

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 392 508**

51 Int. Cl.:

C12N 15/63 (2006.01)
C12N 15/85 (2006.01)
C12N 15/12 (2006.01)
C12N 15/62 (2006.01)
C07K 14/705 (2006.01)
C07K 14/72 (2006.01)
A01H 5/00 (2006.01)
C12Q 1/68 (2006.01)
G01N 33/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **02714990 .5**
96 Fecha de presentación: **20.02.2002**
97 Número de publicación de la solicitud: **1572862**
97 Fecha de publicación de la solicitud: **14.09.2005**

54

Título: **Receptores X retinoides quiméricos y su uso en un sistema inducible de expresión génica basado en receptores de ecdisona novedoso**

30

Prioridad:

20.02.2001 US 269799 P
31.05.2001 US 294814 P
31.05.2001 US 294819 P

45

Fecha de publicación de la mención BOPI:

11.12.2012

45

Fecha de la publicación del folleto de la patente:

11.12.2012

73

Titular/es:

INTREXON CORPORATION (100.0%)
1750 Kraft Drive, Suite 1400
Blacksburg, VA 24060, US

72

Inventor/es:

PALLI, SUBBA REDDY y
KAPITSKAYA, MARIANNA ZINOVJEVNA

74

Agente/Representante:

PÉREZ BARQUÍN, Eliana

ES 2 392 508 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

Receptores X retinoides quiméricos y su uso en un sistema inducible de expresión génica basado en receptores de ecdisona novedoso

5

Campo de la invención

La presente invención se refiere al campo de la biotecnología o ingeniería genética. Específicamente, la presente invención se refiere al campo de la expresión génica. Más específicamente, la presente invención se refiere a un sistema inducible de expresión génica basado en receptores X retinoides quiméricos/receptores de ecdisona novedoso y a métodos de modulación de la expresión de un gen en una célula huésped usando este sistema inducible de expresión génica.

10

Antecedentes de la invención

15

La mención de cualquier referencia en el presente documento no debe considerarse una admisión de que dicha referencia está disponible como "técnica anterior" de la presente solicitud.

20

En el campo de la ingeniería genética, el control preciso de la expresión génica es una herramienta útil para el estudio, la manipulación y el control del desarrollo y otros procesos fisiológicos. La expresión génica es un proceso biológico complejo que implica una serie de interacciones proteína-proteína específicas. Para desencadenar la expresión génica, de forma que produzca el ARN necesario como la primera etapa de la síntesis de proteínas, debe aproximarse un activador de la transcripción a un promotor que controle la transcripción génica. Normalmente, el propio activador de la transcripción está asociado con una proteína que tiene al menos un dominio de unión a ADN que se une a sitios de unión a ADN presentes en las regiones promotoras de genes. Por tanto, para que se produzca la expresión génica, una proteína que comprenda un dominio de unión a ADN y un dominio de transactivación a una distancia apropiada del dominio de unión a ADN debe situarse en la posición correcta en la región promotora del gen.

25

30

El enfoque transgénico tradicional utiliza un promotor específico de célula para dirigir la expresión de un transgén diseñado. En primer lugar, se incorpora en un genoma huésped una construcción de ADN que contiene el transgén. Cuando se desencadena por un activador de la transcripción, se produce la expresión del transgén en un tipo celular determinado.

35

Otro medio para regular la expresión de genes exógenos en células es por medio de promotores inducibles. Los ejemplos del uso de estos promotores inducibles incluyen el promotor PR1-a, sistemas procariotas de represor-operador, sistemas de inmunofilina inmunosupresores y sistemas de activación de la transcripción de eucariotas superiores, tales como sistemas de receptores de hormonas esteroideas, y se describen a continuación.

40

El promotor PR1-a del tabaco se induce durante la respuesta sistémica de resistencia adquirida tras un ataque patógeno. El uso de PR1-a puede ser limitado porque, frecuentemente, responde a materiales endógenos y 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.*, 1986, Proc. Natl. Acad. Sci. USA 83: 5414-5418; Arnheiter *et al.*, 1990, Cell 62: 51-61; Filmus *et al.*, 1992, Nucleic Acids Research 20: 27550-27560). Sin embargo, estos sistemas presentan limitaciones debido a su efecto sobre la expresión de genes que no son su objetivo. Estos sistemas también son poco consistentes.

45

50

Los sistemas procariotas de represor-operador utilizan proteínas represoras bacterianas y las secuencias de ADN de operador único a las que se unen. Tanto el sistema de represor-operador de tetraciclina ("Tet") como el de lactosa ("Lac") de la bacteria *Escherichia coli* se han usado en plantas y animales para controlar la expresión génica. En el sistema Tet, la tetraciclina se une a la proteína represora TetR, dando lugar a un cambio conformacional que libera la proteína represora del operador, que como consecuencia, permite que se produzca la transcripción. En el sistema Lac, se activa un operón lac en respuesta a la presencia de lactosa o análogos sintéticos tales como isopropil-b-D-tiogalactósido. Desafortunadamente, el uso de estos sistemas está restringido por la química inestable de los ligandos, es decir, tetraciclina y lactosa, su toxicidad, su presencia natural o los niveles relativamente altos necesarios para la inducción o la represión. Por motivos similares, la utilidad en animales de estos sistemas es limitada.

55

60

Las moléculas inmunosupresoras tales como FK506, rapamicina y ciclosporina A se pueden unir a inmunofilinas FKBP12, ciclofilina, etc. Usando esta información se ha ideado una estrategia general para juntar dos proteínas cualesquiera simplemente situando FK506 sobre cada una de las dos proteínas o situando FK506 sobre una y ciclosporina A sobre la otra. Después, se puede usar un homodímero sintético de FK506 (FK1012) o un compuesto resultante de la fusión de FK506-ciclosporina (FKCsA) para inducir la dimerización de estas moléculas (Spencer *et al.*, 1993, *Science* 262:1019-24; Belshaw *et al.*, 1996, *Proc Natl Acad Sci USA* 93:4604-7). El dominio de unión a ADN de Gal4 fusionado con FKBP12 y el dominio activador VP16 fusionado con ciclofilina, y el compuesto FKCsA se usaron para mostrar la heterodimerización y activación de un gen indicador bajo el control de un promotor que

65

contiene sitios de unión a Gal4. Desafortunadamente, este sistema incluye inmunosupresores que pueden tener efectos secundarios no deseados y, por lo tanto, limita su uso para diversas aplicaciones de interruptores génicos de mamífero.

- 5 También se han empleado sistemas de activación de la transcripción de eucariotas superiores, tales como sistemas de receptores de hormonas esteroideas. Los receptores de hormonas esteroideas son miembros de la superfamilia de receptores nucleares y se encuentran en células de vertebrado e invertebrado. Desafortunadamente, el uso de compuestos esteroideos que activan los receptores para la regulación de la expresión génica, en particular en plantas y mamíferos, está limitado debido a su implicación en muchas otras rutas biológicas naturales en estos organismos. Para superar estas dificultades, se ha desarrollado un sistema alternativo usando receptores de ecdisona (EcR) de insecto.

15 El crecimiento, la muda y el desarrollo de los insectos están regulados por la hormona esteroidea ecdisona (hormona de la muda) y las hormonas juveniles (Dhadialla, *et al.*, 1998, Annu. Rev. Entomol. 43: 545-569). El objetivo molecular de la ecdisona en insectos consiste en al menos un receptor de ecdisona (EcR) y una proteína ultraespiráculo (USP). El EcR es un miembro de la superfamilia de receptores esteroideos nucleares que se caracteriza por dominios de unión a ADN distintivo y a ligando y un dominio de activación (Koelle *et al.* 1991, Cell, 67:59-77). Los receptores EcR son sensibles a una serie de compuestos esteroideos tales como ponasterona A y muristerona A. Recientemente, se han descrito compuestos no esteroideos con actividad ecdisteroide agonista, incluidos los insecticidas comercialmente disponibles tebufenozida y metoxifenozida, que se comercializan en todo el mundo por Rohm and Haas Company (véase la solicitud de patente internacional n.º PCT/EP96/00686 y la patente de EE. UU. 5.530.028). Ambos análogos tienen perfiles de seguridad excepcionales para otros organismos.

25 Las solicitudes de patente internacional N.º PCT/US97/05330 (documento WO 97/38117) y PCT/US99/08381 (documento WO 99/58155) divulgan métodos para modular la expresión de un gen exógeno en los que una construcción de ADN que comprende el gen exógeno y un elemento de respuesta a ecdisona se activa por una segunda construcción de ADN que comprende un receptor de ecdisona que, en presencia de un ligando para ello, y opcionalmente en presencia de un receptor que puede actuar como compañero silencioso, se une al elemento de respuesta a ecdisona para inducir la expresión génica. El receptor de ecdisona elegido se aisló a partir de *Drosophila melanogaster*. Normalmente, estos sistemas requieren la presencia de un compañero silencioso, preferentemente un receptor X retinoide (RXR), para proporcionar una activación óptima. En células de mamífero, el receptor de ecdisona de insecto (EcR) forma un heterodímero con el receptor X retinoide (RXR) y regula la expresión de genes objetivo de manera dependiente de ligando. La solicitud de patente internacional N.º PCT/US98/14215 (documento WO 99/02683) divulga que el receptor de ecdisona aislado a partir de la polilla de seda *Bombyx mori* es funcional en sistemas de mamífero sin necesidad de un compañero de dímero exógeno.

40 La patente de EE. UU. N.º 5.880.333 divulga un sistema de heterodímero de ultraespiráculo (USP) y EcR de *Drosophila melanogaster* usado en plantas en las que el dominio de transactivación y el dominio de unión a ADN están situados en dos proteínas híbridas diferentes. Desafortunadamente, este sistema no es eficaz para inducir la expresión de genes indicadores en células animales (para comparación, véase el ejemplo 1.2, más adelante).

45 En cada uno de estos casos, el dominio de transactivación y el dominio de unión de ADN (bien como EcR nativo como en la solicitud de patente internacional n.º PCT/US98/14215 o bien como EcR modificado como en la solicitud de patente internacional N.º PCT/US97/05330) se incorporaron en una sola molécula y los demás compañeros heterodiméricos, USP o RXR, se usaron en su estado nativo.

50 Las desventajas de los sistemas de regulación génica basados en EcR descritos anteriormente incluyen una considerable actividad de fondo en ausencia de ligandos y la imposibilidad de aplicar estos sistemas para su uso tanto en plantas como en animales (véase la patente de EE. UU. N.º 5.880.333). Para la mayoría de las aplicaciones que se basan en la modulación de la expresión génica, no son deseables estos sistemas basados en EcR. Por lo tanto, existe una necesidad en la técnica de proporcionar sistemas mejorados para modular de forma precisa la expresión de genes exógenos tanto en plantas como en animales. Estos sistemas mejorados serían útiles para aplicaciones tales como tratamiento génico, producción de proteínas y anticuerpos a gran escala, ensayos de rastreo de alto rendimiento basados en células, genómica funcional y regulación de rasgos en animales transgénicos. Los sistemas mejorados simples, compactos y dependientes de ligandos que son relativamente baratos, fácilmente disponibles y de baja toxicidad para el huésped, resultarían útiles para regular sistemas biológicos.

60 Recientemente, los solicitantes han mostrado que un sistema inducible de expresión génica basado en receptores de ecdisona en el que los dominios de transactivación y de unión a ADN se separan entre sí situándolos en dos proteínas diferentes, da lugar a una actividad de fondo enormemente reducida en ausencia de un ligando y significativamente aumentada con respecto al fondo en presencia de un ligando (en la solicitud en trámite PCT/US01/09050). Este sistema de dos híbridos es un sistema inducible de expresión génica significativamente mejorado en comparación con los dos sistemas divulgados en las solicitudes PCT/US97/05330 y PCT/US98/14215.

65 Los solicitantes demostraron anteriormente que un sistema de expresión génica basado en receptores de ecdisona en asociación con una proteína ultraespiráculo (USP) de díptero (*Drosophila melanogaster*) o de lepidóptero

(*Choristoneura fumiferana*) se expresa de forma constitutiva en células de mamífero, mientras que un sistema de expresión génica basado en receptores de ecdisona en asociación con un receptor X retinoide (RXR) de vertebrado es inducible en células de mamífero (en la solicitud en trámite PCT/US01/09050). Recientemente, los solicitantes realizaron el sorprendente descubrimiento de que un RXR de invertebrado distinto de díptero y de lepidóptero puede funcionar de forma similar al RXR de vertebrado en un sistema inducible de expresión génica basado en receptores de ecdisona (en la solicitud de EE. UU. presentada junto al presente documento).

Los solicitantes han mostrado ahora que un dominio de unión a ligando de RXR quimérico, que comprende al menos dos fragmentos polipeptídicos, en el que el primer fragmento polipeptídico es de RXR de una especie de vertebrado/invertebrado y el segundo fragmento polipeptídico es de RXR de una especie diferente de vertebrado/invertebrado, de modo que se produce un dominio de unión a ligando de RXR quimérico de vertebrado/invertebrado, un dominio de unión a ligando de RXR quimérico de vertebrado/vertebrado o un dominio de unión a ligando de RXR quimérico de invertebrado/invertebrado, puede funcionar de forma similar a o mejor que el RXR de vertebrado original o bien el RXR de invertebrado original en un sistema inducible de expresión génica basado en receptores de ecdisona. Como se describe en el presente documento, el sistema inducible de expresión génica basado en receptores X retinoides quiméricos/receptores de ecdisona proporciona un sistema inducible de expresión génica en células de bacterias, hongos, levaduras, animales y mamíferos que se caracteriza por un aumento de la sensibilidad a ligando y de la magnitud de la transactivación.

Sumario de la invención

La presente invención se refiere a un sistema inducible de expresión génica basado en receptores X retinoides quiméricos/receptores de ecdisona novedoso, polinucleótidos y polipéptidos de receptores quiméricos novedosos para su uso en el sistema inducible de expresión génica novedoso y métodos de modulación de la expresión de un gen en una célula huésped usando este sistema inducible de expresión génica. En particular, la invención de los solicitantes se refiere a un sistema novedoso de modulación de la expresión génica que comprende un polinucleótido que codifica un dominio de unión a ligando de RXR quimérico (LBD).

Específicamente, la presente invención se refiere a un sistema de modulación de la expresión génica que comprende: a) un primer casete de expresión génica que se puede expresar en una célula huésped que comprende un polinucleótido que codifica un primer polipéptido híbrido que comprende: i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y ii) un dominio de unión a ligando de receptor de ecdisona; y b) un segundo casete de expresión génica que se puede expresar en la célula huésped que comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende: i) un dominio de transactivación; y ii) un dominio de unión a ligando de receptor X retinoide quimérico.

La presente invención también se refiere a un sistema de modulación de la expresión génica que comprende: a) un primer casete de expresión génica que se puede expresar en una célula huésped que comprende un polinucleótido que codifica un primer polipéptido híbrido que comprende: i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y ii) un dominio de unión a ligando de receptor X retinoide quimérico; y b) un segundo casete de expresión génica que se puede expresar en la célula huésped que comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende: i) un dominio de transactivación; y ii) un dominio de unión a ligando de receptor de ecdisona.

La presente invención también se refiere a un sistema de modulación de la expresión génica de acuerdo con la invención que además comprende c) un tercer casete de expresión génica que comprende: i) un elemento de respuesta al que se une el dominio de unión a ADN del primer polipéptido híbrido; ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y iii) un gen cuya expresión se quiere modular.

La presente invención también se refiere a un casete de expresión génica que se puede expresar en una célula huésped, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende, bien i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular o bien ii) un dominio de transactivación; y un dominio de unión a ligando de receptor X retinoide quimérico.

La presente invención también se refiere a un polinucleótido aislado que codifica un polipéptido híbrido que comprende, o bien i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular o bien ii) un dominio de transactivación; y un dominio de unión a ligando de receptor X retinoide quimérico de vertebrado e invertebrado. La presente invención también se refiere a un polipéptido híbrido aislado codificado por el polinucleótido de acuerdo con la invención.

La presente invención también se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico. En una realización específica, el polinucleótido aislado codifica un LBD truncado de RXR quimérico, en el que la mutación por truncamiento afecta a la actividad de unión a ligando o a la sensibilidad a ligando del LBD de RXR quimérico. En otra realización específica, el polinucleótido aislado codifica un polipéptido truncado de RXR quimérico que comprende una mutación por truncamiento que aumenta la sensibilidad a ligando de un heterodímero

que comprende el polipéptido truncado de RXR quimérico y un compañero de dimerización. En una realización específica, el compañero de dimerización es un polipéptido de receptor de ecdisona.

5 La presente invención también se refiere a un polipéptido aislado codificado por un polinucleótido de acuerdo con la invención de los solicitantes.

10 La presente invención también se refiere a un polipéptido híbrido que comprende, o bien i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular o bien ii) un dominio de transactivación; y un dominio de unión a ligando de receptor X retinoide quimérico.

15 La presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento, en el que el LBD truncado de RXR quimérico está codificado por un polinucleótido de acuerdo con la invención.

20 Por tanto, la presente invención también se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que afecta a la actividad de unión a ligando o a la sensibilidad a ligando de dicho LBD truncado de RXR quimérico.

25 La presente invención también se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que aumenta la sensibilidad a ligando de un heterodímero que comprende el LBD truncado de RXR quimérico y un compañero de dimerización. En una realización específica, el compañero de dimerización es un polipéptido de receptor de ecdisona.

30 La invención de los solicitantes también se refiere a métodos de modulación de la expresión génica en una célula huésped usando un sistema de modulación de la expresión génica de acuerdo con la invención. Específicamente, la invención de los solicitantes proporciona un método de modulación de la expresión de un gen en una célula huésped que comprende las etapas de: a) introducir en la célula huésped un sistema de modulación de la expresión génica de acuerdo con la invención; b) introducir en la célula huésped un casete de expresión génica que comprende i) un elemento de respuesta que comprende un dominio reconocido por el dominio de unión a ADN del primer polipéptido híbrido; ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y iii) un gen cuya expresión se quiere modular; y c) introducir en la célula huésped un ligando; de modo que, tras la introducción del ligando en el huésped, se modula la expresión del gen de b)iii).

35 La invención de los solicitantes también proporciona un método de modulación de la expresión de un gen en una célula huésped que comprende un casete de expresión génica que comprende un elemento de respuesta que comprende un dominio reconocido por el dominio de unión a ADN del primer polipéptido híbrido; un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y un gen cuya expresión se quiere modular; en el que el método comprende las etapas de: a) introducir en la célula huésped un sistema de modulación de la expresión génica de acuerdo con la invención; y b) introducir en la célula huésped un ligando; de modo que, tras la introducción del ligando en el huésped, se modula la expresión del gen.

45 La invención de los solicitantes también proporciona una célula huésped aislada que comprende un sistema inducible de expresión génica de acuerdo con la invención. La presente invención también se refiere a una célula huésped aislada que comprende un casete de expresión génica, un polinucleótido o un polipéptido de acuerdo con la invención. En consecuencia, la invención de los solicitantes también se refiere a un organismo no humano que comprende una célula huésped de acuerdo con la invención.

Breve descripción de los dibujos

50 Figura 1: Datos de expresión de VP16LmUSP-EF, VP16MmRXR α -EF y tres clones independientes de VP16MmRXR α (1-7)-LmUSP (8-12)-EF en células NIH3T3 junto con GAL4CfEcR-CDEF y pFRLuc en presencia de ligando no esteroideo (GSE).

55 Figura 2: Datos de expresión de VP16LmUSP-EF, VP16MmRXR α -EF y dos clones independientes de VP16MmRXR α (1-7)-LmUSP (8-12)-EF en células NIH3T3 junto con GAL4CfEcR-CDEF y pFRLuc en presencia de ligando no esteroideo (GSE).

60 Figura 3: Datos de expresión de VP16LmUSP-EF, VP16MmRXR α -EF y dos clones independientes de VP16MmRXR α (1-7)-LmUSP (8-12)-EF en células A549 junto con GAL4CfEcR-CDEF y pFRLuc en presencia de ligando no esteroideo (GSE).

65 Figura 4: Alineaciones de secuencias de aminoácidos de los dominios EF de seis RXR de vertebrado (A) y seis RXR de invertebrado (B). B6, B8, B9, B10 y B11 indican uniones β quimera. A1 indica unión para α quimera. Las hélices 1-12 se indican como H1-H12 y las láminas β plegadas se indican como S1 y S2. F indica la unión del dominio F.

Figura 5: Datos de expresión de interruptores génicos 1.3-1.6 basados en RXR quimérico GAL4CfEcR-CDEF/VP16 en células NIH3T3 junto con pFRLuc en presencia de ligando no esteroideo (GSE).

Figura 6: Datos de expresión de interruptores génicos que comprenden los dominios DEF de EcR de CfEcR, DmEcR, TmEcR o AmaEcR fusionados con el dominio de unión a ADN de GAL4 y los dominios EF de RXR/USP de CfUSP, DmUSP, LmUSP, MrnRXR α , una quimera entre MmRXR α y LmUSP (quimera), AmaRXR1 o AmaRXR2 fusionados con un dominio de activación VP16 junto con pFRLuc en células NIH3T3 en presencia de ligando esteroideo (PonA) o no esteroideo (GSE). Las diferentes construcciones de RXR/USP se compararon en asociación con GAL4CfEcR-DEF.

Figura 7: Datos de expresión de interruptores génicos que comprenden los dominios DEF de EcR de CfEcR, DmEcR, TmEcR o AmaEcR fusionados con el dominio de unión a ADN de GAL4 y los dominios EF de RXR/USP de CfUSP, DmUSP, LmUSP, MrnRXR α , una quimera entre MmRXR α y LmUSP (Quimera), AmaRXR1 o AmaRXR2 fusionados con un dominio de activación VP16 junto con pFRLuc en células NIH3T3 en presencia de ligando esteroideo (PonA) o no esteroideo (GSE). Las diferentes construcciones de RXR/USP se compararon en asociación con GAL4DmEcR-DEF.

Figura 8: Datos de expresión de interruptores génicos que comprenden los dominios DEF de EcR de CfEcR, DmEcR, TmEcR o AmaEcR fusionados con el dominio de unión a ADN de GAL4 y los dominios EF de RXR/USP de CfUSP, DmUSP, LmUSP, MrnRXR α , una quimera entre MmRXR α y LmUSP (Quimera), AmaRXR1 o AmaRXR2 fusionados con un dominio de activación VP16 junto con pFRLuc en células NIH3T3 en presencia de ligando esteroideo (PonA) o no esteroideo (GSE). Las diferentes construcciones de EcR se compararon en asociación con un RXR-EF quimérico (MmRXR α -(1-7)-LmUSP(8-12)-EF).

Figura 9: Datos de expresión de VP16/MmRXR α -EF (α RXR), VP16/quimera entre MmRXR α -EF y LmUSP-EF (MmRXR α -(1-7)-LmUSP(8-12)-EF; aQ7), VP16/LmUSP-EF (LmUSP) y tres clones independientes de cada una de las cinco VP16/quimeras entre HsRXR β -EF y LmUSP-EF (véase la tabla 1 para construcciones quiméricas de RXR; bRXRQ6, bRXRQ8, bRXRQ9, bRXRQ10 y bRXRQ11) se transfectaron en células NIH3T3 junto con GAL4/CfEcR-DEF y pFRLuc. Las células transfectadas se hicieron crecer en presencia de 0, 0,2, 1 y 10 μ M de ligando no esteroideo (GSE). La actividad indicadora se cuantificó 48 horas después de añadir los ligandos.

Figura 10: Datos de expresión de W16/MmRXR α -EF (α RXR), VP16/quimera entre MmRXR α -EF y LmUSP-EF (MmRXR α -(1-7)-LmUSP(8-12)-EF; aQ7), VP16/LmUSP-EF (LmUSP) y tres clones independientes de cada una de las cinco VP16/quimeras entre HsRXR β -EF y LmUSP-EF (véase la tabla 1 para construcciones quiméricas de RXR; bRXRQ6, bRXRQ8, bRXRQ9, bRXRQ10 y bRXRQ11) se transfectaron en células NIH3T3 junto con GAL4/CfEcR-DEF y pFRLuc. Las células transfectadas se hicieron crecer en presencia de 0, 0,2, 1 y 10 μ M de ligando esteroideo (PonA) o 0, 0,04, 0,2, 1 y 10 μ M de ligando no esteroideo (GSE). La actividad indicadora se cuantificó 48 horas después de añadir los ligandos.

Figura 11: Datos de expresión de VP16/MmRXR α -EF (α RXR), VP16/quimera entre MmRXR α -EF y LmUSP-EF (MmRXR α -(1-7)-LmUSP(8-12)-EF; aQ7), VP16/LmUSP-EF (LmUSP) y tres clones independiente de cada una de las cinco VP16/quimeras entre HsRXR β -EF y LmUSP-EF (véase la tabla 1 para construcciones quiméricas de RXR; bRXRQ6, bRXRQ8, bRXRQ9, bRXRQ10 y bRXRQ11) se transfectaron en células NIH3T3 junto con GAL4/DmEcR-DEF y pFRLuc. Las células transfectadas se hicieron crecer en presencia de 0, 0,2, 1 y 10 μ M de ligando esteroideo (PonA) o 0, 0,04, 0,2, 1 y 10 μ M de ligando no esteroideo (GSE). La actividad indicadora se cuantificó 48 horas después de añadir los ligandos.

Figura 12: Efecto del ácido 9-cis-retinoico sobre el potencial de transactivación del interruptor génico GAL4CfEcR-DEF/VP16HsRXR β -(1-8)-LmUSP-(9-12)-EF (β quimera 9) junto con pFRLuc en células NIH 3T3 en presencia de no esteroide (GSE) y ácido 9-cis-retinoico (9Cis) durante 48 horas.

Descripción detallada de la invención

Los solicitantes han mostrado ahora que los dominios de unión a ligando de RXR quiméricos son funcionales dentro de un sistema inducible de modulación de la expresión génica basado en Ec-R en células de mamífero y que estos LBD de RXR presentan sensibilidades a ligando y capacidades de transactivación ventajosas. Por tanto, la invención de los solicitantes proporciona un sistema inducible de expresión génica basado en receptores de ecdisona novedoso que comprende un dominio de unión a ligando de receptor X retinoide quimérico. En una realización particularmente deseable, la invención de los solicitantes proporciona un sistema inducible de expresión génica que tiene un nivel reducido de expresión génica de fondo y es sensible a concentraciones submicromolares de ligando no esteroideo. Por tanto, el sistema inducible de expresión génica novedoso de los solicitantes y su uso en método de modulación de la expresión génica en una célula huésped superan las limitaciones de los sistemas inducibles de expresión génica disponibles actualmente y proporcionan al experto un medio eficaz para controlar la expresión génica.

La presente invención es útil para aplicaciones tales como tratamiento génico, producción de anticuerpos a gran escala, ensayos de rastreo de alto rendimiento basados en células, genómica funcional, análisis de proteómica y metabolómica y regulación de rasgos en organismos transgénicos, donde es 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 génica y para adaptar los niveles de expresión génica para adecuarlos a las necesidades del usuario.

Definiciones

En la presente divulgación, se usan una serie de términos/expresiones y abreviaturas. Se proporcionan las definiciones siguientes y deberían ser útiles para la comprensión del alcance y la puesta en práctica de la presente invención.

En una realización específica, el término "alrededor de" o "aproximadamente" significa dentro del 20 %, preferentemente dentro del 10 %, más preferentemente dentro del 5 % e incluso más preferentemente dentro del 1 % de un valor o intervalo dado.

La expresión "no tiene sustancialmente" significa que una composición que comprende "A" (donde "A" es una sola proteína, molécula de ADN, vector, célula huésped recombinante, etc.) no tiene, sustancialmente, "B" (donde "B" comprende una o más proteínas, moléculas de ADN, vectores, etc., contaminantes) cuando al menos alrededor del 75 % en peso de las proteínas, ADN, vectores (en función de la categoría de las especies a las que pertenezcan A y B) de la composición es "A". Preferentemente, "A" comprende al menos alrededor del 90 % en peso de las especies A + B de la composición, lo más preferentemente, al menos alrededor del 99 % en peso. También se prefiere que una composición que no tiene, sustancialmente, contaminación, contenga únicamente especies de un sólo peso molecular que tengan la actividad o característica de la especie de interés.

El término "aislado" con los fines de la presente invención designa un material biológico (ácido nucleico o proteína) que se ha sacado de su entorno original (el entorno en el que se encuentra presente de forma natural). Por ejemplo, un polinucleótido presente en el estado natural en una planta o un animal no está aislado, sin embargo, el mismo nucleótido separado de los ácidos nucleicos adyacentes en los que está presente de forma natural, se considera "aislado". El término "purificado" no requiere que el material esté presente en una forma que presente una pureza absoluta, que excluya la presencia de otros compuestos. Es más bien una definición relativa.

Un polinucleótido está en el estado "purificado" después de la purificación del material de partida o del material natural en al menos un orden de magnitud, preferentemente 2 o 3 y preferentemente 4 o 5 órdenes de magnitud.

Un "ácido nucleico" es un compuesto polimérico formado por subunidades enlazadas covalentemente denominadas nucleótidos. Los ácidos nucleicos incluyen ácido polirribonucleico (ARN) y ácido polidesoxirribonucleico (ADN), ambos de los cuales pueden ser monocatenarios o bicatenarios. El ADN incluye, entre otros, ADNc, ADN genómico, ADN plasmídico, ADN sintético y ADN semisintético. El ADN puede ser lineal, circular o superenrollado.

Una "molécula de ácido nucleico" se refiere a la forma polimérica de ésteres de fosfato de ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualquier análogo de fosfoéster de los mismos, tales como fosforotioatos y tioésteres, tanto en forma monocatenaria como en una hélice bicatenaria. Pueden existir hélices bicatenarias de ADN-ADN, ADN-ARN y ARN-ARN. El término molécula de ácido nucleico y, en particular, molécula de ADN o ARN, se refiere únicamente a la estructura primaria y secundaria de la molécula, y no la limita a ninguna forma terciaria en particular. Así, este término incluye ADN bicatenario que se encuentra, entre otros, en moléculas de ADN lineal o circular (p. ej., fragmentos de restricción), plásmidos y cromosomas. En el análisis de la estructura de moléculas de ADN bicatenarias concretas, se pueden describir las secuencias en el presente documento de acuerdo con el consenso normal de dar únicamente la secuencia en la dirección de 5' a 3' a lo largo de la hebra de ADN no transcrita (es decir, la hebra que tiene una secuencia homóloga al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha sufrido manipulación biológica molecular.

Se entenderá que el término "fragmento" significa una secuencia de nucleótidos de longitud reducida con respecto al ácido nucleico de referencia y que comprende, a lo largo de la porción común, una secuencia de nucleótidos idéntica al ácido nucleico de referencia. Un fragmento de ácido nucleico de este tipo de acuerdo con la invención, puede estar, si es apropiado, incluido en un polinucleótido mayor del que forma parte. Estos fragmentos comprenden, o de forma alternativa consisten en, oligonucleótidos cuya longitud varía desde al menos 6, 8, 9, 10, 12, 15, 18, 20, 21, 22, 23, 24, 25, 30, 39, 40, 42, 45, 48, 50, 51, 54, 57, 60, 63, 66, 70, 75, 78, 80, 90, 100, 105, 120, 135, 150, 200, 300, 500, 720, 900, 1000 o 1500 nucleótidos consecutivos de un ácido nucleico de acuerdo con la invención.

Como se usa en el presente documento, un "fragmento de ácido nucleico aislado" es un polímero de ARN o ADN que es mono o bicatenario, que contiene opcionalmente bases de nucleótido sintéticas, artificiales o modificadas. Un fragmento de ácido nucleico aislado en forma de un polímero de ADN puede estar formado por uno o más segmentos de ADNc, ADN genómico o ADN sintético.

Un "gen" se refiere a un conjunto de nucleótidos que codifican un polipéptido, e incluye ácidos nucleicos de ADNc y ADN genómico. "Gen" también se refiere a un fragmento de ácido nucleico que expresa una proteína o un polipéptido específicos, incluidas las secuencias reguladoras anteriores (secuencias no codificantes en 5') y posteriores (secuencias no codificantes en 3') a la secuencia codificante. "Gen nativo" se refiere a un gen tal como se encuentra en la naturaleza con sus propias secuencias reguladoras. "Gen quimérico" se refiere a cualquier gen que no es un gen nativo, que comprende secuencias reguladoras y/o codificantes que no se encuentran juntas en la naturaleza. En consecuencia, un gen quimérico puede comprender secuencias reguladoras y secuencias codificantes que derivan de distintas fuentes, o secuencias reguladoras y secuencias codificantes derivadas de la misma fuente, pero dispuestas de una manera diferente a la que se encuentra en la naturaleza. Un gen quimérico puede comprender secuencias codificantes derivadas de fuentes diferentes y/o secuencias reguladoras derivadas de fuentes diferentes. "Gen endógeno" se refiere a un gen nativo en su ubicación natural en el genoma de un organismo. Un gen "exógeno" o gen "heterólogo" se refiere a un gen que no se encuentra normalmente en el organismo huésped, sino que se introduce en el organismo huésped mediante transferencia génica. Los genes exógenos pueden comprender genes nativos insertados en un organismo no nativo o genes quiméricos. Un "transgén" es un gen que se ha introducido en el genoma mediante un procedimiento de transformación.

ADN "heterólogo" se refiere a ADN ubicado de forma artificial en la célula o en un sitio cromosómico de la célula. Preferentemente, el ADN heterólogo incluye un gen exógeno a la célula.

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

Una molécula de ácido nucleico "se puede hibridar" con otra molécula de ácido nucleico, tal como un ADNc, ADN genómico o ARN, cuando una forma monocatenaria de la molécula de ácido nucleico se puede alinear con la otra molécula de ácido nucleico bajo las condiciones apropiadas de temperatura y fuerza iónica de la solución (véase Sambrook *et al.*, 1989 *infra*). Las condiciones de hibridación y lavado son bien conocidas y se ejemplifican 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), en particular en el capítulo 11 y la tabla 11.1 de dicho documento. Las condiciones de temperatura y fuerza iónica determinan la "rigurosidad" de la hibridación.

Las condiciones de rigurosidad se pueden ajustar para rastrear de fragmentos moderadamente similares, tales como secuencias homólogas de organismos relacionados de forma lejana, a fragmentos altamente similares, tales como genes que duplican enzimas funcionales de organismos estrechamente relacionados. Para el rastreo preliminar de ácidos nucleicos homólogos, se pueden usar condiciones de hibridación poco rigurosas, que corresponden a una T_m de 55 °, p. ej., SSC 5x, SDS al 0,1 %, leche al 0,25 % y sin formamida; o formamida al 30 %, SSC 5x, SDS al 0,5 %). Las condiciones de hibridación moderadamente rigurosas corresponden a una T_m mayor, p. ej., formamida al 40 %, con SSC 5x o 6x. Las condiciones de hibridación altamente rigurosas corresponden a la mayor T_m , p. ej., formamida al 50 %, SSC 5x o 6x. La hibridación requiere que los dos ácidos nucleicos contengan secuencias complementarias, aunque en función de la rigurosidad de la hibridación, pueden existir apareamientos erróneos entre bases.

El término "complementario" se usa para describir la relación entre bases de nucleótidos que pueden hibridarse entre sí. Por ejemplo, con respecto al ADN, la adenosina es complementaria a la timina y la citosina es complementaria a la guanina. En consecuencia, la presente invención también incluye fragmentos de ácido nucleico aislado que son complementarios a las secuencias completas como se divulgan o se usan en el presente documento, así como aquellas secuencias de ácidos nucleicos sustancialmente similares.

En una realización específica, el término "condiciones de hibridación estándar" se refiere a una T_m de 55 °C, y utiliza condiciones expuestas anteriormente. En una realización preferida, la T_m es de 60 °C; en una realización más preferida, la T_m es de 65 °C.

Los lavados posteriores a la hibridación también determinan las condiciones de rigurosidad. Un conjunto de condiciones preferidas usa una serie de lavados que comienzan con SSC 6x, SDS al 0,5 % a temperatura ambiente durante 15 minutos (min), repetido después con SSC 2x, SDS al 0,5 % a 45 °C durante 30 minutos, y repetido después dos veces con SSC 0,2x, SDS al 0,5 % a 50 °C durante 30 minutos. Un conjunto de condiciones rigurosas más preferido usa temperaturas más altas en las que los lavados son idénticos a los anteriores, excepto por la temperatura de los dos lavados finales de 30 min en SSC 0,2x, SDS al 0,5 %, que se aumentó hasta 60 °C. Otro conjunto preferido de condiciones altamente rigurosas usa dos lavados finales en SSC 0,1x, SDS al 0,1 % a 65 °C. La hibridación requiere que los dos ácidos nucleicos comprendan secuencias complementarias, aunque en función de la rigurosidad de la hibridación, pueden existir apareamientos erróneos entre bases.

La rigurosidad apropiada para hibridar ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementación, variables bien conocidas en la técnica. Cuanto mayor es el grado de similitud u homología entre dos secuencias de nucleótidos, mayor es el valor de T_m para híbridos de ácidos nucleicos que tienen esas secuencias. La estabilidad relativa (correspondiente a una T_m mayor) de las hibridaciones de ácidos nucleicos disminuye en el siguiente orden: ARN:ARN, ADN:ARN, ADN:ADN. Para híbridos de más de 100 nucleótidos de longitud, se han derivado ecuaciones para calcular T_m (véase Sambrook *et al.*, *supra*, 9.50-0.51). Para la hibridación con ácidos nucleicos más cortos, es decir, oligonucleótidos, la posición de los apareamientos erróneos adquiere más

importancia y la longitud del oligonucleótido determina su especificidad (véase Sambrook *et al.*, *supra*, 11.7-11.8).

En una realización la longitud de un ácido nucleico hibridable es de al menos 10 nucleótidos. Preferentemente, una longitud mínima para un ácido nucleico hibridable es de al menos alrededor de 15 nucleótidos; más preferentemente de al menos alrededor de 20 nucleótidos; y lo más preferentemente la longitud es de al menos 30 nucleótidos. Además, el experto en la técnica reconocerá que la temperatura y la concentración salina de la solución de lavado se pueden ajustar según sea necesario, de acuerdo con factores tales como la longitud de la sonda.

El término "sonda" se refiere a una molécula de ácido nucleico monocatenaria que puede aparear sus bases con un ácido nucleico objetivo monocatenario complementario para formar una molécula bicatenaria. Como se usa en el presente documento, el término "oligonucleótido" se refiere a un ácido nucleico, en general de al menos 18 nucleótidos, que es hibridable con una molécula de ADN genómico, una molécula de ADNc, un ADN plasmídico o una moléculas de ARNm. Los oligonucleótidos pueden estar marcados, p. ej., con ³²P-nucleótidos o nucleótidos a los que se ha conjugado covalentemente una marca, tal como biotina. Un oligonucleótido marcado se puede usar como sonda para detectar la presencia de un ácido nucleico. Los oligonucleótidos (de los que pueden estar marcados uno o ambos) se pueden usar como cebadores de PCR, para clonar un fragmento de ácido nucleico o uno de longitud completa, o para detectar la presencia de un ácido nucleico. Un oligonucleótido también se puede usar para formar una triple hélice con una molécula de ADN. En general, los oligonucleótidos se preparan sintéticamente, preferentemente en un sintetizador de ácidos nucleicos. En consecuencia, se pueden preparar oligonucleótidos con enlaces fosfoéster análogos artificiales, tales como enlaces tioéster, etc.

Un "cebador" es un oligonucleótido que hibrida con una secuencia de ácidos nucleicos objetivo para crear una región de ácido nucleico bicatenario que puede servir como punto de inicio para la síntesis de ADN bajo condiciones adecuadas. Estos cebadores se pueden usar en la reacción en cadena de la polimerasa.

"Reacción en cadena de la polimerasa" se abrevia como PCR y significa un método *in vitro* para amplificar enzimáticamente secuencias de ácidos nucleicos específicas. La PCR implica una serie repetitiva de ciclos de temperatura, comprendiendo cada ciclo tres etapas: desnaturalización del ácido nucleico molde para separar las hebras de la moléculas objetivo, alineación de un cebador oligonucleotídico de PCR con el ácido nucleico molde y extensión del/de los cebador(es) mediante la polimerasa de ADN. La PCR proporciona un medio para detectar la presencia de la molécula objetivo y, bajo condiciones cuantitativas o semicuantitativas, para determinar la cantidad relativa de esa molécula objetivo en el grupo de ácidos nucleicos de partida.

"La reacción en cadena de la polimerasa de transcripción inversa" se abrevia RT-PCR y significa un método *in vitro* para producir enzimáticamente una molécula o moléculas de ADNc objetivo a partir de una molécula o moléculas de ARN, seguido de la amplificación enzimática de una secuencia o secuencias de ácidos nucleicos específicas de la molécula o moléculas de ADNc descritas anteriormente. La RT-PCR también proporciona un medio para detectar la presencia de la molécula objetivo y, bajo condiciones cuantitativas o semicuantitativas, para determinar la cantidad relativa de esa molécula objetivo en el grupo de ácidos nucleicos de partida.

Una "secuencia codificante" de ADN es una secuencia de ADN bicatenario que se transcribe y se traduce en un polipéptido en una célula *in vitro* o *in vivo* cuando se sitúa bajo el control de secuencias reguladoras adecuadas. "Secuencias reguladoras adecuadas" se refiere a secuencias de nucleótidos situadas corriente arriba (secuencias no codificantes en 5'), dentro o corriente abajo (secuencias no codificantes en 3') de una secuencia codificante, y que influyen en la transcripción, el procesamiento o la estabilidad del ARN o la traducción de la secuencia codificante. Las secuencias reguladoras pueden incluir promotores, secuencias líder de la traducción, intrones, secuencias de reconocimiento de poliadenilación, sitios de procesamiento de ARN, sitios de unión a efectores y estructuras de tallo-lazo. Los límites de la secuencia codificante se determinan por un codón de inicio en el extremo 5' terminal (amino) y un codón de detención de la traducción en el extremo 3' terminal (carboxilo). Una secuencia codificante puede incluir, entre otras, secuencias procariotas, ADNc de ARNm, secuencias de ADN genómico e incluso secuencias de ADN sintéticas. Si la secuencia codificante está destinada a la expresión en una célula eucariota, habitualmente habrá una señal de poliadenilación y una secuencia de terminación de la transcripción situadas en 3' de la secuencia codificante.

