



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 364 923**

51 Int. Cl.:

**C12N 15/86** (2006.01)

**C12N 15/82** (2006.01)

**C07K 14/705** (2006.01)

**C07K 14/72** (2006.01)

**C12N 15/10** (2006.01)

**A01K 67/027** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **01975606 .3**

96 Fecha de presentación : **28.09.2001**

97 Número de publicación de la solicitud: **1334200**

97 Fecha de publicación de la solicitud: **13.08.2003**

54 Título: **Sistema de regulación de múltiples genes inducibles.**

30 Prioridad: **03.10.2000 US 237446 P**  
**27.09.2001 US 965697**

45 Fecha de publicación de la mención BOPI:  
**16.09.2011**

45 Fecha de la publicación del folleto de la patente:  
**16.09.2011**

73 Titular/es: **INTREXON CORPORATION**  
**1872 Pratt Drive Suite 1400**  
**Blacksburg, Virginia 24060, US**

72 Inventor/es: **Dhadialla, Tarlochan, Singh;**  
**Cress, Dean, Ervin;**  
**Carlson, Glenn, Richard;**  
**Hormann, Robert, Eugene;**  
**Palli, Subba, Reddy;**  
**Kudla, Arthur, John;**  
**Herzig, Ronald, Phillip, Jr. y**  
**Philip, Mohan**

74 Agente: **Curell Aguilá, Marcelino**

**ES 2 364 923 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Sistema de regulación de múltiples genes inducibles.

5 **Campo de la invención**

La presente invención se refiere al campo de la biotecnología o de la ingeniería genética. Más específicamente, la presente invención se refiere a un sistema de regulación de múltiples genes inducibles que funciona dentro de las células, tejidos u organismos no humanos controlando simultáneamente la expresión cuantitativa de dos o más genes.

**Antecedentes de la invención**

Se citan en la presente memoria diversas publicaciones. Sin embargo, la cita de cualquier referencia en la presente memoria no debe interpretarse como una admisión de que dicha referencia se encuentra disponible como "técnica anterior" a la presente solicitud.

En el campo de la ingeniería genética, el control preciso de la expresión génica es una herramienta valiosa para estudiar, manipular y controlar el desarrollo y otros procesos fisiológicos. La expresión génica es un proceso biológico complejo que implica varias interacciones específicas proteína-proteína. Con el fin de que la expresión génica resulte inducida, de manera que produzca el ARN necesario como primera etapa en la síntesis de proteínas, un activador transcripcional debe aproximarse a un promotor que controla la transcripción génica. Típicamente, el activador transcripcional mismo se asocia a una proteína que presenta por lo menos un dominio de unión a ADN que se une a sitios de unión a ADN presentes en las regiones promotoras de los genes. De esta manera, para que se produzca la expresión génica, debe transportarse una proteína que comprende un dominio de unión a ADN y un dominio de transactivación situado a una distancia apropiada del dominio de unión a ADN hasta la posición correcta en la región promotora del gen.

El enfoque transgénico tradicional utiliza un promotor que es específico de un tipo celular para controlar la expresión de un transgén de diseño. En primer lugar se incorpora en un genoma huésped un constructo de ADN que contiene el transgén. Al resultar inducido por un activador transcripcional, se produce la expresión del transgén en un tipo celular dado.

Un sistema de regulación de múltiples genes es un sistema que permite la regulación simultánea y cuantitativa de muchos genes diferentes en la misma célula, tejido u organismo. En la actualidad, en aplicaciones que van desde el análisis del genoma humano pasando por la proteómica hasta la producción de cantidades a gran escala de proteínas y las terapias génicas, no existe ninguna tecnología para regular simultáneamente más de un gen en la misma célula. La regulación génica resulta crítica en todas estas aplicaciones debido a que garantiza que, sea cual sea el gen que se analice, éste se encuentre bajo un control preciso y cuantitativo, y por lo tanto sean cuales sean los resultados obtenidos, estos se encuentran vinculados directa y específicamente a ese gen y no a otros. Sin embargo, en la actualidad la regulación génica se encuentra limitada a un gen cada vez, y esto supone una limitación cualitativa y cuantitativa significativa. El control paralelo de múltiples genes en la misma célula permite analizar fenómenos biológicos mucho más complejos, en los que se encuentran implicados múltiples genes, así como crear nuevas aplicaciones terapéuticas.

Otro medio para regular la expresión de genes foráneos en células utiliza promotores inducibles. Entre los ejemplos de este tipo de promotores se incluyen el promotor PR1-a, los sistemas procarióticos de represor-operador, los sistemas basados en moléculas inmunosupresoras y los sistemas de eucariotas superiores de activación de la transcripción.

El promotor PR1-a del tabaco resulta inducido durante la respuesta sistémica de adquisición de resistencia tras el ataque de patógenos. La utilización de PR1-a puede ser limitada debido a que con frecuencia responde a materiales endógenos y a factores externos tales como patógenos, radiación UV-B y contaminantes. Se han descrito sistemas de regulación génica basados en promotores inducidos por choque térmico, interferón y metales pesados (Wurn *et al.*, Proc. Natl. Acad. Sci. USA 83:5414-5418, 1986; Arnheiter *et al.*, Cell 62:51-61, 1990; Filmus *et al.*, Nucleic Acids Research 20:27550-27560, 1992). Sin embargo, estos sistemas presentan limitaciones debido a su efecto sobre la expresión de genes no diana. Estos sistemas también se encuentran débilmente regulados ("leaky").

Los sistemas procarióticos de represor-operador utilizan proteínas bacterianas represoras y las secuencias únicas de ADN operador a las que se unen. Los sistemas de represor-operador tanto tetraciclina ("Tet") como lactosa ("Lac") de la bacteria *Escherichia coli* han sido utilizados en plantas y animales para controlar la expresión génica. En el sistema Tet, la tetraciclina se une a la proteína represora TetR resultando en un cambio conformacional que libera la proteína represora del operador que, como resultado, permite que se produzca la transcripción. En el sistema Lac, un operón Lac resulta activado en respuesta a la presencia de lactosa, o de análogos sintéticos tales como isopropil-b-D-tiogalactósido. Por desgracia, la utilización de dichos sistemas se encuentra restringida por la química inestable de los ligandos, es decir la tetraciclina y la lactosa, su toxicidad, su presencia natural o los niveles

relativamente altos necesarios para la inducción o la represión. Por motivos similares, la utilidad de dichos sistemas en animales es limitada.

Las moléculas inmunosupresoras tales como FK506, la rapamicina y la ciclosporina A pueden unirse a las inmunofilinas FKBP12, a la ciclofilina, etc. Utilizando esta información, se ha diseñado una estrategia general para reunir dos proteínas cualesquiera simplemente colocando FK506 sobre cada una de las dos proteínas o colocando FK506 sobre una y ciclosporina A, sobre otra. A continuación, puede utilizarse un homodímero sintético de FK506 (FK1012) o un compuesto resultante de la fusión de FK506-ciclosporina (FKCsA) para inducir la dimerización de dichas moléculas (Spencer *et al.*, Science 262:1019-24, 1993; Belshaw *et al.*, Proc. Natl. Acad. Sci. USA 93:4604-7, 1996). Se ha utilizado el dominio Gal4 de unión a ADN fusionado a FKBP12 y el dominio activador VP16 fusionado a ciclofilina y el compuesto FKCsA para mostrar la heterodimerización y la activación de un gen informador bajo el control de un promotor que contiene sitios de unión de Gal4. Por desgracia, este sistema incluye inmunosupresores que presentan efectos secundarios no deseados, lo que limita su utilización para diversas aplicaciones de sistema de regulación génica de mamíferos.

Los sistemas de activación transcripcional de los eucariotas superiores tales como los sistemas de receptores de hormonas esteroideas también han sido utilizados. Los receptores de hormonas esteroideas son miembros de la superfamilia de receptores nucleares y se encuentran en células de vertebrado y de invertebrado. Por desgracia, la utilización de compuestos esteroideos que activan los receptores para la regulación de la expresión génica, en particular en plantas y mamíferos, se encuentra limitada por su implicación en muchas otras rutas biológicas naturales en dichos organismos. Para superar estas dificultades, se ha desarrollado un sistema alternativo utilizando los receptores de la ecdisona de los insectos (EcR).

La diana molecular para la ecdisona en los insectos consiste de por lo menos un receptor de ecdisona (EcR) y la proteína Ultraspiracle (USP). El EcR es un miembro de la superfamilia de receptores esteroideos nucleares, que se caracteriza a partir de su signatura de ADN y los dominios de unión a ligandos y por un dominio de activación (Koelle *et al.*, Cell 67:59-77, 1991). El EcR es un miembro de la superfamilia de receptores nucleares y se clasifica en subfamilia 1, grupo H (denominado en la presente memoria "receptores nucleares de grupo H"). Los miembros de cada grupo comparten una identidad de aminoácidos de 40% a 60% en el dominio E (de unión a ligando (Laudet *et al.*, A Unified Nomenclature System for the Nuclear Receptor Subfamily, Cell 97:161-163, 1999). Además del receptor de ecdisona, entre otros miembros de esta subfamilia 1, grupo H de receptores nucleares se incluyen: el receptor ubicuo (UR), el receptor huérfano 1 (OR-1), el receptor nuclear 1 de hormonas esteroideas (NER-1), la proteína de interacción con RXR 15 (RIP-15), el receptor X hepático  $\beta$  (LXR $\beta$ ), la proteína de tipo receptor de hormona esteroidea (RLD-1), el receptor hepático X (LXR), el receptor X hepático  $\alpha$  (LXR $\alpha$ ), el receptor X de farnesoide (FXR), la proteína de interacción con receptores 14 (RIP-14) y el receptor de farnesol (HRR-1).

Los receptores EcR responden a varios compuestos esteroideos, tales como la ponasterona A y la muristerona A. Recientemente se han descrito compuestos no esteroideos con actividad agonista ecdisteroidea, incluyendo los insecticidas disponibles comercialmente denominados tebufenózido y metoxifenózido, comercializados en todo el mundo por Rohm and Haas Company (ver la solicitud de patente internacional n<sup>o</sup> PCT/EP96/00686 y la patente US n<sup>o</sup> 5.530.028). Ambos análogos presentan perfiles de seguridad excepcionales para otros organismos.

El receptor de ecdisona de los insectos (EcR) se heterodimeriza con Ultraspiracle (USP), el homólogo de insecto del RXR de mamífero, y se une a los ecdisteroides y elementos de respuesta a receptores de ecdisona y activa la transcripción de los genes sensibles a la ecdisona (Riddiford *et al.*, Vitam. Horm. 60:1-73, 2000). Los complejos EcR/USP/ligando desempeñan funciones importantes durante el desarrollo y reproducción de los insectos. El EcR es un miembro de la superfamilia de receptores de hormonas esteroideas y presenta cinco dominios modulares: los dominios A/B (transactivación), C (unión a ADN, heterodimerización), D (bisagra, heterodimerización), E (unión a ligandos, heterodimerización y transactivación) y, en algunos casos, F (transactivación). Algunos de estos dominios, tales como A/B, C y E, conservan su función en el caso de que se encuentren fusionados con otras proteínas.

Los sistemas de expresión de genes inducibles regulados estrechamente, o "interruptores génicos", resultan útiles en diversas aplicaciones, tales como la terapia génica, la producción a gran escala de proteínas en células, los ensayos de cribado de alto rendimiento basados en células, la genómica funcional y la regulación de caracteres en plantas y animales transgénicos. La primera versión de un interruptor génico basado en EcR utilizó el EcR de *Drosophila melanogaster* (DmEcR) y el RXR de *Mus musculus* (MmRXR) y demostró que estos receptores en presencia del esteroide ponasterona A, transactivaban genes informadores en líneas celulares de mamífero y en ratones transgénicos (Christopherson *et al.*, PNAS 89(14):6314-8, 1992, y No *et al.*, PNAS 98(8):3346-51, 1996). Se compararon directamente los sistemas de regulación de EcR y de la tetraciclina y se concluyó que el sistema de regulación de EcR presentaba una actividad basal inferior en comparación con cualquiera de las dos versiones del sistema basado en tetraciclina (tTA y rtTA), demostrando que el sistema basado en EcR era menos débil ("leaky"). Posteriormente, Suhr *et al.* (PNAS 95:7999-8004, 1998) demostraron que el agonista no esteroideo de la ecdisona denominado tebufenózido inducía un nivel elevado de transactivación de los genes informadores en las células de mamífero a través del EcR de *Bombyx mori* (BmEcR) en ausencia de pareja exógena de heterodímero (ver también la solicitud de patente internacional n<sup>o</sup> PCT/US98/14215 (patente WO n<sup>o</sup> 99/02683).

Las solicitudes de patente internacional nº PCT/US97/05330 (patente WO nº 97/38117) y nº PCT/US99/08381 (patente WO nº 99/58155) dan a conocer procedimientos para modular la expresión de un gen exógeno en el que un constructo de ADN que comprende el gen exógeno y un elemento de respuesta a ecdisona resulta activado por un segundo constructo de ADN que comprende un receptor de ecdisona que, en presencia de un ligando del mismo, y opcionalmente en presencia de un receptor capaz de actuar como pareja silenciosa, se une al elemento de respuesta a ecdisona, induciendo la expresión génica. El receptor de ecdisona de elección fue aislado a partir de *Drosophila melanogaster*. Típicamente, dichos sistemas requieren la presencia de una pareja silenciosa, preferentemente el receptor X del retinoide (RXR), para proporcionar la activación óptima. En las células de mamífero, el receptor de ecdisona de insecto (EcR) se heterodimeriza con el receptor X del retinoide (RXR) y regula la expresión de los genes diana de un modo dependiente de ligando. La solicitud de patente internacional nº PCT/US98/14215 (patente WO nº 99/02683) da a conocer que el receptor de ecdisona aislado a partir del gusano de la seda *Bombyx mori* resulta funcional en sistemas de mamífero sin necesidad de una pareja exógena del dímero.

La patente US nº 6.265.173 B1 da a conocer que diversos miembros de la superfamilia de receptores esteroideos/tiroideos pueden combinarse con el receptor Ultraspiracle (USP) de *Drosophila melanogaster* o con fragmentos del mismo que comprenden por lo menos el dominio de dimerización de USP, para la utilización en un sistema de expresión génica. La patente US nº 5.880.333 da a conocer un sistema de heterodímero de EcR de *Drosophila melanogaster* y Ultraspiracle (USP) utilizado en plantas en las que el dominio de transactivación y el dominio de unión a ADN se encuentran situados en dos proteínas híbridas diferentes. Por desgracia, estos sistemas basados en la USP son constitutivos en las células animales y, por lo tanto, no resultan efectivos para regular la expresión de los genes informadores.

En cada uno de dichos casos, el dominio de transactivación y el dominio de unión a ADN (en forma de EcR nativo, tal como en la solicitud de patente internacional nº PCT/US98/14215 o de EcR modificado, tal como en la solicitud de patente internacional nº PCT/US97/05330) han sido incorporados en una única molécula y las otras parejas heterodiméricas, USP o RXR, fueron utilizadas en su estado nativo.

Moradpour *et al.* (Biol. Chem. 379:1189-91, 1998) dan a conocer la transfección estable de sistemas de expresión tanto regulados por tetraciclina como por ecdisona en una línea celular humana. El sistema permite la regulación independiente (ortogonal) de dos genes separados.

Entre los inconvenientes de los sistemas de regulación génica basados en EcR que se han descrito se incluyen una actividad de fondo considerable en ausencia de ligandos, la no aplicabilidad de estos sistemas para la utilización tanto en plantas como en animales (ver la patente US nº 5.880.333) y el uso limitado o la incapacidad para regular la expresión de múltiples genes. Por lo tanto, existe una necesidad en la técnica de sistemas basados en EcR mejorados para modular con precisión la expresión de dos o más genes exógenos tanto en plantas como en animales. Estos sistemas mejorados resultarían útiles para aplicaciones tales como la terapia génica, la producción a gran escala de proteína y anticuerpos, los ensayos de cribado de alto rendimiento basados en células, la genómica funcional y la regulación de caracteres en animales transgénicos. Para determinadas aplicaciones tales como la terapia génica, puede resultar deseable disponer de un sistema de expresión génica inducible que responda bien a ligandos no esteroideos sintéticos y que simultáneamente resulte insensible a los esteroides naturales.

Recientemente, los presentes solicitantes han demostrado que un sistema de expresión génica inducible basado en receptores de ecdisona en el que los dominios de transactivación y de unión a ADN se separan uno de otro al introducirlos en dos proteínas diferentes, resulta en una actividad de fondo muy reducida en ausencia de un ligando, y en una actividad significativamente incrementada sobre el fondo en presencia de un ligando (solicitud pendiente de patente nº PCT/US01/09050). Este sistema de dos híbridos es un sistema de modulación de la expresión génica inducible significativamente mejorado en comparación con los dos sistemas dados a conocer en las solicitudes nº PCT/US97/05330 y nº PCT/US98/14215. El sistema de dos híbridos explota la capacidad de una pareja de proteínas interactuantes de llevar el dominio de activación de la transcripción a una posición más favorable respecto al dominio de unión a ADN, de manera que, al unirse el dominio de unión a ADN al sitio de unión a ADN en el gen, el dominio de transactivación activa de modo más efectivo el promotor (ver, por ejemplo, la patente US nº 5.283.173). Brevemente, el sistema de expresión génica de dos híbridos comprende dos casetes de expresión génica: el primero codificante de un dominio de unión a ADN fusionado a un polipéptido receptor nuclear, y el segundo, codificante de un dominio de transactivación fusionado con un polipéptido receptor nuclear diferente. En presencia de ligando, la interacción del primer polipéptido con el segundo polipéptido ancla eficazmente el dominio de unión de ADN al dominio de transactivación. Debido a que los dominios de unión a ADN y de transactivación residen en dos moléculas diferentes, la actividad de fondo en ausencia de ligando se ve muy reducida.

Un sistema de dos híbridos también proporciona una sensibilidad mejorada a ligandos no esteroideos, por ejemplo diacilhidrazinas, en comparación con ligandos esteroideos, por ejemplo la ponasterona A ("Pon A") o la muristerona A ("MurA"). Es decir, en comparación con los esteroides, los ligandos no esteroideos proporcionan una actividad más alta a una concentración inferior. Además, debido a que la transactivación basada en interruptores génicos de EcR con frecuencia es dependiente de la línea celular, resulta más fácil adaptar los sistemas de interruptor para obtener la máxima capacidad de transactivación para cada aplicación. Adicionalmente, el sistema de dos híbridos evita algunos efectos secundarios debidos a la sobreexpresión de RXR, que se producen con frecuencia al utilizar el RXR

no modificado como pareja de interruptor. En un sistema de dos híbridos preferente, los dominios nativos de unión a ADN y de transactivación de EcR o de RXR resultan eliminados y, como resultado, estas moléculas híbridas presentan una menor probabilidad de interactuar con otros receptores de hormona esteroidea presentes en la célula, resultando en menos efectos secundarios.

5 La invención de los solicitantes supera una deficiencia de la técnica y proporciona un medio para modular simultáneamente la expresión de dos o más genes en la misma célula. La invención de los solicitantes proporciona un sistema de regulación de múltiples genes inducibles que permite la regulación simultánea y cuantitativa de dos o más genes diferentes en la misma célula, tejido u organismo. La invención de los solicitantes resulta útil en aplicaciones en las que resulta crítica la regulación de múltiples genes. De esta manera, la invención de los solicitantes supera una deficiencia en el campo de la expresión génica y resulta útil en los campos de la genómica funcional, la proteómica, la metabolómica, el cribado toxicológico, los ensayos de cribado de alto rendimiento basados en células, la producción de proteínas, las terapias génicas y similares. La invención de los solicitantes proporciona un medio para el control en paralelo de múltiples genes en la misma célula y permite al experto en la materia analizar fenómenos biológicos complejos, en los que se encuentran implicados múltiples genes o rutas, así como crear nuevas aplicaciones terapéuticas.

### Breve descripción de los dibujos

20 Figura 1: transactivación ortogonal de genes informadores a través de los constructos GAL4DmEcR-CDEF y LexACfEcR-CDEF transfectados en células NIH3T3 conjuntamente con VP16MmRXR $\alpha$ -LmUSP-Efquimera, p8OPLexARELuc y p6XGALRETTTPSEAP con PonA y/o GS<sup>TM</sup>-E. Los números sobre las columnas indican el factor de incremento respecto a los niveles en DMSO.

25 Figura 2: transactivación de genes informadores a través de GAL4CfEcR-DEF o GAL4NcEcR-CDE transfectados en células CHO conjuntamente con VP16MmRXR $\alpha$ -EF e informador pFRLuc con PonA o GS<sup>TM</sup>-E. Los números sobre las columnas indican el factor de incremento respecto a los niveles en DMSO.

### Descripción detallada de la invención

30 La presente invención resulta útil para aplicaciones tales como la terapia génica, la producción a gran escala de proteína y anticuerpos, los ensayos de cribado de alto rendimiento basados en células, los ensayos de cribado de ligandos ortogonales, la genómica funcional, la proteómica, la metabolómica, el cribado toxicológico y la regulación de caracteres en organismos transgénicos, en los que resulta deseable el control de los niveles de expresión génica. Una ventaja de la invención de los solicitantes es que proporciona un medio para regular la expresión de dos o más genes y para regular los niveles de expresión para adaptarse a los requisitos del usuario.

La presente invención proporciona un sistema de regulación de múltiples genes inducibles que comprende una pluralidad de sistemas de regulación génica individualmente operables, en la que:

- 40 a) cada sistema de regulación génica individualmente operable comprende:
- i) uno o más polinucleótidos codificantes de un complejo de receptores que comprende:
    - 45 A) un dominio de unión a ADN,
    - B) un primer dominio de unión a ligando de receptor nuclear y un segundo dominio de unión a ligando de receptor nuclear, y
    - 50 C) un dominio de transactivación,
  - ii) un ligando,
  - iii) un polinucleótido que comprende:
    - 55 A) un polinucleótido exógeno o endógeno, y
    - B) un elemento de respuesta,
- 60 en los que:
- A) el polinucleótido exógeno o endógeno se encuentra funcionalmente ligado al elemento de respuesta, y
  - 65 B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o ausencia del ligando resulta en la activación o supresión del polinucleótido exógeno o endógeno, y

C) uno de los dominios de unión a ligando de receptor nuclear es un dominio de unión a ligando de receptor nuclear de grupo H y el otro dominio de unión a ligando de receptor nuclear es un dominio de unión a ligando de receptor nuclear capaz de formar un dímero con el dominio de unión a ligando de receptor nuclear de grupo H, y

5 b) cada sistema de regulación génica individualmente operable es ortogonal respecto al otro sistema de regulación génica individualmente operable presente en el sistema de regulación de múltiples genes inducibles.

10 En otro aspecto, la presente invención proporciona un virus que comprende el sistema de regulación de múltiples genes según la presente invención.

En un aspecto adicional, la presente invención proporciona una célula que comprende el sistema de regulación de múltiples genes según la presente invención.

15 La presente invención proporciona en un aspecto adicional un organismo transgénico no humano que comprende una o más células de la presente invención.

En otro aspecto, la presente invención proporciona un procedimiento para desarrollar un sistema de regulación de múltiples genes según la presente invención que comprende las etapas siguientes:

20 a) definir un conjunto de ligandos diversamente modificados basado en cambios progresivos de elementos farmacóforos,

25 b) preparar un primer conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ligando que:

i) es de origen natural

30 ii) se encuentra modificado por delección, inserción o mutación,

iii) es quimérico,

iv) es sintético, o

35 v) una combinación de los anteriores,

c) opcionalmente, introducir los polipéptidos receptores nucleares en una célula,

40 d) realizar una búsqueda entre los polipéptidos receptores nucleares en la célula de c) o *in vitro*, utilizando el conjunto de ligandos diversamente modificados con el fin de encontrar modulación génica o unión de dominio de unión a ligando, o ambos,

45 e) determinar la ortogonalidad de las combinaciones de polipéptido receptor nuclear/ligando con el fin de definir un subconjunto de ligandos con propiedades diferentes de modulación génica,

f) opcionalmente repetir las etapas a) a e) utilizando un conjunto modificado de ligandos y un conjunto modificado de polipéptidos receptores nucleares,

50 g) preparar un segundo conjunto de polipéptidos receptores nucleares,

en el que cada polipéptido receptor nuclear comprende un dominio de unión a ADN que:

i) es de origen natural,

55 ii) se encuentra modificado por delección, inserción o mutación,

iii) es quimérico,

iv) es sintético, o

60 v) es una combinación de los anteriores,

h) preparar un conjunto de constructos de ADN que comprende un gen exógeno o endógeno y elementos de respuesta que:

65 i) son de origen natural,

ii) se encuentran modificados por delección, inserción o mutación,

5 iii) son quiméricos,

iv) son sintéticos, o

v) son una combinación de los anteriores,

10 i) clonar el primer conjunto de polipéptidos receptores nucleares y el segundo conjunto de polipéptidos receptores nucleares y los constructos de ADN y después introducir los clones en una célula,

j) someter a ensayo a la célula para correactividad con los ligandos identificados en la etapa e), y

15 k) seleccionar un conjunto ortogonal de ligandos, dominios de unión a ligando, dominios de unión a ADN y elementos de respuesta, basado en los resultados de las etapas e) y j), con el fin de que comprenda el sistema de regulación de múltiples genes inducibles según la presente invención.

20 En un aspecto adicional, la presente invención proporciona un sistema de regulación de múltiples genes inducibles que comprende una pluralidad de sistemas de regulación génica individualmente operables, en el que:

a) cada sistema de regulación génica individualmente operable comprende:

25 i) uno o más polinucleótidos codificantes de un complejo de receptores que comprende:

A) un dominio de unión a ADN,

B) un dominio de unión a ligando, y

30 C) un dominio de transactivación,

ii) un ligando,

35 iii) un polinucleótido que comprende:

A) un polinucleótido exógeno o endógeno, y

B) un elemento de respuesta,

40 en el que:

A) el polinucleótido exógeno o endógeno se encuentra funcionalmente ligado al elemento de respuesta,

45 B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o en ausencia del ligando resulta en la activación o supresión del polinucleótido exógeno o endógeno, y

C) el complejo de receptores es un complejo de receptores nucleares de grupo H, y

50 b) cada sistema de regulación génica individualmente operable es ortogonal respecto a otros sistemas de regulación génica individualmente operables en el sistema de regulación de múltiples genes inducibles.

La presente invención proporciona, en otro aspecto, un procedimiento para desarrollar un sistema de regulación de múltiples genes según la presente invención que comprende las etapas siguientes:

55 a) definir un conjunto de ligandos diversamente modificados basado en cambios progresivos de elementos farmacóforos,

b) preparar un primer conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ligando que:

60 i) es de origen natural,

ii) se encuentra modificado por delección, inserción o mutación,

65 iii) es quimérico,

iv) es sintético, o

v) es una combinación de los anteriores,

5 c) opcionalmente, introducir los polipéptidos receptores nucleares en una célula,

d) realizar una búsqueda entre los polipéptidos receptores nucleares en la célula de c) o *in vitro*, utilizando el conjunto de ligandos diversamente modificados con el fin de encontrar modulación génica o unión de dominio de unión a ligando, o ambos,

10 e) determinar la ortogonalidad de las combinaciones de polipéptido receptor nuclear/ligando con el fin de definir un subconjunto de ligandos con diferentes propiedades de modulación génica,

15 f) opcionalmente repetir las etapas a) a e) utilizando un conjunto modificado de ligandos y un conjunto modificado de polipéptidos receptores nucleares,

g) preparar un segundo conjunto de polipéptidos receptores nucleares,

en el que cada polipéptido receptor nuclear comprende un dominio de unión a ADN que:

20 i) es de origen natural,

ii) se encuentra modificado por delección, inserción o mutación,

25 iii) es quimérico,

iv) es sintético, o

30 v) es una combinación de los anteriores,

h) preparar un conjunto de constructos de ADN que comprende un gen exógeno o endógeno y elementos de respuesta que:

35 i) son de origen natural,

ii) se encuentran modificados por delección, inserción o mutación,

iii) son quiméricos,

40 iv) son sintéticos, o

v) son una combinación de los anteriores,

45 i) clonar el primer conjunto de polipéptidos receptores nucleares y el segundo conjunto de polipéptidos receptores nucleares y los constructos de ADN y después introducir los clones en una célula,

j) someter a ensayo la célula para correactividad con los ligandos identificados en la etapa e), y

50 k) seleccionar un conjunto ortogonal de ligandos, dominios de unión a ligando, dominios de unión a ADN, y elementos de respuesta basándose en los resultados de las etapas e) y j), con el fin de que comprenda el sistema de regulación de múltiples genes inducibles según la presente invención.

### Definiciones

55 En la presente exposición, se utilizan varios términos, expresiones y abreviaturas. Se proporcionan las definiciones siguientes, que deberían resultar útiles para entender el alcance y la práctica de la presente invención.

60 En una forma de realización específica, el término "alrededor de" o "aproximadamente" se refiere a dentro de un entorno del 20%, preferentemente del 10%, más preferentemente del 5%, y todavía más preferentemente del 1% de un valor o intervalo dado.

65 Tal como se utiliza en la presente memoria, todos los porcentajes son porcentajes en peso y todas las partes son partes en peso, a menos que se indique lo contrario, y son inclusivas y combinables. Todas las proporciones son en peso y todos los intervalos de proporciones son inclusivos y combinables. Todos los intervalos molares son inclusivos y combinables.

La expresión "sustancialmente libre" se refiere a que la composición que comprende "A" (en la que "A" es una única proteína, molécula de ADN, vector, célula huésped recombinante, etc.) se encuentra sustancialmente libre de "B" (en la que "B" comprende una o más proteínas, moléculas de ADN, vectores, etc. contaminantes), en el caso de que por lo menos aproximadamente 75% en peso de las proteínas, ADN o vectores (dependiendo de la categoría de la especie a la que pertenezca A y B) en la composición sea "A". Preferentemente, "A" comprende por lo menos aproximadamente 90% en peso de las especies A+B en la composición, más preferentemente por lo menos aproximadamente 99% en peso. También resulta preferente que una composición, que se encuentra sustancialmente libre de contaminación, contenga únicamente una especie de un único peso molecular que presente la actividad o característica de la especie de interés.

El término "aislado" para los fines de la presente invención se refiere a un material biológico (ácido nucleico o proteína) que ha sido extraído de su ambiente original (el ambiente en el que se encuentra presente naturalmente). Por ejemplo, un polinucleótido presente en el estado natural en una planta o en un animal no ha sido aislado; sin embargo, ese mismo polinucleótido separado de los ácidos nucleicos contiguos con los que se encuentra presente naturalmente, se considera "aislado". El término "purificado" no requiere que el material se encuentre presente en una forma que muestre una pureza absoluta, en exclusión de la presencia de otros compuestos. Por el contrario es una definición relativa.

