc Leu	ata Ile	ggt Gly	ggc Gly 295	aag Lys	tac Tyr	atg Met	ggc Gly	gag Glu 300	ctg Leu	gtg Val	cgg Arg	ctt Leu	912
gtg Val 305	ctg Leu	ctc Leu	agg Arg	ctc Leu	gtg Val 310	gac Asp	gaa Glu	aac Asn	ctg Leu	ctc Leu 315	ttc Phe	cac His	ggg Gly	gag Glu	gcc Ala 320	960
tcc Ser	gag Glu	cag Gln	ctg Leu	cgc Arg 325	aca Thr	cgc Arg	gga Gly	gcc Ala	ttc Phe 330	gag Glu	acg Thr	cgc Arg	ttc Phe	gtg Val 335	tcg Ser	1008
cag Gln	gtg Val	gag Glu	agc Ser 340	gac Asp	acg Thr	ggc Gly	gac Asp	cgc Arg 345	aag Lys	cag Gln	atc Ile	tac Tyr	aac Asn 350	atc Ile	ctg Leu	1056

ES 2 646 388 A1

agc acg ctg ggg ctg cga ccc tcg acc acc gac tgc gac atc gtg cgc 1104
 Ser Thr Leu Gly Leu Arg Pro Ser Thr Thr Asp Cys Asp Ile Val Arg
 355 360 365
 cgc gcc tgc gag agc gtg tct acg cgc gct gcg cac atg tgc tcg gcg 1152
 Arg Ala Cys Glu Ser Val Ser Thr Arg Ala Ala His Met Cys Ser Ala
 370 375 380
 ggg ctg gcg ggc gtc atc aac cgc atg cgc gag agc cgc agc gag gac 1200
 Gly Leu Ala Gly Val Ile Asn Arg Met Arg Glu Ser Arg Ser Glu Asp
 385 390 395 400
 gta atg cgc atc act gtg ggc gtg gat ggc tcc gtg tac aag ctg cac 1248
 Val Met Arg Ile Thr Val Gly Val Asp Gly Ser Val Tyr Lys Leu His
 405 410 415
 ccc agc ttc aag gag cgg ttc cat gcc agc gtg cgc agg ctg acg ccc 1296
 Pro Ser Phe Lys Glu Arg Phe His Ala Ser Val Arg Arg Leu Thr Pro
 420 425
 agc tgc gag atc acc ttc atc gag tcg gag gag ggc agt ggc cgg ggc 1344
 Ser Cys Glu Ile Thr Phe Ile Glu Ser Glu Glu Gly Ser Gly Arg Gly
 435 440 445
 gcg gcc ctg gtc tcg gcg gtg gcc tgt aag aag gcc tgt atg ctg ggc 1392
 Ala Ala Leu Val Ser Ala Val Ala Cys Lys Lys Ala Cys Met Leu Gly
 450 455 460
 cag tga 1398
 Gln
 465

<210> 10
 <211> 465
 <212> PRT
 <213> Homo sapiens

<400> 10

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

ES 2 646 388 A1

Asp Ala Met Thr Gly Thr Ala Glu Met Leu Phe Asp Tyr Ile Ser Glu
 115 120 125
 Cys Ile Ser Asp Phe Leu Asp Lys His Gln Met Lys His Lys Lys Leu
 130 135 140
 Pro Leu Gly Phe Thr Phe Ser Phe Pro Val Arg His Glu Asp Ile Asp
 145 150 155 160
 Lys Gly Ile Leu Leu Asn Trp Thr Lys Gly Phe Lys Ala Ser Gly Ala
 165 170 175
 Glu Gly Asn Asn Val Val Gly Leu Leu Arg Asp Ala Ile Lys Arg Arg
 180 185 190
 Gly Asp Phe Glu Met Asp Val Val Ala Met Val Asn Asp Thr Val Ala
 195 200 205
 Thr Met Ile Ser Cys Tyr Tyr Glu Asp His Gln Cys Glu Val Gly Met
 210 215 220
 Ile Val Gly Thr Gly Cys Asn Ala Cys Tyr Met Glu Glu Met Gln Asn
 225 230 235 240
 Val Glu Leu Val Glu Gly Asp Glu Gly Arg Met Cys Val Asn Thr Glu
 245 250 255
 Trp Gly Ala Phe Gly Asp Ser Gly Glu Leu Asp Glu Phe Leu Leu Glu
 260 265 270
 Tyr Asp Arg Leu Val Asp Glu Ser Ser Ala Asn Pro Gly Gln Gln Leu
 275 280 285
 Tyr Glu Lys Leu Ile Gly Gly Lys Tyr Met Gly Glu Leu Val Arg Leu
 290 295 300
 Val Leu Leu Arg Leu Val Asp Glu Asn Leu Leu Phe His Gly Glu Ala
 305 310 315 320
 Ser Glu Gln Leu Arg Thr Arg Gly Ala Phe Glu Thr Arg Phe Val Ser
 325 330 335
 Gln Val Glu Ser Asp Thr Gly Asp Arg Lys Gln Ile Tyr Asn Ile Leu
 340 345 350
 Ser Thr Leu Gly Leu Arg Pro Ser Thr Thr Asp Cys Asp Ile Val Arg
 355 360 365
 Arg Ala Cys Glu Ser Val Ser Thr Arg Ala Ala His Met Cys Ser Ala
 370 375 380
 Gly Leu Ala Gly Val Ile Asn Arg Met Arg Glu Ser Arg Ser Glu Asp
 18

ES 2 646 388 A1

<220>

<223> Plásmido comercial pACT2

<400> 13

aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttatfff tctaaataca	60
ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa	120
aaggaagagt atgagtattc aacatttccg tgtcgcctt attccctfff ttgcggcatt	180
ttgccttctt gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca	240
gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag	300
ttttcgcccc gaagaacgtt ttccaatgat gagcactfff aaagttctgc tatgtggcgc	360
ggtattatcc cgtgttgacg ccgggcaaga gcaactcggc cggccatac actattctca	420
gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt	480
aagagaatta tgcagtgctg ccataacatc gagtgataac actgcggcca acttacttct	540
gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt	600
aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga	660
caccacgatg cctgcagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact	720
tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc	780
acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga	840
gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt	900
agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgcctga	960
gataggtgcc tcaactgatta agcattggta actgtcagac caagtttact catatatact	1020
ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tcctttttga	1080
taatctcatg accaaaatcc cttaacgtga gttttcgctc cactgagcgt cagaccccgt	1140
agaaaagatc aaaggatcct cttgagatcc ttttttctg cgcgtaatct gctgcttgca	1200
aacaaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct	1260
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta	1320
gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct	1380
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc	1440
aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt cgtgcacaca	1500
gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga	1560
aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg	1620
aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatatctt atagtcctgt	1680
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag	1740
cctatgaaa aacgccagca acgcggcctt tttacggttc ctggcctfff gctggcctff	1800
tgctcacatg ttctttctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt	1860
tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcagat cagtgagcga	1920
ggaagcggaa gagcgcctga tgcggtatff tctccttacg catctgtgcg gtatffcaca	1980

ES 2 646 388 A1

ccgcatatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatac 2040
 actccgctat cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc aacacccgct 2100
 gacgcgccct gacgggcttg tctgctcccg gcatccgctt acagacaagc tgtgaccgctc 2160
 tccgggagct gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gaggcagcaa 2220
 cttcttttct ttttttttct tttctctctc ccccgttggt gtctcaccat atccgcaatg 2280
 acaaaaaaaaa tgatggaaga cactaaagga aaaaattaac gacaaagaca gcaccaacag 2340
 atgtcgttgt tccagagctg atgaggggta tcttcgaaca cacgaaactt tttccttctt 2400
 tcattcacgc aactactct ctaatgagca acggtatagc gccttccttc cagttacttg 2460
 aatttgaaat aaaaaaagtt tgccgctttg ctatcaagta taaatagacc tgcaattatt 2520
 aatcttttgt ttctctgtca ttgttctcgt tcccttttct ccttgtttct ttttctgcac 2580
 aatatttcaa gctataccea gcatacaatc aactccaagc tttgcaaaga tggataaagc 2640
 ggaattaatt cccgagcctc caaaaaagaa gagaaaggtc gaattgggta ccgccgcaa 2700
 ttttaataca agtgggaata ttgctgatag ctattgtcc ttcactttca ctaacagtag 2760
 caacggtccg aacctcataa caactcaaac aaattctcaa gcgctttcac aaccaattgc 2820
 ctctctaac gttcatgata acttcatgaa taatgaaatc acggctagta aaattgatga 2880
 tggaataaat tcaaaaccac tgtcacctgg ttggacggac caaactgcgt ataacgcggt 2940
 tggaatcact acagggatgt ttaataccac tacaatggat gatgtatata actatctatt 3000
 cgatgatgaa gatacccac caaacccaaa aaaagagatc tgtatggctt acccatacga 3060
 tgttccagat tacgctagct tgggtgggtca tatggccatg gaggccccgg ggatccgaat 3120
 tcgagctcga gagatctatg aatcgtagat actgaaaaac cccgcaagtt cacttcaact 3180
 gtgcatcgtg caccatctca atttctttca tttatacatc gttttgcctt cttttatgta 3240
 actatactcc tctaagtttc aatcttggcc atgtaacctc tgatctatag aattttttaa 3300
 atgactagaa ttaatgccca tctttttttt ggacctaat tcttcatgaa aatatattac 3360
 gagggcttat tcagaagctt tggacttctt cgccagaggt ttggtcaagt ctccaatcaa 3420
 ggttgtcggc ttgtctacct tgccagaaat ttacgaaaag atggaaaagg gtcaaactcg 3480
 tggtagatac gttgttgaca cttctaaata agcgaatttc ttatgattta tgatttttat 3540
 tattaataaa gttataaaaa aaataagtgt atacaaattt taaagtgact cttaggtttt 3600
 aaaacgaaaa ttcttgttct tgagtaactc tttctgtag gtcaggttgc tttctcaggt 3660
 atagcatgag gtcgctctta ttgaccacac ctctaccggc atgtgcgagg gacctaataa 3720
 cttcgtatag catacattat acgaagtat attaaggggt ccggatcgcg gccgctcgac 3780
 ctgcagccaa gctagcggac gtaaactcct cttcagacct aataacttcg tatagcatac 3840
 attatacгаа gttatattaa gggttattga atatgatcgg aattggtcga gggacctaat 3900
 aacttcgat agcatacatt atacgaagtt atattaaggg ttccggatcg cggccgctcg 3960
 acctgcagcc aggctagctt ggctggacgt aaactcctct tcagacctaa taacttcgat 4020

ES 2 646 388 A1

agcatacatt atacgaagtt atattaaggg ttattgatat gatcgggaatt ggtcgaactac 4080
 gtcgttaagg ccgtttctga cagagtaaaa ttcttgaggg aactttcacc attatgggaa 4140
 atggttcaag aaggatttga cttaaactcc atcaaaggtt cagggtcattg agtggtttttt 4200
 atttgttgta tttttttttt ttttagagaaa atcctccaat atcaaattag gaatcgtagt 4260
 ttcattgattt tctgttacac ctaacttttt gtgtgggtgcc ctctccttg tcaatattaa 4320
 tgtaaagtg caattctttt tccttatcac gttgagccat tagtatcaat ttgcttacct 4380
 gtattccttt actatcctcc tttttctcct tcttgataaa tgtatgtaga ttgcgtatat 4440
 agtttcgtct accctatgaa catattccat tttgtaattt cgtgtcgttt ctattatgaa 4500
 tttcatttat aaagtttatg tacaatatc ataaaaaag agaacttttt taagcaagga 4560
 ttttctaac ttcttcggcg acagcatcac cgacttcggt ggtactgttg gaaccaccta 4620
 aatcaccagt tctgatacct gcatccaaaa cttttttaac tgcattctca atggccttac 4680
 ctcttcagg caagttcaat gacaatttca acatcattgc agcagacaag atagtggcga 4740
 tagggtcaac ctattctttt ggcaaatctg gagcagaacc gtggcatggt tcgtacaaac 4800
 caaatgcggt gttcttgtct ggcaaagagg ccaaggacgc agatggcaac aaaccaaggg 4860
 aacctgggat aacggaggct tcatcggaga tgatatacacc aaacatggtg ctggtgatta 4920
 taataccatt taggtggggtt gggttcttaa ctaggatcat ggcggcagaa tcaatcaatt 4980
 gatgttgaac ctcaatgta ggaaattcgt tcttgatggt ttctccaca gtttttctcc 5040
 ataactttga agaggccaaa acattagctt tatccaagga ccaaataggc aatggtggct 5100
 catgtttagt ggccatgaaa gcggccattc ttgtgattct ttgcacttct ggaacggtgt 5160
 attgttcaact atccaagcg acaccatcac catcgtcttc ctttctctta ccaaagtaaa 5220
 tacctcccac taattctctg acaacaacga agtcagtacc tttagcaaat tgtggcttga 5280
 ttggagataa gtctaaaaga gagtcggatg caaagttaca tggctttaag ttggcgtaca 5340
 attgaagttc ttacggatt tttagtaaac ctgttccagg tctaacta ccggtacccc 5400
 atttaggacc acccacagca cctaacaaaa cggcatcaac ctctttggag gcttccagcg 5460
 cctcatctgg aagtgggaca cctgtagcat cgatagcagc accaccaatt aaatgatttt 5520
 cgaaatcgaa cttgacattg gaacgaacat cagaaatagc tttagaacc ttaatggctt 5580
 cggctgtgat ttcttgacca acgtgggtcac ctggcaaac gacgatcttc ttaggggag 5640
 acataggggc agacattaga atggtatatc ctgaaatat atatatatat tgctgaaatg 5700
 taaaaggtaa gaaaagttag aaagtaagac gattgctaac cacctattgg aaaaaacaat 5760
 aggtccttaa ataattttgt caacttcaag tattgtgatg caagcattta gtcatgaacg 5820
 ctctctatt ctatatgaaa agccggttcc ggcctctcac ctttctttt tctccaatt 5880
 tttcagttga aaaaggtata tgcgtcaggc gacctctgaa attaacaaaa aatttccagt 5940
 catcgaattt gattctgtgc gatagcggc ctgtgtgttc tcgttatggt gaggaaaaaa 6000
 ataaggttg ctaagagatt cgaactctg catcttacga tacctgagta tccccacagt 6060
 tccttaccac tcttttgta ctctattgat ccagctcagc aaaggcagtg tgatctaaga 6120

ES 2 646 388 A1

ttctatcttc	gcgatgtagt	aaaactagct	agaccgagaa	agagactaga	aatgcaaaaag	6180
gcacttctac	aatggctgcc	atcattatta	tccgatgtga	cgctgcagct	tctcaatgat	6240
attcgaatac	gctttgagga	gatacagcct	aatatccgac	aaactgtttt	acagattttac	6300
gatcgtactt	gttaccatc	attgaatfff	gaacatccga	acctgggagt	ttccctgaa	6360
acagatagta	tatttgaacc	tgtataataa	tatatagtct	agcgctttac	ggaagacaat	6420
gtatgtatff	cggttcctgg	agaaactatt	gcatctattg	cataggtaat	cttgcacgtc	6480
gcatccccgg	ttcattttct	gcgtttccat	cttgcacttc	aatagcatat	ctttgttaac	6540
gaagcatctg	tgcttcattt	tgtagaacaa	aaatgcaacg	cgagagcgct	aatfffftcaa	6600
acaaagaatc	tgagctgcat	ttttacagaa	cagaaatgca	acgcgaaagc	gctatffffac	6660
caacgaagaa	tctgtgcttc	atfffftghaa	aacaaaaatg	caacgcgaga	gcgctaattt	6720
ttcaaaaaa	gaatctgagc	tgcatffffta	cagaacagaa	atgcaacgcg	agagcgctat	6780
ttaccaaca	aagaatctat	acttctffff	tgttctacaa	aaatgcatcc	cgagagcgct	6840
atfffftctaa	caaagcatct	tagattactt	tttttctcct	ttgtgcgctc	tataatgcag	6900
tctcttgata	actfffftgca	ctgtaggtcc	gttaaggfta	gaagaaggct	actttgggtg	6960
ctatfffftctc	ttccataaaa	aaagcctgac	tccacttccc	gcgtttactg	attactagcg	7020
aagctgcggg	tgcatfffft	caagataaag	gcatccccga	ttatattcta	taccgatgtg	7080
gattgcgcat	actttgtgaa	cagaaagtga	tagcgttgat	gattcttcat	tggtcagaaa	7140
attatgaacg	gtttcttcta	ttttgtctct	atatactacg	tataggaaat	gtttacattt	7200
tcgtattgft	ttcgattcac	tctatgaata	gttcttacta	caatfffft	gtctaaagag	7260
taatactaga	gataaacata	aaaaatgtag	aggtcgagtt	tagatgcaag	ttcaaggagc	7320
gaaaggftgga	tggttaggft	atatagggat	atagcacaga	gatatatagc	aaagagatac	7380
ttttgagcaa	tgtttgftgga	agcggatttc	gcaatatfff	agtagctcgt	tacagtccgg	7440
tgcgtttttg	gttttttgaa	agtgcgtctt	cagagcgctt	ttggttttca	aaagcgctct	7500
gaagttccta	tactttctag	agaataggaa	cttcggaata	ggaacttcaa	agcgtttccg	7560
aaaacgagcg	cttccgaaaa	tgcaacgcga	gctg'gcaca	tacagctcac	tgttcacgtc	7620
gcacctatat	ctgcgtgftg	cctgtatata	tatatacatg	agaagaacgg	catagtgcgt	7680
gtttatgctt	aatgcgtac	ttatatgcgt	ctatfttatgt	aggatgaaag	gtagtctagt	7740
acctcctgtg	atattatccc	attccatgcg	gggtatcgta	tgcttcttcc	agcactaccc	7800
tttagctgft	ctatatgctg	ccactcctca	attggattag	tctcatcctt	caatgctatc	7860
atftcctftg	atattggatc	atatgcatag	taccgagaaa	ctagtgcgaa	gtagtgatca	7920
ggatftgctg	ttatctgatg	agtatacgft	gtcctggcca	cggcagaagc	acgcttatcg	7980
ctccaatftc	ccacaacatt	agtcaactcc	gttaggccct	tcattgaaag	aaatgaggftc	8040
atcaaatgftc	ttccaatgftg	agatftttggg	ccatftttfta	tagcaaagat	tgaataaggc	8100
gcatftttct	tcaaagc					8117

ES 2 646 388 A1

<210> 14
 <211> 9594
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Plásmido pACT2-GK-3xPTAP

<400> 14
 aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttatttt tctaaataca 60
 ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 120
 aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 180
 ttgccttctt gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 240
 gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 300
 ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 360
 ggtattatcc cgtgttgacg cggggcaaga gcaactcggg cggccgatac actatttctca 420
 gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 480
 aagagaatta tgcaagtctg ccataaccat gagtgataac actgcggcca acttacttct 540
 gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt 600
 aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 660
 caccacgatg cctgcagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 720
 tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 780
 acttctgctc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga 840
 gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 900
 agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgtctga 960
 gataggtgcc tcaactgatta agcattggta actgtcagac caagtttact catatatact 1020
 ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tcctttttga 1080
 taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 1140
 agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 1200
 aacaaaaaaaa ccaccgctac cagcgggtgg ttgtttgccg gatcaagagc taccaactct 1260
 ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgtgta 1320
 gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 1380
 aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttgactc 1440
 aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggt cgtgcacaca 1500
 gccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga 1560
 aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg 1620
 aacaggagag cgcacgaggg agcttcagg gggaaacgcc tggatcttt atagtcctgt 1680
 cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag 1740
 cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt 1800

ES 2 646 388 A1

tgctcacatg	ttctttcctg	cgttatcccc	tgattctgtg	gataaccgta	ttaccgcctt	1860
tgagtgagct	gataccgctc	gccgcagccg	aacgaccgag	cgagcgaggt	cagtgagcga	1920
ggaagcggaa	gagcgcctga	tgcggtat	tctccttacg	catctgtgcg	gtatttcaca	1980
ccgcatatgg	tgactctca	gtacaatctg	ctctgatgcc	gcatagttaa	gccagtatac	2040
actccgctat	cgctacgtga	ctgggtcatg	gctgcgcccc	gacacccgcc	aacacccgct	2100
gacgcgccct	gacgggcttg	tctgctcccc	gcatccgctt	acagacaagc	tgtgaccgtc	2160
tccgggagct	gcatgtgtca	gaggttttca	ccgtcatcac	cgaaacgcgc	gaggcagcaa	2220
cttcttttct	tttttttctt	tttctctctc	ccccgttggt	gtctcaccat	atccgcaatg	2280
acaaaaaaaa	tgatggaaga	cactaaagga	aaaaattaac	gacaaagaca	gcaccaacag	2340
atgtcgttgt	tccagagctg	atgaggggta	tcttcgaaca	cacgaaactt	tttctttcct	2400
tcattcacgc	acactactct	ctaatgagca	acggtatacg	gccttccttc	cagttacttg	2460
aatttgaaat	aaaaaaagtt	tgccgctttg	ctatcaagta	taaatagacc	tgcaattatt	2520
aatcttttgt	ttcctcgtca	ttgttctcgt	tccctttctt	ccttgtttct	ttttctgcac	2580
aatatttcaa	gctataccaa	gcatacaatc	aactccaagc	tttgcaaaga	tgataaaagc	2640
ggaattaatt	cccgagcctc	caaaaaagaa	gagaaaggtc	gaattgggta	ccgccgcaa	2700
ttttaatcaa	agtgggaata	ttgctgatag	ctcattgtcc	ttcactttca	ctaacagtag	2760
caacgggtccg	aacctcataa	caactcaaac	aaattctcaa	gcgctttcac	aaccaattgc	2820
ctcctctaac	gttcatgata	acttcatgaa	taatgaaatc	acggctagta	aaattgatga	2880
tggtataaat	tcaaaaccac	tgtcacctgg	ttggacggac	caaactgcgt	ataacgcggt	2940
tggaatcact	acagggatgt	ttaataccac	tacaatggat	gatgtatata	actatctatt	3000
cgatgatgaa	gatacccccac	caaaccctaaa	aaaagagatc	tgtatggctt	accatacga	3060
tgttccagat	tacgctagct	tgggtgggtca	tatggccatg	gaggccccgg	ggatcctgat	3120
gctggacgac	agagccagga	tggaggccgc	caagaaggag	aaggtagagc	agatcctggc	3180
agagttccag	ctgcaggagg	aggacctgaa	gaaggtgatg	agacggatgc	agaaggagat	3240
ggaccgcggc	ctgaggctgg	agacctatga	agaggccagt	gtgaagatgc	tgcccaccta	3300
cgctgcgctc	acccagaag	gctcagaagt	cggggacttc	ctctccctgg	acctgggtgg	3360
cactaacttc	agggtgatgc	tgggtgaaggt	gggagaaggt	gaggaggggc	agtggagcgt	3420
gaagaccaa	caccagatgt	actccatccc	cgaggacgcc	atgaccggca	ctgctgagat	3480
gctcttcgac	tacatctctg	agtgcattctc	cgacttcctg	gacaagcatc	agatgaaaca	3540
caagaagctg	cccctgggct	tcaccttctc	ctttcctgtg	aggcacgaag	acatcgataa	3600
gggcatcctt	ctcaactgga	ccaagggtctt	caaggcctca	ggagcagaag	ggaacaatgt	3660
cggtgggctt	ctgcgagacg	ctatcaaacg	gagaggggac	tttgaaatgg	atgtgggtggc	3720
aatggtgaat	gacacggtgg	ccacgatgat	ctcctgctac	tacgaagacc	atcagtgcca	3780
ggtcggcatg	atcgtgggca	cgggctgcaa	tgctgctac	atggaggaga	tgcaaatgt	3840