"Marco de lectura abierto" se abrevia ORF y significa una extensión de secuencia de ácidos nucleicos, bien de ADN, ADNc o ARN, que comprende una señal de inicio o un codón de inicio de la traducción, tal como un ATG o AUG, y un codón de terminación y se puede traducir potencialmente en una secuencia polipeptídica.

El término "cabeza-cabeza" se usa en el presente documento para describir la orientación de dos secuencias polinucleotídicas una con respecto a la otra. Dos polinucleótidos están situados en una orientación cabeza-cabeza cuando el extremo 5' de la hebra codificante de un polinucleótido está adyacente al extremo 5' de la hebra codificante del otro polinucleótido, de modo que la dirección de la transcripción de cada polinucleótido avanza alejándose del extremo 5' del otro polinucleótido. El término "cabeza-cabeza" se puede abreviar como (5')-a-(5') y también se puede indicar mediante los símbolos (←→) o (3'←5'5'→3').

El término "cola-cola" se usa en el presente documento para describir la orientación de dos secuencias

5 polinucleotídicas una con respecto a la otra. Dos polinucleótidos están situados en una orientación cola-cola cuando el extremo 3' de la hebra codificante de un polinucleótido está adyacente al extremo 3' de la hebra codificante del otro polinucleótido, de modo que la dirección de la transcripción de cada polinucleótido avanza hacia el otro polinucleótido. El término "cola-cola" se puede abreviar como (3')-a-(3') y también se puede indicar mediante los símbolos ($\rightarrow\leftarrow$) o ($5'\rightarrow 3'3'\leftarrow 5'$).

10 El término "cabeza-cola" se usa en el presente documento para describir la orientación de dos secuencias polinucleotídicas una con respecto a la otra. Dos polinucleótidos están situados en una orientación cabeza-cola cuando el extremo 5' de la hebra codificante de un polinucleótido está adyacente al extremo 3' de la hebra codificante del otro polinucleótido, de modo que la dirección de la transcripción de cada polinucleótido avanza en la misma dirección que la del otro polinucleótido. El término "cabeza-cola" se puede abreviar como (5')-a-(3') y también se puede indicar mediante los símbolos ($\rightarrow\rightarrow$) o ($5'\rightarrow 5'5'\rightarrow 3'$).

15 El término "corriente abajo" se refiere a una secuencia de nucleótidos que está situada en 3' de la secuencia de nucleótidos de referencia. En particular, las secuencias de nucleótidos corriente abajo se refieren, en general, a secuencias posteriores al punto de inicio de la transcripción. Por ejemplo, el codón de inicio de la traducción de un gen se sitúa corriente abajo del sitio de inicio de la transcripción.

20 El término "corriente arriba" se refiere a una secuencia de nucleótidos que está situada en 5' de la secuencia de nucleótidos de referencia. En particular, las secuencias de nucleótidos corriente arriba se refieren, en general, a secuencias que se sitúan en el lado 5' de una secuencia codificante o punto de inicio de la transcripción. Por ejemplo, la mayor parte de los promotores se sitúan corriente arriba del sitio de inicio de la transcripción.

25 Los términos "endonucleasa de restricción" y "enzima de restricción" se refieren a una enzima que se une y corta en una secuencia de nucleótidos específica de un ADN bicatenario.

30 "Recombinación homóloga" se refiere a la inserción de una secuencia de ADN exógeno en otra molécula de ADN, p. ej., la inserción de un vector en un cromosoma. Preferentemente, el vector se dirige a un sitio cromosómico específico para su recombinación homóloga. Para la recombinación homóloga específica, el vector contendrá regiones de homología con las secuencias del cromosoma lo suficientemente largas como para permitir la unión complementaria y la incorporación del vector en el cromosoma. Regiones de homología más largas y grados de similitud de secuencia mayores pueden aumentar la eficacia de la recombinación homóloga.

35 Se pueden usar varios métodos conocidos en la técnica para propagar un polinucleótido de acuerdo con la invención. Una vez establecidos un sistema huésped y unas condiciones de crecimiento adecuados, se pueden propagar y preparar en cantidad vectores de expresión recombinantes. Como se describe en el presente documento, los vectores de expresión que se pueden usar, incluyen, entre otros, los siguientes vectores o sus derivados: virus humanos o animales tales como virus vaccinia o adenovirus; virus de insectos tales como baculovirus; vectores de levaduras; vectores bacteriófagos (p. ej., lambda) y vectores de ADN plasmídico y cosmídico, por mencionar algunos.

45 Un "vector" es cualquier medio para la clonación y/o la transferencia de un ácido nucleico en una célula huésped. Un vector puede ser un replicón al que se puede unir otro segmento de ADN para que se produzca la duplicación del segmento unido. Un "replicón" es cualquier elemento genético (p. ej., plásmido, fago, cósmido, cromosoma, virus) que funciona como una unidad autónoma de duplicación de ADN *in vivo*, es decir, que puede duplicarse 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*. Se pueden usar una gran cantidad de vectores conocidos en la técnica para manipular ácidos nucleicos, incorporar elementos de respuesta y promotores en genes, etc. Los vectores posibles incluyen, por ejemplo, plásmidos o virus modificados incluidos, por ejemplo bacteriófagos tales como derivados de lambda, o plásmidos tales como derivados de los plásmidos PBR322 o 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 se puede lograr ligando los fragmentos de ADN apropiados en un vector elegido que tenga extremos cohesivos complementarios. De forma alternativa, se pueden modificar enzimáticamente los extremos de las moléculas de ADN o se puede producir cualquier sitio ligando secuencias de nucleótidos (enlazadores) en los extremos del ADN. Estos vectores se pueden genomanipular para que contengan genes marcadores seleccionables que permiten la selección de células que han incorporado el marcador en el genoma celular. Estos marcadores permiten la identificación y/o la selección de células huésped que incorporan y expresan las proteínas codificadas por el marcador.

60 Los vectores víricos, y en particular los vectores retrovíricos, se han usado en una amplia variedad de aplicaciones de administración de genes en células, así como en sujetos animales vivos. Los vectores víricos que se pueden usar incluyen, entre otros, vectores de retrovirus, virus adenoasociados, viruela, baculovirus, vaccinia, herpes simplex, Epstein-Barr, adenovirus, geminivirus and caulimovirus. Los vectores no víricos 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 útiles para seleccionar, medir y monitorizar 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 frecuentemente porta un gen que no forma parte del metabolismo central de la células, y habitualmente en forma de moléculas de ADN bicatenario circular. Estos elementos pueden ser secuencias de duplicación autónoma, secuencias que forman parte del genoma, secuencias de fagos o de nucleótidos, lineales, circulares o superenrolladas de un ADN o ARN mono o bicatenario, derivado de cualquier fuente, en el que las secuencias de nucleótidos se han unido o recombinado en una única construcción que puede introducir un fragmento promotor y la secuencia de ADN para un producto génico seleccionado junto con la secuencia no traducida en 3' apropiada en una célula.

Un "vector de clonación" es un "replicón", que es una extensión unitaria de ácido nucleico, preferentemente ADN, que se duplica de forma secuencial y que comprenden un origen de duplicación, tal como un plásmido, fago o cósmido, al que se puede unir otro segmento de ácido nucleico para que se produzca la duplicación del segmento unido. Los vectores de clonación pueden ser capaces de replicarse en un tipo celular y expresarse en otro ("vector lanzadera").

Los vectores se pueden introducir en las células huésped deseadas mediante métodos conocidos en la técnica, p. ej., transfección, electroporación, microinyección, transducción, fusión celular, DEAE dextrano, precipitación con fosfato de calcio, lipofección (fusión de lisosomas), uso de una pistola génica o un transportador de vectores de ADN (véanse, p. ej., Wu *et al.*, 1992, J. Biol. Chem. 267: 963-967; Wu y Wu, 1988, J. Biol. Chem. 263: 14621-14624; y Hartmut *et al.*, en la solicitud de patente canadiense N.º 2.012.311, presentada el 15 de marzo de 1990).

Un polinucleótido de acuerdo con la invención también se puede introducir *in vivo* mediante lipofección. Durante la última década, ha aumentado el uso de liposomas para la encapsulación y transfección de ácidos nucleicos *in vitro*. Los lípidos catiónicos diseñados para limitar las dificultades y riesgos encontrados con la transfección mediada por liposomas se pueden usar para preparar liposomas para transfección *in vivo* de un gen que codifica un marcador (Feigner *et al.*, 1987, Proc. Natl. Acad. Sci. U.S.A. 84: 7413; Mackey, *et al.*, 1988, Proc. Natl. Acad. Sci. U.S.A 85: 8027-8031; y Ulmer *et al.*, 1993, Science 259: 1745-1748). El uso de lípidos catiónicos puede promover la encapsulación de ácidos nucleicos cargados negativamente, y también promover la fusión con membranas celulares cargadas negativamente (Feigner y Ringold, 1989, Science 337: 387-388). En las publicaciones de patente internacional WO 95/18863 y WO 96/17823 y en la patente de EE. UU. N.º 5.459.127 se describen composiciones y compuestos lipídicos especialmente útiles para la transferencia de ácidos nucleicos. El uso de lipofección para introducir genes exógenos en los órganos específicos *in vivo* presenta determinadas ventajas prácticas. La dirección molecular de liposomas a células específicas representa un área de beneficio. Resulta evidente que dirigir la transfección a tipos celulares concretos se preferiría especialmente en un tejido con heterogeneidad celular, tal como el páncreas, el hígado, el riñón y el cerebro. Los lípidos se pueden acoplar químicamente a otras moléculas con el fin de dirigirlos (Mackey, *et al.*, 1988, *supra*). Se podrían acoplar químicamente a liposomas péptidos dirigidos, p. ej., hormonas o neurotransmisores, y proteínas tales como anticuerpos, o moléculas no peptídicas.

También son útiles para facilitar la transfección de un ácido nucleico *in vivo* otras moléculas, tales como un oligopéptido catiónico (p. ej., en el documento WO 95/21931), péptidos derivados de proteínas de unión a ADN (p. ej., en el documento WO 96/25508) o un polímero catiónico (p. ej., en el documento WO 95/21931).

También se puede introducir un vector *in vivo* como un plásmido de ADN desnudo (véanse las patentes de EE. UU. 5.693.622, 5.589.466 y 5.580.859). También se pueden usar enfoques de administración de ADN mediada por receptor (Curiel *et al.*, 1992, Hum. Gene Ther. 3: 147-154; y Wu y Wu, 1987, J. Biol. Chem. 262: 4429-4432).

El término "transfección" significa la incorporación de ARN o ADN exógeno o heterólogo por una célula. Una célula se ha "transfectado" con ARN o ADN exógeno o heterólogo cuando ese ARN o ADN se ha introducido en la célula. Una célula se ha "transformado" con ARN o ADN exógeno o heterólogo cuando el ARN o ADN transfectado efectúa un cambio fenotípico. El ARN o ADN transformante se puede integrar (enlazado covalentemente) en ADN cromosómico que constituye el genoma de la célula.

"Transformación" se refiere a la transferencia de un fragmento de ácido nucleico al genoma de un organismo huésped, dando lugar a una herencia genéticamente estable. Los organismos huésped que contiene los fragmentos de ácido nucleico transformados se denominan organismos "transgénicos" o "recombinantes" o "transformados".

El término "región genética" se referirá a una región de una molécula de ácido nucleico o una secuencia de nucleótidos que comprende un gen que codifica un polipéptido.

Además, el vector recombinante que comprende un polinucleótido de acuerdo con la invención puede incluir uno o más orígenes para la duplicación en los huéspedes celulares en los que se desea su amplificación o su expresión, marcadores o marcadores seleccionables.

El término "marcador seleccionable" significa un factor identificador, habitualmente un gen de resistencia química o a antibiótico, que se puede seleccionar basándose en el efector del gen marcador, es decir, resistencia a un

antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes y similares, en los que el efecto se usa para seguir la herencia de un ácido nucleico de interés y/o para identificar una célula u organismo que ha heredado el ácido nucleico de interés. Los ejemplos de genes marcadores seleccionables conocidos en la técnica incluyen: genes que proporcionan resistencia a ampicilina, estreptomina, gentamicina, kanamicina, higromicina, herbicida bialafós, sulfonamida y similares; y genes que se usan como marcadores fenotípicos, es decir, genes reguladores de antocianina, gen de la isopentil transferasa y similares.

El término "gen indicador" significa un ácido nucleico que codifica un factor identificador que se puede identificar basándose en el efecto del gen indicador, en el que el efecto se usa para seguir la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés y/o para medir la inducción de la expresión o la transcripción del gen. Los ejemplos de genes indicadores conocidos y usados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT), β -galactosidasa (LacZ), β -glucuronidasa (Gus) y similares. Los genes marcadores seleccionables también se pueden considerar genes indicadores.

"Promotor" se refiere a una secuencia de ADN que puede controlar la expresión de una secuencia codificante o un ARN funcional. En general, una secuencia codificante se sitúa en 3' con respecto a una secuencia promotora. Los promotores pueden derivar en su totalidad de un gen nativo o estar formados por diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintético. Los expertos en la técnica entienden que promotores diferentes pueden dirigir la expresión de un gen en tejidos o tipos celulares diferentes, o en etapas diferentes del desarrollo, o en respuesta a condiciones ambientales o fisiológicas diferentes. Los promotores que hacen que un gen se exprese en la mayoría de los tipos celulares en la mayoría de las ocasiones se denominan comúnmente "promotores constitutivos". Los promotores que hacen que un gen se exprese en un tipo celular específico se denominan comúnmente "promotores específicos de célula" o "promotores específicos de tejido". Los promotores que hacen que un gen se exprese en una etapa del desarrollo o la diferenciación celular específica se denominan comúnmente "promotores específicos del desarrollo" o "promotores específicos de la diferenciación celular". Los promotores que se inducen y hacen 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, dado que en la mayoría de los casos no se han definido totalmente los límites exactos de las secuencias reguladoras, fragmentos de ADN de longitudes diferentes pueden tener una actividad promotora idéntica.

Una "secuencia promotora" es una región reguladora de ADN que se puede unir a la polimerasa de ARN en una célula e iniciar la transcripción de una secuencia codificante corriente abajo (en dirección 3'). Para el propósito de definir la presente invención, la secuencia promotora está unida en su extremo 3' por el sitio de inicio de la transcripción y se extiende corriente arriba (en dirección 5') para incluir el mínimo número de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del nivel de fondo. Dentro de la secuencia promotora se encontrará un sitio de inicio de la transcripción (convenientemente definido, por ejemplo, mapeando con la nucleasa S1), así como dominios de unión a proteína (secuencias consenso) responsables de la unión de la polimerasa de ARN.

Una secuencia codificante está "bajo el control" de secuencias de control de la transcripción y la traducción en una célula cuando la polimerasa de ARN transcribe la secuencia codificante en ARNm, que después sufre un ajuste de trans-ARN (si la secuencia codificante contiene intrones) y se traduce en la proteína codificada por la secuencia codificante.

"Secuencias de control de la transcripción y la traducción" son secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que permiten la expresión de una secuencia codificante en una célula huésped. En células eucariotas, las señales de poliadenilación son secuencias de control.

El término "elemento de respuesta" significa uno o más elementos de ADN que actúan en cis que confieren sensibilidad en un promotor mediada a través de 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 formado por motivos de secuencia o semisitios separados por un número variable de nucleótidos. Los semisitios pueden ser similares o idénticos y estar dispuestos como repeticiones directas o inversas o como un solo semisitio o multímeros de semisitios adyacentes en tándem. El elemento de respuesta puede comprender un promotor mínimo aislado a partir de diferentes organismos en función de la naturaleza de la célula u organismo en el que se va a incorporar el elemento de respuesta. El dominio de unión a ADN de la primera proteína híbrida se une, en presencia o ausencia de un ligando, a la secuencia de ADN de un elemento de respuesta para iniciar o suprimir la transcripción de genes corriente abajo bajo la regulación de este elemento de respuesta. Los ejemplos de secuencias de ADN para elementos de respuesta del receptor de ecdisona natural incluyen: RRGG/TTCANTGAC/ACY (véase Cherbas L., *et. al.*, (1991), *Genes Dev.* 5,120-131); AGGTCAN_(n)AGGTCA, donde N_(n) puede ser uno o más nucleótidos espaciadores (véase D'Avino PP., *et. al.*, (1995), *Mol. Cell. Endocrinol.* 113, 1-9); y GGGTTGAATGAATTT (véase Antoniewski C., *et. al.*, (1994), *Mol. Cell Biol.* 14, 4465-4474).

La expresión "enlazado de manera funcional" se refiere a la asociación de secuencias de ácidos nucleicos en un

solo fragmento de ácido nucleico, de forma que la función de uno se ve afectada por el otro. Por ejemplo, un promotor está enlazado de manera funcional con una secuencia codificante cuando puede afectar a la expresión de esa secuencia codificante (es decir, que la secuencia codificante está bajo el control transcripcional del promotor). Las secuencias codificantes se pueden enlazar de manera funcional a secuencias reguladoras en orientación sentido o antisentido.

El término "expresión", como se usa en el presente documento, se refiere a la transcripción y acumulación estable de ARN sentido (ARNm) o antisentido derivado a partir de un ácido nucleico o polinucleótido. Expresión también se puede referir a la traducción de ARNm en una proteína o polipéptido.

Los términos "casete", "casete de expresión" y "casete de expresión génica" se refieren a un segmento de ADN que se puede insertar 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 los sitios de restricción se diseñan para garantizar la inserción del casete en el marco de lectura correcto para la transcripción y la traducción. "Casete de transformación" se refiere a un vector específico que comprende un polinucleótido que codifica un polipéptido de interés y que tiene elementos además del polinucleótido que facilitan la transformación de una célula huésped en particular. Los casetes, casetes de expresión, casetes de expresión génica y casetes de transformación de la invención también pueden comprender elementos que permitan la expresión potenciada de un polinucleótido que codifica un polipéptido de interés en una célula huésped. Estos elementos puede incluir, entre otros: un promotor, un promotor mínimo, un potenciador, un elemento de respuesta, una secuencia de terminación, una secuencia de poliadenilación y similares.

Para los propósitos de la presente invención, el término "interruptor génico" se refiere a la combinación de un elemento de respuesta asociado con 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 están incorporados el elemento de respuesta y el promotor.

Los términos "modulan" y "modula" significan inducir, reducir o inhibir la expresión de un gen o un ácido nucleico, dando lugar a la correspondiente inducción, reducción o inhibición de la producción de proteína o polipéptido.

Los plásmidos o vectores de acuerdo con la invención pueden comprender además al menos un promotor adecuado para dirigir la expresión de un gen en una célula huésped. El término "vector de expresión" significa 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, se sitúa habitualmente bajo el control de elementos de control tales como un promotor, un promotor mínimo, un potenciador o similares. Las regiones de control de la iniciación o promotores, que son útiles para dirigir la expresión de un ácido nucleico en la célula huésped deseada son numerosos y conocidos por los expertos en la técnica. Prácticamente cualquier promotor que pueda dirigir estos genes es adecuado para la presente invención, incluidos, entre otros: promotores víricos, promotores bacterianos, promotores animales, promotores de mamíferos, promotores sintéticos, promotores constitutivos, promotores específicos de tejido, promotores específicos del desarrollo, promotores inducibles, promotores regulados por luz; *CYC1*, *HIS3*, *GAL1*, *GAL4*, *GAL10*, *ADH1*, *PGK*, *PH05*, *GAPDH*, *ADC1*, *TRP1*, *URA3*, *LEU2*, *ENO*, *TPI*, promotores de fosfatasa alcalina (útiles para la expresión en *Saccharomyces*); promotor *AOX1* (útil para la expresión en *Pichia*); β -lactamasa, promotores *lac*, *ara*, *tet*, *trp*, *IP_L*, *IP_R*, *T7*, *tac*, y *trc* (útiles para la expresión en *Escherichia coli*); promotores regulados por luz; los promotores animales y de mamíferos conocidos en la técnica incluyen, entre otros, la región del promotor temprano del SV40 (SV40e), el promotor contenido en la repetición terminal larga (RTL) en 3' del virus del sarcoma de Rous (VSR), los genes de los promotores del E1A o de los promotores tardíos principales (MLP) de adenovirus (Ad), el promotor temprano del citomegalovirus (CMV), el promotor de timidina cinasa (TK) del virus herpes simplex (VHS), un promotor del factor de elongación 1 alfa (EF1), un promotor de fosfoglicerato cinasa (PGK), un promotor de ubiquitina (Ubc), un promotor de albúmina, las secuencias reguladoras del promotor de la metalotioneína-L de ratón y regiones de control de la transcripción, los promotores ubicuos (HPRT, vimentina, α -actina, tubulina y similares), los promotores de los filamentos intermedios (desmina, neurofilamentos, queratina, GFAP y similares), los promotores de genes terapéuticos (de tipo MDR, CFTR o factor VIII y similares), promotores relacionados con la patogénesis o enfermedades y promotores que presentan especificidad de tejido y que se han utilizado en animales transgénicos, tales como la región de control del gen de la elastasa I que está activa en células acinares del páncreas; la región de control del gen de la insulina activa en células beta del páncreas, la región de control del gen de la inmunoglobulina activa en células linfoides, la región de control del virus de tumores mamarios en ratones activa en células de testículo, mama, linfoides y mastocitos; las regiones de control del gen de la albúmina, Apo AI y Apo AII activas en el hígado, la región de control del gen de la alfa-fetoproteína activa en el hígado, la región de control del gen de la alfa 1-antitripsina activa en el hígado, la región de control del gen de la beta-globina activa en células mieloides, la región de control del gen de la proteína básica de mielina activa en oligodendrocitos del cerebro, la región de control del gen 2 de cadena ligera de la miosina activa en músculo esquelético y la región de control del gen de liberación de la hormona gonadotropina activa en el hipotálamo, el promotor de piruvato cinasa, el promotor de vilina, el promotor de la proteína intestinal de unión a ácidos grasos, el promotor de la α -actina de células de músculo liso y similares. Además, estas secuencias de expresión se pueden modificar mediante la adición de un potenciador o secuencias reguladoras y similares.

Los potenciadores que se puede usar en la realizaciones de la invención incluyen, entre otros: un potenciador del SV40, un potenciador del citomegalovirus (CMV), un potenciador del factor de elongación 1 (EF1), potenciadores de levaduras, potenciadores génicos víricos y similares.

5 La regiones de control de la terminación, es decir, secuencias de terminación o poliadenilación, también pueden derivar de diversos genes nativos de los huéspedes preferidos. Opcionalmente, puede no ser necesario un sitio de terminación, aunque lo más preferido es que se incluya. En una realización preferida de la invención, la región de control de la terminación puede estar comprendida o derivar de una secuencia sintética, una señal de poliadenilación sintética, una señal de poliadenilación tardía del SV40, una señal de poliadenilación del SV40, una señal de poliadenilación de la hormona de crecimiento bovina (BGH), secuencias de terminación víricas o similares.

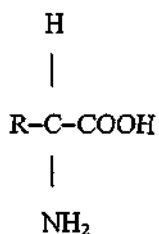
10 Las expresiones "secuencias no codificantes en 3'" o "región no traducida (UTR) en 3'" se refieren a secuencias de ADN situadas corriente abajo (3') de una secuencia codificante y pueden comprender secuencias de reconocimiento de poliadenilación [poly(A)] y otras secuencias que codifican señales reguladoras que pueden afectar al procesamiento del ARNm o a la expresión génica. Habitualmente, la señal de poliadenilación se caracteriza por afectar a la adición de tramos de ácido poliadenílico al extremo 3' del precursor del ARNm.

15 "Región reguladora" significa una secuencia de ácidos nucleicos que regula la expresión de una segunda secuencia de ácidos nucleicos. Una región reguladora puede incluir secuencias que son responsables de forma natural de la expresión de un ácido nucleico en particular (una región homóloga) o puede incluir secuencias de un origen diferente que son responsables de la expresión de diferentes proteínas o incluso proteínas sintéticas (una región heteróloga). En particular, las secuencias pueden ser secuencias de genes procariotas, eucariotas o víricos, o secuencias derivadas que estimulan o reprimen la transcripción de un gen de manera específica o inespecífica y de manera inducible o no inducible. Las regiones reguladoras incluyen orígenes de duplicación, sitios de ajuste de ARN, promotores, potenciadores, secuencias de terminación de la transcripción y secuencias señal que dirigen al polipéptido hacia las rutas de secreción de la célula objetivo.

20 Una región reguladora de una "fuente heteróloga" es una región reguladora que no está asociada de forma natural con el ácido nucleico expresado. Entre las regiones reguladoras heterólogas se incluyen regiones reguladoras de una especie distinta, regiones reguladoras de un gen diferente, secuencias reguladoras híbridas y secuencias reguladoras que no existen en la naturaleza, sino que están diseñadas por un experto en la técnica.

25 "Transcrito de ARN" se refiere al producto resultante de la transcripción catalizada por la polimerasa de ARN de una secuencia de ADN. Cuando el transcrito de ARN es una copia complementaria perfecta de la secuencia de ADN, se denomina transcrito primario o puede ser una secuencia de ARN derivada del procesamiento postranscripcional del transcrito primario y se denomina ARN maduro. "ARN mensajero (ARNm)" se refiere al ARN que no tiene intrones y que se puede traducir en una proteína por la célula. "ADNc" se refiere a un ADN bicatenario que es complementario con y deriva del ARNm. ARN "sentido" se refiere a un transcrito de ARN que incluye el ARNm y, por lo tanto, se puede traducir en una proteína por la célula. "ARN antisentido" se refiere a un transcrito de ARN que es complementario a la totalidad o a parte de un transcrito primario o ARNm objetivo y que bloquea la expresión de un gen objetivo. La complementariedad de un ARN antisentido puede ser con cualquier parte del transcrito del gen específico, es decir, en la secuencia no codificante en 5', en la secuencia no codificante en 3' o en la secuencia codificante. "ARN funcional" se refiere a ARN antisentido, ARN de ribozima u otro ARN que no se traduce pero tiene un efecto sobre los procesos celulares.

30 Un "polipéptido" es un compuesto polimérico formado por residuos de aminoácido enlazados covalentemente. Los aminoácido tienen la siguiente estructura general:



35 40 45 50 55 Los aminoácidos se clasifican en siete grupos en función de 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 está fusionado al grupo amino. Preferentemente, un polipéptido de la invención comprende al menos alrededor de 14 aminoácidos.

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

Un "polipéptido aislado" o "proteína aislada" es un polipéptido o proteína que no tiene, sustancialmente, los compuestos que normalmente están asociados con él/ella en su estado natural (p. ej., otras proteínas o polipéptidos, ácidos nucleicos, hidratos de carbono, lípidos). Con "aislado" no se pretende excluir mezclas artificiales o sintéticas con otros compuestos, o la presencia de impurezas que no interfieren con la actividad biológica, y que pueden estar presentes, por ejemplo, debido a una purificación incompleta, la adición de estabilizantes, o la formación de compuestos en una preparación farmacéuticamente aceptable.

Se entenderá que "fragmento" de un polipéptido de acuerdo con la invención significa un polipéptido cuya secuencia de aminoácidos es más corta que la del polipéptido de referencia y que comprende, en la totalidad de la porción con estos polipéptidos de referencia, una secuencia de aminoácidos idéntica. Estos fragmentos pueden incluirse, cuando sea apropiado, en un polipéptido mayor del que forman parte. Estos fragmentos de un polipéptido de acuerdo con la invención pueden tener una longitud de al menos 2, 3, 4, 5, 6, 8, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 30, 35, 40, 45, 50, 100, 200, 240 o 300 aminoácidos.

Una "variante" de un polipéptido o proteína es cualquier análogo, fragmento, derivado o mutante que deriva de un polipéptido o proteína y que mantiene al menos una propiedad biológica del polipéptido o la 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 que codifica la proteína, o pueden implicar ajuste diferencial o modificación postraducciona. El experto en la técnica puede producir variantes con sustituciones, deleciones, adiciones o reemplazos de aminoácidos individuales o múltiples. Estas variantes pueden incluir, entre otras: (a) variantes en las que se sustituyen uno o más residuos de aminoácido con 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 seroalbúmina. Las técnicas para obtener estas variantes, que incluyen técnicas genéticas (supresiones, deleciones, mutaciones, etc.), químicas y enzimáticas, son conocidas por los expertos en la técnica. Preferentemente, un polipéptido variante comprende al menos alrededor de 14 aminoácidos.

Una "proteína heteróloga" se refiere a una proteína que no se produce de forma natural en la célula.

Una "proteína madura" se refiere a un polipéptido procesado postraduccionalmente; es decir, uno del cual se han eliminado todos los pre- o propéptidos presentes en el producto primario de la traducción. Proteína "precursora" se refiere al producto primario de la traducción del ARNm; es decir, con pre- y propéptidos presentes todavía. Los pre- y propéptidos pueden ser, pero no se limitan a señales de localización intracelular.

El término "péptido señal" se refiere a un polipéptido amino terminal que precede a la proteína madura secretada. El péptido señal se escinde de ella y, por lo tanto, no está presente en la proteína madura. Los péptidos señal tienen la función de dirigir y translocar proteínas secretadas a través de membranas celulares. El péptido señal también se denomina proteína señal.

Una "secuencia señal" se incluye en el comienzo de la secuencia codificante de una proteína que se va a expresar en la superficie de una célula. Esta secuencia codifica un péptido señal, en N-terminal del polipéptido maduro, que dirige a la célula huésped para translocar el polipéptido. En el presente documento se usa el término "secuencia señal de translocación" para referirse a este tipo de secuencia señal. Las secuencias señal de translocación se pueden encontrar asociadas con una variedad de proteínas nativas de eucariotas y procariotas y, con frecuencia, son funcionales en ambos tipos de organismos.

El término "homología" se refiere al porcentaje de identidad entre dos restos de polinucleótido o dos de polipéptido. La correspondencia entre la secuencia de un resto y otro se puede determinar mediante técnicas conocidas en la técnica. Por ejemplo, la homología se puede determinar mediante una comparación directa de la información de secuencia entre dos moléculas polipeptídicas alineando la información de secuencia y usando programas de ordenador fácilmente disponibles. De forma alternativa, la homología se puede determinar mediante la hibridación de polinucleótidos bajo condiciones que forman dúplex estables entre regiones homólogas, seguida de digestión con nucleasas específicas de una hebra y determinación del tamaño de los fragmentos digeridos.

Como se usa en el presente documento, el término "homólogo" es todas sus formas gramaticales y variantes ortográficas se refiere a la relación entre proteínas que poseen un "origen evolutivo común", incluidas proteínas de superfamilias (p. ej., la superfamilia de las inmunoglobulinas) y proteínas homólogas de especies diferentes (p. ej., cadena ligera de la miosina, etc.) (Reeck *et al.*, 1987, Cell 50:667.). Estas proteínas (y sus genes codificantes) tienen homología de secuencia, como refleja su alto grado de similitud de secuencia. Sin embargo, en el uso común y en la presente solicitud, el término "homólogo", cuando se modifica con un adverbio tal como "altamente", se puede referir a similitud de secuencia y no a un origen evolutivo común.

En consecuencia, el término "similitud de secuencia" 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 (véase Reeck *et al.*, 1987, Cell 50:667).

5 En una realización específica, dos secuencias de ADN son "sustancialmente homólogas" o "sustancialmente similares" cuando al menos alrededor del 50 % (preferentemente al menos alrededor del 75 % y lo más preferentemente al menos alrededor del 90 o el 95 %) de los nucleótidos coincidan a lo largo de la extensión definida de las secuencias de ADN. Se pueden identificar secuencias que son sustancialmente homólogas comparando las secuencias usando programas informáticos estándar disponibles en bancos de datos de secuencias, o en un experimento de hibridación de bandas southern bajo, por ejemplo, condiciones rigurosas definidas para ese sistema en particular. La definición de condiciones de hibridación apropiadas forma parte del conocimiento de la técnica.
10 Véase, p. ej., Sambrook *et al.*, 1989, *supra*.

15 Como se usa en el presente documento, "sustancialmente similar" se refiere a fragmentos de ácido nucleico en los que los cambios en una o más bases de nucleótido dan lugar a la sustitución de uno o más aminoácidos, pero no afectan a las propiedades funcionales de la proteína codificada por la secuencia de ADN. "Sustancialmente similar" también se refiere a fragmentos de ácido nucleico en los que los cambios en una o más bases de nucleótido no afectan a la capacidad del fragmento de ácido nucleico para mediar la modificación de la expresión génica mediante tecnología antisentido o de cosupresión. "Sustancialmente similar" también se refiere a modificaciones de los fragmentos de ácido nucleico de la presente invención tales como delección o inserción de una o más bases de nucleótido que no afectan sustancialmente a las propiedades funcionales del transcrito resultante. Por lo tanto, se entiende que la invención engloba más que las secuencias ejemplares. Cada una de las modificaciones propuestas forma parte de los conocimientos rutinarios de la técnica, como lo está la determinación del mantenimiento de la actividad biológica de los productos codificados.

25 Además, el experto en la técnica reconoce que las secuencias sustancialmente similares englobadas por la presente invención también se definen por su capacidad para hibridar, bajo condiciones rigurosas (SSC 0,1x, SDS al 0,1 %, 65 °C y lavado con SSC 2x, SDS al 0,1 % seguido de SSC 0,1x, SDS al 0,1 %), con las secuencias ejemplificadas en el presente documento. Los fragmentos de ácido nucleico sustancialmente similares de la presente invención son aquellos fragmentos de ácido nucleico cuyas secuencias de ADN son al menos un 70 % idénticas a la secuencia de ADN de los fragmentos de ácido nucleico de los que se informa en el presente documento. Los fragmentos de ácido nucleico sustancialmente similares preferidos de la presente invención son aquellos fragmentos de ácido nucleico cuyas secuencias de ADN son al menos un 80 % idénticas a la secuencia de ADN de los fragmentos de ácido nucleico de los que se informa en el presente documento. Se prefieren más los fragmentos de ácido nucleico que son al menos un 90 % idénticos a la secuencia de ADN de los fragmentos de ácido nucleico de los que se informa en el presente documento. Se prefieren aún más los fragmentos de ácido nucleico que son al menos un 95 % idénticos a la secuencia de ADN de los fragmentos de ácido nucleico de los que se informa en el presente documento.
30
35

40 Dos secuencias de aminoácidos son "sustancialmente homólogas" o "sustancialmente similares" cuando más de alrededor del 40 % de los aminoácidos son idénticos, o más del 60 % son similares (funcionalmente idénticos). Preferentemente, las secuencias similares u homólogas se identifican mediante alineación usando, por ejemplo, el programa Pileup de GCG (Genetics Computer Group, Program Manual for the GCG Package, Versión 7, Madison, Wisconsin).

45 La expresión "correspondiente a" se usa en el presente documento para referirse a secuencias similares u homólogas, tanto si la posición exacta es idéntica como si es diferente de la de la molécula frente a la que se mide la similitud o la homología. Una alineación de secuencias de ácidos nucleicos o aminoácidos puede incluir espacios. Así, el término "correspondiente a" se refiere a la similitud de secuencia y no a la numeración de los residuos de aminoácidos o las bases de nucleótido.

50 Una "porción sustancial" de una secuencia de aminoácidos o nucleótidos comprende lo suficiente de la secuencia de aminoácidos de un polipéptido o la secuencia de nucleótidos de un gen para identificar teóricamente ese polipéptido o gen, bien mediante evaluación manual de la secuencia por un experto en la técnica o bien mediante comparación de secuencias automatizada por ordenador e identificación usando algoritmos tales como BLAST (Basic Local Alignment Search Tool; Altschul, S. F., *et al.*, (1993) *J. Mol. Biol.* 215: 403-410; véase también www.ncbi.nlm.nih.gov/BLAST/). En general, es necesaria una secuencia de diez o más aminoácidos contiguos para identificar teóricamente una secuencia polipeptídica o de ácidos nucleicos como homóloga a una proteína o un gen conocido. Además, con respecto a las secuencias de nucleótidos, se pueden usar sondas oligonucleotídicas específicas de genes que comprenden 20-30 nucleótidos contiguos en métodos dependientes de secuencia de identificación (p. ej., hibridación de bandas southern) y aislamiento (p. ej., hibridación *in situ* de colonias bacteriana o placas de bacteriófagos) de genes. Además, se pueden usar oligonucleótidos cortos de 12-15 bases como cebadores de amplificación en PCR para obtener un fragmento de ácido nucleico en particular que comprenda los cebadores. En consecuencia, una "porción sustancial" de una secuencia de nucleótidos comprende lo suficiente de la secuencia para identificar y/o aislar específicamente un fragmento de ácido nucleico que comprenda la secuencia.
55
60

65 El término "porcentaje de identidad", como se conoce en la técnica, es una relación entre dos o más secuencias polipeptídicas o entre dos o más secuencias polinucleotídicas, determinada comparando las secuencias. En la

técnica, "identidad" también significa el grado de relación entre secuencias de polipéptidos o polinucleótidos, según sea el caso, determinado por la coincidencia entre cadenas de tales secuencias. La "identidad" y la "similitud" se pueden calcular fácilmente mediante métodos conocidos, incluidos, entre otros, los descritos en: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, Nueva York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, Nueva York (1993); *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., y Griffin, H. G., ed.) Humana Press, Nueva Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); y *Sequence Analysis Primer* (Gribskov, M. y Devereux, J., ed.) Stockton Press, Nueva York (1991). Los métodos preferidos para determinar la identidad de secuencia se diseñan para dar la mejor coincidencia entre las secuencias probadas. Los métodos para determinar la identidad y la similitud están codificados en programas de ordenador disponibles públicamente. Las alineaciones de secuencias y los cálculos de porcentaje de identidad se pueden realizar usando el programa Megalign del paquete de aplicaciones de bioinformática de LASERGENE (DNASTAR Inc., Madison, WI). Se puede realizar una alineación múltiple de las secuencias usando el método Clustal de alineación (Higgins y Sharp (1989) *CABIOS*. 5:151-153) con los parámetros por defecto (PENALIZACIÓN POR HUECO=10, PENALIZACIÓN POR LONGITUD DE HUECO=10). Los parámetros por defecto para alineaciones por parejas usando el método Clustal se pueden seleccionar: KTUPLE 1, PENALIZACIÓN POR HUECO=3, VENTANA=5 y DIAGONALES GUARDADAS=5.

El término "programa informático de análisis de secuencias" se refiere a cualquier algoritmo de ordenador o programa informático que es útil para el análisis de secuencias de nucleótidos o aminoácidos. El "programa informático de análisis de secuencias" puede estar comercialmente disponible o desarrollarse de forma independiente. Los programas informáticos de análisis de secuencias típicos incluirán, entre otros, el conjunto de programas de GCG (Wisconsin Package Versión 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul *et al.*, *J. Mol. Biol.* 215:403-10 (1990) y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE. UU.). En el contexto de la presente solicitud, se entenderá que cuando se use un programa informático de análisis de secuencias para el análisis, los resultados del análisis se basarán en los "valores por defecto" del programa mencionado, a menos que se especifique lo contrario. Como se usa en el presente documento, "valores por defecto" significará cualquier conjunto de valores o parámetros que se cargan originalmente con el programa informático cuando se inicia por primera vez.

Se pueden ensamblar "genes sintéticos" a partir de bloques oligonucleotídicos de construcción que se sintetizan químicamente usando procedimientos conocidos por los expertos en la técnica. Estos bloques de construcción se enlazan y alinean para formar segmentos génicos que después se ensamblan enzimáticamente para construir el gen completo. "Sintetizada químicamente", en relación con una secuencia de ADN, significa que los nucleótidos que la componen se ensamblaron *in vitro*. La síntesis química manual de ADN se puede lograr usando procedimientos bien establecidos, o se puede realizar una síntesis química automatizada usando una de una serie de máquinas comercialmente disponibles. En consecuencia, se pueden adaptar los genes para una expresión génica óptima basada en la optimización de la secuencia de nucleótidos para reflejar el sesgo de los codones de la célula huésped. El experto en la técnica aprecia la probabilidad de una expresión génica satisfactoria si el uso de los codones está sesgado hacia aquellos codones favorecidos por el huésped. La determinación de los codones preferidos se puede basar en un estudio de genes derivados de la célula huésped donde está disponible la información de secuencia.

Sistema de modulación de la expresión génica de la invención

Los solicitantes han mostrado anteriormente que separar los dominios de transactivación y de unión a ADN situándolos en dos proteínas diferentes da lugar a una actividad de fondo enormemente reducida en ausencia de un ligando y significativamente aumentada con respecto al fondo en presencia de un ligando (en la solicitud en trámite PCT/US01/09050). Este sistema de dos híbridos es un sistema inducible de modulación de la expresión génica significativamente mejorado en comparación con los dos sistemas divulgados en las solicitudes de patente internacional PCT/US97/05330 y PCT/US98/14215. El sistema de dos híbridos aprovecha la capacidad de un par de proteínas que interactúan para llevar el dominio de activación de la transcripción a una posición más favorable con respecto al dominio de unión a ADN, de forma que cuando el dominio de unión a ADN se une al sitio de unión a ADN en el gen, el dominio de transactivación activa el promotor de forma más eficaz (véase, por ejemplo, la patente de EE. UU. N.º 5.283.173). Brevemente, el sistema de expresión génica de dos híbridos comprende dos casetes de expresión génica; el primero codifica un dominio de unión a ADN fusionado con un polipéptido de receptor nuclear y el segundo codifica un dominio de transactivación fusionado con un polipéptido de receptor nuclear diferente. En presencia de un ligando, la interacción del primer polipéptido con el segundo polipéptido une de forma eficaz el dominio de unión de ADN al dominio de transactivación. Dado que los dominios de unión a ADN y de transactivación se encuentran en dos moléculas diferentes, la actividad de fondo en ausencia de un ligando disminuye enormemente.

