



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 306 802**

51 Int. Cl.:
C07K 14/52 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **02788225 .7**

86 Fecha de presentación : **20.12.2002**

87 Número de publicación de la solicitud: **1463754**

87 Fecha de publicación de la solicitud: **06.10.2004**

54 Título: **Proteína de pliegue con nudo de cistina.**

30 Prioridad: **21.12.2001 GB 0130738**

45 Fecha de publicación de la mención BOPI:
16.11.2008

45 Fecha de la publicación del folleto de la patente:
16.11.2008

73 Titular/es: **ARES TRADING S.A.**
Zone Industrielle de l'Ourietaz
1170 Aubonne, CH

72 Inventor/es: **Davies, Mark Douglas;**
Phelps, Christopher Benjamin;
Fagan, Richard Joseph;
Power, Christine;
Yorke-Smith, Melanie y
Ibberson, Mark

74 Agente: **Izquierdo Faces, José**

ES 2 306 802 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

Proteína de pliegue con nudo de cistina.

5 Esta invención se refiere a una proteína nueva (INSP002), aquí identificada como una proteína segregada que es un miembro de la familia Dan de la superfamilia de pliegue de nudo con cistina y al uso de esta proteína y las secuencias de ácido nucleico de los genes codificadores en el diagnóstico, prevención y tratamiento de la enfermedad.

Contexto

10 El proceso del descubrimiento del fármaco actualmente está sufriendo una revolución fundamental debido a que la era de la genómica funcional está llegando a la mayoría de edad. El término genómica funcional se aplica a la práctica que emplea herramientas bioinformáticas para atribuir funciones a secuencias proteicas de interés. Tales herramientas se están convirtiendo en muy necesarias ya que la velocidad de generación de datos secuenciales está rápidamente dejando atrás la habilidad de los laboratorios de investigación para asignar funciones a estas secuencias proteicas.

15 A medida que las herramientas bioinformáticas aumentan su potencia y precisión, estas herramientas rápidamente están sustituyendo a las técnicas convencionales de caracterización bioquímica. De hecho, las herramientas bioinformáticas avanzadas empleadas para identificar la presente invención son ahora capaces de producir resultados en los cuales se puede depositar un alto grado de confianza.

20 Varias instituciones y organizaciones comerciales están examinando los datos secuenciales cuando se encuentran disponibles y se están realizando significantes descubrimientos sobre las bases actuales. Sin embargo, existe una necesidad continua por identificar y caracterizar genes adicionales y los polipéptidos que codifican, como objetivos para investigación y para el descubrimiento de fármacos.

Contexto de proteína segregada

30 La habilidad de las células para hacer y segregar proteínas extracelulares es central para muchos procesos biológicos. Las enzimas, los factores de crecimiento, las proteínas de matriz extracelulares y las moléculas señalizadoras son segregados por células por medio de fusión de una vesícula secretora con la membrana plasmática. En la mayoría de los casos, pero no en todos, las proteínas se dirigen al retículo endoplásmico y a la vesícula secretora mediante un péptido señalizador. Los péptidos señalizadores son secuencias que actúan en cis que afectan al transporte de cadenas de polipéptidos a partir del citoplasma a un compartimiento enlazado con la membrana tal como la vesícula secretora. Los polipéptidos que se dirigen a las vesículas secretoras o se segregan a la matriz celular o se retienen en la membrana plasmática. Los polipéptidos que se retienen en la membrana plasmática tendrán uno o más dominios transmembránicos. Ejemplos de proteínas segregadas que juegan un papel central en el funcionamiento de una célula son las citoquinas, las hormonas, las proteínas matrices extracelulares (moléculas de adhesión), proteasas, y factores de crecimiento y de diferenciación.