Un polinucleótido se encuentra en el estado "purificado" tras la purificación del material de partida o del material natural en por lo menos un orden de magnitud, preferentemente 2 ó 3, y preferentemente 4 ó 5 órdenes de magnitud.

Un "ácido nucleico" es un compuesto polimérico que comprende subunidades unidas covalentemente denominadas nucleótidos. La expresión "ácido nucleico" incluye ácido polirribonucleico (ARN) y ácido polidesoxirribonucleico (ADN), los cuales pueden ser de una cadena o de doble cadena. El ADN incluye, aunque sin limitación, ADNc, ADN genómico, ADN plasmídico, ADN sintético y ADN semisintético. El ADN puede ser lineal, circular o superenrollado.

Una "molécula de ácidos nucleicos" se refiere a la forma polimérica del éster fosfato de ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o de desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualesquiera análogos fosfoéster de los mismos, tales como fosforotioatos y tioésteres, en forma de una cadena o de una hélice de doble cadena. Las hélices de doble cadena de ADN-ADN, ADN-ARN y ARN-ARN resultan posibles. La expresión "molécula de ácidos nucleicos", y en particular "molécula de ADN" o "molécula de ARN", se refiere a únicamente las estructuras primaria y secundaria de la molécula, y no se encuentra limitada a ninguna forma terciaria en particular. De esta manera, este término incluye el ADN de doble cadena presente, entre otras, en moléculas de ADN lineal o circular (por ejemplo fragmentos de restricción), plásmidos y cromosomas. En el comentario de la estructura de moléculas particulares de ADN de doble cadena, en la presente memoria puede hacerse referencia a secuencias según la convención normal de proporcionar únicamente la secuencia en la dirección 5' a 3' a lo largo de la cadena no transcrita de ADN (es decir, la cadena que presenta una secuencia homóloga a la del ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha sido sometida a manipulación biológica molecular.

El término "fragmento" se entiende que se refiere a una secuencia de nucleótidos de longitud reducida respecto al ácido nucleico de referencia, y que comprende, en la parte común, una secuencia de nucleótidos idéntica a la del ácido nucleico de referencia. Dicho fragmento de ácidos nucleicos según la invención puede incluirse, en caso apropiado, en un polinucleótido de mayor tamaño del que es un constituyente. Dichos fragmentos comprenden, o alternativamente consisten de oligonucleótidos de longitud comprendida entre por lo menos 6 y 1.500 nucleótidos consecutivos de un polinucleótido mencionado en la presente memoria.

Tal como se utiliza en la presente memoria, un "fragmento aislado de ácidos nucleicos" es un polímero de ADN o ARN que es de una cadena o de doble cadena, que opcionalmente contiene bases nucleotídicas sintéticas, no naturales o alteradas. Un fragmento de ácido nucleico aislado en forma de un polímero de ADN puede comprender uno o más segmentos de ADNc, ADN genómico o ADN sintético.

Un "gen" se refiere a un conjunto de nucleótidos que codifica un polipéptido, e incluye ácidos nucleicos de ADNc y de ADN genómico. El término "gen" también se refiere a un fragmento de ácidos nucleicos que expresa una proteína o polipéptido específico, incluyendo secuencias reguladoras anteriores (secuencias 5' no codificantes) y posteriores (secuencias 3' no codificantes) a la secuencia codificante. La expresión "gen nativo" se refiere a un gen tal como se encuentra presente en la Naturaleza, con sus propias secuencias reguladoras. La expresión "gen quimérico" se refiere a cualquier gen que no sea un gen nativo, que comprende secuencias reguladoras y/o codificantes que no se encuentran conjuntamente en la Naturaleza. De acuerdo con lo anterior, un gen quimérico puede comprender secuencias reguladoras y secuencias codificantes que se derivan a partir de diferentes fuentes, o secuencias reguladoras y secuencias codificantes derivadas a partir de la misma fuente, aunque organizadas de una manera diferente a la observada en la Naturaleza. Un gen quimérico puede comprender secuencias codificantes derivadas a partir de diferentes fuentes y/o secuencias reguladoras derivadas a partir de diferentes fuentes. La expresión "gen endógeno" se refiere a un gen nativo en su localización natural en el genoma de un organismo. Un gen "foráneo" o "heterólogo" se refiere a un gen no presente normalmente en el organismo huésped, sino que ha sido introducido en

el organismo huésped mediante transferencia génica. Los genes foráneos pueden comprender genes nativos insertados en un organismo no nativo, o genes quiméricos. Un "transgén" es un gen que ha sido introducido en el genoma mediante un procedimiento de transformación.

- 5 El término "heterólogo" referido al ADN se refiere a ADN no situado naturalmente en la célula, o en un sitio cromosómico de la célula. Preferentemente, el ADN heterólogo incluye un gen que es foráneo respecto a la célula.

El término "genoma" incluye ADN o ARN cromosómico, así como mitocondrial, de cloroplasto y vírico.

- 10 Una molécula de ácidos nucleicos es "hibridable" con otra molécula de ácidos nucleicos tal como ADNc, ADN genómico o ARN, en el caso de que una forma de una sola cadena de la molécula de ácidos nucleicos pueda aparearse con la otra molécula de ácidos nucleicos bajo las condiciones apropiadas de temperatura y fuerza iónica de la solución (ver Sambrook *et al.*, 1989, *infra*). Las condiciones de hibridación y de lavado son bien conocidas y se encuentran ejemplificadas en Sambrook J., Fritsch E.F. y Maniatis T., *Molecular Cloning: A Laboratory Manual*, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989, particularmente en el capítulo 11 y en la Tabla 11.1 en el mismo. Las condiciones de temperatura y de fuerza iónica determinan la "astringencia" de la hibridación.

- 20 Pueden ajustarse las condiciones de astringencia para el cribado de identificación de fragmentos moderadamente similares, tales como secuencias homólogas de organismos distantemente relacionados, de fragmentos altamente similares, tales como genes que duplican enzimas funcionales procedentes de organismos estrechamente relacionados. Para el cribado preliminar para ácidos nucleicos homólogos, pueden utilizarse condiciones de hibridación de baja astringencia, correspondientes a una  $T_m$  de 55°C, por ejemplo 5xSSC, SDS al 0,1%, leche al 0,25% y en ausencia de formamida; o formamida al 30%, 5x SSC, SDS al 0,5%). Las condiciones de hibridación de astringencia moderada corresponden a una  $T_m$  más alta, por ejemplo formamida al 40%, con 5x ó 6x SCC. Las condiciones de hibridación de alta astringencia corresponden a la  $T_m$  más alta, por ejemplo formamida al 50% y 5x ó 6x SCC.

- 30 La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque dependiendo de la astringencia de la hibridación, son posibles desapareamientos entre bases. El término "complementario" se utiliza para referirse a la relación entre bases nucleotídicas que son capaces de hibridarse entre sí. Por ejemplo, con respecto al ADN, la adenosina es complementaria a la timina y la citosina es complementaria a la guanina. De acuerdo con lo anterior, la presente descripción también incluye fragmentos de ácidos nucleicos aislados que son complementarios a las secuencias completas dadas a conocer o utilizadas en la presente memoria, así como aquellas secuencias de ácidos nucleicos sustancialmente similares.

- 40 En una forma de realización específica, los polinucleótidos se detectan mediante la utilización de condiciones de hibridación que comprenden una etapa de hibridación a una  $T_m$  de 55°C, y utilizando condiciones tales como las indicadas anteriormente. En una forma de realización preferente, la  $T_m$  es 60°C; en una forma de realización más preferente, la  $T_m$  es 63°C; en una forma de realización todavía más preferente, la  $T_m$  es de 65°C.

- 45 Los lavados posteriores a la hibridación también determinan las condiciones de astringencia. Un conjunto de condiciones preferentes utiliza una serie de lavados partiendo de 6X SSC, SDS al 0,5% a temperatura ambiente durante 15 minutos (min), repetidos después con 2X SSC, SDS al 0,5% a 45°C durante 30 minutos, y repetidos posteriormente dos veces con 0,2X SSC, SDS al 0,5% a 50°C durante 30 minutos. Un conjunto más preferente de condiciones astringentes utiliza temperaturas más altas en las que los lavados son idénticos a los indicados anteriormente, excepto por la temperatura de los dos lavados finales de 30 minutos en 0,2X SSC, SDS al 0,5%, que se incrementó a 60°C. Otro conjunto preferente de condiciones altamente astringentes utiliza dos lavados finales en 0,1X SSC, SDS al 0,1% a 65°C. La hibridación requiere que los dos ácidos nucleicos comprendan secuencias complementarias, aunque dependiendo de la astringencia de la hibridación, resultan posibles desapareamientos entre bases.

- 55 La astringencia apropiada para la hibridación de ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementariedad, variables que son bien conocidas de la técnica. A medida que se incrementa el grado de similitud o homología entre dos secuencias de nucleótidos, se incrementa el valor de la  $T_m$  para los híbridos de los ácidos nucleicos que presentan dichas secuencias. La estabilidad relativa (correspondiente a una  $T_m$  más alta) de las hibridaciones de ácidos nucleicos se reduce en el orden siguiente: ARN:ARN, ADN:ARN y ADN:ADN. Para los híbridos de longitud superior a 100 nucleótidos, se han derivado ecuaciones para calcular la  $T_m$  (ver Sambrook *et al.*, *supra*, 9:50-0.51). Para la hibridación con ácidos nucleicos más cortos, es decir oligonucleótidos, la posición de los desapareamientos adquiere importancia, y la longitud del oligonucleótido determina su especificidad (ver Sambrook *et al.*, *supra*, 11.7-11.8).

- 65 En una forma de realización específica, los polinucleótidos se detectan mediante la utilización de condiciones de hibridación que comprenden una etapa de hibridación en sal de concentración inferior a 500 mM y a por lo menos 37 grados Celsius, y una etapa de lavado en 2X SSPE a por lo menos 63 grados Celsius. En una forma de realización preferente, las condiciones de hibridación comprenden una concentración salina inferior a 200 mM y por lo menos 37

grados Celsius en la etapa de hibridación. En una forma de realización más preferente, las condiciones de hibridación comprenden 2X SSPE y 63 grados Celsius para las etapas tanto de hibridación como de lavado.

En una forma de realización, la longitud de un ácido nucleico hibridable es de por lo menos aproximadamente 10 nucleótidos. Preferentemente, una longitud mínima de un ácido nucleico hibridable es de por lo menos aproximadamente 15 nucleótidos; más preferentemente de por lo menos aproximadamente 20 nucleótidos, y todavía más preferentemente la longitud es de por lo menos 30 nucleótidos. Además, el experto en la materia reconocerá que la temperatura y la concentración salina de la solución de lavado pueden ajustarse según resulte necesario según factores tales como la longitud de la sonda.

El término "sonda" se refiere a una molécula de ácidos nucleicos de una sola cadena las bases de la cual pueden aparearse con un ácido nucleico diana de una cadena complementario, formando una molécula de doble cadena.

Tal como se utiliza en la presente memoria, el término "oligonucleótido" se refiere a un ácido nucleico, generalmente de por lo menos 18 nucleótidos, que puede hibridarse con una molécula de ADN genómico, una molécula de ADNc, un ADN plasmídico o una molécula de ARN. Los oligonucleótidos pueden marcarse con, por ejemplo, nucleótidos-<sup>32</sup>P o con nucleótidos a los que puede conjugarse covalentemente un marcaje, tal como biotina. Puede utilizarse un oligonucleótido marcado a modo de sonda para detectar la presencia de un ácido nucleico. Los oligonucleótidos (encontrándose marcados uno o los dos) pueden utilizarse como cebadores de PCR, para la clonación de longitud completa o de un fragmento de un ácido nucleico, o para detectar la presencia de un ácido nucleico. También puede utilizarse un oligonucleótido para formar una triple hélice con una molécula de ADN. Generalmente los oligonucleótidos se preparan sintéticamente, preferentemente en un sintetizador de ácidos nucleicos. De acuerdo con lo anteriormente expuesto, pueden prepararse oligonucleótidos con enlaces de análogo de fosfoéster de origen no natural, tales como enlaces tioéster, etc.

Un "cebador" es un oligonucleótido que se hibrida con una secuencia diana de ácidos nucleicos creando una región de ácidos nucleicos de doble cadena que puede servir como punto de inicio para la síntesis de ADN bajo condiciones adecuadas. Dichos cebadores pueden utilizarse en una reacción en cadena de polimerasa.

La "reacción en cadena de la polimerasa" se abrevia como PCR y se refiere a un procedimiento *in vitro* para amplificar enzimáticamente secuencias específicas de ácidos nucleicos. La PCR implica una serie repetida de ciclos térmicos, comprendiendo cada ciclo tres etapas: la desnaturalización del ácido nucleico molde para separar las cadenas de la molécula diana, el apareamiento de un cebador oligonucleótido monocatenario de PCR con el ácido nucleico molde, y la extensión del cebador o cebadores hibridados por parte de la ADN polimerasa. La PCR proporciona un medio para detectar la presencia de una molécula diana y, bajo condiciones cuantitativas o semicuantitativas, para detectar la cantidad relativa de la molécula diana dentro del grupo de partida de ácidos nucleicos.

La "reacción en cadena de la polimerasa en transcripción inversa" se abrevia RT-PCR y se refiere a un procedimiento *in vitro* para producir enzimáticamente una o más moléculas diana de ADNc a partir de una o más moléculas de ARN, seguido de la amplificación enzimática de una o más secuencias específicas de ácidos nucleicos dentro de la molécula o moléculas diana de ADNc tal como se ha indicado anteriormente. La RT-PCR también proporciona un medio para detectar la presencia de la molécula diana y, bajo condiciones cuantitativas o semicuantitativas, para determinar la cantidad relativa de la molécula diana dentro del grupo de partida de ácidos nucleicos.

Una "secuencia codificante" de ADN es una secuencia de ADN de doble cadena que es transcrita y traducida en un polipéptido en una célula *in vitro* o *in vivo* al situarla bajo el control de las secuencias reguladoras apropiadas. La expresión "secuencias reguladoras adecuadas" se refiere a secuencias de nucleótidos situadas cadena arriba (secuencias 5' no codificantes), dentro o cadena abajo (secuencias 3' no codificantes) de una secuencia codificante, y que influyen sobre la transcripción, el procesamiento o la estabilidad del ARN, o la traducción de la secuencia codificante asociada. Las secuencias reguladoras pueden incluir promotores, secuencias líder de traducción, intrones, secuencias de reconocimiento de poliadenilación, sitios de procesamiento del ARN, sitios de unión de efectores y estructuras de tallo-bucle. Los límites de la secuencia codificante están determinados por un codón de inicio en el extremo 5' terminal (amino) y por un codón de parada de traducción en el extremo 3' terminal (carboxilo). Una secuencia codificante puede incluir, aunque sin limitación, secuencias procarióticas, ADNc del ARNm, secuencias de ADN genómico e incluso secuencias de ADN sintético. En el caso de que la secuencia codificante esté destinada a la expresión en una célula eucariótica, habitualmente se encontrará una señal de poliadenilación y una secuencia de terminación de la transcripción 3' respecto a la secuencia codificante.

La expresión "marco de lectura abierto" se abrevia ORF y se refiere a un segmento de secuencia de ácidos nucleicos, de ADN, ADNc o ARN, que comprende una señal de inicio de traducción o un codón de inicio, tal como ATG o AUG, y un codón de terminación y que potencialmente puede traducirse en una secuencia de polipéptido.

La expresión "cabeza con cabeza" se utiliza en la presente memoria para referirse a la orientación mutua de dos secuencias polipeptídicas. Dos polinucleótidos se encuentran posicionados en una orientación de cabeza con

cabeza en el caso de que el extremo 5' de la cadena codificante de un polinucleótido sea contigua al extremo 5' de la cadena codificante del otro polinucleótido, de manera que la dirección de transcripción de cada polinucleótido continúe de manera que se aleje del extremo 5' del otro polinucleótido. La expresión "cabeza con cabeza" puede abreviarse (5')-a-(5') y también puede indicarse con los símbolos ( $\leftarrow\rightarrow$ ) o ( $3'\leftarrow 5'5'\rightarrow 3'$ ).

5 La expresión "cola con cola" se utiliza en la presente memoria para referirse a la orientación mutua de dos secuencias polinucleótidas. Dos polinucleótidos se encuentran posicionados en una orientación de cola con cola en el caso de que el extremo 3' de la cadena codificante de un polinucleótido sea contigua al extremo 3' de la cadena codificante del otro polinucleótido, de manera que la dirección de transcripción de cada polinucleótido transcurre hacia el otro polinucleótido. La expresión "cola con cola" puede abreviarse (3')-a-(3') y también puede indicarse con los símbolos ( $\rightarrow\leftarrow$ ) o ( $5'\rightarrow 3'3'\leftarrow 5'$ ).

15 La expresión "cabeza con cola" se utiliza en la presente memoria para referirse a la orientación mutua de dos secuencias polinucleótidas. Dos polinucleótidos se encuentran posicionados en una orientación de cabeza con cola en el caso de que el extremo 5' de la cadena codificante de un polinucleótido sea contigua al extremo 3' de la cadena codificante del otro polinucleótido, de manera que la dirección de transcripción de cada polinucleótido transcurre en la misma dirección que la del otro polinucleótido. La expresión "cabeza con cola" puede abreviarse (5')-a-(3') y también puede indicarse con los símbolos ( $\rightarrow\rightarrow$ ) o ( $5'\rightarrow 3'5'\rightarrow 3'$ ).

20 La expresión "cadena abajo" se refiere a una secuencia de nucleótidos que se encuentra situada en orientación 3' respecto a la secuencia de nucleótidos de referencia. En particular, las secuencias de nucleótidos cadena abajo generalmente se refieren a las secuencias siguientes al punto de inicio de transcripción. Por ejemplo, el codón de inicio de traducción de un gen se encuentra situado cadena abajo del sitio de inicio de transcripción.

25 La expresión "cadena arriba" se refiere a una secuencia de nucleótidos que se encuentra situada en orientación 5' respecto a la secuencia de nucleótidos. En particular, las secuencias de nucleótidos situadas cadena arriba generalmente se refieren a secuencias que se encuentran situadas en el lado 5' de una secuencia codificante o de un punto de inicio de transcripción. Por ejemplo, la mayoría de promotores se encuentra situada cadena arriba del sitio de inicio de transcripción.

30 Las expresiones "endonucleasa de restricción" y "enzima de restricción" se refieren a un enzima que se une y corta dentro de una secuencia específica de nucleótidos dentro del ADN de doble cadena.

35 La expresión "recombinación homóloga" se refiere a la inserción de una secuencia foránea de ADN en otra molécula de ADN, por ejemplo la inserción de un vector en un cromosoma. Preferentemente, el vector presenta como diana un sitio cromosómico específico para la recombinación homóloga. Para la recombinación homóloga específica, el vector contendrá regiones suficientemente largas de homología respecto a secuencias del cromosoma para permitir la unión complementaria y la incorporación del vector en el cromosoma. Una mayor longitud de las regiones de homología y mayores grados de similitud de secuencias pueden incrementar la eficiencia de la recombinación homóloga.

40 Pueden utilizarse varios procedimientos conocidos de la técnica para propagar un polinucleótido mencionado en la presente memoria. Tras establecer un sistema huésped y condiciones de cultivo adecuados, los vectores de expresión recombinante pueden propagarse y prepararse en grandes cantidades. Tal como se describe en la presente memoria, entre los vectores de expresión que pueden utilizarse se incluyen, aunque sin limitación, los vectores siguientes o derivados de los mismos: virus humanos o animales tales como el virus Vaccinia o adenovirus; virus de insecto tales como baculovirus, vectores de levadura, vectores bacteriófagos (por ejemplo lambda) y vectores ADN plásmidos y cósmidos, entre otros.

50 Un "vector" se refiere a cualquier medio para la clonación y/o transferencia de un ácido nucleico al interior de una célula huésped. Un vector puede ser un replicón al que puede unirse otro segmento de ADN de manera que se produzca la replicación del segmento unido. Un "replicón" es cualquier elemento genético (por ejemplo plásmido, fago, cósmido, cromosoma, virus) que funciona como unidad autónoma de replicación del ADN *in vivo*, es decir, que es capaz de replicación bajo su propio control. El término "vector" incluye medios tanto víricos como no víricos para introducir el ácido nucleico en una célula *in vitro*, *ex vivo* o *in vivo*. Puede utilizarse un gran número de vectores conocidos de la técnica para manipular los ácidos nucleicos, para incorporar elementos de respuesta y promotores dentro de genes, etc. Entre los posibles vectores se incluyen, por ejemplo, plásmidos o virus modificados, incluyendo, por ejemplo bacteriófagos tales como derivados de lambda, o plásmidos tales como derivados del plásmido pBR322 o el pUC, o el vector Bluescript. Por ejemplo, la inserción de los fragmentos de ADN correspondientes a elementos de respuesta y promotores en un vector adecuado puede llevarse a cabo mediante ligación de los fragmentos de ADN apropiados en un vector seleccionado que presente extremos cohesivos complementarios. Alternativamente, los extremos de las moléculas de ADN pueden modificarse enzimáticamente o puede producirse cualquier sitio mediante ligación de las secuencias de nucleótidos (conectores) a los extremos de ADN. Este tipo de vectores puede manipularse para que contenga genes marcadores seleccionables que permitan la selección de las células que han incorporado el marcador en el genoma celular. Estos marcadores permiten la

identificación y/o selección de las células huésped que incorporan y expresan las proteínas codificadas por el marcador.

Los vectores víricos, y particularmente los vectores retrovíricos, han sido utilizados en una amplia diversidad de aplicaciones de administración génica en células, así como en sujetos animales vivos. Entre los vectores víricos que pueden utilizarse se incluyen, aunque sin limitación, vectores retrovirus, virus adenoasociados, virus pox, baculovirus, virus Vaccinia, virus herpes simplex, virus de Epstein-Barr, adenovirus, geminivirus y caulimovirus. Entre los vectores no víricos se incluyen plásmidos, liposomas, lípidos cargados eléctricamente (citofectinas), complejos de ADN-proteína y biopolímeros. Además de un ácido nucleico, un vector también puede comprender una o más regiones reguladoras y/o marcadores seleccionables que resultan útiles para seleccionar, medir y realizar un seguimiento de los resultados de la transferencia de ácidos nucleicos (transferencia a qué tejidos, duración de la expresión, etc.).

El término "plásmido" se refiere a un elemento extracromosómico que con frecuencia porta un gen que no es parte del metabolismo central de la célula, y que habitualmente se encuentra en forma de moléculas de ADN circular de doble cadena. Dichos elementos pueden ser secuencias de replicación autónoma, secuencias que se integran en el genoma, secuencias de fago o de nucleótidos, ADN o ARN lineal, circular o superenrollado de una cadena o de doble cadena, derivado a partir de cualquier fuente, en el que se han unido varias secuencias de nucleótidos o se han recombinado en una única construcción que es capaz de introducir un fragmento promotor y una secuencia de ADN para un producto génico seleccionado conjuntamente con la secuencia 3' no traducida apropiada en una célula.

Un "vector de clonación" es un "replicón", que es una longitud unitaria de un ácido nucleico, preferentemente ADN, que se replica secuencialmente y que comprende un origen de replicación, tal como un plásmido, fago o cósmido, al que puede unirse otro segmento de ácidos nucleicos de manera que se produzca la replicación del segmento unido. Los vectores de clonación pueden ser capaces de replicarse en un tipo celular y expresarse en otro ("vector lanzadera").

Los vectores pueden introducirse en las células huésped deseadas mediante procedimientos conocidos de la técnica, por ejemplo transfección, electroporación, microinyección, transducción, fusión celular, DEAE-dextrano, precipitación con fosfato cálcico, lipofección (fusión de lisosomas), utilización de una pistola génica, o de un transportador de vector de ADN (ver, por ejemplo, Wu *et al.*, J. Biol. Chem. 267:963-967, 1992; Wu y Wu, J. Biol. Chem. 263:14621-14624, 1988; y Hartmut *et al.*, solicitud de patente canadiense nº 2.012.311, presentada el 15 de marzo de 1990).

También puede introducirse *in vivo* un polinucleótido según la descripción mediante lipofección. Durante la última década, se ha producido un uso creciente de los liposomas para el encapsulado y transfección *in vitro* de ácidos nucleicos. Los lípidos catiónicos sintéticos diseñados para limitar las dificultades y peligros durante la transfección mediada por liposomas pueden utilizarse para preparar liposomas para la transfección *in vivo* de un gen codificante de un marcador (Felgner *et al.*, PNAS 84:7413, 1987; Mackey *et al.*, Proc. Natl. Acad. Sci. U.S.A. 85:8027-8031, 1988; y Ulmer *et al.*, Science 259:1745-1748, 1993). La utilización de lípidos catiónicos puede promover el encapsulado de ácidos nucleicos de carga negativa, y también la fusión con membranas celulares cargadas negativamente (Felgner y Ringold, Science 337:387-388, 1989). Se describen compuestos y composiciones de lípidos particularmente útiles para la transferencia de ácidos nucleicos en las publicaciones de patente internacional nº WO95/18863 y nº WO96/17823 y en la patente US nº 5.459.127. La utilización de lipofección para introducir genes exógenos en los órganos específicos *in vivo* presenta ciertas ventajas prácticas. El direccionamiento molecular de los liposomas a células específicas representa una ventaja. Resulta evidente que dirigir la transfección a tipos celulares particulares resultaría particularmente preferente en un tejido con heterogeneidad celular, tal como el páncreas, el hígado, el riñón y el cerebro. Los lípidos pueden acoplarse químicamente a otras moléculas para el fin de direccionamiento (Mackey *et al.*, *supra*, 1988). Los péptidos que se dirigen, por ejemplo hormonas o neurotransmisores, y proteínas tales como anticuerpos, o moléculas no peptídicas, podrían acoplarse a los liposomas químicamente.

Otras moléculas también resultan útiles para facilitar la transfección de un ácido nucleico *in vivo*, tales como un oligopéptido catiónico (por ejemplo en la patente nº WO95/21931), los péptidos derivados de proteínas de unión a ADN (por ejemplo en la patente nº WO96/25508) o un polímero catiónico (por ejemplo la patente nº WO95/21931).

También resulta posible introducir un vector *in vivo* en forma de plásmido de ADN desnudo (ver las patentes US nº 5.693.622, nº 5.589.466 y nº 5.580.859). También pueden utilizarse enfoques de direccionamiento de ADN mediado por receptores (Curiel *et al.*, Hum. Gene Ther. 3:147-154, 1992; y Wu y Wu, J. Biol. Chem. 262:4429-4432, 1987).

El término "transfección" se refiere a la incorporación de ARN o ADN exógeno o heterólogo por parte de una célula. Una célula ha sido "transfectada" por ARN o ADN exógeno o heterólogo en el caso de que dicho ARN o ADN haya sido introducido en el interior de la célula. Una célula ha sido "transformada" por ARN o ADN exógeno o heterólogo en el caso de que el ARN o ADN comporte un cambio fenotípico. El ARN o ADN transformante puede integrarse (unirse covalentemente) en ADN cromosómico que constituye el genoma de la célula.

El término "transformación" se refiere a la transferencia de un fragmento de ácidos nucleicos al interior del genoma de un organismo huésped, resultando en la herencia genéticamente estable. Los organismos huésped que contienen los fragmentos de ácidos nucleicos transformados se denominan organismos "transgénicos" o "recombinantes" o "transformados".

5 La expresión "región genética" se refiere a una región de una molécula de ácidos nucleicos o de una secuencia de nucleótidos que comprende un gen codificante de un polipéptido.

10 Además, el vector recombinante que comprende un polinucleótido según la descripción puede incluir uno o más orígenes de replicación en los huéspedes celulares en los que se busca su amplificación o su expresión, marcadores o marcadores seleccionables.

15 La expresión "marcador seleccionable" se refiere a un factor de identificación, habitualmente un gen de resistencia a antibiótico o de resistencia química, que puede ser seleccionado basándose en el efecto de dicho gen marcador, es decir, en la resistencia a un antibiótico, la resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes y similares, en los que el efecto se utiliza para seguir el rastro de la herencia de un ácido nucleico de interés y/o para identificar una célula u organismo que ha heredado el ácido nucleico de interés. Entre los ejemplos de genes de marcador seleccionable conocidos y utilizados en la técnica se incluyen: genes que proporcionan resistencia a la ampicilina, estreptomycin, gentamicina, canamicina, higromicina, herbicida bialafos, sulfonamida y similares, y genes que se utilizan como marcadores fenotípicos, es decir, genes reguladores de la antocianina, gen de la isopentanol-transferasa y similares.

20 La expresión "gen informador" se refiere a un ácido nucleico codificante de un factor de identificación que puede identificarse basándose en el efecto del gen informador, en el que el efecto se utiliza para seguir el rastro de la herencia de un ácido nucleico de interés, o para identificar una célula u organismo que ha heredado el ácido nucleico de interés y/o para medir la inducción de la expresión génica o la transcripción. Entre los ejemplos de genes informadores conocidos y utilizados en la técnica se incluyen: luciferasa (Luc), proteína fluorescente verde (GFP), cloranfenicol acetiltransferasa (CAT),  $\beta$ -galactosidasa (LacZ),  $\beta$ -glucuronidasa (Gus) y similares. Los genes de marcador seleccionable también pueden considerarse genes informadores.

25 El término "promotor" se refiere a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una secuencia codificante se encuentra situada en el lado 3' de una secuencia de promotor. Los promotores pueden derivarse en su totalidad de un gen nativo, o pueden estar compuestos de diferentes elementos derivados de diferentes promotores presentes en la Naturaleza, o incluso puede comprender segmentos de ADN sintético. El experto en la materia entenderá que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos celulares, o en diferentes etapas del desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas. Los promotores que causan que un gen se expresa en la mayoría de tipos celulares en la mayoría de ocasiones se denominan comúnmente "promotores constitutivos". Los promotores que causan que un gen se expresa en un tipo celular específico se denominan comúnmente "promotores específicos de un tipo celular" o "promotores específicos de un tejido". Los promotores que causan que un gen se expresa en una etapa específica del desarrollo o de la diferenciación celular se denominan comúnmente "promotores específicos del desarrollo" o "promotores específicos de la diferenciación celular". Los promotores que resultan inducidos y que causan que un gen se exprese tras la exposición o el tratamiento de la célula con un agente, molécula biológica, compuesto químico, ligando, luz o similar, que induce el promotor, se denominan comúnmente "promotores inducibles" o "promotores regulables". Se reconoce además que, debido a que en la mayoría de casos no se han definido por completo los límites exactos de las secuencias reguladoras, algunos fragmentos de ADN de longitudes diferentes pueden presentar una actividad promotora idéntica.