El sistema de modulación de la expresión génica basado en receptores de ecdisona de dos híbridos puede ser heterodimérico u homodimérico. En general, un complejo de EcR funcional se refiere a un complejo proteico heterodimérico que consiste en dos miembros de la familia de receptores de esteroides, una proteína de receptor de ecdisona obtenida a partir de diversos insectos y una proteína ultraspiráculo (USP) o el homólogo de vertebrado de la USP, la proteína de receptor X retinoide (véase Yao, *et al.* (1993) *Nature* 366,476-479; Yao, *et al.*, (1992) *Cell* 71,63-72). Sin embargo, el complejo también puede ser un homodímero como se detalla a continuación. El complejo

receptor de ecdisteroides funcional también puede incluir proteína(s) adicional(es) tales como inmunofilinas. Otros miembros de la familia de proteínas de receptores de esteroides, conocidos como factores transcripcionales (tales como DHR38 o *betaFTZ-1*), también puede ser compañeros dependientes de ligando o independientes para EcR, USP y/o RXR. Adicionalmente, pueden ser necesarios otros cofactores tales como proteínas conocidas de forma general como coactivadores (también denominadas adaptadores o mediadores). Estas proteínas no se unen de forma específica de secuencia al ADN y no participan en la transcripción basal. Pueden ejercer su efecto sobre la activación de la transcripción a través de diversos mecanismos, incluida la estimulación de la unión al ADN de activadores, afectando a la estructura de la cromatina o mediando las interacciones activador-complejo de iniciación. Los ejemplos de coactivadores de este tipo incluyen RIP140, TIF1, RAP46/Bag-1, ARA70, SRC-1/NCoA-1, TIF2/GRIP/NCoA-2, ACTR/AIB1/RAC3/pCIP, así como el coactivador promiscuo proteína de unión B al elemento de respuesta C, CBP/p300 (para una revisión, véase Glass *et al.*, *Curr. Opin. Cell Biol.* 9: 222-232, 1997). Asimismo, pueden ser necesarios cofactores proteicos conocidos de forma general como correpresores (también conocidos como represores, silenciadores o mediadores de silenciamiento) para inhibir de forma eficaz la activación transcripcional en ausencia de ligando. Estos correpresores pueden interaccionar con el receptor de ecdisona sin ligando para silenciar la actividad en el elemento de respuesta. Las pruebas actuales indican que la unión del ligando cambia la conformación del receptor, lo que da lugar a la liberación del correpresor y al reclutamiento de los coactivadores descritos anteriormente, eliminando de este modo su actividad silenciadora. Los ejemplos de correpresores incluyen N-CoR y SMRT (para una revisión, véase Horwitz *et al.* *Mol Endocrinol.* 10: 1167-1177, 1996). Estos cofactores pueden ser endógenos, en el interior de la célula u organismo, o se pueden añadir de forma exógena como transgenes para que se expresen de manera regulada o no regulada. Los complejos homodiméricos de la proteína de receptor de ecdisona, USP o RXR también pueden ser funcionales bajo algunas circunstancias.

Normalmente, el complejo receptor de ecdisona incluye proteínas que son miembros de la superfamilia de receptores nucleares en la que, en general, todos los miembros se caracterizan por la presencia de un dominio de transactivación aminoterminal, un dominio de unión a ADN ("DBD") y un dominio de unión a ligando ("LBD") separado del DBD por una región bisagra. Como se usa en el presente documento, el término "dominio de unión a ADN" comprende una secuencia polipeptídica mínima de una proteína de unión a ADN, hasta la longitud completa de una proteína de unión a ADN, siempre que el dominio de unión a ADN funcione para asociarse con un elemento de respuesta en 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, F (véanse la patente de EE. UU. 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 tener otro dominio de transactivación en el lado carboxilterminal del LBD, que corresponde a "F".

El DBD se caracteriza por la presencia de dos dedos de cinc de cisteínas entre los que se encuentran dos motivos de aminoácidos, la caja P y la caja D, que confieren especificidad por los elementos de respuesta a ecdisona. Estos dominios pueden ser nativos, modificados o quimeras de diferentes dominios de proteínas de receptor heterólogas. Este receptor EcR, como un subconjunto de la familia de receptores de esteroides, también posee regiones peor definidas 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 se pueden intercambiar.

Se conocen sistemas de interruptores génicos que incorporan componentes del complejo receptor de ecdisona. Sin embargo, en estos sistemas conocidos, siempre que se usa el EcR está asociado con dominios de unión a ADN y dominios de transactivación nativos o modificados en la misma molécula. Normalmente se usan USP o RXR como compañeros silenciosos. Los solicitantes han mostrado anteriormente que cuando los dominios de unión a ADN y los dominios de transactivación están en la misma molécula, la actividad de fondo en ausencia de un ligando es alta y que esta actividad disminuye drásticamente cuando los dominios de unión a ADN y los dominios de transactivación están en moléculas diferentes, es decir, en cada uno de los dos compañeros de un complejo heterodimérico u homodimérico (véase el documento PCT/US01/09050). Este sistema de dos híbridos también proporciona una sensibilidad mejorada a ligandos no esteroideos, por ejemplo, diacilhidrazinas, en comparación con ligandos esteroideos, por ejemplo, ponasterona A ("PonA") o muristerona A ("MurA"). Es decir, en comparación con los esteroideos, los ligandos no esteroideos proporcionan una actividad mayor a una concentración menor. Además, dado que la transactivación basada en interruptores génicos del EcR depende frecuentemente de la línea celular, es más fácil adaptar los sistemas de interruptores para obtener la máxima capacidad de transactivación para cada aplicación. Además, el sistema de dos híbridos evita algunos efectos secundarios debidos a la sobreexpresión de RXR que ocurren con frecuencia cuando se usa RXR no modificado como compañero de interruptor. En una realización específica del sistema de dos híbridos, se eliminan el dominio de unión a ADN y los dominios de transactivación nativos de EcR o RXR y, como consecuencia, estas moléculas quiméricas tienen menos posibilidades de interaccionar con otros receptores de hormonas esteroideas presentes en la línea celular, dando lugar a una reducción de los efectos secundarios.

Los solicitantes han mostrado anteriormente que un receptor de ecdisona en asociación con una proteína ultraespiráculo (USP) de díptero (mosca de la fruta *Drosophila melanogaster*) o de lepidóptero (gusano de las yemas de la picea *Choristoneura fumiferana*) se expresa de forma constitutiva en células de mamífero, mientras que un receptor de ecdisona en asociación con un receptor X retinoide (RXR) de vertebrado es inducible en células de

mamífero (en la solicitud en trámite PCT/US01/09050). Recientemente, los solicitantes realizaron el sorprendente descubrimiento de que la proteína ultraespiráculo de *Locusta migratoria* ("LmUSP") y el homólogo 1 de RXR y el homólogo 2 de RXR de la garrapata ixódida *Amblyomma americanum* ("AmaRXR1" y "AmaRXR2", respectivamente) y sus homólogos que no son de díptero ni de lepidóptero, incluidos, entre otros: un homólogo de RXR de cangrejo violinista *Celuca pugilator* ("CpRXR"), un homólogo de RXR de escarabajo *Tenebrio molitor* ("TmRXR"), un homólogo de RXR de abeja melífera *Apis mellifera* ("AmRXR") y un homólogo de RXR de pulgón *Myzus persicae* ("MpRXR"), todos los cuales se denominan conjuntamente en el presente documento RXR de invertebrado, pueden funcionar de forma similar al receptor X retinoide (RXR) de vertebrado en un sistema inducible de expresión génica basado en receptores de ecdisona inducible en células de mamífero (en la solicitud de EE. UU. presentada junto al presente documento).

Como se describe en el presente documento, los solicitantes han descubierto ahora que un dominio de unión a ligando de RXR quimérico que comprende al menos dos fragmentos polipeptídicos, en el que el primer fragmento polipeptídico es de RXR de una especie de vertebrado/invertebrado y el segundo fragmento polipeptídico es de RXR de una especie diferente de vertebrado/invertebrado, de modo que se produce un dominio de unión a ligando de RXR quimérico de vertebrado/invertebrado, un dominio de unión a ligando de RXR quimérico de vertebrado/vertebrado o un dominio de unión a ligando de RXR quimérico de invertebrado/invertebrado, puede funcionar en un sistema inducible de expresión génica basado en receptores de ecdisona. Sorprendentemente, el sistema inducible de expresión génica basado en EcR/RXR quiméricos novedoso de los solicitantes puede funcionar de forma similar a o mejor que el sistema de expresión génica basado en EcR/RXR de vertebrado (en el documento PCT/US01/09050) y el sistema de expresión génica basado en EcR/RXR de invertebrado (en la solicitud de EE. UU. presentada junto al presente documento) en términos de sensibilidad a ligando y magnitud de inducción génica. Por tanto, la presente invención proporciona un sistema inducible de expresión génica basado en EcR para su uso en células bacterianas, fúngicas, de levaduras, de animales y de mamíferos.

En particular, los solicitantes describen en el presente documento un sistema de dos híbridos novedosos que comprende un dominio de unión a ligando de RXR quimérico. Este sistema de expresión génica novedoso demuestra por primera vez que un polipéptido que comprende un dominio de unión a ligando de RXR quimérico puede funcionar como componente de un sistema inducible de expresión génica basado en EcR inducible en células de levadura y de mamífero. Como se analiza en el presente documento, este descubrimiento es tanto inesperado como sorprendente.

Específicamente, la invención de los solicitantes se refiere a un sistema de modulación de la expresión génica que comprende: a) un primer casete de expresión génica que se puede expresar en una célula huésped, en el que el primer casete de expresión génica comprende un polinucleótido que codifica un primer polipéptido híbrido que comprende i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y ii) un dominio de unión a ligando de receptor de ecdisona; y b) un segundo casete de expresión génica que se puede expresar en la célula huésped, en el que el segundo casete de expresión génica comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende i) un dominio de transactivación; y ii) un dominio de unión a ligando de receptor X retinoide quimérico.

La presente invención también se refiere a un sistema de modulación de la expresión génica que comprende: a) un primer casete de expresión génica que se puede expresar en una célula huésped, en el que el primer casete de expresión génica comprende un polinucleótido que codifica un primer polipéptido híbrido que comprende i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y ii) un dominio de unión a ligando de receptor X retinoide quimérico; y b) un segundo casete de expresión génica que se puede expresar en la célula huésped, en el que el segundo casete de expresión génica comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende i) un dominio de transactivación; y ii) un dominio de unión a ligando de receptor de ecdisona.

La presente invención también se refiere a un sistema de modulación de la expresión génica de acuerdo con la presente invención que además comprende c) un tercer casete de expresión génica que comprende: i) un elemento de respuesta al que se une el dominio de unión a ADN del primer polipéptido híbrido; ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y iii) un gen cuya expresión se quiere modular.

En una realización específica, el gen cuya expresión se quiere modular es un gen homólogo con respecto a la célula huésped. En otra realización específica, el gen cuya expresión se quiere modular es un gen heterólogo con respecto a la célula huésped.

Los ligandos para su uso en la presente invención como se describe a continuación, cuando se combinan con un dominio de unión a ligando de EcR y un dominio de unión a ligando de RXR quimérico, que a su vez están unidos al elemento de respuesta enlazado a un gen, proporcionan los medios para la regulación temporal externa de la expresión del gen. El mecanismo de unión o el orden en el que los diversos componentes de la presente invención se unen entre sí, es decir, por ejemplo, ligando a receptor, primer polipéptido híbrido a elemento de respuesta, segundo polipéptido híbrido a promotor, etc., no es crucial. La unión del ligando al dominio de unión a ligando del EcR y al dominio de unión a ligando del RXR quimérico permite la expresión o supresión del gen. Este mecanismo

no excluye el potencial de unión de ligando a EcR o RXR quimérico y la consiguiente formación de complejos homodiméricos activos (p. ej. EcR + EcR o RXR quimérico + RXR quimérico). Preferentemente, se varían uno o más de dominios del receptor, produciendo un interruptor génico híbrido. Normalmente, se pueden escoger uno o más de los tres dominios, DBD, LBD y el dominio de transactivación, de una fuente diferente de la fuente de los demás dominios de forma que los genes híbridos y las proteínas híbridas resultantes se optimizan en la célula u organismo huésped escogido para la actividad de transactivación, la unión complementaria del ligando y el reconocimiento de un elemento de respuesta específico. Además, el propio elemento de respuesta se puede modificar o sustituir con elementos de respuesta para otros dominios proteicos de unión a ADN tales como la proteína GAL-4 de levadura (véase Sadowski, *et al.* (1988), *Nature* 335: 563-564) o la proteína LexA de *Escherichia coli* (véase Brent y Ptashne (1985), *Cell* 43: 729-736) o elementos de respuesta sintéticos específicos para interacciones dirigidas con proteínas diseñadas, modificadas y seleccionadas para estas interacciones específicas (véase, por ejemplo, Kim, *et al.* (1997), *Proc. Natl. Acad. Sci., USA* 94: 3616-3620) para adaptar los receptores híbridos. Otra ventaja de los sistemas de dos híbridos es que permiten escoger un promotor usado para dirigir la expresión génica de acuerdo con un resultado final deseado. Este doble control puede ser particularmente importante en áreas de tratamiento génico, especialmente cuando se producen proteínas citotóxicas, porque se pueden controlar tanto el momento de la expresión como las células en las que se produce la expresión. Cuando se introducen genes, enlazados de manera funcional a un promotor adecuado, en las células del sujeto, la expresión de los genes exógenos se controla por la presencia del sistema de la presente invención. Los promotores se pueden regular de forma constitutiva o inducible o pueden ser específicos de tejido (es decir, expresados sólo en un tipo de células en particular) o específicos de determinadas etapas del desarrollo del organismo.

Casetes de expresión génica de la invención

El sistema inducible de expresión génica basado en EcR/RXR quiméricos novedoso de la invención comprende casetes de expresión génica que se pueden expresar en una célula huésped, en los que los casetes de expresión génica comprenden cada uno un polinucleótido que codifica un polipéptido híbrido. Así, la invención de los solicitantes también proporciona casetes de expresión génica para su uso en el sistema de expresión génica de la invención.

Específicamente, la presente invención proporciona un casete de expresión génica que comprende un polinucleótido que codifica un polipéptido híbrido. En particular, la presente invención proporciona un casete de expresión génica que se puede expresar en una célula huésped, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende o bien i) un dominio de unión a ADN que reconoce un elemento de respuesta o bien ii) un dominio de transactivación; y un dominio de unión a ligando de receptor de ecdisona o un dominio de unión a ligando de receptor X retinoide quimérico.

En una realización específica, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN que reconoce un elemento de respuesta y un dominio de unión a ligando de EcR.

En otra realización específica, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN que reconoce un elemento de respuesta y un dominio de unión a ligando de RXR quimérico.

En otra realización específica, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación y un dominio de unión a ligando de EcR.

En otra realización específica, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación y un dominio de unión a ligando de RXR quimérico.

En una realización preferida, el dominio de unión a ligando (LBD) es un LBD de EcR, un LBD de RXR quimérico o un LBD o un LBD quimérico de un miembro de la familia de receptores nucleares de hormonas esteroideas/tiroideas relacionado, o análogo, combinación o modificación de los mismos. En una realización específica, el LBD es un LBD de EcR o un LBD de RXR quimérico. En otra realización específica, el LBD es de un LBD truncado de EcR o un LBD truncado de RXR quimérico. Una mutación por truncamiento se puede realizar mediante cualquier método usado en la técnica, incluidos, entre otros, la digestión/delección con endonucleasas de restricción, la delección mediada por PCR/dirigida por oligonucleótidos, la mutagénesis química, la rotura de hebras de ADN y similares.

El EcR puede ser un EcR de invertebrado, preferentemente seleccionado de la clase de los artrópodos. Preferentemente, el EcR se selecciona del grupo que consiste en un EcR de lepidóptero, un EcR de díptero, un EcR de ortóptero, un EcR de homóptero y un EcR de hemíptero. Más preferentemente, el EcR para su uso es un EcR del gusano de las yemas de la picea *Choristoneura fumiferana* ("CfEcR"), un EcR del escarabajo *Tenebrio molitor* ("TmEcR"), un EcR de *Manduca sexta* ("MsEcR"), a EcR de *Heliothis virescens* ("HvEcR"), un EcR del mosquito pequeño *Chironomus tentans* ("CtEcR"), un EcR de la polilla de la seda *Bombyx mori* ("BmEcR"), un EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR"), un EcR del mosquito *Aedes aegypti* ("AaEcR"), un EcR de la moscarda *Lucilia capitata* ("LcEcR"), un EcR de la moscarda *Lucilia cuprina* ("LucEcR"), un EcR de la mosca mediterránea de la fruta *Ceratitis capitata* ("CcEcR"), un EcR de la langosta *Locusta migratoria* ("LniEcR"), un EcR del

pulgón *Myzus persicae* ("MpEcR"), un EcR del cangrejo violinista *Celuca pugilator* ("CpEcR"), un EcR de la garrapata ixódida *Amblyomma americanum* ("AmaEcR"), un EcR de la mosca blanca *Bemisia argentifoli* ("BaEcR", SEQ ID NO: 68) o un EcR de la cigarra *Nephotetix cincticeps* ("NcEcR", SEQ ID NO: 69). En una realización específica, el LBD es del EcR del gusano de las yemas de la picea (*Choristoneura fumiferana*) ("CfEcR") o del EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR").

En una realización específica, el LBD de EcR comprende dominios EF de longitud completa. En una realización preferida, los dominios EF de longitud completa están codificados por un polinucleótido que comprende una secuencia de ácidos nucleicos de la SEQ ID NO: 1 o la SEQ ID NO: 2.

En una realización específica, el LBD es de un LBD de EcR truncado. El truncamiento del LBD de EcR da lugar a una delección de al menos 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235 o 240 aminoácidos. En otra realización específica, el truncamiento del LBD de EcR da lugar a una delección de al menos un dominio polipeptídico parcial. En otra realización específica, el truncamiento del LBD de EcR da lugar a una delección de al menos un dominio polipeptídico entero. Más preferentemente, el truncamiento del polipéptido de EcR da lugar a una delección de al menos un dominio A/B, un dominio C, un dominio D, un dominio F, unos dominios A/B/C, unos dominios A/B/1/2-C, unos dominios A/B/C/D, unos dominios A/B/C/D/F, unos dominios A/B/F, unos dominios A/B/C/F, un dominio E parcial o un dominio F parcial. También se puede realizar una combinación de varias delecciones parciales y/o totales.

En una realización, el dominio de unión a ligando de receptor de ecdisona está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 1 (CfEcR-EF), SEQ ID NO: 2 (DmEcR-EF), SEQ ID NO: 3 (CfEcR-DE) y SEQ ID NO: 4 (DmEcR-DE).

En una realización preferida, el dominio de unión a ligando de receptor de ecdisona está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF), SEQ ID NO: 67 (DmEcR-DEF), SEQ ID NO: 71 (TmEcR-DEF) y SEQ ID NO: 73 (AmaEcR-DEF).

En una realización, el dominio de unión a ligando de receptor de ecdisona comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 5 (CfEcR-EF), SEQ ID NO: 6 (DmEcR-EF), SEQ ID NO: 7 (CfEcR-DE) y SEQ ID NO: 8 (DmEcR-DE).

En una realización preferida, el dominio de unión a ligando de receptor de ecdisona comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 72 (TmEcR-DEF) y SEQ ID NO: 74 (AmaEcR-DEF).

Preferentemente, el dominio de unión a ligando de RXR quimérico comprende al menos dos fragmentos polipeptídicos seleccionados del grupo que consiste en un fragmento polipeptídico de RXR de una especie de vertebrado, un fragmento polipeptídico de RXR de una especie de invertebrado y un fragmento polipeptídico de homólogo de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero. Un dominio de unión a ligando de RXR quimérico de acuerdo con la invención puede comprender al menos dos fragmentos polipeptídicos de RXR de especies diferentes, o cuando la especie es la misma, los dos o más fragmentos polipeptídicos pueden ser de dos o más isoformas diferentes del fragmento polipeptídico de RXR de la especie.

En una realización específica, el fragmento polipeptídico de RXR de una especie de vertebrado es de un RXR de ratón *Mus musculus* ("MmRXR") o un RXR de ser humano *Homo sapiens* ("HsRXR"). El polipéptido de RXR puede ser una isoforma RXR α , RXR β o RXR γ .

En una realización preferida, el fragmento polipeptídico de RXR de una especie de vertebrado es de un dominio EF de RXR de una especie de vertebrado codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 y SEQ ID NO: 14. En otra realización preferida, el fragmento polipeptídico de RXR de una especie de vertebrado es de un dominio EF de RXR de una especie de vertebrado que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19 y SEQ ID NO: 20.

En otra realización específica, el fragmento polipeptídico de RXR de una especie de invertebrado es de un polipéptido de espiráculo de la langosta *Locusta migratoria* ("LmUSP"), un homólogo 1 de RXR de la garrapata ixódida *Amblyomma americanum* ("AmaRXR1"), un homólogo 2 de RXR de la garrapata ixódida *Amblyomma americanum* ("AmaRXR2"), un homólogo de RXR del cangrejo violinista *Celuca pugilator* ("CpRXR"), un homólogo de RXR del escarabajo *Tenebrio molitor* ("TmRXR"), un homólogo de RXR de abeja melífera *Apis mellifera* ("AmRXR") y un homólogo de RXR del pulgón *Myzus persicae* ("MpRXR").

En una realización preferida, el fragmento polipeptídico de RXR de una especie de invertebrado es de un dominio EF

- de RXR de una especie de invertebrado codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25 y SEQ ID NO: 26. En otra realización preferida, el fragmento polipeptídico de RXR de una especie de invertebrado es de un dominio EF de RXR de una especie de invertebrado que comprende una
- 5 secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31 y SEQ ID NO: 32.
- En otra realización específica, el fragmento polipeptídico de RXR de una especie de invertebrado es de un homólogo de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero.
- 10 En una realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de vertebrado y un fragmento polipeptídico de RXR de una especie de invertebrado.
- 15 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de vertebrado y un fragmento polipeptídico de homólogo de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero.
- 20 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de invertebrado y un fragmento polipeptídico de homólogo de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero.
- 25 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de vertebrado y un fragmento polipeptídico de RXR de una especie de vertebrado diferente.
- 30 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de invertebrado y un fragmento polipeptídico de RXR de una especie de invertebrado diferente.
- 35 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende al menos un fragmento polipeptídico de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero y un fragmento polipeptídico de RXR de una especie de invertebrado que no es ni díptero ni lepidóptero diferente.
- 40 En una realización específica, el LBD de RXR quimérico comprende un dominio LBD de RXR que comprende al menos un fragmento polipeptídico seleccionado del grupo que consiste en una hélice 1 de dominio EF, una hélice 2 de dominio EF, una hélice 3 de dominio EF, una hélice 4 de dominio EF, una hélice 5 de dominio EF, una hélice 6 de dominio EF, una hélice 7 de dominio EF, una hélice 8 de dominio EF, una hélice 9 de dominio EF, una hélice 10 de dominio EF, una hélice 11 de dominio EF, una hélice 12 de dominio EF, un dominio F y un dominio EF (lámina β plegada, en el que el fragmento polipeptídico es de RXR de especies diferentes, es decir, quimérico con respecto al dominio LBD de RXR, de las del dominio LBD de RXR.
- 45 En otra realización específica, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-6, las hélices 1-7, las hélices 1-8, las hélices 1-9, las hélices 1-10, las hélices 1-11 o las hélices 1-12 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 7-12, las hélices 8-12, las hélices 9-12, las hélices 10-12, las hélices 11-12, la hélice 12 o el dominio F del RXR de una segunda especie de acuerdo con la invención, respectivamente.
- 50 En una realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-6 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 7-12 del RXR de una segunda especie de acuerdo con la invención.
- 55 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-7 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 8-12 del RXR de una segunda especie de acuerdo con la invención.
- 60 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-8 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 9-12 del RXR de una segunda especie de acuerdo con la invención.
- 65 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-9 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento

polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 10-12 del RXR de una segunda especie de acuerdo con la invención.

5 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-10 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 11-12 del RXR de una segunda especie de acuerdo con la invención.

10 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-11 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende la hélice 12 del RXR de una segunda especie de acuerdo con la invención.

15 En otra realización preferida, el primer fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende las hélices 1-12 del RXR de una primera especie de acuerdo con la invención, y el segundo fragmento polipeptídico del dominio de unión a ligando de RXR quimérico comprende un dominio F del RXR de una segunda especie de acuerdo con la invención.

20 En otra realización específica, el LBD de de un dominio de unión a ligando truncado de RXR quimérico. El truncamiento del LBD de RXR quimérico da lugar a una delección de al menos 1, 2, 3, 4, 5, 6, 8, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235 o 240 aminoácidos. Preferentemente, el truncamiento del LBD de RXR quimérico da lugar a una delección de al menos un dominio polipeptídico parcial. Más preferentemente, el truncamiento del LBD de RXR quimérico da lugar a una delección de al menos un dominio polipeptídico entero. En una realización preferida, el truncamiento del LBD de RXR quimérico da lugar a una delección de al menos un dominio E parcial, un dominio E completo, un dominio F parcial, un dominio F completo, una hélice 1 de dominio EF, una hélice 2 de dominio EF, una hélice 3 de dominio EF, una hélice 4 de dominio EF, una hélice 5 de dominio EF, una hélice 6 de dominio EF, una hélice 7 de dominio EF, una hélice 8 de dominio EF, una hélice 9 de dominio EF, una hélice 10 de dominio EF, una hélice 11 de dominio EF, una hélice 12 de dominio EF o una lámina β plegada de dominio EF. También se puede realizar una combinación de varias delecciones parciales y/o totales.

35 En una realización preferida, el dominio de unión a ligando truncado de RXR quimérico está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos de la SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37 o SEQ ID NO: 38. En otra realización preferida, el dominio de unión a ligando truncado de RXR quimérico comprende una secuencia de ácidos nucleicos de la SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 42, SEQ ID NO: 43, o SEQ ID NO: 44.

40 En una realización preferida, el dominio de unión a ligando de RXR quimérico está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en a) la SEQ ID NO: 45, b) los nucleótidos 1-348 de la SEQ ID NO: 13 y los nucleótidos 268-630 de la SEQ ID NO: 21, c) los nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, d) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21, e) los nucleótidos 1-555 de la SEQ ID NO: 13 y los nucleótidos 490-630 de la SEQ ID NO: 21, f) los nucleótidos 1-624 de la SEQ ID NO: 13 y los nucleótidos 547-630 de la SEQ ID NO: 21, g) los nucleótidos 1-645 de la SEQ ID NO: 13 y los nucleótidos 601-630 de la SEQ ID NO: 21 y h) los nucleótidos 1-717 de la SEQ ID NO: 13 y los nucleótidos 613-630 de la SEQ ID NO: 21.

50 En otra realización preferida, el dominio de unión a ligando de RXR quimérico comprende una secuencia de aminoácidos seleccionada del grupo que consiste en a) la SEQ ID NO: 46, b) los aminoácidos 1-116 de la SEQ ID NO: 13 y los aminoácidos 90-210 de la SEQ ID NO: 21, c) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, d) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21, e) los aminoácidos 1-185 de la SEQ ID NO: 13 y los aminoácidos 164-210 de la SEQ ID NO: 21, f) los aminoácidos 1-208 de la SEQ ID NO: 13 y los aminoácidos 183-210 de la SEQ ID NO: 21, g) los aminoácidos 1-215 de la SEQ ID NO: 13 y los aminoácidos 201-210 de la SEQ ID NO: 21 y h) los aminoácidos 1-239 de la SEQ ID NO: 13 y los aminoácidos 205-210 de la SEQ ID NO: 21.

60 Para los propósitos de la presente invención, EcR, RXR de vertebrado, RXR de invertebrado y RXR quimérico también incluyen EcR, RXR de vertebrado, RXR de invertebrado y RXR quiméricos sintéticos e híbridos y sus homólogos.

65 El dominio de unión a ADN puede ser cualquier dominio de unión a ADN con un elemento de respuesta conocido, incluidos dominios de unión a ADN sintéticos y quiméricos, o análogo, combinación o modificación de los mismos. Preferentemente, el DBD es un DBD de GAL4, un DBD de LexA, un DBD de un factor de transcripción, un DBD de un miembro de la superfamilia de receptores nucleares de hormonas esteroideas/tiroideas, un DBD de LacZ bacteriano o un DBD de put de levadura. Más preferentemente, el DBD es un DBD de GAL4 [SEQ ID NO: 47 (polinucleótido) o SEQ ID NO: 48 (polipéptido)] o un DBD de LexA [(SEQ ID NO: 49 (polinucleótido) o SEQ ID NO:

50 (polipéptido)].

El dominio de transactivación (abreviado "AD" o "TA") puede ser cualquier AD de receptor nuclear de hormonas esteroideas/tiroideas, AD sintético o quimérico, AD de poliglutamina, AD de aminoácidos básico o ácidos, un AD de VP16, un AD de GAL4, un AD de NF-kB, un AD de BP64, un dominio de activación ácido de B42 (B42AD), o un análogo, combinación o modificación de los mismos. En una realización específica, el AD es un AD sintético o quimérico o se obtiene a partir de un AD de VP16, GAL4, NF-kB o un dominio de activación ácido de B42. Preferentemente, el AD es un AD de VP16 [SEQ ID NO: 51 (polinucleótido) o SEQ ID NO: 52 (polipéptido)] o un AD de B42 [SEQ ID NO: 53 (polinucleótido) o SEQ ID NO: 54 (polipéptido)].

En una realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 47) y un DBD de LexA (SEQ ID NO: 49) y un dominio de unión a ligando de EcR codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF), SEQ ID NO: 67 (DmEcR-DEF), SEQ ID NO: 71 (TmEcR-DEF) y SEQ ID NO: 73 (AmaEcR-DEF).

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 48) y un DBD de LexA (SEQ ID NO: 50) y un dominio de unión a ligando de EcR que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 72 (TmEcR-DEF) y SEQ ID NO: 74 (AmaEcR-DEF).

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 47) y un DBD de LexA (SEQ ID NO: 49) y un dominio de unión a ligando de RXR quimérico codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en a) la SEQ ID NO: 45, b) los nucleótidos 1-348 de la SEQ ID NO: 13 y los nucleótidos 268-630 de la SEQ ID NO: 21, c) los nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, d) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21, e) los nucleótidos 1-555 de la SEQ ID NO: 13 y los nucleótidos 490-630 de la SEQ ID NO: 21, f) los nucleótidos 1-624 de la SEQ ID NO: 13 y los nucleótidos 547-630 de la SEQ ID NO: 21, g) los nucleótidos 1-645 de la SEQ ID NO: 13 y los nucleótidos 601-630 de la SEQ ID NO: 21 y h) los nucleótidos 1-717 de la SEQ ID NO: 13 y los nucleótidos 613-630 de la SEQ ID NO: 21.

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de unión a ADN que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 48) y un DBD de LexA (SEQ ID NO: 50) y un dominio de unión a ligando de RXR quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en a) la SEQ ID NO: 46, b) los aminoácidos 1-116 de la SEQ ID NO: 13 y los aminoácidos 90-210 de la SEQ ID NO: 21, c) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, d) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21, e) los aminoácidos 1-185 de la SEQ ID NO: 13 y los aminoácidos 164-210 de la SEQ ID NO: 21, f) los aminoácidos 1-208 de la SEQ ID NO: 13 y los aminoácidos 183-210 de la SEQ ID NO: 21, g) los aminoácidos 1-215 de la SEQ ID NO: 13 y los aminoácidos 201-210 de la SEQ ID NO: 21, y h) los aminoácidos 1-239 de la SEQ ID NO: 13 y los aminoácidos 205-210 de la SEQ ID NO: 21.

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos de la SEQ ID NO: 51 o la SEQ ID NO: 53 y un dominio de unión a ligando de EcR codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en la SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF), SEQ ID NO: 67 (DmEcR-DEF), SEQ ID NO: 71 (TmEcR-DEF) y SEQ ID NO: 73 (AmaEcR-DEF).

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación que comprende una secuencia de aminoácidos de la SEQ ID NO: 52 o la SEQ ID NO: 54 y un dominio de unión a ligando de EcR que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 72 (TmEcR-DEF) y SEQ ID NO: 74 (AmaEcR-DEF).

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos de la SEQ ID NO: 51 o la SEQ ID NO: 53 y un dominio de unión a ligando de RXR quimérico codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en a) la SEQ ID NO: 45, b) los nucleótidos 1-348 de la SEQ ID NO: 13 y los nucleótidos 268-630 de la SEQ ID NO: 21, c) los

nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, d) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21, e) los nucleótidos 1-555 de la SEQ ID NO: 13 y los nucleótidos 490-630 de la SEQ ID NO: 21, f) los nucleótidos 1-624 de la SEQ ID NO: 13 y los nucleótidos 547-630 de la SEQ ID NO: 21, g) los nucleótidos 1-645 de la SEQ ID NO: 13 y los nucleótidos 601-630 de la SEQ ID NO: 21 y h) los nucleótidos 1-717 de la SEQ ID NO: 13 y los nucleótidos 613-630 de la SEQ ID NO: 21.

En otra realización preferida, el casete de expresión génica codifica un polipéptido híbrido que comprende un dominio de transactivación que comprende una secuencia de aminoácidos de la SEQ ID NO: 52 o la SEQ ID NO: 54 y un dominio de unión a ligando de RXR quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en a) la SEQ ID NO: 46, b) los aminoácidos 1-116 de la SEQ ID NO: 13 y los aminoácidos 90-210 de la SEQ ID NO: 21, c) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, d) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21, e) los aminoácidos 1-185 de la SEQ ID NO: 13 y los aminoácidos 164-210 de la SEQ ID NO: 21, f) los aminoácidos 1-208 de la SEQ ID NO: 13 y los aminoácidos 183-210 de la SEQ ID NO: 21, g) los aminoácidos 1-215 de la SEQ ID NO: 13 y los aminoácidos 201-210 de la SEQ ID NO: 21 y h) los aminoácidos 1-239 de la SEQ ID NO: 13 y los aminoácidos 205-210 de la SEQ ID NO: 21.

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 de los mismos. En la presente invención se puede emplear un único RE o varios RE, ya sean varias copias del mismo RE o dos o más RE diferentes. En una realización específica, el RE es un RE de GAL4 ("GAL4RE"), LexA, un RE de un receptor nuclear de hormonas esteroideas/tiroideas o un RE sintético que reconoce un dominio de unión a ADN sintético. Preferentemente, el RE es un GAL4RE que comprende una secuencia polinucleotídica de la SEQ ID NO: 55 o un LexARE (operón "op") que comprende una secuencia polinucleotídica de la SEQ ID NO: 56 (2XLexAop). Preferentemente, la primera proteína híbrida no tiene, sustancialmente, un dominio de transactivación y la segunda proteína híbrida no tiene, sustancialmente, un dominio de unión a ADN. Para los propósitos de la presente invención, "no tiene, sustancialmente" significa que la proteína en cuestión no contiene una secuencia suficiente del dominio en cuestión para proporcionar una actividad de activación o unión.

Por tanto, la presente invención también se refiere a un casete de expresión génica que comprende: i) un elemento de respuesta que comprende un dominio al que se une un polipéptido que comprende un dominio de unión a ADN; ii) un promotor que se activa por un polipéptido que comprende un dominio de transactivación; y iii) un gen cuya expresión se quiere modular.

Los genes de interés para su uso en los casetes de expresión génica de los solicitantes pueden ser genes endógenos o genes heterólogos. La información de la secuencia de ácidos nucleicos o de aminoácidos para una proteína o gen deseado se puede localizar en una de las muchas bases de datos de acceso público, por ejemplo, GENBANK, EMBL, Swiss-Prot y PIR, o en muchas publicaciones periódicas relacionadas con la biología. Así, los expertos en la técnica tienen acceso a la información de secuencia de ácidos nucleicos para prácticamente todos los genes conocidos. Después, se puede usar esta información para construir las construcciones deseadas para la inserción del gen de interés en el interior de los casetes de expresión génica usados en los métodos de los solicitantes descritos en el presente documento.

Los ejemplos de genes de interés para su uso en los casetes de expresión génica de los solicitantes incluyen, entre otros: genes que codifican polipéptidos terapéuticamente deseables o productos que se pueden usar para tratar una afección, una enfermedad, un trastorno, una disfunción, un defecto genético, tales como anticuerpos monoclonales, enzimas, proteasas, citocinas, interferones, insulina, eritropoyetina, factores de coagulación, otros factores o componentes sanguíneos, vectores víricos para tratamiento génico, virus para vacunas, objetivos para el descubrimiento de fármacos, genómica funcional y aplicaciones de análisis de proteómica y similares.

Polinucleótidos de la invención

El sistema inducible de expresión génica basado en receptores X retinoides quiméricos/receptores de ecdisona novedoso de la invención comprende un casete de expresión génica que comprende un polinucleótido que codifica un polipéptido híbrido que comprende a) un dominio de unión a ADN o un dominio de transactivación y b) un dominio de unión a ligando de EcR o un dominio de unión a ligando de RXR quimérico. Estos casetes de expresión génica, los polinucleótidos que comprenden y los polipéptidos híbridos que codifican, son útiles como componentes de un sistema de expresión génica basado en EcR para modular la expresión de un gen en una célula huésped.

Por tanto, la presente invención proporciona un polinucleótido aislado que codifica un polipéptido híbrido que comprende a) un dominio de unión a ADN o un dominio de transactivación de acuerdo con la invención y b) un dominio de unión a ligando de EcR o un dominio de unión a ligando de RXR quimérico de acuerdo con la invención.

La presente invención también se refiere a un polinucleótido aislado que codifica un dominio de unión a ligando de RXR quimérico de acuerdo con la invención.

La presente invención también se refiere a un polinucleótido aislado que codifica un LBD truncado de EcR o un LBD truncado de RXR quimérico que comprende una mutación por truncamiento de acuerdo con la invención. Específicamente, la presente invención se refiere a un polinucleótido aislado que codifica un dominio de unión a ligando truncado de EcR o RXR quimérico que comprende una mutación por truncamiento que afecta a la actividad de unión a ligando o a la sensibilidad a ligando que es útil en la modulación de la expresión génica en una célula huésped.

En una realización específica, el polinucleótido aislado que codifica un LBD de EcR comprende una secuencia polinucleotídica seleccionada del grupo que consiste en la SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF), SEQ ID NO: 67 (DmEcR-DEF), SEQ ID NO: 71 (TmEcR-DEF) y SEQ ID NO: 73 (AmaEcR-DEF).

En otra realización específica, el polinucleótido aislado codifica un LBD de EcR que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 72 (TmEcR-DEF) y SEQ ID NO: 74 (AmaEcR-DEF).

En otra realización específica, el polinucleótido aislado que codifica un LBD de RXR quimérico comprende una secuencia polinucleotídica seleccionada del grupo que consiste en la SEQ ID NO: 45, b) los nucleótidos 1-348 de la SEQ ID NO: 13 y los nucleótidos 268-630 de la SEQ ID NO: 21, c) los nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, d) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21, e) los nucleótidos 1-555 de la SEQ ID NO: 13 y los nucleótidos 490-630 de la SEQ ID NO: 21, f) los nucleótidos 1-624 de la SEQ ID NO: 13 y los nucleótidos 547-630 de la SEQ ID NO: 21, g) los nucleótidos 1-645 de la SEQ ID NO: 13 y los nucleótidos 601-630 de la SEQ ID NO: 21 y h) los nucleótidos 1-717 de la SEQ ID NO: 13 y los nucleótidos 613-630 de la SEQ ID NO: 21.

En otra realización específica, el polinucleótido aislado codifica un LBD de RXR quimérico que comprende una secuencia de aminoácidos que consiste en a) la SEQ ID NO: 46, b) los aminoácidos 1-116 de la SEQ ID NO: 13 y los aminoácidos 90-210 de la SEQ ID NO: 21, c) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, d) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21, e) los aminoácidos 1-185 de la SEQ ID NO: 13 y los aminoácidos 164-210 de la SEQ ID NO: 21, f) los aminoácidos 1-208 de la SEQ ID NO: 13 y los aminoácidos 183-210 de la SEQ ID NO: 21, g) los aminoácidos 1-215 de la SEQ ID NO: 13 y los aminoácidos 201-210 de la SEQ ID NO: 21 y h) los aminoácidos 1-239 de la SEQ ID NO: 13 y los aminoácidos 205-210 de la SEQ ID NO: 21.

En particular, la presente invención se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento, en el que la mutación reduce la actividad de unión a ligando o la sensibilidad a ligando del LBD truncado de RXR quimérico. En una realización específica, la presente invención se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento que reduce la actividad de unión a esteroides o la sensibilidad a esteroides del LBD truncado de RXR quimérico.

En otra realización específica, la presente invención se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento que reduce la actividad de unión a no esteroide o la sensibilidad a no esteroide del LBD truncado de RXR quimérico.

La presente invención también se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento, en el que la mutación potencia la actividad de unión a ligando o la sensibilidad a ligando del LBD truncado de RXR quimérico. En una realización específica, la presente invención se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento que potencia la actividad de unión a esteroides o la sensibilidad a esteroides del LBD truncado de RXR quimérico.

En otra realización específica, la presente invención se refiere a un polinucleótido aislado que codifica un LBD truncado de RXR quimérico que comprende una mutación por truncamiento que potencia la actividad de unión a no esteroide o la sensibilidad a no esteroide del LBD truncado de RXR quimérico.

La presente invención también se refiere a un polinucleótido aislado que codifica un LBD truncado de receptor X retinoide quimérico que comprende una mutación por truncamiento que aumenta la sensibilidad a ligando de un heterodímero que comprende el LBD truncado de receptor X retinoide quimérico y un compañero de dimerización. En una realización específica, el compañero de dimerización es un polipéptido de receptor de ecdisona. Preferentemente, el compañero de dimerización es un polipéptido truncado de EcR. Más preferentemente, el compañero de dimerización es un polipéptido de EcR en el que se han eliminado los dominios A/B. Incluso más preferentemente, el compañero de dimerización es un polipéptido de EcR que comprende una secuencia de aminoácidos de la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 72 (TmEcR-DEF) o SEQ ID NO: 74 (AmaEcR-DEF).