ES 2 646 388 A1

ggagctggtg gagggggacg agggccgcat gtgctgcaat accgagtggg ggcgcttcgg 3900
 ggactccggc gagctggacg agttcctgct ggagtatgac cgcctggtgg acgagagctc 3960
 tgcaaaccce ggtcagcagc tgtatgagaa gctcataggt ggcaagtaca tgggagagct 4020
 ggtgctggctt gtgctgctca ggctcgtgga cgaaaacctg ctcttcacg gggaggcctc 4080
 cgagcagctg cgcacacgcy gagccttcga gacgcgcttc gtgtcgcagg tggagagcga 4140
 cacgggcyac cgcaagcaga tctacaacat cctgagcacg ctggggctgc gaccctcgac 4200
 caccgactgc gacatcgtgc gccgcgctg cgagagcgtg tctacgcgcy ctgcyacat 4260
 gtgctcggcy gggctggcgg gcyctatcaa ccgcatgcy gagagccgca gcyaggagct 4320
 aatgcyatc actgtgggcy tggatggctc cgtgtacaag ctgcaccca gcttcaagga 4380
 gcggttccat gccagcgtgc gcaggctgac gccagctgc gagatcacct tcatcgagtc 4440
 ggaggagggc agtggccggg gcygcyccct ggtctcggcy gtggcctgta agaaggcctg 4500
 tatgctgggc cagtggatcc gaattcgagc tcgagagcca gaaccacag caccgcctga 4560
 gcctaccgcc ccaccgaac cgacggcgc tccagctgag taatcgagag atctatgaat 4620
 cgtagatact gaaaaacccc gcaagttcac ttcaactgtg catcgtgcac catctcaatt 4680
 tctttcattt atacatcgtt ttgccttctt ttatgtaact atactcctct aagtttcaat 4740
 cttggccatg taacctctga tctatagaat tttttaaag actagaatta atgcccattt 4800
 tttttttgga cctaaattct tcatgaaaat atattacgag ggcttattca gaagctttgg 4860
 acttcttcgc cagaggtttg gtcaagtctc caatcaaggt tgtcggcttg tctaccttg 4920
 cagaaattta cgaaaagatg gaaaagggtc aaatcgttg tagatacgtt gttgacactt 4980
 ctaaataagc gaatttctta tgatttatga tttttattat taaataagtt ataaaaaaaa 5040
 taagtgtata caaattttaa agtgactctt aggtttttaa acgaaaattc ttgttcttga 5100
 gtaactctt cctgtaggtc aggttgctt ctcaggata gcatgaggtc gctcttattg 5160
 accacacctc taccggcatg tgcgagggac ctaataactt cgtatagcat acattatacg 5220
 aagttatatt aagggttccg gatcgcggcc gctcgacctg cagccaagct agcggagcga 5280
 aactcctctt cagacctaata aacttcgtat agcatacatt atacgaagtt atattaaggg 5340
 ttattgaata tgatcggaat tggctgaggg acctaaataac ttcgtatagc atacattata 5400
 cgaagttata ttaagggttc cggatcgcgg ccgctcgacc tgcagccagg ctagcttggc 5460
 tggacgtaaa ctctcttca gacctaaata cttcgatagc atacattata cgaagttata 5520
 ttaagggtta ttgatatgat cggatttggc cgactacgct gtttaaggccg tttctgacag 5580
 agtaaaattc ttgagggaac tttcaccatt atgggaaatg gttcaagaag gtattgactt 5640
 aaactccatc aaatggctcag gtcattgagt gttttttatt tttttttttt 5700
 agagaaaatc ctccaatc aaattaggaa tcgtagtttc atgattttct gttacaccta 5760
 actttttgtg tggctccctc ctcttctca atattaatgt taaagtcaa tctttttcc 5820
 ttatcacgtt gagccattag tatcaatttg cttacctgta ttcctttact atcctcctt 5880
 tctccttct tgataaatgt atgtagattg cgtatatagt ttcgtctacc ctatgaacat 5940

ES 2 646 388 A1

attccatfff gtaatttcgt gtcgtttcta ttatgaatff catttataaa gtttatgtac 6000
 aatatcata aaaaaagaga atctttttta gcaaggatff tcttaacttc ttcggcgaca 6060
 gcatcaccga cttcgggtgg actgttggaa ccacctaaat caccagttct gataacctgca 6120
 tccaaaacct ttttaactgc atcttcaatg gccttacctt cttcaggcaa gttcaatgac 6180
 aatttcaaca tcattgcagc agacaagata gtggcgatag ggtcaacctt attctttggc 6240
 aaatctggag cagaaccgtg gcatggttcg taaaaaccaa atgcggtggt cttgtctggc 6300
 aaagaggcca aggacgcaga tggcaacaaa cccaaggaac ctgggataac ggaggcttca 6360
 tcggagatga tatcaccaaa catgttgctg gtgattataa taccatttag gtgggttggg 6420
 ttcttaacta ggatcatggc ggcagaatca atcaattgat gttgaacctt caatgtagga 6480
 aattcgttct tgatggtttc ctccacagtt tttctccata atcttgaaga ggccaaaaca 6540
 tttagctttat ccaaggacca aataggcaat ggtggctcat gttgtagggc catgaaagcg 6600
 gccattcttg tgattctttg cacttctgga acggtgtatt gttcactatc ccaagcgaca 6660
 ccatcaccat cgtcttcctt tctcttacca aagtaaatac ctcccactaa ttctctgaca 6720
 acaacgaagt cagtacctt agcaaatgt ggcttgattg gagataagtc taaaagagag 6780
 tcggatgcaa agttacatgg tcttaagttg gcgtacaatt gaagttcttt acggatffff 6840
 agtaaacctt gttcaggtct aacactaccg gtaccccat taggaccacc cacagcacct 6900
 acaaaaacgg catcaacctt cttggaggct tccagcgcct catctggaag tgggacacct 6960
 gtagcatcga tagcagcacc accaattaaa tgatfffctga aatcgaactt gacattggaa 7020
 cgaacatcag aatagcttt aagaacctta atggcttcgg ctgtgatttc ttgaccaacg 7080
 tggtcacctg gcaaaacgac gatcttctta ggggcagaca taggggcaga cattagaatg 7140
 gtatatcctt gaaatatata tatatattgc tgaaatgtaa aaggtaagaa aagttagaaa 7200
 gtaagacgat tgctaaccac ctattggaaa aaacaatagg tccttaata atattgtcaa 7260
 cttcaagtat tgtgatgcaa gcatttagtc atgaacgctt ctctattcta tatgaaaagc 7320
 cggttccggc ctctcacctt tcctttttct cccaatfff cagttgaaaa aggtatatgc 7380
 gtcaggcgac ctctgaaatt acaaaaaat ttccagtcac cgaatffgat tctgtgcat 7440
 agcggccctg tgtgttctcg ttatgttgag gaaaaaata atggttgcta agagattcga 7500
 actcttgcat cttacgatac ctgagtattc ccacagttcc ttaccactct tttgttactc 7560
 tattgatcca gctcagcaaa ggcagtgatga tctaagattc tatcttcgag atgtagtaaa 7620
 actagctaga ccgagaaaaga gactagaaat gcaaaaggca cttctacaat ggctgccatc 7680
 attattatcc gatgtgacgc tgcagcttct caatgatatt cgaatacgct ttgaggagat 7740
 acagccta atccgacaaa ctgttttaca gatttacgat cgtacttggt acccatcatt 7800
 gaatfffcaa catccgaacc tgggagtttt ccctgaaaca gatagtatat ttgaacctgt 7860
 ataataatat atagtctagc gctttacgga agacaatgta tgtatffctgg ttcctggaga 7920
 aactattgca tctattgcat aggtaatctt gcacgtcgca tccccggttc atffttctgcg 7980

ES 2 646 388 A1

tttccatctt gcacttcaat agcatatctt tgttaacgaa gcatctgtgc ttcattttgt 8040
 agaacaaaaa tgcaacgcga gagcgctaatt ttttcaaaca aagaatctga gctgcatttt 8100
 tacagaacag aaatgcaacg cgaagcgct attttaccaa cgaagaatct gtgcttcatt 8160
 tttgtaaaac aaaaatgcaa cgcgagagcg ctaatttttc aaacaaagaa tctgagctgc 8220
 atttttacag aacagaaatg caacgcgaga gcgctatttt accaacaag aatctatact 8280
 tctttttgt tctacaaaaa tgcaccccga gagcgctatt tttctaaca agcatcttag 8340
 attacttttt ttctccttg tgcgctctat aatgcagtct cttgataact ttttgcactg 8400
 taggtccgtt aagggttaga gaaggctact ttgggtgtcta ttttctcttc cataaaaaaa 8460
 gcctgactcc acttcccgcg tttactgatt actagcgaag ctgcggtgc attttttcaa 8520
 gataaaggca tccccgatta tattctatac cgatgtggat tgcgcatact ttgtgaacag 8580
 aaagtgatag cgttgatgat tcttcattgg tcagaaaatt atgaacgggt tcttctattt 8640
 tgtctctata tactacgtat aggaaatggt tacattttcg tattgttttc gattcactct 8700
 atgaatagtt ctactacaa tttttttgtc taaagagtaa tactagagat aaacataaaa 8760
 aatgtagagg tcgagtttag atgcaagttc aaggagcgaa aggtggatgg gtaggttata 8820
 tagggatata gcacagagat atatagcaa gagatacttt tgagcaatgt ttgtggaagc 8880
 ggtattcgca atattttagt agctcgttac agtccggtgc gtttttggtt ttttgaaagt 8940
 gcgtcttcag agcgcttttg gttttcaaaa gcgctctgaa gttcctatac tttctagaga 9000
 ataggaactt cggaatagga acttcaaagc gtttccgaaa acgagcgctt ccgaaaatgc 9060
 aacgcgagct gcgcacatac agctcactgt tcacgtcgca cctatatctg cgtgttgctt 9120
 gtatatatat atacatgaga agaacggcat agtgcgtggt tatgcttaaa tgcgtactta 9180
 tatgctgcta tttatgtagg atgaaaggta gtctagtacc tcctgtgata ttatcccatt 9240
 ccattgctggg tatcgtatgc ttccttcagc actacccttt agctgttcta tatgctgcca 9300
 ctctcaatt ggattagtct catccttcaa tgctatcatt tcctttgata ttggatcata 9360
 tgcatagtac cgagaaacta gtgcgaagta gtgatcaggt attgctgtta tctgatgagt 9420
 atacgttgc ctggccacgg cagaagcag cttatcgctc caatttcca caacattagt 9480
 caactccgtt aggcccttca ttgaaagaaa tgaggctatc aaatgtcttc caatgtgaga 9540
 ttttgggcca tttttatag caaagattga ataaggcgca tttttcttca aagc 9594

<210> 15
 <211> 342
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Dominio de activación transcripcional de Gal4 (GAD-Gal4 Activating Domain)

<400> 15
 gccaatTTTA atcaaagtgg gaatattgct gatagctcat tgccttcac tttcactaac 60
 agtagcaacg gtccgaacct cataacaact caacaaatt ctcaagcgct ttcacaacca 120

ES 2 646 388 A1

attgcctcct ctaacgttca tgataacttc atgaataatg aaatcacggc tagtaaaatt 180
 gatgatggta ataattcaaa accactgtca cctggttgga cggaccaaac tgcgtataac 240
 gcgtttgaa tcactacagg gatgtttaat accactacaa tggatgatgt atataactat 300
 ctattcgatg atgaagatac cccaccaaac ccaaaaaaag ag 342

<210> 16
 <211> 398
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> promotor de ADH1 en el plásmido pACT2

<400> 16
 gcaacttctt ttcttttttt ttcttttctc tctccccgt tgttgtctca ccatatccgc 60
 aatgacaaaa aaaatgatgg aagacactaa aggaaaaaat taacgacaaa gacagcacca 120
 acagatgtcg ttgttccaga gctgatgagg ggtatcttcg aacacacgaa actttttcct 180
 tccttcattc acgcacacta ctctctaata agcaacggta tacggccttc cttccagtta 240
 cttgaatttg aaataaaaa agtttgccgc tttgctatca agtataaata gacctgcaat 300
 tattaatctt ttgtttcctc gtcattgttc tcgttcctt tcttccttgt ttctttttct 360
 gcacaatatt tcaagctata ccaagcatac aatcaact 398

<210> 17
 <211> 73
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> oligonucleótido OV570

<400> 17
 tcgagagcca gaaccacag caccgcctga gcctaccgcc ccacccgaac cgacggcgcc 60
 tccagctgag taa 73

<210> 18
 <211> 73
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> oligonucleótido OV571

<400> 18
 tcgattactc agctggaggc gccgtcggtt cgggtggggc ggtaggctca ggcggtgctg 60
 tgggttctgg ctc 73

<210> 19
 <211> 8190
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Plásmido pACT2-3xPTAP

ES 2 646 388 A1

<400> 19
 aggtggcact tttcggggaa atgtgcgcg aaccctatt tgtttatfff tctaaataca 60
 ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 120
 aaggaagagt atgagtattc aacatttccg tgtcgcccctt attccctfff ttgcggcatt 180
 ttgccttctt gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 240
 gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 300
 ttttcgcccc gaagaacgtt ttccaatgat gagcactfff aaagttctgc tatgtggcgc 360
 ggtattatcc cgtgttgacg ccgggcaaga gcaactcggc cgccgatac actatttctca 420
 gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 480
 aagagaatta tgcagtgctg ccataacatc gagtgataac actgcggcca acttacttct 540
 gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaaatgg gggatcatgt 600
 aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 660
 caccacgatg cctgcagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 720
 tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 780
 acttctgctc tcggcccttc cggctggctg gtttattgct gataaatctg gagccgggtga 840
 gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 900
 agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgtgga 960
 gataggtgcc tcaactgatta agcattggta actgtcagac caagtttact catatatact 1020
 ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tcctttttga 1080
 taatctcatg accaaaatcc cttaacgtga gttttcgctc cactgagcgt cagaccccgt 1140
 agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct gctgcttgca 1200
 aacaaaaaaaa ccaccgctac cagcgggtgg ttgtttgccg gatcaagagc taccaactct 1260
 ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta 1320
 gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 1380
 aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg gggtggactc 1440
 aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggt cgtgcacaca 1500
 gccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agctatgaga 1560
 aagcgccacg ctccccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg 1620
 aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatcttt atagtcctgt 1680
 cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag 1740
 cctatggaaa aacgccagca acgcccctt tttacggctc ctggcctfff gctggccttt 1800
 tgctcacatg ttctttctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 1860
 tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcagat cagtgagcga 1920
 ggaagcggaa gagcgcctga tgcggtatft tctccttacg catctgtgcg gtatttcaca 1980
 ccgcatatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatac 2040

ES 2 646 388 A1

actccgctat cgctacgtga ctgggtcatg gctgcgcccc gacacccgcc aacacccgct 2100
 gacgcgccct gacgggcttg tctgctcccc gcatccgctt acagacaagc tgtgaccgctc 2160
 tccgggagct gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gaggcagcaa 2220
 cttcttttct ttttttttct tttctctctc ccccgttggt gtctcaccat atccgcaatg 2280
 acaaaaaaaaa tgatggaaga cactaaagga aaaaattaac gacaaagaca gcaccaacag 2340
 atgtcgttgt tccagagctg atgaggggta tcttcgaaca cacgaaactt tttccttctt 2400
 tcattcacgc aactactct ctaatgagca acggtatagc gccttccttc cagttacttg 2460
 aatttgaat aaaaaagtt tgccgctttg ctatcaagta taaatagacc tgcaattatt 2520
 aatcttttgt ttctctgtca ttgttctcgt tccctttctt ccttgtttct ttttctgcac 2580
 aatatttcaa gctatacaa gcatacaatc aactccaagc tttgcaaaga tggataaagc 2640
 ggaattaatt cccgagcctc caaaaaagaa gagaaaggtc gaattgggta ccgccgcaa 2700
 ttttaatcaa agtgggaata ttgctgatag ctattgtcc ttcactttca ctaacagtag 2760
 caacgggccg aacctcataa caactcaaac aaattctcaa gcgctttcac aaccaattgc 2820
 ctctctaac gttcatgata acttcatgaa taatgaaatc acggctagta aaattgatga 2880
 tggaataaat tcaaaaccac tgtcacctgg ttggacggac caaactgcgt ataacgcggt 2940
 tggaatcact acagggatgt ttaataccac tacaatggat gatgtatata actatctatt 3000
 cgatgatgaa gatacccac caaacccaaa aaaagagatc tgtatggctt acccatacga 3060
 tgttcagat tacgctagct tgggtggtca tatggccatg gaggccccgg ggatccgaat 3120
 tcgagctcga gagccagaac ccacagcacc gcctgagcct accgccccac ccgaaccgac 3180
 ggcgcctcca gctgagtaat cgagagatct atgaatcgta gatactgaaa aaccccgcaa 3240
 gttcacttca actgtgcatc gtgcaccatc tcaatttctt tcatttatac atcgttttgc 3300
 cttcttttat gtaactatac tcctctaagt ttcaatcttg gccatgtaac ctctgatcta 3360
 tagaattttt taaatgacta gaattaatgc ccatcttttt tttggacctt aattcttcat 3420
 gaaaatatat tacgagggct tattcagaag ctttggactt cttcgccaga ggtttgggtca 3480
 agtctccaat caaggttgtc ggcttgctta ccttgccaga aatttacgaa aagatggaaa 3540
 agggctcaat cgttggtaga tacgttgttg acacttctaa ataagcgaat ttcttatgat 3600
 ttatgatfff tattattaaa taagttataa aaaaaataag tgtatacaaa ttttaaagtg 3660
 actcttaggt tttaaaacga aaattcttgt tcttgagtaa ctctttcctg taggtcaggt 3720
 tgctttctca ggtatagcat gaggtcgctc ttattgacca cacctctacc ggcatgtgctg 3780
 agggacctaa taacttcgta tagcatacat tatacgaagt tatattaagg gttccggatc 3840
 gcggccgctc gacctgcagc caagctagcg gacgtaaact cctcttcaga cctaataact 3900
 tcgtatagca tacattatac gaagttatat taagggttat tgaatatgat cggaattggt 3960
 cgagggacct aataacttcg tatagcatac attatacгаа gttatattaa gggttccgga 4020
 tcgcgccgc tcgacctgca gccaggctag cttggctgga cgtaaactcc tcttcagacc 4080
 taataacttc gatagcatac attatacгаа gttatattaa gggttattga tatgatcgga 4140

ES 2 646 388 A1

attggtcgac tacgtcgta aggccgtttc tgacagagta aaattcttga gggaaactttc 4200
 accattatgg gaaatggttc aagaaggat tgacttaaac tccatcaaat ggtcagggtca 4260
 ttgagtgttt tttatttggt gtatTTTTTT ttttttagag aaaatcctcc aatatcaaat 4320
 taggaatcgt agtttcatga ttttctgtta cacctaactt tttgtgtggt gccctcctcc 4380
 ttgtcaatat taatgttaaa gtgcaattct ttttccttat cacgttgagc cattagtatc 4440
 aatttgctta cctgtattcc tttactatcc tcctTTTTTt ctttcttgat aaatgtatgt 4500
 agattgcgta tatagtttcg tctaccctat gaacatattc cattttgtaa tttcgtgctg 4560
 tttctattat gaatttcatt tataaagttt atgtacaaat atcataaaaa aagagaatct 4620
 ttttaagcaa ggattttctt aacttcttcg gcgacagcat caccgacttc ggtgggtactg 4680
 ttggaaccac ctaaatacacc agttctgata cctgcatcca aaacctTTTT aactgcatct 4740
 tcaatggcct taccttcttc aggcaagttc aatgacaatt tcaacatcat tgcagcagac 4800
 aagatagtgg cgatagggtc aaccttattc tttggcaaat ctggagcaga accgtggcat 4860
 ggttcgtaca aaccaaatgc ggtgttcttg tctggcaaag aggccaagga cgcagatggc 4920
 aacaaacca aggaacctgg gataacggag gcttcatcgg agatgatatc accaaacatg 4980
 ttgctggtga ttataatacc atttaggtgg gttgggttct taactaggat catggcggca 5040
 gaatcaatca attgatgttg aaccttcaat gtaggaaatt cgttcttgat ggtttctctc 5100
 acagtttttc tccataatct tgaagaggcc aaaacattag ctttatccaa ggaccaata 5160
 ggcaatggtg gctcatgttg tagggccatg aaagcggcca ttcttgtgat tctttgact 5220
 tctggaacgg tgtattgttc actatcccaa gcgacaccat caccatcgtc ttcctttctc 5280
 ttaccaaagt aaatacctcc cactaattct ctgacaacaa cgaagtcagt acctttagca 5340
 aattgtggct tgattggaga taagtctaaa agagagtcgg atgcaaagtt acatggtctt 5400
 aagttggcgt acaattgaag ttctttacgg attttttagta aaccttgttc aggtctaaca 5460
 ctaccggtac cccatttagg accaccaca gcacctaaca aaacggcatc aaccttcttg 5520
 gaggttcca gcgcctcatc tggaagtggg acacctgtag catcgatagc agcaccacca 5580
 attaaatgat tttcgaatc gaacttgaca ttggaacgaa catcagaaat agctttaaga 5640
 acctaatgg cttcggctgt gatttcttga ccaacgtggt cacctggcaa aacgacgatc 5700
 ttcttagggg cagacatagg ggcagacatt agaatggtat atccttgaaa tatatatata 5760
 tattgctgaa atgtaaaagg taagaaaagt tagaaagtaa gacgattgct aaccacctat 5820
 tggaaaaaac aataggtcct taaataatat tgtcaacttc aagtattgtg atgcaagcat 5880
 ttagtcatga acgcttctct attctatatg aaaagccggt tccggcctct cacctttcct 5940
 ttttctcca atttttcagt tgaaaaaggt atatgctca ggcgacctct gaaattaaca 6000
 aaaaatttcc agtcatcgaa tttgattctg tgcgatagcg cccctgtgtg ttctcgttat 6060
 gttgaggaag aaaataatgg ttgctaagag attcgaactc ttgcatctta cgatacctga 6120
 gtattccac agttccttac cactcttttg ttactctatt gatccagctc agcaaaggca 6180