30 Una "secuencia promotora" es una región reguladora del ADN capaz de unirse a una ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante situada cadena abajo (dirección 3'). Para los fines de definición de la presente invención, la secuencia promotora se encuentra limitada en su extremo 3' por el sitio de inicio de transcripción y se extiende cadena arriba (dirección 5'), incluyendo el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables sobre el nivel de fondo. Dentro de la secuencia del promotor se encuentra un sitio de inicio de la transcripción (definido convenientemente, por ejemplo mediante mapeado con nucleasa S1), así como dominios de unión de proteína (secuencias de consenso) responsables de la unión de la ARN polimerasa.

35 Una secuencia codificante se encuentra "bajo el control" de secuencias de control transcripcional y traduccional en una célula en el caso de que la ARN polimerasa transcriba la secuencia codificante en ARNm, que después es cortado y ARN trans-empalmado (en el caso de que la secuencia codificante contenga intrones) y traducido en la proteína codificada por la secuencia codificante.

40 Las "secuencias de control transcripcional y traduccional" son secuencias reguladoras del ADN, tales como promotores, intensificadores, terminadores y similares, que permiten la expresión de una secuencia codificante en una célula huésped. En las células eucarióticas, las señales de poliadenilación son secuencias de control.

La expresión "elemento de respuesta" se refiere a uno o más elementos de ADN de acción en cis que proporcionan sensibilidad a un promotor mediada por la interacción con los dominios de unión a ADN del primer gen quimérico. Este elemento de ADN puede ser de secuencia palindrómica (perfecta o imperfecta) o estar compuesta de motivos de secuencia o hemisitios separados por un número variable de nucleótidos. Los hemisitios pueden ser similares o idénticos, y encontrarse separados por un número variable de nucleótidos. Los hemisitios pueden ser similares o idénticos y encontrarse dispuestos en forma de repeticiones directas o inversas o en forma de hemisitio único o multímeros de hemisitios contiguos en tándem. El elemento de respuesta puede comprender un promotor mínimo aislado a partir de diferentes organismos dependiendo de la naturaleza de la célula u organismo en el que se incorporará el elemento de respuesta. El dominio de unión a ADN de la primera proteína híbrida se une, en presencia o en ausencia de un ligando, a la secuencia de ADN de un elemento de respuesta para iniciar o suprimir la transcripción de uno o más genes situados cadena abajo bajo la regulación de dicho elemento de respuesta. Entre los ejemplos de secuencias de ADN para elementos de respuesta del receptor natural de la ecdisona se incluyen: RRGG/TTCANTGAC/ACY (ver Cherbas L., *Genes Dev.* 5:120-131, 1991), AGGTCAN(n)AGGTCA, en el que N(n) puede ser uno o más nucleótidos espaciadores (ver D'Avino PP. *et al.*, *Mol. Cell. Endocrinol.* 113:1-9, 1995) y GGGTTGAATGAATTT (ver Antoniewski C. *et al.*, *Mol. Cell. Biol.* 14:4465-4474, 1994).

La expresión "funcionalmente ligado" se refiere a la asociación de secuencias de ácidos nucleicos en un único fragmento de ácidos nucleicos de manera que la función de una se encuentra afectada por la otra. Por ejemplo, un promotor se encuentra funcionalmente ligado a una secuencia codificante en el caso de que pueda presentar un efecto sobre la expresión de esa secuencia codificante (es decir, que la secuencia codificante se encuentre bajo el control transcripcional del promotor). Las secuencias codificantes pueden encontrarse funcionalmente ligadas a secuencias reguladoras en orientación de sentido o antisentido.

El término "expresión", tal como se utiliza en la presente memoria, se refiere a la transcripción y acumulación estable de ARN de sentido (ARNm) o antisentido derivado de un ácido nucleico o polinucleótido. La expresión también puede referirse a la traducción de ARNm en una proteína o polipéptido.

Las expresiones "casete", "casete de expresión" y "casete de expresión génica" se refieren a un segmento de ADN que puede insertarse en un ácido nucleico o polinucleótido en sitios de restricción específicos o mediante recombinación homóloga. El segmento de ADN comprende un polinucleótido que codifica un polipéptido de interés, y el casete y sitios de restricción están diseñados para garantizar la inserción del casete en el marco de lectura correcto para la transcripción y la traducción. La expresión "casete de transformación" se refiere a un vector específico que comprende un polinucleótido que codifica un polipéptido de interés y que presenta elementos además del polinucleótido que facilitan la transformación de una célula huésped particular. Los casetes, casetes de expresión, casetes de expresión génica y casetes de transformación de la descripción también pueden comprender elementos que permitan la expresión incrementada de un polinucleótido codificante de un polipéptido de interés en una célula huésped. Entre estos elementos pueden incluirse, aunque sin limitación, un promotor, un promotor mínimo, un intensificador, un elemento de respuesta, una secuencia de terminador, una secuencia de poliadenilación y similares.

Para los fines de la presente invención, la expresión "interruptor génico" se refiere a la combinación de un elemento de respuesta asociado a un promotor, y un sistema basado en EcR que, en presencia de uno o más ligandos, modula la expresión de un gen en el que se han incorporado el elemento de respuesta y el promotor.

Los términos "modulan" y "modula" se refieren a que induce, reduce o inhibe la expresión de los ácidos nucleicos o genes, resultando en la inducción, reducción o inhibición respectiva de la producción de proteínas o polipéptidos.

Los plásmidos o vectores según la descripción pueden comprender además por lo menos un promotor adecuado para controlar la expresión de un gen en una célula huésped. La expresión "vector de expresión" se refiere a un vector, plásmido o vehículo diseñado para permitir la expresión de una secuencia de ácidos nucleicos insertada tras la transformación en el huésped. El gen clonado, es decir, la secuencia de ácidos nucleicos insertada, habitualmente se sitúa bajo el control de elementos de control tales como un promotor, un promotor mínimo, un intensificador o similar. Las regiones de control del inicio o los promotores, que resultan útiles para controlar la expresión de un ácido nucleico en la célula huésped deseada, son numerosos y resultarán familiares para el experto en la materia. Prácticamente cualquier promotor capaz de controlar estos genes resulta adecuado para la presente invención, incluyendo, aunque sin limitación, promotores víricos, promotores bacterianos, promotores animales, promotores de mamífero, promotores sintéticos, promotores constitutivos, promotores específicos de tejido, promotores específicos del desarrollo, promotores inducibles, promotores regulados por la luz, *CYC1*, *HIS3*, *GAL1*, *GAL4*, *GAL10*, *ADH1*, *PGK*, *PHO5*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO*, *TPI*, promotores fosfatasa alcalina (útiles para la expresión en *Saccharomyces*), promotor *AOX1* (útil para la expresión en *Pichia*); promotores b-lactamasa, *lac*, *ara*, *tet*, *trp*, *IP<sub>L</sub>*, *IP<sub>R</sub>*, *T7*, *tac* y *trc* (útiles para la expresión en *Escherichia coli*); regulados por la luz, específicos de semillas, específicos de polen, específicos ováricos, relacionados con la patogénesis o con enfermedades, 35S del virus del mosaico de la coliflor, mínimo de 35S del CMV, virus del mosaico de la vena de la yuca (CsVMV), proteína de unión a la clorofila a/b, ribulosa-1,5-bifosfato carboxilasa, específico de brote, específico de raíz, quitinasa, inducible por estrés, virus baciliforme del tungro del arroz, super-promotor vegetal, leucina aminopeptidasa de la patata, nitrato reductasa, manopina sintasa, nopalina sintasa, ubiquitina, proteína zeína y promotores antocianina (útiles para la

expresión en células vegetales); entre los promotores animales y de mamífero conocidos de la técnica se incluyen, aunque sin limitación, la región promotora temprana de SV40 (SV40e), el promotor contenido en la repetición terminal larga 3' (LTR) del virus del sarcoma de Rous (RSV), los promotores del gen E1A o promotor tardío mayor (MLP) de los adenovirus (Ad), el promotor temprano de citomegalovirus (CMV), el promotor timidina quinasa (TK) del virus del herpes simplex (HSV), un promotor IE1 de baculovirus, un promotor de factor I alfa de elongación (EF1), un promotor fosfoglicerato-quinasa (PGK), un promotor ubiquitina (Ubc), un promotor albúmina, las secuencias reguladoras del promotor metalotioneína-L de ratón y las regiones de control transcripcional, los promotores ubicuos (HPRT, vimentina,  $\alpha$ -actina, tubulina y similares), los promotores de los filamentos intermedios (desmina, neurofilamentos, queratina, GFAP y similares), los promotores de genes terapéuticos (del tipo MDR, CFTR o factor VIII, y similares), promotores relacionados con patogénesis o enfermedades, y promotores que muestran especificidad de tejido y han sido utilizados en animales transgénicos, tales como la región de control génico de la elastasa I, que es activa en las células acinares pancreáticas; la región de control génico de la insulina activa en las células beta pancreáticas, la región de control génico de inmunoglobulinas activa en las células linfoides, la región de control vírico del tumor mamario de ratón activa en las células testiculares, mamarias, linfoides y mastocitos; el gen albúmina, las regiones de control Apo AI y Apo AII activas en el hígado, la región de control génico de la alfa-fetoproteína activa en el hígado, las regiones de control génico de alfa1-antitripsina activas en el hígado, la región activa de control génico de la beta-globina en las células mieloides, la región de control génico de la proteína básica de la mielina activas en células oligodendrocíticas en el cerebro, la región de control génico de la cadena ligera 2 de la miosina activa en el músculo esquelético y la región de control génico de la hormona liberadora de gonadotropina activa en el hipotálamo, el promotor piruvato quinasa, el promotor vilina, el promotor de la proteína intestinal de unión a ácidos grasos, el promotor de la  $\alpha$ -actina de las células musculares lisas, y similares. Además, dichas secuencias de expresión pueden modificarse mediante la adición de secuencias intensificadores o reguladoras, y similares.

Entre los intensificadores que pueden utilizarse en formas de realización de la invención se incluyen, aunque sin limitación: un intensificador de SV40, un intensificador de citomegalovirus (CMV), un intensificador del factor de elongación 1 (EF1), los intensificadores de levaduras, los intensificadores de gene víricos, y similares.

Las regiones de control de la terminación, es decir, las secuencias de terminador o de poliadenilación, también pueden derivarse a partir de diversos genes nativos a los huéspedes preferentes. Opcionalmente, puede resultar innecesario un sitio de terminación; sin embargo, resulta más preferente su inclusión. En una forma de realización preferente de la invención, la región de control de terminación puede comprender o puede haberse derivado a partir de una secuencia sintética, señal sintética de poliadenilación, una señal de poliadenilación tardía de SV40, una señal de poliadenilación de SV40, una señal de poliadenilación de la hormona de crecimiento bovina (BGH), secuencias de terminador vírico, o similares.

Las expresiones "secuencias 3' no codificantes" o "región 3' no traducida (UTR)" se refieren a secuencias de DAN situadas cadena abajo (3') de una secuencia codificante y pueden comprender secuencias de reconocimiento de poliadenilación [poli(A)] y otras secuencias codificantes de señales reguladoras capaces de afectar al procesamiento del ARNm o a la expresión génica. La señal de poliadenilación habitualmente se caracteriza por afectar a la adición de segmentos de ácido poliadenílico al extremo 3' del precursor de ARNm.

La "región reguladora" se refiere a una secuencia de ácidos nucleicos que regula la expresión de una segunda secuencia de ácidos nucleicos. Una región reguladora puede incluir secuencias que son naturalmente responsables de la expresión de un ácido nucleico particular (una región homóloga) o puede incluir secuencias de un origen diferente que son responsables de la expresión de diferentes proteínas o incluso puede incluir secuencias de un origen diferente que son responsables de la expresión de diferentes proteínas o incluso de proteínas sintéticas (una región heteróloga). En particular, las secuencias pueden ser secuencias de genes procarióticos, eucarióticos o víricos o secuencias derivadas que estimulan o reprimen la transcripción de un gen de un modo específico o no específico y de un modo inducible o no inducible. Entre las regiones reguladoras se incluyen orígenes de replicación, sitios de procesamiento del ARN, promotores, intensificadores, secuencias de terminación de la transcripción y secuencias de señal que dirigen al polipéptido a las rutas secretorias de la célula diana.

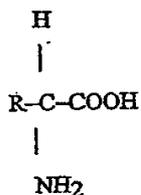
Una región reguladora de una "fuente heteróloga" es una región reguladora que no se encuentra naturalmente asociada al ácido nucleico expresado. SE incluyen entre las regiones reguladoras heterólogas, regiones reguladoras de una especie diferente, regiones reguladoras de un gen diferente, secuencias reguladoras híbridas, y secuencias reguladoras no presentes naturalmente sino diseñadas por un experto ordinario en la materia.

La expresión "transcrito de ARN" se refiere al producto resultante de la transcripción catalizada por ARN polimerasa de una secuencia de ADN. En el caso de que el transcrito de ARN sea una copia complementaria perfecta de la secuencia de ADN, se denomina transcrito primario o puede ser una secuencia de ARN derivada del procesamiento post-transcripcional del transcrito primario y se denomina ARN maduro. La expresión "ARN mensajero (ARNm)" se refiere al ARN que no presenta intrones y que puede ser traducido en proteína por la célula. El término "ADNc" se refiere a un ADN de doble cadena que es complementario y ha sido derivado a partir de ARNm. El ARN "de sentido" se refiere al transcrito de ARN que incluye el ARNm y que por lo tanto puede ser traducido en proteína por la célula.

La expresión "ARN antisentido" se refiere a un transcrito de ARN que es complementario a la totalidad o parte de un transcrito primario diana o ARNm, y que bloquea la expresión de un gen diana.

5 La complementariedad de un ARN antisentido puede ser respecto a cualquier parte del transcrito génico específico, es decir, en la secuencia 5' no codificante, en la secuencia 3' no codificante, o en la secuencia codificante. La expresión "ARN funcional" se refiere a ARN antisentido, ARN ribozima u otro ARN que no se traduce pero que presenta un efecto sobre procesos celulares.

10 Un "polipéptido" es un compuesto polimérico que consta de residuos aminoácidos unidos covalentemente. Los aminoácidos presentan la estructura general siguiente:



15 Los aminoácidos se clasifican en siete grupos basándose en la cadena lateral R: (1) cadenas laterales alifáticas, (2) cadenas laterales que contienen un grupo hidroxílico (OH), (3) cadenas laterales que contienen átomos de azufre, (4) cadenas laterales que contienen un grupo ácido o amida, (5) cadenas laterales que contienen un grupo básico, (6) cadenas laterales que contienen un anillo aromático, y (7) prolina, un iminoácido en el que la cadena lateral se encuentra fusionado con el grupo amino. Un polipéptido de la descripción preferentemente comprende por lo menos aproximadamente 14 aminoácidos.

20 Una "proteína" es un polipéptido que realiza un papel estructural o funcional en una célula viva.

25 Un "polipéptido aislado" o una "proteína aislada" es un polipéptido o proteína que se encuentra sustancialmente libre de aquellos compuestos que normalmente se encuentran asociados a los mismos en su estado natural (por ejemplo otras proteínas o polipéptidos, ácidos nucleicos, carbohidratos y lípidos). El término "aislado" no pretende excluir mezclas artificiales o sintéticas con otros compuestos, o la presencia de impurezas que no interfieren con la actividad biológica, y que pueden encontrarse presentes, por ejemplo, debido a una purificación incompleta, la adición de estabilizadores o la formación de un compuesto en una preparación farmacéuticamente aceptable.

30 Un "fragmento" de un polipéptido según la descripción se entiende que hace referencia a un polipéptido cuya secuencia de aminoácidos es más corta que la del polipéptido de referencia y que comprende, a lo largo de la porción completa con dichos polipéptidos de referencia, una secuencia de aminoácidos idéntica. Dichos fragmentos pueden, en caso apropiado, incluirse en un polipéptido de mayor tamaño del que forman una parte. Dichos fragmentos de un polipéptido según la descripción pueden presentar una longitud de por lo menos 2 a 300 aminoácidos.

35 Una "variante" de un polipéptido o proteína es cualquier análogo, fragmento, derivado o mutante que se deriva de un polipéptido o proteína y que conserva por lo menos una propiedad biológica del polipéptido o proteína. Pueden existir diferentes variantes del polipéptido o proteína en la Naturaleza. Estas variantes pueden ser variaciones alélicas caracterizadas por diferencias en las secuencias de nucleótidos del gen estructural codificante de la proteína, o pueden implicar una modificación diferente durante el procesamiento o después de la traducción. El experto en la materia puede producir variantes que presentan sustituciones, deleciones, adiciones o reemplazamientos de un único o múltiples aminoácidos. Estas variantes pueden incluir, entre otros, (a) variantes en las que se sustituyen uno o más residuos aminoácidos por aminoácidos conservadores o no conservadores, (b) variantes en las que se añaden uno o más aminoácidos al polipéptido o proteína, (c) variantes en las que uno o más de los aminoácidos incluye un grupo sustituyente, y (d) variantes en las que el polipéptido o proteína se fusiona con otro polipéptido, tal como albúmina sérica. Las técnicas para obtener estas variantes, incluyendo las técnicas genéticas (supresiones, deleciones, mutaciones, etc.), químicas y enzimáticas, son conocidas por el experto ordinario en la materia. Un polipéptido variante preferentemente comprende por lo menos aproximadamente 14 aminoácidos.

40 Una "proteína heteróloga" se refiere a una proteína no producida naturalmente en la célula.

55 Una "proteína madura" se refiere a un polipéptido procesado post-traduccionalmente, es decir, uno del que cualquier prepéptido o propéptido presente en el producto primario de traducción ha sido eliminado. La proteína "precursora" se refiere al producto primario de la traducción del ARNm, es decir, con los prepéptidos y propéptidos todavía presentes. Los prepéptidos y propéptidos pueden encontrarse limitados, aunque no necesariamente, a señales de localización intracelular.

La expresión "péptido de señal" se refiere a un polipéptido aminoterminal que precede a la proteína madura secretada. El péptido de señal se escinde de la proteína madura y por lo tanto no se encuentra presente en la misma. Los péptidos de señal presentan la función de dirigir y traslocar las proteínas secretadas a través de las membranas celulares. El péptido de señal también se denomina proteína de señal.

Se incluye una "secuencia de señal" al inicio de la secuencia codificante de una proteína que debe expresarse sobre la superficie de una célula. Esta secuencia codifica un péptido de señal, en posición N-terminal respecto al polipéptido maduro, que dirige que la célula huésped trasloque el polipéptido. La expresión "secuencia de señal de traslocación" se utiliza en la presente memoria para referirse a este tipo de secuencia de señal. Las secuencias de señal de traslocación pueden encontrarse asociadas a una diversidad de proteínas nativas a los eucariotas y procariontes, y con frecuencia son funcionales en ambos tipos de organismo.

El término "homología" se refiere al porcentaje de identidad entre dos polinucleótidos o entre dos grupos polipeptídicos. La correspondencia entre la secuencia de un grupo y otro puede determinarse mediante técnicas conocidas de la técnica. Por ejemplo, la homología puede determinarse mediante una comparación directa entre la información de secuencia entre dos moléculas de polipéptido mediante la alineación de la información de secuencia y la utilización de programas informáticos fácilmente disponibles. Alternativamente, la homología puede determinarse mediante hibridación de los polinucleótidos bajo condiciones que forman dúplex estables entre regiones homólogas, seguido de la digestión con una o más nucleasas específicas de la forma de cadena individual y la determinación de los tamaños de los fragmentos digeridos.

Tal como se utiliza en la presente memoria, el término "homólogo" en todas sus formas gramaticales y variaciones ortográficas se refiere a la relación entre proteínas que presentan un "origen evolutivo común", incluyendo proteínas de superfamilias (por ejemplo la superfamilia de las inmunoglobulinas) y las proteínas homólogas de diferentes especies (por ejemplo la cadena ligera de la miosina, etc. (Reeck *et al.*, Cell 50:667, 1987). Estas proteínas (y los genes codificantes de las mismas) presentan homología de secuencia, tal como se pone de manifiesto por su elevado grado de similitud de secuencias. Sin embargo, en su utilización habitual y en la presente solicitud, el término "homólogo", en el caso de que se modifique con un adverbio tal como "altamente", puede referirse a la similitud de secuencias y no a un origen evolutivo común.

De acuerdo con lo anteriormente expuesto, la expresión "similitud de secuencias" en todas sus formas gramaticales se refiere al grado de identidad o correspondencia entre secuencias de ácidos nucleicos o de aminoácidos de proteínas que pueden compartir o no un origen evolutivo común (ver Reeck *et al.*, Cell 50:667, 1987).

En una forma de realización específica, dos secuencias de ADN son "sustancialmente homólogas" o "sustancialmente similares" en el caso de que por lo menos aproximadamente 50% (preferentemente por lo menos aproximadamente 75%, y todavía más preferentemente por lo menos aproximadamente 90% ó 95%) de los nucleótidos son correspondientes a lo largo de la longitud definida de las secuencias de ADN. Las secuencias que son sustancialmente homólogas pueden identificarse mediante la comparación de las secuencias utilizando programas informáticos estándares disponibles en bancos de datos de secuencias, o en un experimento de hibridación southern bajo, por ejemplo, condiciones astringentes definidas para ese sistema particular. La definición de las condiciones apropiadas de hibridación se encuentra comprendida dentro de los conocimientos del experto en la materia (ver, por ejemplo, Sambrook *et al.*, *supra*, 1989).

Tal como se utiliza en la presente memoria, la expresión "sustancialmente similar" se refiere a fragmentos de ácidos nucleicos en los que los cambios en una o más bases nucleotídicas resulta en la sustitución de uno o más aminoácidos, pero que no afectan a las propiedades funcionales de la proteína codificada por la secuencia de ADN. La expresión "sustancialmente similar" también se refiere a fragmentos de ácidos nucleicos en los que los cambios de una o más bases nucleotídicas no afecta a la capacidad del fragmento de ácidos nucleicos de mediar en la alteración de la expresión génica mediante tecnología antisentido o de cosupresión. La expresión "sustancialmente similar" también se refiere a modificaciones de los fragmentos de ácidos nucleicos de la descripción, tales como la delección o la inserción de una o más bases nucleotídicas que no afecta sustancialmente a las propiedades funcionales del transcrito resultante. Por lo tanto, se entiende que la descripción comprende más que las secuencias ejemplares específicas. Cada una de las modificaciones propuestas se encuentra perfectamente comprendida dentro de los conocimientos del experto rutinario en la materia, al igual que la determinación de la conservación de la actividad biológica de los productos codificados.

Además, el experto en la materia reconoce que secuencias sustancialmente similares comprendidas en la presente descripción también se definen a partir de su capacidad para hibridarse, bajo condiciones astringentes (0,1X SSC, SDS al 0,1%, 65°C, y lavado con 2X SSC, SDS al 0,1% seguido de 0,1X SSC, SDS al 0,1%) con las secuencias ejemplificadas en la presente memoria. Los fragmentos de ácidos nucleicos sustancialmente similares de la descripción son aquellos fragmentos de ácidos nucleicos cuyas secuencias de ADN son por lo menos 70% idénticas a la secuencia de ADN de los fragmentos de ácidos nucleicos indicados en la presente memoria. Los fragmentos de ácidos nucleicos sustancialmente similares preferentes de la descripción son aquellos fragmentos de ácidos nucleicos cuyas secuencias de ADN son por lo menos 80% idénticas a la secuencia de ADN de los fragmentos de ácidos nucleicos indicados en la presente memoria. Los fragmentos de ácidos nucleicos más preferentes son por lo

menos 90% idénticos a la secuencia de ADN de los fragmentos de ácidos nucleicos indicados en la presente memoria. Resultan todavía más preferentes los fragmentos de ácidos nucleicos que son por lo menos 95% idénticos a la secuencia de ADN de los fragmentos de ácidos nucleicos indicados en la presente memoria.

5 Dos secuencias de aminoácidos son "sustancialmente homólogas" o "sustancialmente similares" en el caso de que más de aproximadamente 40% de los aminoácidos sean idénticos, o más de 60% sean similares (funcionalmente idénticos). Preferentemente, las secuencias similares u homólogas se identifican mediante alineamiento utilizando, por ejemplo, el programa Pileup del GCG (Genetics Computer Group, Program Manual for the GCG Package, versión 7, Madison, Wisconsin).

10 La expresión "correspondiente a" se utiliza en la presente memoria para referirse a secuencias similares u homólogas, sea o no la posición exacta idéntica o diferente de la molécula para la que se mide la similitud o la homología. Una alineación de secuencias de ácidos nucleicos o de aminoácidos puede incluir espacios. De esta manera, el término "correspondiente a" se refiere a la similitud de secuencias, y no a la numeración de los residuos aminoácidos o bases nucleotídicas.

15 Una "parte sustancial" de una secuencia de aminoácidos o de nucleótidos comprende una parte suficiente de la secuencia de aminoácidos de un polipéptido o de la secuencia de nucleótidos de un gen para identificar putativamente el polipéptido o gen, mediante evaluación manual de la secuencia por parte del experto en la materia, o mediante la comparación automática computerizada de las secuencias y la identificación utilizando algoritmos tales como BLAST (Basic Local Alignment Search Tool; Altschul S.F. *et al.*, J. Mol. Biol. 215:403-410, 1993; ver también [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). En general, una secuencia de diez o más aminoácidos contiguos o de treinta o más nucleótidos resulta necesaria para identificar putativamente una secuencia polipeptídica o de ácidos nucleicos como homóloga respecto a una proteína o gen conocido. Además, con respecto a las secuencias de nucleótidos, pueden utilizarse sondas oligonucleótidas específicas de un gen que comprenden 20 a 30 nucleótidos contiguos, en procedimientos dependientes de secuencia para la identificación (por ejemplo la hibridación southern) y aislamiento de genes (por ejemplo la hibridación *in situ* de colonias bacterianas o placas de bacteriófagos). Además, pueden utilizarse oligonucleótidos cortos de 12 a 15 bases a modo de cebadores de amplificación en la PCR con el fin de obtener un fragmento de ácidos nucleicos particular que comprende los cebadores. Por consiguiente, una "parte sustancial" de una secuencia de nucleótidos comprende una parte suficiente de la secuencia para identificar y/o aislar específicamente un fragmento de ácidos nucleicos que comprende la secuencia.

20 La expresión "porcentaje de identidad", tal como se conoce en la técnica, es una relación entre dos o más secuencias polipeptídicas o entre dos o más secuencias polinucleótidas, determinada mediante la comparación de las secuencias. En la técnica, "identidad" también se refiere al grado de relación entre secuencias de polipéptido o polinucleótida, según sea el caso, determinado como la correspondencia entre cadenas de dichas secuencias. La "identidad" o "similitud" puede calcularse fácilmente mediante procedimientos conocidos, incluyendo, aunque sin limitación, los descritos en: *Computational Molecular Biology* (Lesk A.M., editor), Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects* (Smith D.W., editor), Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I* (Griffin A.M. y Griffin H.G., editores), Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology* (von Heinje G., editor), Academic Press, 1987; y *Sequence Analysis Primer* (Gribskov M. y Devereux J., editores), Stockton Press, New York, 1991. Los procedimientos preferentes para determinar la identidad están diseñados para proporcionar la correspondencia óptima entre las secuencias sometidas a ensayo. Los procedimientos para determinar la identidad y similitud se encuentran codificados en programas informáticos públicamente disponibles. Pueden llevarse a cabo cálculos de alineación de secuencias y de porcentaje de identidad utilizando el programa Megalign del conjunto de análisis bioinformático LASERGENE (DNASTAR Inc., Madison, WI). Pueden llevarse a cabo alineaciones múltiples de las secuencias utilizando el procedimiento Clustal de alineación (Higgins y Sharp, CABIOS 5:151-153, 1989) con los parámetros por defecto (PENALIZACIÓN POR HUECO=10, PENALIZACIÓN POR LONGITUD DE HUECO=10). Pueden seleccionarse los parámetros por defecto para las alineaciones por parejas utilizando el procedimiento Clustal: KTUPLO 1, PENALIZACIÓN POR HUECO=3, VENTANA=5 y DIAGONALES SALVADAS=5.

25 La expresión "programas de análisis de secuencias" se refiere a cualquier algoritmo de ordenador o programa informático que resulte útil para el análisis de secuencias de nucleótidos o de aminoácidos. Los "programas de análisis de secuencias" pueden encontrarse comercialmente disponibles o desarrollarse independientemente. Los programas de análisis de secuencias típicos incluirán, aunque sin limitación, el conjunto GCG de programas (Wisconsin Package versión 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul *et al.*, J. Mol. Biol. 215:403-410, 1990) y DNASTAR (DNASTAR, Inc., 1228, S. Park St., Madison, WI 53715, USA). Dentro del contexto de la presente solicitud se entenderá que, en el caso de que se utilicen programas de análisis de secuencias para el análisis, los resultados del análisis se basarán en los "valores por defecto" del programa referenciado, a menos que se indique lo contrario. Tal como se utiliza en la presente memoria, los "valores por defecto" se refiere a cualquier conjunto de valores o parámetros que se cargan originalmente con los programas al inicializarlos.

30 Los "genes sintéticos" pueden construirse a partir de bloques constructivos de oligonucleótidos que se sintetizan químicamente utilizando procedimientos conocidos por el experto en la materia. Estos bloques constructivos se ligan

y se hibridan formando segmentos génicos que seguidamente se ensamblan enzimáticamente para construir el gen completo. La expresión "sintetizado químicamente", referida a una secuencia de ADN, se refiere a que los nucleótidos componentes han sido ensamblados *in vitro*. La síntesis química manual de ADN puede llevarse a cabo utilizando procedimientos bien establecidos, o la síntesis química automática puede llevarse a cabo utilizando uno de entre varios aparatos disponibles comercialmente. De acuerdo con lo anterior, los genes pueden adaptarse para la expresión génica óptima basándose en la optimización de la secuencia de nucleótidos de manera que refleje el sesgo en el uso de codones de la célula huésped. El experto en la materia apreciará la probabilidad de que la expresión génica se produzca con éxito si el uso de codones se encuentra sesgado hacia aquellos codones favorecidos por el huésped. La determinación de los codones preferentes puede basarse en un estudio de los genes derivados de la célula huésped de la que se disponga de información de secuencia.