Polipéptidos de la invención

- El sistema inducible de expresión génica basado en receptores X retinoides quiméricos/receptores de ecdisona novedoso de la invención comprende un casete de expresión génica que comprende un polinucleótido que codifica un polipéptido híbrido que comprende a) un dominio de unión a ADN o un dominio de transactivación y b) un dominio de unión a ligando de EcR o un dominio de unión a ligando de RXR quimérico. Estos casetes de expresión génica, los polinucleótidos que comprenden y los polipéptidos híbridos que codifican, son útiles como componentes de un sistema de expresión génica basado en EcR/RXR quiméricos para modular la expresión de un gen en una célula huésped.
- Por tanto, la presente invención también se refiere a un polipéptido híbrido que comprende a) un dominio de unión a ADN o un dominio de transactivación de acuerdo con la invención y b) un dominio de unión a ligando de EcR o un dominio de unión a ligando de RXR quimérico de acuerdo con la invención.
- La presente invención también se refiere a un polipéptido aislado que comprende un dominio de unión a ligando de RXR quimérico de acuerdo con la invención.
- La presente invención también se refiere a un LBD truncado de EcR aislado o un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento de acuerdo con la invención. Específicamente, la presente invención se refiere a un LBD truncado de EcR aislado o un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que afecta a la actividad de unión a ligando o a la sensibilidad a ligando.
- En una realización específica, el polipéptido de LBD de EcR aislado está codificado por un polinucleótido que comprende una secuencia polinucleotídica seleccionada del grupo que consiste en la SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF), SEQ ID NO: 67 (DmEcR-DEF), SEQ ID NO: 71 (TmEcR-DEF) y SEQ ID NO: 73 (AmaEcR-DEF).
- En otra realización específica, el polipéptido de LBD de EcR aislado comprende una secuencia de aminoácidos seleccionada del grupo que consiste en la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 72 (TmEcR-DEF) y SEQ ID NO: 74 (AmaEcR-DEF).
- En otra realización específica, el LBD truncado de RXR quimérico aislado está codificado por un polinucleótido que comprende una secuencia polinucleotídica seleccionada del grupo que consiste en la SEQ ID NO: 45, b) los nucleótidos 1-348 de la SEQ ID NO: 13 y los nucleótidos 268-630 de la SEQ ID NO: 21, c) los nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, d) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21, e) los nucleótidos 1-555 de la SEQ ID NO: 13 y los nucleótidos 490-630 de la SEQ ID NO: 21, f) los nucleótidos 1-624 de la SEQ ID NO: 13 y los nucleótidos 547-630 de la SEQ ID NO: 21, g) los nucleótidos 1-645 de la SEQ ID NO: 13 y los nucleótidos 601-630 de la SEQ ID NO: 21 y h) los nucleótidos 1-717 de la SEQ ID NO: 13 y los nucleótidos 613-630 de la SEQ ID NO: 21.
- En otra realización preferida, el LBD truncado de RXR quimérico aislado comprende una secuencia de aminoácidos seleccionada del grupo que consiste en a) la SEQ ID NO: 46, b) los aminoácidos 1-116 de la SEQ ID NO: 13 y los aminoácidos 90-210 de la SEQ ID NO: 21, c) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, d) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21, e) los aminoácidos 1-185 de la SEQ ID NO: 13 y los aminoácidos 164-210 de la SEQ ID NO: 21, f) los aminoácidos 1-208 de la SEQ ID NO: 13 y los aminoácidos 183-210 de la SEQ ID NO: 21, g) los aminoácidos 1-215 de la SEQ ID NO: 13 y los aminoácidos 201-210 de la SEQ ID NO: 21 y h) los aminoácidos 1-239 de la SEQ ID NO: 13 y los aminoácidos 205-210 de la SEQ ID NO: 21.
- La presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que reduce la actividad de unión a ligando o la sensibilidad a ligando de dicho LBD truncado de RXR quimérico.
- Por tanto, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que reduce la actividad de unión a ligando o la sensibilidad a ligando de dicho LBD truncado de RXR quimérico.
- En una realización específica, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que reduce la actividad de unión a esteroides o la sensibilidad a esteroides del LBD truncado de RXR quimérico.
- En otra realización específica, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que reduce la actividad de unión a no esteroide o la sensibilidad a no esteroide del LBD truncado de RXR quimérico.
- Además, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que potencia la actividad de unión a ligando o la sensibilidad a ligando del LBD truncado

de RXR quimérico.

La presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que potencia la actividad de unión a ligando o la sensibilidad a ligando del LBD truncado de RXR quimérico. En una realización específica, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que potencia la actividad de unión a esteroides o la sensibilidad a esteroides del LBD truncado de RXR quimérico.

En otra realización específica, la presente invención se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que potencia la actividad de unión a no esteroide o la sensibilidad a no esteroide del LBD truncado de RXR quimérico.

La presente invención también se refiere a un LBD truncado de RXR quimérico aislado que comprende una mutación por truncamiento que aumenta la sensibilidad de un heterodímero que comprende el LBD truncado de RXR quimérico y un compañero de dimerización.

En una realización específica, el compañero de dimerización es un polipéptido de receptor de ecdisona. Preferentemente, el compañero de dimerización es un polipéptido truncado de EcR. Preferentemente, el compañero de dimerización es un polipéptido de EcR en el que se han eliminado los dominios A/B o A/B/C. Incluso más preferentemente, el compañero de dimerización es un polipéptido de EcR que comprende una secuencia de aminoácidos de la SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 58 (DmEcR-DEF), SEQ ID NO: 70 (CfEcR-CDEF), SEQ ID NO: 72 (TmEcR-DEF) o SEQ ID NO: 74 (AmaEcR-DEF).

Método de modulación de la expresión génica de la invención

La invención de los solicitantes también se refiere a métodos de modulación de la expresión génica en una célula huésped usando un sistema de modulación de la expresión génica de acuerdo con la invención. Específicamente, la invención de los solicitantes proporciona un método de modulación de la expresión de un gen en una célula huésped que comprende las etapas de: a) introducir en la célula huésped un sistema de modulación de la expresión génica de acuerdo con la invención; y b) introducir en la célula huésped un ligando; en el que el gen que se quiere modular es un componente de un casete de expresión génica que comprende: i) un elemento de respuesta que comprende un dominio reconocido por el dominio de unión a ADN del primer polipéptido híbrido; ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y iii) un gen cuya expresión se quiere modular, de modo que, tras la introducción del ligando en la célula huésped, se modula la expresión del gen.

La invención también proporciona un método de modulación de la expresión de un gen en una célula huésped que comprende las etapas de: a) introducir en la célula huésped un sistema de modulación de la expresión génica de acuerdo con la invención; b) introducir en la célula huésped un casete de expresión génica que comprende i) un elemento de respuesta que comprende un dominio reconocido por el dominio de unión a ADN del primer polipéptido híbrido; ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y iii) un gen cuya expresión se quiere modular; y c) introducir en la célula huésped un ligando; de modo que se modula la expresión del gen en la célula huésped.

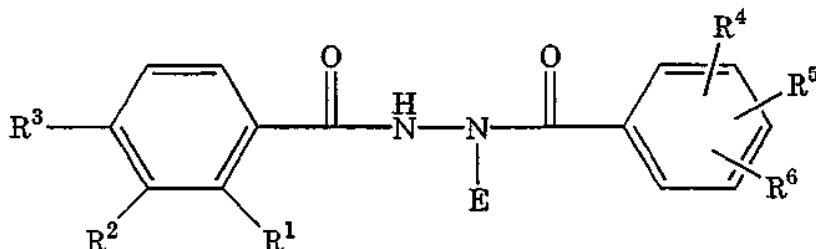
Los genes de interés para su expresión en una célula huésped usando los métodos de los solicitantes pueden ser genes endógenos o genes heterólogos. La información de la secuencia de ácidos nucleicos o de aminoácidos para una proteína o gen deseado se puede localizar en una de las muchas bases de datos de acceso público, por ejemplo, GENBANK, EMBL, Swiss-Prot y PIR, o en muchas publicaciones periódicas relacionadas con la biología. Así, los expertos en la técnica tienen acceso a la información de secuencia de ácidos nucleicos para prácticamente todos los genes conocidos. Después, se puede usar esta información para construir las construcciones deseadas para la inserción del gen de interés en el interior de los casetes de expresión génica usados en los métodos de los solicitantes descritos en el presente documento.

Los ejemplos de genes de interés para su uso en una célula huésped usando los métodos de los solicitantes incluyen, entre otros: genes que codifican polipéptidos terapéuticamente deseables o productos que se pueden usar para tratar una afección, una enfermedad, un trastorno, una disfunción, un defecto genético, tales como anticuerpos monoclonales, enzimas, proteasas, citocinas, interferones, insulina, eritropoyetina, factores de coagulación, otros factores o componentes sanguíneos, vectores víricos para tratamiento génico, virus para vacunas, objetivos para el descubrimiento de fármacos, genómica funcional y aplicaciones de análisis de proteómica y similares.

Son ligandos aceptables todos los que modulan la expresión del gen cuando la unión del dominio de unión a ADN del sistema de dos híbridos al elemento de respuesta en presencia del ligando da lugar a la activación o supresión de la expresión de los genes. Los ligandos preferidos incluyen ponasterona, muristerona A, ácido 9-cis-retinoico, análogos sintéticos del ácido retinoico, N,N'-diacilhidrazinas tales como las divulgadas en las patentes de EE. UU. N.º 6013836, 5117057, 5530028 y 5378726; dibenzoilalquil cianohidrazinas tales como las divulgadas en la solicitud europea N.º 461809; N-alquil-N,N'-diaroilhidrazinas tales como la divulgadas en la patente de EE. UU. N.º 5225443; N-acil-N-alquilcarbonilhidrazinas tales como las divulgadas

en la solicitud europea N.º 234994; N-aróil-N-alquil-N'-aróilhidrazinas tales como las descritas en la patente de EE. UU. N.º 4985461; y otros materiales similares, incluidos 3,5-di-terc-butil-4-hidroxi-N-isobutil-benzamida, 8-O-acetilharpágido y similares.

5 En una realización preferida, el ligando para su uso en el método de los solicitantes de modulación de la expresión de un gen es un compuesto de la fórmula:



10 en la que:

E es un alquilo (C₄-C₆) que contiene un carbono terciario o un cianoalquilo (C₃-C₅) que contiene un carbono terciario; R¹ es H, Me, Et, i-Pr, F, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OH, OMe, OEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, SCN o SCHF₂;

R² es H, Me, Et, n-Pr, i-Pr, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe₂, NEt₂, SMe, SEt, SOCF₃, OCF₂CF₂H, COEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, OCF₃, OCHF₂, O-i-Pr, SCN, SCHF₂, SOMe, NH-CN, o se une con R³ y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R³ es H, Et, o se une con R² y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R⁴, R⁵ y R⁶ son independientemente H, Me, Et, F, Cl, Br, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OMe, OEt, SMe o SEt.

30 En otra realización preferida, se puede usar un segundo ligando además del primer ligando analizado anteriormente en el método de los solicitantes de modulación de la expresión de un gen, en el que el segundo ligando es ácido 9-cis-retinoico o un análogo sintético del ácido retinoico.

35 La invención de los solicitantes permite la modulación de la expresión génica en células huésped procariontas y eucariotas. Por tanto, la presente invención también se refiere a un método de modulación de la expresión génica en una célula huésped seleccionada del grupo que consiste en una célula bacteriana, una célula fúngica, una célula de levadura, una célula animal y una célula de mamífero. Preferentemente, la célula huésped es una célula de levadura, una célula de hámster, una célula de ratón, una célula de mono o una célula de ser humano.

40 La expresión en células huésped transgénicas puede ser útil para la expresión de diversos polipéptidos de interés, incluidos, entre otros, polipéptidos terapéuticos, intermedios de rutas; para la modulación de rutas ya existentes en el huésped para la síntesis de nuevos productos que hasta ahora era imposible usando el huésped; ensayos basados en células; ensayos de genómica funcional, producción de proteínas bioterapéuticas, ensayos de proteómica y similares. Adicionalmente, los productos génicos pueden ser útiles para conferir mayores rendimientos de crecimiento del huésped o para permitir que se utilice un modo de crecimiento alternativo.

Células huésped y organismos no humanos de la invención

50 Como se describe anteriormente, el sistema de modulación de la expresión génica de la presente invención se puede usar para modular la expresión génica en una célula huésped. La expresión en células huésped transgénicas puede ser útil para la expresión de diversos genes de interés. Por tanto, la invención de los solicitantes proporciona una célula huésped aislada que comprende un sistema de expresión génica de acuerdo con la invención. La presente invención también proporciona una célula huésped aislada que comprende un casete de expresión génica de acuerdo con la invención. La invención de los solicitantes también proporciona una célula huésped aislada que comprende un polinucleótido o un polipéptido de acuerdo con la invención. La célula huésped aislada puede ser una célula huésped procarionta o eucariota.

Preferentemente, la célula huésped se selecciona del grupo que consiste en una célula bacteriana, una célula fúngica, una célula de levadura, una célula animal y una célula de mamífero. Los ejemplos de células huésped preferidas incluyen, entre otras, células huésped de especies fúngicas o de levaduras tales como *Aspergillus*, *Trichoderma*, *Saccharomyces*, *Pichia*, *Candida*, *Hansenula*, o de especies bacterianas tales como las de los géneros *Synechocystis*, *Synechococcus*, *Salmonella*, *Bacillus*, *Acinetobacter*, *Rhodococcus*, *Streptomyces*, *Escherichia*, *Pseudomonas*, *Methylomonas*, *Methylobacter*, *Alcaligenes*, *Synechocystis*, *Anabaena*, *Thiobacillus*, *Methanobacterium* y *Klebsiella*, animales y de mamífero.

En una realización específica, la célula huésped es una célula de levadura seleccionada del grupo que consiste en una célula huésped de *Saccharomyces*, de *Pichia* y de *Candida*.

En otra realización específica, la célula huésped es una célula de hámster.

En otra realización específica, la célula huésped es una célula murina.

En otra realización específica, la célula huésped es una célula de mono.

En otra realización específica, la célula huésped es una célula de ser humano.

La transformación de células huésped se conoce bien en la técnica y se puede lograr mediante una variedad de métodos, incluidos, entre otros, electroporación, infección vírica, transfección de plásmido/vector, transfección mediada por vector no vírico, bombardeo de partículas y similares. La expresión de 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 procariontes y eucariotas son bien conocidos en la técnica (véanse la sección de métodos generales de los ejemplos). Se pueden recoger las células y aislar los productos génicos de acuerdo con protocolos específicos para el producto génico.

Además, se puede escoger una célula huésped que module la expresión del polinucleótido insertado o que modifique y procese el producto polipeptídico de la manera específica deseada. Las diferentes células huésped presentan características y mecanismos específicos para el procesamiento y la modificación traduccional y postraduccional [p. ej., glucosilación, escisión (p. ej., de la secuencia señal)] de proteínas. Se pueden escoger líneas celulares o sistemas huéspedes apropiados para garantizar la modificación y el procesamiento deseados de la proteína exógena expresada. Por ejemplo, se puede usar la expresión en un sistema bacteriano para producir un producto de proteína del núcleo no glucosilada. Sin embargo, un polipéptido expresado en bacterias puede no plegarse correctamente. La expresión en levaduras puede producir un producto glucosilado. La expresión en células eucariotas puede aumentar la probabilidad de glucosilación y plegamiento "nativo" de una proteína heteróloga. Además, la expresión en células de mamífero puede proporcionar una herramienta para reconstituir, o constituir, la actividad del polipéptido. Además, los diferentes sistemas de expresión de vector/huésped pueden afectar a las reacciones de procesamiento, tales como escisiones proteolíticas, en diferente medida.

La invención de los solicitantes también se refiere a un organismo no humano que comprende una célula huésped aislada de acuerdo con la invención. Preferentemente, el organismo no humano se selecciona del grupo que consiste en una bacteria, un hongo, una levadura, un animal y un mamífero. Más preferentemente, el organismo no humano es una levadura, un ratón, una rata, un conejo, un gato, un perro, un bóvido, una cabra, un cerdo, un caballo, una oveja, un mono o un chimpancé.

En una realización específica, el organismo no humano es una levadura seleccionada del grupo que consiste en *Saccharomyces*, *Pichia*, y *Candida*.

En otra realización específica, el organismo no humano es un ratón *Mus musculus*.

Medida de la transcripción/expresión génica

Una medida útil de los métodos de la invención de los solicitantes es la del estado transcripcional de la célula, incluidas las identidades y abundancias de ARN, preferentemente de especies de ARNm. Estas medidas se llevan a cabo convenientemente midiendo las abundancias de ADNc mediante cualquiera de las diversas tecnologías de expresión génica existentes.

La tecnología de matriz de ácido nucleico es una técnica útil para determinar la expresión diferencial de ARNm. Esta tecnología incluye, por ejemplo, chips oligonucleotídicos y micromatrices de ADN. Estas técnicas se basan en fragmentos de ADN u oligonucleotídos que corresponden a diferentes genes o ADNc que están inmovilizados sobre un soporte sólido e hibridados con sondas preparadas a partir de conjuntos de ARNm extraído a partir de células, tejidos u organismos completos y convertidos en ADNc. Los chips oligonucleotídicos son matrices de oligonucleotídos sintetizadas sobre un sustrato usando técnicas fotolitográficas. Se han producido chips que pueden analizar hasta 1700 genes. Las micromatrices de ADN son matrices de muestras de ADN, normalmente productos

de PCR, que se imprimen robóticamente sobre un portaobjetos de microscopio. Cada gen se analiza mediante una secuencia de ADN de longitud completa o parcial. Actualmente se preparan comercialmente de forma rutinaria micromatrices con hasta 10.000 genes. La principal diferencia entre estas dos técnicas es que, normalmente, los chips oligonucleotídicos utilizan oligonucleótidos 25-meros que permiten el fraccionamiento de moléculas de ADN cortas, mientras que los objetivos de ADN mayores de las micromatrices, aproximadamente 1000 pares de bases, pueden proporcionar una mayor sensibilidad al fraccionar mezclas complejas de ADN.

Otra medida útil de los métodos de la invención de los solicitantes es la de determinar el estado traduccional de la célula midiendo las abundancias de las especies proteicas constituyentes presentes en la célula usando procesos bien conocidos en la técnica.

Cuando se desea identificar los genes asociados con diversas funciones fisiológicas, se puede emplear un ensayo en el que se midan cambios en 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.

Además, se puede usar la expresión de un gen indicador o un marcador seleccionable para medir la modulación de la expresión génica usando la invención de los solicitantes.

En la técnica se conocen bien otros métodos para detectar los productos de expresión génica e incluyen análisis por transferencia de bandas southern (detección de ADN), transferencia en mancha o en ranura (ADN, ARN), transferencia de bandas northern (ARN), RT-PCR (ARN), transferencia de bandas western (detección de polipéptidos) y ELISA (polipéptidos). Aunque es menos preferido, se pueden usar proteínas marcadas para detectar una secuencia de ácido nucleico en particular con la que hibrida.

En algunos casos, es necesario amplificar la cantidad de una secuencia de ácido nucleico. Esto se puede llevar a cabo usando uno o más de una serie de métodos adecuados, incluidos, por ejemplo, la reacción en cadena de la polimerasa ("PCR"), la reacción en cadena de la ligasa ("LCR"), amplificación por desplazamiento de hebra ("SDA"), amplificación basada en la transcripción y similares. La PCR se lleva a cabo de acuerdo con técnicas conocidas en las que, por ejemplo, se trata una muestra de ácido nucleico en presencia de una polimerasa de ADN termoestable, bajo condiciones de hibridación, con un par de cebadores oligonucleotídicos, hibridando un cebador con una hebra (molde) de la secuencia específica que se quiere detectar. Los cebadores son lo suficientemente complementarios con cada hebra molde de la secuencia específica como para hibridar con ellas. Se sintetiza un producto de extensión de cada cebador y es complementario con la hebra molde de ácidos nucleicos con la que hibridó. El producto de extensión sintetizado a partir de cada cebador también puede servir como molde para la síntesis adicional de productos de extensión usando los mismos cebadores. Tras un número suficiente de ciclos de síntesis de productos de extensión, se puede analizar la muestra como se describe anteriormente para evaluar si la secuencia o secuencias que se quieren detectar están presentes.

La presente invención se puede comprender mejor por referencia a los siguientes ejemplos no limitantes, que se proporcionan como ejemplares de la invención.

Ejemplos

Métodos generales

Las técnicas estándar de ADN recombinante y de clonación molecular usadas en el presente documento son bien conocidas en la técnica y se describen por Sambrook, J., Fritsch, E. F. y Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. (1989) (Maniatis) y por T. J. Silhavy, M. L. Bannan, y L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) y por Ausubel, F. M. *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Assoc y Wiley-Interscience (1987).

En la técnica se conocen bien materiales y métodos adecuados para el mantenimiento y el crecimiento de cultivos bacterianos. Se pueden encontrar técnicas adecuadas para su uso en los ejemplos siguientes expuestas en *Manual of Methods for General Bacteriology* (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg y G. Briggs Phillips, ed), American Society for Microbiology, Washington, DC. (1994) o por Thomas D. Brock en *Biotechnology: A Textbook of Industrial Microbiology*, segunda edición, Sinauer Associates, Inc., Sunderland, MA (1989). Todos los reactivos, enzimas de restricción y materiales usados para el crecimiento y el mantenimiento de las células huésped se obtuvieron de Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GDBCO/BRL (Gaithersburg, MD) o Sigma Chemical Company (St. Louis, MO), a menos que se especifique lo contrario.

Las manipulaciones de secuencias genéticas se pueden lograr usando el conjunto de programas disponible de Genetics Computer Group Inc. (Wisconsin Package Versión 9.0, Genetics Computer Group (GCG), Madison, WI).

5 Cuando se usa el programa "Pileup" de GCG, se pueden usar el valor por defecto de 12 por creación de huecos y el valor por defecto de 4 por extensión de huecos. Cuando se usa el programa "Gap" o "Bestfit" de CGC, se pueden usar la penalización por defecto de 50 por creación de huecos y la penalización por defecto de 3 por extensión de huecos. En cualquier caso, cuando no se solicitan los parámetros del programa de GCG, en estos programas o en cualquier otro de GCG, se pueden usar los valores por defecto.

10 El significado de las abreviaturas es el siguiente: "h" significa hora(s), "min" significa minuto(s), "s" 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, "x g" significa multiplicado por gravedad, "nt" significa nucleótido(s), "aa" significa aminoácido(s), "pb" significa par(es) de bases, "kb" significa kilobase(s), "k" significa kilo, "μ" significa micro y "°C" significa grados Celsius.

15 Ejemplo 1

20 El sistema inducible de modulación de la expresión génica basado en EcR/RXR quiméricos de los solicitantes es útil en diversas aplicaciones que incluyen el tratamiento génico, la expresión de proteínas de interés en células huésped, la producción de organismos transgénicos y los ensayos basados en células. Los solicitantes han realizado el sorprendente descubrimiento de que un dominio de unión a ligando de receptor X retinoide quimérico puede sustituir a cualquiera de los polipéptidos de RXR originales y funcionar de forma inducible en un sistema de modulación de la expresión génica basado en EcR/RXR quiméricos tras la unión de ligando. Además, el polipéptido de RXR quimérico también puede funcionar mejor que cualquier dominio de unión a ligando de RXR original/donador. El sorprendente descubrimiento de los solicitantes y los mejores resultados inesperados proporcionan un sistema inducible de expresión génica novedoso para aplicaciones de células bacterianas, fúngicas, de levadura, de animal y de mamífero. Este ejemplo describe la construcción de varios casetes de expresión génica para su uso en el sistema inducible de expresión génica basado en EcR/RXR quiméricos de la invención.

30 Los solicitantes construyeron varios casetes de expresión génica basados en EcR basados en el EcR del gusano de las yemas de la picea *Choristoneura fumiferana* ("CfEcR"), la proteína ultraespiráculo de *C. fumiferana* ("CfUSP"), el EcR de *Drosophila melanogaster* ("DmEcR"), la USP de *D. melanogaster* ("DmUSP"), el EcR de *Tenebrio molitor* ("TmEcR"), el EcR de *Amblyomma americanum* ("AmaEcR"), el homólogo 1 de RXR de *A. americanum* ("AmaRXR1"), el homólogo 2 de RXR de *A. americanum* ("AmaRXR2"), la isoforma α del receptor X retinoide del ratón *Mus musculus* ("MmRXRα"), la isoforma β del receptor X retinoide de ser humano *Homo sapiens* ("HsRXRβ") y la proteína ultraespiráculo de la langosta *Locusta migratoria* ("LmUSP").

35 Las construcciones de receptor preparadas comprenden 1) un dominio de unión a ligando (LBD) de EcR, un LBD de RXR de vertebrado (MmRXRα o HsRXRβ), un LBD de USP de invertebrado (CfUSP o DmUSP), un LBD de RXR de invertebrado (LmUSP, AmaRXR1 o AmaRXR2) o un LBD de RXR quimérico que comprende un fragmento de LBD de RXR de vertebrado y un fragmento de LBD de RXR de invertebrado; y 2) un dominio de unión a ADN (DBD) de GAL4 o LexA o un dominio de transactivación (AD) activador ácido B42 o VP16. Las construcciones indicadoras incluyen un gen indicador, luciferasa o LacZ, enlazado de forma funcional a una construcción de promotor sintético que comprende un elemento de respuesta de GAL4 o un elemento de respuesta de LexA a los que unen el DBD de Gal4 o el DBD de LexA, respectivamente. Diversas combinaciones de estas construcciones de receptor e indicadoras se cotransfectaron en células de mamífero como se describe en los ejemplos 2-6 más adelante.

45 Casetes de expresión génica: Se construyeron pares de casetes (interruptores) de expresión génica basados en receptores de ecdisona como sigue, usando métodos estándar de clonación disponibles en la técnica. A continuación figura una breve descripción de la preparación y la composición de cada interruptor usado en los ejemplos descritos en el presente documento.

50 1.1 GAL4CfEcR-CDEF/VP16MmRXRα-EF: Los dominios C, D, E y F del EcR del gusano de las yemas de la picea *Choristoneura fumiferana* ("CfEcR-CDEF"; SEQ ID NO: 59) se fusionaron con un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD"; SEQ ID NO: 47) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Los dominios EF del RXRα del ratón *Mus musculus* ("MmRXRα-EF"; SEQ ID NO: 9) se fusionaron con el dominio de transactivación de VP16 ("VP16AD"; SEQ ID NO: 51) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Cinco sitios de unión de elementos de respuesta de GAL4 ("5XGAL4RE"; que comprenden 5 copias de un GAL4RE que comprende la SEQ ID NO: 55) se fusionaron con un promotor mínimo de E1b (SEQ ID NO: 61) y se situaron corriente arriba del gen de la luciferasa (SEQ ID NO: 62).

60 1.2 Gal4CfEcR-CDEF/VP16LmUSP-EF: Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque MmRXRα-EF se reemplazó con los dominios EF de la proteína ultraespiráculo de *Locusta migratoria* ("LmUSP-EF"; SEQ ID NO: 21).

65 1.3 Gal4CfEcR-CDEF/VP16MmRXRα (1-7)-LmUSP(8-12)-EF: Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque MmRXRα-EF se reemplazó con las hélices 1 a 7 de MmRXRα-EF y

las hélices 8 a 12 de LmUSP-EF (SEQ ID NO: 45).

- 5 1.4 Gal4CfEcR-CDBF/VP16MmRXR α (1-7)-LmUSP(8-12)-EF-MmRXR α -F: Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque MmRXR α -EF se reemplazó con las hélices 1 a 7 de MmRXR α -EF y las hélices 8 a 12 de LmUSP-EF (SEQ ID NO: 45) y en la que los últimos 18 nucleótidos C-terminales de la SEQ ID NO: 45 (dominio F) se reemplazaron con el dominio F de MmRXR α ("MmRXR α -F", SEQ ID NO: 63).
- 10 1.5 Gal4CfEcR-CDEFATP16MmRXR α (1-12)-EF-LmUSP-F: Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque MmRXR α -EF se reemplazó con las hélices 1 a 12 de MmRXR α -EF (SEQ ID NO: 9) y en la que los últimos 18 nucleótidos C-terminales de la SEQ ID NO: 9 (dominio F) se reemplazaron con el dominio F de LmUSP ("LmUSP-F", SEQ ID NO: 64).
- 15 1.6 Gal4CfEcR-CDEF/VP16LmUSP(1-12)EF-MmRXR α -F: Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque MmRXR α -EF se reemplazó con las hélices 1 a 12 de LmUSP-EF (SEQ ID NO: 21) y en la que los últimos 18 nucleótidos C-terminales de la SEQ ID NO: 21 (dominio F) se reemplazaron con el dominio F de MmRXR α ("MmRXR α -F", SEQ ID NO: 63).
- 20 1.7 GAL4CfEcR-DEF/VP16CfUSP-EF: Los dominios D, E y F del EcR del gusano de las yemas de la píceca *Choristoneura fumiferana* ("CfEcR-DEF"; SEQ ID NO: 65) se fusionaron con un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD"; SEQ ID NO: 47) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Los dominios EF de la USP de *C. fumiferana* ("CfUSP-EF"; SEQ ID NO: 66) se fusionaron con el dominio de transactivación de VP16 ("VP16AD"; SEQ ID NO: 51) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Cinco sitios de unión de elementos de respuesta de GAL4 ("5XGAL4RE"; que comprenden 5 copias de un GAL4RE que comprende la SEQ ID NO: 55) se fusionaron con un promotor mínimo de E1b (SEQ ID NO: 61) y se situaron corriente arriba del gen de la luciferasa (SEQ ID NO: 62).
- 25 1.8 GAL4CfEcR-DEF/VP16DmUSP-EF: Esta construcción se preparó de la misma manera que el interruptor 1.7 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF correspondientes de la USP de la mosca de la fruta *Drosophila melanogaster* ("DmUSP-EF", SEQ ID NO: 75).
- 30 1.9 Gal4CfEcR-DEF/VP16LmUSP-EF: Esta construcción se preparó de la misma manera que el interruptor 1.7 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF de la USP de *Locusta migratoria* ("LmUSP-EF"; SEQ ID NO: 21).
- 35 1.10 GAL4CfEcR-DEF/VP16MmRXR α -EF: Esta construcción se preparó de la misma manera que el interruptor 1.7 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF de MmRXR α de *M. musculus* ("MmRXR α -EF", SEQ ID NO: 9).
- 40 1.11 GAL4CfEcR-DEF/VP16AmaRXR1-EF: Esta construcción se preparó de la misma manera que el interruptor 1.7 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF del homólogo 1 de RXR de la garrapata *Amblyomma americanum* ("AmaRXR1-EF", SEQ ID NO: 22).
- 45 1.12 GAL4CfEcR-DEF/VP16AmaRXR2-EF: Esta construcción se preparó de la misma manera que el interruptor 1.7 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF del homólogo 2 de RXR de la garrapata *A. americanum* ("AmaRXR2-EF", SEQ ID NO: 23).
- 50 1.13 Gal4CfEcR-DEF/VP16MmRXR α (1-7)-LmUSP(8-12)-EF (" α quimera n.^o 7"): Esta construcción se preparó de la misma manera que el interruptor 1.1 anterior, excepto porque CfUSP-EF se reemplazó con las hélices 1 a 7 de MmRXR α -EF y las hélices 8 a 12 de LmUSP-EF (SEQ ID NO 45).
- 55 1.14 GAL4DmEcR-DEF/VP16CfUSP-EF: Los dominios D, E y F del EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR-DEF"; SEQ ID NO: 67) se fusionaron con un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD"; SEQ ID NO: 47) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Los dominios EF de la USP de *C. fumiferana* ("CfUSP-EF"; SEQ ID NO: 66) se fusionaron con el dominio de transactivación de VP16 ("VP16AD"; SEQ ID NO: 51) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Cinco sitios de unión de elementos de respuesta de GAL4 ("5XGAL4RE"; que comprenden 5 copias de un GAL4RE que comprende la SEQ ID NO: 55) se fusionaron con un promotor mínimo de E1b (SEQ ID NO: 61) y se situaron corriente arriba del gen de la luciferasa (SEQ ID NO: 62).
- 60 1.15 GAL4DmEcR-DEF/VP16DmUSP-EF: Esta construcción se preparó de la misma manera que el interruptor 1.14 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF correspondientes de la USP de la mosca de la fruta *Drosophila melanogaster* ("DmUSP-EF", SEQ ID NO: 75).
- 65 1.16 Gal4DmEcR-DEF/VP16LmUSP-EF: Esta construcción se preparó de la misma manera que el interruptor 1.14

anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF de la USP de *Locusta migratoria* ("LmUSP-EF"; SEQ ID NO: 21).

5 1.17 GAL4DmEcR-DEF/VP16MmRXR α -EF: Esta construcción se preparó de la misma manera que el interruptor 1.14 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF de MmRXR α de *Mus musculus* ("MmRXR α -EF", SEQ ID NO: 9).

10 1.18 GAL4DmEcR-DEF/VP16AmaRXR1-EF: Esta construcción se preparó de la misma manera que el interruptor 1.14 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF del homólogo 1 de RXR de la garrapata ixódida *Amblyomma americanum* ("AmaRXR1-EF", SEQ ID NO: 22).

15 1.19 GAL4DmEcR-DEF/VP16AmaRXR2-EF: Esta construcción se preparó de la misma manera que el interruptor 1.14 anterior, excepto porque CfUSP-EF se reemplazó con los dominios EF del homólogo 2 de RXR de la garrapata ixódida *A. americanum* ("AmaRXR2-EF", SEQ ID NO: 23).

1.20 Gal4DmEcR-DEF/VP16MmRXR α (1-7)-LmUSP(8-12)-EF: Esta construcción se preparó de la misma manera que el interruptor 1.14 anterior, excepto porque CfUSP-EF se reemplazó con las hélices 1 a 7 de MmRXR α -EF y las hélices 8 a 12 de LmUSP-EF (SEQ ID NO 45).

20 1.21 GAL4TmEcR-DEF/VP16MmRXR α (1-7)-LmUSP(8-12)-EF: Esta construcción se preparó de la misma manera que el interruptor 1.20 anterior, excepto porque DmEcR-DEF se reemplazó con los dominios D, E y F correspondientes del EcR del escarabajo *Tenebrio molitor* ("TmEcR-DEF", SEQ ID NO: 71), se fusionó con un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD"; SEQ ID NO: 47) y se situó bajo el control de un promotor de SV40e (SEQ ID NO: 60). Dominios EF quiméricos que comprenden las hélices 1 a 7 de MmRXR α -EF y las hélices 8 a 12 de LmUSP-EF (SEQ ID NO: 45) se fusionaron con el dominio de transactivación de VP16 ("VP16AD"; SEQ ID NO: 51) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Cinco sitios de unión de elementos de respuesta de GAL4 ("5XGAL4RE"; que comprenden 5 copias de un GAL4RE que comprende la SEQ ID NO: 55) se fusionaron con un promotor mínimo de E1b (SEQ ID NO: 61) y se situaron corriente arriba del gen de la luciferasa (SEQ ID NO: 62).

30 1.22 Gal4AmaEcR-DEF/VP16MmRXR α (1-7)-LmUSP(8-12)-EF: Esta construcción se preparó de la misma manera que el interruptor 1.21 anterior, excepto porque TmEcR-DEF se reemplazó con los dominios DEF correspondientes del EcR de la garrapata *Amblyomma americanum* ("AmaRXR1-EF", SEQ ID NO: 73).

35 1.23 GAL4CfBcR-CDEF/VP16HsRXR β -EF: Los dominios C, D, E y F del EcR del gusano de las yemas de la píceo *Choristoneura fumiferana* ("CfEcR-CDEF"; SEQ ID NO: 59) se fusionaron con un dominio de unión a ADN de GAL4 ("Gal4DNABD" o "Gal4DBD"; SEQ ID NO: 47) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Los dominios EF del RXR β de ser humano *Homo sapiens* ("HsRXR β -EF"; SEQ ID NO: 13) se fusionaron con el dominio de transactivación de VP16 ("VP16AD"; SEQ ID NO: 51) y se situaron bajo el control de un promotor de SV40e (SEQ ID NO: 60). Cinco sitios de unión de elementos de respuesta de GAL4 ("5XGAL4RE"; que comprenden 5 copias de un GAL4RE que comprende la SEQ ID NO: 55) se fusionaron con un promotor mínimo de E1b (SEQ ID NO: 61) y se situaron corriente arriba del gen de la luciferasa (SEQ ID NO: 62).

40 1.24 GAL4CfEcR-DEF/VP16HsRXR β -EF: Esta construcción se preparó de la misma manera que el interruptor 1.23 anterior, excepto porque CfEcR-CDEF se reemplazó con los dominios DEF de EcR de *C. fumiferana* ("CfEcR-DEF", SEQ ID NO: 65).

45 1.25 GAL4CfEcR-DEF/VP16HsRXR β (1-6)-LmUSP(7-12)-EF (" β quimera n.º 6"): Esta construcción se preparó de la misma manera que el interruptor 1.24 anterior, excepto porque HsRXR β -EF se reemplazó con las hélices 1 a 6 de HsRXR β -EF (nucleótidos 1-348 de la SEQ ID NO: 13) y las hélices 7 a 12 de LmUSP-EF (nucleótidos 268-630 de la SEQ ID NO: 21).

50 1.26 GAL4CfEcR-DEF/VP16HsRXR β (1-7)-LmUSP(8-12)-EF (" β quimera n.º 8"): Esta construcción se preparó de la misma manera que el interruptor 1.24 anterior, excepto porque HsRXR β -EF se reemplazó con las hélices 1 a 7 de HsRXR β -EF (nucleótidos 1-408 de la SEQ ID NO: 13) y las hélices 8 a 12 de LmUSP-EF (nucleótidos 337-630 de la SEQ ID NO: 21).

55 1.27 GAL4CfEcR-DEF/VP16HsRXR β (1-8)-LmUSP(9-12)-EF (" β quimera n.º 9"): Esta construcción se preparó de la misma manera que el interruptor 1.24 anterior, excepto porque HsRXR β -EF se reemplazó con las hélices 1 a 8 de HsRXR β -EF (nucleótidos 1-465 de la SEQ ID NO: 13) y las hélices 9 a 12 de LmUSP-EF (nucleótidos 403-630 de la SEQ ID NO: 21).

60 1.28 GAL4CfEcR-DEF/VP16HsRXR β (1-9)-LmUSP(11-12)-EF (" β quimera n.º 10"): Esta construcción se preparó de la misma manera que el interruptor 1.24 anterior, excepto porque HsRXR β -EF se reemplazó con las hélices 1 a 9 de

HsRXR β -EF (nucleótidos 1-555 de la SEQ ID NO: 13) y las hélices 10 a 12 de LmUSP-EF (nucleótidos 490-630 de la SEQ ID NO: 21).

5 1.29 GAL4CfEcR-DEF/VP16HsRXR β (1-10)-LmUSP(11-12)-EF (" β quimera n.º 11"): Esta construcción se preparó de la misma manera que el interruptor 1.24 anterior, excepto porque HsRXR β -EF se reemplazó con las hélices 1 a 10 de HsRXR β -EF (nucleótidos 1-624 de la SEQ ID NO: 13) y las hélices 11 a 12 de LmUSP-EF (nucleótidos 547-630 de la SEQ ID NO: 21).

10 1.30 GAL4DmEcR-DEF/VP16HsRXR β (1-6)-LmUSP(7-12)-EF (" β quimera n.º 6"): Esta construcción se preparó de la misma manera que el interruptor 1.25 anterior, excepto porque CfEcR-DEF se reemplazó con DmEcR-DEF (SEQ ID NO: 67).

15 1.31 GAL4DmEcR-DEF/VP16HsRXR β (1-7)-LmUSP(8-12)-EF (" β quimera n.º 8"): Esta construcción se preparó de la misma manera que el interruptor 1.26 anterior, excepto porque CfEcR-DEF se reemplazó con DmEcR-DEF (SEQ ID NO:67).

20 1.32 GAL4DmEcR-DEF/VP16HsRXR β (1-8)-LmUSP(9-12)-EF (" β quimera n.º 9"): Esta construcción se preparó de la misma manera que el interruptor 1.27 anterior, excepto porque CfEcR-DEF se reemplazó con DmEcR-DEF (SEQ ID NO:67).

1.33 GAL4DmEcR-DEF/VP16HsRXR β (1-9)-LmUSP(10-12)-EF (" β quimera n.º 10"): Esta construcción se preparó de la misma manera que el interruptor 1.28 anterior, CfEcR-DEF se reemplazó con DmEcR-DEF (SEQ ID NO: 67).

25 1.34 GAL4DmEcR-DEF/VP16HsRXR β (1-10)-LmUSP(11-12)-EF (" β quimera n.º 11"): Esta construcción se preparó de la misma manera que el interruptor 1,29 anterior, excepto porque CfEcR-DEF se reemplazó con DmEcR-DEF (SEQ ID NO: 67).

Ejemplo 2

30 Recientemente, los solicitantes han realizado el sorprendente descubrimiento de que los RXR de invertebrado y sus homólogos de RXR que no son ni de díptero ni de lepidóptero pueden funcionar de forma similar a o mejor que los RXR de vertebrado en un sistema inducible de modulación de la expresión génica basado en receptores de ecdisona en células tanto de levadura como de mamífero (en la solicitud de EE. UU. provisional con número de serie 60/294.814). De hecho, los solicitantes han demostrado que LmUSP es un compañero mejor para CfEcR que el RXR de ratón en células de mamífero. Aun así, para la mayoría de las aplicaciones de sistemas de expresión génica, en particular los destinados a células de mamífero, es deseable tener un RXR de vertebrado como compañero. Para identificar una región mínima de LmUSP necesaria para esta mejora, los solicitantes han construido y analizado quimeras de RXR de vertebrado/RXR de invertebrado (denominadas indistintamente en el presente documento "RXR quiméricos" o "quimeras de RXR") en un sistema inducible de modulación de la expresión génica basado en EcR. Brevemente, se estudiaron el potencial de inducción génica (magnitud de la inducción) y la especificidad y sensibilidad a ligando usando un ligando no esteroideo en una inducción dependiente de dosis de la expresión del gen indicador en las células NIH3T3 y las células A549 transfectadas.

45 En el primer conjunto de quimeras de RXR, las hélices 8 a 12 de MmRXR α -EF se reemplazaron con las hélices 8 a 12 de LmUSP-EF (interruptor 1.3 como se preparó en el ejemplo 1). Se tomaron tres clones independientes (quimeras de RXR Q n.º 1, Q n.º 2 y Q n.º 3 en las figuras 1-3) y se compararon con los interruptores originales MmRXR α -EF y LmUSP-EF (interruptores 1.1 y 1.2, respectivamente, como se prepararon en el ejemplo 1). La quimera de RXR y los ADN originales se transfectaron en células NTH3T3 de ratón junto con Gal4/CfEcR-CDEF y el plásmido indicador pFRLuc. Las células transfectadas se hicieron crecer en presencia de 0, 0,2, 1, 5 y 10 μ M de ligando no esteroideo N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina (ligando GS-ETM). Las células se recogieron 48 horas después del tratamiento y se sometió a ensayo la actividad indicadora. Los números de la parte superior de las barras corresponden a la multiplicación máxima de la activación/inducción para ese tratamiento.