ES 2 646 388 A1

gtgtgatcta agattctatc ttcgcatgt agtaaaacta gctagaccga gaaagagact 6240
 agaaatgcaa aaggcacttc tacaatggct gccatcatta ttatccgatg tgacgctgca 6300
 gcttctcaat gatattcgaa tacgctttga ggagatacag cctaatatcc gacaaactgt 6360
 ttacagatt tacgatcgta cttgttacc atcattgaat tttgaacatc cgaacctggg 6420
 agttttccct gaaacagata gtatatttga acctgtataa taatatatag tctagcgctt 6480
 tacggaagac aatgtatgta tttcggttcc tggagaaact attgcatcta ttgcataggt 6540
 aatcttgac gtcgcatccc cggttcattt tctgctttc catcttgac ttcaatagca 6600
 tatctttgtt aacgaagcat ctgtgcttca tttttagaa caaaaatgca acgagagagc 6660
 gctaattttt caaacaaga atctgagctg cttttttaca gaacagaaat gcaacgcgaa 6720
 agcgctattt taccaacgaa gaatctgtgc ttcatttttg taaaacaaaa atgcaacgcg 6780
 agagcgctaa tttttcaaac aaagaatctg agctgcattt ttacagaaca gaaatgcaac 6840
 gcgagagcgc tttttacca acaagaatc tatacttctt ttttgttcta caaaaatgca 6900
 tcccagagc gctatttttc taacaaagca tcttagatta cttttttct cttttgtgcg 6960
 ctctataatg cagtctcttg ataactttt gcaactgtagg tccgttaagg ttagaagaag 7020
 gctactttgg tgtctatttt ctctccata aaaaaagcct gactccactt cccgctttta 7080
 ctgattacta gcgaagctgc ggggtcattt tttcaagata aaggcatccc cgattatatt 7140
 ctataccgat gtggattgcg catactttgt gaacagaaag tgatagcggt gatgattctt 7200
 cattggtcag aaaattatga acggtttctt ctattttgtc tctatatact acgtatagga 7260
 aatgtttaca ttttcgtatt gttttcgatt cactctatga atagttctta ctacaatttt 7320
 tttgtctaaa gagtaatact agagataaac ataaaaaatg tagaggtcga gtttagatgc 7380
 aagttcaagg agcgaaaggt ggatgggtag gttatatagg gatatagcac agagatatat 7440
 agcaaagaga tacttttgag caatgtttgt ggaagcggta ttcgcaatat tttagtagct 7500
 cgttacagtc cgggtcgttt ttggtttttt gaaagtgcgt cttcagagcg cttttggttt 7560
 tcaaaagcgc tctgaagttc ctatactttc tagagaatag gaacttcgga ataggaactt 7620
 caaagcgttt ccgaaaacga gcgcttccga aaatgcaac cgagctgcg acatacagct 7680
 cactgttcac gtcgcaccta tatctgctg ttgcctgtat atatatatac atgagaagaa 7740
 cggcatagtg cgtgtttatg cttaaagcgc tacttatatg cgtctattta tgtaggatga 7800
 aaggtagtct agtacctcct gtgatattat cccattccat gcggggatc gtatgcttcc 7860
 ttcagacta cccttagct gttctatatg ctgccactcc tcaattggat tagtctcatc 7920
 cttcaatgct atcatttctt ttgatattgg atcatatgca tagtaccgag aaactagtgc 7980
 gaagtagtga tcaggattg ctgttatctg atgagtatac gttgtcctgg ccacggcaga 8040
 agcacgctta tcgctcaat ttcccacaac attagtcaac tccgtaggc cttcattga 8100
 aagaaatgag gtcatcaat gtcttcaat gtgagatttt gggccatttt ttatagcaaa 8160
 gattgaataa ggcgcatttt tcttcaagc 8190

ES 2 646 388 A1

<210> 20
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV500
 <400> 20
 gggatcctga tgctggacga cagagccagg 30

<210> 21
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV562
 <400> 21
 gggatccact ggcccagcat acaggccttc 30

<210> 22
 <211> 7001
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Plásmido pRS402-LexA
 <400> 22
 tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
 accatagatc tgaattaatt cttgaataat acataacttt tcttaaaaga atcaaagaca 240
 gataaaattt aagagatatt aatattagt gagaagccga gaattttgta acaccaacat 300
 aacctgaca tctttaaca cttttaatta tgatacatct cttacgtcat gattgattat 360
 tacagctatg ctgacaaatg actcttggtg catggctacg aaccgggtaa tactaagtga 420
 ttgactcttg ctgacctttt attaagaact aatgggaca tattatggag catttcatgt 480
 ataaattggt gcgtaaaatc gttggatctc tcttctaagt acatcctact ataacaatca 540
 agaaaaaca gaaaatcggg caaacaatc aagtatggat tctagaacag ttggtatatt 600
 aggaggggga caattgggac gtatgattgt tgaggcagca aacaggctca acattaagac 660
 ggtaatacta gatgctgaaa attctcctgc caaacaata agcaactcca atgaccacgt 720
 taatggctcc tttccaatc ctcttgatat cgaaaaacta gctgaaaaat gtgatgtgct 780
 aacgattgag attgagcatg ttgatgttcc tactactaaag aatcttcaag taaaacatcc 840
 caaattaaaa atttaccctt ctccagaaac aatcagattg atacaagaca aatatattca 900
 aaaagagcat ttaatcaaaa atggatagc agttaccaa agtgttcctg tggaacaagc 960
 cagtgagacg tccctattga atggtggaag agatttgggt tttcattcg tcttgaagtc 1020
 gaggactttg gcatacgatg gaagaggtaa cttcgttgta aagaataagg aaatgattcc 1080

ES 2 646 388 A1

ggaagctttg gaagtactga aggatcgtcc tttgtacgcc gaaaaatggg caccatttac 1140
 taaagaatta gcagtcatga ttgtgaggtc tgттаacggt ttagtgtttt cttaccaat 1200
 tgtagagact atccacaagg acaatatttg tgacttatgt tatgсgсctg ctagagttcc 1260
 ggactccgtt caacttaagg cgaagttggt ggcagaaaat gcaatcaaat cttttcccg 1320
 ttgtggтata tttggтgтgg ааatgттcta tttagaaaca ggggaattgc ttattaacga 1380
 aattgccccа агgсctcaca actctggaca ttataccatt gatgсttgсg tcactttctca 1440
 atttgaagct catttgagat caatattgga tttgсcaatg ccaaagaatt tcacatcttt 1500
 ctccaccatt аcaacgaacg ccattatgct ааatgттctt ggagacaaac atacaaaaga 1560
 taaagagcta gaaacttgсg ааagagcatt ggcgactcca ggttcctcag tgtacttata 1620
 tggaaaagag tctagaccta acagaaaagt aggtcacata аatattattg cctccagtat 1680
 ggcggaatgt gaacaaaggc tgaactacat tacaggtaga actgatattc caatcaaaat 1740
 ctctgtсgct caaaagttgg acttggaaгc аatggтcaaa ccattggттg gaatcatcat 1800
 gggatcagac tctgacttgс сggтаatgtc tgссgсatgt gсggттttaa аagattттgg 1860
 сgttcattt gaagtгaca tagtctctgс tсatagaact ccacatagga tgtcagcата 1920
 tgctatttcc gcaagcaagc gtggaattaa аacaattatc gctggagctg gtggggctgс 1980
 tcacttgcca ggtatggтgg ctgcaatgac accacttсct gtcatсggтg tgссcgтаaa 2040
 агgttcttgt ctagatggag tagattcttt аattcaatt gtgcaaatgc ctagaggtgt 2100
 tccagtagct accgtсgcta tтаataatag tacgaacgct gсgctgттgg ctgtcagact 2160
 gcttggсgct tatgattcaa gttatacaac gaaaatggaa сagттттtat taaagcaaga 2220
 agaagaagtt ctgtcaaaг cacaaaagtt agaaactgtc ggttacgaag cttatctaga 2280
 аacaagтаa tatataagtt tattgatata ctgtacagc аaataattat аaatgatat 2340
 acctattttt taggctttgt tatgattaca tcaaatgtgg acttcataca tagaaatcaa 2400
 сgcttacagg tgtсcttttt таagaatttc аacataaga tctatгсggт gtgaaatacc 2460
 gcacagatgc gтаaggagaa аataccgсat сaggaaattg таaacgtтаa tattttgtta 2520
 аaattсgсgt таaattттtg tтаaatсagc tсattттtta accaataggc сgaaatсggс 2580
 ааaatссctt атаaatcaaa агаatagacc gagatagggt tgagtгттgt tсcagттtg 2640
 аacaagagtc cactatтаaа gaacgtggac tсcaacgtca аaggгсgaaa аaccgtctat 2700
 сaggгсgatg gсссactacg tgaaccatca ссctaatcaа gттттттggg gtcgaggtgс 2760
 сgтаaagсac таaatсggaa ссctaaaggg агсссссgat ttagagcttg агggggaaag 2820
 ссggсgaaсg tggсgagaaa ggaagggaag ааagсgaaaг gagсggгсgс tagggсgctg 2880
 gcaagtгtag сggтсagсct gсgсgтаacc accacaccg ссgсgctтаa tgсgссgcta 2940
 сaggгсgсgt сgсgсcattc gccattcagg ctgсgсaact gттgggaagg гсgatсggтg 3000
 сggгсctctt сgctattacg ссagctggсg ааaggгggat gtgctgcaag гсgattaagt 3060
 tgggтаacгc сaggгттttc ссagтсagгa сgttgтаaaa сgacггссag tgagсгсгсg 3120
 таatacgact cactataggg сgaattgggt асctттгтт gттtссgggt gtacaaatg 3180

ES 2 646 388 A1

gacttcctct tttctggcaa ccaaaccat acatcgggat tcctataata ccttcgttgg 3240
tctccctaac atgtaggtgg cggaggggag atatacaata gaacagatac cagacaagac 3300
ataatgggct aaacaagact acaccaatta cactgcctca ttgatggtgg tacataacga 3360
actaatactg tagccctaga cttgatagcc atcatcatat cgaagtttca ctaccctttt 3420
tccatttgcc atctattgaa gtaataatag gcgcatgcaa cttcttttct ttttttttct 3480
tttctctctc ccccgttggt gtctcacat atccgcaatg acaaaaaaaaa tgatggaaga 3540
cactaaagga aaaaattaac gacaaagaca gcaccaacag atgtcgttgt tccagagctg 3600
atgaggggta tcttcgaaca cacgaaactt tttccttctc tcattcacgc acactactct 3660
ctaatgagca acggtatacg gccttcctc cagttacttg aatttgaaat aaaaaaagtt 3720
tgccgctttg ctatcaagta taaatagacc tgcaattatt aatcttttgt ttcctcgta 3780
ttgttctcgt tccctttctt ccttgtttct ttttctgcac aatatttcaa gctataccaa 3840
gcatacaatc aactccaagc ttgaattaat tccgggcgga atgaaagcgt taacggccag 3900
gcaacaagag gtgtttgatc tcatccgtga tcacatcagc cagacaggta tgccgccgac 3960
gcgtgcgga atcgcgcagc gtttggggtt ccgttccca aacgcggctg aagaacatct 4020
gaaggcgctg gcacgcaaag gcgttattga aattgtttcc ggcgcatcac gcgggattcg 4080
tctgttgag gaagaggaag aagggttgcc gctggtaggt cgtgtggctg ccggtgaacc 4140
acttctggcg caacagcata ttgaaggta ttatcaggtc gatccttctt tattcaagcc 4200
gaatgctgat ttctgctgc gcgtcagcgg gatgtcagtg aaagatatcg gcattatgga 4260
tggtgacttg ctggcagtc ataaaactca ggatgtacgt aacggtcagg tcgttgtcgc 4320
acgtattgat gacgaagtta ccgttaagcg cctgaaaaaa cagggcaata aagtcgaact 4380
gttgccagaa aatagcgagt ttaaaccaat tgctcgtagat cttcgtcagc agagcttcac 4440
cattgaagg ctggcgggtg gggttattcg caacggcgac tggctggaat tcccggggat 4500
ccgtcgacct gcagccaagc taattccggg cgaatttctt atgatttatg atttttatta 4560
ttaaataagt tataaaaaaa ataagtgtat acaatttta aagtgactct taggttttaa 4620
aacgaaaatt cttgttcttg agtaactctt tcctgtaggt caggttgctt tctcaggtat 4680
agcatgaggt cgctcttatt gaccacacct ctaccggcat gcgagctcca gcttttgttc 4740
cctttagtga gggttaattg cgcgcttggc gtaatcatgg tcatagctgt ttctgtgtg 4800
aaattgttat ccgctcaca ttccacaca catacgagcc ggaagcataa agtgtaaagc 4860
ctggggtgcc taatgagtga gctaactcac attaattgct ttgcgctcac tgcccgttt 4920
ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg 4980
cggtttgctg attgggcgct cttccgctt ctcgctcact gactcgctgc gctcggctg 5040
tcggctgctg cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc 5100
aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa 5160
aaaggccgcg ttgctggcgt tttccatag gctccgccc cctgacgagc atcacaaaa 5220

ES 2 646 388 A1

tcgacgctca agtcagaggt ggcgaaacct gacaggacta taaagatacc aggcgtttcc 5280
 ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc 5340
 cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag 5400
 ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaacccccg ttcagcccga 5460
 ccgctgcgcc ttatccggtg actatcgtct tgagtccaac ccggtaagac acgacttatac 5520
 gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac 5580
 agagtcttg aagtgggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg 5640
 cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca 5700
 aaccaccgct ggtagcgggt gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa 5760
 aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa 5820
 ctcacgttaa gggatttttg tcatgagatt atcaaaaagg atcttcacct agatcctttt 5880
 aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag 5940
 ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat 6000
 agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc 6060
 cagtgctgca atgataccgc gagaccacg ctcaccggct ccagatttat cagcaataaa 6120
 ccagccagcc ggaagggccg agcgcagaag tggctctgca actttatccg cctccatcca 6180
 gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa 6240
 cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggtg tggcttcatt 6300
 cagctccggt tccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc 6360
 ggtagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact 6420
 catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc 6480
 tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg 6540
 ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgtc 6600
 catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc 6660
 cagttcgatg taaccactc gtgcaccaa ctgatcttca gcatctttta ctttcaccag 6720
 cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac 6780
 acggaaatgt tgaatactca tactcttctc ttttcaatat tattgaagca tttatcaggg 6840
 ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac aaataggggt 6900
 tccgcgaca tttccccgaa aagtccacc tgacgtctaa gaaaccatta ttatcatgac 6960
 attaacctat aaaaataggc gtatcacgag gccctttcgt c 7001

<210> 23
 <211> 606
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Secuencia que codifica para el dominio de unión al ADN del represor transcripcional LexA

ES 2 646 388 A1

<220>
 <221> CDS
 <222> (1)..(606)

<400> 23

atg	aaa	gcg	tta	acg	gcc	agg	caa	caa	gag	gtg	ttt	gat	ctc	atc	cgt		48
Met	Lys	Ala	Leu	Thr	Ala	Arg	Gln	Gln	Glu	Val	Phe	Asp	Leu	Ile	Arg		
1				5					10					15			
gat	cac	atc	agc	cag	aca	ggt	atg	ccg	ccg	acg	cgt	gcg	gaa	atc	gcg		96
Asp	His	Ile	Ser	Gln	Thr	Gly	Met	Pro	Pro	Thr	Arg	Ala	Glu	Ile	Ala		
			20					25					30				
cag	cgt	ttg	ggg	ttc	cgt	tcc	cca	aac	gcg	gct	gaa	gaa	cat	ctg	aag		144
Gln	Arg	Leu	Gly	Phe	Arg	Ser	Pro	Asn	Ala	Ala	Glu	Glu	His	Leu	Lys		
		35					40					45					
gcg	ctg	gca	cgc	aaa	ggc	gtt	att	gaa	att	gtt	tcc	ggc	gca	tca	cgc		192
Ala	Leu	Ala	Arg	Lys	Gly	Val	Ile	Glu	Ile	Val	Ser	Gly	Ala	Ser	Arg		
	50					55					60						
ggg	att	cgt	ctg	ttg	cag	gaa	gag	gaa	gaa	ggg	ttg	ccg	ctg	gta	ggt		240
Gly	Ile	Arg	Leu	Leu	Gln	Glu	Glu	Glu	Glu	Gly	Leu	Pro	Leu	Val	Gly		
65					70					75					80		
cgt	gtg	gct	gcc	ggt	gaa	cca	ctt	ctg	gcg	caa	cag	cat	att	gaa	ggt		288
Arg	Val	Ala	Ala	Gly	Glu	Pro	Leu	Leu	Ala	Gln	Gln	His	Ile	Glu	Gly		
				85					90					95			
cat	tat	cag	gtc	gat	cct	tcc	tta	ttc	aag	ccg	aat	gct	gat	ttc	ctg		336
His	Tyr	Gln	Val	Asp	Pro	Ser	Leu	Phe	Lys	Pro	Asn	Ala	Asp	Phe	Leu		
			100					105					110				
ctg	cgc	gtc	agc	ggg	atg	tcg	atg	aaa	gat	atc	ggc	att	atg	gat	ggt		384
Leu	Arg	Val	Ser	Gly	Met	Ser	Met	Lys	Asp	Ile	Gly	Ile	Met	Asp	Gly		
		115					120					125					
gac	ttg	ctg	gca	gtg	cat	aaa	act	cag	gat	gta	cgt	aac	ggt	cag	gtc		432
Asp	Leu	Leu	Ala	Val	His	Lys	Thr	Gln	Asp	Val	Arg	Asn	Gly	Gln	Val		
	130					135					140						
gtt	gtc	gca	cgt	att	gat	gac	gaa	gtt	acc	gtt	aag	cgc	ctg	aaa	aaa		480
Val	Val	Ala	Arg	Ile	Asp	Asp	Glu	Val	Thr	Val	Lys	Arg	Leu	Lys	Lys		
	145				150					155					160		
cag	ggc	aat	aaa	gtc	gaa	ctg	ttg	cca	gaa	aat	agc	gag	ttt	aaa	cca		528
Gln	Gly	Asn	Lys	Val	Glu	Leu	Leu	Pro	Glu	Asn	Ser	Glu	Phe	Lys	Pro		
				165					170					175			
att	gtc	gta	gat	ctt	cgt	cag	cag	agc	ttc	acc	att	gaa	ggg	ctg	gcg		576
Ile	Val	Val	Asp	Leu	Arg	Gln	Gln	Ser	Phe	Thr	Ile	Glu	Gly	Leu	Ala		
			180					185					190				
gtt	ggg	gtt	att	cgc	aac	ggc	gac	tgg	ctg								606
Val	Gly	Val	Ile	Arg	Asn	Gly	Asp	Trp	Leu								
		195					200										

<210> 24
 <211> 202
 <212> PRT
 <213> Artificial sequence

<220>
 <223> Synthetic Construct

<400> 24

ES 2 646 388 A1

Met Lys Ala Leu Thr Ala Arg Gln Gln Glu Val Phe Asp Leu Ile Arg
 1 5 10 15

Asp His Ile Ser Gln Thr Gly Met Pro Thr Arg Ala Glu Ile Ala
 20 25 30

Gln Arg Leu Gly Phe Arg Ser Pro Asn Ala Ala Glu Glu His Leu Lys
 35 40 45

Ala Leu Ala Arg Lys Gly Val Ile Glu Ile Val Ser Gly Ala Ser Arg
 50 55 60

Gly Ile Arg Leu Leu Gln Glu Glu Glu Glu Gly Leu Pro Leu Val Gly
 65 70 75 80

Arg Val Ala Ala Gly Glu Pro Leu Leu Ala Gln Gln His Ile Glu Gly
 85 90 95

His Tyr Gln Val Asp Pro Ser Leu Phe Lys Pro Asn Ala Asp Phe Leu
 100 105 110

Leu Arg Val Ser Gly Met Ser Met Lys Asp Ile Gly Ile Met Asp Gly
 115 120 125

Asp Leu Leu Ala Val His Lys Thr Gln Asp Val Arg Asn Gly Gln Val
 130 135 140

Val Val Ala Arg Ile Asp Asp Glu Val Thr Val Lys Arg Leu Lys Lys
 145 150 155 160

Gln Gly Asn Lys Val Glu Leu Leu Pro Glu Asn Ser Glu Phe Lys Pro
 165 170 175

Ile Val Val Asp Leu Arg Gln Gln Ser Phe Thr Ile Glu Gly Leu Ala
 180 185 190

Val Gly Val Ile Arg Asn Gly Asp Trp Leu
 195 200

<210> 25
 <211> 703
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Promotor de ADH1 en el plásmido pLexA(1-202)PL

<400> 25
 ccttttggtg tttccgggtg tacaatatgg acttcctctt ttctggcaac caaacccata 60
 catcgggatt cctataatac cttcgttggt ctccctaaca tgtaggtggc ggaggggaga 120
 tatacaatag aacagatacc agacaagaca taatgggcta aacaagacta caccaattac 180
 actgcctcat tgatggtggt acataacgaa ctaatactgt agccctagac ttgatagcca 240