Tal como se utiliza en la presente memoria, dos o más sistemas de regulación génica individualmente operables son "ortogonales" en el caso de que: a) la modulación de cada uno de los sistemas dados por su ligando respecto, a una concentración seleccionada, resulte en un cambio medible de la magnitud de expresión del gen de ese sistema, y b) el cambio es diferente en grado estadísticamente significativo respecto al cambio de expresión de todos los demás sistemas operables simultáneamente en la célula, tejido u organismo, con independencia de la simultaneidad o secuencialidad de la modulación producida. Preferentemente, la modulación de cada sistema de regulación génica individualmente operable lleva a cabo un cambio en la expresión génica por lo menos 2 veces superior que todos los demás sistemas operables en la célula, tejido u organismo. Más preferentemente, el cambio es por lo menos 5 veces superior. Todavía más preferentemente, el cambio es por lo menos 10 veces superior. Todavía más preferentemente, el cambio es por lo menos 100 veces superior. Todavía más preferentemente, el cambio es por lo menos 500 veces superior. Idealmente, la modulación de cada uno de los sistemas dados por parte de su ligando respecto a una concentración seleccionada resulta en un cambio medible de la magnitud de expresión del gen de ese sistema y ningún cambio medible de la expresión de todos los demás sistemas operables en la célula, tejido u organismo. En estos casos, se dice que el sistema de regulación de múltiples genes inducibles es "totalmente ortogonal".

#### **Sistema de modulación de la expresión de múltiples genes de la invención**

Tal como se indica en la presente memoria, la invención de los solicitantes proporciona un sistema de regulación de múltiples genes inducibles que permite la regulación simultánea y cuantitativa de dos o más genes diferentes en la misma célula, tejido u organismo. Los solicitantes han descubierto que los sistemas basados en receptores pueden modificarse y combinarse para crear un sistema de regulación de múltiples genes inducibles que comprende una pluralidad de sistemas de regulación génica individualmente operables.

En una forma de realización específica, el sistema de regulación de múltiples genes inducibles comprende una pluralidad de sistemas de regulación génica individualmente operables, en la que:

a) cada sistema de regulación génica individualmente operable comprende:

i) uno o más polinucleótidos codificante de un complejo de receptores que comprende:

A) un dominio de unión a ADN,

B) un dominio de unión a ligando, y

C) un dominio de transactivación,

ii) un ligando,

iii) un polinucleótido que comprende:

A) un gen exógeno o endógeno, y

B) un elemento de respuesta,

en el que:

A) el gen exógeno o endógeno se encuentra bajo el control del elemento de respuesta, y

B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o en ausencia del ligando resulta en la activación o supresión del gen, y

b) cada sistema de regulación génica individualmente operable es ortogonal respecto a los demás sistemas de regulación génica individualmente operables presentes en el sistema de regulación de múltiples genes inducibles.

En otra forma de realización, se proporciona un sistema de regulación de múltiples genes inducibles que comprende una pluralidad de sistemas de regulación génica individualmente operables, en el que:

a) cada sistema de regulación génica individualmente operable comprende:

i) uno o más complejos de receptores, comprendiendo cada uno:

A) un dominio de unión a ADN,

B) un dominio de unión a ligando, y

C) un dominio de transactivación,

ii) un ligando,

iii) un polinucleótido que comprende:

A) un gen exógeno o endógeno, y

B) un elemento de respuesta,

en el que:

A) el gen exógeno o endógeno se encuentra bajo el control del elemento de respuesta, y

B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o en ausencia del ligando resulta en la activación o supresión del gen, y

b) cada sistema de regulación génica individualmente operable es ortogonal respecto a los demás sistemas de regulación génica individualmente operables presentes en el sistema de regulación de múltiples genes inducibles.

Los solicitantes han encontrado que los receptores nucleares son receptores preferentes para la utilización en los sistemas de expresión de múltiples genes inducibles de la presente descripción. Entre los receptores nucleares preferentes se incluyen los receptores nucleares de Grupo H. Entre los receptores nucleares más preferentes se incluyen los receptores de ecdisona.

En la Naturaleza, el sistema de regulación de EcR utiliza pulsos de 20-hidroiecdisona (20E), una hormona esteroidea, para regular la muda y otros procesos del desarrollo en insectos. La 20E transduce su señal a través de un complejo heterodimérico de proteínas que incluye el receptor de ecdisona (EcR) y la proteína Ultraspiracle (USP). El EcR controla la expresión de los genes sensibles a la ecdisona mediante la unión de los elementos de respuesta a la ecdisona (EcRE) presentes en sus promotores. El ADNc de EcR se clonó por primera vez a partir de *D. melanogaster*. Se encontró que tanto EcR como USP eran miembros de la superfamilia de receptores nucleares debido a que contienen los dominios característicos: A/B (transactivación), C (unión a ADN), D (bisagra) y E (unión a ligando).

En total, se han clonado veinte secuencias de EcR de especies insecto, cangrejo y garrapata (ver posteriormente). La comparación entre las secuencias de aminoácidos deducidas de dichos ADNc demostró que el dominio de unión a ADN de 66 aminoácidos se encontraba bien conservado en los EcRs, mientras que los dominios A/B, D y F no se encontraban muy bien conservados. Los residuos críticos en el dominio de unión a ligando se encontraban bien conservados. Existe una similitud de aminoácidos de aproximadamente 90% en los dominios de unión a ligando dentro de los grupos de secuencias de EcR, pero este nivel cae a 50-60% al comparar dos grupos.

De esta manera, los receptores preferentes para la utilización en los sistemas de expresión de múltiples genes inducibles de los solicitantes incluyen receptores nucleares; entre los receptores más preferentes se incluyen los receptores nucleares de Grupo H seleccionados de entre el grupo constituido por receptor de la ecdisona, receptor ubicuo (UR), receptor huérfano 1 (OR-1), receptor nuclear 1 de hormona esteroidea (NER-1), proteína 15 de interacción con RXR (RIP-15), el receptor X hepático  $\beta$  (LXR $\beta$ ), la proteína de tipo receptor de hormona esteroidea (RLD-1), el receptor hepático X (LXR), el receptor X hepático  $\alpha$  (LXR $\alpha$ ), el receptor X de farnesoide (FXR), la proteína de interacción con receptores 14 (RIP-14) y el receptor de farnesol (HRR-1); entre los receptores todavía más preferentes se incluyen los receptores de la ecdisona.

La presente invención resulta útil para aplicaciones tales como la terapia génica, la producción a gran escala de proteínas y anticuerpos, los ensayos de cribado de alto rendimiento basados en células, los ensayos de cribado de ligandos ortogonales, la genómica funcional, la proteómica, la metabolómica, los biosensores y la regulación de caracteres en organismos transgénicos, en los que el control de los niveles de expresión génica resulta deseable.

Una ventaja de la invención de los solicitantes es que proporciona un medio para regular la expresión de múltiples genes y de ajustar los niveles de expresión para adaptarse a las necesidades del usuario.

En particular, los solicitantes describen en la presente memoria un nuevo sistema de expresión de múltiples genes inducibles que comprende por lo menos dos sistemas de expresión génica individualmente operables. Cada sistema de expresión génica individualmente operable comprende por lo menos un primer casete de expresión génica que comprende un elemento de respuesta, un promotor funcionalmente ligado a un polinucleótido o gen de interés que debe expresarse y el polinucleótido del gen de interés que debe expresarse. La inducción del primer casete de expresión génica puede llevarse a cabo utilizando por lo menos un segundo casete de expresión génica.

En una forma de realización específica, el segundo casete de expresión génica comprende un polinucleótido que codifica un polipéptido que comprende un dominio de unión a ADN que se une al elemento de respuesta del primer casete de expresión génica, un dominio de transactivación que transactiva el promotor del primer casete de expresión génica y un dominio de unión a ligando. La presente forma de realización utiliza un sistema de expresión génica basado en un "único interruptor" para expresar el primer casete de expresión génica que comprende el polinucleótido o gen de interés. Un sistema de expresión génica basado en un "único interruptor" es uno en el que el dominio de transactivación, el dominio de unión a ADN y el dominio de unión a ligando se encuentran en un polipéptido codificado.

Alternativamente, el sistema de modulación de la expresión génica puede ser un sistema de modulación de la expresión génica basado en un "doble interruptor" o en "dos híbridos", en el que el dominio de transactivación y el dominio de unión a ADN se encuentran situados en dos polipéptidos codificados diferentes. En la presente forma de realización específica, la inducción del primer casete de expresión génica puede llevarse a cabo utilizando por lo menos un segundo casete de expresión génica y un tercer casete de expresión. Preferentemente, el segundo casete de expresión génica comprende un polinucleótido codificante de un polipéptido que comprende un dominio de unión a ADN que se une al elemento de respuesta del primer casete de expresión génica y un dominio de unión a ligando; y el tercer casete de expresión génica comprende un polinucleótido codificante de un polipéptido que comprende un dominio de transactivación que transactiva el promotor del primer casete de expresión génica y un dominio de unión a ligando.

En una forma de realización preferente, el sistema de expresión de múltiples genes inducibles de la descripción comprende por lo menos dos sistemas de modulación de la expresión génica, en el que cada sistema de modulación de la expresión génica modulable comprende:

a) i) un primer casete de expresión génica que comprende un polinucleótido que codifica un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse, y un dominio de unión a ligando de receptor nuclear,

ii) un ligando, y

iii) un segundo casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del polipéptido codificado del primer casete de expresión génica, B) un promotor que resulta activado por el dominio de transactivación del polipéptido codificado del primer casete de expresión génica, y C) un gen cuya expresión debe modularse,

b) i) un primer casete de expresión génica que comprende un polinucleótido que codifica un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse, y un dominio de unión a ligando de receptor nuclear,

ii) un segundo dominio de unión a ligando de receptor nuclear seleccionado de entre el grupo constituido por un dominio de unión a ligando de receptor X del retinoide de vertebrado, un dominio de unión a ligando de receptor X del retinoide de invertebrado, un dominio de unión a ligando de la proteína Ultraspiracle, y un dominio de unión a ligando quimérico que comprende dos fragmentos polipeptídicos, en el que el primer fragmento polipeptídico procede de un dominio de unión a ligando de receptor X del retinoide de vertebrado, un dominio de unión a ligando del receptor X del retinoide de invertebrado, o un dominio de unión a ligando de la proteína Ultraspiracle, y el segundo fragmento polipeptídico procede de un dominio diferente de unión a ligando de receptor X del retinoide de vertebrado, de unión a ligando de receptor X del retinoide de invertebrado o de unión a ligando de la proteína Ultraspiracle,

iii) un ligando, y

iv) un segundo casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del polipéptido codificado del primer casete de expresión génica, B) un promotor que resulta activado por el dominio de transactivación del polipéptido codificado del primer casete de expresión génica, y C) un gen cuya expresión debe modularse, o

c) i) un primer casete de expresión génica que comprende un polinucleótido que codifica un primer polipéptido que comprende un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse y un dominio de unión a ligando de receptor nuclear,

5 ii) un segundo casete de expresión génica que comprende un polinucleótido que codifica un segundo polipéptido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor nuclear,

10 iii) un ligando, y

15 iv) un tercer casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido del primer casete de expresión génica, B) un promotor que resulta activado por el dominio de transactivación del segundo polipéptido del segundo casete de expresión génica, y C) un gen cuya expresión debe modularse,

en el que uno de los dominios de unión a ligando de receptor nuclear de c) i) o de c) ii) es un dominio de unión a ligando de receptor nuclear de grupo H.

20 En otra forma de realización preferente, el sistema de expresión de múltiples genes inducibles de la descripción comprende por lo menos dos sistemas de modulación de la expresión génica, en los que sistema de modulación de la expresión de gen operable comprende:

25 a) i) un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse, y un dominio de unión a ligando de receptor nuclear,

ii) un ligando, y

30 iii) un casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del polipéptido de a) i), B) un promotor que resulta activado por el dominio de transactivación del polipéptido de a) i), y C) un gen cuya expresión debe modularse,

35 b) i) un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse, y un dominio de unión a ligando de receptor nuclear,

40 ii) un segundo dominio de unión a ligando de receptor nuclear, seleccionado de entre el grupo constituido por un dominio de unión a ligando de receptor X del retinoide de vertebrado, un dominio de unión a ligando del receptor X del retinoide de invertebrado, un dominio de unión a ligando de proteína Ultraspiracle, y un dominio de unión a ligando quimérico que comprende dos fragmentos polipeptídicos, en el que el primer fragmento polipeptídico procede de un dominio de unión a ligando de receptor X del retinoide de vertebrado, de un dominio de unión a ligando de receptor X del retinoide de invertebrado o de un dominio de unión a ligando de proteína Ultraspiracle, y el segundo fragmento polipeptídico procede de un dominio diferente de unión a ligando de receptor X del retinoide de vertebrado, de unión a ligando de receptor X del retinoide de invertebrado, o de un dominio de unión a ligando de proteína Ultraspiracle,

45 iii) un ligando, y

50 iv) un casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido de b) i), B) un promotor que resulta activado por el dominio de transactivación del segundo polipéptido de b) i), y C) un gen cuya expresión debe modularse; o

55 c) i) un primer polipéptido que comprende un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se debe modular y un dominio de unión a ligando de receptor nuclear,

ii) un segundo polipéptido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor,

60 iii) un ligando, y

65 iv) un casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido de c)i); B) un promotor que resulta activado por el dominio de transactivación del segundo polipéptido de c)ii); y C) un gen cuya expresión debe modularse,

en el que uno de los dominios de unión a ligando de receptor nuclear de c) i) o c) ii) es un dominio de unión a ligando de receptor nuclear de grupo H.

En una forma de realización preferente, en el caso de que el sistema de modulación de la expresión génica comprenda C), el primer polipéptido se encuentra sustancialmente libre de dominio de transactivación y el segundo polipéptido se encuentra sustancialmente libre de dominio de unión a ADN. Para los fines de la descripción, "sustancialmente libre" se refiere a que la proteína en cuestión no contiene una secuencia suficiente del dominio en cuestión para proporcionar activación o actividad de unión, en donde en el caso de que únicamente un dominio de unión a ligando de receptor nuclear sea un dominio de unión a ligando de grupo H, el otro dominio de unión a ligando de receptor nuclear puede ser cualquiera otro receptor nuclear que forme un dímero con el dominio de unión a ligando de grupo H. Por ejemplo, en el caso de que el dominio de unión a ligando de receptor nuclear de grupo H sea un dominio de unión a ligando de receptor de la ecdisona, el otro dominio de unión a ligando de receptor nuclear ("pareja") puede ser de un receptor de ecdisona, de un receptor del retinoide X de vertebrado (RXR), de un RXR de invertebrado, de una proteína Ultraspiracle (USP), o de un receptor nuclear quimérico que comprenda por lo menos dos fragmentos polipeptídicos diferentes que sean dominios de unión a ligando de receptor nuclear, seleccionados de entre el grupo constituido por un RXR de vertebrado, un RXR de invertebrado, y una USP (ver las solicitudes copendientes nº PCT/US01/09050, nº US60/294.814 y nº US60/294.819. El dominio de unión a ligando de receptor nuclear de "pareja" puede comprender además una mutación por truncado, una mutación por delección, una mutación por sustitución u otra modificación.

Preferentemente, el dominio de unión a ligando de RXR de vertebrado es de un RXR de ser humano (*Homo sapiens*), de ratón (*Mus musculus*), de rata (*Rattus norvegicus*), de pollo (*Gallus gallus*), de cerdo (*Sus scrofa domestica*), de rana (*Xenopus laevis*), de pez cebra (*Danio rerio*), de tunicado *Polyandrocarpa misakiensis* o de medusa *Tripedalia cysophora*.

Preferentemente, el dominio de unión a ligando de RXR de invertebrado es de polipéptido Ultraspiracle de langosta *Locusta migratoria* ("LmUSP"), un homólogo 1 de RXR de garrapata ixodida *Amblyomma americanum* ("AmaRXR1"), un homólogo 2 de RXR de garrapata ixodida *Amblyomma americanum* ("AmaRXR2"), un homólogo de RXR de cangrejo violinista (*Celaca pugilator*) ("CpRXR"), un homólogo de RXR de escarabajo *Tenebrio molitor* ("TmRXR"), un homólogo de RXR de abeja melífera (*Apis mellifera*) ("AmRXR"), un homólogo de RXR de áfido *Myzus persicae* ("MpRXR") o un homólogo de RXR de no díptero/no lepidóptero. Ver la solicitud copendiente provisional de patente US nº 60/294.814, presentada el 31 de mayo de 2001.

Preferentemente, el dominio de unión a ligando de RXR quimérico comprende por lo menos dos fragmentos polipeptídicos seleccionados de entre el grupo constituido por un fragmento polipeptídico de RXR de especie de vertebrado, un fragmento polipeptídico de RXR de especie de invertebrado y un fragmento polipeptídico de homólogo de RXR de especie de invertebrado no díptero/no lepidóptero. Un dominio de unión a ligando de RXR quimérico para la utilización en la presente invención puede comprender por lo menos dos fragmentos polipeptídicos de RX de dos especies diferentes, o en el caso de que la especie sea la misma, los dos o más fragmentos polipeptídicos pueden proceder de dos o más isoformas diferentes del fragmento polipeptídico del RXR de la especie (ver la solicitud copendiente provisional de patente US nº 60/294.819, presentada el 31 de mayo de 2001).

En una forma de realización específica, el gen cuya expresión debe modularse es un gen homólogo respecto a la célula huésped. En otra forma de realización específica, el gen cuya expresión debe modularse es un gen heterólogo respecto a la célula huésped.

Preferentemente, se modifica uno o más de los dominios de receptor, produciendo un interruptor génico híbrido. Típicamente, uno o más de los tres dominios, DBD, LBD y dominio de transactivación, pueden seleccionarse de una fuente diferente de la fuente de los demás dominios, de manera que los genes híbridos y las proteínas híbridas resultantes se optimicen en la célula u organismo huésped seleccionado para su actividad transactivadora, la unión complementaria del ligando y el reconocimiento de un elemento de respuesta específico. Además, el elemento de respuesta mismo puede modificarse o sustituirse con elementos de respuesta para otros dominios de proteína de unión a ADN tales como la proteína GAL-4 de levadura (ver Sadowski *et al.*, Nature 335:563-564, 1988) o la proteína LexA de *Escherichia coli* (ver Brent y Ptashne, Cell 43:729-736, 1985) o los elementos de respuesta sintéticos específicos para interacciones diana con proteínas diseñadas, modificadas y seleccionadas para dichas interacciones específicas (ver, por ejemplo, Kim *et al.*, Proc. Natl. Acad. Sci. USA 94:3616-3620, 1997) que incluyan receptores híbridos. Otra ventaja de los sistemas de dos híbridos es que permiten la elección de un promotor utilizado para controlar la expresión génica según un resultado final deseado. Este control doble puede resultar particularmente importante en áreas de terapia génica, especialmente en el caso de que se produzcan proteínas citotóxicas, debido a que puede controlarse tanto la temporización de la expresión como las células en que se produce la expresión. Al introducir genes, ligados funcionalmente a un promotor adecuado, en las células del sujeto, se controla la expresión de los genes exógenos a partir de la presencia del sistema de la presente invención. Los promotores puede estar regulados constitutiva o induciblemente o pueden ser específicos de tejido (es decir, expresados únicamente en un tipo particular de células) o ser específicos de determinadas etapas del desarrollo del organismo.

Un sistema de modulación de la expresión génica basado en receptores de la ecdisona de la presente invención puede ser heterodimérico u homodimérico. Un complejo de EcR funcional generalmente se refiere a un complejo heterodimérico de proteínas que consiste de dos miembros de la familia de los receptores esteroides, una proteína de receptor de ecdisona obtenido de diversos insectos, y una proteína Ultraspiracle (USP) o el homólogo de vertebrado de USP, la proteína receptor X del retinoide (ver Yao *et al.*, Nature 366:476-479, 1993; Yao *et al.*, Cell 71:63-72, 1992). Sin embargo, el complejo también puede ser un homodímero, tal como se indica posteriormente. El complejo funcional de receptores ecdisteroides también puede incluir una o más proteínas adicionales, tales como inmunofilinas. Algunos miembros adicionales de la familia de proteínas receptores de esteroides, conocidos como factores transcripcionales (tales como DHR38 o *beta*FTZ-1), también pueden ser parejas dependientes o independientes de ligando para EcR, USP y/o RXR. Además, pueden resultar necesarios otros cofactores, tales como proteínas generalmente conocidas como coactivadores (también denominados adaptadores o mediadores). Estas proteínas no se unen de manera específica de secuencia a ADN y no se encuentran implicadas en la transcripción basal. Pueden ejercer su efecto sobre la activación de la transcripción a través de diversos mecanismos, incluyendo la estimulación de la unión de ADN de los activadores, los efectos sobre la estructura de la cromatina o la mediación en las interacciones de activador-complejo de inicio. Entre los ejemplos de dichos coactivadores se incluyen RIP140, TIF1, RAP46/Bag-1, ARA70, SRC-1/NCoA-1, TIF2/GRIP/NCoA-2, ACTR/AIB1/RAC3/pCIP, así como la proteína B de unión al elemento de respuesta al coactivador C promiscuo, CBP/p300 (para una revisión ver Glass *et al.*, Curr. Opin. Cell Biol. 9:222-232, 1997). Además, los cofactores proteicos generalmente conocidos como correpresores (también conocidos como represores, silenciadores o mediadores silenciadores) pueden resultar necesarios para inhibir efectivamente la activación transcripcional en ausencia de ligando. Estos correpresores pueden interactuar con el receptor de ecdisona sin ligando silenciando la actividad en el elemento de respuesta. Los datos disponibles en la actualidad sugieren que la unión del ligando modifica la conformación del receptor, resultando en la liberación del correpresor y en el reclutamiento de los activadores anteriormente indicados, eliminando de esta manera su actividad silenciadora. Entre los ejemplos de correpresores se incluyen N-CoR y SMRT (para una revisión, ver Horwitz *et al.*, Mol. Endocrinol. 10:1167-1177, 1996). Estos cofactores pueden ser endógenos de la célula u organismo, o pueden añadirse exógenamente en forma de transgenes que se expresan de un modo regulado o no regulado. Los complejos homodímeros de la proteína de receptor de la ecdisona, USP o RXR también pueden ser funcionales bajo algunas circunstancias.

El complejo de receptor de ecdisona típicamente incluye proteínas que son miembros de la superfamilia de los receptores nucleares, en la que todos los miembros se caracterizan generalmente por la presencia de un dominio de transactivación amino-terminal, un dominio de unión a ADN ("DBD") y un dominio de unión a ligando ("LBD") separado del DBD por una región bisagra. Tal como se utiliza en la presente memoria, la expresión "dominio de unión a ADN" comprende una secuencia polipeptídica mínima de una proteína de unión a ADN, e incluso hasta la longitud completa de una proteína de unión a ADN, con la condición de que el dominio de unión a ADN funcione asociándose a un elemento de respuesta particular. Los miembros de la superfamilia de receptores nucleares también se caracterizan por la presencia de cuatro o cinco dominios: A/B, C, D, E y, en algunos miembros, de F (ver la patente US nº 4.981.784 y Evans, Science 240:889-895, 1988). El dominio "A/B" corresponde al dominio de transactivación, "C" corresponde al dominio de unión a ADN, "D" corresponde a la región bisagra y "E" corresponde al dominio de unión a ligando. Algunos miembros de la familia también pueden presentar otro dominio de transactivación en el lado carboxi-terminal del LBD, correspondiente a "F".

El DBD se caracteriza por la presencia de dos dedos de cinc-cisteína entre los que se encuentran dos motivos aminoácidos, la caja P y la caja D, que proporcionan especificidad a los elementos de respuesta a la ecdisona. Estos dominios pueden ser de origen natural, o modificados por delección, inserción o mutación; sintéticos; quimeras de diferentes dominios de proteínas receptores heterólogas, o una combinación de los mismos. Este receptor, al igual que un subconjunto de la familia de receptores de esteroides, también presenta regiones peor definidas que son responsables de las propiedades de heterodimerización. Debido a que los dominios de EcR, USP y RXR son de naturaleza modular, los dominios LBD, DBD y de transactivación son intercambiables.

### Casetes de expresión génica

El nuevo sistema de expresión de múltiples genes inducibles de la invención comprende casetes de expresión génica que son capaces de expresarse en una célula huésped, en la que los casetes de expresión génica comprenden, cada uno, un polinucleótido que codifica un polipéptido de interés, en forma de un polipéptido "interruptor" que induce la expresión de un polipéptido o gen de interés, o el polipéptido o gen de interés que se desea expresar mediante el sistema de expresión de múltiples genes inducibles de la invención. De esta manera, la invención de los solicitantes también proporciona casetes de expresión génica para la utilización en el sistema de expresión de múltiples genes inducibles de la invención.

En una forma de realización específica, el casete de expresión génica que es capaz de expresarse en una célula huésped comprende un polinucleótido que codifica un polipéptido seleccionado de entre el grupo constituido por: a) un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN, y un dominio de unión a ligando de receptor nuclear, b) un polipéptido que comprende un dominio de unión a ADN y un dominio de unión a ligando de receptor nuclear, y c) un polipéptido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor nuclear.

En otra forma de realización específica, la presente descripción proporciona un casete de expresión génica que es capaz de expresarse en una célula huésped, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido seleccionado de entre el grupo constituido por: a) un polipéptido híbrido que comprende un dominio de transactivación, un dominio de unión a ADN, y un dominio de unión a ligando de receptor nuclear, b) un polipéptido híbrido que comprende un dominio de unión a ADN y un dominio de unión a ligando de receptor nuclear, y c) un polipéptido híbrido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor nuclear. Un polipéptido híbrido según la descripción comprende por lo menos dos fragmentos polipeptídicos, en el que cada fragmento polipeptídico procede de una fuente diferente, es decir, un polipéptido diferente, un receptor nuclear diferente, una especie diferente, etc. El polipéptido híbrido según la descripción puede comprender por lo menos dos dominios polipeptídicos, en el que cada dominio polipeptídico procede de una fuente diferente.

En una forma de realización específica, el dominio de unión a ligando de receptor nuclear es un receptor nuclear de grupo H seleccionado de entre el grupo constituido por un receptor de ecdisona, un receptor ubicuo, un receptor huérfano 1, un NER-1, un receptor nuclear 1 de hormona esteroidea, una proteína 15 de interacción con el receptor X del retinoide, un receptor X hepático  $\beta$ , una proteína de tipo receptor de hormona esteroidea, un receptor X hepático, un receptor X hepático  $\alpha$ , un receptor X farnesoide, una proteína 14 de interacción con receptor, y un receptor de farnesol. En una forma de realización preferente, el dominio de unión a ligando de receptor nuclear es de un receptor de ecdisona.

De esta manera, la presente descripción también proporciona un casete de expresión génica que comprende un polinucleótido que codifica un polipéptido seleccionado de entre el grupo constituido por: a) un polipéptido que comprende un dominio de transactivación, un dominio de unión a ADN, y un dominio de unión a ligando de receptor de ecdisona, b) un polipéptido que comprende un dominio de unión a ADN y un dominio de unión a ligando de receptor de ecdisona, y c) un polipéptido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor de ecdisona. Preferentemente, el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido seleccionado de entre el grupo constituido por: a) un polipéptido híbrido que comprende un dominio de transactivación, un dominio de unión a ADN, y un dominio de unión a ligando de receptor de ecdisona, b) un polipéptido híbrido que comprende un dominio de unión a ADN y un dominio de unión a ligando de receptor de ecdisona, y c) un polipéptido híbrido que comprende un dominio de transactivación y un dominio de unión a ligando de receptor de ecdisona, en el que el polipéptido híbrido codificado comprende por lo menos dos fragmentos polipeptídicos, en el que cada fragmento polipeptídico procede de una fuente diferente.

El dominio de unión a ligando (LBD) del receptor de ecdisona (EcR) puede ser un EcR de invertebrado seleccionado de entre el grupo constituido por un EcR de Lepidoptera, un EcR de Diptera, un EcR de artrópodo, un EcR de ortóptero, un EcR de homóptero y un EcR de hemíptero. Preferentemente, el dominio de unión a ligando de EcR para la utilización en la presente invención procede del EcR del gusano del abeto *Choristoneura fumiferana* ("CfEcR" Kothapalli *et al.*, Dev. Genet. 17:319-30, 1995), un EcR del escarabajo amarillo de la harina, *Tenebrio molitor* ("TmEcR", Mouillet *et al.*, Eur. J. Biochem. 248:856-863, 1997), EcR de la oruga del cuerno del tabaco *Manduca sexta* ("MsEcR", Fujiwara *et al.*, Insect Biochem. Molec. Biol. 25:845-856, 1995), EcR del gusano del capullo del tabaco *Heliothis virescens* ("HvEcR", Martínez *et al.*, Insect Biochem. Mol. Biol. 29:915-30, 1999), un EcR de mosquito quironómido *Chironomus tentans* ("CtEcR", Imhof *et al.*, Insect Biochem. Molec. Biol. 23:115-124, 1993), un EcR de gusano de la seda *Bombyx mori* ("BmEcR", Swevers *et al.*, Insect Biochem. Molec. Biol. 25:857-866, 1995), un EcR de mariposa *Bicyclus anynana* ("BanEcR"), un EcR de mariposa color castaño común *Junonia coernia* ("JcEcR"), un EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR", Koelle *et al.*, Cell 67:59-77, 1991), un EcR del mosquito de la fiebre amarilla *Aedes aegypti* ("AeEcR", Cho *et al.*, Insect Biochem. Molec. Biol. 25:19-27, 1995), EcR de la moscarda *Lucilia capitata* ("LcEcR"), una EcR de la moscarda de las ovejas *Lucilia cuprina* ("LucEcR", Hannan e Hill, Insect Biochem. Molec. Biol. 27:479-488, 1997), un EcR de la moscarda azul *Calliphora vicina* ("CvEcR"), un EcR de la mosca de la fruta mediterránea *Ceratitis capitata* ("CcEcR", Verras *et al.*, Eur. J. Biochem. 265:798-808, 1999), un EcR de langosta *Locusta migratoria* ("LmEcR", Saleh *et al.*, Mol. Cell Endocrinol. 143:91-9, 1998), un EcR de áfido *Myzus persicae* ("MpEcR", solicitud publicada de patente internacional n<sup>o</sup> WO99/36520), un EcR de cangrejo violinista *Celaca pugillator* ("CpEcR", Chung *et al.*, Mol. Cell Endocrinol. 139:209-27, 1998), un EcR de garrapata ixódida *Amblyomma americanum* ("AmaEcR", Guo *et al.*, Insect Biochem. Molec. Biol. 27:945-962, 1997), un EcR de mosca blanca *Bemisia argentifoli* ("BaEcR"), solicitud provisional de patente US presentada el 26 de septiembre de 2001, o un EcR de saltahojas verde *Nephotetix cincticeps* ("NcEcR", Palli, solicitud provisional de patente US presentada el 26 de septiembre de 2001). Más preferentemente, el LBD es de un CfEcR, un DmEcR o un NcEcR.