55 Transfecciones: Se transfectaron ADN correspondientes a las diversas construcciones de interruptores descritas en el ejemplo 1, específicamente los interruptores del 1.1 al 1.6, en células NIH3T3 de ratón (ATCC) y células A549 humanas (ATCC) como sigue. Las células se recogieron cuando alcanzaron el 50 % de confluencia y se plaquearon en placas de 6, 12 o 24 pocillos a 125.000, 50.000 o 25.000 células, respectivamente, en 2,5, 1,0 o 0,5 ml de medio de crecimiento que contenía suero fetal bovino al 10 % (FBS), respectivamente. Las células NIH3T3 se hicieron crecer en medio Eagle modificado de Dulbecco (DMEM; LifeTechnologies) y las células A549 se hicieron crecer en mezcla de nutrientes F12K (LifeTechnologies). Al día siguiente, se aclararon las células con medio de crecimiento y se transfectaron durante cuatro horas. Se descubrió que SuperfectTM (Qiagen Inc.) era el mejor reactivo de transfección para células 3T3 y para células A549. Para placas de 12 pocillos, se mezclaron 4 μ l de SuperfectTM con 100 μ l de medio de crecimiento. Se añadieron a la mezcla de transfección 1,0 μ g de la construcción indicadora y 65 0,25 μ g de cada construcción de receptor del par de receptores que se quería analizar. Se añadió una segunda

construcción indicadora [pTKRL (Promega), 0,1 µg/mezcla de transfección] que comprendía un gen de luciferasa de *Renilla* enlazado de forma funcional y situado bajo el control de un promotor constitutivo de timidina cinasa (TK) y se usó para la normalización. Los contenidos de la mezcla de transfección se mezclaron en un mezclador de vórtex y se dejaron reposar a temperatura ambiente durante 30 min. Al final de la incubación, se añadió la mezcla de transfección a las células mantenidas en 400 µl de medio de crecimiento. Las células se mantuvieron a 37 °C y con CO₂ al 5 % durante cuatro horas. Al final de la incubación, se añadieron 500 µl de medio de crecimiento que contenía FBS al 20 % y, o bien dimetilsulfóxido (DMSO; control) o una solución de DMSO de 0,2, 1, 5, 10 y 50 µM de ligando no esteroideo N-(2-etil-3-metoxibenzoil)N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina, y las células se mantuvieron a 37 °C y con CO₂ al 5 % durante 48 horas. Se recogieron las células y se sometió a ensayo la actividad indicadora. Para las placas de 6 y 24 pocillos se siguió el mismo procedimiento, excepto porque todos los reactivos se duplicaron para las placas de 6 pocillos y se redujeron a la mitad para las placas de 24 pocillos.

Ligando: El ligando no esteroideo N-(2-etil-3-metoxibenzoil)-N'-(3,5-dimetilbenzoil)-N'-t-butilhidrazina (ligando no esteroideo GSTTM-E) es un ligando ecdisteroide sintético estable sintetizado en Rohm and Haas Company. Los ligandos se disolvieron en DMSO y la concentración final de DMSO se mantuvo al 0,1 % tanto en los controles como en los tratamientos.

Ensayos indicadores: Las células se recogieron 48 horas después de añadir los ligandos. Se añadieron 125, 250 o 500 µl de tampón de lisis pasiva (parte del sistema de ensayo indicador Dual-luciferaseTM de Promega Corporation) a cada pocillo de placas de 24 o 12 o 6 pocillos, respectivamente. Las placas se dispusieron en un agitador rotatorio durante 15 minutos. Se sometieron a ensayo veinte µl de lisado. La actividad luciferasa se midió usando el sistema de ensayo indicador Dual-luciferaseTM de Promega Corporation siguiendo las instrucciones del fabricante. La β-galactosidasa se midió usando el kit de ensayo Galacto-StarTM de TROPK siguiendo las instrucciones del fabricante. Todas las actividades luciferasa y β-galactosidasa se normalizaron usando la luciferasa de *Renilla* como estándar. Se calculó la multiplicación de las actividades dividiendo las unidades lumínicas relativas ("RLU") normalizadas en células tratadas con ligando con RLU normalizadas en células tratadas con DMSO (control no tratado).

Resultados: Sorprendentemente, los tres clones independientes de las quimeras de RXR probadas (interruptor 1.3) fueron mejores que cualquiera de los interruptores basados en los originales, MmRXRα-EF (interruptor 1.1) y LmUSP-EF (interruptor 1.2), véase la figura 1. En particular, el RXR quimérico demostró un aumento de la sensibilidad a ligando y un aumento de la magnitud de la inducción. Por tanto, los solicitantes han realizado el sorprendente descubrimiento de que se puede usar un dominio de unión a ligando de RXR quimérico en lugar de un RXR de vertebrado o un RXR de invertebrado en un sistema inducible de modulación de la expresión génica basado en EcR. Este sistema de expresión génica basado en EcR/RXR quiméricos novedoso proporciona un sistema mejorado caracterizado por un aumento tanto de la sensibilidad a ligando como de la magnitud de la inducción.

Los dos mejores clones de quimeras de RXR del interruptor 1.3 ("Q n.º 1" y "Q n.º 2" de la figura 2) se compararon con los interruptores basados en los originales 1.1 y 1.2 es un experimento repetido ("Quim-1" y "Quim-2" en la figura 2, respectivamente). En este experimento, el interruptor basado en RXR quimérico fue de nuevo más sensible a ligando no esteroideo que cualquiera de los interruptores basados en los originales (véase la figura 2). Sin embargo, en este experimento, el interruptor basado en RXR quimérico fue mejor que el interruptor basado en RXR (MmRXRα-EF) de vertebrado en cuanto a la magnitud de inducción, pero fue similar al interruptor de RXR (LmUSP-EF) de invertebrado.

Los mismos interruptores basados en RXR quiméricos y en RXR originales también se estudiaron en un línea celular de carcinoma de pulmón humano A549 (ATCC) y se observaron resultados similares (figura 3).

Por tanto, los solicitantes han demostrado por primera vez que un dominio de unión a ligando de RXR quimérico puede funcionar de forma eficaz en asociación con un receptor de ecdisona en un sistema inducible de expresión génica en células de mamífero. Sorprendentemente, el sistema inducible de expresión génica basado en EcR/RXR quiméricos de la presente invención es una mejora con respecto a los sistemas de modulación de la expresión génica basados tanto en EcR/RXR de vertebrado como en EcR/RXR de invertebrado, ya que se requiere menos ligando para la transactivación y se puede lograr un aumento de los niveles de transactivación.

Basándose en el descubrimiento de los solicitantes descrito en el presente documento, un experto en la técnica puede predecir que otro dominio de unión a ligando de RXR quimérico que comprende al menos dos fragmentos polipeptídicos de RXR de especies diferentes de un LBD de RXR de vertebrado, un LBD de RXR de invertebrado o un LBD de homólogo de RXR que no es de díptero ni de lepidóptero también funcionará en el sistema inducible de expresión génica basado en EcR/RXR quiméricos de los solicitantes. Basándose en la invención de los solicitantes, los medios para preparar realizaciones adicionales de LBD de RXR quiméricos dentro del alcance de la presente invención se encuentran en la técnica y no es necesaria ninguna experimentación indebida. De hecho, un experto en la técnica puede clonar y secuenciar de forma rutinaria un polinucleótido que codifica un LBD de RXR de vertebrado o invertebrado o de un homólogo de RXR, y basándose en análisis de homología de secuencia similares a los presentados en la figura 4, determinar el polinucleótido y los fragmentos polinucleotídicos correspondientes del LBD del RXR de esa especie en particular que se engloban en el alcance de la presente invención.

Un experto en la técnica también puede predecir que el sistema inducible de expresión génica novedoso de los solicitantes también funcionará para modular la expresión génica en células de levadura. Dado que los sistemas de expresión génica basados de EcR/homólogos de RXR de díptero/y homólogos de RXR de lepidóptero funcionan de forma constitutiva en células de levadura (datos no mostrados), de forma similar a como funcionan en células de mamífero, y que los RXR de invertebrado que no son dípteros ni lepidópteros funcionan de forma inducible en asociación con un EcR en células de mamífero, se predice que el sistema inducible de modulación de la expresión génica basado en EcR/RXR quiméricos funciona de forma inducible en células de levadura, de forma similar a como funcionan en células de mamífero. Por tanto, el sistema inducible de expresión génica de EcR/RXR quiméricos de la presente invención es útil en aplicaciones en las que se desea la modulación de los niveles de expresión génica en células tanto de levadura como de mamífero. Además, también se contempla que la invención de los solicitantes funcione en otras células, incluidas, entre otras, células bacterianas, células fúngicas y células animales.

Ejemplo 3

En el extremo C-terminal del LBD hay seis aminoácidos que son diferentes entre MmRXR α y LmUSP (véanse las alineaciones de secuencia presentadas en la figura 4). Para comprobar si estos seis aminoácidos contribuyen a las diferencias observadas entre las capacidades de transactivación de MmRXR α y LmUSP, los solicitantes construyeron quimeras de RXR en las que los seis aminoácidos C-terminales, denominados en el presente documento dominio F, de un RXR original se sustituyeron por el dominio F del otro RXR original. Se construyeron interruptores génicos que comprendían LmUSP-EF fusionado con MmRXR α -F (VP16/LmUSP-EF-MmRXR α -F, interruptor 1.6), MmRXR α -EF fusionado con LmUSP-F (VP16/MmRXR α EF-LmUSP-F, interruptor 1.5) y MmRXR α -EF(1-7)-LmUSP-EF(8-12) fusionado con MmRXR α -F (Quimera/RXR-F, interruptor 1.4) como se describe en el ejemplo 1. Estas construcciones se transflectaron en células NIH3T3 y se sometió a ensayo el potencial de transactivación en presencia de 0, 0,2, 1 y 10 μ M de ligando no esteroideo N-(2-etil-3-metoxibenzoil)N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina. Las quimeras de dominio F (interruptores 1.4-1.6) se compararon con el LBD de RXR quimérico MmRXR α -EF(1-7)-LmUSP-EF(8-12) del interruptor génico 1.3. Se usó el plásmido pFRLUC (Stratagene) que codifica un polipéptido de luciferasa como construcción de gen indicador y se usó pTKRL (Promega) que codifica un polipéptido de luciferasa de *Renilla* bajo el control del promotor constitutivo TK para normalizar las transfecciones como se describe anteriormente. Se recogieron las células, se lisaron y se midió la actividad indicadora luciferasa en los lisados celulares. Se presentan las unidades lumínicas relativas de luciferasa de mosca totales. El número de la parte superior de cada barra es la multiplicación máxima de la inducción para ese tratamiento. El análisis se realizó por triplicado y se determinaron los recuentos medios de luciferasa [unidades lumínicas relativas (RLU) totales] como se describe anteriormente.

Como se muestra en la figura 5, los seis aminoácidos del extremo C-terminal del LBD (dominio F) no parecen contribuir a las diferencias observadas entre las capacidades de transactivación de RXR de vertebrado y RXR de invertebrado, lo que indica que lo más probable es que las hélices 8-12 del dominio EF sean responsables de estas diferencias entre RXR de vertebrado e invertebrado.

Ejemplo 4

Este ejemplo describe la construcción de cuatro interruptores génicos basados en EcR-DEF que comprenden los dominios DEF de *Choristoneura fumiferana* (lepidóptero), *Drosophila melanogaster* (díptero), *Tenebrio molitor* (coleóptero) y *Amblyomma americanum* (ixódido) fusionados a un dominio de unión a ADN de GAL4. Además, los dominios EF de RXR de vertebrado, RXR de invertebrado o USP de invertebrado de USP de *Choristoneura fumiferana*, USP de *Drosophila melanogaster*, USP de *Locusta migratoria* (ortóptero), RXR α de *Mus musculus* (vertebrado), una quimera entre MmRXR α y LmUSP (Quimera; del interruptor 1.13), homólogo 1 de RXR de *Amblyomma americanum* (ixódido), homólogo 2 de RXR de *Amblyomma americanum* (ixódido) se fusionaron con un dominio de activación VP16. Las combinaciones de receptores se compararon en cuanto a su capacidad para transactivar el plásmido indicador pFRLuc en células NIH3T3 de ratón en presencia de 0, 0,2, 1 o 10 μ M de ligando esteroideo PonA (Sigma Chemical Company) o 0, 0,04, 0,2, 1 o 10 μ M de ligando no esteroideo N-(2-etil-3-metoxibenzoil)N'-(3,5-dimetilbenzoil)-N'-terc-butilhidrazina como se describe anteriormente. Se recogieron las células, se lisaron y se midió la actividad indicadora luciferasa en los lisados celulares. Se presentan las unidades lumínicas relativas de luciferasa de mosca totales. El número de la parte superior de cada barra es la multiplicación máxima de la inducción para ese tratamiento. El análisis se realizó por triplicado y se determinaron los recuentos medios de luciferasa [unidades lumínicas relativas (RLU) totales] como se describe anteriormente.

Las figuras 6-8 muestran los resultados de estos análisis. La quimera de MmRXR-LmUSP fue el mejor compañero para CfEcR (inducción de 11.000 veces, figura 6), DmEcR (inducción de 1759 veces, figura 7). Para todos los demás EcR probados, la quimera de RXR produjo mayores niveles de fondo en ausencia de ligando (véase la figura 8). El interruptor basado en CfEcR/RXR quimérico (interruptor 1.13) fue más sensible a no esteroide que a PonA, mientras que el interruptor basado en DmEcR/RXR quimérico (interruptor 1.20) fue más sensible a PonA que a no esteroide. Dado que estos dos formatos de interruptor producen niveles de inducción decentes y muestran sensibilidad diferencial a esteroides y no esteroides, se pueden aprovechar para aplicaciones en las que se desean interruptores

de dos o más genes.

Excepto por CfEcR, todos los demás EcR probados en asociación con el RXR quimérico son más sensibles a esteroides que a no esteroides. El interruptor basado en TmEcR/RXR quimérico (interruptor 1.21; figura 8) es más sensible a PonA y menos sensible a no esteroide y funciona mejor asociado con MmRXR α , AmaRXR1 o AmaRXR2. El interruptor basado en AmaEcR/RXR quimérico (interruptor 1.22; figura 8) también es más sensible a PonA y menos sensible a no esteroide y funciona mejor asociado con un casete de expresión génica basado en LmUSP, MmRXR, AmaRXR1 o AmaRXR2. Por tanto, parece que los interruptores génicos basado en TmEcR/ y AmaEcR/RXR quimérico forman un grupo de receptores de ecdisona que es diferente del grupo de interruptores génicos basados en EcR de lepidóptero y díptero/RXR quimérico (CfEcR/RXR quimérico y DmEcR/RXR quimérico, respectivamente). Como se indica anteriormente, las diferentes sensibilidades a ligando de los interruptores génicos basados en EcR/RXR quiméricos de los solicitantes son ventajosos para su uso en aplicaciones en las que se desean interruptores de dos o más genes.

5 Ejemplo 5

Este ejemplo describe el análisis adicional de los solicitantes de casetes de expresión génica que codifican diversos polipéptidos de RXR quiméricos que comprenden un fragmento polipeptídico de isoforma RXR α de ratón o un fragmento polipeptídico de isoforma RXR β humana y un fragmento polipeptídico de LmUSP en células NIH3T3 de ratón. Estas quimeras de RXR se construyeron en un esfuerzo para identificar la hélice o hélices del dominio EF que contribuyen a las diferencias de transactivación observadas entre RXR de vertebrado e invertebrado. Brevemente, se construyeron cinco casetes de expresión génica diferentes que codifican un dominio de unión a ligando de RXR quimérico como se describe en el ejemplo 1. Los cinco dominios de unión a ligando de RXR quimérico codificados por estos casetes de expresión génica y los correspondientes fragmentos de RXR de vertebrado y de RXR de invertebrado que comprenden se representan en la tabla 1.

Tabla 1 RXR quiméricos de dominio EF de HsRXR β /LmUSP

Nombre de la quimera	Fragmento(s) polipeptídico(s) de HsRXR β -EF	Fragmento(s) polipeptídico(s) de LmUSP-EF
β Quimera n.º 6	Hélices 1-6	Hélices 7-12
β Quimera n.º 8	Hélices 1-7	Hélices 8-12
β Quimera n.º 9	Hélices 1-8	Hélices 9-12
β Quimera n.º 10	Hélices 1-9	Hélices 10-12
β Quimera n.º 11	Hélices 1-10	Hélices 11-12

Se transfectaron tres clones individuales de cada LBD de RXR quimérico de la tabla 1 en células NIH3T3 de ratón junto con GAL4CfEcR-DEF (interruptores 1.25-1.29 del ejemplo 1; figuras 9 y 10) o GAL4DmEcR-DEF (interruptores 1.30-1.34 del ejemplo 1; figura 11) y el plásmido indicador pFRLuc como se describe anteriormente. Las células transfectadas se cultivaron en presencia de a) 0, 0,2, 1 o 10 μ M de ligando no esteroideo (figura 9) o b) 0, 0,2, 1 o 10 μ M de ligando esteroide PonA o 0, 0,4, 0,2, 1 o 10 μ M de ligando no esteroide (figuras 10 y 11) durante 48 horas. Se midió la actividad del gen indicador y se muestran la RLU totales. El número de la parte superior de cada barra es la multiplicación máxima de la inducción para ese tratamiento y es la media de tres repeticiones.

Como se muestra en la figura 9, los mejores resultados se obtuvieron cuando se usó un dominio de unión a ligando de RXR quimérico HsRXR β H1-8 y LmUSP H9-12 (del interruptor 1.27), lo que indica que la hélice 9 de LmUSP puede ser responsable de la sensibilidad y la magnitud de la inducción.

Usando CfEcR como compañero, la quimera 9 demostró la inducción máxima (véase la figura 10). Las quimeras 6 y 8 también produjeron una buena inducción y menor fondo, como consecuencia la multiplicación de la inducción fue mayor para estar dos quimeras en comparación con la quimera 9. Las quimeras 10 y 11 produjeron niveles menores de actividad indicadora.

Usando DmEcR como compañero, la quimera 8 produjo la actividad indicadora (véase la figura 11). La quimera 9 también tuvo un buen rendimiento, mientras que las quimeras 6, 10 y 11 demostraron niveles menores de actividad indicadora.

La selección de un dominio de unión a ligando de RXR quimérico en particular también puede influir en el rendimiento del EcR en respuesta a un ligando en particular. Específicamente, CfEcR en combinación con la quimera 11 respondió bien a no esteroide, pero no a PonA (véase la figura 10). Por el contrario, DmEcR en combinación con la quimera 11 respondió bien a PonA, pero no a no esteroide (véase la figura 11).

5 Ejemplo 6

5 Este ejemplo demuestra el efecto de la introducción de un segundo ligando en la célula huésped que comprende un sistema inducible de modulación de la expresión génica basado en EcR/RXR quiméricos de la invención. En particular, los solicitantes han determinado el efecto del ácido 9-cis-retinoico sobre el potencial de transactivación del interruptor génico GAL4CfEcR-DEF/VP16HsRXRβ-(1-8)-LmUSP-(9-12)-EF (βquimera 9; interruptor 1.27) junto con pFRLuc en células NIH 3T3 en presencia de no esteroide (GSE) durante 48 horas.

10 Brevemente, se transfectaron GAL4CfEcR-DEF, pFRLuc y VP16HsRXRβ-(1-8)-LmUSP-(9-12)-EF (quimera n.º 9) en células NIH3T3 y las células transfectadas se trataron con 0, 0,04, 0,2, 1, 5 y 25 µM de ligando no esteroideo (GSE) y 0, 1, 5 y 25 µM de ácido 9-cis-retinoico (Sigma Chemical Company). La actividad indicadora se midió 48 horas después de añadir los ligandos.

15 Como se muestra en la figura 12, la presencia de ácido retinoico aumentó la sensibilidad del CfEcR-DEF a ligando no esteroideo. A una concentración de ligando no esteroideo de 0,04 µM, hay muy poca inducción en ausencia de ácido 9-cis-retinoico, pero cuando se añade ácido 9-cis-retinoico 1 µM además del no esteroide 0,04 µM, la inducción aumenta enormemente.

Listado de secuencias

20 <110> Rohm and Haas Company

Palli, Subba R.

25 Kapitskaya, Marianna Z.

<120> Receptores X retinoides quiméricos y su uso en un sistema inducible de expresión génica basado en receptores de ecdisona novedoso

30 <130> A01238

<150> Documento US 60/294.819

<151> 2001-05'31

35 <160> 75

<170> PatentIn versión 3,1

40 <210> 1

<211> 735

<212> ADN

45 <213> Choristoneura fumiferana

<400> 1

```

taccaggacg ggtacgagca gccttctgat gaagatttga agaggattac gcagacgtgg      60
cagcaagcgg acgatgaaaa cgaagagtct gacactcoct tccgccagat cacagagatg      120
actatcctca cggccaact tatcgtggag ttcgcgaagg gattgccagg gttcgccaag      180
atctcgcagc ctgatcaaat tacgctgctt aaggcttgct caagtgaggt aatgatgctc      240
cgagtcgcgc gacgatacga tgcggcctca gacagtgttc tgttcgcgaa caaccaagcg      300
tacactcgcg acaactaccg caaggctggc atggcctacg tcatcgagga totactgcac      360
ttctgccggt gcatgtactc tatggcgctt gacaacatcc attacgcgct gctcacgget      420

```

ES 2 392 508 T3

gtcgtcatct tttctgaccg gccaggggtg gaggagccgc aactggtgga agaaatccag 480
 cgggtactacc tgaatacgcct ccgcatctat atcctgaacc agctgagcgg gtccggcgcgt 540
 tcgtccgtca tatacggcaa gatcctctca atcctctctg agctacgcac gctcggcatg 600
 caaaactcca acatgtgcat ctccctcaag ctcaagaaca gaaagctgoc gcctttcctc 660
 gaggagatct gggatgtggc ggacatgtcg cacacccaac cgcgcctat cctcgagtcc 720
 cccacgaatc tctag 735

<210> 2

5 <211> 1338

<212> ADN

<213> Drosophila melanogaster

10

<400> 2

tatgagcagc catctgaaga ggatctcagg cgtataatga gtcaaccoga tgagaacgag 60
 agccaaaacgg acgtcagctt tcggcatata accgagataa ccatactcac ggtccagttg 120
 attggttgagt ttgctaaagg tctaccagcg tttacaaaga taccocagga ggaccagatc 180
 acgttactaa aggcctgctc gtcggagggtg atgatgctgc gtatggcacg acgctatgac 240
 cacagctcgg actcaatatt cttegcgaat aatagatcat atacgcggga ttcttcaaaa 300
 atggccggaa tggctgataa cattgaagac ctgctgcatt tctgcccga aatgttctcg 360
 atgaagggtg acaacgtcga atacgcgctt ctcactgcca ttgtgatctt ctccgaccgg 420
 ccgggccttg agaaggccca actagtcgaa gcgatccaga gctactacat cgacacgcta 480
 cgcatttata tactcaaccg ccactgcggc gactcaatga gcctcgtctt ctacgcaaag 540
 ctgctctoga tctcaccga gctgcgtacg ctgggcaacc agaacgcga gatgtgtttc 600
 tcactaaagc tcaaaaaccg caaactgccc aagttcctcg aggagatctg ggacgttcat 660
 gccatcccgc catcgggtcca gtcgcacctt cagattacc aggaggagaa cgagcgtctc 720
 gagcgggctg agcgtatgcg ggcacgggtt gggggcgcca ttaccgcccg cattgattgc 780
 gactctgctt ccaactcggc ggcggcagcc gcggcccagc atcagcctca gcctcagccc 840
 cagccccaac cctcctccct gaccagaac gattcccagc accagacaca gccgcagcta 900
 caacctcagc taccacctca gctgcaaggc caactgcaac cccagctcca accacagctt 960
 cagacgcaac tccagccaca gattcaacca cagccacagc tcttcccgct ctccgctccc 1020
 gtgcccgcct ccgtaaccgc acctggttcc ttgtccgagg tcagtacgag cagcgaatac 1080
 atgggcggaa gtgcggccat aggaccatc acgcccggaa ccaccagcag taccacggct 1140

ES 2 392 508 T3

gcccgttaccg ctagctccac cacatcagcg gtaccgatgg gcaacggagt tggagtcggt 1200
 gttgggggtgg ggggcaacgt cagcatgtat gcgaacgccc agatggcgat ggccttgatg 1260
 ggtgtagccc tgcattcgca ccaagagcag cttatcgggg gagtggcggg taagtcggag 1320
 cactcgacga ctgcatag 1338

<210> 3

5 <211> 960

<212> ADN

<213> Choristoneura fumiferana

10

<400> 3

cctgagtgcg tagtaccgga gactcagtgc gccatgaagc ggaaagagaa gaaagcacag	60
aaggagaagg acaaactgcc tgtcagcacg acgacgggtg acgaccacat gccgccatt	120
atgcagtgtg aacctccacc tcctgaagca gcaaggattc acgaagtggg cccaaggttt	180
ctctccgaca agctggttga 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
ttcgcgaagg gattgccagg gttcgccaaag atctcgcagc ctgatcaaat tacgctgctt	480
aaggcttgct caagtgaggt aatgatgctc cgagtgcgac gacgatacga tgcggcctca	540
gacagtgttc tgttcgcgaa caaccaagcg tacactcgcg acaactaccg caaggctggc	600
atggcctacg tcatcgagga tctactgcac ttctgcccgt gcatgtactc tatggcgttg	660
gacaacatcc attacgcgct gotcaaggct gtcgtcatct tttctgaccg gccagggttg	720
gagcagccgc aactggtgga agaaatccag cggtaactacc tgaatacget ccgcatctat	780
atcctgaacc agctgagcgg gtcgggcgct tcgtccgtca tatacggcaa gatcctctca	840
atcctctctg agctacgcac gctcggcatg caaaaactcca acatgtgcat ctccctcaag	900
ctcaagaaca gaaagctgcc gcctttcctc gaggagatct gggatgtggc ggacatgtcg	960

15

<210> 4

<211> 969

20 <212> ADN

<213> Drosophila melanogaster

<400> 4

ES 2 392 508 T3

cggccggaat gcgtcgtccc ggagaaccaa tgtgcgatga agcggcgcga aaagaaggcc 60
 cagaaggaga aggacaaaat gaccacttcg ccgagctctc agcatggcgg caatggcagc 120
 ttggcctctg gtggcggcca agactttggt aagaaggaga ttcttgacct tatgacatgc 180
 gagccgcccc agcatgccac tattcogcta ctacctgatg aaatattggc caagtgtcaa 240
 gcgcgcaata taccttcctt aacgtacaat cagttggcgg ttatatacaa gttaatttgg 300
 taccaggatg gctatgagca gccatctgaa gaggatctca ggcgtataat gagtcaaccc 360
 gatgagaacg agagccaaac ggacgtcagc tttcggcata taaccgagat aaccatactc 420
 acgggccagt tgattggtga gtttgctaaa ggtctaccag cgtttataaaa gataccccag 480
 gaggaccaga tcacgttact aaaggcctgc tcgtcggagg tgatgatgct gcgtatggca 540
 cgacgctatg accacagctc ggactcaata ttcttcgcga ataatagatc atatacgcgg 600
 gattcttaca aaatggccgg aatggctgat aacattgaag acctgctgca tttctgccgc 660
 caaatgttct cgatgaaggt ggacaacgct gaatacgcgc ttctcaactgc cattgtgatc 720
 ttctcggacc ggccgggctt ggagaaggcc caactagtcg aagcgatcca gagctactac 780
 atcgacacgc tacgcattta tatactcaac cgccactgcg gcgactcaat gagcctcgtc 840
 ttctacgcaa agctgctctc gatcctcacc gagctgcgta cgctgggcaa ccagaacgcc 900
 gagatgtggt tctcactaaa gctcaaaaac cgcaaactgc ccaagttcct cgaggagatc 960
 tgggacggt 969

<210> 5

5 <211> 244

<212> PRT

<213> Choristoneura fumiferana

10

<400> 5

Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile
 1 5 10 15

Thr Gln Thr Trp Gln Gln Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr
 20 25 30

Pro Phe Arg Gln Ile Thr Glu Met Thr Ile Leu Thr Val Gln Leu Ile
 35 40 45

Val Glu Phe Ala Lys Gly Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro
 50 55 60

ES 2 392 508 T3

Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu
65 70 75 80

Arg Val Ala Arg Arg Tyr Asp Ala Ala Ser Asp Ser Val Leu Phe Ala
85 90 95

Asn Asn Gln Ala Tyr Thr Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala
100 105 110

Tyr Val Ile Glu Asp Leu Leu His Phe Cys Arg Cys Met Tyr Ser Met
115 120 125

Ala Leu Asp Asn Ile His Tyr Ala Leu Leu Thr Ala Val Val Ile Phe
130 135 140

Ser Asp Arg Pro Gly Leu Glu Gln Pro Gln Leu Val Glu Glu Ile Gln
145 150 155 160

Arg Tyr Tyr Leu Asn Thr Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser
165 170 175

Gly Ser Ala Arg Ser Ser Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu
180 185 190

Ser Glu Leu Arg Thr Leu Gly Met Gln Asn Ser Asn Met Cys Ile Ser
195 200 205

Leu Lys Leu Lys Asn Arg Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp
210 215 220

Asp Val Ala Asp Met Ser His Thr Gln Pro Pro Pro Ile Leu Glu Ser
225 230 235 240

Pro Thr Asn Leu

<210> 6

5 <211> 445

<212> PRT

10 <213> Drosophila melanogaster

<400> 6

ES 2 392 508 T3

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

ES 2 392 508 T3

Phe Tyr Ala Lys Leu Leu Ser Ile Leu Thr Glu Leu Arg Thr Leu Gly
 180 185 190

Asn Gln Asn Ala Glu Met Cys Phe Ser Leu Lys Leu Lys Asn Arg Lys
 195 200 205

Leu Pro Lys Phe Leu Glu Glu Ile Trp Asp Val His Ala Ile Pro Pro
 210 215 220

Ser Val Gln Ser His Leu Gln Ile Thr Gln Glu Glu Asn Glu Arg Leu
 225 230 235 240

Glu Arg Ala Glu Arg Met Arg Ala Ser Val Gly Gly Ala Ile Thr Ala
 245 250 255

Gly Ile Asp Cys Asp Ser Ala Ser Thr Ser Ala Ala Ala Ala Ala Ala
 260 265 270

Gln His Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Ser Ser Leu Thr
 275 280 285

Gln Asn Asp Ser Gln His Gln Thr Gln Pro Gln Leu Gln Pro Gln Leu
 290 295 300

Pro Pro Gln Leu Gln Gly Gln Leu Gln Pro Gln Leu Gln Pro Gln Leu
 305 310 315 320

Gln Thr Gln Leu Gln Pro Gln Ile Gln Pro Gln Pro Gln Leu Leu Pro
 325 330 335

Val Ser Ala Pro Val Pro Ala Ser Val Thr Ala Pro Gly Ser Leu Ser
 340 345 350

Ala Val Ser Thr Ser Ser Glu Tyr Met Gly Gly Ser Ala Ala Ile Gly
 355 360 365

Pro Ile Thr Pro Ala Thr Thr Ser Ser Ile Thr Ala Ala Val Thr Ala
 370 375 380

Ser Ser Thr Thr Ser Ala Val Pro Met Gly Asn Gly Val Gly Val Gly
 385 390 395 400

Val Gly Val Gly Gly Asn Val Ser Met Tyr Ala Asn Ala Gln Thr Ala
 405 410 415

ES 2 392 508 T3

Met Ala Leu Met Gly Val Ala Leu His Ser His Gln Glu Gln Leu Ile
 420 425 430

Gly Gly Val Ala Val Lys Ser Glu His Ser Thr Thr Ala
 435 440 445

<210> 7

5 <211> 320

<212> PRT

<213> Choristoneura fumiferana

10

<400> 7

Pro Glu Cys Val Val Pro Glu Thr Gln Cys Ala Met Lys Arg Lys Glu
 1 5 10 15

Lys Lys Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser Thr Thr Thr
 20 25 30

Val Asp Asp His Met Pro Pro Ile Met Gln Cys Glu Pro Pro Pro Pro
 35 40 45

Glu Ala Ala Arg Ile His Glu Val Val Pro Arg Phe Leu Ser Asp Lys
 50 55 60

Leu Leu Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn
 65 70 75 80

Gln Gln Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu
 85 90 95

Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln
 100 105 110

Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr
 115 120 125

Glu Met Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly
 130 135 140

Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu
 145 150 155 160

Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr
 165 170 175

Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr
 180 185 190

ES 2 392 508 T3

Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu
 195 200 205

Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His
 210 215 220

Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu
 225 230 235 240

Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr
 245 250 255

Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser
 260 265 270

Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu
 275 280 285

Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu Lys Asn Arg
 290 295 300

Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp Asp Val Ala Asp Met Ser
 305 310 315 320

<210> 8

5 <211> 323

<212> PRT

<213> Drosophila melanogaster

10

<400> 8

Arg Pro Glu Cys Val Val Pro Glu Asn Gln Cys Ala Met Lys Arg Arg
 1 5 10 15

Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Met Thr Thr Ser Pro Ser
 20 25 30

Ser Gln His Gly Gly Asn Gly Ser Leu Ala Ser Gly Gly Gly Gln Asp
 35 40 45

Phe Val Lys Lys Glu Ile Leu Asp Leu Met Thr Cys Glu Pro Pro Gln
 50 55 60

His Ala Thr Ile Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys Cys Gln
 65 70 75 80

ES 2 392 508 T3

Ala Arg Asn Ile Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val Ile Tyr
85 90 95

Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu Glu Asp
100 105 110

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

Val Ser Phe Arg His Ile Thr Glu Ile Thr Ile Leu Thr Val Gln Leu
130 135 140

Ile Val Glu Phe Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile Pro Gln
145 150 155 160

Glu Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met
165 170 175

Leu Arg Met Ala Arg Arg Tyr Asp His Ser Ser Asp Ser Ile Phe Phe
180 185 190

Ala Asn Asn Arg Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala Gly Met
195 200 205

Ala Asp Asn Ile Glu Asp Leu Leu His Phe Cys Arg Gln Met Phe Ser
210 215 220

Met Lys Val Asp Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile
225 230 235 240

Phe Ser Asp Arg Pro Gly Leu Glu Lys Ala Gln Leu Val Glu Ala Ile
245 250 255

Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn Arg His
260 265 270

Cys Gly Asp Ser Met Ser Leu Val Phe Tyr Ala Lys Leu Leu Ser Ile
275 280 285

Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ala Glu Met Cys Phe
290 295 300

Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Lys Phe Leu Glu Glu Ile
305 310 315 320

Trp Asp Val

ES 2 392 508 T3

<210> 9

<211> 714

5

<212> ADN

<213> Mus musculus

10

<400> 9

```

gccaacgagg acatgcctgt agagaagatt ctggaagccg agcttgctgt cgagcccaag      60
actgagacat acgtggagge aaacatgggg ctgaacccca gctcaccaaa tgaccctggt      120
accaacatct gtcaagcagc agacaagcag ctcttcactc ttgtggagtg ggccaagagg      180
atcccacact tttctgagct gcccttagac gaccaggtca tcctgctacg ggcaggctgg      240
aacgagctgc tgatcgctc cttctccac cgctccatag ctgtgaaaga tgggattctc      300
ctggccaccg gcctgcacgt acaccggaac agcgctcaca gtgctggggg gggcgccatc      360
tttgacaggg tgctaacaga gctggtgtct aagatgcgtg acatgcagat ggacaagacg      420
gagctgggct gcctgcgagc cattgtcctg ttcaacctg actotaaggg gctctcaaac      480
cctgctgagg tggaggcggt gagggagaag gtgtatgcgt cactagaagc gtactgcaaa      540
cacaagtacc ctgagcagcc gggcaggttt gccaaagctgc tgctccgctt gcctgcactg      600
cgttccatcg ggctcaagtg cctggagcac ctgtttctct tcaagctcat cggggacacg      660
cccategaca ccttctcat ggagatgctg gaggcaccac atcaagccac ctag          714
    
```

<210> 10

15

<211> 720

<212> ADN

20

<213> Mus musculus

<400> 10

```

gccctgagg agatgcctgt ggacaggatc ctggaggcag agcttgctgt ggagcagaag      60
agtgaccaag gcgttgaggg tcctggggcc accgggggtg gtggcagcag cccaaatgac      120
ccagtgacta acatctgcc aagcagctgac aaacagctgt tcacactcgt tgagtgggca      180
aagaggatcc cgcacttctc ctccctacct ctggacgata aggtcactat gctgcgggca      240
ggctggaacg agctcctcat tgcgtcctc tccatcgggt ccattgatgt ccgagatggc      300
atcctcctgg ccacgggtct tcatgtgcac agaaactcag cccattccgc aggcgtggga      360
gccatctttg atcgggtgct gacagagcta gtgtccaaaa tgcgtgacat gaggatggac      420
aagacagagc ttggctgcct gcgggcaatc atcatgttta atccagacgc caagggcctc      480
tccaaccctg gagaggtgga gatccttcgg gagaaggtgt acgcctcact ggagacctat      540
    
```

ES 2 392 508 T3

tgcaagcaga agtaccctga gcagcagggc cggtttgcca agctgctgtt acgtcttctt 600
 gccctccgct ccatcggect caagtgtctg gagcacctgt tcttcttcaa gctcattggc 660
 gacaccccca ttgacacctt cctcatggag atgcttgagg ctccccacca gctagcctga 720

<210> 11

5 <211> 705

<212> ADN

10 <213> Mus musculus

<400> 11

agccacgaag acatgcccggt ggagaggatt ctagaagccg aacttgctgt ggaaccaaag	60
acagaatcct acggtgacat gaacgtggag aactcaacaa atgaccctgt taccaacata	120
tgccatgctg cagataagca acttttcacc ctcgttgagt gggccaaacg catccccac	180
ttctcagatc tcaccttggg ggaccaggtc attctactcc gggcagggtg gaatgaactg	240
ctcattgcct ccttctccca ccgctcgggt tccgtccagg atggcatcct gctggccacg	300
ggcctccacg tgcacaggag cagcgcctac agccggggag tcggctccat cttegacaga	360
gtccttacag agttgggtgt caagatgaaa gacatgcaga tggataagtc agagctgggg	420
tgcctacggg ccatcggtgt gtttaaccca gatgccaaagg gtttatccaa cccctctgag	480
gtggagactc ttcgagagaa ggtttatgcc accctggagg cctataccaa gcagaagtat	540
ccggaacagc caggcaggtt tgccaagctt ctgctgcgtc tccctgctct gcgctccatc	600
ggcttgaaat gcctggaaca cctcttcttc ttcaagctca ttggagacac tcccatcgac	660
agcttctca tggagatggt ggagacccca ctgcagatca cctga	705

15 <210> 12

<211> 850

20 <212> ADN

<213> Homo sapiens

<400> 12

ES 2 392 508 T3

gccaacgagg acatgccggt ggagaggatc ctggaggctg agctggccgt ggagcccaag 60
 accgagacct acgtggaggc aaacatgggg ctgaaccca gctgccgaa cgacctgtc 120
 accaacattt gccaaagcgc cgacaaacag cttttcacc tggtggagtg ggccaagcgg 180
 atcccacact tctcagagct gcccctggac gaccaggtea tcttgcctgc ggcaggctgg 240
 aatgagctgc tcatgcctc cttctcccac cgctccatcg ccgtgaagga cgggatcctc 300
 ctggccaccg ggctgcacgt ccaccggaac agcgcaccaca gcgcaggggt gggcgccatc 360
 tttgacaggg tgctgacgga gcttgtgtcc aagatgcggg acatgcagat ggacaagacg 420
 gagctgggct gcctgcgcgc catcgtcctc tttaacctcg actccaaggg gctctogaac 480
 ccggccgagg tggagggcgt gagggagaag gtctatgcgt ccttggaggc ctactgcaag 540
 cacaagtacc cagagcagcc ggaaggttc gctaagctct tgctccgct gccggctctg 600
 cgctccatcg ggctcaaatg cctggaacat ctcttctct tcaagctcat cggggacaca 660
 ccattgaca ccttcttat ggagatgctg gaggcgcgc accaaatgac ttaggcctgc 720
 gggcccatec tttgtgccc cccgttctgg ccaccctgcc tggacgccag ctgttctct 780
 cagcctgagc cctgtccctg ccttctctg cctggcctgt ttggactttg gggcacagcc 840
 tgtcactgct 850

<210> 13

5

<211> 720

<212> ADN

10

<213> Homo sapiens

<400> 13

gccccgagg agatgcctgt ggacaggatc ctggaggcag agcttgctgt ggaacagaag 60
 agtgaccagg gcgttgaggg tcctggggga accgggggta gcggcagcag cccaaatgac 120
 cctgtgacta acatctgtca ggcagctgac aaacagctat tcaogcttgt tgagtgggcg 180
 aagaggatcc cacactttc ctcttgcct ctggatgato aggtcatatt gctgcgggca 240
 ggctggaatg aactcctcat tgctccttt tcacaccgat ccattgatgt tcgagatggc 300
 atctccttg ccacaggtct tcacgtgcac cgcaactcag ccattcagc aggagtagga 360
 gccatctttg atcgggtgct gacagagcta gtgtccaaa tgogtgacat gaggatggac 420
 aagacagagc ttggctgcct gagggcaatc attctgttta atccagatgc caagggcctc 480
 tccaacccta gtgaggtgga ggtcctgcgg gagaaagtgt atgcatcact ggagacctac 540
 tgcaaacaga agtaccctga gcagcagggc cggtttgcoa agctgctgct acgtcttct 600
 gcctccggt ccattggcct taagtgtcta gagcatctgt ttttcttcaa gctcattggt 660