40 Los factores de crecimiento representan un grupo relativamente grande de polipéptidos que comparten la propiedad de inducir la multiplicación celular tanto *in vivo* como *in vitro*. Los factores de crecimiento se diferencian de las hormonas clásicas endocrinas como la insulina o la hormona de crecimiento de dos importantes maneras. En primer lugar, las hormonas endocrinas normalmente se sintetizan en glándulas especializadas (como el páncreas, en el caso de la insulina) mientras que los factores de crecimiento a menudo se sintetizan en múltiples tipos de células y tejidos. En segundo lugar, las hormonas clásicas endocrinas se liberan en los fluidos corporales en el punto de la síntesis y se transportan a su tejido objetivo en el flujo sanguíneo. Un sello distintivo de los factores de crecimiento es que, en la mayoría de los casos, actúan localmente en el interior de los tejidos en los que están sintetizados (recogido en Heath, JK. (1993) Growth Factors, Oxford University Press, UK, pp.15-33).

50 A pesar de que el nivel de similitud (secuencia) entre factores de crecimiento no es alto, pueden clasificarse en superfamilias en base a sus similitudes estructurales y funcionales. Ejemplos de estas superfamilias incluyen: (a) factores de crecimiento hematopoyéticos, como la hormona de crecimiento, IL-2, IL-4, G-CSF, y CNTF, que todas poseen un motivo estructural con un conjunto de cuatro hélices; (b) los miembros de la familia beta-trifolia, como IL-1 beta, IL-1 alpha, FGF, y el factor de crecimiento queratinocito, (c) los factores de crecimiento similares a EGF tales como EGF y TGF alpha, que tienen un dominio similar a la inmunoglobulina; y (d) el factor de crecimiento de pliegue de nudo de cistina que incluye NGF, TGF beta, PDGF, y hormonas de glicoproteína.

60 Los factores de crecimiento son extracelulares y con el fin de ejercer un efecto biológico, interactúan con receptores específicos de elevada afinidad localizados en las membranas plasmáticas de las células objetivo. La caracterización molecular de una variedad de diferentes receptores del factor de crecimiento ha revelado que pueden clasificarse en familias definidas: los receptores tirosina quinasa, los receptores de siete transmembranas asociados a la proteína G, y los receptores serina/treonina quinasa. Los receptores tirosina quinasa se caracterizan por un dominio extracelular, un dominio en la transmembrana y un dominio intracelular que posee actividad tirosina quinasa. Estos receptores del factor de crecimiento serina/treonina quinasa son similares a los receptores tirosina quinasa con un dominio extracelular, un dominio en la transmembrana, y un dominio intracelular. El dominio intracelular tiene actividad intrínseca serina/treonina quinasa.

La desregulación de los factores de crecimiento se encuentra implicada en una variedad de estados enfermos, incluyendo, aunque no limitando, las enfermedades oncológicas (Bartucci M *et al.*, (2001) Cancer Res. Sep 15; 61 (18):6747-54; Dias S *et al.*, (2001) Proc Natl Acad Sci USA Sep 11; 98 (19): 10857-62, Djavan B *et al.* (2001) World J Urol. 19(4): 225-33), enfermedades inflamatorias (Fiocchi C. (2001) J Clin Invest. Aug; 108 (4): 523-6, Hodge S *et al.* (2001) Respirology. Sep; 6(3): 205-211, Fenwick SA *et al.*, (2001) J. Anat. Sep; 199 (Pt 3): 231-40), enfermedades neurológicas (Cooper JD *et al.*, (2001) Proc Natl Acad Sci USA 98 (18): 10439-44, Fahnstock M *et al.*, (2001) Mol Cell Neurosci 18(2): 210-20), y enfermedades metabólicas (Vickers MH *et al.*, (2001) Endocrinology. 142(9):3964-73).

10 Superfamilia de pliegue con nudo de cistina

La típica estructura vista en la superfamilia con nudo de cistina se basa en la presencia de 6 residuos de cisteína creando 3 enlaces disulfuro. Dos de los enlaces disulfuro crean una estructura similar a un anillo que es penetrada por el tercer enlace disulfuro (Sun *et al.*, 1995). Los dominios con nudo de cistina a menudo se encuentran con más de 15 residuos de cisteína. Los residuos de cisteína extra normalmente se emplean para crear enlaces disulfuros adicionales en el interior del dominio con el nudo cistina o intercambiar enlaces disulfuro durante la dimerización.