El dominio de unión a ADN puede ser cualquier dominio de unión a ADN con un elemento de respuesta conocido, incluyendo dominios de unión a ADN sintéticos y quiméricos, o análogos, combinaciones o modificaciones de los mismos. Preferentemente, el DBD es un DBD de GAL4, un DBD de LexA, un DBD de factor de transcripción, un DBD de miembro de receptores nucleares del grupo H, un DBD de miembro de la superfamilia de los receptores nucleares de hormonas esteroideas/tiroideas, o un DBD de LacZ bacteriano. Más preferentemente, el DBD es un DBD de EcR, un DBD de GAL4 o un DBD de LexA.

El dominio de transactivación (abreviado "AD" o "TA") puede ser cualquier AD de miembro de receptores nucleares de grupo H, AD de receptores nucleares de hormona esteroidea/tiroidea, AD sintético o quimérico, AD poliglutamina, AD aminoácido básico o ácido, un AD de VP16, un AD de GAL4, un AD de NF- $\kappa$ B, un AD de BP64, un AD de dominio de activación ácido de B42 (B42AD), un dominio de transactivación de p65 (p65AD), o un análogo, combinación, o su modificación. En una forma de realización particular, el AD es AD sintético o quimérico o se obtiene de un EcR, un receptor glucocorticoide, UP16, GAL4, NF- $\kappa$ b, o un AD de dominio de activación ácido de B42. Preferentemente, el AD es un AD de EcR, un AD de VP16, un AD de B42 o un AD de p65.

La presente descripción también proporciona un casete de expresión génica que comprende: i) un elemento de respuesta que comprende un dominio reconocido por un polipéptido que comprende un dominio de unión a ADN, ii) un promotor que resulta activado por un polipéptido que comprende un dominio de transactivación, e iii) un gen cuya expresión debe modularse.

El elemento de respuesta ("RE") puede ser cualquier elemento de respuesta con un dominio de unión a ADN conocido, o un análogo, combinación o modificación del mismo. En la presente invención puede utilizarse un único RE o múltiples REs; o copias múltiples del mismo RE o dos o más REs diferentes. En una forma de realización específica, el RE es un RE de GAL4 ("GAL4RE"), LexA, un RE de receptor nuclear de grupo H, un RE de receptor nuclear de hormona esteroidea/tiroidea, o un RE sintético que reconoce un dominio de unión a ADN sintético. Preferentemente, el RE es un elemento de respuesta a ecdisona (EcRE), un GAL4RE, o un RE de LexA (operón, "op").

Puede obtenerse un dominio de unión a ADN de receptor nuclear de hormona esteroidea/tiroidea, un dominio de activación o un elemento de respuesta según la descripción a partir de un receptor nuclear de hormona esteroidea/tiroidea seleccionado de entre el grupo constituido por receptor  $\alpha$  de hormona tiroidea (TR $\alpha$ ), receptor 1 tiroideo (c-erbA-1), receptor  $\alpha$  de hormona tiroidea (THRA), receptor  $\beta$  de hormona tiroidea (TR $\beta$ ), receptor  $\beta$  de hormona tiroidea (THRB), receptor  $\alpha$  de ácido retinoico (RAR $\alpha$ ), receptor  $\beta$  de ácido retinoico (RAR $\beta$ ), hepatoma (HAP), receptor  $\gamma$  de ácido retinoico (RAR $\gamma$ ), receptor de tipo gamma del ácido retinoico (RARD), receptor  $\alpha$  activado por el proliferador de peroxisomas (PPAR $\alpha$ ), receptor  $\beta$  activado por proliferador de peroxisomas (PPAR $\beta$ ), receptor relacionado con el activador del proliferador de peroxisomas (NUC-1), receptor  $\delta$  activado por proliferador de peroxisomas (PPAR $\delta$ ), receptor relacionado con el activador del proliferador de peroxisomas (FFAR), receptor  $\gamma$  activado por proliferador de peroxisomas (PPAR $\gamma$ ), receptor huérfano codificado por la cadena no codificante del receptor  $\alpha$  de la hormona tiroidea (REVERB $\alpha$ ), receptor relacionado con v-erb A (EAR-1), receptor relacionado con v-erb (EAR-1A),  $\gamma$ , receptor huérfano codificado por la cadena no codificante del receptor  $\beta$  de la hormona tiroidea (REVERB $\beta$ ), receptor relacionado con v-erb (EAR-1 $\beta$ ), receptor nuclear huérfano BD73 (BD73), receptor relacionado con rev-erbA (RVR), proteína 126 de dedo de cinc (HZF2), proteína E75 inducible por ecdisona (E75), proteína E78 inducible por ecdisona (E78), receptor 78 de *Drosophila* (DR-78), receptor  $\alpha$  huérfano relacionado con el retinoide (ROR $\alpha$ ), receptor X del retinoide  $\alpha$  (RZR $\alpha$ ), receptor huérfano  $\beta$  relacionado con el retinoide (ROR $\beta$ ), receptor Z del retinoide  $\beta$  (RZR $\beta$ ), receptor huérfano  $\gamma$  relacionado con el retinoide (ROR $\gamma$ ), receptor Z del retinoide  $\gamma$  (RZR $\gamma$ ), receptor huérfano relacionado con el retinoide (TOR), receptor hormonal 3 (HR-3), receptor hormonal 3 de *Drosophila* (DHR-3), miohemeritina (MHR-3), receptor 3 de la hormona del crecimiento (GHR-3), receptor nuclear 3 de *C. elegans* (CNR-3), receptor hormonal 3 de *C. elegans* (CHR-3), receptor nuclear 14 de *C. elegans* (CNR-14), receptor de la ecdisona (ECR), receptor ubicuo (UR), receptor nuclear huérfano (OR-1), NER-1, proteína 15 de interacción con receptor (RIP-15), receptor X hepático  $\beta$  (LXR $\beta$ ), proteína de tipo receptor de hormona esteroidea (RLD-1), receptor X hepático (LXR), receptor X hepático  $\alpha$  (LXR $\alpha$ ), receptor X de farnesoide (FXR), proteína 14 de interacción con receptor (RIP-14), HRR-1, receptor de vitamina D (VDR), receptor nuclear huérfano (ONR-1), receptor X de pregnano (PXR), receptor esteroide y xenobiótico (SXR), receptor X de benzoato (BXR), receptor nuclear (MB-67), receptor 1 constitutivo de androstano (CAR-1), receptor  $\alpha$  constitutivo de androstano (CAR $\alpha$ ), receptor 2 constitutivo de androstano (CAR-2), receptor  $\beta$  constitutivo de androstano (CAR $\beta$ ), receptor hormonal 96 de *Drosophila* (DHR-96), receptor hormonal nuclear 1 (NHR-1), factor nuclear 4 de hepatocitos (HNF-4), factor nuclear 4G de hepatocitos (HNF-4G), factor nuclear 4B de hepatocitos (HNF-4B), DHNF-4, factor nuclear 4D de hepatocitos (HNF-4D), receptor X del retinoide  $\alpha$  (RXR $\alpha$ ), receptor X del retinoide  $\beta$  (RXR $\beta$ ), proteína de unión II de región II (H-2RIIBP), corregulador 1 de receptor nuclear (RCoR-1), receptor X del retinoide  $\gamma$  (RXR $\gamma$ ), Ultraspiracle (USP), 2C1, factor coriónico 1 (CF-1), receptor testicular (TR-2), receptor testicular (TR2-11), TR4, TAK-1, receptor hormonal de *Drosophila* (DHR78), tailless (TLL), homólogo de tailless (TLX), XTLL, factor transcripcional I de promotor cadena arriba de ovoalbúmina de pollo (COUP-TFI), factor transcripcional A del promotor situado cadena arriba de la ovoalbúmina de pollo (COUP-TFA), EAR-3, SVP-44, factor transcripcional II del promotor situado cadena arriba de la ovoalbúmina de pollo (COUP-TFII), factor transcripcional B del promotor situado cadena arriba de la ovoalbúmina de pollo (COUP-TFB), ARP-1, SVP-40, SVP, factor transcripcional III del promotor situado cadena arriba de la ovoalbúmina de pollo (COUP-TFIII), factor transcripcional G del promotor situado cadena arriba de la ovoalbúmina de pollo (COUP-TFG), SVP-46, EAR-2, receptor  $\alpha$  de estrógeno (ER $\alpha$ ), receptor  $\beta$  de estrógeno (ER $\beta$ ), receptor 1 relacionado con estrógeno (ERR1), receptor  $\alpha$  relacionado con estrógeno (ERR $\alpha$ ), receptor 2 relacionado con estrógeno (ERR2), receptor  $\beta$  relacionado con estrógeno (ERR $\beta$ ), receptor de glucocorticoide (GR), receptor mineralocorticoide (MR), receptor de progesterona (PR), receptor de andrógeno (AR), gen B inducido por el factor de crecimiento nervioso (NGFI-B), receptor nuclear similar a Nur-77 (TRS), N10, receptor huérfano (NUR-77), gen de respuesta temprana humana (NAK-1), factor 1 relacionado con Nurr (NURR-1), gen de respuesta inmediato-temprano humano (NOT), receptor nuclear 1 hepático regenerador (RNR-1), dedo de cinc 3 hematopoyético (HZF-

3), proteína 1 relacionada con Nur (TINOR), receptor huérfano nuclear 1 (NOR-1), receptor relacionado con NOR1 (MINOR), receptor hormonal 38 de *Drosophila* (DHR-38), receptor nuclear 8 de *C. elegans* (CNR-8), C48D5, factor esteroideogénico 1 (SF1), péptido de tipo endozepina (ELP), factor 1 de Fushi-Tarazu (FTZ-F1), proteína de unión 4 adrenal (AD4BP), homólogo de receptor hepático (LRH-1), factor transcripcional de fetoproteína (FTF), factor nuclear de células germinales (GCNFM), receptor asociado al testículo relacionado con el receptor del retinoide (RTR), knirps (KNI), relacionado con knirps (KNRL), gónada embrionaria (EGON), gen de *Drosophila* del receptor nuclear dependiente de ligando (EAGLE), receptor nuclear similar a tritórax (ODR7), Tritórax, gen de la región crítica de la hipoplasia congénita adrenal de reversión del sexo sensible a dosis del cromosoma X (DAX-1), hipoplasia congénita adrenal e hipogonadismo hipogonadotrópico (AHCH) y pareja de heterodímero corto (SHP).

Para los fines de la presente invención, los receptores nucleares, los receptores nucleares de grupo H, EcR, USP y RXR también incluyen receptores nucleares sintéticos y quiméricos, receptores nucleares de grupo H, receptores de ecdisona, EcR, USP, RXR y los homólogos de los mismos.

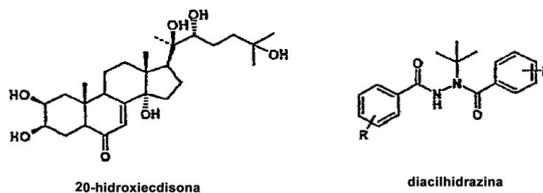
Los genes de interés para la utilización en los casetes de expresión génica de los solicitantes pueden ser genes endógenos o genes heterólogos. La información de secuencia de ácidos nucleicos o de aminoácidos de un gen o proteína deseado puede encontrarse en una de las muchas bases de datos de acceso público, por ejemplo GENBANK, EMBL, Swiss-Prot y PIR, o en muchas publicaciones de revistas relacionadas con la Biología. De esta manera, el experto en la materia dispone de acceso a información de secuencia de ácidos nucleicos para prácticamente la totalidad de los genes conocidos. Esta información puede utilizarse seguidamente para construir los constructos deseados para la inserción del gen de interés dentro de los casetes de expresión génica utilizados en los procedimientos de los solicitantes descritos en la presente memoria.

Entre los ejemplos de genes de interés para la utilización en los casetes de expresión génica de los solicitantes se incluyen, aunque sin limitación, genes codificantes de polipéptidos o productos terapéuticamente deseables que pueden utilizarse para tratar una condición, una enfermedad, un trastorno, una disfunción, un defecto genético, tales como anticuerpos monoclonales, enzimas, proteasas, citoquinas, interferones, insulina, eritropoyetina, factores de coagulación, otros factores o componentes sanguíneos, vectores víricos para la terapia génica, virus para vacunas, dianas para el descubrimiento de fármacos, análisis y aplicaciones de genómica funcional y de proteómica, y similares.

Los procedimientos siguientes se utilizan para preparar los sistemas de regulación de múltiples genes de la presente invención:

los sistemas de regulación de múltiples genes requieren el desarrollo inicial de los ligandos del sistema de regulación, los cuales se utilizan para cribar nuevos dominios de unión a ligando (LBDs). A continuación, se crean dominios de unión a ADN (DBDs) únicos a partir de los cuales se aíslan los elementos de respuesta (REs) a ADN de alta afinidad correspondientes. Finalmente, se crea una colección única de receptores nucleares (NRs) mediante la fusión de los nuevos LBDs y DBDs a dominios de activación transcripcional (ADs) bien caracterizados.

Con el fin de desarrollar un conjunto de parejas de ligando/receptor sin interacciones cruzadas ("totalmente ortogonales"), las estructuras de cabeza de serie de ligando y de receptor presentan una diversidad estructural máxima. Para los receptores basados en ecdisona, dos quimotipos son ideales para la utilización como ligandos: ecdisteroides naturales tales como, por ejemplo, la 20-hidroxiectdisona y las diacilhidrazinas.



Los ecdisteroides naturales son potentes (Kds de tan sólo aproximadamente 1  $\mu$ M) pero aparentemente presentan bastante reactividad cruzada en diferentes especies de insecto, según los datos disponibles de ensayos con insectos completos y basados en células. Las diacilhidrazinas (Kds de tan sólo aproximadamente 0,5 nM), en su mayor parte aparentemente también presentan reactividad cruzada con los EcRs con los que no son activos en absoluto (ver Dhadialla *et al.*, Annu. Rev. Entomol. 43:545-69, 1998). No existe ningún conjunto de ligando/receptor ortogonal en estas dos familias estructurales. Para conseguir el objetivo de un sistema de regulación ortogonal de múltiples genes, la identificación de ligandos requiere tanto la identificación de farmacóforos correspondientes para el receptor especificado como la identificación de farmacóforos no correspondientes para los receptores no interactuantes. Los presentes inventores han descubierto exactamente este tipo de sistema ortogonal.

Los ligandos aceptables son cualesquiera que modulen la expresión del gen en el caso de que la unión del DBD al elemento de respuesta en presencia del ligando resulte en la activación o supresión de la expresión de uno de los genes en el sistema de regulación de múltiples genes y que no activan o suprimen los demás genes del sistema de

regulación múltiple, es decir, el sistema es ortogonal. Entre los ligandos preferentes se incluyen las hormonas naturales ponasterona y muristerona A, los derivados y/o análogos de las mismas, así como las N,N'-diacilhidrazinas, tales como las dadas a conocer en las patentes US nº 6.013.836, nº 5.117.057, nº 5.530.021 y nº 5.378.726; las dibenzoilalquil-cianohidrazinas, tales como las dadas a conocer en la solicitud de patente europea nº 461.809; las N-alquil-N,N'-diaroilhidrazinas, tales como las dadas a conocer en la patente US nº 5.225.443; las N-acil-N-alquilcarbonilhidrazinas, tales como las dadas a conocer en la solicitud de patente europea nº 234.994; las N-aroil-N-alquil-N'-aroilhidrazinas, tales como las descritas en la patente US nº 4.985.461, y otros materiales similares, incluyendo la 3,5-di-terc-butil-4-hidroxi-N-isobutil-benzamida, el 8-O-acetilharpágido, la ecdisona, la 20-hidroxiectidisona, la ponasterona A, la muristerona A, los oxisteroles, el 22(R)-hidroxicolesterol, el 24(S)-hidroxicolesterol, el 25-epoxicolesterol, T0901317, el 5-alfa-6-alfa-epoxicolesterol-3-sulfato (ECHS), el 7-cetocolesterol-3-sulfato, el farnesol, los ácidos biliares, los ésteres de 1,1-bifosfonato, la hormona juvenil III, y similares.

Debido a que un sistema de regulación de múltiples genes requiere ligandos discretos que no reaccionen cruzadamente entre ellos o con otros receptores dentro de la célula, pero que sean específicos e induzcan únicamente un receptor específico, se utilizan varias estrategias para definir los ligandos apropiados para cada combinación de sistema de regulación de múltiples genes.

La complementación de ligando se inicia a partir de un ligando altamente activo conocido y continua de una de las tres maneras siguientes:

1) modificación en etapas de la identidad del elemento farmacóforo individual (es decir, el sitio activo) en el ligando, en el que la hipótesis planteada es de un farmacóforo ligando, se altera drásticamente un elemento dentro del farmacóforo y se realiza una búsqueda de una biblioteca de receptores mutantes para identificar una alteración complementaria. Tras identificar una combinación de mutante/ligando con éxito, se utiliza una secuencia iterativa de diseño de modelo proteico-ligando para optimizar la interacción ligando/receptor, maximizando la respuesta o minimizando la respuesta (en el caso de que resulte deseable suprimir la expresión génica en lugar de inducir la expresión).

2) Adición de un nuevo "dominio variable" del ligando, en el que el farmacóforo y el locus de unión complementario permanecen más o menos constantes. Se une al ligando nuclear un grupo adicional, no esencial pero potencialmente perjudicial para la unión a los receptores naturales. El tamaño y naturaleza de este grupo permite la modificación variegada y la funcionalización. Tal como anteriormente, seguidamente se realiza una búsqueda en la biblioteca de receptores mutantes.

3) La eliminación total de una agrupación de elementos farmacóforos del ligando y la sustitución con un nuevo mapa de PE (análogamente al concepto de estructuras químicas), en el que se conserva aproximadamente la mitad del farmacóforo conocido, y sustitución de la agrupación de farmacóforos faltante con diversas entidades. Estos nuevos fragmentos moleculares proporcionan patrones alternativos de PE o de otro modo replican alternativamente (aunque no totalmente) el patrón original. Las bibliotecas de receptores mutantes, los miembros de las cuales portan modificaciones de residuos en los loci de unión a PE y/o una modificación de la forma de la cavidad, se someten a una búsqueda para complementariedad con el farmacóforo nuevamente perturbado.

Desde el punto de vista del ligando, el procedimiento es el siguiente:

1. Definición de un conjunto de ligandos modificados de modo diverso basándose en cambios graduales de PE, la adición de un nuevo "dominio variable" de ligando y sustitución completa de la agrupación de PE. Los moldes de ligando de partida incluyen diacilhidrazinas y los ecdisteroides naturales.

2. Preparación de un conjunto de receptores en el que los LBDs de los receptores son naturales; se modifican mediante delección, inserción o mutación; sintéticos; quimeras de diferentes dominios de proteínas receptores heterólogos, o una combinación de los mismos. Las modificaciones pueden realizarse mediante mezcla de ADNs, ITCHY o mutagénesis a partir de una pluralidad de receptores naturales. Las mutaciones de LBD deberían sondear regiones del bolsillo de unión e idealmente deberían muestrear residuos de carga +/-, de carácter lipófilico y que puedan actuar como donadores/aceptores de enlaces de H de puntos de unión sospechados.

3. Opcionalmente, introducción de receptores mutantes en células.

4. Búsqueda del conjunto de receptores con un conjunto de ligandos para la modulación y/o unión de genes, en la que tanto la modulación génica como las búsquedas de unión pueden llevarse a cabo *in vivo* (en células en las que han sido introducidos los receptores mutantes) o *in vitro*. Preferentemente, la búsqueda de modulación génica se lleva a cabo *in vivo* en células y la búsqueda de unión se lleva a cabo *in vitro*.

5. Análisis de los datos - tabulado de la magnitud de la inducción/unión como función del receptor y el ligando. Examen de la rejilla para identificar ortogonalidad (combinaciones de receptor/ligando que son mutuamente no productivas como sistemas de regulación génica).

- 5 6. Optimización. Se repiten las etapas 1 a 5 con una modificación más focalizada del ligando y mutaciones del LBD específicas de sitio basadas en los resultados de estructura/actividad de la primera ronda e información de modelado de homología de proteínas.

10 Los ligandos apropiados para estos enfoques deberían: 1) ser fácilmente accesibles a la síntesis, 2) mostrar potencial de una farmacocinética aceptable para fármacos, y 3) prestarse a la modificación estructural. Tanto los ecdisteroides como las diacilhidrazinas cumplen los requisitos, aunque la modificación de esteroides es sintéticamente más exigente y, para que funcione óptimamente, debería eliminar las funcionalidades químicas que se prestan al metabolismo.

15 Tal como se utiliza en la presente memoria, el término "conjunto" se refiere a uno o más. Preferentemente, sin embargo, un "conjunto" o "biblioteca" incluye dos o más miembros. Típicamente, un conjunto incluye muchos más miembros que el número total de sistemas de regulación génica individualmente operables en el sistema de regulación de múltiples genes inducibles.

20 Cada sistema individual que comprende los sistemas de regulación de múltiples genes dentro de una célula requiere un receptor apropiado. Para los fines de la presente invención, el término "célula" incluye los virus. Aunque muchos receptores resultan aplicables al sistema de la presente descripción, resultan preferentes los receptores nucleares. Los sistemas de regulación génica basados en EcR resultan ideales para la utilización en la presente invención debido a que cada sistema de regulación ofrece una regulación muy estrecha de la expresión génica y existe  
25 suficiente variabilidad en la familia de EcR para permitir la generación de múltiples nuevos receptores EcR. Son conocidas de la técnica varias técnicas para introducir mutaciones en una secuencia de ADN, incluyendo la mutagénesis sitio-dirigida, la PCR propensa a errores, la utilización de la cepa mutadora Red AL-1, la mezcla de ADNs (ver Chang C.C. *et al.*, Nat. Biotechnol. 17(8):793-7, 1999, y Stemmer W.P., Proc. Natl. Acad. Sci. USA 91(22):10747-51, 1994) y el truncado gradual para la creación de enzimas híbridos, también conocido como ITCHY (ver Michnick S.W. y Arnold F.H., Nat. Biotechnol. 17(12):1159-60, 1999, y Ostermeier M. *et al.*, Nat. Biotechnol. 17(12):1205-9, 1999).

35 El LBD de EcR de *Choristeneuria fumiferana* (CfEcR) es un candidato ideal para la mutagénesis debido a que muestra una elevada afinidad de unión a ligando con determinadas diacilhidrazinas. Sin embargo, otros receptores de ecdisona, así como otros receptores nucleares pueden mutarse para la utilización dentro del sistema de la presente invención. La mutagénesis aleatoria mediante mutagénesis sitio-dirigida, PCR propensa a errores y mediante la utilización de la cepa mutadora *Epicurian coli* XL1-Red, produce varias mutaciones aleatorias en el LBD del CfEcR. La cepa mutadora XL-1 Red se manipula para que sea deficiente en tres genes implicados en la reparación del ADN: mutS, mutD y mutT. La transformación de un plásmido en la cepa resulta en la generación de  
40 mutaciones integradas aleatoriamente en toda la secuencia debido a que la cepa presenta una tasa 5.000 veces más alta de mutación. Los polinucleótidos resultantes codificantes de los LBDs mutados seguidamente se clonan en vectores apropiados y se crean bibliotecas. Las bibliotecas seguidamente se criban utilizando los procedimientos siguientes. Los receptores de ecdisona que comprenden mutaciones por truncado y las mutaciones por sustitución que afectan a la unión, especificidad y/o sensibilidad de ligandos se han obtenido recientemente (ver la solicitud de patente internacional copendiente nº PCT/US01/19050 y la solicitud provisional de patente US titulada "Novel Substitution Mutant Receptors and Their Use in a Nuclear Receptor-Based Inducible Gene Expression System", de Palli *et al.*, nº de serie todavía no asignado, presentada el 21 de agosto de 2001).

45 Las mutaciones por sustitución pueden realizarse mediante cualquier técnica de mutagénesis conocida de la técnica, incluyendo, aunque sin limitación, la mutagénesis sitio-dirigida *in vitro* (Hutchinson C. *et al.*, J. Biol. Chem. 253:6551, 1978; Zoller y Smith, DNA 3:479-488, 1984; Oliphant *et al.*, Gene 44:177, 1986; Hutchinson *et al.*, Proc. Natl. Acad. Sci. U.S.A. 83:710, 1986), la utilización de conectores TAB<sup>®</sup> (Pharmacia), la digestión con endonucleasa de restricción/delección de fragmento y sustitución, la mutagénesis mediada por PCR/dirigida por oligonucleótidos, y similares. Resultan preferentes las técnicas basadas en la PCR para la mutagénesis sitio-dirigida (ver Higuchi, "Using PCR to Engineer DNA", en: PCR Technology: Principles and Applications for DNA Amplification, H. Erlich, editor, Stockton Press, capítulo 6, páginas 61 a 70, 1989).

50 La mezcla de ADNs es una técnica que aprovecha la homología de secuencias que comparte una familia de genes. En este procedimiento, se clona la familia de genes en un vector común y se amplifica por PCR utilizando un conjunto de cebadores específico para las secuencias del vector que flanquean la inserción. De esta manera, todos los genes amplificados comparten las mismas secuencias 5' y 3'. Los productos génicos amplificados por PCR seguidamente se tratan con ADNasaI y los fragmentos resultantes en el intervalo de tamaños de 10 a 60 pares de bases ("bp") se purifican en gel. Estos fragmentos se utilizan seguidamente en una reacción de PCR que no contiene oligonucleótidos a modo de cebadores. De este modo, el cebado para la extensión con la ADN polimerasa Taq procede de la hibridación de regiones homólogas en la familia de genes. Tras esta ronda de mezcla del ADN, puede repetirse sucesivamente el procedimiento para generar una biblioteca saturada de secuencias reorganizadas.

Al completarse la reorganización, se lleva a cabo una amplificación por PCR con los cebadores que se utilizaron para amplificar las secuencias clonadas del vector. De esta manera, se amplifica una biblioteca de genes reorganizados, cuyos miembros típicamente presentan el mismo tamaño que sus genes parentales. A continuación, se clona la biblioteca y se somete a ensayo para el fenotipo deseado (ver Stemmer, 1994, anteriormente).  
 5 Nuevamente, los receptores nucleares y EcR en particular son una buena elección para este procedimiento, en el aspecto de que la superfamilia de receptores nucleares, de la que EcR es un miembro, contiene más de 300 miembros reconocidos procedentes de vertebrados, artrópodos y nemátodos.

Como alternativa al desarrollo de sistemas de regulación de múltiples genes basados en receptores nucleares de grupo H simplemente mediante la utilización de la serie de receptores nucleares de grupo H de tipo salvaje disponibles actualmente de diferentes especies y la manipulación de un único ligando para cada uno de ellos, también pueden manipularse nuevos LBDs de receptores nucleares de grupo H que todavía respetan la arquitectura básica de receptor/receptor nuclear/receptor nuclear de grupo H, preferentemente EcR (ver la solicitud copendiente provisional de patente US, número de serie todavía no asignado, titulada "Novel Substitution Mutant Receptors And Their Use In A Nuclear Receptor-Based Inducible Gene Expression System", inventores: Palli *et al.*, presentada el 21 de agosto de 2001, que describe dominios de unión a ligando de receptor nuclear de grupo H que comprenden mutaciones por sustitución que modifican la sensibilidad, especificidad o magnitud de transactivación de ligando del dominio de unión a ligando. Con el fin de llevar a cabo lo anterior, se reorganizaron secuencias de ADN procedentes de una multitud de las secuencias más divergentes. La ventaja reside en el hecho de que existe una homología de secuencia de ácidos nucleicos significativa entre muchos receptores conocidos y clonados. Lo anterior implica una mayor probabilidad de generar una biblioteca de receptores que contenga nuevos LBDs.

Un aspecto importante de la presente invención es que, mediante la utilización de receptores nucleares humanos en el presente procedimiento, pueden desarrollarse sistemas de regulación de múltiples genes con aplicaciones de terapia génica en el ser humano. Además, la reorganización del ADN puede utilizar secuencias de receptor nuclear humano para generar nuevos DBDs. Al igual que la reorganización de EcR para obtener nuevos LBDs, existen ventajas y desventajas a la mezcla de diferentes receptores nucleares humanos para obtener nuevos DBDs. Las ventajas son dos. En primer lugar, la utilización de secuencias humanas reduce en gran medida la probabilidad de antigenicidad de los nuevos receptores en las aplicaciones de terapia génica en el ser humano. En segundo lugar, existe una mayor probabilidad de generar nuevos DBDs mediante la mezcla de diferentes receptores hormonales nucleares humanos procedentes de la superfamilia que mediante la mezcla de EcRs procedentes de diferentes especies, dada la mayor diversidad de secuencias de dominio C. Sin embargo, con el fin de que la mezcla de ADN presente éxito, debe existir una cantidad mínima de homología.

Los receptores nucleares se unen al ADN en forma de heterodímeros con RXR, en forma de homodímeros o en forma de monómeros. El dominio de dimerización de la superfamilia de receptores nucleares es bipartita: se encuentra dividida entre los dominios C y E del receptor. La presente invención únicamente utiliza el dominio C de los receptores nucleares humanos mezclados. Por lo tanto, resulta lógico seleccionar la mezcla de aquellos receptores humanos que se unen a ADN en forma de monómero, tales como SF-1, NGF1-B, ERR, ROR, TLL y Rev-Erb. Al optar por mezclar únicamente receptores de este tipo, el sistema de regulación génica de la presente invención en primer lugar permite que el nuevo receptor EcR basado en LBD funcione sin RXR. De este modo, el receptor EcR basado en LBD no utiliza RXR endógeno, apartándolo de esta manera de receptores endógenos. En segundo lugar, la cantidad de secuencia codificante que resulta necesario dirigir a un genoma huésped se reduce esencialmente a la mitad al evitar la necesidad de administrar RXR. Dadas las restricciones de tamaño de los actualmente conocidos vehículos de direccionamiento génico, cualquier reducción de la cantidad de ADN que deba administrarse en un huésped en una aplicación de terapia génica resultaría muy beneficiosa. Para las aplicaciones de terapia génica en el ser humano, se fusiona un nuevo DBD con un nuevo LBD basado en EcR. Para completar el receptor, se añade un dominio de activación procedente de un factor transcripcional humano.