ES 2 392 508 T3

gacaccccca tcgacacctt cctcatggag atgcttgagg ctccccatca actggcctga 720

<210> 14

5 <211> 705

<212> ADN

<213> Homo sapiens

10

<400> 14

ggtcataaag acatgcctgt ggagaggatt ctagaagctg aacttgctgt tgaaccaaag 60
 acagaatcct atggtgacat gaatatggag aactcgacaa atgaccctgt taccaacata 120
 tgtcatgctg ctgacaagca gcttttcacc ctcgttgaat gggccaagcg tattccccac 180
 ttctctgacc tcacctgga ggaccaggtc attttgcttc gggcaggggtg gaatgaattg 240
 ctgattgect ctttctccca ccgetcagtt tccgtgcagg atggcatcct tctggccacg 300
 ggtttacatg tccacoggag cagtgccac agtgctgggg tcggctccat ctttgacaga 360
 gttctaactg agctggtttc caaaatgaaa gacatgcaga tggacaagtc ggaactggga 420
 tgctgagag ccattgtact ctttaaccca gatgccaaagg gctgtccaa cccctctgag 480
 gtggagactc tgcgagagaa ggtttatgcc acccttgagg cctacaccaa gcagaagtat 540
 ccggaacagc caggcaggtt tgccaagctg ctgctgcgcc tccagctct gcgttccatt 600
 ggcttgaaat gectggagca cctcttcttc ttcaagctca tcggggacac cccattgac 660
 accttctca tggagatggt ggagaccccg ctgcagatca cctga 705

15

<210> 15

<211> 237

20

<212> PRT

<213> Mus musculus

<400> 15

25

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

ES 2 392 508 T3

Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp
65 70 75 80

Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys
85 90 95

Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala
100 105 110

His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu
115 120 125

Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys
130 135 140

Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn
145 150 155 160

Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu
165 170 175

Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys
180 185 190

Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu
195 200 205

Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr
210 215 220

Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Ala Thr
225 230 235

<210> 16

5

<211> 239

<212> PRT

10 <213> Mus musculus

<400> 16

ES 2 392 508 T3

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

Val Glu Gln Lys Ser Asp Gln Gly Val Glu Gly Pro Gly Ala Thr Gly
20 25 30

Gly Gly Gly Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala
35 40 45

Ala Asp Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro
50 55 60

His Phe Ser Ser Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala
65 70 75 80

Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Asp
85 90 95

Val Arg Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn
100 105 110

Ser Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr
115 120 125

Glu Leu Val Ser Lys Met Arg Asp Met Arg Met Asp Lys Thr Glu Leu
130 135 140

Gly Cys Leu Arg Ala Ile Ile Met Phe Asn Pro Asp Ala Lys Gly Leu
145 150 155 160

Ser Asn Pro Gly Glu Val Glu Ile Leu Arg Glu Lys Val Tyr Ala Ser
165 170 175

Leu Glu Thr Tyr Cys Lys Gln Lys Tyr Pro Glu Gln Gln Gly Arg Phe
180 185 190

Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys
195 200 205

Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile
210 215 220

Asp Thr Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Leu Ala
225 230 235

ES 2 392 508 T3

<211> 234

<212> PRT

5

<213> Mus musculus

<400> 17

Ser His Glu Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu Leu Ala
1 5 10 15

Val Glu Pro Lys Thr Glu Ser Tyr Gly Asp Met Asn Val Glu Asn Ser
20 25 30

Thr Asn Asp Pro Val Thr Asn Ile Cys His Ala Ala Asp Lys Gln Leu
35 40 45

Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe Ser Asp Leu
50 55 60

Thr Leu Glu Asp Gln Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu
65 70 75 80

Leu Ile Ala Ser Phe Ser His Arg Ser Val Ser Val Gln Asp Gly Ile
85 90 95

Leu Leu Ala Thr Gly Leu His Val His Arg Ser Ser Ala His Ser Arg
100 105 110

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

Met Lys Asp Met Gln Met Asp Lys Ser Glu Leu Gly Cys Leu Arg Ala
130 135 140

Ile Val Leu Phe Asn Pro Asp Ala Lys Gly Leu Ser Asn Pro Ser Glu
145 150 155 160

Val Glu Thr Leu Arg Glu Lys Val Tyr Ala Thr Leu Glu Ala Tyr Thr
165 170 175

Lys Gln Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Leu
180 185 190

Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu
195 200 205

ES 2 392 508 T3

Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Ser Phe Leu Met
 210 215 220

Glu Met Leu Glu Thr Pro Leu Gln Ile Thr
 225 230

<210> 18

5 <211> 237

<212> PRT

<213> Homo sapiens

10

<400> 18

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

Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn
 20 25 30

Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp
 35 40 45

Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe
 50 55 60

Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp
 65 70 75 80

Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys
 85 90 95

Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala
 100 105 110

His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu
 115 120 125

Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys
 130 135 140

Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn
 145 150 155 160

Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu
 165 170 175

ES 2 392 508 T3

Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys
 180 185 190

Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu
 195 200 205

Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr
 210 215 220

Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Met Thr
 225 230 235

<210> 19

5 <211> 239

<212> PRT

<213> Homo sapiens

10

<400> 19

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

Val Glu Gln Lys Ser Asp Gln Gly Val Glu Gly Pro Gly Gly Thr Gly
 20 25 30

Gly Ser Gly Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala
 35 40 45

Ala Asp Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro
 50 55 60

His Phe Ser Ser Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala
 65 70 75 80

Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Asp
 85 90 95

Val Arg Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn
 100 105 110

Ser Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr
 115 120 125

Glu Leu Val Ser Lys Met Arg Asp Met Arg Met Asp Lys Thr Glu Leu
 130 135 140

ES 2 392 508 T3

Gly Cys Leu Arg Ala Ile Ile Leu Phe Asn Pro Asp Ala Lys Gly Leu
 145 150 155 160

Ser Asn Pro Ser Glu Val Glu Val Leu Arg Glu Lys Val Tyr Ala Ser
 165 170 175

Leu Glu Thr Tyr Cys Lys Gln Lys Tyr Pro Glu Gln Gln Gly Arg Phe
 180 185 190

Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys
 195 200 205

Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile
 210 215 220

Asp Thr Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Leu Ala
 225 230 235

<210> 20

5 <211> 234

<212> PRT

<213> Homo sapiens

10

<400> 20

Gly His Glu Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu Leu Ala
 1 5 10 15

Val Glu Pro Lys Thr Glu Ser Tyr Gly Asp Met Asn Met Glu Asn Ser
 20 25 30

Thr Asn Asp Pro Val Thr Asn Ile Cys His Ala Ala Asp Lys Gln Leu
 35 40 45

Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe Ser Asp Leu
 50 55 60

Thr Leu Glu Asp Gln Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu
 65 70 75 80

Leu Ile Ala Ser Phe Ser His Arg Ser Val Ser Val Gln Asp Gly Ile
 85 90 95

Leu Leu Ala Thr Gly Leu His Val His Arg Ser Ser Ala His Ser Ala
 100 105 110

ES 2 392 508 T3

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

Met Lys Asp Met Gln Met Asp Lys Ser Glu Leu Gly Cys Leu Arg Ala
 130 135 140

Ile Val Leu Phe Asn Pro Asp Ala Lys Gly Leu Ser Asn Pro Ser Glu
 145 150 155 160

Val Glu Thr Leu Arg Glu Lys Val Tyr Ala Thr Leu Glu Ala Tyr Thr
 165 170 175

Lys Gln Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Leu
 180 185 190

Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu
 195 200 205

Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met
 210 215 220

Glu Met Leu Glu Thr Pro Leu Gln Ile Thr
 225 230

<210> 21

5 <211> 635

<212> ADN

<213> Locusta migratoria

10

<400> 21

tgcatacaga catgcctggt gaacgcatac ttgaagctga aaaacgagtg gagtgcmaaag	60
cagaaaacca agtgggaatat gagctggtgg agtgggctaa acacatcccg cacttcacat	120
ccctacctct ggaggaccag gttctcctcc tcagagcagg ttggaatgaa ctgctaattg	180
cagcattttc acatcgatct gtagatgta aagatggcat agtacttgcc actggtctca	240
cagtgcacgc aaattctgcc catcaagctg gagtcggcac aatatttgac agagtttga	300
cagaactggt agcaaagatg agagaaatga aatggataa aactgaactt ggctgcttgc	360
gatctgttat tcttttcaat ccagaggtga ggggtttgaa atccgcccag gaagttgaac	420
ttctacgtga aaaagtatat gccgctttgg aagaatatac tagaacaaca catcccgatg	480
aaccaggaag atttgcaaaa cttttgcttc gtctgccttc tttacgttcc ataggcctta	540
agtgtttggg gcatttgttt ttctttcgcc ttattggaga tgttccaatt gatacgttcc	600

tgatggagat gottgaaatca ccttctgatt cataa

635

<210> 22

5 <211> 687

<212> ADN

<213> Amblyomma americanum

10

<400> 22

cctcctgaga tgcctctgga ggcatactg gaggcagagc tgcgggttga gtcacagacg	60
gggaccctct cggaaagcgc acagcagcag gatccagtga gcagcatctg ccaagctgca	120
gaccgacagc tgcaccagct agttcaatgg gccaaagcaca ttccacattt tgaagagctt	180
ccccttgagg accgcatggt gttgctcaag gctggctgga acgagctgct cattgctgct	240
ttctcccacc gttctgttga cgtgogtgat ggcattgtgc tcgctacagg tcttgtggtg	300
cagcggcata gtgctcatgg ggctggcgtt ggggccatat ttgatagggt tctcaactgaa	360
ctggtagcaa agatgcgtga gatgaagatg gacgcgactg agcttgatg cctgcttgcct	420
gtggtacttt ttaatcctga ggccaagggg ctgcggacct gcccaagtgg aggccctgag	480
ggagaaagtg tatctgcctt ggaagagcac tgccggcagc agtaccacaga ccagcctggg	540
cgctttgcca agctgctgct gcggttgcca gctctgcgca gtattggcct caagtgcctc	600
gaacatctct tttctctcaa gctcatcggg gacacgcca togacaactt tcttcttcc	660
atgctggagg cccctctga cccctaa	687

15 <210> 23

<211> 693

<212> ADN

20

<213> Amblyomma americanum

<400> 23

tctccggaca tgccactcga acgcattctc gaagccgaga tgcgcgtoga gcagccggca	60
ccgtccgttt tggcgcagac ggccgcctcg ggccgcgacc ccgtcaacag catgtgccag	120
gctgccccgc cacttcacga gctcgtacag tgggccccgc gaattccgca cttcgaagag	180
cttcccatcg aggatcgcac cgcgctgctc aaagccggct ggaacgaact gcttattgcc	240

ES 2 392 508 T3

gccttttcgc accgttctgt ggcgggtgcgc gacggcatcg ttctggccac cgggctggtg 300
 gtgcagcggc acagcgcaca cggcgcaggc gttggcgaca tcttcgaccg cgtactagcc 360
 gagctggtgg ccaagatgcg cgacatgaag atggacaaaa cggagctcgg ctgcctgcgc 420
 gccgtggtgc tcttcaatcc agacgccaaag ggtctccgaa acgccaccag agtagaggcg 480
 ctccgcgaga aggtgtatgc ggcgctggag gagcactgcc gtccggacca cccggaccaa 540
 ccgggtcgcct tcggcaagct gctgctgcgg ctgcctgcct tgcgcagcat cgggctcaaa 600
 tgctcgagc atctgttctt cttcaagctc atcggagaca ctccataga cagcttcctg 660
 ctcaacatgc tggaggcacc ggcagacccc tag 693

<210> 24

5 <211> 801

<212> ADN

<213> Celuca pugilator

10

<400> 24

tcagacatgc caattgccag catacgggag gcagagctca gcgtggatcc catagatgag 60
 cagccgctgg accaaggggt gaggettccag gttccactcg cacctcctga tagtgaaaag 120
 tgtagcttta ctttaccttt tcateccgctc agtgaagtat cctgtgctaa ccctctgcag 180
 gatgtggtga gcaacatatg ccaggcagct gacagacatc tgggtgcagct ggtggagtgg 240
 gccaaagcaca tcccacactt cacagacett cccatagagg accaagtggg attactcaaa 300
 gccgggtgga acgagttgct tattgcctca ttctcacacc gtagcatggg cgtggaggat 360
 ggcacgtgtc tggccacagg gctcgtgatc cacagaagta gtgctcacca ggctggagtg 420
 ggtgccaatc ttgatcgtgt cctctctgag ctgggtggcca agatgaagga gatgaagatt 480
 gacaagacag agctgggctg ccttcgctcc atcgtcctgt tcaaccaga tgccaaagga 540
 ctaaactgcg tcaatgatgt ggagatcttg cgtgagaagg tgtatgctgc cctggaggag 600
 tacacaogaa ccacttacc tgatgaacct ggacgcttg ccaagttgct tetgcgactt 660
 cctgcactca ggtctatagg cctgaagtgt cttgagtacc tcttctggt taagctgatt 720
 ggagacactc ccctggacag ctacttgatg aagatgctcg tagacaaccc aaatacaagc 780
 gtcactcccc ccaaccagcta g 801

15 <210> 25

<211> 690

<212> ADN

20

<213> Tenebrio molitor

ES 2 392 508 T3

<400> 25

gccgagatgc ccctcgacag gataatcgag gcggagaaac ggatagaatg cacacccgct 60
 ggtggctctg gtgggtgctgg agagcaacac gacgggggtga acaacatctg tcaagccact 120
 aacaagcagc tgttccaact ggtgcaatgg gctaagctca tacctcactt tacctcgttg 180
 ccgatgtcgg accaggtgct tttattgagg gcaggatgga atgaattgct catcgccgca 240
 ttctcgcaca gatotataca ggcgcaggat gccatcgttc tagccacggg gttgacagtt 300
 aacaaaacgt cggcgcacgc cgtgggcgtg ggcaacatct acgaccgct cctctccgag 360
 ctggtgaaca agatgaaaga gatgaagatg gacaagacgg agctgggctg cttgagagcc 420
 atcatcctct acaacccccac gtgtcgcggc atcaagtccg tgcaggaagt ggagatgctg 480
 cgtgagaaaa tttacggcgt gctggaagag tacaccagga ccaccaccc gaacgagccc 540
 ggcaggttcg ccaaactgct tctgcgcctc cgggcctca ggtccatcgg gttgaaatgt 600
 tccgaacacc tctttttctt caagctgac ggtgatgttc caatagacac gttcctgatg 660
 5 gagatgctgg agtctccggc ggacgcttag 690

<210> 26

<211> 681

10

<212> ADN

<213> Apis mellifera

15

<400> 26

cattcggaca tgccgatcga gcgtatcctg gaggccgaga agagagtcca atgtaagatg 60
 gagcaacagg gaaattacga gaatgcagtg togcacattt gcaacgccac gaacaaacag 120
 ctgttccagc tggtagcatg ggogaaacac atccccgatt ttacctcgtt gccactggag 180
 gatcaggtac ttctgctcag ggccggttgg aacgagttgc tgatagcctc cttttcccac 240
 cgttccatcg acgtgaagga cggtatcgtg ctggcgacgg ggatcacctg gcatcggaac 300
 teggcgcagc aggccggcgt gggcacgata ttcgaccgtg tcctctcggg gcttgtctcg 360
 aaaatgcgtg aatgaagat ggacaggaca gagcttggct gtctcagatc tataatactc 420
 ttcaatcccg aggttcgagg actgaaatcc atccaggaag tgaccctgct ccgtgagaag 480
 atctacggcg ccctggaggg ttattgccgc gtagcttggc ccgacgacgc tggaagattc 540
 gcgaaattac ttctacgcct gcccgccatc cgctcgatcg gattaaagtg cctcgagtac 600
 ctgttcttct tcaaaatgat cggtgacgta ccgatcgaag attttctcgt ggagatgtta 660
 gaatcgcgat cagatcctta g 681

ES 2 392 508 T3

<210> 27

<211> 210

5 <212> PRT

<213> *Locusta migratoria*

<400> 27

10

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

ES 2 392 508 T3

<210> 28

<211> 228

5

<212> PRT

<213> Amblyomma americanum

10 <400> 28

Pro Pro Glu Met Pro Leu Glu Arg Ile Leu Glu Ala Glu Leu Arg Val
1 5 10 15

Glu Ser Gln Thr Gly Thr Leu Ser Glu Ser Ala Gln Gln Gln Asp Pro
20 25 30

Val Ser Ser Ile Cys Gln Ala Ala Asp Arg Gln Leu His Gln Leu Val
35 40 45

Gln Trp Ala Lys His Ile Pro His Phe Glu Glu Leu Pro Leu Glu Asp
50 55 60

Arg Met Val Leu Leu Lys Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala
65 70 75 80

Phe Ser His Arg Ser Val Asp Val Arg Asp Gly Ile Val Leu Ala Thr
85 90 95

Gly Leu Val Val Gln Arg His Ser Ala His Gly Ala Gly Val Gly Ala
100 105 110

Ile Phe Asp Arg Val Leu Thr Glu Leu Val Ala Lys Met Arg Glu Met
115 120 125

Lys Met Asp Arg Thr Glu Leu Gly Cys Leu Leu Ala Val Val Leu Phe
130 135 140

Asn Pro Glu Ala Lys Gly Leu Arg Thr Cys Pro Ser Gly Gly Pro Glu
145 150 155 160

Gly Glu Ser Val Ser Ala Leu Glu Glu His Cys Arg Gln Gln Tyr Pro
165 170 175

Asp Gln Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu
180 185 190

Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys Leu
195 200 205

ES 2 392 508 T3

Ile Gly Asp Thr Pro Ile Asp Asn Phe Leu Leu Ser Met Leu Glu Ala
 210 215 220

Pro Ser Asp Pro
 225

<210> 29

5 <211> 230

<212> PRT

<213> Amblyomma americanum

10

<400> 29

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

Glu Gln Pro Ala Pro Ser Val Leu Ala Gln Thr Ala Ala Ser Gly Arg
 20 25 30

Asp Pro Val Asn Ser Met Cys Gln Ala Ala Pro Pro Leu His Glu Leu
 35 40 45

Val Gln Trp Ala Arg Arg Ile Pro His Phe Glu Glu Leu Pro Ile Glu
 50 55 60

Asp Arg Thr Ala Leu Leu Lys Ala Gly Trp Asn Glu Leu Leu Ile Ala
 65 70 75 80

Ala Phe Ser His Arg Ser Val Ala Val Arg Asp Gly Ile Val Leu Ala
 85 90 95

Thr Gly Leu Val Val Gln Arg His Ser Ala His Gly Ala Gly Val Gly
 100 105 110

Asp Ile Phe Asp Arg Val Leu Ala Glu Leu Val Ala Lys Met Arg Asp
 115 120 125

Met Lys Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Val Val Leu
 130 135 140

Phe Asn Pro Asp Ala Lys Gly Leu Arg Asn Ala Thr Arg Val Glu Ala
 145 150 155 160

Leu Arg Glu Lys Val Tyr Ala Ala Leu Glu Glu His Cys Arg Arg His
 165 170 175

ES 2 392 508 T3

His Pro Asp Gln Pro Gly Arg Phe Gly Lys Leu Leu Leu Arg Leu Pro
 180 185 190

Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe
 195 200 205

Lys Leu Ile Gly Asp Thr Pro Ile Asp Ser Phe Leu Leu Asn Met Leu
 210 215 220

Glu Ala Pro Ala Asp Pro
 225 230

<210> 30

5

<211> 266

<212> PRT

10 <213> Celuca pugilator

<400> 30

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

Pro Ile Asp Glu Gln Pro Leu Asp Gln Gly Val Arg Leu Gln Val Pro
 20 25 30

Leu Ala Pro Pro Asp Ser Glu Lys Cys Ser Phe Thr Leu Pro Phe His
 35 40 45

Pro Val Ser Glu Val Ser Cys Ala Asn Pro Leu Gln Asp Val Val Ser
 50 55 60

Asn Ile Cys Gln Ala Ala Asp Arg His Leu Val Gln Leu Val Glu Trp
 65 70 75 80

Ala Lys His Ile Pro His Phe Thr Asp Leu Pro Ile Glu Asp Gln Val
 85 90 95

Val Leu Leu Lys Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser
 100 105 110

His Arg Ser Met Gly Val Glu Asp Gly Ile Val Leu Ala Thr Gly Leu
 115 120 125

ES 2 392 508 T3

Val Ile His Arg Ser Ser Ala His Gln Ala Gly Val Gly Ala Ile Phe
 130 135 140

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

Asp Lys Thr Glu Leu Gly Cys Leu Arg Ser Ile Val Leu Phe Asn Pro
 165 170 175

Asp Ala Lys Gly Leu Asn Cys Val Asn Asp Val Glu Ile Leu Arg Glu
 180 185 190

Lys Val Tyr Ala Ala Leu Glu Glu Tyr Thr Arg Thr Thr Tyr Pro Asp
 195 200 205

Glu Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg
 210 215 220

Ser Ile Gly Leu Lys Cys Leu Glu Tyr Leu Phe Leu Phe Lys Leu Ile
 225 230 235 240

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

Pro Asn Thr Ser Val Thr Pro Pro Thr Ser
 260 265

<210> 31

5 <211> 229

<212> PRT

10 <213> Tenebrio molitor

<400> 31

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

Cys Thr Pro Ala Gly Gly Ser Gly Gly Val Gly Glu Gln His Asp Gly
 20 25 30

Val Asn Asn Ile Cys Gln Ala Thr Asn Lys Gln Leu Phe Gln Leu Val
 35 40 45

ES 2 392 508 T3

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

Gln Val Leu Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala
65 70 75 80

Phe Ser His Arg Ser Ile Gln Ala Gln Asp Ala Ile Val Leu Ala Thr
85 90 95

Gly Leu Thr Val Asn Lys Thr Ser Ala His Ala Val Gly Val Gly Asn
100 105 110

Ile Tyr Asp Arg Val Leu Ser Glu Leu Val Asn Lys Met Lys Glu Met
115 120 125

Lys Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Ile Leu Tyr
130 135 140

Asn Pro Thr Cys Arg Gly Ile Lys Ser Val Gln Glu Val Glu Met Leu
145 150 155 160

Arg Glu Lys Ile Tyr Gly Val Leu Glu Glu Tyr Thr Arg Thr Thr His
165 170 175

Pro Asn Glu Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala
180 185 190

Leu Arg Ser Ile Gly Leu Lys Cys Ser Glu His Leu Phe Phe Phe Lys
195 200 205

Leu Ile Gly Asp Val Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu
210 215 220

Ser Pro Ala Asp Ala
225

<210> 32

5

<211> 226

<212> PRT

10 <213> Apis mellifera

<400> 32

ES 2 392 508 T3

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

Glu Cys Lys Met Glu Gln Gln Gly Asn Tyr Glu Asn Ala Val Ser His
 20 25 30

Ile Cys Asn Ala Thr Asn Lys Gln Leu Phe Gln Leu Val Ala Trp Ala
 35 40 45

Lys His Ile Pro His Phe Thr Ser Leu Pro Leu Glu Asp Gln Val Leu
 50 55 60

Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His
 65 70 75 80

Arg Ser Ile Asp Val Lys Asp Gly Ile Val Leu Ala Thr Gly Ile Thr
 85 90 95

Val His Arg Asn Ser Ala Gln Gln Ala Gly Val Gly Thr Ile Phe Asp
 100 105 110

Arg Val Leu Ser Glu Leu Val Ser Lys Met Arg Glu Met Lys Met Asp
 115 120 125

Arg Thr Glu Leu Gly Cys Leu Arg Ser Ile Ile Leu Phe Asn Pro Glu
 130 135 140

Val Arg Gly Leu Lys Ser Ile Gln Glu Val Thr Leu Leu Arg Glu Lys
 145 150 155 160

Ile Tyr Gly Ala Leu Glu Gly Tyr Cys Arg Val Ala Trp Pro Asp Asp
 165 170 175

Ala Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Ile Arg Ser
 180 185 190

Ile Gly Leu Lys Cys Leu Glu Tyr Leu Phe Phe Phe Lys Met Ile Gly
 195 200 205

Asp Val Pro Ile Asp Asp Phe Leu Val Glu Met Leu Glu Ser Arg Ser
 210 215 220

Asp Pro
 225

ES 2 392 508 T3

<211> 516

<212> ADN

5

<213> *Locusta migratoria*

<400> 33

```

atccctacct ctggaggacc aggttctcct cctcagagca ggttggaatg aactgctaata 60
tgcagcattt tcacatcgat ctgtagatgt taaagatggc atagtacttg ccaactggctc 120
cacagtgcac cgaaattctg cccatcaagc tggagtcggc acaatatttg acagagtttt 180
gacagaactg gtagcaaaga tgagagaaat gaaaatggat aaaactgaac ttggctgctt 240
gcgatctggt attcttttca atccagaggt gaggggtttg aaatccgccc aggaagttga 300
acttctacgt gaaaaagtat atgccgcttt ggaagaatat actagaacaa cacatcccga 360
tgaaccagga agatttgcaa aacttttgc tgcctctgct tctttacggt ccataggcct 420
taagtgtttg gagcatttgt tttctttcgc cttattggag atgttccaat tgatacgttc 480
ctgatggaga tgcttgaatc accttctgat tcataa 516

```

10

<210> 34

<211> 528

15

<212> ADN

<213> *Amblyomma americanum*

<400> 34

20

```

attccacatt ttgaagagct tccccttgag gaccgcatgg tgttgctcaa ggcctggctgg 60
aacgagctgc tcattgctgc tttctcccac cgttctggtg acgtgcgtga tggcattgtg 120
ctcgtacag gtcttggtg gcagcggcat agtgctcatg gggctggcgt tggggccata 180
tttgataggg ttctcactga actggtagca aagatgcgtg agatgaagat ggaccgcact 240
gagcttggat gcctgcttgc tgtggtactt tttaatcctg aggccaaggg gctgcggacc 300
tgcccaagtg gaggcctga gggagaaagt gtatctgctt tggaagagca ctgccggcag 360
cagtaccag accagcctgg gcgctttgcc aagctgctgc tgcggttgcc agctctgcgc 420
agtattggcc tcaagtgcct cgaacatctc tttttcttca agctcatcgg ggacaagccc 480
atcgacaact ttcttcttcc catgctggag gccccctctg acccctaa 528

```

<210> 35

25

<211> 531

<212> ADN

<213> *Amblyomma americanum*

30

ES 2 392 508 T3

<400> 35

attccgcact	tcgaagagct	tcccatcgag	gatcgcaccg	cgctgctcaa	agccggctgg	60
aacgaactgc	ttattgcegc	cttttcgcac	cgttctgtgg	cggtgcgcga	cggcacgctt	120
ctggccaccg	ggctggtggt	gcagcggcac	agcgcacacg	gcgagggcgt	tggcgacatc	180
ttcgaccgcg	tactagccga	gctggtggcc	aagatgcgcg	acatgaagat	ggacaaaacg	240
gagctcggct	gcctgcgcgc	cgtggtgctc	ttcaatccag	acgccaaagg	tctccgaaac	300
gccaccagag	tagagggcgt	ccgcgagaag	gtgtatgcgg	cgctggagga	gcactgcogt	360
cggcaccacc	cggaccaacc	gggtcgcttc	ggcaagctgc	tgctgcggct	gcctgccttg	420
cgcagcatcg	ggctcaaatg	cctcgagcat	ctgttcttct	tcaagctcat	cggagacact	480
cccatagaca	gcttctgct	caacatgctg	gaggcaccgg	cagacccta	g	531

5

<210> 36

<211> 552

10 <212> ADN

<213> Celuca pugilator

<400> 36

15

atcccacact	tcacagacct	tcccatagag	gaccaagtgg	tattactcaa	agccgggtgg	60
aacgagttgc	ttattgcctc	attctcacac	cgtagcatgg	gcgtggagga	tggcatcgtg	120
ctggccacag	ggctcgtgat	ccacagaagt	agtgtctacc	aggctggagt	gggtgccata	180
tttgatcgtg	tcctctctga	gctggtggcc	aagatgaagg	agatgaagat	tgacaagaca	240
gagctgggct	gccttcgctc	catcgtctctg	ttcaaccag	atgccaaagg	actaaactgc	300
gtcaatgatg	tggagatctt	gcgtgagaag	gtgtatgctg	ccctggagga	gtacacacga	360
accaettacc	ctgatgaacc	tggacgcttt	gccaagttgc	ttctgcgact	tcctgcactc	420
aggtctatag	gcctgaagtg	tcttgagtac	ctcttcctgt	ttaagctgat	tggagacact	480
ccctggaca	gctacttgat	gaagatgctc	gtagacaacc	caaatacaag	cgtcactccc	540
cccaccagct	ag					552

20 <210> 37

<211> 531

<212> ADN

25

<213> Tenebrio molitor

<400> 37

ES 2 392 508 T3

atacctcact ttacctcggt gccgatgtcg gaaccaggtgc ttttattgag ggcaggatgg 60
 aatgaattgc tcatcgccgc attctcgcac agatctatac aggcgcagga tgccatcggt 120
 ctagccacgg gggtgacagt taacaaaacg tcggcgcaoc ccgtgggctg gggcaacatc 180
 tacgaccgcy tectctccga gctggtgaac aagatgaaag agatgaagat ggacaagacg 240
 gagctgggct gcttgagagc catcatcctc tacaacocca cgtgtcgcgg catcaagtcc 300
 gtgcaggaag tggagatgct gcgtgagaaa atttacggcg tgctggaaga gtacaccagg 360
 accaccacc cgaacgagcc cggcaggttc gccaaactgc ttctgcgcct cccggccctc 420
 aggtccatcg gggtgaaatg ttccgaacac ctcttttct tcaagctgat cggatgatgtt 480
 ccaatagaca cgttctgat ggagatgctg gactctccgg cggacgctta g 531

<210> 38

5 <211> 531

<212> ADN

<213> Apis mellifera

10

<400> 38

atcccgcat tttacctcggt gccactggag gatcaggtac ttctgctcag ggccgggttg 60
 aacgagttgc tgatagcctc cttttccac cgttccatcg acgtgaagga cggatcctgt 120
 ctggcgacgg ggatcaccgt gcctcggaac tcggcgcaoc aggcgggctg gggcacgata 180
 ttgaccgctg tectctccga gcttgtctcg aaaatgctg aaatgaagat ggacaggaca 240
 gagcttggct gtctcagatc tataatactc ttcaatcccg aggttcgagg actgaaatcc 300
 atccaggaag tgaccctgct ccgtgagaag atctacggcg ccctggaggg ttattgccgc 360
 gtagcttggc ccgacgagc tggagattc gcgaaattac ttctacgcct gcccgccatc 420
 cgctcgatcg gattaaagt cctcgagtac ctgttctct tcaaatgat cggatgacgt 480
 ccgatcgacg atttctcgt ggagatgtta gaatcgcat cagatcctta g 531

15 <210> 39

<211> 176

<212> PRT

20

<213> Locusta migratoria

<400> 39

ES 2 392 508 T3

Ile Pro His Phe Thr Ser Leu Pro Leu Glu Asp Gln Val Leu Leu Leu
 1 5 10 15

Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala Phe Ser His Arg Ser
 20 25 30

Val Asp Val Lys Asp Gly Ile Val Leu Ala Thr Gly Leu Thr Val His
 35 40 45

Arg Asn Ser Ala His Gln Ala Gly Val Gly Thr Ile Phe Asp Arg Val
 50 55 60

Leu Thr Glu Leu Val Ala Lys Met Arg Glu Met Lys Met Asp Lys Thr
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ser Val Ile Leu Phe Asn Pro Glu Val Arg
 85 90 95

Gly Leu Lys Ser Ala Gln Glu Val Glu Leu Leu Arg Glu Lys Val Tyr
 100 105 110

Ala Ala Leu Glu Glu Tyr Thr Arg Thr Thr His Pro Asp Glu Pro Gly
 115 120 125

Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ser Leu Arg Ser Ile Gly
 130 135 140

Leu Lys Cys Leu Glu His Leu Phe Phe Phe Arg Leu Ile Gly Asp Val
 145 150 155 160

Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ser Pro Ser Asp Ser
 165 170 175

<210> 40

5 <211> 175

<212> PRT

<213> Amblyomma americanum

10

<400> 40

Ile Pro His Phe Glu Glu Leu Pro Leu Glu Asp Arg Met Val Leu Leu
 1 5 10 15

ES 2 392 508 T3

Lys Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala Phe Ser His Arg Ser
 20 25 30

Val Asp Val Arg Asp Gly Ile Val Leu Ala Thr Gly Leu Val Val Gln
 35 40 45

Arg His Ser Ala His Gly Ala Gly Val Gly Ala Ile Phe Asp Arg Val
 50 55 60

Leu Thr Glu Leu Val Ala Lys Met Arg Glu Met Lys Met Asp Arg Thr
 65 70 75 80

Glu Leu Gly Cys Leu Leu Ala Val Val Leu Phe Asn Pro Glu Ala Lys
 85 90 95

Gly Leu Arg Thr Cys Pro Ser Gly Gly Pro Glu Gly Glu Ser Val Ser
 100 105 110

Ala Leu Glu Glu His Cys Arg Gln Gln Tyr Pro Asp Gln Pro Gly Arg
 115 120 125

Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu
 130 135 140

Lys Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro
 145 150 155 160

Ile Asp Asn Phe Leu Leu Ser Met Leu Glu Ala Pro Ser Asp Pro
 165 170 175

<210> 41

5 <211> 176

<212> PRT

<213> Amblyomma americanum

10

<400> 41

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

ES 2 392 508 T3

Arg His Ser Ala His Gly Ala Gly Val Gly Asp Ile Phe Asp Arg Val
 50 55 60

Leu Ala Glu Leu Val Ala Lys Met Arg Asp Met Lys Met Asp Lys Thr
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ala Val Val Leu Phe Asn Pro Asp Ala Lys
 85 90 95

Gly Leu Arg Asn Ala Thr Arg Val Glu Ala Leu Arg Glu Lys Val Tyr
 100 105 110

Ala Ala Leu Glu Glu His Cys Arg Arg His His Pro Asp Gln Pro Gly
 115 120 125

Arg Phe Gly Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly
 130 135 140

Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr
 145 150 155 160

Pro Ile Asp Ser Phe Leu Leu Asn Met Leu Glu Ala Pro Ala Asp Pro
 165 170 175

<210> 42

5 <211> 183

<212> PRT

<213> Celuca pugilator

10

<400> 42

ES 2 392 508 T3

Ile Pro His Phe Thr Asp Leu Pro Ile Glu Asp Gln Val Val Leu Leu
 1 5 10 15

Lys Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser
 20 25 30

Met Gly Val Glu Asp Gly Ile Val Leu Ala Thr Gly Leu Val Ile His
 35 40 45

Arg Ser Ser Ala His Gln Ala Gly Val Gly Ala Ile Phe Asp Arg Val
 50 55 60

Leu Ser Glu Leu Val Ala Lys Met Lys Glu Met Lys Ile Asp Lys Thr
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ser Ile Val Leu Phe Asn Pro Asp Ala Lys
 85 90 95

Gly Leu Asn Cys Val Asn Asp Val Glu Ile Leu Arg Glu Lys Val Tyr
 100 105 110

Ala Ala Leu Glu Glu Tyr Thr Arg Thr Thr Tyr Pro Asp Glu Pro Gly
 115 120 125

Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly
 130 135 140

Leu Lys Cys Leu Glu Tyr Leu Phe Leu Phe Lys Leu Ile Gly Asp Thr
 145 150 155 160

Pro Leu Asp Ser Tyr Leu Met Lys Met Leu Val Asp Asn Pro Asn Thr
 165 170 175

Ser Val Thr Pro Pro Thr Ser
 180

5 <210> 43

<211> 176

<212> PRT

10

<213> Tenebrio molitor

<400> 43

ES 2 392 508 T3

Ile Pro His Phe Thr Ser Leu Pro Met Ser Asp Gln Val Leu Leu Leu
 1 5 10 15

Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala Phe Ser His Arg Ser
 20 25 30

Ile Gln Ala Gln Asp Ala Ile Val Leu Ala Thr Gly Leu Thr Val Asn
 35 40 45

Lys Thr Ser Ala His Ala Val Gly Val Gly Asn Ile Tyr Asp Arg Val
 50 55 60

Leu Ser Glu Leu Val Asn Lys Met Lys Glu Met Lys Met Asp Lys Thr
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ala Ile Ile Leu Tyr Asn Pro Thr Cys Arg
 85 90 95

Gly Ile Lys Ser Val Gln Glu Val Glu Met Leu Arg Glu Lys Ile Tyr
 100 105 110

Gly Val Leu Glu Glu Tyr Thr Arg Thr Thr His Pro Asn Glu Pro Gly
 115 120 125

Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly
 130 135 140

Leu Lys Cys Ser Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Val
 145 150 155 160

Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ser Pro Ala Asp Ala
 165 170 175

5 <210> 44

<211> 176

<212> PRT

10

<213> Apis mellifera

<400> 44

ES 2 392 508 T3

Ile Pro His Phe Thr Ser Leu Pro Leu Glu Asp Gln Val Leu Leu Leu
 1 5 10 15

Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser
 20 25 30

Ile Asp Val Lys Asp Gly Ile Val Leu Ala Thr Gly Ile Thr Val His
 35 40 45

Arg Asn Ser Ala Gln Gln Ala Gly Val Gly Thr Ile Phe Asp Arg Val
 50 55 60

Leu Ser Glu Leu Val Ser Lys Met Arg Glu Met Lys Met Asp Arg Thr
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ser Ile Ile Leu Phe Asn Pro Glu Val Arg
 85 90 95

Gly Leu Lys Ser Ile Gln Glu Val Thr Leu Leu Arg Glu Lys Ile Tyr
 100 105 110

Gly Ala Leu Glu Gly Tyr Cys Arg Val Ala Trp Pro Asp Asp Ala Gly
 115 120 125

Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Ile Arg Ser Ile Gly
 130 135 140

Leu Lys Cys Leu Glu Tyr Leu Phe Phe Phe Lys Met Ile Gly Asp Val
 145 150 155 160

Pro Ile Asp Asp Phe Leu Val Glu Met Leu Glu Ser Arg Ser Asp Pro
 165 170 175

5 <210> 45

<211> 711

<212> ADN

10

<213> Secuencia artificial

<220>

15 <223> Dominio de unión a ligando de RXR quimérico

<400> 45

ES 2 392 508 T3

gccaacgagg acatgcctgt agagaagatt ctggaagccg agcttgctgt cgagcccaag 60
 actgagacat acgtggagge aaacatgggg ctgaacccca gctcaccaaa tgaccctggt 120
 accaacatct gtcaagcagc agacaagcag ctcttcactc ttgtggagtg ggccaagagg 180
 atcccacact tttctgagct gccctagac gaccaggtoa tcoctgctacg ggcaggetgg 240
 aacgagctgc tgatgcctc cttctcccac cgctccatag ctgtgaaaga tgggattctc 300
 ctggccaccg gcctgcacgt acaccggaac agcgctcaca gtgctggggg gggcgccatc 360
 tttgacaggg tgctaacaga gctgggtgtc aagatgcgtg acatgcagat ggacaagact 420
 gaacttggct gcttgcgatc tgttattctt ttcaatccag aggtgagggg tttgaaatcc 480
 gccaggaag ttgaacttct acgtgaaaaa gtatatgccg ctttgggaaga atatactaga 540
 acaacacatc ccgatgaacc aggaagattt gcaaaacttt tgcttctgtc gocttcttta 600
 cgtccatag gccttaagtg tttggagcat ttgtttttct ttgcottat tggagatggt 660
 ccaattgata cgttcctgat ggagatgctt gaatcacctt ctgattcata a 711

<210> 46

5 <211> 236

<212> PRT

<213> Secuencia artificial

10

<220>

<223> Dominio de unión a ligando de RXR quimérico

15 <400> 46

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

Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn
 20 25 30

Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp
 35 40 45

ES 2 392 508 T3

Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe
 50 55 60

Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp
 65 70 75 80

Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys
 85 90 95

Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala
 100 105 110

His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu
 115 120 125

Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys
 130 135 140

Leu Arg Ser Val Ile Leu Phe Asn Pro Glu Val Arg Gly Leu Lys Ser
 145 150 155 160

Ala Gln Glu Val Glu Leu Leu Arg Glu Lys Val Tyr Ala Ala Leu Glu
 165 170 175

Glu Tyr Thr Arg Thr Thr His Pro Asp Glu Pro Gly Arg Phe Ala Lys
 180 185 190

Leu Leu Leu Arg Leu Pro Ser Leu Arg Ser Ile Gly Leu Lys Cys Leu
 195 200 205

Glu His Leu Phe Phe Phe Arg Leu Ile Gly Asp Val Pro Ile Asp Thr
 210 215 220

Phe Leu Met Glu Met Leu Glu Ser Pro Ser Asp Ser
 225 230 235

<210> 47

5 <211> 441

<212> ADN

<213> *Saccharomyces cerevisiae*

10

<400> 47

ES 2 392 508 T3

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

<210> 48

5

<211> 147

<212> PRT

10 <213> Saccharomyces cerevisiae

<400> 48

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

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

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

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

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

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

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

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

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

ES 2 392 508 T3

Thr Val Ser
145

<210> 49

<211> 606

5

<212> ADN

<213> Escherichia coli

10

<400> 49

atgaaagcgt taacggccag gcaacaagag gtgtttgata tcatacgtga tcacatcagc	60
cagacaggta tgccgccgac gcggtcggaa atcgcgcagc gtttgggggt cggttcccca	120
aacgcggctg aagaacatct gaaggcgctg gcacgcaaag gcgttattga aattgtttcc	180
ggcgcacac gcgggattcg tctgttgacg gaagaggaag aagggttgcc gctggtaggt	240
cgtgtggctg ccggtgaacc acttctggcg caacagcata ttgaaggcca ttatcaggtc	300
gataccttct tattcaagcc gaatgctgat ttctgctgc gcgtcagcgg gatgtcgatg	360
aaagatatcg gcattatgga tggtagcttg ctggcagtcg ataaaactca ggatgtacgt	420
aacggtcagg tcgttgctgc acgtattgat gacgaagtta ccgtaagcg cctgaaaaaa	480
cagggcaata aagtcgaact gttgccagaa aatagcgagt ttaaaccaat tgtcgtagat	540
cttcgtcagc agagcttcac cattgaaggg ctggcgggtg gggttattcg caacggcgac	600
tggtctg	606

<210> 50

15

<211> 202

<212> PRT

20

<213> Escherichia coli

<400> 50

ES 2 392 508 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

5 <210> 51

<211> 271

<212> ADN

10

<213> Virus herpes simplex 7

<400> 51

ES 2 392 508 T3

atgggcccta aaaagaagcg taaagtcgcc cccccgaaccg atgtcagcct gggggacgag 60
 ctccacttag acggcgagga cgtggcgatg ggcgatgccg acgcgctaga cgatttcgat 120
 ctggacatgt tgggggacgg ggattccccg gggccgggat ttacccccca cgactccgcc 180
 ccctacggcg ctctggatat ggccgacttc gagtttgagc agatgtttac cgatgccott 240
 ggaattgacg agtacggtgg ggaattccccg g 271