ES 2 646 388 A1

tcatcatatc gaagtttcac tacccttttt ccatttgcca tctattgaag taataatagg 300
 cgcatgcaac ttcttttctt tttttttctt ttctctctcc cccgttggtg tctcaccata 360
 tccgcaatga caaaaaaaaa gatggaagac actaaaggaa aaaattaacg acaaagacag 420
 caccaacaga tgtcgttggt ccagagctga tgaggggtat cttcgaacac acgaaacttt 480
 ttcttctctt cattcagca cactactctc taatgagcaa cggtatacgg ctttcttcc 540
 agttacttga atttgaaata aaaaaagttt gccgctttgc tatcaagtat aaatagacct 600
 gcaattatta atcttttggt tcctcgtcat tgttctcgtt ccctttcttc cttgtttctt 660
 tttctgcaca atatttcaag ctataccaag catacaatca act 703

<210> 26
 <211> 193
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Terminador de ADH1 en el plásmido pLexA(1-202)PL

<400> 26
 gcgaatttct tatgatttat gatttttatt attaaataag ttataaaaa aataagtgta 60
 taaaaattt aaagtgactc ttaggtttta aaacgaaaat tcttgttctt gagtaactct 120
 ttctgtagg tcaggttgct ttctcaggta tagcatgagg tcgctcttat tgaccacacc 180
 tctaccggca tgc 193

<210> 27
 <211> 10144
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Plásmido pLexA(1-200)PL

<400> 27
 agctgcatgt gtcagagggt ttcaccgtca tcaccgaaac gcgcgaggca ggatgatccg 60
 ggatcgaaga aatgatggta aatgaaatag gaaatcaagg agcatgaagg caaaagacaa 120
 atataagggt cgaacgaaaa ataaagtga aagtgttgat atgatgtatt tggctttgcg 180
 gcgccgaaaa aacgagttta cgcaattgca caatcatgct gactctgtgg cggacccgcg 240
 ctcttgccgg cccggcgata acgctgggcg tgaggctgtg cccggcggag tttttgcbc 300
 ctgcattttc caaggtttac cctgcgctaa ggggagat tggagaagca ataagaatgc 360
 cggttggggt tgcgatgatg acgaccacga caactggtgt cattatttaa gttgccgaaa 420
 gaacctgagt gcatgtgcaa catgagtata ctagaagaat gagccaagac ttgagagacg 480
 cgagtttgcc ggtggtgca acaatagagc gaccatgacc ttgaagggtga gacgagcaca 540
 accgctagag tactttgaag aggaaacagc aataggggtg ctaccagtat aaatagacag 600
 gtacatacaa cactggaaat ggttgtctgt ttgagtacgc tttcaattca tttgggtgtg 660
 cactttatta tgttacaata tgaaggaa ctttacactt ctctatgca catatattaa 720

ES 2 646 388 A1

ttaaagtcca atgctagtag agaagggggg taacaccctt ccgcgctctt ttccgatttt 780
 tttctaaacc gtggaatatt tcggatatcc ttttgttggt tccgggtgta caatatggac 840
 ttctcttttt ctggcaacca aaccataca tcgggattcc tataatacct tcgttggtct 900
 ccctaacatg taggtggcgg aggggagata tacaatagaa cagataccag acaagacata 960
 atgggctaaa caagactaca ccaattacac tgcctcattg atggtggtac ataacgaact 1020
 aatactgtag ccctagactt gatagccatc atcatatcga agtttcacta ccctttttcc 1080
 atttgccatc tattgaagta ataataggcg catgcaactt cttttctttt tttttctttt 1140
 ctctctcccc cgttgttgtc tcaccatata cgcaatgaca aaaaaaatga tggaagacac 1200
 taaaggaaaa aattaacgac aaagacagca ccaacagatg tcgttgttcc agagctgatg 1260
 aggggtatct tcgaacacac gaaacttttt ctttccttca ttcacgcaca ctactctcta 1320
 atgagcaacg gtatacggcc ttccttccag ttacttgaat ttgaaataaa aaaagtttgc 1380
 cgctttgcta tcaagtataa atagacctgc aattattaat cttttgtttc ctcgctcattg 1440
 ttctcgttcc ctttcttctt tgtttctttt tctgcacaat atttcaagct ataccaagca 1500
 tacaatcaac tccaagcttg aattaattcc gggcggaatg aaagcgtaa cggccaggca 1560
 acaagaggtg tttgatctca tccgtgatca catcagccag acaggtatgc cgccgacgcg 1620
 tgcggaaatc gcgcagcgtt tggggttccg ttccccaac gcggctgaag aacatctgaa 1680
 ggcgctggca cgcaaaggcg ttattgaaat tgtttccggc gcatcacgcg ggattcgtct 1740
 gttgcaggaa gaggaagaag ggttgccgct ggtaggtcgt gtggctgccg gtgaaccact 1800
 tctggcgcaa cagcatattg aaggtcatta tcaggtcgat ctttccttat tcaagccgaa 1860
 tgctgatttc ctgctgcgcg tcagcgggat gtcgatgaaa gatatcggca ttatggatgg 1920
 tgacttgctg gcagtgcata aaactcagga tgtacgtaac ggtcaggtcg ttgtcgcacg 1980
 tattgatgac gaagttaccg ttaagcgcct gaaaaaacag ggcaataaag tcgaactggt 2040
 gccagaaaat agcgagttta aaccaattgt cgtagatctt cgtcagcaga gcttcaccat 2100
 tgaagggctg gcggttgggg ttattcgcaa cggcgactgg ctggaattcc cggggatccg 2160
 tcgacctgca gccaaagctaa ttccgggcga atttcttatg atttatgatt tttattatta 2220
 aataagttat aaaaaaata agtgtataca aattttaaag tgactcttag gttttaaac 2280
 gaaaattctt gttcttgagt aactctttcc tgtaggtcag gttgctttct caggtatagc 2340
 atgaggtcgc tcttattgac cacacctcta ccggcatgcc gagcaaatgc ctgcaaatcg 2400
 ctccccattt cacccaattg tagatatgct aactccagca atgagttgat gaatctcggg 2460
 gtgtatttta tgcctcaga ggacaacacc tgttgtaatc gttcttccac acggatcgat 2520
 ccacaggacg ggtgtggtcg ccatgatcgc gtagtcgata gtggctccaa gtagcgaagc 2580
 gagcaggact gggcggcggc caaagcggtc ggacagtgct ccgagaacgg gtgcgcatag 2640
 aaattgcatc aacgcatata gcgctagcag cacgccatag tgactggcga tgctgtcggg 2700
 atggacgata tcccgcaaga ggcccggcag taccggcata accaagccta tgcctacagc 2760
 atccaggggtg acggtgccga ggatgacgat gagcgcattg ttagatttca tacacggtgc 2820

ES 2 646 388 A1

ctgactgcgt tagcaattta actgtgataa actaccgcat taaagctagc tttgaagaaa 2880
aatgcgccctt attcaatctt tgctataaaa aatggcccaa aatctcacat tgggaagacat 2940
ttgatgacct catttctttc aatgaagggc ctaacggagt tgactaatgt tgtgggaaat 3000
tggagcgata agcgtgcttc tgccgtggcc aggacaacgt atactcatca gataacagca 3060
atacctgatc actacttcgc actagtttct cggctactatg catatgatcc aatatcaaag 3120
gaaatgatag cattgaagga tgagactaat ccaattgagg agtggcagca tatagaacag 3180
ctaaagggta gtgctgaagg aagcatacga taccgccat ggaatgggat aatatcacag 3240
gaggtactag actaccttc atcctacata aatagacgca tataagtacg catttaagca 3300
taaacacgca ctatgccgtt cttctcatgt atatatatat acaggcaaca cgcagatata 3360
ggtgcgacgt gaacagtgag ctgtatgtgc gcagctcgcg ttgcattttc ggaagcgctc 3420
gttttcggaa acgctttgaa gttcctattc cgaagttcct attctctaga aagtatagga 3480
acttcagagc gcttttgaaa accaaaagcg ctctgaagac gcactttcaa aaaacaaaaa 3540
acgcaccgga ctgtaacgag ctactaaaat attgcaata ccgcttcac aaacattgct 3600
caaaagtatc tctttgctat atatctctgt gctatatccc tatataacct acccatccac 3660
ctttcgctcc ttgaacttgc atctaaactc gacctctaca ttttttatgt ttatctctag 3720
tattactctt tagacaaaa aattgtagta agaactattc atagagtgaa tcgaaaacaa 3780
tacgaaaatg taaacatttc ctatacgtag tatatagaga caaaatagaa gaaaccgttc 3840
ataattttct gaccaatgaa gaatcatcaa cgctatcact ttctgttcac aaagtatgcg 3900
caatccacat cggtagagaa tataatcggg gatgccttta tcttgaaaa atgcacccgc 3960
agcttcgcta gtaatcagta aacgcgggaa gtggagtcag gcttttttta tgggaagagaa 4020
aatagacacc aaagtagcct tcttctaacc ttaacggacc tacagtgcaa aaagttatca 4080
agagactgca ttatagagcg cacaaaggag aaaaaagta atctaagatg ctttgttaga 4140
aaaatagcgc tctcgggatg ctttttgta gaacaaaaa gaagtataga ttctttgttg 4200
gtaaaatagc gctctcgcgt tgcatctctg ttctgtaaaa atgcagctca gattctttgt 4260
ttgaaaaatt agcgcctctg cgttgcattt ttgttttaca aaaatgaagc acagattctt 4320
cgttggtaaa atagcgcttt cgcggtgcat ttctgttctg taaaaatgca gctcagattc 4380
tttgtttgaa aaattagcgc tctcgcgttg catttttggt ctacaaaatg aagcacagat 4440
gcttcgttaa caaagatatg ctattgaagt gcaagatgga aacgcagaaa atgaaccggg 4500
gatgcgacgt gcaagattac ctatgcaata gatgcaatag tttctccagg aaccgaaata 4560
catacattgt cttccgtaaa gcgctagact atatattatt atacaggttc aaatatacta 4620
tctgtttcag ggaaaactcc caggttcggg tgttcaaaat tcaatgatgg gtaacaagta 4680
cgatcgtaaa tctgtaaaac agtttgctcg atattaggct gtatctcctc aaagcgtatt 4740
cgaatatcat tgagaagctg cagcaggcgt gaagttagac gacaacttct ctctggaaac 4800
gcataccgat attcaggctg ctgcaaaggc acaggctagt gcccggtcga gtgcatccgg 4860

ES 2 646 388 A1

taccaccca gatgctgtag tagcttctgg tagcactgca atgagccatg cttatcaaga 4920
 aaacacaggt tttggtactc gtcccatata tcttgacatg caagccacta caccaacaga 4980
 ccctaggggt ttggatacga tgttgaagtt ttatacggga ctttatggta atcctcattc 5040
 caacactcac tcttacgggt gggaaacaaa tactgctgtg gaaaatgcta gagctcacgt 5100
 agcaaagatg atcaatgccg accccaagga aataatattc acttcgggag cgaccgaatc 5160
 taataatatg gttcttaagg gtgtccaag attttataag aagactaaga aacacatcat 5220
 caccactaga acggaacaca agtgtgtctt ggaagccgca cgggccatga tgaaggaggg 5280
 atttgaagtc actttcctaa atgtggacga tcaaggtctt atcgatttga aggaattgga 5340
 agatgccatt agaccagata cctgtctcgt ctctgtgatg gctgtcaata atgaaatcgg 5400
 tgtcattcaa cctattaaag aaattggagc aattttaga aagaataaga tcctcgggga 5460
 caccaaata ggcgatctcg gccttttctg ttcttggagc tgggacatgt ttgccatcga 5520
 tccatctacc accagaacgg ccgtagatc tgctgccacc gttgtttcca ccgaagaaac 5580
 caccgttgcc gtaaccacca cgacggttgt tgctaaagaa gctgccaccg ccacggccac 5640
 cgttgtagcc gccgttggtt ttattgtagt tgctactggt atttctggca cttcttggtt 5700
 ttctcttaa gtgaggagga acataacat tctcgttgtt gtcgttgatg cttaaatfff 5760
 gcacttgttc gctcagttca gccataatat gaaatgcttt tcttgttgtt cttacggaat 5820
 accacttgcc acctatcacc acaactaact ttttcccgtt cctccatctc ttttatatff 5880
 tttttctcga tcgagttcaa gagaaaaaa aagaaaaagc aaaaagaaaa aaggaaagcg 5940
 cgcctcgttc agaatgacac gtatagaatg atgcattacc ttgtcatctt cagtatcata 6000
 ctgttcgat acatacttac tgacattcat aggtatacat atatacacat gtatatatat 6060
 cgtatgctgc agctttaa atcgggtgct actacataag aacacctttg gtggagggaa 6120
 catcgttggt accattgggc gaggtggctt ctcttatggc aaccgcaaga gccttgaacg 6180
 cactctcact acggtgatga tcattcttgc ctgcagaca atcaacgtgg agggtaattc 6240
 tgctagcctc tgcaaagctt tcaagaaaat gcgggatcat ctgcaagag agatctccta 6300
 ctttctccct ttgcaaacca agttcgaca ctgcgtacgg cctgttcgaa agatctacca 6360
 ccgctctgga aagtgcctca tccaaaggcg caaatcctga tccaaacctt tttactccac 6420
 gcgccagtag ggcctcttta aaagcttgac cgagagcaat cccgcagtct tcagtggtgt 6480
 gatggtcgtc tatgtgtaag tcaccaatgc actcaacgat tagcgaccag ccggaatgct 6540
 tggccagagc atgtatcata tggccagaa accctatacc tgtgtggacg ttaatcactt 6600
 gcgattgtgt ggcctgttct gctactgctt ctgcctcttt ttctgggaag atcgagtgt 6660
 ctatcgctag gggaccacc tttaaagaga tcgcaatctg aatcttgggt tcatttgtaa 6720
 tacgctttac tagggctttc tgctctgtca tctttgcctt cgtttatctt gcctgctcat 6780
 tttttagtat attcttcgaa gaaatcacat tactttatat aatgtataat tcattatgtg 6840
 ataatgcaa tcgctaagaa aaaaaagag tcatccgcta ggtggaaaaa aaaaaatgaa 6900
 aatcattacc gaggcataaa aaaatataga gtgtactaga ggaggccaag agtaatagaa 6960

ES 2 646 388 A1

aaagaaaatt gcgggaaagg actgtgttat gacttccctg actaatgccg tgttcaaacg 7020
atacctggca gtgactccta gcgctcacca agctcttaaa acgagaatta agaaaaagtc 7080
gtcatctttc gataagtttt tcccacagca aagcaatagt agaaaaacaa tgggaaacgt 7140
tgaatgaaga caaagcgtcg tggtttaaaa ggaaatacgc tcacgtacat gctagggaac 7200
aggaccgtgc agcggatcta atgaatccat ttgttagtta atagtttaaa tgtttttatc 7260
ggaagagggt ttgtcatcac atcagcaatg ttcttcttgg tctcgatgta gtatacgtat 7320
aaattattac ctgatacttc atctctaagt ctattgcct ttgtgcaaaa aaatctgttt 7380
ctaaatttct cttcatttgt agacttaatt atactgatcg ttgatctact atcagtaagt 7440
aagcctttaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa ctgtaacaat agcaataccc 7500
caaataccta atgtagtcc agcaagcaag ctaaaaagta aagcaacaac ataactcacc 7560
cctgcatctg cagcttttgc ccgggcagcc tgctctgcct gtgttttctt taattgagca 7620
gtagaccatt tagcagttgc atgaatagct gcagcgtcac atcggataat aatgatggca 7680
gccattgtag aagtgccttt tgcatttcta gtctctttct cggcttagct agttttacta 7740
catcgcgaag atagaatctt agatcacact gcctttgctg agctggatca atagagtaac 7800
aaaagagtgg taaggcctcg ttaaaggaca aggacctgag cggaagtgta tcgtacagta 7860
gacggagtat actagtatag tctatagtcc gtggaattaa ttcttgaaga cgaaagggcc 7920
tcgtgatacg cctatthtta taggttaatg tcatgataat aatggtttct tagacgtcag 7980
gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatthttc taaatacatt 8040
caaataatgta tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa 8100
ggaagagtat gagtattcaa catttccgtg tcgcccttat tccctthttt gcggcatttt 8160
gccttctgt ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt 8220
tgggtgcacg agtgggttac atcgaactgg atctcaacag cggttaagatc cttgagagtt 8280
ttcgccccga agaacgtttt ccaatgatga gcactthtaa agttctgcta tgtggcgcgg 8340
tattatcccg tgttgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga 8400
atgacttggg tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 8460
gagaattatg cagtgtgccc ataaccatga gtgataacac tgcggccaac ttacttctga 8520
caacgatcgg aggaccgaag gagctaaccg cttthttgca caacatgggg gatcatgtaa 8580
ctcgccttga tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 8640
ccacgatgcc tgcagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta 8700
cttagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 8760
ttctgcgctc ggcccttccg gctggctggg ttattgctga taaatctgga gccggtgagc 8820
gtgggtctcg cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag 8880
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga 8940
taggtgcctc actgattaag cattggtaac tgtcagacca agtttactca tatatacttt 9000

ES 2 646 388 A1

agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 9060
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag 9120
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa 9180
caaaaaaacc accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt 9240
ttccgaaggt aactggcttc agcagagcgc agatacctaaa tactgtcctt ctagtgtagc 9300
cgtagttagg ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa 9360
tcctgttacc agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa 9420
gacgatagtt accggataag gcgagcggg cgggctgaac ggggggttcg tgcacacagc 9480
ccagcttggg gcgaacgacc tacaccgaac tgagatacct acagcgtgag ctatgagaaa 9540
gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggctcgaa 9600
caggagagcg cacgaggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg 9660
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 9720
tatggaaaaa cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg 9780
ctcacatggt ctttctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg 9840
agtgagctga taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg 9900
aagcgggaaga ggcctgatg cggtatcttc tccttacgca tctgtgcggg atttcacacc 9960
gcatatggtg cactctcagt acaatctgct ctgatgccgc atagttaagc cagtatacac 10020
tccgctatcg ctacgtgact gggcatggc tgcgccccga caccgcca caccgctga 10080
cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg tgaccgtctc 10140
cggg 10144

<210> 28
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Cebador OV502

<400> 28
gggtaccttt tgttgtttcc ggggtgtac 28

<210> 29
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Cebador OV503

<400> 29
ggagctcgca tgccggtaga ggtgtggtc 29

<210> 30
<211> 8182
<212> DNA
<213> Artificial Sequence

ES 2 646 388 A1

<220>

<223> Plásmido pRS402-LexA-Tsg101

<400> 30

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accatagatc	tgaattaatt	cttgaataat	acataacttt	tcttaaaaga	atcaaagaca	240
gataaaattt	aagagatatt	aaatattagt	gagaagccga	gaattttgta	acaccaacat	300
aacactgaca	tctttaacaa	cttttaatta	tgatacattt	cttacgtcat	gattgattat	360
tacagctatg	ctgacaaatg	actcttgttg	catggctacg	aaccgggtaa	tactaagtga	420
ttgactcttg	ctgacctttt	attaagaact	aaatggacaa	tattatggag	catttcatgt	480
ataaattggt	gcgtaaaatc	gttggatctc	tcttctaagt	acatcctact	ataacaatca	540
agaaaaacaa	gaaaatcggg	caaaacaatc	aagtatggat	tctagaacag	ttggtatatt	600
aggaggggga	caattgggac	gtatgattgt	tgaggcagca	aacaggctca	acattaagac	660
ggtaataacta	gatgctgaaa	attctcctgc	caaacaataa	agcaactcca	atgaccacgt	720
taatggctcc	ttttccaatc	ctcttgatat	cgaaaaacta	gctgaaaaat	gtgatgtgct	780
aacgattgag	attgagcatg	ttgatgttcc	tactactaaag	aatcttcaag	taaaacatcc	840
caaattaaaa	atctaccctt	ctccagaaac	aatcagattg	atacaagaca	aatatattca	900
aaaagagcat	ttaatcaaaa	atggtatagc	agttacccaa	agtgttcctg	tggaacaagc	960
cagtgagacg	tccctattga	atggttgaag	agattttgggt	tttcattcgc	tcttgaagtc	1020
gaggactttg	gcatacgatg	gaagaggtaa	cttcgttgta	aagaataagg	aaatgattcc	1080
ggaagctttg	gaagtactga	aggatcgtcc	tttgtacgcc	gaaaaatggg	caccatttac	1140
taaagaatta	gcagtcatga	ttgtgaggtc	tgtaaacggt	ttagtgtttt	cttacccaat	1200
tgtagagact	atccacaagg	acaatatttg	tgacttatgt	tatgcgctcg	ctagagttcc	1260
ggactccggt	caacttaagg	cgaagttggt	ggcagaaaat	gcaatcaaat	cttttcccgg	1320
ttgtggtata	tttgggtggt	aaatgttcta	tttagaaaca	ggggaattgc	ttattaacga	1380
aattgcccc	aggcctcaca	actctggaca	ttataccatt	gatgcttgcg	tcacttctca	1440
atgtgaagct	catttgagat	caatattgga	tttgccaatg	ccaaagaatt	tcacatcttt	1500
ctccaccatt	acaacgaacg	ccattatgct	aaatgttctt	ggagacaaac	atacaaaaga	1560
taaagagcta	gaaacttgcg	aaagagcatt	ggcgactcca	ggttcctcag	tgtacttata	1620
tggaaaagag	tctagaccta	acagaaaagt	aggtcacata	aatattattg	cctccagtat	1680
ggcggaatgt	gaacaaaggc	tgaactacat	tacaggtaga	actgatattc	caatcaaaat	1740
ctctgtcgct	caaaagttgg	acttggagc	aatgggtcaa	ccattggttg	gaatcatcat	1800
gggatcagac	tctgacttgc	cggtaatgtc	tgccgcatgt	gcggttttaa	aagattttgg	1860
cgttccattt	gaagtgacaa	tagtctctgc	tcatagaact	ccacatagga	tgtcagcata	1920