La superfamilia con nudo de cistina se divide en subfamilias que incluyen las hormonas glicoproteína (por ejemplo, folículos que estimulan la hormona), el factor beta de crecimiento transformador (TGFBeta) proteínas (por ejemplo, 20 la proteína morfogenética ósea 4), las proteínas similares al factor (similares a PDGF) de crecimiento derivado de plaquetas (por ejemplo, el factor de crecimiento A derivado de plaqueta), factores de crecimiento nervioso (NGF) (por ejemplo, factor neurotrófico derivado del cerebro) y el gen anómalo seleccionado en el análisis diferencial en la familia del neuroblastoma (DAN) (por ejemplo, cerberus). La subfamilia DAN incluye Cer1, Cerberus, Caronte, Drm/Kremlin, PRDC, DAN, Dante y CeCan1 (Massague *et al.*, Genes Dev 2000 Mar 15; 14(6):627-44; Massague & 25 Wotton, EMBO J. 2009 Apr 17; 19(8):1745-54).

Se piensa que los miembros de la subfamilia DAN pueden ser capaces de modular las acciones de los miembros de la subfamilia TGFBeta de proteínas (Pearce *et al.*, Dev. Biol. 1999 Mayo 1; 209(1):98-110). Más específicamente, es posible que los miembros de la subfamilia DAN sean capaces de modular las acciones de proteínas morfogenéticas 30 óseas (BMPs) durante el desarrollo.

Se ha descubierto que los miembros de la subfamilia DAN actúan como antagonistas de proteínas morfogenéticas óseas (BMP), que son miembros de la subfamilia TGFBeta de la superfamilia con nudo de cistina (Stanley *et al.*, Mech Dev. 1998 Oct; 77(2):173-84; Massague *et al* 2000 (*supra*); Massague J & Wotton D, 2002 (*supra*). Los mo- 35 nómeros BMP homo o heterodiméricos, por medio del enlace de sus dominios con nudo de cisteína, interactúan con los receptores de la superficie celular. Se cree que los miembros de la subfamilia DAN son capaces de enlazarse con BMPs por medio de sus propios dominios con nudo de cistina. Esto evita que BMP se enlace con su pareja natural de dimerización y como resultado, el BMP ya no es capaz de interactuar con su receptor señalizador de superficie celular. Experimentos enfocados específicamente a DAN, Cer1, y DRM han demostrado que inhiben la acción de BMP4 40 (Pearce *et al.*, 1999 (*supra*)).

Se ha logrado una mayor comprensión de la función del cerberus como un resultado de estudios de enlace (Piccolo S. *et al.*, Nature, 1999 Feb 25; 397(6271):707-10). Los primeros estudios funcionales llevados a cabo sobre el cerberus usaron la proteína cerberus *Xenopus laevis* (cer). La microinyección de *Xenopus* cerberus mRNA en embriones *Xeno-* 45 *pus* reveló que la proteína cer producía la formación de cabezas ectópicas en el endodermo anterior del organizador de Spemann (Bouwmeester *et al.*, Nature 1996 Aug 15; 382(6592):595-601; Bouwmeester T., INT J Dev Biol. 2000, 145 (1 Spec No):251-8). Los estudios sobre enlaces realizados por Piccolo y sus colaboradores revelaron que la proteína cerberus *Xenopus* se enlaza e inhibe las acciones de Nodal, BMP y proteínas Wnt a través de puntos independientes. Más específicamente, descubrieron que el cerberus tiene una elevada afinidad específica y un efecto inhibitorio sobre 50 Xnr-1 (miembro de la familia Nodal), BMP4 (miembro de la familia BMP) y Xwmt-8 (miembro de la familia Wnt). Este trabajo se une con el cerberus, y a partir de ahí con otros miembros de la familia DAN, para las rutas de desarrollo y diferenciación de tejido.