Al contrario que en la mezcla de ADNs, ITCHY no se basa en la homología de secuencias para la generación de genes quiméricos. ITCHY genera una biblioteca de truncados N-terminales y C-terminales de dos genes mediante la digestión gradual de sus extremos con exonucleasa III (ExoIII). ExoIII cataliza las deleciones de mononucleótidos de extremos romos ó 5' protuberantes. Por lo tanto, un extremo 3' protuberante protegerá un extremo de una secuencia frente a la digestión. Esta propiedad de ExoIII puede aprovecharse para obtener deleciones direccionales de los genes de interés. Por ejemplo, en el caso de que se llevase a cabo una ITCHY en los genes de los receptores de estrógeno ("ER") y de progesterona ("PR"), la biblioteca de mutantes deseada no presentaría un dominio A/B de ER ligada a un dominio A/B de PR. El producto deseado presentaría la región 5' (amino-terminal) del ADNc de ER ligada a la región 3' (carboxi-terminal) del ADNc de PR o la región 5' del ADNc de PR ligada a la región 3' del ADNc de ER. La deleción direccional de las secuencias parentales evita que se produzcan ligaciones no deseadas. Durante la reacción de ExoIII, se extraen alícuotas pequeñas a intervalos cortos y se inactiva el enzima de manera que al terminar el procedimiento completo, se obtiene una biblioteca completa de deleciones de 1 par de bases a lo largo de la longitud de los ADNc. Esta biblioteca de fragmentos seguidamente se liga entre sí creando una nueva biblioteca de genes quiméricos. Los productos de ITCHY pueden mostrar un gran abanico de variación de tamaños. Para los fines de la presente invención, las recombinaciones que tienen lugar en los dominios DEF de las secuencias de EcR y el dominio C de los receptores nucleares humanos monoméricos mencionados anteriormente resultan importantes para las aplicaciones de terapia génica en el ser humano. Entre los dominios de unión a ADN

humano que pueden resultar útiles en este procedimiento se incluyen, por ejemplo, el factor transcripcional mitocondrial A ("mtTFA"). También puede resultar útil anclar dos DBDs que normalmente no heterodimerizan con secuencias cortas de aminoácidos con el fin de producir DBDs únicos.

5 En comparación con la reorganización del ADN, anteriormente, a pesar del hecho de que ITCHY genera únicamente un entrecruzamiento o recombinación entre dos genes, resulta en la generación de todos los posibles entrecruzamientos debido a que la técnica no se basa en la homología de secuencias. Lo anterior resulta en un espacio de secuencias más diverso en el que cribar para encontrar funciones mejoradas en comparación con la reorganización del ADN. Dado que existen regiones diferentes dentro de los dominios C y E que comparten homología de secuencia, la reorganización del ADN únicamente podrá generar un número limitado de regiones reorganizadas. Sin embargo, ITCHY no se basa en la homología de secuencias y puede resultar en todas las posibilidades de recombinación. Lo anterior es un hecho muy significativo a considerar en la construcción de nuevos LBDs y DBDs. Además, existe la posibilidad de llevar a cabo en primer lugar ITCHY en varias parejas génicas diferentes y después utilizar esa biblioteca, o un subconjunto de esa biblioteca, en la reorganización del ADN. De esta manera se obtendrán todas las posibles recombinaciones y se incrementa la probabilidad de identificar nuevos dominios de unión a ligando y a ADN.

Tras desarrollar bibliotecas de ligandos potenciales, LBDs y DBDs de receptor, deben evaluarse para sus interacciones funcionales. Tras ensamblar los componentes receptor y ligando para el sistema de regulación de múltiples genes, resulta necesario someter a ensayo y validar el sistema en líneas celulares y ensayos apropiados.

Se utiliza el procedimiento siguiente para dicha evaluación. En primer lugar, se introducen sitios de restricción únicos en las secuencias de ADN de partida en ambos extremos de los LBDs y en ambos extremos de los DBDs antes de la mutagénesis/reorganización/ITCHY. Los LBDs y DBDs seguidamente se extraen de la biblioteca de ADNs, resultando de la mutagénesis/reorganización/ITCHY. Las bibliotecas de LBD y DBD resultantes seguidamente se clonan en vectores apropiados y se evalúan mediante la identificación de combinaciones funcionales de los ligandos, LBDs, DBDs y REs utilizando uno o más de los procedimientos siguientes:

a. Cribado de LBDs utilizando un LBD acoplado traduccionalmente a un gen de resistencia a antibiótico mediante la fusión de un gen de resistencia a antibiótico, tal como canamicina, a los LBDs mutados/reorganizados/ITCHY. El ADN codificante de los LBDs de los receptores en las bibliotecas se extrae utilizando los sitios de restricción únicos introducidos entre los dominios C y D y en el extremo 3' de los transcritos de la biblioteca. Estos LBDs se insertarán en un casete de expresión contenido dentro de un plásmido y se cribará una biblioteca de plásmidos para proteínas traducidas de longitud completa mediante transformación de la biblioteca en *E. coli* y la selección para resistencia a la canamicina. Se aísla el ADN plasmídico de todas las colonias resistentes y se utiliza para el cribado de células de mamífero.

b. Cribado de LBDs utilizando un LBD de célula de mamífero y un sistema de un híbrido, ambos para un ligando complementario y para la capacidad del receptor unido a ligando de activar la transcripción en forma de homodímero o de monómero. Este procedimiento utiliza una línea celular establemente transfectada con un constructo informador que consiste de un RE Gal4 multimerizado y un promotor mínimo que controla la expresión de GFP. A partir de la biblioteca de plásmidos aislada a partir de las colonias resistentes a la canamicina en la etapa a., anteriormente, la región que incluye el dominio de activación VP16, DBD Gal4 y el LBD nuevo se subclona en un vector de expresión retroviral. Este retrovirus también codifica un gen de resistencia a antibiótico seleccionable. Se prepara una biblioteca de retrovirus que contiene el ADN codificante de estas proteínas de fusión y se utiliza para infectar la línea celular indicada anteriormente que contiene una GFP informadora bajo el control de un RE Gal4 concatamérico. Las células con ADN retroviral establemente integrado se seleccionan a partir de la resistencia a antibiótico. Además, estas células pueden separarse mediante FACS para la expresión de GFP. Las células que expresan GFP en ausencia de cualquier ligando exógeno se descartan. La población celular restante se expande y se divide en grupos en placas multipocillo. En un ensayo de alto rendimiento, se incuba cada grupo de células con un ligando diferente procedente de la biblioteca de ligandos. Se seleccionan aquellos ligandos que activan GFP al nivel máximo y se utilizan para cribar la biblioteca de LBDs en una segunda ocasión. Tal como se ha indicado anteriormente, se divide la biblioteca de receptores mutantes en grupos y se incuba cada grupo con un ligando diferente. Sin embargo, debido a que ahora hay un número limitado de ligandos, se separa cada grupo de células mediante FACS. Las células con los niveles más altos de expresión de GFP en respuesta al ligando se recogen, se siembran en placas a densidad baja y se seleccionan los clones individuales. El ADN codificante del LBD en cada uno de estos clones se aísla mediante amplificación por PCR y se secuencia. Estos transcritos codifican un LBD capaz de mediar en la activación transcripcional en respuesta a la exposición al ligando con el que se cribó. A continuación, estos ligandos se modifican para optimizar la afinidad y especificada para su LBD complementario. Los ligandos modificados se evalúan al igual que se evaluó la biblioteca de ligandos de partida indicada anteriormente. En el caso de que se sustituya GFP por un gen de resistencia a antibiótico en el plásmido informador, el cribado de células de mamífero también puede realizarse basándose en la selección con antibióticos. En este caso, las células infectadas retroviralmente se cultivan en presencia de antibiótico y se aíslan las células supervivientes. Tal como se ha indicado anteriormente, el ADN codificante del LBD en cada uno de los clones celulares se aísla y se secuencia.

65

c. Cribado de DBD utilizando un DBD acoplado traduccionalmente a un gen de resistencia a antibiótico siguiendo una estrategia similar a la descrita anteriormente para el cribado de LBD con el fin de seleccionar contra las proteínas traducidas incompletamente. El ADN codificante de los DBDs de los receptores en las bibliotecas se extrae utilizando los sitios de restricción únicos introducidos entre los dominios A/B y C y los dominios C y D. A continuación, estos DBDs se insertan en el sitio de clonación múltiple de los diversos casetes de expresión contenidos dentro de plásmidos.

d. Cribado de DBD para REs afines utilizando un ensayo de un híbrido de levadura en el que la biblioteca de DBDs seleccionada anteriormente se criba con una biblioteca de REs. La cepa de levadura utilizada en este ensayo no debería ser sensible al ganciclovir en ausencia de timidina quinasa y también debería presentar los requerimientos nutricionales siguientes: leucina (leu), histidina (his) y uracilo (ura) (es decir, la cepa es deficiente en LEU, HIS y URA). La biblioteca de REs es una agrupación parcialmente degenerado de oligonucleótidos basada en el RE de consenso de los receptores nucleares monoméricos utilizados en los procedimientos de mutagénesis/reorganización/ITCHY. Se sintetiza esta biblioteca con sitios de restricción en cada extremo de los oligonucleótidos. Los oligonucleótidos digeridos con enzimas de restricción se clonan en un vector de expresión levadura en el extremo 3' de un informador que consiste de una proteína de fusión LEU2/timidina quinasa. Esta proteína de fusión permite la selección tanto positiva como negativa de activación transcripcional. El vector informador también contiene un marcador seleccionable URA expresado constitutivamente para la selección positiva de transformantes estables en medio ura<sup>-</sup>. Se utiliza la expresión de timidina quinasa para seleccionar negativamente para la activación del RE por parte de factores endógenos de levadura mediante la adición del análogo nucleósido ganciclovir al medio de cultivo. Tras la transformación, se cultivan las células en medio ura-ganciclovir<sup>+</sup>. Las células que contienen REs capaces de unirse a factores de transcripción de levadura activados no sobreviven. Las células transformadas supervivientes se agrupan. El casete de DBD de receptor nuclear de VP16 de los plásmidos de expresión bacterianos aislados anteriormente se extrae y se clona bajo el control de un promotor constitutivo en un vector de expresión levadura. Este vector contiene un marcador seleccionable HIS para la transformación. La levadura transformada y seleccionada para la presencia del informador RE-LEU2/timidina quinasa seguidamente se transforma una segunda vez con esta biblioteca de proteínas de fusión de DBD de receptor nuclear de VP16. Se cultiva la levadura en medio ura<sup>-</sup>his<sup>-</sup>leu<sup>-</sup> para seleccionar para la presencia del plásmido informador, para el plásmido que contiene la biblioteca de DBD y para la unión de la proteína DBD de VP16 al RE en el informador, respectivamente. Las células que expresan una proteína VP16-DBD que se une al RE del informador también expresan timidina quinasa activa. Sin embargo, no se añade ganciclovir al medio en este caso, de manera que no se selecciona contra estas células. Por el contrario, las células se encuentran bajo una selección positiva del crecimiento debido a la expresión del marcador LEU2.

e. Cribado de DBDs de factor de transcripción anclados utilizando un RE de ADN al que se une la proteína. Los factores transcripcionales anclados se expresan en *E. coli* en forma de proteínas de fusión con una etiqueta epítipo. La etiqueta epítipo se utiliza para construir una columna de afinidad de los DBDs anclados. Esta columna se utiliza para seleccionar para un RE de ADN. Los REs candidatos que deben cribarse se encuentran contenidos dentro de una biblioteca sintetizada de oligonucleótidos. Los oligonucleótidos en esta biblioteca contienen los REs para cada uno de los DBDs que comprenden la proteína anclada con un espaciador de secuencia aleatoria de longitud variable entre los REs. Ambos extremos de esta biblioteca de oligonucleótidos presentan una secuencia definida. Se sintetiza una biblioteca degenerada diferente para cada proteína quimérica. Se utiliza un procedimiento de cribado basado en la PCR para aislar las secuencias de ADN con afinidad de unión más alta para los DBDs anclados. La columna de afinidad de proteína anclada-DBD se incuba con la biblioteca de REs, se eliminan mediante lavado los oligonucleótidos no unidos y se eluyen los oligonucleótidos unidos. Los oligonucleótidos eluidos son REs candidatos y se amplifican mediante PCR utilizando cebadores que se hibridan a los extremos definidos de los oligonucleótidos. Los productos de PCR resultantes se aplican a la columna de afinidad. Este procedimiento de selección de oligonucleótidos y la PCR se repiten varias veces, preferentemente 10 a 12 veces. Tras el cribado repetido, la mezcla final de oligonucleótidos se clona en un vector y el grupo resultante de vectores se transforma en *E. coli*. La secuencia de oligonucleótidos en clones individuales se determina mediante secuenciación de ADN. Las secuencias de múltiples clones se alinean para determinar un sitio de unión de consenso para cada uno de los DBDs anclados. Se sintetiza un oligonucleótido basado en este sitio de unión de consenso y se utiliza con el DBD anclado correspondiente en el ensayo de levadura descrito anteriormente. De esta manera, se determinan las afinidades relativas entre las combinaciones RE/DBD aisladas utilizando los productos mutagenizados/reorganizados/ITCHY y se determinan los DBDs anclados.

Se aparea cada pareja de las variantes de LBD identificadas en los cribados indicados anteriormente, con un DBD único identificado anteriormente. El ADN codificante de cada una de estas nuevas variantes quiméricas LBD/DBD se sitúa bajo el control de un promotor constitutivo en un vector de expresión de mamífero. El elemento de respuesta a ADN para cada uno de los DBDs de las proteínas quiméricas LBD/DBD se inserta cadena arriba de un gen informador en este mismo plásmido. Cada plásmido incluye un marcador de resistencia a antibiótico bajo el control de un promotor constitutivo. Las variantes de receptor quimérico y los REs en cada una de estas líneas celulares estables individuales se caracteriza con respecto al nivel de expresión de informador en ausencia de ligando, así como el factor de inducción de actividad de informador en presencia de ligando.

El grupo de plásmidos codificantes de receptores quiméricos funcionales y sus REs correspondientes seguidamente se transfectan establemente de modo secuencial en células de mamífero. Tras la transfección de cada plásmido, se someten a ensayo las células para su respuesta al ligando o ligandos para el receptor o receptores quiméricos que contienen las células. La selección de las células con integración estable de estos plásmidos requiere la utilización de genes codificantes de proteínas que conviertan a las células en resistentes a múltiples antibióticos (higromicina, neomicina, puromicina, bleomicina y blasticidina). El ensayo para la expresión de genes inducibles requiere múltiples proteínas informadoras tales como, por ejemplo, la luciferasa de luciérnaga, la luciferasa de *Renilla*,  $\beta$ -galactosidasa, fosfatasa alcalina, cloranfenicol acetiltransferasa y hormonas de crecimiento. Estos materiales se encuentran fácilmente disponibles.

Tras integrar establemente en el genoma el ADN codificante de todos los receptores y sus elementos de respuesta correspondiente, la línea celular resultante se utiliza para evaluar la reactividad cruzada (ortogonalidad) entre diferentes ligandos y diferentes receptores quiméricos en el contexto de la misma célula. Las células se incuban con ligandos individuales y se somete a ensayo la actividad de todas las proteínas informadoras. Las células también se incuban con diferentes combinaciones de los ligandos con ensayo posterior de todas las proteínas informadoras.

Preferentemente, la interacción ligando/LBD para cada una de las parejas de ligando/LBD presenta Kds de entre 0,1 y 1.000 nM. Más preferentemente, la interacción ligando/LBD para cada una de las parejas de ligando/LBD presenta Kds de entre 0,1 y 100 nM. Todavía más preferentemente, la interacción de ligando/LBD para cada una de las parejas de ligando/LBD presenta Kds de entre 0,1 y 10 nM. Más preferentemente, la interacción de ligando/LBD para cada una de las parejas de ligando/LBD presenta Kds de entre 0,1 y 1,0 nM.

Los sistemas de regulación de múltiples genes de la presente invención resultan útiles no sólo en el área de la modulación génica misma, sino también en otras áreas importantes, tales como, por ejemplo, la proteómica, la genómica funcional, las terapias génicas, los ensayos de alto rendimiento basados en células, los biosensores, el cribado toxicológico y la producción de proteínas a gran escala.

Específicamente, la genómica funcional y la proteómica en la actualidad se encuentran limitadas por su incapacidad para tratar fenotipos multifactoriales en los que se encuentran implicados múltiples genes. Existen numerosos ejemplos. Entre algunos de los más llamativos se incluyen, por ejemplo, las cascadas de transducción de señales, tales como la ruta wnt/catenina, en la que se encuentran implicadas más de 20 proteínas. Esta ruta ha sido implicada en el cáncer, en enfermedades neurodegenerativas, en disfunciones del sistema inmunológico y en otros. La disección de las interacciones entre los miembros de la ruta en estudios de genómica funcional y de proteómica podría verse muy facilitada por la aparición de un sistema de control de múltiples genes. A continuación, el investigador podría regular múltiples factores simultáneamente y determinar interacciones clave de manera más exacta que con los enfoques del estado actual de la técnica que se basan en estrategias de inactivación génica, activación génica o mutagénesis.

De manera similar, determinadas terapias génicas requieren la regulación de más de un gen. Por ejemplo, la generación de respuestas inmunológicas frente a un cáncer a través de la introducción de cócteles de genes de citoquina y de antígenos requeriría un sistema de regulación múltiple. Esto permite que la terapia funcione con seguridad y de un modo integrado cuantitativo.

Los sistemas de regulación de múltiples genes de la presente invención presentan muchas ventajas sobre otros sistemas de inducción génica utilizados en la actualidad, dependiendo de la aplicación. En la proteómica y genómica funcional, cambia significativamente el modo en el que se analizan los fenotipos celulares y la función génica. En lugar de un gen/proteína cada vez, permite el análisis de rutas moleculares enteras en células, lo que se encuentra mucho más próximo a lo que realmente sucede en un organismo real. En la producción de proteínas o cribado de alto rendimiento, la tecnología es una nueva infraestructura para la producción de proteínas en paralelo o para el cribado frente a múltiples dianas simultáneamente.

### **Células huésped y organismos no humanos**

Otro aspecto de la presente invención se refiere a una célula huésped aislada que comprende un sistema de regulación de múltiples genes según la invención. Tal como se ha indicado anteriormente, los sistemas de regulación de múltiples genes de la presente invención pueden utilizarse para modular la expresión génica en una célula huésped. La expresión en células huésped transgénicas puede resultar útil para la expresión de diversos genes de interés. La invención de los solicitantes permite la modulación de la expresión génica en células huésped procarióticas y eucarióticas. La expresión en células huésped transgénicas resulta útil para la expresión de diversos polipéptidos de interés, incluyendo, aunque sin limitación, antígenos producidos en plantas a modo de vacunas, enzimas como la alfa-amilasa, la fitasa, la glucanasa y la xilanas, genes para la resistencia frente a insectos, nemátodos, hongos, bacterias, virus y perturbaciones abióticas, antígenos, nutracéuticos, farmacéuticos, vitaminas, genes para modificar el contenido de aminoácidos, resistencia a herbicidas, frío, sequía y tolerancia al calor, productos industriales, aceites, proteínas, carbohidratos, antioxidantes, plantas estériles masculinas, flores, combustibles, otros caracteres exteriores, polipéptidos terapéuticos, intermediarios de rutas; para la modulación de rutas preexistentes en el huésped, para la síntesis de nuevos productos hasta ahora no posibles utilizando el

huésped; ensayos basados en células; ensayos de genómica funcional, producción de proteínas bioterapéuticas, ensayos de proteómica, y similares. Además, los productos génicos podrían resultar útiles para proporcionar rendimientos de crecimiento más altos del huésped o para permitir la utilización de un modo de crecimiento alternativo.

5 En una forma de realización específica, la célula huésped aislada es una célula huésped procariótica o una célula huésped eucariótica. En otra forma de realización específica, la células huésped aislada es una célula huésped de invertebrado o una célula huésped de vertebrado. Preferentemente, la células huésped aislada se selecciona de entre el grupo constituido por una célula bacteriana, una célula fúngica, una célula de levadura, una célula de nemátodo, una célula de insecto, una célula de pez, una célula vegetal, una célula de ave, una célula animal y una célula de mamífero. Más preferentemente, la célula huésped aislada es una célula de levadura, una célula de nemátodo, una célula de insecto, una célula vegetal, una célula de pez cebra, una célula de pollo, una célula de hámster, una célula de ratón, una célula de rata, una célula de conejo, una célula de gato, una célula de perro, una célula bovina, una célula de cabra, una célula de vaca, una célula de cerdo, una célula de caballo, una célula de oveja, una célula de simio, una célula de mono, una célula de chimpancé o una célula humana.

Entre los ejemplos de células huésped preferentes se incluyen, aunque sin limitación, especies fúngicas o de levadura tales como *Aspergillus*, *Trichoderma*, *Saccharomyces*, *Pichia*, *Candida*, *Hansenula*, o especies bacterianas tales como aquéllas en los géneros *Synechocystis*, *Synechococcus*, *Salmonella*, *Bacillus*, *Acinetobacter*, *Rhodococcus*, *Streptomyces*, *Escherichia*, *Pseudomonas*, *Methylomonas*, *Methylobacter*, *Alcaligenes*, *Synechocystis*, *Anabaena*, *Thiobacillus*, *Methanobacterium* y *Klebsiella*; especies vegetales seleccionadas de entre el grupo constituido por manzana, *Arabidopsis*, bajra, plátano, cebada, alubias, remolacha, soja verde, garbanzo, chili, pepino, berenjena, habas verdes, maíz, melón, mijo, alubia, avena, okra, *Panicum*, papaya, cacahuete, guisante, pimiento, guisante de paloma, piña, *Phaseolus*, patata, calabaza, arroz, sorgo, soja, calabacín, caña de azúcar, remolacha azucarera, girasol, batata, té, tomate, tabaco, sandía y trigo; células huésped animales y de mamífero.

En una forma de realización específica, la célula huésped aislada es una célula de levadura seleccionada de entre el grupo constituido por una célula huésped de *Saccharomyces*, de *Pichia* y de *Candida*.

En otra forma de realización específica, la célula huésped aislada es una célula de nemátodo *Caenorhabdus elegans*.

En otra forma de realización específica, la célula huésped aislada es una célula vegetal seleccionada de entre el grupo constituido por manzana, *Arabidopsis*, bajra, plátano, cebada, alubias, remolacha, soja verde, garbanzo, chili, pepino, berenjena, habas verdes, maíz, melón, mijo, alubia, avena, okra, *Panicum*, papaya, cacahuete, guisante, pimiento, guisante de paloma, piña, *Phaseolus*, patata, calabaza, arroz, sorgo, soja, calabacín, caña de azúcar, remolacha azucarera, girasol, batata, té, tomate, tabaco, sandía y célula de trigo.

En otra forma de realización específica, la célula huésped aislada es una célula de pez cebra.

En otra forma de realización específica, la célula huésped aislada es una célula de pollo.

En otra forma de realización específica, la célula huésped aislada es una célula de mamífero seleccionada de entre el grupo constituido por una célula de hámster, una célula de ratón, una célula de rata, una célula de conejo, una célula de gato, una célula de perro, una célula bovina, una célula de cabra, una célula de vaca, una célula de cerdo, una célula de caballo, una célula de oveja, una célula de mono, una célula de chimpancé y una célula humana.

La transformación de células huésped es bien conocida de la técnica y puede llevarse a cabo mediante una diversidad de procedimientos, incluyendo, aunque sin limitación, electroporación, infección vírica, transfección de plásmido/vector, transfección mediada por un vector no vírico, transformación mediada por *Agrobacterium*, bombardeo de partículas, y similares. La expresión de los productos génicos deseados implica cultivar las células huésped transformadas bajo condiciones adecuadas e inducir la expresión del gen transformado. Las condiciones de cultivo y los protocolos de expresión génica en células procarióticas y eucarióticas son bien conocidas de la técnica (ver la sección de procedimientos generales de los Ejemplos). Las células pueden recolectarse y los productos génicos, aislarse, según protocolos específicos para el producto génico.

Además, puede seleccionarse una célula huésped que modula la expresión de los polinucleótidos transfectados, o modifica y procesa los productos polipeptídicos de un modo específico deseado. Las diferentes células huésped presentan mecanismos característicos y específicos para el procesamiento y modificación traduccionales y post-traduccionales [por ejemplo, glucosilación, corte (por ejemplo de secuencia de señal)] de las proteínas. Las líneas celulares o sistemas huésped apropiados pueden seleccionarse para que garanticen la modificación y procesamiento deseados de la proteína foránea expresada. Por ejemplo, puede utilizarse la expresión en un sistema bacteriano para producir un producto proteico central no glucosilado. Sin embargo, un polipéptido expresado en bacterias puede no plegarse correctamente. La expresión en levaduras puede producir un producto glucosilado. La expresión en células eucarióticas puede incrementar la probabilidad de una glucosilación y plegamiento "nativos" de

una proteína heteróloga. Además, la expresión en células de mamífero puede proporcionar una herramienta para reconstituir, o constituir, la actividad del polipéptido. Además, diferentes sistemas de expresión de vector/huésped pueden afectar a las reacciones de procesamiento, tales como los cortes proteolíticos, en un grado diferente.

5 La invención de los solicitantes también se refiere a un organismo no humano que comprende una célula huésped aislada según la invención. En una forma de realización específica, el organismo no humano es un organismo procaríótico o un organismo eucariótico. En otra forma de realización específica, el organismo no humano es un organismo invertebrado o un organismo vertebrado.

10 Preferentemente, el organismo no humano se selecciona de entre el grupo constituido por una bacteria, un hongo, una levadura, un nematodo, un insecto, un pez, un planta, un ave, un animal y un mamífero. Más preferentemente, el organismo no humano es una levadura, un nemátodo, un insecto, una planta, un pez cebrá, un pollo, un hámster, un ratón, una rata, un conejo, un gato, un perro, un bovino, una cabra, una vaca, un cerdo, un caballo, una oveja, un simio, un mono o un chimpancé.

15 En una forma de realización específica, el organismo no humano es una levadura seleccionada de entre el grupo constituido por *Saccharomyces*, *Pichia* y *Candida*.

20 En otra forma de realización específica, el organismo no humano es un nemátodo *Caenorhabdus elegans*.

En otra forma de realización específica, el organismo no humano es una planta seleccionada de entre el grupo constituido por manzana, *Arabidopsis*, bajra, plátano, cebada, alubias, remolacha, soja verde, garbanzo, chili, pepino, berenjena, habas verdes, maíz, melón, mijo, alubia, avena, okra, *Panicum*, papaya, cacahuete, guisante, pimienta, guisante de paloma, piña, *Phaseolus*, patata, calabaza, arroz, sorgo, soja, calabacín, caña de azúcar, remolacha azucarera, girasol, batata, té, tomate, tabaco, sandía y trigo.

25 En otra forma de realización específica, el organismo no humano es un ratón (*Mus musculus*).

### Medición de la expresión/transcripción génicas

30 Una medida útil de los sistemas y procedimientos de regulación de múltiples genes de la presente invención es la del estado transcripcional de la célula, incluyendo las identidades y abundancias de ARN, preferentemente de especies de ARNm. Dichas medidas convenientemente se llevan a cabo mediante la medición de las abundancias de ADNc mediante cualquiera de entre varias técnicas existentes de expresión génica.

35 La tecnología de matrices de ácidos nucleicos es una técnica útil para determinar la expresión diferencial de ARNm. Dicha tecnología incluye, por ejemplo, chips de oligonucleótidos y micromatrices de ADN. Estas técnicas se basan en fragmentos de ADN u oligonucleótidos que corresponden a diferentes genes o ADNc que se inmovilizan sobre un soporte sólido y se hibridan con sondas preparadas a partir de grupos de ARNm total extraídos de células, tejidos u organismos completos y se convierten en ADNc. Los chips de oligonucleótidos son matrices de oligonucleótidos sintetizados en un sustrato utilizando técnicas fotolitográficas. Se han producido chips que pueden analizar hasta 1.700 genes. Las micromatrices de ADN son matrices de muestras de ADN, típicamente de productos de PCR, que se graban robóticamente sobre un portaobjetos de microscopía. Se analiza cada gen con una secuencia de ADN diana de longitud total o parcial. Las micromatrices con hasta 10.000 genes ahora se preparan de modo rutinario comercialmente. La diferencia principal entre estas dos técnicas es que los chips de oligonucleótidos típicamente utilizan oligonucleótidos 25-meros que permiten el fraccionamiento de moléculas cortas de ADN, mientras que las dianas de ADN más grandes de las micromatrices, de aproximadamente 1.000 pares de bases, pueden proporcionar más sensibilidad en el fraccionamiento de mezclas complejas de ADN.

50 Otra medida útil de los sistemas y procedimientos de regulación de múltiples genes de los solicitantes de la invención es la de la determinación del estado de traducción de la célula mediante la medición de las abundancias de las especies proteicas constituyentes presentes en la célula utilizando procedimientos bien conocidos de la técnica.

55 En el caso de que se desee la identificación de los genes asociados a diversas funciones fisiológicas, puede utilizarse un ensayo en el que se miden funciones tales como el crecimiento celular, la apoptosis, la senescencia, la diferenciación, la adhesión, la unión a moléculas específicas, la unión a otra célula, la organización celular, la organogénesis, el transporte intracelular, la facilitación del transporte, la conversión energética, el metabolismo, la miogénesis, la neurogénesis y/o la hematopoyesis.

60 Además, puede utilizarse la expresión de un marcador seleccionable o gen informador para medir la expresión génica utilizando la invención del solicitante.

65 Otros procedimientos para detectar los productos de expresión génica son bien conocidos de la técnica y entre ellos se incluyen los análisis de transferencia southern (detección de ADN), de membranas de puntos o ranuras (ADN, ARN), de transferencia northern (ARN), RT-PCR (ARN), de transferencia western (detección de polipéptidos) y

ELISA (polipéptidos). Aunque menos preferente, pueden utilizarse proteínas marcadas para detectar una secuencia particular de ácidos nucleicos con la que se hibridan.