ES 2 646 388 A1

tgctatttcc gcaagcaagc gtggaattaa aacaattatc gctggagctg gtggggctgc 1980
 tcacttgcca ggtatggtgg ctgcaatgac accacttcct gtcacggtg tgcccgtaaa 2040
 aggttcttgt ctagatggag tagattcttt acattcaatt gtgcaaatgc ctagaggtgt 2100
 tccagtagct accgtcgcta ttaataatag tacgaacgct gcgctgttgg ctgtcagact 2160
 gcttggcgct tatgattcaa gttatacaac gaaaatggaa cagtttttat taaagcaaga 2220
 agaagaagtt cttgtcaaag cacaaaagtt agaaactgtc ggttacgaag cttatctaga 2280
 aaacaagtaa tatataagtt tattgatata cttgtacagc aaataattat aaaatgatat 2340
 acctatTTTT taggctttgt tatgattaca tcaaatgtgg acttcataca tagaaatcaa 2400
 cgcttacagg tgtccttttt taagaatttc atacataaga tctatgCGGT gtgaaatacc 2460
 gcacagatgc gtaaggagaa aataccgcat caggaaattg taaacgttaa tattttgtta 2520
 aaattcgcgt taaatTTTTg ttaaactcagc tcatttttta accaataggc cgaaatcggc 2580
 aaaatccctt ataaatcaaa agaatagacc gagatagggT tgagtgttgt tccagtttgg 2640
 aacaagagtc cactatataa gaacgtggac tccaacgtca aagggcgaaa aaccgtctat 2700
 cagggcgatg gccactacg tgaaccatca ccctaataca gttttttggg gtcgaggtgc 2760
 cgtaaagcac taaatcggaa ccctaaaggg agccccgat ttagagcttg acggggaaag 2820
 ccggcgaacg tggcgagaaa ggaagggaaG aaagcgaaag gagcgggCGC tagggcgctg 2880
 gcaagtgtag cggtcacgct gcgcgtaacc accacaccCG cgcgcttaa tgcgccgcta 2940
 cagggcgcgt cgcgccattc gccattcagg ctgcgcaact gttgggaagg gcgatcgggtg 3000
 cgggcctctt cgctattacg ccagctggcg aaagggggat gtgctgcaag gcgattaagt 3060
 tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa cgacggccag tgagcgcgcg 3120
 taatacgact cactataggg cgaattgggt accttttgtt gtttccgggt gtacaatatg 3180
 gacttcctct tttctggcaa ccaaaccat acatcgggat tcctataata cttcgttgg 3240
 tctccctaac atgtaggtgg cggagggggag atatacaata gaacagatac cagacaagac 3300
 ataatgggct aaacaagact acaccaatta cactgcctca ttgatggtgg tacataacga 3360
 actaactctg tagccctaga cttgatagcc atcatcatat cgaagtttca ctaccctttt 3420
 tccatttgcc atctattgaa gtaataatag gcgcatgcaa cttcttttct tttttttct 3480
 tttctctctc ccccgttggt gtctcacat atccgcaatg acaaaaaaaaa tgatggaaga 3540
 cactaaagga aaaaattaac gacaaagaca gcaccaacag atgtcgttgt tccagagctg 3600
 atgaggggta tcttgaaca cacgaaactt tttccttctc tcattcacgc aactactct 3660
 ctaatgagca acggatagc gccttcctc cagttacttg aatttgaaat aaaaaagtt 3720
 tgccgctttg ctatcaagta taaatagacc tgcaattatt aatcttttgt ttcctcgta 3780
 ttgttctcgt tccctttctt cttgtttct ttttctgcac aatatttcaa gctataccaa 3840
 gcatacaatc aactccaagc ttgaattaat tccgggcgga atgaaagcgt taacggccag 3900
 gcaacaagag gtgtttgatc tcatccgtga tcacatcagc cagacaggta tgccgccgac 3960
 gcgtgcggaa atcgcgcagc gtttgggggt cgttcccca aacgcggctg aagaacatct 4020

ES 2 646 388 A1

gaaggcgctg gcacgcaaag gcgttattga aattgtttcc ggcgcatcac gcgggattcg 4080
 tctgttgag gaagaggaag aagggttgcc gctggtaggc cgtgtggctg ccggtgaacc 4140
 acttctggcg caacagcata ttgaaggcca ttatcaggtc gatccttcct tattcaagcc 4200
 gaatgctgat ttctgctgc gcgtcagcgg gatgtcagtg aaagatatcg gcattatgga 4260
 tgggtgactg ctggcagtc ataaaactca ggatgtacgt aacggtcagg tcgtttgctgc 4320
 acgtattgat gacgaagtta ccgttaagcg cctgaaaaaa cagggcaata aagtcgaact 4380
 gttgccagaa aatagcgagt ttaaaccaat tgctgtagat cttcgtcagc agagcttcac 4440
 cattgaaggg ctggcgggtg gggttattcg caacggcgac tggctggaat tcccggggat 4500
 cctcatggcg gtgtcggaga gccagctcaa gaaaatggtg tccaagtaca aatacagaga 4560
 cctaactgta cgtgaaactg tcaatgttat tactctatac aaagatctca aacctgtttt 4620
 ggattcatat gtttttaacg atggcagttc cagggaaacta atgaacctca ctggaacaat 4680
 ccctgtgcct tatagaggta atacatacaa tattccaata tgcctatggc tactggacac 4740
 ataccatata aatcccccta tctgttttgt taagcctact agttcaatga ctattaaaac 4800
 aggaaagcat gttgatgcaa atgggaagat atatcttcct tatctacatg aatggaaaca 4860
 cccacagtca gacttgttgg ggcttattca ggtcatgatt gtggtatttg gagatgaacc 4920
 tccagtcttc tctcgtccta tttcggcatc ctatccgcca taccaggcaa cggggccacc 4980
 aaatacttcc tacatgccag gcatgccagg tggaatctct ccatacccat ccggataccc 5040
 tcccaatccc agtggttacc caggctgtcc ttaccacact ggtggtccat atcctgccac 5100
 aacaagttct cagtaccctt ctcagcctcc tgtgaccact gttggtcca gtagggatgg 5160
 cacaatcagc gaggacacca tccgagcctc tctcatctct gcggtcagtg acaaactgag 5220
 atggcggatg aaggaggaaa tggatcgtgc ccaggcagag ctcaatgcct tgaaacgaac 5280
 agaagaagac ctgaaaagg gtcaccagaa actggaagag atggttacc gtttagatca 5340
 agaagtagcc gaggttgata aaaacataga acttttgaaa aagaaggatg aagaactcag 5400
 ttctgctctg gaaaaaatgg aaaatcagtc tgaaaacaat gatatcgatg aagttatcat 5460
 tcccacagct cccttataca aacagatcct gaatctgtat gcagaagaaa acgctattga 5520
 agacactatc ctttacttgg gagaagcctt gagaaggggc gtgatagacc tggatgtcctt 5580
 cctgaagcat gtacgtcttc tgtcccgtaa acagttccag ctgagggcac taatgcaaaa 5640
 agcaagaaag actgccggtc tcagtacact ctactgagga tccgctcagc tgcagccaag 5700
 ctaattccgg gcgaatttct tatgatttat gatttttatt attaaataag ttataaaaaa 5760
 aataagtgta taaaattttt aaagtgactc ttaggtttta aaacgaaaat tcttgttctt 5820
 gagtaactct ttctgtagg tcaggttgct ttctcaggta tagcatgagg tcgctcttat 5880
 tgaccacacc tctaccggca tgcgagctcc agcttttgtt cccttttagtg agggttaatt 5940
 gcgcgcttgg cgtaatcatg gtcatactctg tttcctgtgt gaaattgtta tccgctcaca 6000
 attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg 6060

ES 2 646 388 A1

agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg 6120
 tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggagc 6180
 tcttccgctt cctcgtcac tgactcgtg cgctcggctg ttcggctgcg gcgagcggta 6240
 tcagctcact caaaggcggg aatacgggta tccacagaat caggggataa cgcaggaaag 6300
 aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 6360
 tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg 6420
 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 6480
 cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga 6540
 agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtgta ggctgctcgc 6600
 tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcg cttatccggt 6660
 aactatcgtc ttgagtcca cccggtaga cagcacttat cgccactggc agcagccact 6720
 ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtgggtg 6780
 cctaactacg gctacactag aaggacagta tttggatctt gcgctctgct gaagccagtt 6840
 accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggg 6900
 ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 6960
 ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgta agggattttg 7020
 gtcagagat tatcaaaaag gatcttcacc tagatccttt taaattaa atgaagtttt 7080
 aatcaatct aaagtatata tgagtaaact tggctgaca gttaccaatg cttaatcagt 7140
 gaggcaccta tctcagcgtc ctgtctatct cgttcaccca tagttgctg actccccgctc 7200
 gtgtagataa ctacgatacg ggagggctta ccatctggcc ccagtgctgc aatgataccg 7260
 cgagaccac gtcaccggc tccagattta tcagcaataa accagccagc cggaagggcc 7320
 gagcgagaa gtggctctgc aactttatcc gcctccatcc agtctattaa ttggtgcccg 7380
 gaagctagag taagtagttc gccagttaat agtttgcgca acgttgctg cattgctaca 7440
 ggcatcgtgg tgtcacgctc gtcgtttggg atggcttcat tcagctccgg ttcccaacga 7500
 tcaaggcgag ttacatgat ccccatggtg tgcaaaaaag cggttagctc cttcggctct 7560
 ccgatcgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcactg 7620
 cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg tgagtactca 7680
 accaagtcac tctgagaata gtgtatgagg cgaccgagtt gctcttgccc ggcgtcaata 7740
 cgggataata ccgcccaca tagcagaact taaaagtgc tcatcattgg aaaacgttct 7800
 tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaaccact 7860
 cgtgcacca actgatcttc agcatctttt actttacca gcgcttctgg gtgagcaaaa 7920
 acaggaaggc aaaatgccgc aaaaaaggga ataagggcga cacggaaatg ttgaatactc 7980
 atactcttcc tttttcaata ttattgaagc atttatcagg gttattgtct catgagcggg 8040
 tacatatttg aatgtattta gaaaaataa caaatagggg ttccgagcac atttccccga 8100
 aaagtgccac ctgacgtcta agaaaccatt attatcatga cattaaccta taaaaatagg 8160

ES 2 646 388 A1

cgtatcacga ggccctttcg tc

8182

<210> 31
 <211> 1173
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)..(1173)

<400> 31
 atg gcg gtg tcg gag agc cag ctc aag aaa atg gtg tcc aag tac aaa 48
 Met Ala Val Ser Glu Ser Gln Leu Lys Lys Met Val Ser Lys Tyr Lys
 1 5 10 15
 tac aga gac cta act gta cgt gaa act gtc aat gtt att act cta tac 96
 Tyr Arg Asp Leu Thr Val Arg Glu Thr Val Asn Val Ile Thr Leu Tyr
 20 25 30
 aaa gat ctc aaa cct gtt ttg gat tca tat gtt ttt aac gat ggc agt 144
 Lys Asp Leu Lys Pro Val Leu Asp Ser Tyr Val Phe Asn Asp Gly Ser
 35 40 45
 tcc agg gaa cta atg aac ctc act gga aca atc cct gtg cct tat aga 192
 Ser Arg Glu Leu Met Asn Leu Thr Gly Thr Ile Pro Val Pro Tyr Arg
 50 55 60
 ggt aat aca tac aat att cca ata tgc cta tgg cta ctg gac aca tac 240
 Gly Asn Thr Tyr Asn Ile Pro Ile Cys Leu Trp Leu Leu Asp Thr Tyr
 65 70 75 80
 cca tat aat ccc cct atc tgt ttt gtt aag cct act agt tca atg act 288
 Pro Tyr Asn Pro Pro Ile Cys Phe Val Lys Pro Thr Ser Ser Met Thr
 85 90 95
 att aaa aca gga aag cat gtt gat gca aat ggg aag ata tat ctt cct 336
 Ile Lys Thr Gly Lys His Val Asp Ala Asn Gly Lys Ile Tyr Leu Pro
 100 105 110
 tat cta cat gaa tgg aaa cac cca cag tca gac ttg ttg ggg ctt att 384
 Tyr Leu His Glu Trp Lys His Pro Gln Ser Asp Leu Leu Gly Leu Ile
 115 120 125
 cag gtc atg att gtg gta ttt gga gat gaa cct cca gtc ttc tct cgt 432
 Gln Val Met Ile Val Val Phe Gly Asp Glu Pro Pro Val Phe Ser Arg
 130 135 140
 cct att tcg gca tcc tat ccg cca tac cag gca acg ggg cca cca aat 480
 Pro Ile Ser Ala Ser Tyr Pro Pro Tyr Gln Ala Thr Gly Pro Pro Asn
 145 150 155 160
 act tcc tac atg cca ggc atg cca ggt gga atc tct cca tac cca tcc 528
 Thr Ser Tyr Met Pro Gly Met Pro Gly Gly Ile Ser Pro Tyr Pro Ser
 165 170 175
 gga tac cct ccc aat ccc agt ggt tac cca ggc tgt cct tac cca cct 576
 Gly Tyr Pro Pro Asn Pro Ser Gly Tyr Pro Gly Cys Pro Tyr Pro Pro
 180 185 190
 ggt ggt cca tat cct gcc aca aca agt tct cag tac cct tct cag cct 624
 Gly Gly Pro Tyr Pro Ala Thr Thr Ser Ser Gln Tyr Pro Ser Gln Pro
 195 200 205
 cct gtg acc act gtt ggt ccc agt agg gat ggc aca atc agc gag gac 672
 Pro Val Thr Thr Val Gly Pro Ser Arg Asp Gly Thr Ile Ser Glu Asp
 50

ES 2 646 388 A1

210	215	220	
acc atc cga gcc tct ctc atc tct gcg gtc agt gac aaa ctg aga tgg			720
Thr Ile Arg Ala Ser Leu Ile Ser Ala Val Ser Asp Lys Leu Arg Trp			240
225	230	235	
cgg atg aag gag gaa atg gat cgt gcc cag gca gag ctc aat gcc ttg			768
Arg Met Lys Glu Glu Met Asp Arg Ala Gln Ala Glu Leu Asn Ala Leu			255
245		250	
aaa cga aca gaa gaa gac ctg aaa aag ggt cac cag aaa ctg gaa gag			816
Lys Arg Thr Glu Glu Asp Leu Lys Lys Gly His Gln Lys Leu Glu Glu			270
260		265	
atg gtt acc cgt tta gat caa gaa gta gcc gag gtt gat aaa aac ata			864
Met Val Thr Arg Leu Asp Gln Glu Val Ala Glu Val Asp Lys Asn Ile			285
275	280	285	
gaa ctt ttg aaa aag aag gat gaa gaa ctc agt tct gct ctg gaa aaa			912
Glu Leu Leu Lys Lys Lys Asp Glu Glu Leu Ser Ser Ala Leu Glu Lys			295
290	295	300	
atg gaa aat cag tct gaa aac aat gat atc gat gaa gtt atc att ccc			960
Met Glu Asn Gln Ser Glu Asn Asn Asp Ile Asp Glu Val Ile Ile Pro			320
305	310	315	
aca gct ccc tta tac aaa cag atc ctg aat ctg tat gca gaa gaa aac			1008
Thr Ala Pro Leu Tyr Lys Gln Ile Leu Asn Leu Tyr Ala Glu Glu Asn			335
325	330	335	
gct att gaa gac act atc ctt tac ttg gga gaa gcc ttg aga agg ggc			1056
Ala Ile Glu Asp Thr Ile Leu Tyr Leu Gly Glu Ala Leu Arg Arg Gly			345
340	345	350	
gtg ata gac ctg gat gtc ttc ctg aag cat gta cgt ctt ctg tcc cgt			1104
Val Ile Asp Leu Asp Val Phe Leu Lys His Val Arg Leu Leu Ser Arg			365
355	360	365	
aaa cag ttc cag ctg agg gca cta atg caa aaa gca aga aag act gcc			1152
Lys Gln Phe Gln Leu Arg Ala Leu Met Gln Lys Ala Arg Lys Thr Ala			380
370	375	380	
ggt ctc agt gac ctc tac tga			1173
Gly Leu Ser Asp Leu Tyr			390
385	390		

<210> 32
 <211> 390
 <212> PRT
 <213> Homo sapiens

<400> 32

Met Ala Val Ser Glu Ser Gln Leu Lys Lys Met Val Ser Lys Tyr Lys	
1 5 10 15	
Tyr Arg Asp Leu Thr Val Arg Glu Thr Val Asn Val Ile Thr Leu Tyr	
20 25 30	
Lys Asp Leu Lys Pro Val Leu Asp Ser Tyr Val Phe Asn Asp Gly Ser	
35 40 45	
Ser Arg Glu Leu Met Asn Leu Thr Gly Thr Ile Pro Val Pro Tyr Arg	
50 55 60	

ES 2 646 388 A1

Gly 65 Asn Thr Tyr Asn 70 Ile Pro Ile Cys Leu 75 Trp Leu Leu Asp Thr Tyr 80
 Pro Tyr Asn Pro 85 Pro Ile Cys Phe Val Lys 90 Pro Thr Ser Ser Met 95 Thr
 Ile Lys Thr Gly 100 Lys His Val Asp Ala 105 Asn Gly Lys Ile Tyr 110 Leu Pro
 Tyr Leu His 115 Glu Trp Lys His Pro 120 Gln Ser Asp Leu Leu 125 Gly Leu Ile
 Gln Val 130 Met Ile Val Val Phe 135 Gly Asp Glu Pro Pro 140 Val Phe Ser Arg
 Pro 145 Ile Ser Ala Ser Tyr 150 Pro Pro Tyr Gln Ala 155 Thr Gly Pro Pro Asn 160
 Thr Ser Tyr Met Pro 165 Gly Met Pro Gly Gly 170 Ile Ser Pro Tyr Pro 175 Ser
 Gly Tyr Pro Pro 180 Asn Pro Ser Gly Tyr 185 Pro Gly Cys Pro Tyr 190 Pro Pro
 Gly Gly Pro 195 Tyr Pro Ala Thr Thr 200 Ser Ser Gln Tyr Pro 205 Ser Gln Pro
 Pro Val 210 Thr Thr Val Gly Pro 215 Ser Arg Asp Gly Thr 220 Ile Ser Glu Asp
 Thr 225 Ile Arg Ala Ser Leu 230 Ile Ser Ala Val Ser 235 Asp Lys Leu Arg Trp 240
 Arg Met Lys Glu 245 Glu Met Asp Arg Ala Gln 250 Ala Glu Leu Asn Ala 255 Leu
 Lys Arg Thr Glu 260 Glu Asp Leu Lys Lys 265 Gly His Gln Lys Leu 270 Glu Glu
 Met Val Thr 275 Arg Leu Asp Gln Glu 280 Val Ala Glu Val Asp 285 Lys Asn Ile
 Glu 290 Leu Leu Lys Lys Lys Asp 295 Glu Glu Leu Ser Ser 300 Ala Leu Glu Lys
 Met 305 Glu Asn Gln Ser Glu 310 Asn Asn Asp Ile Asp 315 Glu Val Ile Ile Pro 320
 Thr Ala Pro Leu Tyr 325 Lys Gln Ile Leu Asn 330 Leu Tyr Ala Glu Glu 335 Asn

ES 2 646 388 A1

Ala Ile Glu Asp Thr Ile Leu Tyr Leu Gly Glu Ala Leu Arg Arg Gly
 340 345 350

Val Ile Asp Leu Asp Val Phe Leu Lys His Val Arg Leu Leu Ser Arg
 355 360 365

Lys Gln Phe Gln Leu Arg Ala Leu Met Gln Lys Ala Arg Lys Thr Ala
 370 375 380

Gly Leu Ser Asp Leu Tyr
 385 390

<210> 33
 <211> 11325
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Plásmido pLexA(1-200)PL-Tsg101

<400> 33
 agctgcatgt gtcagagggtt ttcaccgtca tcaccgaaac gcgcgaggca ggatgatccg 60
 ggatcgaaga aatgatggta aatgaaatag gaaatcaagg agcatgaagg caaaagacaa 120
 atataagggt cgaacgaaaa ataaagttaa aagtgttgat atgatgtatt tggctttgcg 180
 gcgccgaaaa aacgagttta cgcaattgca caatcatgct gactctgtgg cggacccgcg 240
 ctcttgccgg cccggcgata acgctgggcg tgaggctgtg cccggcggag ttttttgcgc 300
 ctgcattttc caaggtttac cctgcgctaa ggggcgagat tggagaagca ataagaatgc 360
 cggttggggt tgcgatgatg acgaccacga caactggtgt cattatttaa gttgccgaaa 420
 gaacctgagt gcatttgcaa catgagtata ctagaagaat gagccaagac ttgcgagacg 480
 cgagtttgcc ggtggtgcga acaatagagc gaccatgacc ttgaagggtga gacgcgcata 540
 accgctagag tactttgaag aggaaacagc aatagggttg ctaccagtat aaatagacag 600
 gtacatacaa cactggaaat ggttgtctgt ttgagtacgc tttcaattca tttgggtgtg 660
 cactttatta tgttacaata tggaagggaa ctttacactt ctctatgca catatattaa 720
 ttaaagtcca atgctagtag agaagggggg taacaccctt ccgcgctctt ttccgatttt 780
 tttctaaacc gtggaatatt tcggatatcc ttttgttggt tccgggtgta caatatggac 840
 ttctctttt ctggcaacca aaccataca tcgggattcc tataatacct tcgttggtct 900
 ccctaacatg taggtggcgg aggggagata tacaatagaa cagataccag acaagacata 960
 atgggctaaa caagactaca ccaattacac tgcctcattg atggtggtac ataacgaact 1020
 aatactgtag ccctagactt gatagccatc atcatatcga agtttcacta ccctttttcc 1080
 atttgccatc tattgaagta ataataggcg catgcaactt cttttctttt tttttctttt 1140
 ctctctccc cgttgttgtc tcaccatatc cgcaatgaca aaaaaaatga tggaagacac 1200
 taaaggaaaa aattaacgac aaagacagca ccaacagatg tcgttggtcc agagctgatg 1260
 aggggtatct tcgaacacac gaaacttttt ctttccttca ttcacgcaca ctactctcta 1320