La esclerostina, codificada por el gen SOST, es también un miembro de la subfamilia DAN (Brumkow *et al.*, 2001, 55 Am. J. Hum. Genet. 68:577-589). SOST se ha unido con la esclerosteosis, una displasia ósea autosómica recesiva. El fenotipo asociado por la esclerosteosis es un sobrecrecimiento óseo progresivo, que puede dar lugar a gigantismo, deformación de las facies y atropamiento del séptimo u octavo nervio craneal (Brumkow *et al*, 2001 (*supra*). La unión entre esclerosteosis y SOST se determinó a través de mapeo de homocigosidad en familias que están afectadas por la enfermedad. Brumkow y colaboradores identificaron una similitud entre el fenotipo asociado con la esclerosteosis y 60 los efectos asociado con otros miembros de la subfamilia DAN. Esta unión se fortaleció por medio de la sugerencia de que la esclerosteosis puede surgir debido a la pérdida de un regulador negativo del miembro de la subfamilia TGFBeta, más especialmente BMP.

La identificación de proteínas segregadas y en particular los factores de crecimiento, como los miembros de su 65 subfamilia de pliegue con nudo cistina, y en particular miembros de la subfamilia DAN, son por lo tanto de extrema importancia en el aumento de la comprensión de las rutas subyacentes que dan lugar a los estados 5 de enfermedad y estados de enfermedad asociada, mencionados anteriormente, y para desarrollar genes más efectivos o terapias farmacéuticas para tratar estos desórdenes.

La invención

La invención se basa en el descubrimiento de que la proteína INSP002 funciona como una proteína segregada y además como una proteína segregada de la subfamilia DAN de la superfamilia citoquina de pliegue y nudo de cistina.

En un primer aspecto, la invención proporciona un polipéptido que tiene la función de un miembro de la subfamilia DAN de la citoquinas con pliegue y nudo de cistina, que:

- (i) comprende o consiste en la secuencia de aminoácido tal y como se señala en SEQ ID NO:6, SEQ ID NO:8, o SEQ ID NO:14
- (ii) es un fragmento de la misma;
- (iii) es un equivalente funcional de (i) o (ii) que posee más del 70% de identidad secuencial con (i) o (ii).

El polipéptido que tiene la secuencia señalada en SEQ ID NO:6 se refiere a partir de ahora como “el polipéptido INSP002”. Los primeros 22 aminoácidos del polipéptido INSP002 son un péptido señalizador y la secuencia del polipéptido INSP002 sin la secuencia señalizadora se indica en la SEQ ID NO:8. El polipéptido que tiene la secuencia indicada en SEQ ID NO:8 se refiere a partir de ahora como “el polipéptido INSP002 sin péptido señalizador”.

El polipéptido que tiene la secuencia indicada en SEQ ID NO:14 es una variante del polipéptido INSP002. Es idéntico al polipéptido INSP002 excepto en que contiene una eliminación de dos aminoácidos en las posiciones 107 y 108 y una única sustitución de aminoácido en la posición 110 en comparación con el polipéptido INSP002. El polipéptido que tiene la secuencia señalada en SEQ ID NO:14 se refiere a partir de ahora como “la variante polipéptido INSP002”.

El polipéptido de acuerdo con el primer aspecto de la invención funciona como un miembro de la subfamilia DAN de la superfamilia de citoquina de pliegue con nudo de cistina. El término “citoquina de pliegue con nudo de cistina” es bien comprendido en la técnica y el experto profesional fácilmente será capaz de distinguir si un polipéptido funciona como un miembro de la superfamilia de citoquina de pliegue con nudo de cistina usando uno u otro del conjunto de ensayos conocidos en la técnica.