En algunos caso resulta necesario amplificar la cantidad de una secuencia de ácidos nucleicos. Lo anterior puede llevarse a cabo utilizando uno o más de entre varios procedimientos adecuados, incluyendo, por ejemplo, la reacción en cadena de la polimerasa ("PCR"), la reacción en cadena de la ligasa ("LCR"), la amplificación por desplazamiento de cadena ("SDA"), la amplificación basada en la transcripción, y similares. La PCR se lleva a cabo según técnicas conocidas, en las que, por ejemplo, se trata una muestra de ácidos nucleicos en presencia de una ADN polimerasa termolábil, bajo condiciones de hibridación, con un par de cebadores oligonucleótidos, con un cebador que se hibrida con una cadena (molde) de la secuencia específica que debe detectarse. Los cebadores son suficientemente complementarios a cada cadena molde de la secuencia específica para hibridarse con la misma. Se sintetiza un producto de extensión de cada cebador, que es complementario a la cadena molde de ácidos nucleicos a la que se hibrida. El producto de extensión sintetizado a partir de cada cebador también puede servir como molde para la síntesis posterior de productos de extensión utilizando los mismos cebadores. Tras un número suficiente de rondas de síntesis de productos de extensión, puede analizarse la muestra tal como se ha descrito anteriormente con el fin de evaluar si la secuencia o secuencias que deben detectarse se encuentran presentes.

La presente invención podrá entenderse mejor haciendo referencia a los Ejemplos no limitativos siguientes, que se proporcionan a título ejemplar de la invención.

## Ejemplos

### Procedimientos generales

Según la presente invención pueden utilizarse técnicas convencionales de la biología molecular, microbiología y de ADN recombinante comprendidas dentro de los conocimientos del experto en la materia. Dichas técnicas se explican exhaustivamente en la literatura. Ver, por ejemplo, Sambrook, Fritsch y Maniatis, *Molecular Cloning: A Laboratory Manual*, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989 (en la presente memoria, "Sambrook *et al.*, 1989"); DNA Cloning: A Practical Approach, volúmenes I y II (D.N. Glover, editor, 1985); *Oligonucleotide Synthesis* (M.J. Gait, editor, 1984); *Nucleic Acid Hybridization* [B.D.Hames y S.J. Higgins, editores, 1985]; *Transcription And Translation* [B.D. Hames y S.J. Higgins, editores, 1984]; *Animal Cell Culture* [R.I. Freshney, ed. (1986)]; *Immobilized Cells And Enzymes* [IRL Press, (1986)]; B. Perbal, *A Practical Guide To Molecular cloning*, 1984; F.M. Ausubel *et al.* (editores), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc., 1984.

Entre los vehículos de clonación convencionales se incluyen los plásmidos de tipo pBR322 y pUC y fagos de la serie de M13. Estos pueden obtenerse comercialmente (Bethesda Research Laboratories).

Para la ligación, los fragmentos de ADN pueden separarse según su tamaño mediante electroforesis en gel de agarosa o de acrilamida, extraerse con fenol o con una mezcla de fenol/cloroformo, precipitarse con etanol y después incubarse en presencia de ADN ligasa de fago T4 (Biolabs) siguiendo las recomendaciones del proveedor.

El rellenado de los extremos 5' protuberantes puede llevarse a cabo con el fragmento Klenow de la ADN polimerasa I de *E. coli* (Biolabs) siguiendo las especificaciones del proveedor. La destrucción de los extremos 3' protuberantes se lleva a cabo en presencia de ADN polimerasa de T4 (Biolabs) utilizada siguiendo las recomendaciones del fabricante. La destrucción de los extremos 5' protuberantes se lleva a cabo mediante un tratamiento controlado con nucleasa S1.

La mutagénesis dirigida *in vitro* por oligodesoxinucleótidos sintéticos puede llevarse a cabo según el procedimiento desarrollado por Taylor *et al.* [Nucleic Acids Res. 13:8749-8764, 1985] utilizando el kit distribuido por Amersham.

La amplificación enzimática de fragmentos de ADN mediante la técnica de la PCR [Polymerase-catalyzed Chain Reaction, Saiki R.K. *et al.*, Science 230:1350-1354, 1985; Mullis K.B. y Faloona F.A., Meth. Enzym. 155:335-350, 1987] puede llevarse a cabo utilizando un "ciclador térmico de ADN" (Perkin Elmer Cetus) siguiendo las especificaciones del fabricante.

La verificación de las secuencias de nucleótidos puede llevarse a cabo mediante el procedimiento desarrollado por Sanger *et al.* [Proc. Natl. Acad. Sci. USA 74:5463-5467, 1977] utilizando el kit distribuido por Amersham.

Los ADNs plasmídicos pueden purificarse mediante el sistema de purificación de plásmidos de Qiagen siguiendo las instrucciones del fabricante.

Las manipulaciones de secuencias genéticas pueden llevarse a cabo utilizando el conjunto de programas disponibles del Genetics Computer Group Inc. (Wisconsin Package versión 9.0, Genetics Computer Group (GCG), Madison, WI). En el caso de que se utilice el programa "Pileup" del GCG, puede utilizarse un valor por defecto de creación de huecos de 12, y un valor por defecto de extensión de hueco de 4. En el caso de que se utilicen los programas "Gap" o "Bestfit" del GCG, pueden utilizarse valores por defecto de penalización por creación de hueco

de 50 y de penalización por extensión de hueco de 3. En cualquier caso, en el caso de que no se soliciten los parámetros del programa del CGC, en estos y en otros programas del CGC, pueden utilizarse los valores por defecto.

5 Los significados de las abreviaturas son los siguientes: "h" significa hora(s), "min" significa minuto(s), "sec" significa segundo(s), "d" significa día(s), "µl" significa microlitro(s), "ml" significa mililitro(s), "L" significa litro(s), "µM" significa micromolar, "mM" significa milimolar, "µg" significa microgramo(s), "mg" significa miligramo(s), "A" significa adenina o adenosina, "T" significa timina o timidina, "G" significa guanina o guanosina, "C" significa citidina o citosina, "xg" significa número de veces el valor de la gravedad, "nt" significa nucleótido(s), "aa" significa aminoácido(s), "bp" significa par(es) de bases, "kb" significa kilobase(s), "k" significa kilo, "µ" significa micro y "°C" significa grados Celsius.

### Ejemplo 1

15 El presente ejemplo describe la construcción de varios casetes de expresión génica para la utilización en un sistema de expresión de múltiples genes inducibles según la invención. Los solicitantes construyeron varios casetes de expresión génica basados en el EcR del gusano del abeto *Choristoneura fumiferana* ("CfEcR"), el EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR"), el receptor de ecdisona del saltahoja verde *Nephotetix cincticeps* ("NcEcR"), el receptor X de la isoforma α del retinoide de ratón *Mus musculus* ("MmRXRα") y la proteína Ultraspiracle homóloga de RXRT de invertebrado de langosta *Locusta migratoria* ("LmUSP"). Los constructos de receptor preparados comprendían un dominio de unión a ligando de un EcR o de un RXR de vertebrado, y un dominio de unión a ADN (DBD) de GAL4 o un dominio de transactivación (AD) de VP16. Entre los constructos informadores se incluían un gen informador, luciferasa (Luc) o fosfatasa alcalina secretada (SEAP), funcionalmente ligados a un constructo sintético de promotor que comprendía un elemento de respuesta a GAL4 al que se une el DBD de GAL4. Se cotransfectaron diversas combinaciones de estos constructos de receptor e informador en células de mamífero tal como se ha descrito anteriormente.

**Casetes de expresión génica:** se construyeron casetes de expresión génica para la utilización en dos sistemas de expresión génica inducible de doble interruptor basados en receptores de ecdisona, de la manera siguiente utilizando procedimientos estándares de clonación disponibles de la técnica. A continuación se proporciona una breve descripción de la preparación y composición de cada interruptor utilizado en los Ejemplos descritos en la presente memoria.

35 1.1 - Quimera GAL4DmEcR-CDEF/VP16MmRXR-LmUSP-EF y quimera LexACfEcR-CDEF/VP16MmRXR-LmUSP: se fusionó un polinucleótido codificante de los dominios C, D, E y F del EcR de la mosca de la fruta (*Drosophila melanogaster*) ("DmEcR-CDEF", SEC ID nº 1) con un polinucleótido codificante de un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD", SEC ID nº 2) y se situó bajo el control de un promotor/intensificador de citomegalovirus (CMV) (SEC ID nº 3). Se fusionó un polinucleótido codificante de los dominios C, D, E y F del EcR del gusano del abeto (*Choristoneura fumiferana*) ("CfEcR-CDEF", SEC ID nº 4) con un polinucleótido codificante de un dominio de unión a ADN de LexA ("LexADNABD" o "LexADBD", SEC ID nº 5) y se situó bajo el control de un promotor/intensificador de citomegalovirus (CMV) (SEC ID nº 3). Se fusionaron un polinucleótido codificante de un polipéptido quimérico de dominios EF del receptor X del retinoide isoforma α de ratón (*Mus musculus*) ("MmRXRα") y proteína Ultraspiracle de langosta (*Locusta migratoria*) ("LmUSP-EF") (SEC ID nº 6) con un polinucleótido codificante de un dominio de transactivación de VP16 ("VP16AD", SEC ID nº 7) y se situó bajo el control de un promotor/intensificador de CMV (SEC ID nº 3). Se fusionaron seis sitios de unión al elemento de respuesta Gal4 de consenso ("6XGAL4RE", SEC ID nº 8) con un promotor mínimo de la albúmina (SEC ID nº 9) y se situaron bajo el control del gen de la fosfatasa alcalina secretada (SEAP) (SEC ID nº 10). Se fusionaron ocho sitios de unión a elemento de respuesta LexA de consenso ("8opLexARE", SEC ID nº 11) con un TATAA sintético (SEC ID nº 12) y se situaron cadena arriba del gen de la luciferasa (SEC ID nº 13).

50 1.2 - GAL4CfEcR-DEF/VP16MmRXRα-EF y GAL4NcEcR-CDE/VP16MmRXRα-EF: este constructo se preparó de la manera siguiente. Se fusionó un polinucleótido codificante de los dominios D, E y F del EcR del gusano del abeto (*Choristoneura fumiferana*) ("CfEcR-DEF", SEC ID nº 14) con un polinucleótido codificante de un dominio de unión a ADN de GAL4 ("GAL4DBD", SEC ID nº 2) y se situó bajo el control de un promotor/intensificador de citomegalovirus (CMV) (SEC ID nº 3). Se fusionó un polinucleótido codificante de los dominios C, D y E del receptor de ecdisona del saltahoja verde (*Nephotetix cincticeps*) ("NcEcR-CDE", SEC ID nº 15) con un polinucleótido codificante de un dominio de unión a ADN de GAL4 ("GAL4DBD", SEC ID nº 2) y se situó bajo el control de un promotor/intensificador de citomegalovirus (CMV) (SEC ID nº 3). Se fusionó un polinucleótido codificante de los dominios E y F del receptor X del retinoide isoforma α de ratón (*Mus musculus*) ("MmRXRα", SEC ID nº 16) con un polinucleótido codificante de un dominio de transactivación de VP16 ("VP16AD", SEC ID nº 7) y se situó bajo el control de un promotor/intensificador de CMV (SEC ID nº 3). Se fusionaron seis sitios de unión a elemento de respuesta GAL4 de consenso ("6xGAL4RE", SEC ID nº 8) con un TATAA sintético (SEC ID nº 12) y se situaron cadena arriba del gen de la luciferasa de luciérnaga (SEC ID nº 13).

65 Los sistemas de doble interruptor resultantes de los Ejemplos 1.1 y 1.2 se sometieron a ensayo para identificar actividades mediante la transfección de los mismos en células NIH3T3 o CHO en presencia de ligando esteroideo

ponasterona A (PonA) y ligando no esteroideo N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina (GS<sup>TM</sup>-E).

5 **Ligandos:** se obtuvo el ligando esteroideo ponasterona A de Sigma Chemical Company. El ligando no esteroideo N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina (ligando no esteroideo GS<sup>TM</sup>-E) es un ligando ecdisteroide estable sintético sintetizado en Rohm and Haas Company. Se disolvieron ambos ligandos en DMSO y se mantuvo la concentración final de DMSO a 0,1% en tanto controles como tratamientos.

10 **Transfecciones:** los ADN correspondientes a los constructos de doble interruptor descritos de manera general en los Ejemplos 1.1 y 1.2 se transfectaron en células NIH3T3 de ratón (ATCC, Ejemplo 1.1) o en células CHO (ATCC, Ejemplo 1.2), de la manera siguiente. Se siguieron procedimientos estándares para el cultivo y mantenimiento de las células. Las células se recolectaron al alcanzar una confluencia de 50% y se sembraron en placas de 6, 12 ó 24 pocillos a razón de 125.000, 50.000 ó 25.000 células, respectivamente en 2,5, 1,0 ó 0,5 ml de medio de cultivo que contenía suero de feto bovino (FBS) al 10%, respectivamente. Al día siguiente, se enjuagaron las células con medio de cultivo y se transfectaron durante cuatro horas. Se utilizaron como reactivos de transfección Superfect<sup>TM</sup> (Qiagen Inc.) para las células 3T3 y LipofectAMINE<sup>TM</sup> (Life Technologies) para las células CHO. Para las placas de 12 pocillos, se mezclaron 4 µl de Superfect<sup>TM</sup> o LipofectAMINE<sup>TM</sup> con 100 µl de medio de cultivo. Se añadieron a la mezcla de transfección un µg de constructo informador y 0,25 µg de cada constructo de receptor de la pareja de receptores que debía analizarse. Se añadió un segundo constructo de informador [pTKRL (Promega), 0,1 µg/mezcla de transfección] que comprendía un gen de luciferasa de *Renilla* funcionalmente ligado y situado bajo el control de un promotor constitutivo de timidina quinasa (TK) y se utilizó para la normalización. Se mezcló el contenido de la mezcla de transfección en un mezclador vórtex y se dejó reposar a temperatura ambiente durante 30 minutos. Al final de la incubación, se añadió la mezcla de transfección a las células mantenidas en 400 µl de medio de cultivo. Se mantuvieron las células a 37°C y con 5% de CO<sub>2</sub> durante cuatro horas. Al final de la incubación, se añadieron 500 µl de medio de cultivo que contenía FBS al 20% y dimetilsulfóxido (DMSO, control) o una solución en DMSO de ligando esteroideo o no esteroideo, y las células se mantuvieron a 37°C y con 5% de CO<sub>2</sub> durante 48 horas. Se recolectaron las células y se sometió a ensayo la actividad de informador. Se siguió el mismo procedimiento para las placas de 6 y de 24 pocillos, excepto en que se dobló la cantidad de todos los reactivos para las placas de 6 pocillos y se redujo a la mitad para las placas de 24 pocillos.

30 **Ensayos de informador:** se recolectaron las células 48 horas después de añadir ligandos y se cuantificaron las actividades de informador utilizando el sistema de ensayo de informador Dual-luciferasa<sup>TM</sup> de Promega Corporation. Se añadieron 125 µl de tampón de lisis pasivo (parte del sistema de ensayo de informador Dual-luciferasa<sup>TM</sup> de Promega Corporation) a cada pocillo de la placa de 24 pocillos. Se introdujeron las placas en un agitador rotatorio durante 15 minutos. Se sometieron a ensayo veinte µl del lisado. Se midió la actividad de luciferasa utilizando el sistema de ensayo de informador Dual-luciferasa<sup>TM</sup> de Promega Corporation siguiendo las instrucciones del fabricante. Se midió la actividad de fosfatasa alcalina utilizando el kit de ensayo Phospholight<sup>TM</sup> de TROPIX siguiendo las instrucciones del fabricante. Se normalizaron todas las actividades de luciferasa y de fosfatasa alcalina utilizando la luciferasa de *Renilla* como estándar. Se calculó el factor de variación de las actividades dividiendo las unidades lumínicas relativas normalizadas ("RLU") en células tratadas con ligando por las RLU normalizadas en células tratadas con DMSO (control no tratado).

## Ejemplo 2

45 El presente Ejemplo describe la capacidad de un sistema de expresión génica de doble interruptor de la invención de los solicitantes de modular la expresión de dos casetes informadores de expresión génica, en el que dos casetes informadores de expresión génica se regulan independientemente con dos ligandos diferentes. Específicamente, un casete informador de expresión génica se regula induciblemente con un ligando esteroide y el otro casete informador de expresión génica se regula induciblemente con un ligando no esteroideo. Brevemente, los solicitantes prepararon un sistema de expresión génica inducible de doble interruptor tal como se ha descrito anteriormente, en el Ejemplo 1.1. El sistema de doble interruptor resultante seguidamente se sometió a ensayo en células de mamífero NIH3T3, de la manera siguiente.

55 Los ADN correspondientes a los constructos de doble interruptor descritos de manera general en el Ejemplo 1.1 se transfectaron en células NIH3T3 de ratón (ATCC) tal como se describe en el Ejemplo 1. Al final del periodo de incubación de transfección, se añadieron 250 µl de medio de cultivo que contenía FBS al 20% y dimetilsulfóxido (DMSO, control) o una solución en DMSO de ligando esteroideo PonA 0,02, 0,1, 0,5 ó 2,5 µM y/o ligando no esteroideo GS<sup>TM</sup>-E [N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina] y las células se mantuvieron a 37°C y con 5% de CO<sub>2</sub> durante 48 horas. Se recolectaron las células y se sometieron a ensayo las actividades de informador tal como se ha indicado anteriormente.

65 Tal como se muestra en la figura 1, al tratar las células con ligando no esteroideo solo, únicamente se indujo la actividad de luciferasa (ver la figura 1A). Al tratar las células con ligando esteroideo solo, únicamente se indujo actividad de informador SEAP (ver la figura 1B). Al tratar las células con tanto ligando esteroideo como ligando no esteroideo, se indujeron ambas actividades génicas de informador (ver la figura 1C). El presente Ejemplo muestra un

sistema de expresión de múltiples genes inducibles que comprende dos sistemas de expresión génica individualmente operables, uno basado en EcR de díptero (DmEcR) y el otro basado en EcR de lepidóptero (CfEcR).

### Ejemplo 3

5 El presente Ejemplo describe la capacidad de un sistema de expresión génica de doble interruptor de la invención de los solicitantes de modular la expresión de dos casetes informadores de expresión génica, en el que los dos casetes informadores de expresión génica se regulan independientemente con dos ligandos diferentes. En particular, un casete informador de expresión génica resulta induciblemente regulado por un ligando esteroideo y el otro casete informador de expresión génica resulta induciblemente regulado por un ligando no esteroideo. Brevemente, los solicitantes prepararon un sistema de expresión génica inducible de doble interruptor tal como se ha descrito anteriormente, en el Ejemplo 1.2. El sistema de doble interruptor resultante seguidamente se sometió a ensayo en células CHO de ovario de hámster chino, de la manera siguiente.

15 Los ADN correspondientes a los constructos de doble interruptor descritos de manera general en el Ejemplo 1.2 se transfectaron en células CHO de hámster (ATCC) tal como se describe en el Ejemplo 1. Las células CHO se transfectaron con: 1) GAL4CfEcR-DEF/VP16MmRXR $\alpha$ -EF y pFRLuc, ó 2) GAL4NcEcR-CDE/VP16MmRXR $\alpha$ -EF y pFRLuc. Al final del periodo de incubación de transfección, se añadieron 250  $\mu$ l de medio de cultivo que contenía FBS al 20% y dimetilsulfóxido (DMSO, control) o una solución en DMSO de ligando esteroideo PonA 0,1, 1,5 ó 10  $\mu$ M o ligando no esteroideo GS TM-E [N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina], y las células se mantuvieron a 37°C y con 5% de CO<sub>2</sub> durante 48 horas. Se recolectaron las células y se sometió a ensayo la actividad de informador tal como se ha descrito anteriormente.

25 Tal como se muestra en la figura 2, al tratar con únicamente ligando no esteroideo las células transfectadas con GAL4CfEcR-DEV/VP16MmRXR $\alpha$ -EF y pFRLuc, se indujo actividad de luciferasa regulada por CfEcR-DEF (ver CfEcR/GSE de la figura 2); sin embargo, el tratamiento de estas células transfectadas con el ligando esteroideo PonA no indujo la expresión gen informador (ver CfEcR/PonA de la figura 2). Al tratar con únicamente ligando esteroideo las células transfectadas con GAL4NcEcR-CDE/VP16MmRXR $\alpha$ -EF y pFRLuc, se indujo actividad de informador luciferasa regulada por NcEcR-DE (ver NcEcR/PonA de la figura 2); sin embargo, el tratamiento de estas células transfectadas con ligando no esteroideo no indujo la expresión de gen informador (ver NcEcR/GSE de la figura 2). La insensibilidad de CfEcR-DEF a PonA y la insensibilidad de NcEcR-CDE a GSTM-E permite que los dos sistemas de modulación de la expresión génica descritos en el presente ejemplo sean modulador ortogonalmente. De esta manera, el presente Ejemplo muestra dos sistemas de expresión génica individualmente operables, uno basado en EcR de lepidóptero (CfEcR) y el otro, basado en EcR de homóptero (NcEcR), para la utilización en un sistema de expresión de múltiples genes inducibles de la invención.

### LISTADO DE SECUENCIAS

40 <110> Rohm and Haas Company Hormann, Robert E. Palli, Subba Reddy Carlson, Glenn R. Cress, Dean E. Dhadialla, Tarlochan S. Herzig, Ronald P. Kudla, Arthur J. Philip, Mohan

<120> Sistema de regulación de múltiples genes inducibles

<130> RH-0037 (DN A01115A)

45 <150> 60/237,446

<151> 2000-10-03

<160> 16

50 <170> Patente en versión 3.1

<210> 1

<211> 1878

55 <212> ADN

<213> Drosophila melanogaster

<400> 1

```

-----
ggacctgctg caccgggtgca agaggagctg tgcctggttt gcggcgacag ggcctccggc      60
taccactaca acgcccctcac ctgtgagggc tgcaaggggt tctttcgcag cagcgttacg      120
aagagcgcgc tctactgctg caagttcggg cgcgcctgcg aaatggacat gtacatgagg      180
cgaaagtgtc aggagtgcgc cctgaaaaag tgcctggccg tgggtatgcg gccggaatgc      240
gtcgtcccg  agaaccaatg tgcgatgaag cggcgcgaaa agaaggccca gaaggagaag      300
gacaaaatga cacttcgcc  gagctctcag catggcgca  atggcagctt ggcctctggt      360
ggcggccaag actttgttaa gaaggagatt cttgacctta tgacatgca gccgcccag      420
catgccacta ttccgctact acctgatgaa atattggcca agtgtcaagc gcgcaatata      480
ccttccttaa cgtacaatca gttggccgtt atatacaagt taatttgta ccaggatggc      540
tatgagcagc catctgaaga ggatctcagg cgtataatga gtcaaccga tgagaacgag      600
agccaaacgg acgtcagctt tcggcatata accgagataa ccatactcac ggtccagttg      660
attgttgagt ttgctaaagg tctaccagcg tttacaaaga taccacagga ggaccagatc      720
acgttactaa aggcctgctc gtcggaggtg atgatgctgc gtatggcacg acgctatgac      780
cacagctcgg actcaatatt cttcgcgaat aatagatcat atacgcggga ttcttataaa      840
atggccggaa tggctgataa cattgaagac ctgctgcatt tctgccgcca aatgttctcg      900
atgaaggtgg acaacgtcga atacgcgctt ctcactgcca ttgtgatctt ctcggaccgg      960
ccgggcctgg agaaggccca actagtcgaa gcgatccaga gctactacat cgacacgcta     1020
cgcatttata tactcaaccg cactgcggc  gactcaatga gcctcgtctt ctacgaaaag     1080
ctgctctcga tcctcaccga gctgcgtacg ctgggcaacc agaacgccga gatgtgttc     1140
tcactaaagc tcaaaaaccg caaactgcc  aagttcctcg aggagatctg ggacgttcac     1200

gccatcccg  catcgggtcca gtcgcacctt cagattacc  aggaggagaa cgagcgtctc     1260
gagcgggctg agcgtatgcg ggcacggtt  gggggcgcca ttaccgccg  cattgattgc     1320
gactctgctt cacttcggc  ggcggcagcc gcggcccagc atcagcctca gcctcagccc     1380
cagccccaac cctcctccct gaccagaac  gattcccagc accagacaca gccgcagcta     1440
caacctcagc taccacctca gctgcaaggt caactgcaac cccagctcca accacagett     1500
cagacgcaac tccagccaca gattcaacca cagccacagc tccttcccgt ctccgctccc     1560
gtgcccgcct ccgtaaccgc acctggttcc ttgtccgcyg tcagtacgag cagcgaatac     1620
atgggcggaa gtgcggccat aggaccatc  acgcccggca ccaccagcag taccacggct     1680
gccgttaccg ctagctccac cacatcagcg gtaccgatgg gcaacggagt tggagtcggg     1740
gttgggtgg  gcggcaacgt cagcatgtat gcgaacgccc agacggcgat ggccttgatg     1800
ggtgtagccc tgcattcgca ccaagagcag cttatcgggg gagtggcggg taagtcggag     1860
cactcgacga ctgcatag                                     1878

```

- 5 <210> 2
- <211> 441
- <212> ADN
- <213> Saccharomyces cerevisiae
- <400> 2

atgaagctac tgtcttctat cgaacaagca tgcgatattt gccgacttaa aaagctcaag 60  
 tgc tccaaaag aaaaaccgaa gtgcgccaag tgtctgaaga acaactggga gtgtcgctac 120  
 tctcccaaaa ccaaaaggtc tccgctgact agggcacatc tgacagaagt ggaatcaagg 180  
 ctagaaagac tggacagct atttctactg atttttcctc gagaagacct tgacatgatt 240  
 ttgaaaatgg attcctttaca ggatataaaa gcattgttaa caggattatt tgtacaagat 300  
 aatgtgaata aagatgccgt cacagataga ttggcttcag tggagactga tatgcctcta 360  
 acattgagac agcatagaat aagtgcgaca tcatcatcgg aagagagtag taacaaaggt 420  
 caaagacagt tgactgtatc g 441

5 <210> 3  
 <211> 750  
 <212> ADN  
 <213> cytomegalovirus

<400> 3  
 tcaatattgg ccattagcca tattattcat tgggtatata gcataaatca atattggcta 60  
 ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120  
 aatatgaccg ccattgtggc attgattatt gactagtatt taatagtaat caattacggg 180  
 gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240  
 gcctggctga ccgcccaacg acccccggcc attgacgtca ataatgacgt atgttcccat 300  
 agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360  
 ccacttgcca gtacatcaag tgtatcatat gccaaagtcg ccccctattg acgtcaatga 420  
  
 cggtaaatgg cccgcctggc attatgccca gtacatgacc ttacgggact ttcctacttg 480  
 gcagtacatc tacgtattag tcatcgctat taccatgggtg atgcggtttt ggcagtacac 540  
 caatggcggt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600  
 caatggggagt ttgttttggc accaaaatca acgggacttt ccaaaatgac gtaacaactg 660  
 cgatcggccc ccccggtgac gcaaatgggc ggtaggcgtg tacggtgga ggtctatata 720  
 agcagagctc gtttagtgaa ccgtcagatc 750

10  
 <210> 4  
 <211> 1290  
 <212> ADN  
 15 <213> Choristoneura fumiferana

<400> 4

```

agaaggcccc tgctgaccgt cagcaagagg aactgtgtct ggtatgcggg gacagagcct    60
ccggatacca ctacaatgcg ctcacgtgtg aagggtgtaa agggttcttc agacggagtg    120
ttacacaaaa tgcggtttat atttghaaat tcggtcacgc ttgcgaaatg gacatgtaca    180
tgcgacggaa atgccaggag tgcgcctga agaagtgtct agctgtaggc atgaggcctg    240
agtgcgtagt acccgagact cagtgcgcca tgaagcggaa agagaagaaa gcacagaagg    300
agaaggacaa actgcctgtc agcacgacga cggtgacga ccacatgccg cccattatgc    360
agtgtgaacc tccacctcct gaagcagcaa ggattcacga agtggttcca aggtttctct    420
ccgacaagct gttggagaca aaccggcaga aaaacatccc ccagttgaca gccaaccagc    480
agttccttat cgccaggctc atctggtacc aggacgggta cgagcagcct tctgatgaag    540
atltgaagag gattacgcag acgtggcagc aagcggacga tgaaaacgaa gagtctgaca    600
ctcccttccg ccagatcaca gagatgacta tcctcacggc ccaacttate gtggagttcg    660
cgaagggatt gccagggttc gccaaagatct cgcagcctga tcaaattacg ctgcttaagg    720
cttgctcaag tgaggtaatg atgctccgag tcgctcgacg atacgatgcg gcctcagaca    780
gtgttctgtt cgcaacaac caagcgtaca ctcgcgacaa ctaccgcaag gctggcatgg    840
cctacgtcat cgaggatcta ctgcacttct gccggtgcat gtactctatg gcgttgga    900
acatccatta cgcgctgctc acggctgtcg tcacttttcc tgaccggcca gggttggagc    960
agccgcaact ggtggaagaa atccagcggc actacctgaa tacgctccgc atctatatcc   1020
tgaaccagct gagcgggtcg gcgcgttctg ccgtcatata cggcaagatc ctctcaatcc   1080
tctctgagct acgcacgctc ggcatgcaaa actccaacat gtgcatctcc ctcaagctca   1140
agaacagaaa gctgccgcct tcctcagagg agatctggga tgtggcagga catgtcgcac   1200
acccaaccgc cgcctatctc gagtccccca cgaatctcta gccctcgcgc gcacgcatcg   1260
ccgatgccgc gtccggccgc gctgctctga                                     1290

```

<210> 5  
 <211> 606  
 <212> ADN

5 <213> Escherichia coli

<400> 5

```

atgaaagcgt taacggccag gcaacaagag gtgtttgatc tcatccgtga tcacatcagc    60
cagacaggta tgccgccgac gcgtgcggaa atcgcgcagc gtttgggggt ccgttccccca   120
aacgcggctg aagaacatct gaaggcgtg gcacgcaaag gcgttattga aattgtttcc   180
ggcgcacac gcgggattcg tctgttgacg gaagaggaag aagggttggc gctggtaggc   240
cgtgtggctg ccggtgaacc acttctggcg caacagcata ttgaaggcca ttatcaggtc   300
gatccttctt tattcaagcc gaatgctgat ttcctgctgc gcgtcagcgg gatgtcgatg   360
aaagatatcg gcattatgga tggtagcttg ctggcagtyc ataaaactca ggatgtacgt   420
aacggtcagg tcgttgctgc acgtattgat gacgaagtta ccgttaagcg cctgaaaaaa   480
cagggcaata aagtcgaact gttgccagaa aatagcgagt ttaaaccaat tgtcgtagat   540
cttcgtcagc agagcttcac cattgaaggg ctggcgggtg gggttattcg caacggcgac   600
tggtctg                                     606

```

10 <210> 6  
 <211> 711  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <223> Chimeric MmRXR/LmUSP-EF