ES 2 646 388 A1

atgagcaacg gtatacggcc ttccttccag ttacttgaat ttgaaataaa aaaagtttgc 1380
cgctttgcta tcaagtataa atagacctgc aattattaat cttttgtttc ctcgtcattg 1440
ttctcgttcc cttttcttct tgtttctttt tctgcacaat atttcaagct ataccaagca 1500
tacaatcaac tccaagcttg aattaattcc gggcggaatg aaagcgtaa cggccaggca 1560
acaagaggty tttgatctca tccgtgatca catcagccag acaggatgac cgccgacgcy 1620
tgcggaaatc gcgcagcgtt tggggttccg ttccccaac gcggctgaag aacatctgaa 1680
ggcgctggca cgcaaaggcy ttattgaaat tgtttccggc gcatcacgcy ggattcgtct 1740
gttgaggaa gaggaagaag ggttgccgct ggtaggtcgt gtggctgccc gtgaaccact 1800
tctggcgcaa cagcatattg aaggtcatta tcaggtcgat ctttccttat tcaagccgaa 1860
tgctgatttc ctgctgcgcy tcagcgggat gtcgatgaaa gatatcggca ttatggatgg 1920
tgacttgctg gcagtgcata aaactcagga tgtacgtaac ggtcaggctc ttgtcgcacg 1980
tattgatgac gaagttaccg ttaagcgcct gaaaaaacag ggcaataaag tcgaactggt 2040
gccagaaaat agcgagtta aaccaattgt cgtagatctt cgtcagcaga gcttcacat 2100
tgaagggctg gcggttgggg ttattcgcaa cggcgactgg ctggaattcc cggggatcct 2160
catggcggty tcggagagcc agctcaagaa aatggtgtcc aagtacaaat acagagacct 2220
aactgtacgt gaaactgtca atgttattac tctatacaaa gatctcaaac ctgttttggaa 2280
ttcatatggt ttaacgatg gcagttccag ggaactaatg aacctactg gaacaatccc 2340
tgtgccttat agaggtaata catacaatat tccaatatgc ctatggctac tggacacata 2400
cccatataat ccccctatct gttttgttaa gcctactagt tcaatgacta ttaaaacagg 2460
aaagcatggt gatgcaaatg ggaagatata tcttccttat ctacatgaat ggaaacaccc 2520
acagtcagac ttgttggggc ttattcaggt catgattgtg gtatttggag atgaacctcc 2580
agtcttctct cgtcctatth cggcatccta tccgccatac caggcaacgg ggccaccaa 2640
tacttctac atgccaggca tgccaggtyg aatctctcca taccatccg gataacctcc 2700
caatcccagt ggttaccag gctgtcctta cccacctggt ggtccatata ctgccacaac 2760
aagttctcag tacccttctc agcctcctgt gacctggtt ggtcccagta gggatggcac 2820
aatcagcag gacaccatcc gagcctctct catctctgcy gtcagtgaca aactgagatg 2880
gcggatgaag gaggaaatgg atcgtgccca ggcagagctc aatgccttga aacgaacaga 2940
agaagacctg aaaaagggtc accagaaact ggaagagatg gttaccctgt tagatcaaga 3000
agtagccgag gttgataaaa acatagaact tttgaaaaag aaggatgaag aactcagttc 3060
tgctctggaa aaaatggaaa atcagcttga aaacaatgat atcgatgaag ttatcattcc 3120
cacagctccc ttatacaaac agatcctgaa tctgtatgca gaagaaaacg ctattgaaga 3180
cactatcctt tacttgggag aagccttgag aaggggcytg atagacctgg atgtcttctc 3240
gaagcatgta cgtcttctgt cccgtaaaca gttccagctg agggcactaa tgcaaaaagc 3300
aagaaagact gccggtctca gtgacctcta ctgaggatcc gtcgacctgc agccaagcta 3360
attccgggcy aatttcttat gatttatgat ttttattatt aaataagtta taaaaaaat 3420

ES 2 646 388 A1

aagtgtatac aaattttaa gtgactctta ggttttataa cgaaaattct tgttcttgag 3480
taactctttc ctgtaggtca ggttgctttc tcaggatag catgaggctg ctcttattga 3540
ccacacctct accggcatgc cgagcaaag cctgcaaag gctccccatt tcaccaatt 3600
gtagatatgc taactccagc aatgagttga tgaatctcgg tgtgtatttt atgtcctcag 3660
aggacaacac ctggtgtaat cgttcttcca cacggatcga tccacaggac ggggtgtggtc 3720
gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggaggcgg 3780
ccaaagcggc cggacagtgc tccgagaacg ggtgagcata gaaattgcat caacgcatat 3840
agcgctagca gcacgcata gtgactggcg atgctgctgg aatggacgat atcccgaag 3900
aggcccggca gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg 3960
aggatgacga tgagcgcatt gttagatttc atacacgggt cctgactgcg ttagcaattt 4020
aactgtgata aactaccgca ttaaagctag ctttgaagaa aaatgagcct tattcaatct 4080
ttgctataaa aaatggcca aaatctcaca ttggaagaca tttgatgacc tcatttcttt 4140
caatgaaggg cctaacggag ttgactaatg ttgtgggaaa ttggagcgat aagcgtgctt 4200
ctgccgtggc caggacaacg tatactcatc agataacagc aatacctgat cactacttcg 4260
cactagtctc tcggtactat gcatatgatc caatatcaaa ggaaatgata gcattgaagg 4320
atgagactaa tccaattgag gagtggcagc atatagaaca gctaaagggt agtgctgaag 4380
gaagcatacg ataccccga tggaaatggga taatatcaca ggaggacta gactaccttt 4440
catcctacat aaatagacgc atataagtac gcatttaagc ataaacacgc actatgccgt 4500
tcttctcatg tatatatata tacaggcaac acgcagatat aggtgagcag tgaacagtga 4560
gctgtatgtg cgcagctcgc gttgcatttt cggaagcgtc cgttttcgga aacgctttga 4620
agttcctatt ccgaagtcc tattctctag aaagtatagg aacttcagag cgcttttgaa 4680
aaccaaaagc gctctgaaga cgcactttca aaaaaccaa aacgcaccgg actgtaacga 4740
gctactaaaa tattgcgaat accgcttcca caaacattgc tcaaaagtat ctctttgcta 4800
tatatctctg tgctatatcc ctatataacc taccatcca cctttcgctc cttgaacttg 4860
catctaaact cgacctctac attttttatg tttatctcta gtattactct ttagacaaaa 4920
aaattgtagt aagaactatt catagagtga atcgaaaaca atacgaaaat gtaaacattt 4980
cctatacgta gtatatagag acaaaataga agaaaccgtt cataattttc tgaccaatga 5040
agaatcatca acgctatcac tttctgttca caaagtatgc gcaatccaca tcggtataga 5100
atataatcgg ggatgccttt atcttgaaaa aatgcacccg cagcttcgct agtaatcagt 5160
aaacgaggga agtgagtcga ggcttttttt atggaagaga aaatagacac caaagtagcc 5220
ttcttctaac cttaacggac ctacagtgc aaaagttatc aagagactgc attatagagc 5280
gcacaaagga gaaaaaagt aatctaagat gctttgttag aaaaatagcg ctctcgggat 5340
gcatttttgt agaacaaaa agaagtatag attctttgtt ggtaaaatag cgctctcgcg 5400
ttgcatttct gttctgtaaa aatgcagctc agattctttg tttgaaaaat tagcgccttc 5460

ES 2 646 388 A1

gcgttgcatt tttgttttac aaaaatgaag cacagattct tcgttggtaa aatagcgctt 5520
 tcgcgttgca tttctgttct gtaaaaatgc agctcagatt ctttgtttga aaaattagcg 5580
 ctctcgcggt gcatttttgt tctacaaaat gaagcacaga tgcttcgtta acaaagatat 5640
 gctattgaag tgcaagatgg aaacgcagaa aatgaaccgg ggatgcgacg tgcaagatta 5700
 cctatgcaat agatgcaata gtttctccag gaaccgaaat acatacattg tcttccgtaa 5760
 agcgctagac tatatattat tatacagggt caaatatact atctgtttca gggaaaactc 5820
 ccaggttcgg atgttcaaaa ttcaatgatg ggtaacaagt acgatcgtaa atctgtaaaa 5880
 cagtttgtcg gatattaggc tgtatctcct caaagcgtat tcgaatatca ttgagaagct 5940
 gcagcaggcg tgaagttaga cgacaacttc tctctggaaa cgcataccga tattcaggct 6000
 gctgcaaagg cacaggctag tgcccgtgcg agtgcacccg gtaccacccc agatgctgta 6060
 gtagcttctg gtagcactgc aatgagccat gcttatcaag aaaacacagg ttttggtact 6120
 cgtcccatat atcttgacat gcaagccact acaccaacag accctaggggt tttggatacg 6180
 atgttgaagt tttatacggg actttatggt aatcctcatt ccaacactca ctcttacggt 6240
 tgggaaacia aactgctgt ggaaaatgct agagctcacg tagcaaagat gatcaatgcc 6300
 gacccaagg aaataatatt cacttcggga gcgaccgaat ctaataatat ggttcttaag 6360
 ggtgtcccaa gatattataa gaagactaag aaacacatca tcaccactag aacggaacac 6420
 aagtgtgtct tggaagccgc acgggccatg atgaaggagg gatttgaagt cactttccta 6480
 aatgtggacg atcaaggctt tatcgatttg aaggaattgg aagatgcat tagaccagat 6540
 acctgtctcg tctctgtgat ggctgtcaat aatgaaatcg gtgtcattca acctatataa 6600
 gaaattggag caattttaga aaagaataag atcctcgggg acaccaaata tggcgatctc 6660
 ggcttttctg tttcttgtag ctgggacatg tttgccatcg atccatctac caccagaacg 6720
 gccgttagat ctgctgccac cgttgtttcc accgaagaaa ccaccgttgc cgtaaccacc 6780
 acgacggttg ttgctaaaga agctgccacc gccacggcca ccgttgtagc cgccgttggt 6840
 gttattgtag ttgctactgt ttttctggc acttcttgggt tttcctctta agtgaggagg 6900
 aacataacca ttctcgttgt tgtcgttgat gcttaaattt tgcacttggt cgctcagttc 6960
 agccataata tgaaatgctt ttcttgttgt tcttacggaa taccacttgc cacctatcac 7020
 cacaactaac tttttccgtt tcctccatct cttttatatt ttttttctcg atcgagttca 7080
 agagaaaaaa aaagaaaaag caaaaagaaa aaaggaaagc ggcctcgtt cagaatgaca 7140
 cgtatagaat gatgcattac cttgtcatct tcagtatcat actgttcgta tacatactta 7200
 ctgacattca taggtataca tatatacaca tgtatatata tcgtatgctg cagctttaaa 7260
 taatcgggtg cactacataa gaacacctt ggtggagggg acatcgttgg taccattggg 7320
 cgaggtggct tctcttatgg caaccgcaag agccttgaac gcactctcac tacggtgatg 7380
 atcattcttg cctcgcagac aatcaacgtg gagggttaatt ctgctagcct ctgcaaagct 7440
 ttcaagaaaa tgcgggatca tctcgaaga gagatctcct actttctccc tttgcaaacc 7500
 aagttcgaca actgctgacg gcctgttcga aagatctacc accgctctgg aaagtgcctc 7560

ES 2 646 388 A1

atccaaaggc gcaaatcctg atccaaacct ttttactcca cgcgccagta gggcctcttt 7620
aaaagcttga ccgagagcaa tcccgcagtc ttcagtgggtg tgatgggtcgt ctatgtgtaa 7680
gtcaccaatg cactcaacga ttagcgacca gccggaatgc ttggccagag catgtatcat 7740
atggtccaga aaccctatac ctgtgtggac gttaatcact tgcgattgtg tggcctgttc 7800
tgctactgct tctgcctctt tttctgggaa gatcgagtgc tctatcgcta ggggaccacc 7860
ctttaaagag atcgcaatct gaatcttggg ttcatttgta atacgcttta ctagggcttt 7920
ctgctctgtc atctttgcct tcgtttatct tgcctgctca ttttttagta tattcttcga 7980
agaaatcaca ttactttata taatgtataa ttcattatgt gataatgcca atcgctaaga 8040
aaaaaaaaaga gtcacccgct aggtggaaaa aaaaaaatga aaatcattac cgaggcataa 8100
aaaaatatag agtgtactag aggaggccaa gagtaataga aaaagaaaaat tgcgggaaag 8160
gactgtgtta tgacttcctt gactaatgcc gtgttcaaac gatacctggc agtgactcct 8220
agcgctcacc aagctcttaa aacgagaatt aagaaaaagt cgtcatcttt cgataagttt 8280
tccccacagc aaagcaatag tagaaaaaca atgggaaacg ttgaatgaag acaaagcgtc 8340
gtggtttaa aggaaatag ctcacgtaca tgctagggaa caggaccgtg cagcggatct 8400
aatgaatcca tttgttagtt aatagtttaa atgtttttat cggaagaggt tttgtcatca 8460
catcagcaat gttcttcttg gtctcgatgt agtatacgta taaattatta cctgatactt 8520
catctctaag tctcattgcc tttgtgccaa aaaatctggt tctaaatttc tcttcatttg 8580
tagacttaat tatactgatc gttgatctac tatcagtaag taagccttta aaaaaaaaaa 8640
aaaaaaaaaa aaaaaaaaaa cctgtaacaa tagcaatacc ccaaatacct aatgtagttc 8700
cagcaagcaa gctaaaaagt aaagcaacaa cataactcac ccctgcatct gcagcttttg 8760
cccgggcagc ctgctctgcc tgtgttttct ttaattgagc agtagaccat ttagcagttg 8820
catgaatagc tgcagcgta catcggataa taatgatggc agccattgta gaagtgcctt 8880
ttgcatttct agtctctttc tcggcttagc tagttttact acatcgcgaa gatagaatct 8940
tagatcacac tgcctttgct gagctggatc aatagagtaa caaaagagtg gtaaggcctc 9000
gttaaaggac aaggacctga gcggaagtgt atcgtacagt agacggagta tactagtata 9060
gtctatagtc cgtggaatta attcttgaag acgaaagggc ctcgtgatac gcctatTTTT 9120
ataggttaat gtcatgataa taatggtttc ttagacgtca ggtggcactt ttcggggaaa 9180
tgtgcgcgga acccctatTT gtttattttt ctaaatacat tcaaataatgt atccgctcat 9240
gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca 9300
acatttccgt gtcgccctta ttcctttttt tgcggcattt tgccttctctg tttttgctca 9360
cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta 9420
catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt 9480
tccaatgatg agcactttta aagttctgct atgtggcgcg gtattatccc gtgttgacgc 9540
cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc 9600

ES 2 646 388 A1

accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgtctgc 9660
cataacccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa 9720
ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga 9780
accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgcagcaat 9840
ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca 9900
attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc 9960
ggctggctgg tttattgctg ataaatctgg agccgggtgag cgtgggtctc gcggtatcat 10020
tgcagcactg gggccagatg gtaagccctc ccgtatcgtg gttatctaca cgacggggag 10080
tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa 10140
gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca 10200
tttttaattt aaaaggatct aggtgaagat cttttttgat aatctcatga ccaaaatccc 10260
ttaacgtgag ttttcgttcc actgagcgtc agaccccgtg gaaaagatca aaggatcttc 10320
ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc 10380
agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 10440
cagcagagcg cagataccea atactgtcct tctagtgtag ccgtagttag gccaccactt 10500
caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 10560
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 10620
ggcgagcgg tcgggctgaa cgggggggtc gtgcacacag cccagcttgg agcgaacgac 10680
ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcggccacgc ttcccgaagg 10740
gagaaaggcg gacaggtatc cggtaaagcg cagggtcgga acaggagagc gcacgagggg 10800
gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 10860
tgagcgtcga tttttgtgat gctcgtcagg ggggaggagc ctatggaaaa acgccagcaa 10920
cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc 10980
gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg 11040
ccgcagccga acgaccgagc gcagcagatc agtgagcag gaagcgggag agcgcctgat 11100
gcggtatfff ctcttacgc atctgtgagg tatttcacac cgcatatggt gactctcag 11160
tacaatctgc tctgatgccg catagttaag ccagtataca ctccgctatc gctacgtgac 11220
tgggtcatgg ctgcccccg acaccgcca acaccgctg acgcgccctg acgggcttgt 11280
ctgctcccgg catccgctta cagacaagct gtgaccgtct ccggg 11325

<210> 34
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Cebador OV310

<400> 34
gggatcctca tggcgggtgtc ggagagccag c

31

<210> 35
 <211> 31
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV311
 <400> 35
 gggatcctca gtagagggtca ctgagaccgg c 31

<210> 36
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV713
 <400> 36
 gcgaggcata tttatggtga agg 23

<210> 37
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV714
 <400> 37
 catttccgtg caaagggtact aac 23

<210> 38
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV731
 <400> 38
 ctgtatataa aaccagtggt tatatgtac 29

<210> 39
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV732
 <400> 39
 tcgagtgctc tatcgctagg g 21

<210> 40
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cebador OV715

<400> 40
ccataggatg ataatgcgat tag 23

<210> 41
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Cebador OV741

<400> 41
cgcttctggt gccggaacc 20

<210> 42
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Cebador OV508

<400> 42
cagattgtac tgagagtga cc 22

<210> 43
<211> 22
<212> DNA
<213> Artificial sequence

<220>
<223> Cebador OV747

<400> 43
attccttgct tcttggtact gg 22

<210> 44
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Cebador OV621

<400> 44
cactgtcacc tggttggacg g 21

<210> 45
<211> 23
<212> DNA
<213> Artificial sequence

<220>
<223> Cebador OV622

<400> 45
ctatagatca gaggttacat ggc 23

<210> 46
<211> 21
<212> DNA
<213> Artificial sequence

ES 2 646 388 A1

<220>
 <223> Cebador OV284

<400> 46
 cgatgatgaa gatacccccac c 21

<210> 47
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Cebador OV285

<400> 47
 gagatggtgc acgatgcaca g 21

<210> 48
 <211> 1797
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Secuencia nucleotídica que codifica para la proteína de fusión
 LexA-TSG101

<220>
 <221> CDS
 <222> (1)..(1797)

<400> 48
 atg aaa gcg tta acg gcc agg caa caa gag gtg ttt gat ctc atc cgt 48
 Met Lys Ala Leu Thr Ala Arg Gln Gln Glu Val Phe Asp Leu Ile Arg
 1 5 10 15

gat cac atc agc cag aca ggt atg ccg ccg acg cgt gcg gaa atc gcg 96
 Asp His Ile Ser Gln Thr Gly Met Pro Pro Thr Arg Ala Glu Ile Ala
 20 25 30

cag cgt ttg ggg ttc cgt tcc cca aac gcg gct gaa gaa cat ctg aag 144
 Gln Arg Leu Gly Phe Arg Ser Pro Asn Ala Ala Glu Glu His Leu Lys
 35 40 45

gcg ctg gca cgc aaa ggc gtt att gaa att gtt tcc ggc gca tca cgc 192
 Ala Leu Ala Arg Lys Gly Val Ile Glu Ile Val Ser Gly Ala Ser Arg
 50 55 60

ggg att cgt ctg ttg cag gaa gag gaa gaa ggg ttg ccg ctg gta ggt 240
 Gly Ile Arg Leu Leu Gln Glu Glu Glu Glu Gly Leu Pro Leu Val Gly
 65 70 75 80

cgt gtg gct gcc ggt gaa cca ctt ctg gcg caa cag cat att gaa ggt 288
 Arg Val Ala Ala Gly Glu Pro Leu Leu Ala Gln Gln His Ile Glu Gly
 85 90 95

cat tat cag gtc gat cct tcc tta ttc aag ccg aat gct gat ttc ctg 336
 His Tyr Gln Val Asp Pro Ser Leu Phe Lys Pro Asn Ala Asp Phe Leu
 100 105 110

ctg cgc gtc agc ggg atg tcg atg aaa gat atc ggc att atg gat ggt 384
 Leu Arg Val Ser Gly Met Ser Met Lys Asp Ile Gly Ile Met Asp Gly
 115 120 125

gac ttg ctg gca gtg cat aaa act cag gat gta cgt aac ggt cag gtc 432
 Asp Leu Leu Ala Val His Lys Thr Gln Asp Val Arg Asn Gly Gln Val
 61

ES 2 646 388 A1

130				135				140								
gtt Val 145	gtc Val	gca Ala	cg Arg	att Ile	gat Asp 150	gac Asp	gaa Glu	gtt Val	acc Thr	gtt Val 155	aag Lys	cg Arg	ctg Leu	aaa Lys	aaa Lys 160	480
cag Gln	ggc Gly	aat Asn	aaa Lys	gtc Val 165	gaa Glu	ctg Leu	ttg Leu	cca Pro	gaa Glu 170	aat Asn	agc Ser	gag Glu	ttt Phe	aaa Lys 175	cca Pro	528
att Ile	gtc Val	gta Val	gat Asp 180	ctt Leu	cg Arg	cag Gln	cag Gln	agc Ser 185	ttc Phe	acc Thr	att Ile	gaa Glu	ggg Gly 190	ctg Leu	gcg Ala	576
gtt Val	ggg Gly	gtt Val 195	att Ile	cg Arg	aac Asn	ggc Gly	gac Asp 200	tgg Trp	ctg Leu	gaa Glu	ttc Phe	ccg Pro 205	ggg Gly	atc Ile	ctc Leu	624
atg Met	gcg Ala 210	gtg Val	tcg Ser	gag Glu	agc Ser	cag Gln 215	ctc Leu	aag Lys	aaa Lys	atg Met	gtg Val 220	tcc Ser	aag Lys	tac Tyr	aaa Lys	672
tac Tyr 225	aga Arg	gac Asp	cta Leu	act Thr	gta Val 230	cg Arg	gaa Glu	act Thr	gtc Val	aat Asn 235	gtt Val	att Ile	act Thr	cta Leu	tac Tyr 240	720
aaa Lys	gat Asp	ctc Leu	aaa Lys	cct Pro 245	gtt Val	ttg Leu	gat Asp	tca Ser	tat Tyr 250	gtt Val	ttt Phe	aac Asn	gat Asp	ggc Gly 255	agt Ser	768
tcc Ser	agg Arg	gaa Glu	cta Leu 260	atg Met	aac Asn	ctc Leu	act Thr	gga Gly 265	aca Thr	atc Ile	cct Pro	gtg Val	cct Pro 270	tat Tyr	aga Arg	816
ggg Gly	aat Asn	aca Thr 275	tac Tyr	aat Asn	att Ile	cca Pro	ata Ile 280	tgc Cys	cta Leu	tgg Trp	cta Leu	ctg Leu 285	gac Asp	aca Thr	tac Tyr	864
cca Pro	tat Tyr 290	aat Asn	ccc Pro	cct Pro	atc Ile	tgt Cys 295	ttt Phe	gtt Val	aag Lys	cct Pro	act Thr 300	agt Ser	tca Ser	atg Met	act Thr	912
att Ile 305	aaa Lys	aca Thr	gga Gly	aag Lys	cat His 310	gtt Val	gat Asp	gca Ala	aat Asn	ggg Gly 315	aag Lys	ata Ile	tat Tyr	ctt Leu	cct Pro 320	960
tat Tyr	cta Leu	cat His	gaa Glu	tgg Trp 325	aaa Lys	cac His	cca Pro	cag Gln	tca Ser 330	gac Asp	ttg Leu	ttg Leu	ggg Gly 335	ctt Leu 335	att Ile	1008
cag Gln	gtc Val	atg Met	att Ile 340	gtg Val	gta Val	ttt Phe	gga Gly	gat Asp 345	gaa Glu	cct Pro	cca Pro	gtc Val	ttc Phe 350	tct Ser	cg Arg	1056
cct Pro	att Ile	tcg Ser 355	gca Ala	tcc Ser	tat Tyr	ccg Pro	cca Pro 360	tac Tyr	cag Gln	gca Ala	acg Thr	ggg Gly 365	cca Pro	cca Pro	aat Asn	1104
act Thr	tcc Ser 370	tac Tyr	atg Met	cca Pro	ggc Gly	atg Met 375	cca Pro	ggg Gly	gga Gly	atc Ile	tct Ser 380	cca Pro	tac Tyr	cca Pro	tcc Ser	1152
gga Gly 385	tac Tyr	cct Pro	ccc Pro	aat Asn	ccc Pro 390	agt Ser	ggg Gly	tac Tyr	cca Pro	ggc Gly 395	tgt Cys	cct Pro	tac Tyr	cca Pro	cct Pro 400	1200
ggg Gly	ggg Gly	cca Pro	tat Tyr	cct Pro 405	gcc Ala	aca Thr	aca Thr	agt Ser	tct Ser 410	cag Gln	tac Tyr	cct Pro	tct Ser	cag Gln 415	cct Pro	1248