En particular, el experto en la técnica puede ser capaz de distinguir si un polipéptido funciona como un miembro de la subfamilia DAN analizando si es un antagonista de los miembros de la superfamilia TGFβ y en particular, si es un antagonista BMP. El embrión *Xenopus* puede emplearse como un sistema para analizar si un polipéptido funciona como un antagonista BMP ya que varios BMPs se expresan en el embrión *Xenopus* (Chang C. *et al.*, 1999, Development 126:3347-3357, Hawley S. *et al.*, 1995, Genes Dev. 9:2923-2935, Hemmati-Brivanlou, A., y G. H. Thomsen. 1995, Dev. Genet. 17:78-89, Jones C. M. *et al.*, 1992, Development 115:639-647). La sobreexpresión de las señales BMP-2/4-clase o BMP-7-clase en el mesodermo temprano produce destinos ventrales, mientras que los inhibidores de estas señales (tales como *Noggin*, *Xnr3*, *Chordin*, o *Follistatin*) producen destinos dorsales. El efecto de un polipéptido sobre el desarrollo embrionario puede por lo tanto emplearse para determinar si el polipéptido es un antagonista BMP.

El término “polipéptidos INSP002” como aquí se emplea incluye polipéptidos que comprenden el polipéptido INSP002, el polipéptido INSP002 sin péptido señalizador o el polipéptido variante INSP002.

En un segundo aspecto, la invención proporciona una molécula de ácido nucleico purificado que codifica un polipéptido del primer aspecto de la invención.

Preferentemente, la molécula de ácido nucleico purificado comprende la secuencia de ácido nucleico tal y como se indica en SEQ ID NO:5 (que codifica el polipéptido INSP002), o SEQ ID NO:13 (que codifica el polipéptido variante INSP002).

La invención además proporciona que la molécula de ácido nucleico purificado consiste en la secuencia de ácido nucleico tal y como se indica en SEQ ID NO:5 (que codifica el polipéptido INSP002), o SEQ ID NO:13 (que codifica el polipéptido variante INSP002).

De acuerdo con una realización de este aspecto de la invención, la molécula de ácido nucleico purificado no contiene la región no traducida 5' localizada en dirección ascendente de la secuencia de ácido nucleico que codifica el polipéptido INSP002 (nucleótidos 1 a 151 de la SEQ ID NO:5). De acuerdo con esta realización, la molécula de ácido nucleico purificado preferentemente comprende los nucleótidos 152 a 721 de la SEQ ID NO:5. La invención además proporciona una molécula de ácido nucleico purificado consistente en los nucleótidos 152 a 721 de la SEQ ID NO:5. La secuencia de nucleótidos que codifica el polipéptido INSP002 sin la región no traducida 5' (nucleótidos 152 a 721 de la SEQ ID NO:5) se da en la región SEQ ID NO:11.

De acuerdo con una realización adicional de este aspecto, la molécula de ácido nucleico purificado no codifica el péptido señalizador localizado en el inicio del polipéptido INSP002 (nucleótidos 152 a 217 de la SEQ ID NO:5). De acuerdo con esta realización, la molécula de ácido nucleico purificado preferentemente comprende los nucleótidos

218 a 721 de SEQ ID NO:5 (codificando el polipéptido INSP002 sin péptido señalizador). La invención además proporciona una molécula de ácido nucleico purificado consistente en los nucleótidos 218 a 721 de la SEQ ID NO:5 (codificando el polipéptido INSP002 sin péptido señalizador). La secuencia de nucleótidos que codifica el polipéptido madura INSP002 (SEQ ID NO:7) se da en la SEQ ID NO:9.

5

De acuerdo con una realización adicional de este aspecto de la invención, la molécula de ácido nucleico no contiene la región no traducida 5' localizada en sentido ascendente de la secuencia de ácido nucleico que codifica el polipéptido variante INSP002 (nucleótidos 1 a 68 de SEQ ID NO:13). De acuerdo con esta realización, la molécula de ácido nucleico purificado preferentemente comprende o consiste en los nucleótidos 69 a 719 de SEQ ID NO:13. La secuencia de nucleótidos que codifica el polipéptido variante INSP002 sin la región no traducida 5' (nucleótidos 69 a 719 de SEQ ID NO:13) se da en la SEQ ID NO:15.