<400> 6  
 gccaacgagg acatgcctgt agagaagatt ctggaagccg agcttgctgt cgagcccaag 60  
 actgagacat acgtggaggc aaacatgggg ctgaaccca gctcaccaaa tgaccctgtt 120  
 accaacatct gtcaagcagc agacaagcag ctcttcactc ttgtggagtg ggccaagagg 180  
 atccccacact tttctgagct gccctagac gaccaggcca tcctgctacg ggcaggctgg 240  
 aacgagctgc tgatcgcctc cttctccac cgctccatag ctgtgaaaga tgggattctc 300  
 ctggccaccg gcctgcacgt acaccggaac agcgcctaca gtgctggggg gggcgccatc 360  
 tttgacaggg tgctaacaga gctgggtgtct aagatgctg acatgcagat ggacaagact 420  
 gaacttggtc gcttgcgacg tgttattctt ttcaatccag aggtgagggg tttgaaatcc 480  
 gccagggaag ttgaacttct acgtgaaaa gtatatgccg ctttggaga atatactaga 540  
 acaacacatc ccgatgaacc aggaagattt gcaaaaactt tgcttcgtct gccttcttta 600  
 cgttccatag gccttaagtg tttggagcat ttgttttct ttcgccttat tggagatgtt 660  
 ccaattgata cgttcctgat ggagatgctt gaatcacctt ctgattcata a 711

5 <210> 7  
 <211> 681  
 <212> ADN  
 <213> virus del herpes simple 7

10 <400> 7  
 atgggcceta aaaagaaga gcgtaaggtc aaagcgttaa cggccaggct tgaattaatt 60

ccgggaggaa taaaagcgtt aacggccagg caacaagagg tgtttgatct catccgtgat 120  
 cacatcagcc agacaggtat gccgccgacg cgtgcggaaa tcgcgacg tttggggttc 180  
 cgttcccaaa acgcggtga agaacatctg aaggcgctgg cacgcaaagg cgttattgaa 240  
 attgtttccg gcgcatcacg cgggattcgt ctggtgcagg aagaggaaga aggggtgccg 300  
 ctggtaggtc gtgtggctgc cgggaacca cttctggcgc aacagcatat tgaaggatcat 360  
 tatcaggtcg atccttcctt attcaagccg aatgctgatt tcctgctgcg cgtcagcggg 420  
 atgtcgatga aagatatcgg cattatggat ggtgacttgc tggcagtgca taaaactcag 480  
 gatgtacgta acggtcaggt cgttgtcga cgtattgatg acgaagttac cgtaagcgc 540  
 ctgaaaaaac agggcaataa agtcgaactg ttgccagaaa atagcgagt taaaccaatt 600  
 gtcgtagatc ttcgtcagca gagcttacc attgaagggc tggcgggttg ggttattcgc 660  
 aacggcgact ggctggaatt c 681

15 <210> 8  
 <211> 117  
 <212> ADN  
 <213> Saccharomyces cerevisiae

<400> 8  
 gcggagtact gtcctccgag cggagtactg tcctccgagc ggagtactgt cctccgagcg 60  
 gagtactgtc ctccgagcgg agtactgtcc tccgagcggg gtaactgtcct ccgagcgg 117

20 <210> 9  
 <211> 138  
 <212> ADN  
 <213> Mus musculus

<400> 9

ES 2 364 923 T3

atcttttgtt gactaagtca ataatcagaa tcagcaggtt tggagtcagc ttggcagggg 60  
 tcagcagcct gggttggaag gagggggtat aaaagcccct tcaccaggag aagccgtcac 120  
 acagatccac aagctcct 138

- <210> 10
- <211> 1560
- <212> ADN
- 5 <213> Homo sapiens
- <400> 10

atgctgctgc tgctgctget gctgggcctg aggctacagc tctccctggg catcatccca 60  
 gttgaggagg agaaccggga cttctggaac cgcgaggcag ccgaggccct gggtgccgcc 120  
 aagaagctgc agcctgcaca gacagccgcc aagaacctca tcatcttctt gggcgatggg 180  
 atgggggtgt ctacggtgac agctgccagg atcctaaaag ggcagaagaa ggacaaactg 240  
 gggcctgaga taccctggc catggaccgc tccccatag tggctctgtc caagacatac 300  
 aatgtagaca aacatgtgcc agacagtgga gccacagcca cggcctacct gtgcggggtc 360  
 aagggcaact tccagaccat tggcttgagt gcagccgccc gctttaacca gtgcaacacg 420  
 acacgcgcca acgaggtcat ctccgtgatg aatcgggcca agaaagcagg gaagtcagtg 480

ggagtggtaa ccaccacagc agtgcagcac gcctcgccag ccggcaccta cgcccacacg 540  
 gtgaaccgca actggtactc ggacgccgac gtgctgcctt cggcccgcca ggaggggtgc 600  
 caggacatcg ctacgcagct catctccaac atggacattg acgtgatcct aggtggaggc 660  
 cgaaagtaca tgtttcgcac gggaacccca gaccctgagt acccagatga ctacagccaa 720  
 ggtgggacca ggtggacgg gaagaatctg gtgcaggaat ggctggcga gcccagggt 780  
 gcccgtatg tgtggaaccg cactgagctc atgcaggctt ccctggacc gtctgtgacc 840  
 catctcatgg gtctcttga gcctggagac atgaaatacg agatccaccg agactccaca 900  
 ctggacccct ccctgatgga gatgacagag gctgccctgc gcctgctgag caggaacccc 960  
 cgcggcttct tcctcttctg ggaggggtgt cgcacgacc atggtcatca tgaaagcagg 1020  
 gcttaccggg cactgactga gacgatcatg ttcgacgacg ccattgagag ggcgggccag 1080  
 ctcaccagcg aggaggacac gctgagcctc gtcactgccg accactccca cgtcttctcc 1140  
 ttcggaggct acccctgcg agggagctcc atcttcgggc tggcccctgg caaggcccgg 1200  
 gacaggaag cctacacggt cctcctatac gaaacggtc caggctatgt gctcaaggac 1260  
 ggcgccggc cggatgttac cgagagcgag agcgggagcc ccgagtatcg gcagcagtca 1320  
 gcagtcccc tggacgaaga gaccacgca ggcgaggacg tggcgggtgt cgcgcgcgcc 1380  
 ccgcagcgcg acctggttca cggcgtgacg gacgagacct tcatagcgca cgtcatggcc 1440  
 ttcgccgctt gcctggagcc ctacaccgce tgcgacctgg cgcctcccgc cggcaccacc 1500  
 gacgccgccc acccgggtta ctctagagtc ggggcggccc gccgcttcca gcagacatga 1560

- 10
- <210> 11
- <211> 206
- <212> ADN
- 15 <213> Escherichia coli
- <400> 11

aagcttgcac gcctgcaggt ccaggccat atctaactctt acctcgactg ctgtatataa 60  
aaccagtggt tatatgtaca gtactgctgt atataaaacc agtggttata tgtacagtac 120  
gtcgcactgct gtatataaaa ccagtggtta tatgtacagt actgctgtat ataaaaccag 180  
tggttatatg tacagtacgt cgactc 206

<210> 12  
<211> 30  
<212> ADN  
5 <213> Secuencia artificial

<220>  
<223> Sintético TATAA

10 <400> 12 tagagggtat ataatggatc cccgggtacc 30

<210> 13  
<211> 1705  
15 <212> ADN  
<213> Photinus pyralis

<400> 13  
atggaagacg ccaaaaacat aaagaaagge ccggcgccat tctatcctct agaggatgga 60  
accgctggag agcaactgca taaggctatg aagagatag ccctggttcc tggaacaatt 120  
gcttttacag atgcacatat cgagggtgac atcacgtacg cggaatactt cgaaatgtcc 180  
gttcggttgg cagaagctat gaaacgatat gggctgaata caaatcacag aatcgtcgta 240  
tgcagtgaaa actctcttca attctttatg ccgggtgttg gcgcgttatt ttcggagtt 300  
gcagttgcgc ccgcgaacga catttataat gaacgtgaat tgctcaacag tatgaacatt 360  
tcgcagccta ccgtagtggt tgtttccaaa aaggggttgc aaaaaatctt gaacgtgcaa 420  
aaaaattac caataatcca gaaaattatt atcatggatt ctaaaacgga ttaccagggga 480  
tttcagtcga tgtacacggt cgtcacatct catctacctc ccgggtttta tgaatacgat 540  
tttgtaccag agtcctttga tcgtgacaaa acaattgcac tgataatgaa ttcctctgga 600  
tctactgggt tacctaaggg tgtggccctt ccgcatagaa ctgcctgcgt cagattctcg 660  
catgccagag atcctatctt tgcaatcaa atcattccgg aactgcgat ttaagtgtt 720  
gttcatttcc atcacggttt tggaatggtt actacactcg gatatttgat atgtggattt 780  
cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatccct tcaggattac 840  
aaaaatcaaa gtgcgttgct agtaccaccc ctatcttcat tcttcgcca aagcactctg 900  
attgacaaat acgatttate taatttacac gaaattgctt ctgggggccc acctctttcg 960  
aaagaagtcg gggaagcggg tgcaaacgc ttccatcttc cagggatagc acaaggatat 1020  
gggctcactg agactacatc agctattctg attacacccg agggggatga taaaccgggc 1080  
gcggtcggta aagttgttcc attttttgaa gcgaaggttg tggatctgga taccgggaaa 1140  
acgctgggag ttaatcagag aggcgaatta tgtgtcagag gacctatgat tatgtccggt 1200  
tatgtaaaca atccggaagc gaccaacgcc ttgattgaca aggatggatg gctacattct 1260  
ggagacatag cttactggga cgaagacgaa cacttcttca tagttgaccg cttgaagtct 1320  
ttaatthaaat acaaagata tcaggtggcc cccgctgaat tggaatcgat attgttacia 1380  
caccccaaca tcttcgacgc gggcgtggca ggtcttccc acgatgacgc cgttgaactt 1440  
cccgccgccc ttgttgtttt ggagcacgga aagacgatga cggaaaaaga gatcgtggat 1500  
tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagtgtt gtttgtggac 1560  
gaagtaccga aaggtcttac cgaaaactc gacgcaagaa aaatcagaga gatcctcata 1620  
aaggccaaga agggcgaaa gtccaaattg taaaatgtaa ctgtattcag cgatgacgaa 1680  
attcttagct attgtaatac tctag 1705

<210> 14  
 <211> 1073  
 <212> ADN  
 <213> Choristoneura fumiferana

5

<400> 14

```

cctgagtgcg tagtaccgga gactcagtcg gccatgaagc ggaagagaa gaaagcacag    60
aaggagaagg acaaactgcc tgtcagcacg acgacggcgg acgaccacat gccgcccatt    120
atgcagtggt aacctccacc tcctgaagca gcaaggattc acgaagtggg tccaaggttt    180
ctctccgaca agctgttggg gacaaaccgg cagaaaaaca tccccagtt gacagccaac    240
cagcagttcc ttatcgccag gctcatctgg taccaggacg ggtacgagca gccttctgat    300
gaagatttga agaggattac gcagacgtgg cagcaagcgg acgatgaaaa cgaagagtct    360
gacactccct tccgccagat cacagagatg actatcctca cggccaact tatcgtggag    420
ttegcgaagg gattgccagg gttcgccaag atctcgacg ctgatcaaat tacgctgctt    480
aaggcttgct caagtgaggt aatgatgctc cgagtcgcca gatacagatg ggctcagac    540
agtgttctgt tcgcgaacaa ccaagcgtac actcgcgaca actaccgcaa ggctggcatg    600
gcctacgtca tcgaggatct actgcacttc tgccggtgca tgtactctat ggctgtggac    660
aacatccatt acgcgctgct cacggctgtc gtcactcttt ctgaccggcc aggggtggag    720
cagccgcaac tggtggaaga aatccagcgg tactacctga atacgctccg catctatctc    780
ctgaaccagc tgagcggggtc ggcgcttcg tccgcatat acggcaagat cctctcaatc    840
ctctctgagc tacgcacgct cggcatgcaa aactccaaca tgtgcatctc cctcaagctc    900
aagaacagaa agctgccgcc tttcctcgag gagatctggg atgtggcagg acatgtcgca    960
cacccaaccg ccgcctatct cgagtccccc acgaatctct agcccctgcg cgcacgcatc   1020
gccgatgccg cgtccggccg cgctgctctg agaattcgat atcaagcttc tag           1073
    
```

<210> 15  
 <211> 1109  
 <212> ADN  
 <213> Nephrotetix cincticeps

10

<400> 15

```

caggaggagc tctgcctggt gtgcggagac cgagcgtcgg gataccacta caacgctctc    60
acctgcgaag gatgcaaggg cttctttcgg aggagtatca ccaaaaacgc agtgtaccag    120
tccaaatacg gcaccaattg tgaatatagc atgtatatgc ggcgcaagtg ccaggagtgc    180
cgactcaaga agtgcctcag tgtagggatg aggccagaat gtgtagtacc tgagtatcaa    240
tgtgccgtaa aaaggaaaga gaaaaaagct caaaaaggaca aagataaacc tgtctcttca    300
accaatggct cgctgaaat gagaatagac caggacaacc gttgtgtggt gttgcagagt    360
gaagacaaca ggtacaactc gactacgccc agtttcggag tcaaaccctc cagtccagaa    420
caagaggagc tcatccacag gctcgtctac ttccagaacg agtacgaaca ccctgccgag    480
gaggatctca agcggatcga gaacctcccc tgtgacgacg atgacctgtg tgatgttcgc    540
tacaacacac ttacggagat cacaatactc acagtccagc tcatcgtgga gtttgcaaaa    600
aaactgcctg gtttcgacaa actactgaga gaggaccaga tcgtgttgct caaggcgtgt    660
tcgagcgagg tgatgatgct gcggatggcg cggaggtacg acgtccagac agactcgatc    720
ctgttcgcca acaaccagcc gtacacgcca gagtcgtaca cgatggcagg cgtgggggaa    780
    
```

ES 2 364 923 T3

gtcatcgaag atctgctgcg gtccggccga ctcatgtgct ccatgaaggt ggacaatgcc 840  
 gagtatgctc tgctcacggc catcgtcacc ttctccgagc ggccgaacct ggcggaagga 900  
 tggaaagtgg agaagatcca ggagatctac ctggaggcgc tcaagtccta cgtggacaac 960  
 cgagtgaac ctccgagtc gaccatcttc gccaaactgc tctccgttct caccgagctg 1020  
 cgaacactcg gcaaccagaa ctccgagatg tgcttctcgt taaactacgc aaccgcaaac 1080  
 atgccaccgt tcctcgaaga aatctggga 1109

<210> 16  
 <211> 714  
 <212> ADN

5 <213> Mus musculus

<400> 16  
 gccaacgagg acatgcctgt agagaagatt ctggaagccg agcttgctgt cgagcccaag 60  
 actgagacat acgtggaggc aaacatgggg ctgaacccca gctcaccaaa tgaccctgtt 120  
 accaacatct gtcaagcagc agacaagcag ctcttcactc ttgtggagtg ggccaagagg 180  
 atcccacact tttctgagct gcccctagac gaccaggta tcctgctacg ggaggctgg 240  
 aacgagctgc tgatgcctc cttctcccac cgctccatag ctgtgaaaga tgggattctc 300  
 ctggccaccg gcctgcacgt acaccggaac agcgcacaca gtgctggggg gggcgccatc 360  
 tttgacaggg tgctaacaga gctggtgtct aagatgcgtg acatgcagat ggacaagacg 420  
 gagctgggct gcctgcgagc cattgtcctg ttcaaccctg actctaaggg gctctcaaac 480  
 cctgctgagg tggaggcgtt gagggagaag gtgtatgctg cactagaagc gtactgcaaa 540  
 cacaagtacc ctgagcagcc gggcaggttt gccaaactgc tgctccgctt gcctgcaactg 600  
 cgttccatcg ggctcaagtg cctggagcac ctgttcttct tcaagctcat cggggacacg 660  
 cccatcgaca ccttcctcat ggagatgctg gaggcaccac atcaagccac ctac 714

## REIVINDICACIONES

1. Sistema de regulación de múltiples genes inducibles, que comprende una pluralidad de sistemas de regulación génica individualmente operables, en el que:

- 5 a) cada sistema de regulación génica individualmente operable comprende:
- i) uno o más polinucleótidos codificantes de un complejo de receptores que comprende:
- 10 A) un dominio de unión a ADN,  
B) un primer dominio de unión a ligando de receptor nuclear y un segundo dominio de unión a ligando de receptor nuclear, y  
C) un dominio de transactivación,
- 15 ii) un ligando,
- iii) un polinucleótido que comprende:
- 20 A) un polinucleótido exógeno o endógeno, y  
B) un elemento de respuesta,

en el que:

- 25 A) el polinucleótido exógeno o endógeno se encuentra funcionalmente ligado al elemento de respuesta, y  
B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o en ausencia del ligando resulta en la activación o en la supresión del polinucleótido exógeno o endógeno, y  
C) uno de los dominios de unión a ligando de receptor nuclear es un dominio de unión a ligando de receptor nuclear de grupo H y el otro dominio de unión a ligando de receptor nuclear es un dominio de unión a ligando de receptor nuclear capaz de formar un dímero con el dominio de unión a ligando de receptor nuclear de grupo H, y
- 30

b) cada sistema de regulación génica individualmente operable es ortogonal respecto a los demás sistemas de regulación génica individualmente operables presentes en el sistema de regulación de múltiples genes inducibles.

35 2. Sistema de regulación de múltiples genes inducibles según la reivindicación 1, en el que cada sistema de regulación génica individualmente operable comprende:

- 40 i) un primer casete de expresión génica que comprende un polinucleótido que codifica un primer polipéptido que comprende un dominio de unión a ADN que reconoce un elemento de respuesta asociado a un gen cuya expresión debe modularse y un primer dominio de unión a ligando de receptor nuclear,
- 45 ii) un segundo casete de expresión génica que comprende un polinucleótido que codifica un segundo polipéptido que comprende un dominio de transactivación y un segundo dominio de unión a ligando de receptor nuclear,
- iii) un ligando, y
- 50 iv) un tercer casete de expresión génica que comprende: A) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido del primer casete de expresión génica, B) un promotor que resulta activado por el dominio de transactivación del segundo polipéptido del segundo casete de expresión génica, y C) un gen cuya expresión debe modularse.

3. Sistema de regulación de múltiples genes inducibles según la reivindicación 1 ó 2, en el que el dominio de unión a ligando de receptor nuclear de grupo H es un dominio de unión a ligando de receptor de la ecdisona y el otro dominio de unión a ligando de receptor nuclear capaz de formar un dímero con el dominio de unión a ligando de receptor nuclear de grupo H se selecciona de entre el grupo constituido por un dominio de unión a ligando de receptor de la ecdisona, un dominio de unión a ligando del receptor X del retinoide de vertebrado, un dominio de unión a ligando del receptor X del retinoide de invertebrado, un dominio de unión a ligando de proteína Ultraspiracle, y un dominio de unión a ligando quimérico, que comprende por lo menos dos fragmentos polipeptídicos diferentes de dominio de unión a ligando de receptor nuclear seleccionados de entre el grupo constituido por un dominio de unión a ligando de receptor X del retinoide de vertebrado, un dominio de unión a ligando del receptor X del retinoide de invertebrado, y un dominio de unión a ligando de la proteína Ultraspiracle.

60

4. Virus que comprende el sistema de regulación de múltiples genes según la reivindicación 1 ó 2.

65 5. Virus según la reivindicación 4, en el que el virus es el virus Vaccinia, un adenovirus o un baculovirus.

6. Virus según la reivindicación 5, en el que dicho virus es un adenovirus.
7. Célula que comprende el sistema de regulación de múltiples genes inducibles según cualquiera de las reivindicaciones 1 a 3.
8. Célula según la reivindicación 7, en la que dicha célula es una célula vegetal.
9. Célula según la reivindicación 8, en la que la célula es una célula vegetal seleccionada de entre el grupo constituido por manzana, *Arabidopsis*, bajra, plátano, cebada, alubias, remolacha, soja verde, garbanzo, chili, pepino, berenjena, habas verdes, maíz, melón, mijo, alubia, avena, okra, *Panicum*, papaya, cacahuate, guisante, pimiento, guisante de paloma, piña, *Phaseolus*, patata, calabaza, arroz, sorgo, soja, calabacín, caña de azúcar, remolacha azucarera, girasol, batata, té, tomate, tabaco, sandía y trigo.
10. Organismo transgénico no humano que comprende una o más células según cualquiera de las reivindicaciones 7 a 9.
11. Sistema de regulación de múltiples genes inducibles según la reivindicación 1 ó 2, en el que uno o más de los polinucleótidos codificantes de un complejo de receptores codifica un complejo de receptores no de mamífero.
12. Sistema de regulación de múltiples genes inducibles según cualquiera de las reivindicaciones 1 a 3 ó 11, en el que el dominio de unión a ligando de receptor nuclear de grupo H es un dominio de unión a ligando de receptor de la ecdisona.
13. Procedimiento para desarrollar un sistema de regulación de múltiples genes inducibles según cualquiera de las reivindicaciones 1 a 3, que comprende las etapas siguientes:
- a) definir un conjunto de ligandos modificados diversamente basado en cambios graduales de elemento farmacóforo,
  - b) preparar un primer conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ligando que es:
    - i) natural,
    - ii) modificado mediante delección, inserción o mutación,
    - iii) quimérico,
    - iv) sintético, o
    - v) una combinación de los mismos,
  - c) opcionalmente, introducir los polipéptidos receptores nucleares en una célula,
  - d) realizar una búsqueda entre los polipéptidos receptores nucleares en la célula de c) o *in vitro*, utilizando el conjunto de ligandos diversamente modificados para la modulación génica, o la unión a dominio de unión de ligando, o ambos,
  - e) determinar la ortogonalidad de las combinaciones de polipéptido receptor nuclear/ligando con el fin de definir un subconjunto de ligandos con diversas propiedades de modulación génica,
  - f) repetir opcionalmente las etapas a) a e) utilizando un conjunto modificado de ligandos y un conjunto modificado de polipéptidos receptores nucleares;
  - g) preparar un segundo conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ADN que es:
    - i) natural,
    - ii) modificado mediante delección, inserción o mutación,
    - iii) quimérico,
    - iv) sintético, o
    - v) una combinación de los mismos,
  - h) preparar un conjunto de constructos de ADN que comprende un gen exógeno o endógeno y elementos de respuesta que son:
    - i) naturales,
    - ii) modificados mediante delección, inserción o mutación,
    - iii) quiméricos,

- iv) sintéticos, o
- v) una combinación de los mismos,

5 i) clonar el primer conjunto de polipéptidos receptores nucleares y el segundo conjunto de polipéptidos receptores nucleares y los constructos de ADN y a continuación, introducir los clones en una célula,

j) someter a ensayo la célula para correactividad con los ligandos identificados en la etapa e), y,

10 k) seleccionar un conjunto ortogonal de ligandos, dominios de unión a ligando, dominios de unión a ADN, y elementos de respuesta, basándose en los resultados de las etapas e) y j), de manera que comprenda el sistema de regulación de múltiples genes inducibles según cualquiera de las reivindicaciones 1 a 3.

14. Sistema de regulación de múltiples genes inducibles que comprende una pluralidad de sistemas de regulación génica individualmente operables, en el que:

15 a) cada sistema de regulación génica individualmente operable comprende:

i) uno o más polinucleótidos codificantes de un complejo de receptores que comprende:

- 20 A) un dominio de unión a ADN,  
B) un dominio de unión a ligando, y  
C) un dominio de transactivación,

25 ii) un ligando,  
iii) un polinucleótido que comprende:

- A) un polinucleótido exógeno o endógeno, y  
B) un elemento de respuesta,

30 en el que:

- 35 A) el polinucleótido exógeno o endógeno se encuentra funcionalmente ligado al elemento de respuesta, y  
B) la unión del dominio de unión a ADN al elemento de respuesta en presencia o en ausencia del ligando resulta en la activación o la supresión del polinucleótido exógeno o endógeno, y  
C) el complejo de receptores es un complejo de receptores nucleares de grupo H, y

40 b) cada sistema de regulación génica individualmente operable es ortogonal respecto a los demás sistemas de regulación génica individualmente operables presentes en el sistema de regulación de múltiples genes inducibles.

15. Sistema de regulación de múltiples genes inducibles según la reivindicación 14, en el que el complejo de receptores nucleares de grupo H es un complejo de receptores de la ecdisona.

45 16. Procedimiento para desarrollar un sistema de regulación de múltiples genes según la reivindicación 14 ó 15, que comprende las etapas de:

a) definir un conjunto de ligandos modificados de modo diverso, basándose en cambios graduales de elementos farmacóforos,

50 b) preparar un primer conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ligando que es:

- 55 i) natural  
ii) modificado mediante delección, inserción o mutación,  
iii) quimérico,  
iv) sintético, o  
v) una combinación de los mismos,

60 c) opcionalmente, introducir los polipéptidos receptores nucleares en una célula,

d) realizar una búsqueda de los polipéptidos receptores nucleares en la célula de c) o *in vitro*, utilizando el conjunto de ligandos diversamente modificados para la modulación génica, o la unión a dominio de unión a ligando, o ambos,

65 e) determinar la ortogonalidad de las combinaciones de polipéptido receptor nuclear/ligando con el fin de definir un subconjunto de ligandos con propiedades de modulación génica diversas,

f) opcionalmente repetir las etapas a) a e) utilizando un conjunto modificado de ligandos y un conjunto modificado de polipéptidos receptores nucleares,

5 g) preparar un segundo conjunto de polipéptidos receptores nucleares, en el que cada polipéptido receptor nuclear comprende un dominio de unión a ADN que es:

- 10 i) natural,  
ii) modificado mediante delección, inserción o mutación,  
iii) quimérico,  
iv) sintético, o  
v) una combinación de los anteriores,

15 h) preparar un conjunto de constructos de ADN que comprende un gen exógeno o endógeno y elementos de respuesta que son:

- 20 i) naturales,  
ii) modificados mediante delección, inserción o mutación,  
iii) quiméricos,  
iv) sintéticos, o  
v) una combinación de los mismos,

25 i) clonar el primer conjunto de polipéptidos receptores nucleares y el segundo conjunto de polipéptidos receptores nucleares y los constructos de ADN y después introducir los clones en una célula,

j) someter a ensayo la célula para correactividad con los ligandos identificados en la etapa e), y

30 k) seleccionar un conjunto ortogonal de ligandos, dominios de unión a ligando, dominios de unión a ADN, y elementos de respuesta basándose en los resultados de las etapas e) y j), de manera que comprenda el sistema de regulación de múltiples genes inducibles según la reivindicación 14 ó 15.

17. Vector de expresión que comprende el sistema de regulación de múltiples genes inducibles según la reivindicación 1 ó 2.

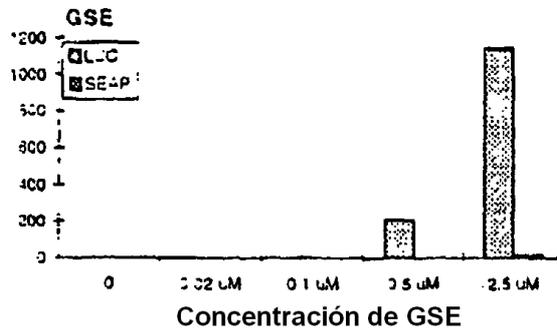


Figura 1A

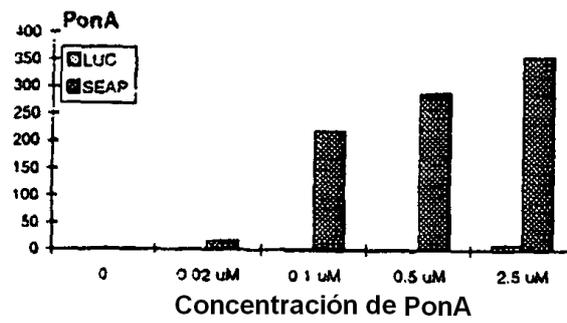


Figura 1B

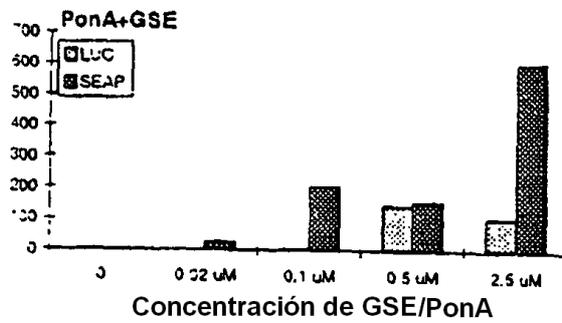


Figura 1C

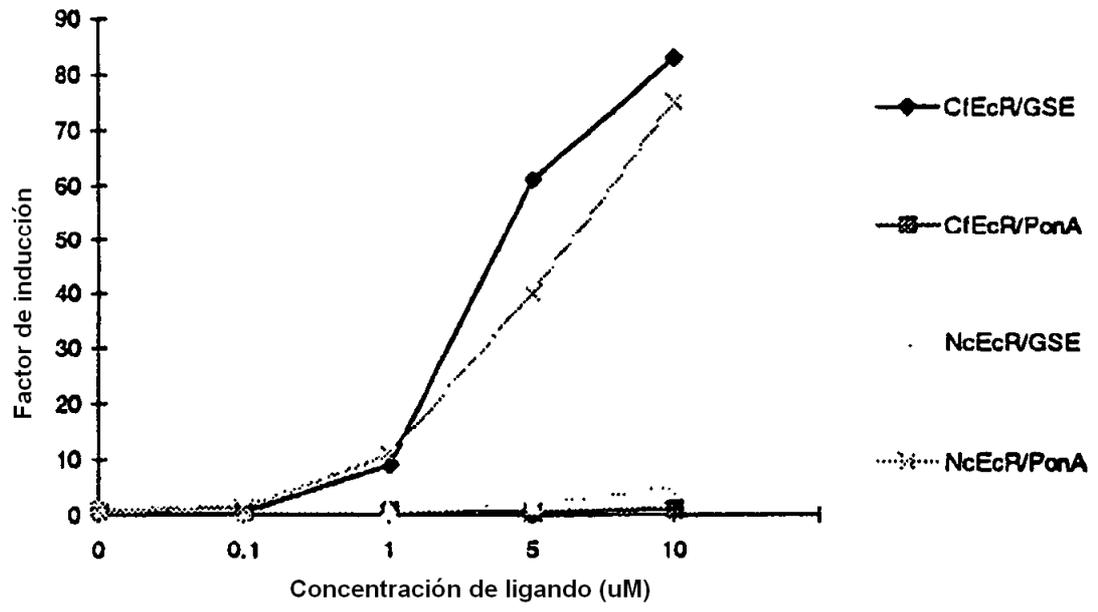


Figura 2