ES 2 646 388 A1

cct gtg acc act gtt ggt ccc agt agg gat ggc aca atc agc gag gac 1296
 Pro Val Thr Thr 420 Val Gly Pro Ser Arg 425 Asp Gly Thr Ile Ser 430 Glu Asp

acc atc cga gcc tct ctc atc tct gcg gtc agt gac aaa ctg aga tgg 1344
 Thr Ile Arg 435 Ala Ser Leu Ile Ser 440 Ala Val Ser Asp Lys 445 Leu Arg Trp

cgg atg aag gag gaa atg gat cgt gcc cag gca gag ctc aat gcc ttg 1392
 Arg Met 450 Lys Glu Glu Met Asp 455 Arg Ala Gln Ala Glu 460 Leu Asn Ala Leu

aaa cga aca gaa gaa gac ctg aaa aag ggt cac cag aaa ctg gaa gag 1440
 Lys Arg Thr 465 Glu Glu Asp 470 Leu Lys Lys Gly His 475 Gln Lys Leu Glu 480 Glu

atg gtt acc cgt tta gat caa gaa gta gcc gag gtt gat aaa aac ata 1488
 Met Val Thr Arg 485 Leu Asp Gln Glu Val 490 Ala Glu Val Asp Lys 495 Asn Ile

gaa ctt ttg aaa aag aag gat gaa gaa ctc agt tct gct ctg gaa aaa 1536
 Glu Leu Leu 500 Lys Lys Lys Asp Glu Glu 505 Leu Ser Ser Ala 510 Leu Glu Lys

atg gaa aat cag tct gaa aac aat gat atc gat gaa gtt atc att ccc 1584
 Met Glu Asn 515 Gln Ser Glu Asn Asn 520 Asp Ile Asp Glu Val 525 Ile Ile Pro

aca gct ccc tta tac aaa cag atc ctg aat ctg tat gca gaa gaa aac 1632
 Thr Ala 530 Pro Leu Tyr Lys Gln 535 Ile Leu Asn Leu Tyr 540 Ala Glu Glu Asn

gct att gaa gac act atc ctt tac ttg gga gaa gcc ttg aga agg ggc 1680
 Ala Ile Glu Asp Thr 550 Ile Leu Tyr Leu Gly 555 Glu Ala Leu Arg Arg Gly 560

gtg ata gac ctg gat gtc ttc ctg aag cat gta cgt ctt ctg tcc cgt 1728
 Val Ile Asp Leu 565 Asp Val Phe Leu Lys 570 His Val Arg Leu Leu 575 Ser Arg

aaa cag ttc cag ctg agg gca cta atg caa aaa gca aga aag act gcc 1776
 Lys Gln Phe 580 Gln Leu Arg Ala Leu Met 585 Gln Lys Ala Arg Lys 590 Thr Ala

ggt ctc agt gac ctc tac tga 1797
 Gly Leu Ser 595 Asp Leu Tyr

<210> 49
 <211> 598
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 49

Met Lys Ala Leu Thr Ala Arg Gln Gln Glu Val Phe Asp Leu Ile Arg
 1 5 10 15

Asp His Ile Ser Gln Thr Gly Met Pro Pro Thr Arg Ala Glu Ile Ala
 20 25 30

Gln Arg Leu Gly Phe Arg Ser Pro Asn Ala Ala Glu Glu His Leu Lys
 63

ES 2 646 388 A1

35 40 45

Ala Leu Ala Arg Lys Gly Val Ile Glu Ile Val Ser Gly Ala Ser Arg
50 55 60

Gly Ile Arg Leu Leu Gln Glu Glu Glu Gly Leu Pro Leu Val Gly
65 70 75 80

Arg Val Ala Ala Gly Glu Pro Leu Leu Ala Gln Gln His Ile Glu Gly
85 90 95

His Tyr Gln Val Asp Pro Ser Leu Phe Lys Pro Asn Ala Asp Phe Leu
100 105 110

Leu Arg Val Ser Gly Met Ser Met Lys Asp Ile Gly Ile Met Asp Gly
115 120 125

Asp Leu Leu Ala Val His Lys Thr Gln Asp Val Arg Asn Gly Gln Val
130 135 140

Val Val Ala Arg Ile Asp Asp Glu Val Thr Val Lys Arg Leu Lys Lys
145 150 155 160

Gln Gly Asn Lys Val Glu Leu Leu Pro Glu Asn Ser Glu Phe Lys Pro
165 170 175

Ile Val Val Asp Leu Arg Gln Gln Ser Phe Thr Ile Glu Gly Leu Ala
180 185 190

Val Gly Val Ile Arg Asn Gly Asp Trp Leu Glu Phe Pro Gly Ile Leu
195 200 205

Met Ala Val Ser Glu Ser Gln Leu Lys Lys Met Val Ser Lys Tyr Lys
210 215 220

Tyr Arg Asp Leu Thr Val Arg Glu Thr Val Asn Val Ile Thr Leu Tyr
225 230 235 240

Lys Asp Leu Lys Pro Val Leu Asp Ser Tyr Val Phe Asn Asp Gly Ser
245 250 255

Ser Arg Glu Leu Met Asn Leu Thr Gly Thr Ile Pro Val Pro Tyr Arg
260 265 270

Gly Asn Thr Tyr Asn Ile Pro Ile Cys Leu Trp Leu Leu Asp Thr Tyr
275 280 285

Pro Tyr Asn Pro Pro Ile Cys Phe Val Lys Pro Thr Ser Ser Met Thr
290 295 300

Ile Lys Thr Gly Lys His Val Asp Ala Asn Gly Lys Ile Tyr Leu Pro
305 310 315 320

ES 2 646 388 A1

Tyr Leu His Glu Trp Lys His Pro Gln Ser Asp Leu Leu Gly Leu Ile
 325 330 335
 Gln Val Met Ile Val Val Phe Gly Asp Glu Pro Pro Val Phe Ser Arg
 340 345 350
 Pro Ile Ser Ala Ser Tyr Pro Pro Tyr Gln Ala Thr Gly Pro Pro Asn
 355 360 365
 Thr Ser Tyr Met Pro Gly Met Pro Gly Gly Ile Ser Pro Tyr Pro Ser
 370 375 380
 Gly Tyr Pro Pro Asn Pro Ser Gly Tyr Pro Gly Cys Pro Tyr Pro Pro
 385 390 395 400
 Gly Gly Pro Tyr Pro Ala Thr Thr Ser Ser Gln Tyr Pro Ser Gln Pro
 405 410 415
 Pro Val Thr Thr Val Gly Pro Ser Arg Asp Gly Thr Ile Ser Glu Asp
 420 425 430
 Thr Ile Arg Ala Ser Leu Ile Ser Ala Val Ser Asp Lys Leu Arg Trp
 435 440 445
 Arg Met Lys Glu Glu Met Asp Arg Ala Gln Ala Glu Leu Asn Ala Leu
 450 455 460
 Lys Arg Thr Glu Glu Asp Leu Lys Lys Gly His Gln Lys Leu Glu Glu
 465 470 475 480
 Met Val Thr Arg Leu Asp Gln Glu Val Ala Glu Val Asp Lys Asn Ile
 485 490 495
 Glu Leu Leu Lys Lys Lys Asp Glu Glu Leu Ser Ser Ala Leu Glu Lys
 500 505 510
 Met Glu Asn Gln Ser Glu Asn Asn Asp Ile Asp Glu Val Ile Ile Pro
 515 520 525
 Thr Ala Pro Leu Tyr Lys Gln Ile Leu Asn Leu Tyr Ala Glu Glu Asn
 530 535 540
 Ala Ile Glu Asp Thr Ile Leu Tyr Leu Gly Glu Ala Leu Arg Arg Gly
 545 550 555 560
 Val Ile Asp Leu Asp Val Phe Leu Lys His Val Arg Leu Leu Ser Arg
 565 570 575
 Lys Gln Phe Gln Leu Arg Ala Leu Met Gln Lys Ala Arg Lys Thr Ala
 580 585 590

ES 2 646 388 A1

Gly Leu Ser Asp Leu Tyr
595

<210> 50
<211> 2421
<212> DNA
<213> Artificial sequence

<220>
<223> Secuencia nucleotídica que codifica GBD-GKRP

<220>
<221> CDS
<222> (1)..(2421)

<400> 50
atg aag cta ctg tct tct atc gaa caa gca tgc gat att tgc cga ctt 48
Met Lys Leu Leu Ser Ser Ile Glu Gln Ala Cys Asp Ile Cys Arg Leu
1 5 10 15
aaa aag ctc aag tgc tcc aaa gaa aaa ccg aag tgc gcc aag tgt ctg 96
Lys Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu
20 25 30
aag aac aac tgg gag tgt cgc tac tct ccc aaa acc aaa agg tct ccg 144
Lys Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro
35 40 45
ctg act agg gca cat ctg aca gaa gtg gaa tca agg cta gaa aga ctg 192
Leu Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu
50 55 60
gaa cag cta ttt cta ctg att ttt cct cga gaa gac ctt gac atg att 240
Glu Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile
65 70 75 80
ttg aaa atg gat tct tta cag gat ata aaa gca ttg tta aca gga tta 288
Leu Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu
85 90 95
ttt gta caa gat aat gtg aat aaa gat gcc gtc aca gat aga ttg gct 336
Phe Val Gln Asp Asn Val Asn Lys Asp Ala Val Thr Asp Arg Leu Ala
100 105 110
tca gtg gag act gat atg cct cta aca ttg aga cag cat aga ata agt 384
Ser Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gln His Arg Ile Ser
115 120 125
gcg aca tca tca tcg gaa gag agt agt aac aaa ggt caa aga cag ttg 432
Ala Thr Ser Ser Ser Glu Glu Ser Ser Asn Lys Gly Gln Arg Gln Leu
130 135 140
act gta tcg ccg gaa ttt gta ata cga ctc act ata ggg cga gcc gcc 480
Thr Val Ser Pro Glu Phe Val Ile Arg Leu Thr Ile Gly Arg Ala Ala
145 150 155 160
atc atg gag gag cag aag ctg atc tca gag gag gac ctg cat atg gcc 528
Ile Met Glu Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu His Met Ala
165 170 175
atg gag gcc gaa ttc atg cca ggc aca aaa cgg ttt caa cat gtc att 576
Met Glu Ala Glu Phe Met Pro Gly Thr Lys Arg Phe Gln His Val Ile
180 185 190
gag acc ccg gag cct ggc aag tgg gag ttg tct ggg tac gag gca gct 624
Glu Thr Pro Glu Pro Gly Lys Trp Glu Leu Ser Gly Tyr Glu Ala Ala
66

ES 2 646 388 A1

195			200			205										
gtg Val	cca Pro 210	atc Ile	acg Thr	gag Glu	aag Lys	tca Ser 215	aac Asn	cca Pro	ctg Leu	acc Thr	cag Gln 220	gat Asp	cta Leu	gac Asp	aaa Lys	672
gca Ala 225	gat Asp	gct Ala	gag Glu	aac Asn	att Ile 230	gtt Val	cga Arg	ctg Leu	cta Leu	ggg Gly 235	caa Gln	tgt Cys	gat Asp	gct Ala	gag Glu 240	720
atc Ile	ttc Phe	cag Gln	gag Glu	gag Glu 245	ggg Gly	caa Gln	gcc Ala	ctg Leu	tcc Ser 250	aca Thr	tac Tyr	cag Gln	aga Arg	ctc Leu 255	tac Tyr	768
agc Ser	gaa Glu	tcc Ser	att Ile 260	ctg Leu	acc Thr	acc Thr	atg Met	gta Val 265	cag Gln	gtg Val	gct Ala	ggg Gly 270	aaa Lys 270	gtt Val	cag Gln	816
gaa Glu	gtg Val	ctg Leu 275	aag Lys	gag Glu	cca Pro	gat Asp	ggg Gly 280	ggg Gly	ctg Leu	gtt Val	gtg Val	ctg Leu 285	agt Ser	gga Gly	ggg Gly	864
ggc Gly 290	acc Thr	tct Ser	ggc Gly	cgg Arg	atg Met	gca Ala 295	ttc Phe	ctc Leu	atg Met	tcg Ser	gtg Val 300	tcc Ser	ttt Phe	aat Asn	cag Gln	912
ctg Leu 305	atg Met	aaa Lys	ggt Gly	ctg Leu	gga Gly 310	cag Gln	aaa Lys	cct Pro	ctt Leu	tac Tyr 315	acc Thr	tac Tyr	ctc Leu	att Ile	gca Ala 320	960
ggt Gly	ggt Gly	gac Asp	agg Arg	tct Ser 325	gtg Val	gtg Val	gcc Ala	tct Ser	agg Arg 330	gag Glu	ggg Gly	aca Thr	gaa Glu	gat Asp 335	agt Ser	1008
gcc Ala	ttg Leu	cac His	ggg Gly 340	att Ile	gag Glu	gaa Glu	ctg Leu	aag Lys 345	aag Lys	gtg Val	gct Ala	gcc Ala	ggg Gly 350	aag Lys	aag Lys	1056
aga Arg	gtg Val	att Ile 355	gtc Val	att Ile	ggc Gly	att Ile	tct Ser 360	gtg Val	gga Gly	ctc Leu	tct Ser	gct Ala 365	ccc Pro	ttt Phe	gtg Val	1104
gca Ala 370	ggc Gly	cag Gln	atg Met	gac Asp	tgc Cys	tgc Cys 375	atg Met	aac Asn	aac Asn	aca Thr	gct Ala 380	gtc Val	ttc Phe	ttg Leu	cca Pro	1152
gtc Val 385	ctg Leu	gtt Val	ggc Gly	ttc Phe	aat Asn 390	cca Pro	gtg Val	agc Ser	atg Met	gcc Ala 395	aga Arg	aat Asn	gac Asp	ccc Pro	att Ile 400	1200
gaa Glu	gac Asp	tgg Trp	agt Ser	tca Ser 405	aca Thr	ttc Phe	cga Arg	caa Gln	gta Val 410	gca Ala	gag Glu	cgg Arg	atg Met	cag Gln 415	aaa Lys	1248
atg Met	cag Gln	gag Glu	aaa Lys 420	cag Gln	aaa Lys	gct Ala	ttt Phe	gtg Val 425	ctc Leu	aat Asn	cct Pro	gcc Ala	atc Ile 430	ggg Gly	ccc Pro	1296
gag Glu	ggt Gly	ctc Leu 435	agc Ser	ggc Gly	tcc Ser	tcc Ser	cgg Arg 440	atg Met	aaa Lys	ggt Gly	gga Gly	agt Ser 445	gcc Ala	acc Thr	aag Lys	1344
att Ile 450	ctg Leu	ctg Leu	gaa Glu	acc Thr	ctg Leu	tta Leu 455	tta Leu	gca Ala	gcc Ala	cat His	aag Lys 460	act Thr	gtg Val	gac Asp	cag Gln	1392
ggc Gly 465	att Ile	gca Ala	gca Ala	tct Ser	caa Gln 470	aga Arg	tgc Cys	ctc Leu	ctg Leu	gaa Glu 475	atc Ile	ttg Leu	cgg Arg	aca Thr	ttt Phe 480	1440

ES 2 646 388 A1

gag Glu	cga Arg	gct Ala	cat His	cag Gln 485	gtg Val	acc Thr	tac Tyr	agc Ser	caa Gln 490	agc Ser	ccc Pro	aag Lys	att Ile	gcc Ala 495	acc Thr	1488
ctg Leu	atg Met	aag Lys	agt Ser 500	gtc Val	agc Ser	acc Thr	agt Ser	ctg Leu 505	gag Glu	aag Lys	aaa Lys	ggc Gly	cac His 510	gtg Val	tac Tyr	1536
ctg Leu	gtt Val	ggc Gly 515	tgg Trp	cag Gln	acc Thr	ctg Leu	ggt Gly 520	atc Ile	att Ile	gcc Ala	atc Ile	atg Met 525	gat Asp	gga Gly	gta Val	1584
gag Glu	tgc Cys 530	atc Ile	cac His	acc Thr	ttt Phe	ggt Gly 535	gct Ala	gat Asp	ttc Phe	cga Arg	gat Asp 540	gtc Val	cgt Arg	ggc Gly	ttt Phe	1632
ctc Leu 545	att Ile	ggt Gly	gat Asp	cac His	agt Ser 550	gac Asp	atg Met	ttt Phe	aac Asn	cag Gln 555	aag Lys	gct Ala	gag Glu	ctc Leu	acc Thr 560	1680
aac Asn	cag Gln	ggt Gly	ccc Pro	cag Gln 565	ttc Phe	acc Thr	ttc Phe	tcc Ser	cag Gln 570	gag Glu	gac Asp	ttc Phe	ccg Pro	act Thr 575	tcc Ser	1728
atc Ile	ctt Leu	ccc Pro	tct Ser 580	ctc Leu	acg Thr	gaa Glu	atc Ile	gat Asp 585	act Thr	gtg Val	gtc Val	ttc Phe	att Ile 590	ttc Phe	acc Thr	1776
ctg Leu	gat Asp	gac Asp 595	aac Asn	ctc Leu	acg Thr	gag Glu	gtg Val 600	cag Gln	act Thr	ata Ile	gtg Val	gag Glu 605	cag Gln	gtg Val	aaa Lys	1824
gag Glu	aag Lys 610	acc Thr	aac Asn	cac His	atc Ile	cag Gln 615	gcc Ala	ctg Leu	gca Ala	cac His	agc Ser 620	acc Thr	gtg Val	ggt Gly	cag Gln	1872
acc Thr 625	ttg Leu	ccg Pro	atc Ile	cct Pro	ctg Leu 630	aag Lys	aag Lys	ctc Leu	ttt Phe	ccc Pro 635	tcc Ser	atc Ile	atc Ile	agc Ser	atc Ile 640	1920
aca Thr	tgg Trp	cca Pro	ctg Leu	ctt Leu 645	ttc Phe	ttt Phe	gaa Glu	tat Tyr	gaa Glu 650	ggg Gly	aac Asn	ttc Phe	atc Ile	cag Gln 655	aag Lys	1968
ttc Phe	cag Gln	cgt Arg	gag Glu 660	cta Leu	agc Ser	acc Thr	aaa Lys	tgg Trp 665	gtg Val	ctg Leu	aat Asn	aca Thr	gtg Val 670	agt Ser	aca Thr	2016
ggt Gly	gct Ala	cat His 675	gtg Val	ctt Leu	ctt Leu	ggt Gly	aag Lys 680	atc Ile	cta Leu	caa Gln	aac Asn	cac His 685	atg Met	ttg Leu	gac Asp	2064
ctt Leu	cgg Arg 690	att Ile	agc Ser	aac Asn	tcc Ser	aag Lys 695	ctc Leu	ttc Phe	tgg Trp	cgg Arg	gcg Ala 700	ctg Leu	gcc Ala	atg Met	ctg Leu	2112
cag Gln 705	cgg Arg	ttc Phe	tct Ser	gga Gly 710	cag Gln 710	tcc Ser	aag Lys	gct Ala	cga Arg	tgc Cys 715	atc Ile	gag Glu	agc Ser	ctc Leu	ctc Leu 720	2160
cga Arg	gcg Ala	atc Ile	cac His	ttt Phe 725	ccc Pro	cag Gln	cca Pro	ctg Leu	tca Ser 730	gat Asp	gat Asp	att Ile	cgg Arg	gct Ala 735	gct Ala	2208
ccc Pro	atc Ile	tcc Ser	tgc Cys 740	cgt Arg	gtc Val	cag Gln	gtt Val	gca Ala 745	cat His	gag Glu	aag Lys	gaa Glu	cag Gln 750	gtg Val	ata Ile	2256

ES 2 646 388 A1

ccc atc gcc ttg ctg agc ctc cta ttc cgg tgc tcg atc act gag gct 2304
 Pro Ile Ala Leu Leu Ser Leu Leu Phe Arg Cys Ser Ile Thr Glu Ala
 755 760 765

cag gca cac ctg gct gca gct cct tct gtc tgt gag gct gtc agg agt 2352
 Gln Ala His Leu Ala Ala Ala Pro Ser Val Cys Glu Ala Val Arg Ser
 770 775 780

gct ctt gct ggg cca ggt cag aag cgc act gcg gac ccc ctc gag atc 2400
 Ala Leu Ala Gly Pro Gly Gln Lys Arg Thr Ala Asp Pro Leu Glu Ile
 785 790 800

cta gag cct gac gtt cag tga 2421
 Leu Glu Pro Asp Val Gln
 805

<210> 51
 <211> 806
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 51

Met Lys Leu Leu Ser Ser Ile Glu Gln Ala Cys Asp Ile Cys Arg Leu
 1 5 10 15

Lys Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu
 20 25 30

Lys Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro
 35 40 45

Leu Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu
 50 55 60

Glu Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile
 65 70 75 80

Leu Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu
 85 90 95

Phe Val Gln Asp Asn Val Asn Lys Asp Ala Val Thr Asp Arg Leu Ala
 100 105 110

Ser Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gln His Arg Ile Ser
 115 120 125