<210> 52

5

<211> 90

<212> PRT

10 <213> Virus herpes simplex 7

<400> 52

Met Gly Pro Lys Lys Lys Arg Lys Val Ala Pro Pro Thr Asp Val Ser
 1 5 10 15

Leu Gly Asp Glu Leu His Leu Asp Gly Glu Asp Val Ala Met Ala His
 20 25 30

Ala Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Asp Gly Asp
 35 40 45

Ser Pro Gly Pro Gly Phe Thr Pro His Asp Ser Ala Pro Tyr Gly Ala
 50 55 60

Leu Asp Met Ala Asp Phe Glu Phe Glu Gln Met Phe Thr Asp Ala Leu
 65 70 75 80

Gly Ile Asp Glu Tyr Gly Gly Glu Phe Pro
 85 90

15

<210> 53

<211> 307

20

<212> ADN

<213> Saccharomyces cerevisiae

25 <400> 53

ES 2 392 508 T3

```

atgggtgctc ctccaaaaaa gaagagaaag gtagctggta tcaataaaga tatcgaggag      60
tgcaatgcca tcattgagca gtttatogac tacctgcgca cgggacagga gatgccgatg      120
gaaatggcgg atcaggcgat taacgtggtg ccgggcatga cgccgaaaac cattcttcac      180
gccggggccgc cgatccagcc tgactggctg aaatcgaatg gttttcatga aattgaagcg      240
gatgttaacg ataccagcct cttgctgagt ggagatgctt cctaccotta tgatgtgcca      300
gattatg                                          307

```

<210> 54

5 <211> 102

<212> PRT

<213> Saccharomyces cerevisiae

10

<400> 54

```

Met Gly Ala Pro Pro Lys Lys Lys Arg Lys Val Ala Gly Ile Asn Lys
 1           5           10           15

```

```

Asp Ile Glu Glu Cys Asn Ala Ile Ile Glu Gln Phe Ile Asp Tyr Leu
          20           25           30

```

```

Arg Thr Gly Gln Glu Met Pro Met Glu Met Ala Asp Gln Ala Ile Asn
      35           40           45

```

```

Val Val Pro Gly Met Thr Pro Lys Thr Ile Leu His Ala Gly Pro Pro
 50           55           60

```

```

Ile Gln Pro Asp Trp Leu Lys Ser Asn Gly Phe His Glu Ile Glu Ala
65           70           75           80

```

```

Asp Val Asn Asp Thr Ser Leu Leu Leu Ser Gly Asp Ala Ser Tyr Pro
          85           90           95

```

```

Tyr Asp Val Pro Asp Tyr
          100

```

15

<210> 55

<211> 19

20

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Elemento de respuesta de GAL4

<400> 55

ggagtactgt cctccgagc

19

5 <210> 56

<211> 36

<212> ADN

10

<213> Secuencia artificial

<220>

15 <223> Elemento de respuesta de 2xLexAop

<400> 56

ctgctgtata taaaaccagt gggtatatgt acagta

36

20

<210> 57

<211> 334

25 <212> PRT

<213> Choristoneura fumiferana

<400> 57

30

Pro Glu Cys Val Val Pro Glu Thr Gln Cys Ala Met Lys Arg Lys Glu
1 5 10 15

Lys Lys Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser Thr Thr Thr
20 25 30

Val Asp Asp His Met Pro Pro Ile Met Gln Cys Glu Pro Pro Pro Pro
35 40 45

Glu Ala Ala Arg Ile His Glu Val Val Pro Arg Phe Leu Ser Asp Lys
50 55 60

Leu Leu Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn
65 70 75 80

Gln Gln Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu
85 90 95

Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln
100 105 110

ES 2 392 508 T3

Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr
 115 120 125

Glu Met Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly
 130 135 140

Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu
 145 150 155 160

Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr
 165 170 175

Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr
 180 185 190

Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu
 195 200 205

Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His
 210 215 220

Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu
 225 230 235 240

Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr
 245 250 255

Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser
 260 265 270

Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu
 275 280 285

Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu Lys Asn Arg
 290 295 300

Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp Asp Val Ala Asp Met Ser
 305 310 315 320

His Thr Gln Pro Pro Pro Ile Leu Glu Ser Pro Thr Asn Leu
 325 330

<210> 58

5 <211> 549

<212> PRT

ES 2 392 508 T3

<213> Drosophila melanogaster

<400> 58

5

Arg Pro Glu Cys Val Val Pro Glu Asn Gln Cys Ala Met Lys Arg Arg
 1 5 10 15

Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Met Thr Thr Ser Pro Ser
 20 25 30

Ser Gln His Gly Gly Asn Gly Ser Leu Ala Ser Gly Gly Gly Gln Asp
 35 40 45

Phe Val Lys Lys Glu Ile Leu Asp Leu Met Thr Cys Glu Pro Pro Gln
 50 55 60

His Ala Thr Ile Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys Cys Gln
 65 70 75 80

Ala Arg Asn Ile Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val Ile Tyr
 85 90 95

Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu Glu Asp
 100 105 110

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

Val Ser Phe Arg His Ile Thr Glu Ile Thr Ile Leu Thr Val Gln Leu
 130 135 140

Ile Val Glu Phe Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile Pro Gln
 145 150 155 160

Glu Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met
 165 170 175

Leu Arg Met Ala Arg Arg Tyr Asp His Ser Ser Asp Ser Ile Phe Phe
 180 185 190

Ala Asn Asn Arg Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala Gly Met
 195 200 205

ES 2 392 508 T3

Ala Asp Asn Ile Glu Asp Leu Leu His Phe Cys Arg Gln Met Phe Ser
 210 215 220

Met Lys Val Asp Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile
 225 230 235 240

Phe Ser Asp Arg Pro Gly Leu Glu Lys Ala Gln Leu Val Glu Ala Ile
 245 250 255

Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn Arg His
 260 265 270

Cys Gly Asp Ser Met Ser Leu Val Phe Tyr Ala Lys Leu Leu Ser Ile
 275 280 285

Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ala Glu Met Cys Phe
 290 295 300

Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Lys Phe Leu Glu Glu Ile
 305 310 315 320

Trp Asp Val His Ala Ile Pro Pro Ser Val Gln Ser His Leu Gln Ile
 325 330 335

Thr Gln Glu Glu Asn Glu Arg Leu Glu Arg Ala Glu Arg Met Arg Ala
 340 345 350

Ser Val Gly Gly Ala Ile Thr Ala Gly Ile Asp Cys Asp Ser Ala Ser
 355 360 365

Thr Ser Ala Ala Ala Ala Ala Ala Gln His Gln Pro Gln Pro Gln Pro
 370 375 380

Gln Pro Gln Pro Ser Ser Leu Thr Gln Asn Asp Ser Gln His Gln Thr
 385 390 395 400

Gln Pro Gln Leu Gln Pro Gln Leu Pro Pro Gln Leu Gln Gly Gln Leu
 405 410 415

Gln Pro Gln Leu Gln Pro Gln Leu Gln Thr Gln Leu Gln Pro Gln Ile
 420 425 430

Gln Pro Gln Pro Gln Leu Leu Pro Val Ser Ala Pro Val Pro Ala Ser
 435 440 445

ES 2 392 508 T3

Val Thr Ala Pro Gly Ser Leu Ser Ala Val Ser Thr Ser Ser Glu Tyr
 450 455 460

Met Gly Gly Ser Ala Ala Ile Gly Pro Ile Thr Pro Ala Thr Thr Ser
 465 470 475 480

Ser Ile Thr Ala Ala Val Thr Ala Ser Ser Thr Thr Ser Ala Val Pro
 485 490 495

Met Gly Asn Gly Val Gly Val Gly Val Gly Val Gly Gly Asn Val Ser
 500 505 510

Met Tyr Ala Asn Ala Gln Thr Ala Met Ala Leu Met Gly Val Ala Leu
 515 520 525

His Ser His Gln Glu Gln Leu Ile Gly Gly Val Ala Val Lys Ser Glu
 530 535 540

His Ser Thr Thr Ala
 545

<210> 59

5

<211> 1288

<212> ADN

10 <213> Choristoneura fumiferana

<400> 59

aagggcctg cgccccgtca gcaagaggaa ctgtgtctgg tatgcgggga cagagcctcc 60
 ggataccact acaatgcgct cacgtgtgaa ggggtgtaaag ggttcttcag acggagtgtt 120
 accaaaaatg cggtttatat ttgtaaattc ggtcacgctt gcgaaatgga catgtacatg 180
 cgacggaaat gccaggagtg ccgcctgaag aagtgcttag ctgtaggcat gaggcctgag 240
 tgcgtagtac ccgagactca gtgcgccatg aagcggaaaag agaagaaagc acagaaggag 300
 aaggacaaac tgccctgtcag cacgaacgac gtggacgacc acatgccgcc cattatgcag 360
 tgtgaacctc cacctcctga agcagcaagg attcacgaag tgggcccaag gtttctctcc 420
 gacaagctgt tggagacaaa ccggcagaaa aacatcccc agttgacagc caaccagcag 480
 ttccttatcg ccaggctcat ctggtaaccag gacgggtacg agcagccttc tgatgaagat 540
 ttgaagagga ttacgcagac gtggcagcaa gcggacgatg aaaacgaaga gtctgacact 600
 cccttcgccc agatcacaga gatgactatc ctcacggtcc aacttatcgt ggagttcgcg 660

ES 2 392 508 T3

aagggattgc caggggttcgc caagatctcg cagcctgac aaattacgct gcttaaggct 720
 tgctcaagtg aggtaatgat gctccgagtc gcgcgacgat acgatgcggc ctcagacagt 780
 gttctgttcg cgaacaacca agcgtacact cgcgacaact accgcaaggc tggcatggcc 840
 tacgtcatcg aggatctact gcacttctgc cggatgcatgt actctatggc gttggacaac 900
 atccattacg cgtctctcac ggotgtctgc atcttttctg accggccagg gttggagcag 960
 ccgcaactgg tggaagaaat ccagcggtac tacctgaata cgctccgcat ctatctcctg 1020
 aaccagctga gcgggtcggc gcgttcgtcc gtcataatcg gcaagatcct ctcaatcctc 1080
 totgagctac gcacgctcgg catgcaaac tccaacatgt gcctctccct caagctcaag 1140
 aacagaaagc tgccgccttt cctcaggagg atctgggatg tggcggacat gtcgcacacc 1200
 caaccgcccgc ctatcctcga gtcccccacg aatctctagc cctcgcgcgc acgcatcgcc 1260
 gatgccgcgt ccggccgcgc tgctctga 1288

<210> 60

5

<211> 309

<212> ADN

10

<213> Virus de simio 40

<400> 60

ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg catctcaatt 60
 agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt atgcaaagca 120
 tgcctctcaa ttagtcagca accatagtcg cgcctcctaac tccgcccata ccgcccctaa 180
 ctccgcccag ttccgcccct tctccgcccc atggtgact aatctctctt atttatgcag 240
 aggccgaggc cgcctcggcc tctgagctat tccagaagta gtgaggaggc ttttttggag 300
 gcctaggct 309

15

<210> 61

<211> 24

20

<212> ADN

<213> Secuencia artificial

<220>

25

<223> promotor mínimo E1b sintético

<400> 61

30

tatataatgg atccccgggt accg

24

<210> 62

ES 2 392 508 T3

<211> 1653

<212> ADN

5

<213> Secuencia artificial

<220>

10

<223> gen de luciferasa

<400> 62

```

atggaagacg ccaaaaacat aaagaaaggc cggcgccat tetatcctct agaggatgga      60
accgctggag agcaactgca taaggctatg aagagatacg ccctgggtcc tggaaacaatt    120
gcttttacag atgcacatat cgagggtgaac atcacgtacg cggaaatactt cgaaatgtcc    180
gttcggttgg cagaagctat gaaacgatat gggctgaata caaatcacag aatcgtcgta    240
tgcagtgaaa actctcttca attctttatg cgggtggttg ggcggttatt tateggagtt    300
gcagttgccc ccgcgaacga catttataat gaacgtgaat tgcacaacag tatgaacatt    360
tcgcagccta ccgtagtggt tgtttccaaa aaggggttgc aaaaaatttt gaacgtgcaa    420
aaaaaattac caataatcca gaaaattatt atcatggatt ctaaacgga ttaccagga     480
tttcagtcga tgtacacggt cgtcacatct catctacctc cgggttttaa tgaatacgat    540
tttgtaccag agtccttga tcgtgacaaa acaattgcac tgataatgaa ttctcttgga    600
tctactgggt tacctaaggg tgtggccctt ccgcatagaa ctgcctgcgt cagattctcg    660
catgccagag atcctatttt tggcaatcaa atcattccgg aactgcatgat ttaagtgtt     720
gttccattcc atcaagggtt tggaatgtt actacactcg gatatttgat atgtggattt     780
cgagtcgtct taatgtatag atttgaagaa gagctgtttt tacgatecct tcaggattac    840
aaaattcaaa gtgcgttgct agtaccaacc ctattttcat tcttcgcaa aagcactctg    900
attgacaaat acgatttacc taatttacac gaaattgctt ctgggggccc acctctttcg    960
aaagaagtcg ggaagcgggt tgcaaacgca ttccatctc cagggatagc acaaggatat   1020
gggctcactg agactacatc agctattctg attacaccg aggggatga taaaccgggc   1080
gcggtcggta aagttgttcc attttttgaa gcgaaggttg tggatctgga taccgggaaa   1140
acgctgggcy ttaatcagag aggcgaatta tgtgtcagag gacctatgat tatgtccggt   1200
tatgtaaaca atccggaagc gaccaacgcc ttgattgaca aggatggatg gctacattct   1260
ggagacatag cttactggga cgaagacgaa cacttcttca tagttgaccg cttgaagtct   1320
ttaattaaat acaaaggata tcaggtggcc cccgctgaat tggaatcgat attgttacia   1380
caocccaaca tottogacgc gggcgtggca ggtcttcccg acgatgacgc cggatgaact   1440

```

ES 2 392 508 T3

ccccccaaca tcttgcacgc gggcgtggca ggtcttcccg acgatgacgc cggatgaactt 1440
 cccgcgcgcg ttgttgtttt ggagcacgga aagacgatga cggaaaaaga gatcgtggat 1500
 tacgtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg gaggagtgtg gtttgtggac 1560
 gaagtaccga aaggtcttac cggaaaactc gacgcaagaa aatcagaga gatcctcata 1620
 aaggccaaga agggcggaaa gtccaaattg taa 1653
 <210> 63
 <211> 18
 5 <212> ADN
 <213> Mus musculus
 10 <400> 63
 ccacatcaag ccacctag 18
 <210> 64
 15 <211> 18
 <212> ADN
 20 <213> Locusta migratoria
 <400> 64
 25 tcaccttctg attcataa 18
 <210> 65
 <211> 1054
 30 <212> ADN
 <213> Choristoneura fumiferana
 <400> 65
 35 cctgagtgcg tagtaccoga gactcagtgc gccatgaagc ggaaagagaa gaaagcacag 60
 aaggagaagg acaaactgcc tgtcagcacg acgacggcgg acgaccacat gccgcccatt 120
 atgcagtgtg aacctccacc tcctgaagca gcaaggatte acgaagtggg cccaaggttt 180
 ctctccgaca agctggtgga 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
 ttcgcgaagg gattgccagg gttcgccaag atctcgcagc ctgatcaaat tacgctgctt 480
 aaggcttgct caagtgaggt aatgatgctc cgagtcgcgc gacgatacga tgcggcctca 540

ES 2 392 508 T3

gacagtgttc tgttcgcgaa caaccaagcg tacactcgcg acaactaccg caaggctggc 600
atggcctacg tcatcgagga tctactgcac ttctgcccgt gcatgtactc tatggcgttg 660
gacaacatcc attacgcgct gctcacggct gtcgtcatct tttctgaccg gccagggttg 720
gagcagccgc aactgggtgga agaaatccag cggctactacc tgaatacgcct ccgcatctat 780
atcctgaacc agctgagcgg gtcggcgcgt tcgtccgtca tatacggcaa gatcctctca 840
atcctctctg agctacgcac gctcggcatg caaaactcca acatgtgcat ctccctcaag 900
ctcaagaaca gaaagctgcc gcctttcctc gaggagatct gggatgtggc ggacatgtcg 960
cacaccaaac cgccgcctat cctcgagtcc cccacgaatc tctagccctt gcgcgcaecg 1020
atgccgatg ccgcgtccgg ccgcgctgct ctga 1054

<210> 66

<211> 798

5

<212> ADN

<213> Choristoneura fumiferana

10

<400> 66

tcgggtgcagg taagcgatga gctgtcaatc gagcgcctaa cggagatgga gtctttggtg 60
gcagatccca gcgaggagtt ccagttcctc cgcgtggggc ctgacagcaa cgtgcctcca 120
cgttaccgcg cgcccgctct ctccctctgc caaataggca acaagcaaat agcggcgttg 180
gtggtatggg cgcgcgacat ccctcatttc gggcagctgg agctggacga tcaagtggta 240
ctcatcaagg cctcctggaa tgagetgcta ctcttcgcca tcgctggcg ctctatggag 300
tatttgggaag atgagagggg gaacggggac ggaacgcgga gcaccactca gccacaactg 360
atgtgtctca tgctggcat gacgttgcac cgcaactcgg cgcagcagge gggcgtgggc 420
gccatcttcg accgcgtgct gtccgagctc agtctgaaga tgcgcacctt gcgcatggac 480
caggccgagt acgtcgcgct caaagccatc gtgctgctca accctgatgt gaaaggactg 540
aagaatcggc aagaagttga cgttttgca gaaaaaatgt tctcttgctt ggacgactac 600
tgccggcggg cgcgaaagcaa cgaggaaggc cggtttgctt ccttgcctgc gcggetgcca 660
gctctccgct ccatctcgct caagagcttc gaacacctct acttcttcca cctcgtggcc 720
gaaggctcca tcagcggata catacgagag gcgctccgaa accacgcgcc tccgatcgac 780
gtcaatgcca tgatgtaa 798

15 <210> 67

<211> 1650

<212> ADN

ES 2 392 508 T3

<213> Drosophila melanogaster

<400> 67

cggccggaat gcgtcgtccc ggagaaccaa tgtgcatga agcggcgcga aaagaaggcc	60
cagaaggaga aggacaaaat gaccacttcg ccgagctctc agcatggcgg caatggcagc	120
ttggcctctg gtggcggcca agactttgtt aagaaggaga ttcttgacct tatgacatgc	180
gagccgcccc agcatgccac tattcogcta ctacctgatg aaatattggc caagtgtcaa	240
gcgcgcaata taccttcctt aacgtacaat cagttggcgg ttatatacaa gttaatttgg	300
taccaggatg gctatgagca gccatctgaa gaggatctca ggcgtataat gagtcaacce	360
gatgagaacg agagccaaac ggacgtcagc tttcggcata taaccgagat aaccatactc	420
acggtccagt tgattgttga gtttgctaaa ggtctaccag cgtttacaaa gataccccag	480
gaggaccaga tcacgttact aaaggcctgc tcgtcggagg tgatgatgct gcgtatggca	540
cgacgctatg accacagctc ggactcaata ttcttcgcga ataatagatc atatacgcgg	600
gattcttaca aaatggcggg aatggctgat aacattgaag acctgctgca tttctgccc	660
caaatgttct cgatgaaggt ggacaacgtc gaatacgcgc ttctcactgc cattgtgatc	720
ttctcggacc ggccgggcct ggagaaggcc caactagtgc aagcgatcca gagctactac	780
atcgacacgc taogcattta tatactcaac cgccactgcg gcgactcaat gagcctcgtc	840
ttctacgcaa agctgctctc gatcctcacc gagctgcgta cgetgggcaa ccagaacgcc	900
gagatgtggt tctcactaaa gctcaaaaac cgcaaactgc ccaagttcct cgaggagatc	960
tgggacgttc atgccaatccc gccatcggtc cagtcgcacc ttcagattac ccaggaggag	1020
aacgagcgtc tcgagcgggc tgagcgtatg cgggcatcgg ttgggggccc cattaccgcc	1080
ggcattgatt gcgactctgc ctccacttcg gcggcggcag ccgcggccca gcatcagcct	1140
cagcctcagc cccagcccca acctcctcc ctgaccaga acgattccca gcaccagaca	1200
cagccgcagc tacaacctca gctaccact cagctgcaag gtcaactgca accccagctc	1260
caaccacagc ttcagacgca actccagcca cagattcaac cacagccaca gctcctccc	1320
gtctccgctc ccgtgccgc ctccgtaacc gcaactgggt ccttgteccg ggtcagtagc	1380
agcagogaat acatggcggg aagtgcggcc ataggaccca tcacgccggc aaccaccagc	1440
agtatcacgg ctgcggttac cgctagctcc accacatcag cggatccgat gggcaacgga	1500
gttgagtcg gtggtgggtt gggcggcaac gtcagcatgt atgcgaacgc ccagacggcg	1560
atggccttga tgggtgtagc cctgcattcg caccaagagc agcttatcgg gggagtggcg	1620
gttaagtcgg agcactcgac gactgcatag	1650

<210> 68

ES 2 392 508 T3

<211> 1586

<212> ADN

5

<213> *Bamecia argentifoli*

<400> 68

```

gaattcgcgg ccgctcgc aa acttccgtac ctctcacc cc ctcgccagga ccccccgcca      60
accagttcac cgtcactccc tccaatggat actcatcccc catgtcttcg ggcagctacg      120
acccttatag tcccaccaat ggaagaatag ggaaagaaga gctttcgcgg gcgaatagtc      180
tgaacgggta caacgtggat agctgcgatg cgtcgcggaa gaagaagga ggaacgggtc      240
ggcagcagga ggagctgtgt ctctctctgc gggaccgcgc ctccggctac cactacaacg      300
ccctcacctg cgaaggctgc aagggtctct tccgtcggag catcaccaag aatgccgtct      360
accagtgtaa atatggaaat aattgtgaaa ttgacatgta catgaggcga aaatgccaaag      420
agtgtcgtct caagaagtgt ctcagcgttg gcatgaggcc agaatgtgta gttcccgaa      480
tocagtgtgc tgtgaagcga aaagagaaaa aagcgcaaaa ggacaaagat aaacctaac      540
caacgacgag ttgttctcca gatggaatca aacaagagat agatcctcaa aggctggata      600
cagattcgca gctattgtct gtaaatggag ttaaaccat tactccagag caagaagagc      660
tcatccatag gctagtttat tttcaaaatg aatatgaaca tccatcccca gaggatatca      720
aaaggatagt taatgctgca ccagaagaag aaaatgtagc tgaagaaagg tttaggcata      780
ttacagaaat tacaattctc actgtacagt taattgtgga attttctaag cgattacctg      840
gttttgacaa actaattcgt gaagatcaaa tagctttatt aaaggcatgt agtagtgaag      900
taatgatggt tagaatggca aggaggtatg atgctgaaac agattcgata ttgtttgcaa      960
ctaaccagcc gtatacgaga gaatcataca ctgtagctgg catgggtgat actgtggagg     1020
atctgctccg attttgtcga catatgtgtg ccatgaaagt cgataacgca gaatatgctc     1080
ttctcactgc cattgtaatt ttttcagaac gaccatctct aagtgaaggc tggaagggtg     1140
agaagattca agaaatttac atagaagcat taaaagcata tgttgaaaat cgaaggaaac     1200
catatgcaac aaccattttt gctaagttac tatctgtttt aactgaacta cgaacattag     1260
ggaatatgaa ttcagaaaca tgcttctcat tgaagctgaa gaatagaaag gtgccatcct     1320
tcctcgagga gatttgggat gttgtttcat aaacagtctt acctcaatto catgttactt     1380
ttcatatttg atttatctca gcagggtgct cagtacttat cctcacatta ctgagctcac     1440
ggtatgctca tacaattata acttgtaata tcatatcggg gatgacaaat ttgttacaat     1500
attctttggt accttaacac aatgttgatc tcataatgat gtatgaattt ttctgttttt     1560
gcaaaaaaaaa aagcggccgc gaattc                                         1586

```


ES 2 392 508 T3

<210> 69

<211> 1109

5

<212> ADN

<213> Nephrotetix cincticeps

10

<400> 69

```

caggaggagc tctgcctggt gtgcgagagc cgagcgtcgg gataccacta caacgctctc      60
acctgcgaag gatgcaaggg ottctttcgg aggagtatca ccaaaaaacgc agtgtaccag     120
tccaaatacg gcaccaattg tgaaatagac atgtatatgc ggcgcaagtg ccaggagtgc     180
cgactcaaga agtgcctcag tgtagggatg aggccagaat gtgtagtacc tgagtatcaa     240
tgtgcccgtaa aaaggaaaga gaaaaaagct caaaaggaca aagataaacc tgtctcttca     300
accaatggct cgctgaaat gagaatagac caggacaacc gttgtgtggt gttgcagagt     360
gaagacaaca ggtacaactc gagtacgccc agtttcggag tcaaaccctc cagtccagaa     420
caagaggagc tcatccacag gctcgtctac ttccagaacg agtacgaaca ccctgcogag     480
gaggatctca agcggatcga gaacctcccc tgtgacgacg atgaccogtg tgatgttcgc     540
tacaaacaca ttacggagat cacaatactc acagtccagc tcatcgtgga gtttgcgaaa     600
aaactgcctg gtttcgacaa actactgaga gaggaccaga tcgtgttgct caaggcgtgt     660
tcgagcagag tgatgatgct gcggatggcg cggaggtagc acgtccagac agactcgatc     720
ctgttcgcca acaaccagcc gtacacgcga gagtcgtaca cgatggcagg cgtgggggaa     780
gtcatcgaag atctgctgcg gttcggccga ctcatgtgct ccatgaaggc ggacaatgcc     840
gagtatgctc tgctcagcgc catcgtcatc ttctccgagc ggccgaacct ggcggaagga     900
tggaaggttg agaagatcca ggagatctac ctggaggcgc tcaagtccca cgtggacaac     960

cgagtgaaac ctcgcagtcc gaccatcttc gccaaaactgc tctccgttct caccgagetg    1020
cgaacactcg gcaaccagaa ctccgagatg tgcttctcgt taaactacgc aaccgcaaac    1080
atgccaccgt tcctcgaaga aatctggga                                     1109

```

15

<210> 70

<211> 401

20

<212> PRT

<213> Choristoneura fumiferana

<400> 70

ES 2 392 508 T3

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

ES 2 392 508 T3

Gln Ile Thr Glu Met Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe
 195 200 205

Ala Lys Gly Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile
 210 215 220

Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala
 225 230 235 240

Arg Arg Tyr Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln
 245 250 255

Ala Tyr Thr Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile
 260 265 270

Glu Asp Leu Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp
 275 280 285

Asn Ile His Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg
 290 295 300

Pro Gly Leu Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr
 305 310 315 320

Leu Asn Thr Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala
 325 330 335

Arg Ser Ser Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu
 340 345 350

Arg Thr Leu Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu
 355 360 365

Lys Asn Arg Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp Asp Val Ala
 370 375 380

Asp Met Ser His Thr Gln Pro Pro Pro Ile Leu Glu Ser Pro Thr Asn
 385 390 395 400

Leu

<210> 71

5 <211> 894

ES 2 392 508 T3

<212> ADN

<213> Tenebrio molitor

5 <400> 71

```

aggccggaat gtgtggtacc ggaagtacag tgtgctgtta agagaaaaga gaagaaagcc      60
caaaaggaaa aagataaacc aaacagcact actaacggct caccagacgt catcaaaatt      120
gaaccagaat tgtcagattc agaaaaaaca ttgactaacg gacgcaatag gatatcacca      180
gagcaagagg agtcatact catacatcga ttggtttatt tccaaaacga atatgaacat      240
ccgtctgaag aagacgttaa acggattatc aatcagccga tagatgggtga agatcagtgt      300
gagatacggc ttaggcatac cacggaaatt acgatcctga ctgtgcagct gatcgtggag      360
tttgccaagc ggttaccagg cttcgataag ctctgcagg aagatcaaat tgctctcttg      420
aaggcatggt caagcgaagt gatgatgttc aggatggccc gacgttacga cgtccagtcg      480
gattccatcc tcttcgtaaa caaccagcct tatccgaggg acagttacaa tttggccggc      540
atgggggaaa ccatcgaaga tctcttgcac ttttgcagaa ctatgtactc catgaaggcg      600
gataatgccg aatatgcttt actaacagcc atcgttattt tctcagagcg accgctcgttg      660
atagaaggct ggaagggtgga gaagatccaa gaaatctatt tagaggcatt gcgggcgtac      720
gtcgacaacc gaagaagccc aagccggggc acaatattog cgaaactcct gtcagtacta      780
actgaattgc ggacgttagg caacccaaat tcagagatgt gcattctcgtt gaaattgaaa      840
aacaaaaagt taccgccggt cctggacgaa atctgggacg tcgacttaaa agca          894
    
```

<210> 72

10

<211> 298

<212> PRT

15 <213> Tenebrio molitor

<400> 72

```

Arg Pro Glu Cys Val Val Pro Glu Val Gln Cys Ala Val Lys Arg Lys
 1              5              10              15

Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Pro Asn Ser Thr Thr Asn
    20              25              30

Gly Ser Pro Asp Val Ile Lys Ile Glu Pro Glu Leu Ser Asp Ser Glu
    35              40              45
    
```

ES 2 392 508 T3

Lys Thr Leu Thr Asn Gly Arg Asn Arg Ile Ser Pro Glu Gln Glu Glu
 50 55 60

Leu Ile Leu Ile His Arg Leu Val Tyr Phe Gln Asn Glu Tyr Glu His
 65 70 75 80

Pro Ser Glu Glu Asp Val Lys Arg Ile Ile Asn Gln Pro Ile Asp Gly
 85 90 95

Glu Asp Gln Cys Glu Ile Arg Phe Arg His Thr Thr Glu Ile Thr Ile
 100 105 110

Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Arg Leu Pro Gly Phe
 115 120 125

Asp Lys Leu Leu Gln Glu Asp Gln Ile Ala Leu Leu Lys Ala Cys Ser
 130 135 140

Ser Glu Val Met Met Phe Arg Met Ala Arg Arg Tyr Asp Val Gln Ser
 145 150 155 160

Asp Ser Ile Leu Phe Val Asn Asn Gln Pro Tyr Pro Arg Asp Ser Tyr
 165 170 175

Asn Leu Ala Gly Met Gly Glu Thr Ile Glu Asp Leu Leu His Phe Cys
 180 185 190

Arg Thr Met Tyr Ser Met Lys Val Asp Asn Ala Glu Tyr Ala Leu Leu
 195 200 205

Thr Ala Ile Val Ile Phe Ser Glu Arg Pro Ser Leu Ile Glu Gly Trp
 210 215 220

Lys Val Glu Lys Ile Gln Glu Ile Tyr Leu Glu Ala Leu Arg Ala Tyr
 225 230 235 240

Val Asp Asn Arg Arg Ser Pro Ser Arg Gly Thr Ile Phe Ala Lys Leu
 245 250 255

Leu Ser Val Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ser Glu
 260 265 270

Met Cys Ile Ser Leu Lys Leu Lys Asn Lys Lys Leu Pro Pro Phe Leu
 275 280 285

Asp Glu Ile Trp Asp Val Asp Leu Lys Ala
 290 295

ES 2 392 508 T3

<210> 73

<211> 948

5 <212> ADN

<213> Amblyomma americanum

<400> 73

10

cggccggaat	gtgtggtgcc	ggagtaccag	tgtgccatca	agcgggagtc	taagaagcac	60
cagaaggacc	ggccaaacag	cacaacgcgg	gaaagtccct	cggcgctgat	ggcgccatct	120
tctgtgggtg	gcgtagagccc	caccagccag	cccatgggtg	gcgagggcag	ctccctgggc	180
agcagcaatc	acgaggagga	taagaagcca	gtggtgctca	gcccaggagt	caagcccctc	240
tcttcatctc	aggaggacct	catcaacaag	ctagtctact	accagcagga	gtttgagtgc	300
ccttctgagg	aagacatgaa	gaaaaccacg	cccttcccc	tgggagacag	tgaggaagac	360
aaccagcggc	gattccagca	cattactgag	atcaccatcc	tgacagtgca	gctcattgtg	420
gagttctcca	agcgggtccc	tggetttgac	acgctggcac	gagaagacca	gattactttg	480
ctgaaggcct	gctccagtga	agtgatgatg	ctgagaggtg	cccggaaata	tgatgtgaag	540
acagattcta	tagtgtttgc	caataaccag	ccgtacacga	gggacaacta	ccgcagtgcc	600
agtgtggggg	actctgcaga	tgccctgttc	cgcttctgcc	gcaagatgtg	tcagctgaga	660
gtagacaacg	ctgaatacgc	actcctgaoc	gccattgtaa	ttttctctga	acggccatca	720
ctgggtggacc	cgcacaaggt	ggagcgcate	caggagtact	acattgagac	cctgcgcatg	780
tactccgaga	accaccggcc	cccaggcaag	aactactttg	cccggctgct	gtccatcttg	840
acagagctgc	gcaccttggg	caacatgaac	gccgaaatgt	gcttctcgct	caaggtgcag	900
aacaagaagc	tgccaccggt	cctggctgag	atttgggaca	tccaagag		948

<210> 74

15 <211> 316

<212> PRT

<213> Amblyomma americanum

20

<400> 74

ES 2 392 508 T3

Arg Pro Glu Cys Val Val Pro Glu Tyr Gln Cys Ala Ile Lys Arg Glu
 1 5 10 15
 Ser Lys Lys His Gln Lys Asp Arg Pro Asn Ser Thr Thr Arg Glu Ser
 20 25 30
 Pro Ser Ala Leu Met Ala Pro Ser Ser Val Gly Gly Val Ser Pro Thr
 35 40 45
 Ser Gln Pro Met Gly Gly Gly Gly Ser Ser Leu Gly Ser Ser Asn His
 50 55 60
 Glu Glu Asp Lys Lys Pro Val Val Leu Ser Pro Gly Val Lys Pro Leu
 65 70 75 80
 Ser Ser Ser Gln Glu Asp Leu Ile Asn Lys Leu Val Tyr Tyr Gln Gln
 85 90 95
 Glu Phe Glu Ser Pro Ser Glu Glu Asp Met Lys Lys Thr Thr Pro Phe
 100 105 110
 Pro Leu Gly Asp Ser Glu Glu Asp Asn Gln Arg Arg Phe Gln His Ile
 115 120 125
 Thr Glu Ile Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ser Lys
 130 135 140
 Arg Val Pro Gly Phe Asp Thr Leu Ala Arg Glu Asp Gln Ile Thr Leu
 145 150 155 160
 Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Gly Ala Arg Lys
 165 170 175
 Tyr Asp Val Lys Thr Asp Ser Ile Val Phe Ala Asn Asn Gln Pro Tyr
 180 185 190
 Thr Arg Asp Asn Tyr Arg Ser Ala Ser Val Gly Asp Ser Ala Asp Ala
 195 200 205
 Leu Phe Arg Phe Cys Arg Lys Met Cys Gln Leu Arg Val Asp Asn Ala
 210 215 220
 Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile Phe Ser Glu Arg Pro Ser
 225 230 235 240

ES 2 392 508 T3

Leu Val Asp Pro His Lys Val Glu Arg Ile Gln Glu Tyr Tyr Ile Glu
 245 250 255

Thr Leu Arg Met Tyr Ser Glu Asn His Arg Pro Pro Gly Lys Asn Tyr
 260 265 270

Phe Ala Arg Leu Leu Ser Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn
 275 280 285

Met Asn Ala Glu Met Cys Phe Ser Leu Lys Val Gln Asn Lys Lys Leu
 290 295 300

Pro Pro Phe Leu Ala Glu Ile Trp Asp Ile Gln Glu
 305 310 315

<210> 75

5 <211> 825

<212> ADN

<213> Drosophila melanogaster

10

<400> 75

```

gtgtccaggg atttctcgat cgagcgcac atagaggccg agcagcgagc ggagacccaa      60
tgcggegac gtgcaactgac gttcctgcgc gttggctcct attccacagt ccagccggac      120
tacaaggggtg ccgtgtcggc cctgtgccaa gtggtaaca aacagctctt ccagatggtc      180
gaatacgcgc gcatgatgcc gcactttgcc caggtgccgc tggacgacca ggtgattctg      240
ctgaaagccg cttggatcga gctgctcatt gcgaacgtgg cctgggtgcag catcgttctg      300
ctggatgacg gcggtgccgg cggcgggggc ggtggactag gccacgatgg ctcccttgag      360
cgacgatcac cgggccttca gcccagcag ctgttctca accagagctt ctcgtaecat      420
cgcaacagtg cgatcaaagc cgggtgtgca gccatcttcg accgcatatt gtcggagctg      480
agtgtaaaga tgaagcggct gaatctcgac cgacgcgagc tgtcctgctt gaaggccatc      540
atactgtaca acccggacat acgogggatc aagagccggg cggagatcga gatgtgccgc      600
gagaaggtgt acgcttgctt ggacgagcac tgccgcctgg aacatccggg cgacgatgga      660
cgctttgcgc aactgctgct gcgtctgccc gctttgcgat cgatcagcct gaagtgccag      720
gatcacctgt tctcttccg cattaccagc gaccggccgc tggaggagct ctttctcgag      780
cagctggagg cgccgccgcc acccggcctg gcgatgaaac tggag                          825
    
```


REIVINDICACIONES

1. Un sistema de modulación de la expresión génica que comprende:

5 a) un primer casete de expresión génica que se puede expresar en una célula huésped que comprende una secuencia polinucleotídica que codifica un primer polipéptido híbrido que comprende:

i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y

10 ii) un dominio de unión a ligando de receptor de ecdisona; y

b) un segundo casete de expresión génica que se puede expresar en la célula huésped que comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende:

15 i) un dominio de transactivación; y

ii) un dominio de unión a ligando de receptor X retinoide quimérico que comprende, o bien (A) las hélices 1-7 de un receptor X retinoide de vertebrado y las hélices 8-12 de un RXR de invertebrado, o bien (B) las hélices 1-8 de un receptor X retinoide de vertebrado y las hélices 9-12 de un RXR de invertebrado, en el que el invertebrado es una especie que no es díptero ni es lepidóptero.

2. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, que comprende además un casete de expresión que comprende:

25 i) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido híbrido;

ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y

30 iii) un gen cuya expresión se quiere modular.

3. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el dominio de unión a ligando (LBD) del receptor de ecdisona del polipéptido híbrido es un LBD de EcR de gusano de las yemas de la píceca *Choristoneura fumiferana* ("CfEcR") o un LBD de EcR de la mosca de la fruta *Drosophila melanogaster* ("DmEcR").

4. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el dominio de unión a ligando del receptor de ecdisona del primer polipéptido híbrido está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF) y SEQ ID NO: 67 (DmEcR-DEF).

5. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el dominio de unión a ligando del receptor de ecdisona del primer polipéptido híbrido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 58 (DmEcR-DEF) y SEQ ID NO: 70 (CfEcR-CDEF).

6. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el dominio de unión a ligando del receptor X retinoide quimérico del segundo polipéptido híbrido está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en

50 a) Los nucleótidos 1-408 de la SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y

b) los nucleótidos 1-465 de la SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.

7. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el dominio de unión a ligando del receptor X retinoide quimérico del segundo polipéptido híbrido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en

60 a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

8. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el primer casete de expresión génica comprende una secuencia de polinucleótidos que codifica el primer polipéptido híbrido que comprende un dominio de unión a ADN seleccionado del grupo que consiste en un dominio de unión a ADN GAL4 y un dominio de unión a ADN LexA, y un dominio de unión a ligando del receptor de ecdisona.

65

- 5 9. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el segundo casete de expresión génica comprende un polinucleótido que codifica el segundo polipéptido híbrido que comprende un dominio de transactivación seleccionado del grupo que consiste en un dominio de transactivación VP16 y un dominio de transactivación activador ácido B42, y un dominio de unión a ligando del receptor X retinoide quimérico.
- 10 10. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el segundo casete de expresión génica comprende un polinucleótido que codifica el segundo polipéptido híbrido que comprende un dominio de transactivación codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un AD de VP16 (SEQ ID NO: 51) y un AD de B42 (SEQ ID NO: 53) y un dominio de unión a ligando del receptor X retinoide quimérico codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en
- 15 a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y
b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.
- 20 11. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, en el que el segundo casete de expresión génica comprende un polinucleótido que codifica el segundo polipéptido híbrido que comprende un dominio de transactivación que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un AD de VP16 (SEQ ID NO: 52) y un AD de B42 (SEQ ID NO: 54) y un dominio de unión a ligando del receptor X retinoide quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en
- 25 a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y
b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.
- 30 12. Un sistema de modulación de la expresión génica que comprende:
- a) un primer casete de expresión génica que se puede expresar en una célula huésped que comprende una secuencia polinucleotídica que codifica un primer polipéptido híbrido que comprende:
- 35 i) un dominio de unión a ADN que reconoce un elemento de respuesta asociado con un gen cuya expresión se quiere modular; y
ii) un dominio de unión a ligando del receptor X retinoide quimérico que comprende (A) las hélices 1-7 de un receptor X retinoide de vertebrado y las hélices 8-12 de un RXR de invertebrado, o bien (B) las hélices 1-8 de un receptor X retinoide de vertebrado y las hélices 9-12 de un RXR de invertebrado,
- 40 en el que el invertebrado es una especie que no es díptero ni es lepidóptero; y
- b) un segundo casete de expresión génica que se puede expresar en la célula huésped que comprende una secuencia polinucleotídica que codifica un segundo polipéptido híbrido que comprende:
- 45 i) un dominio de transactivación; y
ii) un dominio de unión a ligando de receptor de ecdisona.
- 50 13. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, que comprende además un casete de expresión que comprende:
- i) un elemento de respuesta que reconoce el dominio de unión a ADN del primer polipéptido híbrido;
- 55 ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y
iii) un gen cuya expresión se quiere modular.
- 60 14. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el dominio de unión a ligando del receptor X retinoide quimérico del primer polipéptido híbrido está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en
- a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y
- 65 b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.
15. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el dominio de

unión a ligando del receptor X retinoide quimérico del primer polipéptido híbrido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en

5 a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

16. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el dominio de unión a ligando del receptor de ecdisona del segundo polipéptido híbrido está codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en SEQ ID NO: 65 (CfEcR-DEF), SEQ ID NO: 59 (CfEcR-CDEF) y SEQ ID NO: 67 (DmEcR-DEF).

17. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el dominio de unión a ligando del receptor de ecdisona del segundo polipéptido híbrido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 57 (CfEcR-DEF), SEQ ID NO: 58 (DmEcR-DEF) y SEQ ID NO: 70 (CfEcR-CDEF).

18. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el primer casete de expresión génica comprende un polinucleótido que codifica el primer polipéptido híbrido que comprende un dominio de unión a ADN seleccionado del grupo que consiste en un dominio de unión a ADN GAL4 y un dominio de unión a ADN LexA, y un dominio de unión a ligando del receptor X retinoide quimérico.

19. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el primer casete de expresión génica comprende un polinucleótido que codifica el primer polipéptido híbrido que comprende un dominio de unión a ADN codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 47) y un DBD de LexA (SEQ ID NO: 49) y un dominio de unión a ligando del receptor X retinoide quimérico codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en

30 a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y

b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.

20. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el primer casete de expresión génica comprende un polinucleótido que codifica el primer polipéptido híbrido que comprende un dominio de unión a ADN que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 48) y un DBD de LexA (SEQ ID NO: 50) y un dominio de unión a ligando del receptor X retinoide quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en

40 a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

21. El sistema de modulación de la expresión génica de acuerdo con la reivindicación 12, en el que el segundo casete de expresión génica comprende un polinucleótido que codifica el segundo polipéptido híbrido que comprende un dominio de transactivación seleccionado del grupo que consiste en un dominio de transactivación VP16 y un dominio de transactivación activador ácido B42, y un dominio de unión a ligando del receptor de ecdisona.

22. Un casete de expresión génica que comprende un polinucleótido que codifica un polipéptido híbrido que comprende

50 a) un dominio de unión a ADN y

55 b) un dominio de unión a ligando de receptor X retinoide quimérico que comprende, o bien i) las hélices 1-7 de un receptor X retinoide de vertebrado y las hélices 8-12 de un RXR de invertebrado, o bien ii) las hélices 1-8 de un receptor X retinoide de vertebrado y las hélices 9-12 de un RXR de invertebrado, en el que el invertebrado es una especie que no es díptero ni es lepidóptero.

23. El casete de expresión génica de acuerdo con la reivindicación 22, en el que el dominio de unión a ADN es un dominio de unión a ADN GAL4 o un dominio de unión a ADN LexA.

24. El casete de expresión génica de acuerdo con la reivindicación 22, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende un dominio de unión a ADN codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 47) y un DBD de LexA (SEQ ID NO: 49) y un dominio de unión a ligando del receptor X retinoide quimérico codificado por un polinucleótido que comprende una secuencia de ácidos

nucleicos seleccionada del grupo que consiste en

a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y

5 b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.

25. El casete de expresión génica de acuerdo con la reivindicación 22, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende un dominio de unión a ADN que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un DBD de GAL4 (SEQ ID NO: 10 48) y un DBD de LexA (SEQ ID NO: 50) y un dominio de unión a ligando del receptor X retinoide quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en

a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

15 b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

26. Un casete de expresión génica que comprende

a) un polinucleótido que codifica un polipéptido híbrido que comprende un dominio de transactivación y

20 b) un dominio de unión a ligando de receptor X retinoide quimérico que comprende, o bien i) las hélices 1-7 de un receptor X retinoide de vertebrado y las hélices 8-12 de un RXR de invertebrado, o bien ii) las hélices 1-8 de un receptor X retinoide de vertebrado y las hélices 9-12 de un RXR de invertebrado, en el que el invertebrado es una especie que no es díptero ni es lepidóptero.

25 27. El casete de expresión génica de acuerdo con la reivindicación 26, en el que el dominio de transactivación es un dominio de transactivación VP16 o un dominio de transactivación activador ácido B42.

28. El casete de expresión génica de acuerdo con la reivindicación 26, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende un dominio de transactivación 30 codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un AD de VP16 (SEQ ID NO: 51) y un AD de B42 (SEQ ID NO: 53) y un dominio de unión a ligando del receptor X retinoide quimérico codificado por un polinucleótido que comprende una secuencia de ácidos nucleicos 35 seleccionada del grupo que consiste en

a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y

b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.

40 29. El casete de expresión génica de acuerdo con la reivindicación 28, en el que el casete de expresión génica comprende un polinucleótido que codifica un polipéptido híbrido que comprende un dominio de transactivación que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en un AD de VP16 (SEQ ID NO: 52) y un AD de B42 (SEQ ID NO: 54) y un dominio de unión a ligando del receptor X retinoide quimérico que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en

45 a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

50 30. Un polinucleótido aislado que codifica un dominio de unión a ligando del receptor X retinoide quimérico, en el que el polinucleótido comprende una secuencia de ácidos nucleicos seleccionada del grupo que consiste en

a) los nucleótidos 1-408 de SEQ ID NO: 13 y los nucleótidos 337-630 de la SEQ ID NO: 21, y

55 b) los nucleótidos 1-465 de SEQ ID NO: 13 y los nucleótidos 403-630 de la SEQ ID NO: 21.

31. Un polipéptido aislado codificado por el polinucleótido aislado de acuerdo con la reivindicación 30.

32. Un polipéptido del receptor X retinoide quimérico aislado que comprende una secuencia de aminoácidos 60 seleccionada del grupo que consiste en

a) los aminoácidos 1-136 de la SEQ ID NO: 13 y los aminoácidos 113-210 de la SEQ ID NO: 21, y

b) los aminoácidos 1-155 de la SEQ ID NO: 13 y los aminoácidos 135-210 de la SEQ ID NO: 21.

65 33. Un método *in vitro* de modulación de la expresión de un gen en una célula huésped que comprende el gen que

se quiere modular que comprende las etapas de:

a) introducir en la célula huésped el sistema de modulación de la expresión génica de acuerdo con la reivindicación 1; y

b) introducir en la célula huésped un ligando; en el que el gen que se quiere modular es un componente de un casete de expresión génica que comprende:

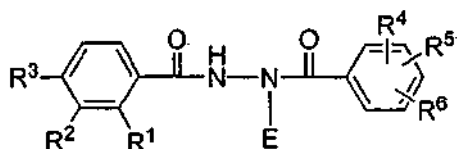
i) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido híbrido;

ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y

iii) un gen cuya expresión se quiere modular;

de modo que después de la introducción del ligando en la célula huésped, se modula la expresión del gen de b)iii).

34. El método de acuerdo con la reivindicación 33, en el que el ligando es un compuesto de la fórmula:



en la que:

E es un alquilo (C₄-C₆) que contiene un carbono terciario o un cianoalquilo (C₃-C₅) que contiene un carbono terciario;

R¹ es H, Me, Et, i-Pr, F, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OH, OMe, OEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, SCN o SCHF₂;

R² es H, Me, Et, n-Pr, i-Pr, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe₂, NEt₂, SMe, SEt, SOCF₃, OCF₂CF₂H, COEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, OCF₃, OCHF₂, O-i-Pr, SCN, SCHF₂, SOMe, NH-CN, o se une con R³ y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R³ es H, Et, o se une con R² y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R⁴, R⁵ y R⁶ son independientemente H, Me, Et, F, Cl, Br, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂O, CN, C≡CH, CH, 1-propinilo, 2-propinilo, vinilo, OMe, OEt, SMe o SEt.

35. El método de acuerdo con la reivindicación 33, que comprende además introducir en la célula huésped un segundo ligando, en el que el segundo ligando es ácido 9-cis-retinoico o un análogo sintético de un ácido retinoico.

36. Un método *in vitro* de modulación de la expresión de un gen en una célula huésped que comprende el gen que se quiere modular que comprende las etapas de:

a) introducir en la célula huésped el sistema de modulación de la expresión génica de la reivindicación 12; y

b) introducir en la célula huésped un ligando; en el que el gen que se quiere modular es un componente de un casete de expresión génica que comprende:

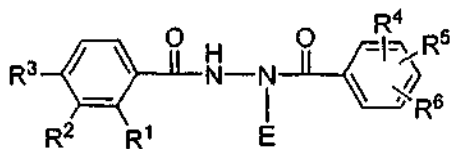
i) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido híbrido;

ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y

iii) un gen cuya expresión se quiere modular;

de modo que después de la introducción del ligando en la célula huésped, se modula la expresión del gen de b)iii).

37. El método de acuerdo con la reivindicación 36, en el que el ligando es un compuesto de la fórmula:



en la que:

- 5 E es un alquilo (C₄-C₆) que contiene un carbono terciario o un cianoalquilo (C₃-C₅) que contiene un carbono terciario;
 R¹ es H, Me, Et, i-Pr, F, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OH, OMe, OEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, SCN o SCHF₂;
- 10 R² es H, Me, Et, n-Pr, i-Pr, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe₂, NEt₂, SMe, SEt, SOCF₃, OCF₂CF₂H, COEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, OCF₃, OCHF₂, O-i-Pr, SCN, SCHF₂, SOMe, NH-CN, o se une con R³ y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxi, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;
- 15 R³ es H, Et, o se une con R² y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxi, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;
- 20 R⁴, R⁵ y R⁶ son independientemente H, Me, Et, F, Cl, Br, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OMe, OEt, SMe o SEt.
38. El método de acuerdo con la reivindicación 36, que comprende además introducir en la célula huésped un segundo ligando, en el que el segundo ligando es ácido 9-cis-retinoico o un análogo sintético de un ácido retinoico.
- 25 39. Una célula huésped aislada que comprende el sistema de modulación de la expresión génica de acuerdo con la reivindicación 1.
40. La célula huésped aislada de acuerdo con la reivindicación 39, en la que la célula huésped se selecciona del grupo que consiste en una célula bacteriana, una célula fúngica, una célula de levadura, una célula animal y una célula de mamífero.
- 30 41. La célula huésped aislada de acuerdo con la reivindicación 40, en la que la célula de mamífero es una célula murina o una célula humana.
- 35 42. Una célula huésped aislada que comprende el sistema de modulación de la expresión génica de acuerdo con la reivindicación 12.
- 40 43. La célula huésped aislada de acuerdo con la reivindicación 42, en la que la célula huésped se selecciona del grupo que consiste en una célula bacteriana, una célula fúngica, una célula de levadura, una célula animal y una célula de mamífero.
- 45 44. La célula huésped aislada de acuerdo con la reivindicación 43, en la que la célula de mamífero es una célula murina o una célula humana.
- 45 45. Un organismo no humano que comprende la célula huésped de la reivindicación 39.
46. El organismo no humano de acuerdo con la reivindicación 45, en el que el organismo no humano se selecciona del grupo que consiste en una bacteria, un hongo, una levadura, un animal y un mamífero.
- 50 47. El organismo no humano de acuerdo con la reivindicación 46, en el que el mamífero se selecciona del grupo que consiste en un ratón, una rata, un conejo, un gato, un perro, un bóvido, una cabra, un cerdo, un caballo, una oveja, un mono y un chimpancé.
- 55 48. Un organismo no humano que comprende la célula huésped de la reivindicación 42.
49. El organismo no humano de acuerdo con la reivindicación 48, en el que el organismo no humano se selecciona del grupo que consiste en una bacteria, un hongo, una levadura, un animal y un mamífero.
- 60 50. El organismo no humano de acuerdo con la reivindicación 49, en el que el mamífero se selecciona del grupo que consiste en un ratón, una rata, un conejo, un gato, un perro, un bóvido, una cabra, un cerdo, un caballo, una oveja,

un mono y un chimpancé.

51. El sistema de modulación de la expresión génica de la reivindicación 1 ó 12, en el que el sistema de modulación de la expresión génica presenta un incremento en la sensibilidad para un ligando no esteroide con respecto a un sistema de modulación de la expresión génica que contiene un dominio de unión a ligando del receptor X retinoide quimérico.

52. El casete de expresión génica de la reivindicación 22 ó 26, en el que el polipéptido presenta un incremento en la sensibilidad para un ligando no esteroide con respecto a un polipéptido que contiene un dominio de unión a ligando del receptor X retinoide quimérico.

53. El sistema de modulación de la expresión génica de una cualquiera de las reivindicaciones 1, 12 ó 51 en el que el sistema presenta un incremento en la magnitud de inducción génica en comparación con un sistema de modulación de la expresión génica que contiene un dominio de unión a ligando del receptor X retinoide quimérico.

54. Los casetes de expresión génica de una cualquiera de las reivindicaciones 22, 26 ó 52 en el que el polipéptido presenta un incremento en la magnitud de inducción génica en comparación con un polipéptido que contiene un dominio de unión a ligando del receptor X retinoide quimérico.

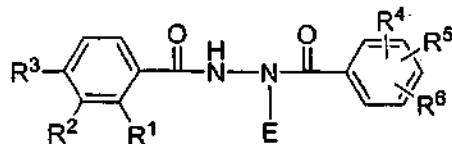
55. Uso de un ligando en la fabricación de un medicamento para modular la expresión de un gen en una célula huésped que comprende el gen que se quiere modular; en el que la célula huésped comprende el sistema de modulación de la expresión génica de acuerdo con la reivindicación 1, y en el que el gen que se quiere modular es un componente de un casete de expresión génica que comprende:

i) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido híbrido;

ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y

iii) un gen cuya expresión se quiere modular.

56. Uso de acuerdo con la reivindicación 55, en el que el ligando es un compuesto de la fórmula:



en la que:

E es un alquilo (C₄-C₆) que contiene un carbono terciario o un cianoalquilo (C₃-C₅) que contiene un carbono terciario;

R¹ es H, Me, Et, i-Pr, F, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OH, OMe, OEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, SCN o SCHF₂;

R² es H, Me, Et, n-Pr, i-Pr, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe₂, SMe, SEt, SOCF₃, OCF₂CF₂H, COEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, OCF₃, OCHF₂, O-i-Pr, SCN, SCHF₂, SMe, NH-CN, o se une con R³ y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R³ es H, Et, o se une con R² y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

R⁴, R⁵ y R⁶ son independientemente H, Me, Et, F, Cl, Br, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂O, CN, OCH, CH, 1-propinilo, 2-propinilo, vinilo, OMe, OEt, SMe o SEt.

57. Uso de acuerdo con la reivindicación 55, en el que dicho medicamento comprende además un segundo ligando, en el que el segundo ligando es un ácido 9-cis-retinoico o un análogo sintético de un ácido retinoico.

58. Uso de un ligando en la fabricación de un medicamento para modular la expresión de un gen en una célula huésped que comprende el gen que se quiere modular; en el que la célula huésped comprende el sistema de modulación de la expresión génica de acuerdo de la reivindicación 12, y en el que el gen que se quiere modular es

un componente de un casete de expresión génica que comprende:

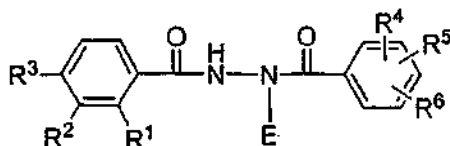
i) un elemento de respuesta reconocido por el dominio de unión a ADN del primer polipéptido híbrido;

5 ii) un promotor que se activa por el dominio de transactivación del segundo polipéptido híbrido; y

iii) un gen cuya expresión se quiere modular.

59. Uso de acuerdo con la reivindicación 58, en el que el ligando es un compuesto de la fórmula:

10



en la que:

15 E es un alquilo (C₄-C₆) que contiene un carbono terciario o un cianoalquilo (C₃-C₅) que contiene un carbono terciario;

R¹ es H, Me, Et, i-Pr, F, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OH, OMe, OEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, SCN o SCHF₂;

20 R² es H, Me, Et, n-Pr, i-Pr, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CH₂OMe, CH₂CN, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe₂, NEt₂, SMe, SEt, SOCF₃, OCF₂CF₂H, COEt, ciclopropilo, CF₂CF₃, CH=CHCN, alilo, azido, OCF₃, OCHF₂, O-i-Pr, SCN, SCHF₂, SMe, NH-CN, o se une con R³ y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

25

R³ es H, Et, o se une con R² y los carbonos de fenilo a los que están unidos R² y R³ para formar un etilendioxilo, un anillo dihidrofurilo con el oxígeno adyacente a un carbono de fenilo o un anillo dihidropirilo con el oxígeno adyacente a un carbono de fenilo;

30

R⁴, R⁵ y R⁶ son independientemente H, Me, Et, F, Cl, Br, formilo, CF₃, CHF₂, CHCl₂, CH₂F, CH₂Cl, CH₂OH, CN, C≡CH, 1-propinilo, 2-propinilo, vinilo, OMe, OEt, SMe o SEt.

60. Uso de acuerdo con la reivindicación 58, en el que dicho medicamento comprende además un segundo ligando, en el que el segundo ligando es un ácido 9-cis-retinoico o un análogo sintético de un ácido retinoico.

35

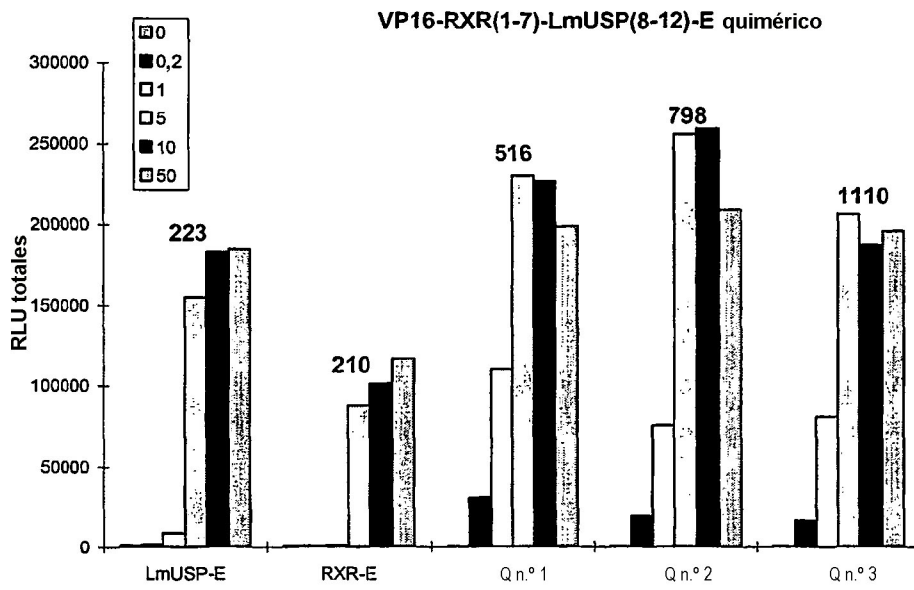


Figura 1

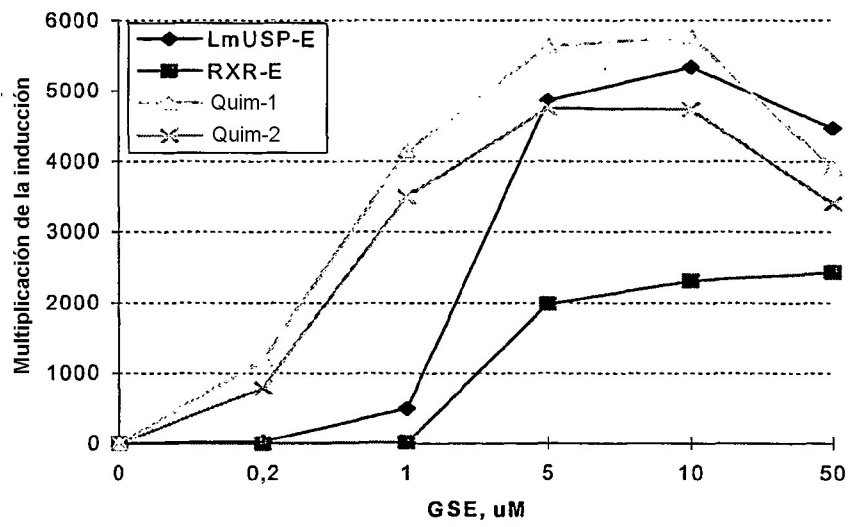


Figura 2

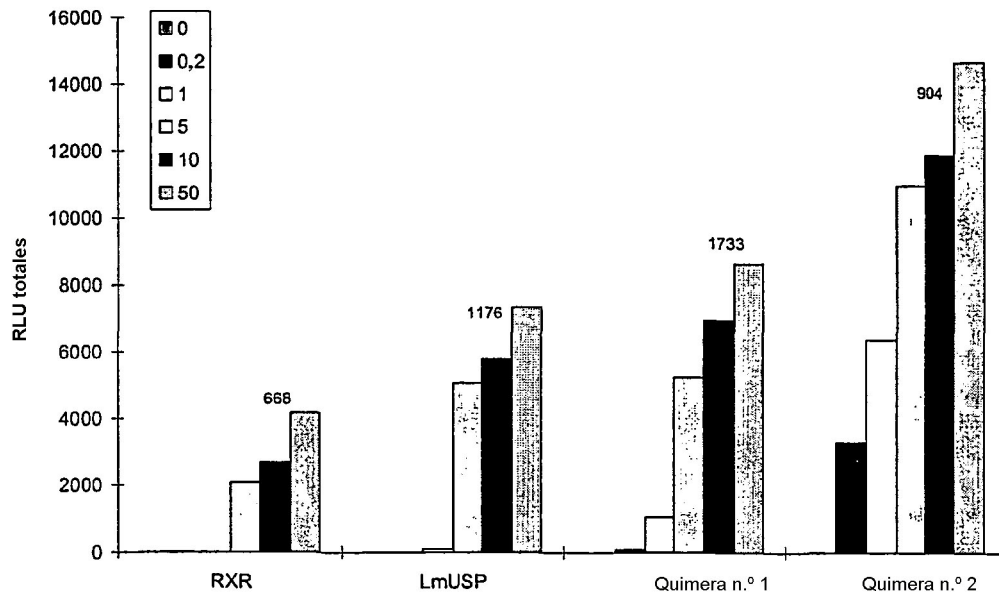


Figura 3

```

HsRXRbEF APEEMPVDRILEAELAVEQKSDQGVGPGGTGGSSPNDPVTNICQAADKQLFTLVEWA 60
MmRXRbEF APEEMPVDRILEAELAVEQKSDQGVGPGATGGGGSSPNDPVTNICQAADKQLFTLVEWA 60
HsRXRaEF ANEDMPVERILEAELAVEPKTETYVEAN--MGLNPSSPNDPVTNICQAADKQLFTLVEWA 58
MmRXRaEF ANEDMPVEKILEAELAVEPKTETYVEAN--MGLNPSSPNDPVTNICQAADKQLFTLVEWA 58
HsRXRgEF GHEDMPVERILEAELAVEPKTESYGDMN----MENSTNDPVTNICHAADKQLFTLVEWA 55
MmRXRgEF SHEDMPVERILEAELAVEPKTESYGDMN----VENSTNDPVTNICHAADKQLFTLVEWA 55
      H1              H3
      B6
HsRXRbEF KRIPHFSSLPLDDQVILLRAGWNELLIASFHRSIDVRDGILLATGLHVHRNSAHSAGVG 120
MmRXRbEF KRIPHFSSLPLDDQVILLRAGWNELLIASFHRSIDVRDGILLATGLHVHRNSAHSAGVG 120
HsRXRaEF KRIPHFSELPLDDQVILLRAGWNELLIASFHRSIAVKDGILLATGLHVHRNSAHSAGVG 118
MmRXRaEF KRIPHFSELPLDDQVILLRAGWNELLIASFHRSIAVKDGILLATGLHVHRNSAHSAGVG 118
HsRXRgEF KRIPHFSDLTLEDQVILLRAGWNELLIASFHRSVSVQDGILLATGLHVHRNSAHSAGVG 115
MmRXRgEF KRIPHFSDLTLEDQVILLRAGWNELLIASFHRSVSVQDGILLATGLHVHRNSAHSAGVG 115
      H4              H5      S1      S2      H6
      B8A1          B9
HsRXRbEF AIFDRVLTTELVS KM RDM RMDKTELGLCLRAIILFNPDAGKLSNPSEVEVLREKVYASLETY 180
MmRXRbEF AIFDRVLTTELVS KM RDM RMDKTELGLCLRAIIMFNPDAGKLSNPGEVEILREKVYASLETY 180
HsRXRaEF AIFDRVLTTELVS KM RDM QMDKTELGLCLRAIVLFPNPSKGLSNPAEVEALREKVYASLEAY 178
MmRXRaEF AIFDRVLTTELVS KM RDM QMDKTELGLCLRAIVLFPNPSKGLSNPAEVEALREKVYASLEAY 178
HsRXRgEF SIFDRVLTTELVS KM KDM QMDKSELGLCLRAIVLFPNPDAGKLSNPSEVETLREKVYATLEAY 175
MmRXRgEF SIFDRVLTTELVS KM KDM QMDKSELGLCLRAIVLFPNPDAGKLSNPSEVETLREKVYATLEAY 175
      H7              H8              H9
      B10          B11
HsRXRbEF CKQKYPEQQGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQLA 239
MmRXRbEF CKQKYPEQQGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQLA 239
HsRXRaEF CKHKYPEQFGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQMT 237
MmRXRaEF CKHKYPEQFGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQAT 237
HsRXRgEF TKQKYPEQFGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLETPLQIT 234
MmRXRgEF TKQKYPEQFGRFAKLLLR L PALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLETPLQIT 234
      H10          H11          H12          F

```

Figura 4A

ES 2 392 508 T3

```

LmRXREF HTDMPVERILEAEKRVECKAENQ-----VEY 26
AmRXREF HSDMPIERILEAEKRVECKMEQQ-----GNY 26
TmRXREF -AEMPLDRIIEAEKRIECPAGGSGG-----VGEQ 29
CpRXREF -SDMPIASIREAELSVDPIEQPLDQGVRLQVPLAPPDSEKCSFTLPFHPVSEVSCANPL 59
AmaRXR1EF PPEMPLEERILEAELRVES-QTGLTSES-----AQQ- 29
AmaRXR2EF SPDMPLEERILEAEMRVEQPAPSVLAQT-----AASG 31
H1

LmRXREF E-----LVEWAKHIPHFTSLPLEDQVLLL RAGWNELLIAAFSHRSVDVK 70
AmRXREF ENAVSHICNATNKQLFQLVAWAKHIPHFTSLPLEDQVLLL RAGWNELLIAAFSHRSIDVK 86
TmRXREF HDGVNNICQATNKQLFQLVQWAKLIPHFTSLPMSDQVLLL RAGWNELLIAAFSHRSIQAQ 89
CpRXREF QDVVSNICQAADRHVLQVLEWAKHIPHFTDLPIEDQVLLKAGWNELLIAAFSHRSMGVE 119
AmaRXR1EF QDPVSSICQAADRQLHQLVQWAKHIPHFEELPLEDRMVLLKAGWNELLIAAFSHRSVDVR 89
AmaRXR2EF RDPVNSMCQAAP-PLHELVQWARRIPHFELPIEDRTALLKAGWNELLIAAFSHRSVAVR 90
H3 H4 H5

LmRXREF DGIVLATGLTVHRNSAHQAGVGTIFDRVLTELVAKMREMMDRTELGCLRSVILFNPEVR 130
AmRXREF DGIVLATGLTVHRNSAQQAGVGTIFDRVLSELVSKMREMMDRTELGCLRSIILFNPEVR 146
TmRXREF DAIVLATGLTVNKTSAHAVGVGNIDRVLSELVNKMREMMDRTELGCLRAIILYNPTCR 149
CpRXREF DGIVLATGLVIHRSSAHQAGVGAI FDRVLSELVAKMKEMKIDRTELGCLRSIVLFNPAK 179
AmaRXR1EF DGIVLATGLVVQRHSAHGAGVGAI FDRVLTELVAKMREMMDRTELGCLLAVVLFNPEAK 149
AmaRXR2EF DGIVLATGLVVQRHSAHGAGVGDIFDRVLAELVAKMRDMKMDRTELGCLRAVVLFNPAK 150
S1 S2 H6 H7 H8

LmRXREF GLKSAQEVLELLREKVYAAL EYTRTTHPDEPGRFAKLLLR LPSLRSIGLKCLEHHLFFFRL 190
AmRXREF GLKSIQEVTLREKIYGALEGYCRVAVPDDAGRFKLLLR LPAIRSIGLKCLEYLFFFFKM 206
TmRXREF GIKSVQEVEMLEREKIYGVLE EYTRTTHPNEPGRFAKLLLR LPAIRSIGLKCS EHLFFFKL 209
CpRXREF GLNCVNDVEILREKVYAAL EYTRTTPDEPGRFAKLLLR LPAIRSIGLKCLEYLF LFKL 239
AmaRXR1EF GLRTCPGGPEGESV-SALEEHCRQY PDQGRFAKLLLR LPAIRSIGLKCLEHHLFFFKL 208
AmaRXR2EF GLRNATRVEALREKVYAAL EECRRHHPDQGRFGKLLLR LPAIRSIGLKCLEHHLFFFKL 210
H9 H10 H11

LmRXREF IGDVPIDTFLMEMLESPSDS----- 210
AmRXREF IGDVPIDDFLVEMLESRSDP----- 226
TmRXREF IGDVPIDTFLMEMLESPADA----- 229
CpRXREF IGDTPLD SYLMKMLVDNPNTSVTPPTS 266
AmaRXR1EF IGDTPIDNFLLSMLEAPSDP----- 228
AmaRXR2EF IGDTPIDSFLLNMLEAPADP----- 230
H12 F

```

Figura 4B

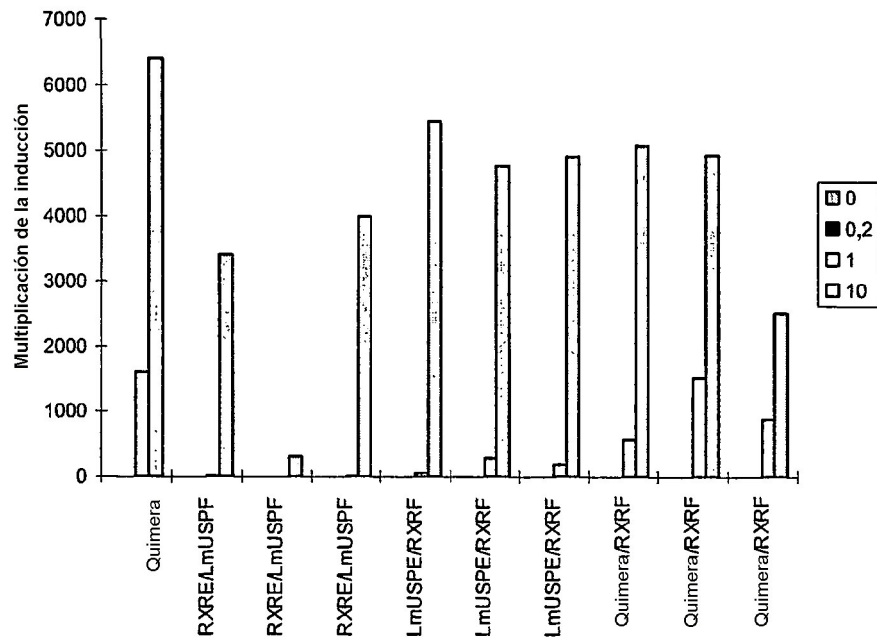


Figura 5

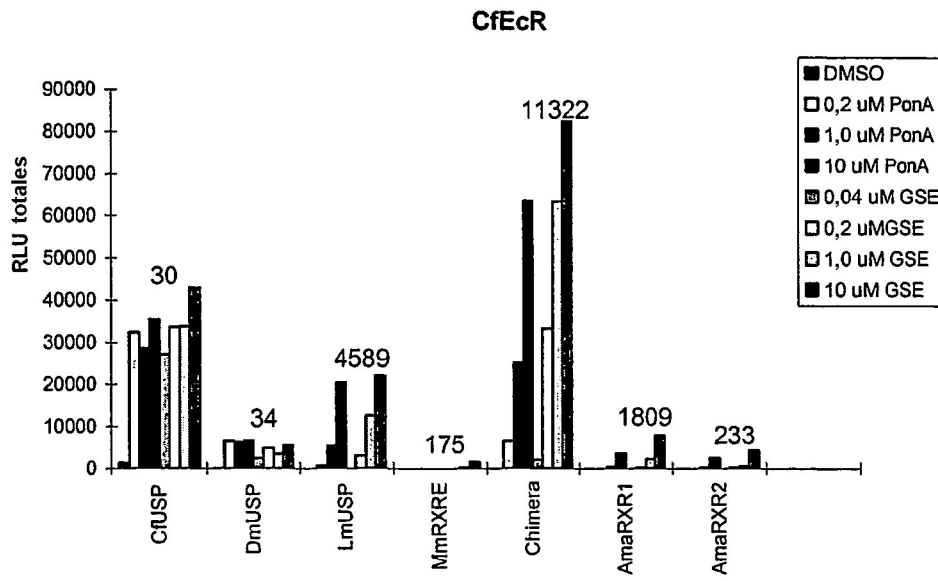


Figura 6

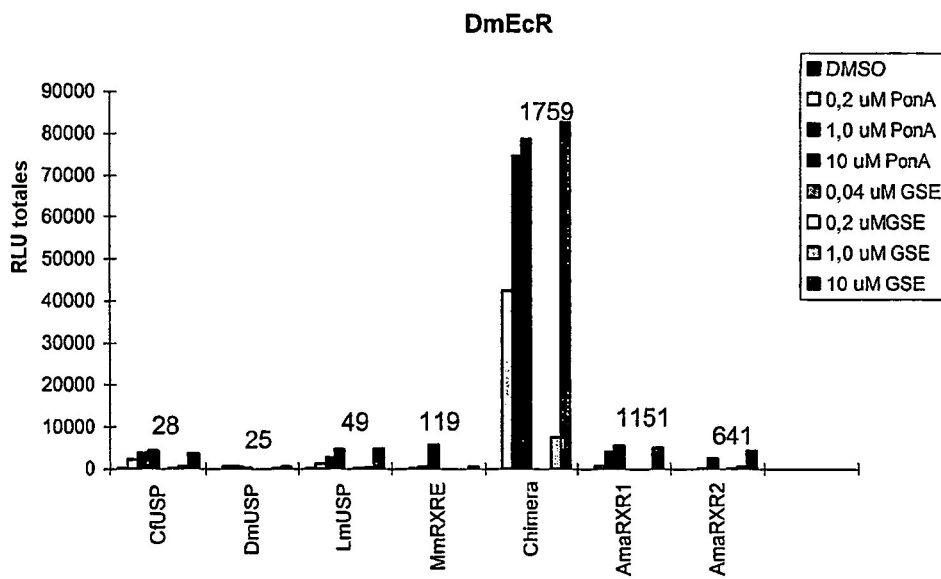


Figura 7

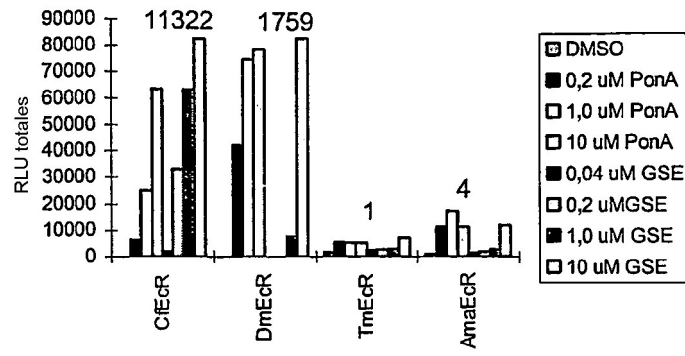


Figura 8

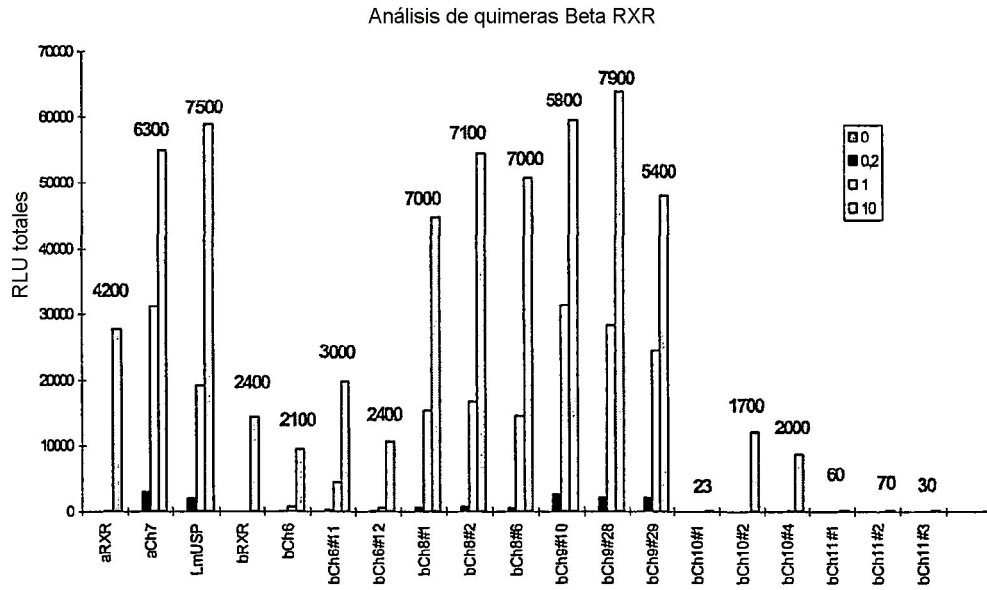


Figura 9

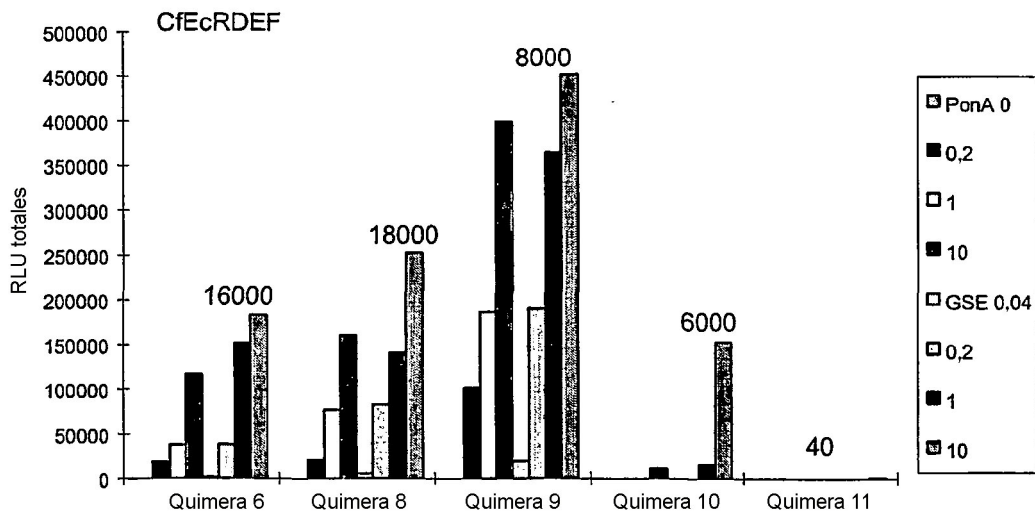


Figura 10

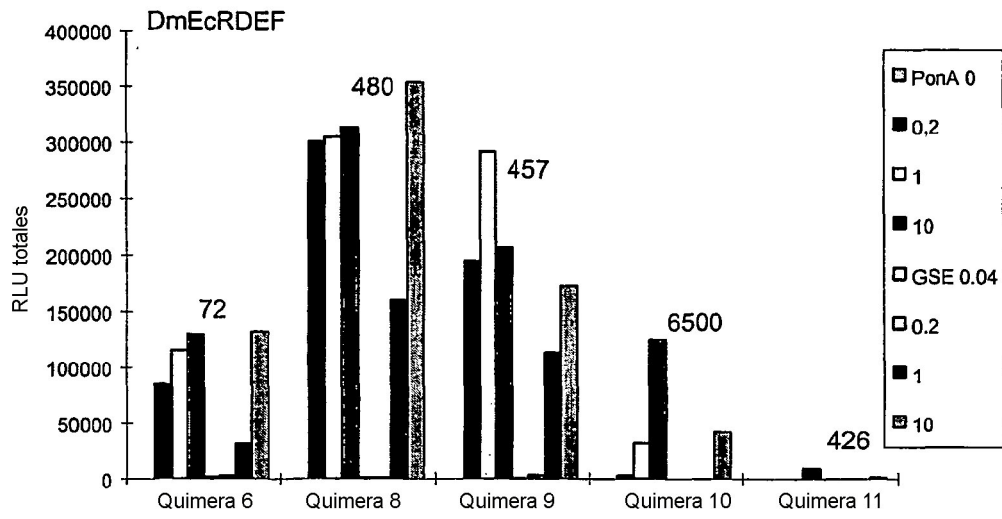


Figura 11

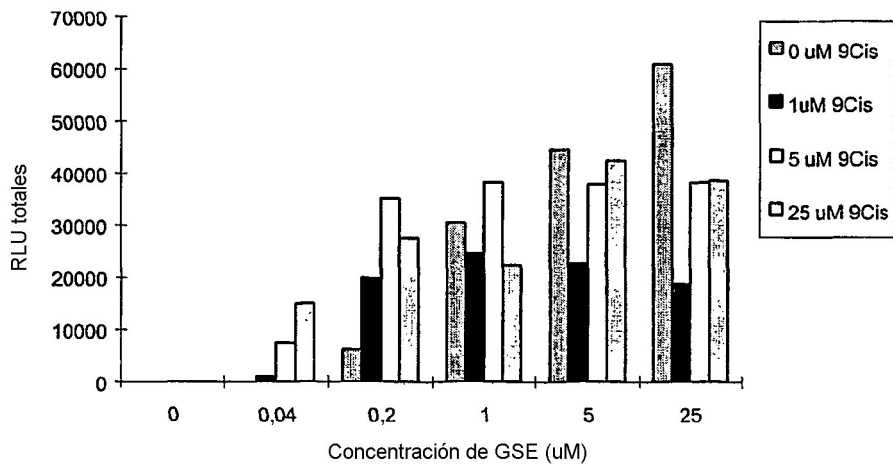


Figura 12