10

En un tercer aspecto, la invención proporciona una molécula de ácido nucleico purificado que codifica un péptido que tiene la función de un miembro de la subfamilia DAN de citoquinas de pliegue con nudo cistina y que se hibridiza bajo condiciones de elevado rigor con una molécula de ácido nucleico del segundo aspecto de la invención.

15

En un cuarto aspecto, la invención proporciona un vector, tal como un vector de expresión, que contiene una molécula de ácido nucleico del segundo y tercer aspecto de la invención.

20

En un quinto aspecto, la invención proporciona una célula huésped transformada con un vector del cuarto aspecto de la invención.

25

En un sexto aspecto de la invención, la invención proporciona un ligando que es un anticuerpo y que se enlaza específicamente con, y que preferiblemente inhibe la actividad de citoquina de pliegue con nudo cistina de un polipéptido del primer aspecto de la invención que es un miembro de la subfamilia DAN de citoquinas de pliegue con nudo cistina.

30

En un séptimo aspecto, la invención proporciona un polipéptido del primer aspecto de la invención, o una molécula de ácido nucleico del segundo o tercer aspecto de la invención, o un vector del cuarto aspecto de la invención, o un ligando del quinto aspecto de la invención, para uso en terapia y diagnóstico. Estas moléculas también pueden emplearse en la fabricación de un medicamento para el tratamiento de desórdenes proliferativos celulares, desórdenes inmunes/inflamatorios, desórdenes cardiovasculares, desórdenes neurológicos, desórdenes de desarrollo, desórdenes metabólicos, infecciones y otras condiciones patológicas.

35

En un octavo aspecto, la invención proporciona un método de diagnóstico de una enfermedad en un paciente, consistente en el análisis del nivel de expresión de un gen natural codificador de un polipéptido del primer aspecto de la invención o la actividad de un polipéptido del primer aspecto de la invención en el tejido de dicho paciente y comparando dicho nivel de expresión o actividad con un nivel de control, donde un nivel que es diferente a dicho nivel de control es indicativo de enfermedad. Tal método preferiblemente se llevará a cabo *in vitro*. Pueden emplearse métodos similares para controlar el tratamiento terapéutico de enfermedad en un paciente, donde la alteración del nivel de expresión o actividad de un polipéptido o de una molécula de ácido nucleico en el periodo de tiempo hacia un nivel de control son indicativos de regresión o retroceso de enfermedad.

40

El desorden o enfermedad en los aspectos séptimo y octavo de la invención es preferiblemente aquel en el que están implicados los niveles aberrantes de un citoquina de pliegue con nudo de cistina, preferiblemente un miembro de la subfamilia DAN. La enfermedad o desorden puede ser uno en el que los niveles aberrantes de los miembros de la superfamilia TGFβ están implicados. En particular, la enfermedad o desorden puede ser aquella en la que se implican BMPs, tales como neuropatías, nefropatías como nefropatía diabética, cáncer, curación de heridas, fibrosis, osteopenia, osteoporosis, fracturas y esclerosteosis.

50

Un método preferente para detectar polipéptidos de un primer aspecto de la invención comprende los pasos de: (a) contactar un ligando, que es un anticuerpo, del sexto aspecto de la invención con una muestra biológica bajo condiciones adecuadas para la formación de un complejo ligando-polipéptido; y (b) detectar dicho complejo.

55

Existe un número de varios de estos métodos de acuerdo con el noveno aspecto de la invención, tal y como el lector especializado se dará cuenta, como los métodos de hibridización de ácido nucleico con investigaciones cortas, análisis de mutación de punto, amplificación de reacción en cadena de polimerasa (PCR) y métodos que emplean anticuerpos para detectar niveles aberrantes de proteína. Se pueden emplear métodos similares en una base de tiempo corto o largo para permitir el tratamiento terapéutico de la enfermedad a ser controlada en el paciente. La invención también comprende kits que son útiles en estos métodos para el diagnóstico de enfermedad.