Ala Thr Ser Ser Ser Glu Glu Ser Ser Asn Lys Gly Gln Arg Gln Leu
 130 135 140

Thr Val Ser Pro Glu Phe Val Ile Arg Leu Thr Ile Gly Arg Ala Ala
 145 150 155 160

Ile Met Glu Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu His Met Ala
 165 170 175

ES 2 646 388 A1

Met Glu Ala Glu Phe Met Pro Gly Thr Lys Arg Phe Gln His Val Ile
180 185 190

Glu Thr Pro Glu Pro Gly Lys Trp Glu Leu Ser Gly Tyr Glu Ala Ala
195 200 205

Val Pro Ile Thr Glu Lys Ser Asn Pro Leu Thr Gln Asp Leu Asp Lys
210 215 220

Ala Asp Ala Glu Asn Ile Val Arg Leu Leu Gly Gln Cys Asp Ala Glu
225 230 235 240

Ile Phe Gln Glu Glu Gly Gln Ala Leu Ser Thr Tyr Gln Arg Leu Tyr
245 250 255

Ser Glu Ser Ile Leu Thr Thr Met Val Gln Val Ala Gly Lys Val Gln
260 265 270

Glu Val Leu Lys Glu Pro Asp Gly Gly Leu Val Val Leu Ser Gly Gly
275 280 285

Gly Thr Ser Gly Arg Met Ala Phe Leu Met Ser Val Ser Phe Asn Gln
290 295 300

Leu Met Lys Gly Leu Gly Gln Lys Pro Leu Tyr Thr Tyr Leu Ile Ala
305 310 315 320

Gly Gly Asp Arg Ser Val Val Ala Ser Arg Glu Gly Thr Glu Asp Ser
325 330 335

Ala Leu His Gly Ile Glu Glu Leu Lys Lys Val Ala Ala Gly Lys Lys
340 345 350

Arg Val Ile Val Ile Gly Ile Ser Val Gly Leu Ser Ala Pro Phe Val
355 360 365

Ala Gly Gln Met Asp Cys Cys Met Asn Asn Thr Ala Val Phe Leu Pro
370 375 380

Val Leu Val Gly Phe Asn Pro Val Ser Met Ala Arg Asn Asp Pro Ile
385 390 395 400

Glu Asp Trp Ser Ser Thr Phe Arg Gln Val Ala Glu Arg Met Gln Lys
405 410 415

Met Gln Glu Lys Gln Lys Ala Phe Val Leu Asn Pro Ala Ile Gly Pro
420 425 430

Glu Gly Leu Ser Gly Ser Ser Arg Met Lys Gly Gly Ser Ala Thr Lys
435 440 445

ES 2 646 388 A1

Ile Leu Leu Glu Thr Leu Leu Leu Ala Ala His Lys Thr Val Asp Gln
 450 455 460

Gly Ile Ala Ala Ser Gln Arg Cys Leu Leu Glu Ile Leu Arg Thr Phe
 465 470 475 480

Glu Arg Ala His Gln Val Thr Tyr Ser Gln Ser Pro Lys Ile Ala Thr
 485 490 495

Leu Met Lys Ser Val Ser Thr Ser Leu Glu Lys Lys Gly His Val Tyr
 500 505 510 515

Leu Val Gly Trp Gln Thr Leu Gly Ile Ile Ala Ile Met Asp Gly Val
 515 520 525

Glu Cys Ile His Thr Phe Gly Ala Asp Phe Arg Asp Val Arg Gly Phe
 530 535 540 545

Leu Ile Gly Asp His Ser Asp Met Phe Asn Gln Lys Ala Glu Leu Thr
 545 550 555 560

Asn Gln Gly Pro Gln Phe Thr Phe Ser Gln Glu Asp Phe Pro Thr Ser
 565 570 575

Ile Leu Pro Ser Leu Thr Glu Ile Asp Thr Val Val Phe Ile Phe Thr
 580 585 590

Leu Asp Asp Asn Leu Thr Glu Val Gln Thr Ile Val Glu Gln Val Lys
 595 600 605

Glu Lys Thr Asn His Ile Gln Ala Leu Ala His Ser Thr Val Gly Gln
 610 615 620

Thr Leu Pro Ile Pro Leu Lys Lys Leu Phe Pro Ser Ile Ile Ser Ile
 625 630 635 640

Thr Trp Pro Leu Leu Phe Phe Glu Tyr Glu Gly Asn Phe Ile Gln Lys
 645 650 655

Phe Gln Arg Glu Leu Ser Thr Lys Trp Val Leu Asn Thr Val Ser Thr
 660 665 670

Gly Ala His Val Leu Leu Gly Lys Ile Leu Gln Asn His Met Leu Asp
 675 680 685

Leu Arg Ile Ser Asn Ser Lys Leu Phe Trp Arg Ala Leu Ala Met Leu
 690 695 700

Gln Arg Phe Ser Gly Gln Ser Lys Ala Arg Cys Ile Glu Ser Leu Leu
 705 710 715 720

ES 2 646 388 A1

Arg Ala Ile His Phe Pro Gln Pro Leu Ser Asp Asp Ile Arg Ala Ala
 725 730 735

Pro Ile Ser Cys Arg Val Gln Val Ala His Glu Lys Glu Gln Val Ile
 740 745 750

Pro Ile Ala Leu Leu Ser Leu Leu Phe Arg Cys Ser Ile Thr Glu Ala
 755 760 765

Gln Ala His Leu Ala Ala Ala Pro Ser Val Cys Glu Ala Val Arg Ser
 770 775 780

Ala Leu Ala Gly Pro Gly Gln Lys Arg Thr Ala Asp Pro Leu Glu Ile
 785 790 795 800

Leu Glu Pro Asp Val Gln
 805

<210> 52
 <211> 1974
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> secuencia nucleotídica que codifica para GAD-GK-3xPTAP

<220>
 <221> CDS
 <222> (1)..(1974)

<400> 52	
atg gat aaa gcg gaa tta att ccc gag cct cca aaa aag aag aga aag	48
Met Asp Lys Ala Glu Leu Ile Pro Glu Pro Pro Lys Lys Lys Arg Lys	
1 5 10 15	
gtc gaa ttg ggt acc gcc gcc aat ttt aat caa agt ggg aat att gct	96
Val Glu Leu Gly Thr Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala	
20 25 30	
gat agc tca ttg tcc ttc act ttc act aac agt agc aac ggt ccg aac	144
Asp Ser Ser Leu Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn	
35 40 45	
ctc ata aca act caa aca aat tct caa gcg ctt tca caa cca att gcc	192
Leu Ile Thr Thr Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala	
50 55 60	
tcc tct aac gtt cat gat aac ttc atg aat aat gaa atc acg gct agt	240
Ser Ser Asn Val His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser	
65 70 75 80	
aaa att gat gat ggt aat aat tca aaa cca ctg tca cct ggt tgg acg	288
Lys Ile Asp Asp Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr	
85 90 95	
gac caa act gcg tat aac gcg ttt gga atc act aca ggg atg ttt aat	336
Asp Gln Thr Ala Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn	
100 105 110	
acc act aca atg gat gat gta tat aac tat cta ttc gat gat gaa gat	384
Thr Thr Thr Met Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp	
115 120 125	

ES 2 646 388 A1

acc Thr	cca Pro 130	cca Pro	aac Asn	cca Pro	aaa Lys	aaa Lys 135	gag Glu	atc Ile	tgt Cys	atg Met	gct Ala 140	tac Tyr	cca Pro	tac Tyr	gat Asp	432
gtt Val 145	cca Pro	gat Asp	tac Tyr	gct Ala	agc Ser 150	ttg Leu	ggt Gly	ggt Gly	cat His	atg Met 155	gcc Ala	atg Met	gag Glu	gcc Ala	ccg Pro 160	480
ggg Gly	atc Ile	ctg Leu	atg Met	ctg Leu 165	gac Asp	gac Asp	aga Arg	gcc Ala	agg Arg 170	atg Met	gag Glu	gcc Ala	gcc Ala	aag Lys 175	aag Lys	528
gag Glu	aag Lys	gta Val	gag Glu 180	cag Gln	atc Ile	ctg Leu	gca Ala	gag Glu 185	ttc Phe	cag Gln	ctg Leu	cag Gln	gag Glu 190	gag Glu	gac Asp	576
ctg Leu	aag Lys	aag Lys 195	gtg Val	atg Met	aga Arg	cgg Arg	atg Met 200	cag Gln	aag Lys	gag Glu	atg Met	gac Asp 205	cgc Arg	ggc Gly	ctg Leu	624
agg Arg	ctg Leu 210	gag Glu	acc Thr	cat His	gaa Glu	gag Glu 215	gcc Ala	agt Ser	gtg Val	aag Lys	atg Met 220	ctg Leu	ccc Pro	acc Thr	tac Tyr	672
gtg Val 225	cgc Arg	tcc Ser	acc Thr	cca Pro	gaa Glu 230	ggc Gly	tca Ser	gaa Glu	gtc Val	ggg Gly 235	gac Asp	ttc Phe	ctc Leu	tcc Ser	ctg Leu 240	720
gac Asp	ctg Leu	ggt Gly	ggc Gly	act Thr 245	aac Asn	ttc Phe	agg Arg	gtg Val	atg Met 250	ctg Leu	gtg Val	aag Lys	gtg Val	gga Gly 255	gaa Glu	768
ggt Gly	gag Glu	gag Glu	ggg Gly 260	cag Gln	tgg Trp	agc Ser	gtg Val	aag Lys 265	acc Thr	aaa Lys	cac His	cag Gln	atg Met 270	tac Tyr	tcc Ser	816
atc Ile	ccc Pro	gag Glu 275	gac Asp	gcc Ala	atg Met	acc Thr	ggc Gly 280	act Thr	gct Ala	gag Glu	atg Met	ctc Leu 285	ttc Phe	gac Asp	tac Tyr	864
atc Ile	tct Ser 290	gag Glu	tgc Cys	atc Ile	tcc Ser	gac Asp 295	ttc Phe	ctg Leu	gac Asp	aag Lys	cat His 300	cag Gln	atg Met	aaa Lys	cac His	912
aag Lys 305	aag Lys	ctg Leu	ccc Pro	ctg Leu	ggc Gly 310	ttc Phe	acc Thr	ttc Phe	tcc Ser	ttt Phe 315	cct Pro	gtg Val	agg Arg	cac His	gaa Glu 320	960
gac Asp	atc Ile	gat Asp	aag Lys	ggc Gly 325	atc Ile	ctt Leu	ctc Leu	aac Asn	tgg Trp 330	acc Thr	aag Lys	ggc Gly	ttc Phe	aag Lys 335	gcc Ala	1008
tca Ser	gga Gly	gca Ala	gaa Glu 340	ggg Gly	aac Asn	aat Asn	gtc Val	gtg Val 345	ggg Gly	ctt Leu	ctg Leu	cga Arg	gac Asp 350	gct Ala	atc Ile	1056
aaa Lys	cgg Arg	aga Arg 355	ggg Gly	gac Asp	ttt Phe	gaa Glu	atg Met 360	gat Asp	gtg Val	gtg Val	gca Ala	atg Met 365	gtg Val	aat Asn	gac Asp	1104
acg Thr	gtg Val 370	gcc Ala	acg Thr	atg Met	atc Ile	tcc Ser 375	tgc Cys	tac Tyr	tac Tyr	gaa Glu	gac Asp 380	cat His	cag Gln	tgc Cys	gag Glu	1152
gtc Val 385	ggc Gly	atg Met	atc Ile	gtg Val	ggc Gly 390	acg Thr	ggc Gly	tgc Cys	aat Asn	gcc Ala 395	tgc Cys	tac Tyr	atg Met	gag Glu	gag Glu 400	1200

ES 2 646 388 A1

atg	cag	aat	gtg	gag	ctg	gtg	gag	ggg	gac	gag	ggc	cgc	atg	tgc	gtc	1248
Met	Gln	Asn	Val	Glu	Leu	Val	Glu	Gly	Asp	Glu	Gly	Arg	Met	Cys	Val	
				405					410					415		
aat	acc	gag	tgg	ggc	gcc	ttc	ggg	gac	tcc	ggc	gag	ctg	gac	gag	ttc	1296
Asn	Thr	Glu	Trp	Gly	Ala	Phe	Gly	Asp	Ser	Gly	Glu	Leu	Asp	Glu	Phe	
			420					425					430			
ctg	ctg	gag	tat	gac	cgc	ctg	gtg	gac	gag	agc	tct	gca	aac	ccc	ggt	1344
Leu	Leu	Glu	Tyr	Asp	Arg	Leu	Val	Asp	Glu	Ser	Ser	Ala	Asn	Pro	Gly	
			435				440					445				
cag	cag	ctg	tat	gag	aag	ctc	ata	ggt	ggc	aag	tac	atg	ggc	gag	ctg	1392
Gln	Gln	Leu	Tyr	Glu	Lys	Leu	Ile	Gly	Gly	Lys	Tyr	Met	Gly	Glu	Leu	
		450				455					460					
gtg	cgg	ctt	gtg	ctg	ctc	agg	ctc	gtg	gac	gaa	aac	ctg	ctc	ttc	cac	1440
Val	Arg	Leu	Val	Leu	Leu	Arg	Leu	Val	Asp	Glu	Asn	Leu	Leu	Phe	His	
					470					475					480	
ggg	gag	gcc	tcc	gag	cag	ctg	cgc	aca	cgc	gga	gcc	ttc	gag	acg	cgc	1488
Gly	Glu	Ala	Ser	Glu	Gln	Leu	Arg	Thr	Arg	Gly	Ala	Phe	Glu	Thr	Arg	
				485					490					495		
ttc	gtg	tcg	cag	gtg	gag	agc	gac	acg	ggc	gac	cgc	aag	cag	atc	tac	1536
Phe	Val	Ser	Gln	Val	Glu	Ser	Asp	Thr	Gly	Asp	Arg	Lys	Gln	Ile	Tyr	
			500					505					510			
aac	atc	ctg	agc	acg	ctg	ggg	ctg	cga	ccc	tcg	acc	acc	gac	tgc	gac	1584
Asn	Ile	Leu	Ser	Thr	Leu	Gly	Leu	Arg	Pro	Ser	Thr	Thr	Asp	Cys	Asp	
		515					520					525				
atc	gtg	cgc	cgc	gcc	tgc	gag	agc	gtg	tct	acg	cgc	gct	gcg	cac	atg	1632
Ile	Val	Arg	Arg	Ala	Cys	Glu	Ser	Val	Ser	Thr	Arg	Ala	Ala	His	Met	
						535					540					
tgc	tcg	gcg	ggg	ctg	gcg	ggc	gtc	atc	aac	cgc	atg	cgc	gag	agc	cgc	1680
Cys	Ser	Ala	Gly	Leu	Ala	Gly	Val	Ile	Asn	Arg	Met	Arg	Glu	Ser	Arg	
					550					555					560	
agc	gag	gac	gta	atg	cgc	atc	act	gtg	ggc	gtg	gat	ggc	tcc	gtg	tac	1728
Ser	Glu	Asp	Val	Met	Arg	Ile	Thr	Val	Gly	Val	Asp	Gly	Ser	Val	Tyr	
				565					570					575		
aag	ctg	cac	ccc	agc	ttc	aag	gag	cgg	ttc	cat	gcc	agc	gtg	cgc	agg	1776
Lys	Leu	His	Pro	Ser	Phe	Lys	Glu	Arg	Phe	His	Ala	Ser	Val	Arg	Arg	
			580					585					590			
ctg	acg	ccc	agc	tgc	gag	atc	acc	ttc	atc	gag	tcg	gag	gag	ggc	agt	1824
Leu	Thr	Pro	Ser	Cys	Glu	Ile	Thr	Phe	Ile	Glu	Ser	Glu	Glu	Gly	Ser	
							600					605				
ggc	cgg	ggc	gcg	gcc	ctg	gtc	tcg	gcg	gtg	gcc	tgt	aag	aag	gcc	tgt	1872
Gly	Arg	Gly	Ala	Ala	Leu	Val	Ser	Ala	Val	Ala	Cys	Lys	Lys	Ala	Cys	
						615					620					
atg	ctg	ggc	cag	tgg	atc	cga	att	cga	gct	cga	gag	cca	gaa	ccc	aca	1920
Met	Leu	Gly	Gln	Trp	Ile	Arg	Ile	Arg	Ala	Arg	Glu	Pro	Glu	Pro	Thr	
					630					635					640	
gca	ccg	cct	gag	cct	acc	gcc	cca	ccc	gaa	ccg	acg	gcg	cct	cca	gct	1968
Ala	Pro	Pro	Glu	Pro	Thr	Ala	Pro	Pro	Glu	Pro	Thr	Ala	Pro	Pro	Ala	
				645					650					655		
gag	taa															1974
Glu																

ES 2 646 388 A1

<210> 53
 <211> 657
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 53

Met Asp Lys Ala Glu Leu Ile Pro Glu Pro Pro Lys Lys Lys Arg Lys
 1 5 10 15

Val Glu Leu Gly Thr Ala Ala Asn Phe Asn Gln Ser Gly Asn Ile Ala
 20 25 30

Asp Ser Ser Leu Ser Phe Thr Phe Thr Asn Ser Ser Asn Gly Pro Asn
 35 40 45

Leu Ile Thr Thr Gln Thr Asn Ser Gln Ala Leu Ser Gln Pro Ile Ala
 50 55 60

Ser Ser Asn Val His Asp Asn Phe Met Asn Asn Glu Ile Thr Ala Ser
 65 70 75 80

Lys Ile Asp Asp Gly Asn Asn Ser Lys Pro Leu Ser Pro Gly Trp Thr
 85 90 95

Asp Gln Thr Ala Tyr Asn Ala Phe Gly Ile Thr Thr Gly Met Phe Asn
 100 105 110

Thr Thr Thr Met Asp Asp Val Tyr Asn Tyr Leu Phe Asp Asp Glu Asp
 115 120 125

Thr Pro Pro Asn Pro Lys Lys Glu Ile Cys Met Ala Tyr Pro Tyr Asp
 130 135 140

Val Pro Asp Tyr Ala Ser Leu Gly Gly His Met Ala Met Glu Ala Pro
 145 150 155 160

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

Glu Lys Val Glu Gln Ile Leu Ala Glu Phe Gln Leu Gln Glu Glu Asp
 180 185 190

Leu Lys Lys Val Met Arg Arg Met Gln Lys Glu Met Asp Arg Gly Leu
 195 200 205

Arg Leu Glu Thr His Glu Glu Ala Ser Val Lys Met Leu Pro Thr Tyr
 210 215 220

Val Arg Ser Thr Pro Glu Gly Ser Glu Val Gly Asp Phe Leu Ser Leu
 225 230 235 240

ES 2 646 388 A1

Asp Leu Gly Gly Thr Asn Phe Arg Val Met Leu Val Lys Val Gly Glu
 245 250 255
 Gly Glu Glu Gly Gln Trp Ser Val Lys Thr Lys His Gln Met Tyr Ser
 260 265 270
 Ile Pro Glu Asp Ala Met Thr Gly Thr Ala Glu Met Leu Phe Asp Tyr
 275 280 285
 Ile Ser Glu Cys Ile Ser Asp Phe Leu Asp Lys His Gln Met Lys His
 290 295 300
 Lys Lys Leu Pro Leu Gly Phe Thr Phe Ser Phe Pro Val Arg His Glu
 305 310 315 320
 Asp Ile Asp Lys Gly Ile Leu Leu Asn Trp Thr Lys Gly Phe Lys Ala
 325 330 335
 Ser Gly Ala Glu Gly Asn Asn Val Val Gly Leu Leu Arg Asp Ala Ile
 340 345 350
 Lys Arg Arg Gly Asp Phe Glu Met Asp Val Val Ala Met Val Asn Asp
 355 360 365
 Thr Val Ala Thr Met Ile Ser Cys Tyr Tyr Glu Asp His Gln Cys Glu
 370 375 380
 Val Gly Met Ile Val Gly Thr Gly Cys Asn Ala Cys Tyr Met Glu Glu
 385 390 395 400
 Met Gln Asn Val Glu Leu Val Glu Gly Asp Glu Gly Arg Met Cys Val
 405 410 415
 Asn Thr Glu Trp Gly Ala Phe Gly Asp Ser Gly Glu Leu Asp Glu Phe
 420 425 430
 Leu Leu Glu Tyr Asp Arg Leu Val Asp Glu Ser Ser Ala Asn Pro Gly
 435 440 445
 Gln Gln Leu Tyr Glu Lys Leu Ile Gly Gly Lys Tyr Met Gly Glu Leu
 450 455 460
 Val Arg Leu Val Leu Leu Arg Leu Val Asp Glu Asn Leu Leu Phe His
 465 470 475 480
 Gly Glu Ala Ser Glu Gln Leu Arg Thr Arg Gly Ala Phe Glu Thr Arg
 485 490 495
 Phe Val Ser Gln Val Glu Ser Asp Thr Gly Asp Arg Lys Gln Ile Tyr
 500 505 510

ES 2 646 388 A1

Asn Ile Leu Ser Thr Leu Gly Leu Arg Pro Ser Thr Thr Asp Cys Asp
 515 520 525

Ile Val Arg Arg Ala Cys Glu Ser Val Ser Thr Arg Ala Ala His Met
 530 535 540

Cys Ser Ala Gly Leu Ala Gly Val Ile Asn Arg Met Arg Glu Ser Arg
 545 550 555 560

Ser Glu Asp Val Met Arg Ile Thr Val Gly Val Asp Gly Ser Val Tyr
 565 570 575

Lys Leu His Pro Ser Phe Lys Glu Arg Phe His Ala Ser Val Arg Arg
 580 585 590

Leu Thr Pro Ser Cys Glu Ile Thr Phe Ile Glu Ser Glu Glu Gly Ser
 595 600 605

Gly Arg Gly Ala Ala Leu Val Ser Ala Val Ala Cys Lys Lys Ala Cys
 610 615 620

Met Leu Gly Gln Trp Ile Arg Ile Arg Ala Arg Glu Pro Glu Pro Thr
 625 630 635 640

Ala Pro Pro Glu Pro Thr Ala Pro Pro Glu Pro Thr Ala Pro Pro Ala
 645 650 655

Glu