60

En un noveno aspecto, la invención proporciona el uso de un polipéptido del primer aspecto de la invención como un miembro de la subfamilia DAN de citoquinas de pliegue con nudo citoquina.

65

En un décimo aspecto, la invención proporciona una composición farmacéutica que comprende un polipéptido del primer aspecto de la invención, o una molécula de ácido nucleico del segundo o tercer aspecto de la invención, o un vector del cuarto aspecto de la invención, o una célula huésped del quinto aspecto de la invención, o un ligando del sexto aspecto de la invención, junto con un portador farmacéuticamente aceptable.

En un undécimo aspecto, la presente invención proporciona un polipéptido del primer aspecto de la invención, o una molécula de ácido nucleico del segundo o tercer aspecto de la invención, o un vector del cuarto aspecto de la invención, o una célula huésped del quinto aspecto de la invención, o un ligando del sexto aspecto de la invención, para uso en la fabricación de un medicamento para el diagnóstico o tratamiento de una enfermedad.

La enfermedad en el undécimo aspecto de la invención es preferentemente una enfermedad en la cual se implican los niveles aberrantes de una citoquina de pliegue con nudo cistina, preferiblemente de un miembro de la subfamilia DAN. La enfermedad también puede ser aquella en la que se implican los niveles aberrantes de un ligando de una citoquina de pliegue con nudo cistina, preferiblemente un ligando de un miembro de la subfamilia DAN. Por ejemplo, la enfermedad puede ser una enfermedad en la que los niveles aberrantes de un miembro de la superfamilia TGFβ se encuentran implicados. En particular, la enfermedad o desorden puede ser aquella en la que se implican BMPs, tales como neuropatías, nefropatías como nefropatía diabética, cáncer, curación de heridas, fibrosis, osteopenia, osteoporosis, fracturas y esclerosteosis.

Para enfermedades en la que la expresión de un gen natural codificador de un polipéptido del primer aspecto de la invención, o en el que la actividad de un polipéptido del primer aspecto de la invención, es inferior en un paciente enfermo en comparación con el nivel de expresión o actividad en un paciente sano, el polipéptido, ácido nucleico o ligando administrado al paciente debería ser un agonista. De manera inversa, para enfermedades en la que la expresión de un gen natural o la actividad de un polipéptido es superior en un paciente enfermo en comparación con el nivel de expresión o actividad en un paciente sano, el polipéptido, ácido nucleico o ligando administrado al paciente debería ser un antagonista. Ejemplos de tales antagonistas incluyen moléculas de ácido nucleico, ribozimas y anticuerpos.

En un duodécimo aspecto, la invención proporciona animales no humanos transgénicos o knockout que se han transformado para expresar niveles superiores, inferiores o niveles ausentes de un polipéptido del primer aspecto de la invención. Tales animales transgénicos son modelos muy útiles para el estudio de enfermedades y pueden también emplearse en regímenes de análisis para la identificación de compuestos que son efectivos en el tratamiento de diagnóstico de tales enfermedades.

Un resumen de técnicas y procedimientos habituales que pueden emplearse con el fin de utilizar la invención se presentan a continuación. Se entenderá que esta invención no está limitada a la metodología, protocolos, líneas celulares, vectores y reagentes particulares aquí descritos. También se entenderá que la terminología aquí empleada tiene únicamente el fin de describir realizaciones particulares y no se pretende que esta terminología se limite al alcance de la presente invención. La extensión de la invención se limita sólo se encuentra limitada por los términos de las reivindicaciones adjuntas.

Las abreviaciones habituales para nucleótidos y aminoácidos se emplean en esta descripción.

La práctica de la presente invención empleará, a menos que se indique lo contrario, técnicas convencionales para biología molecular, microbiología, tecnología de DNA recombinante e inmunobiología, que se encuentran en el ámbito de estudio y práctica de aquellos expertos en la técnica.

Tales técnicas se explican completamente en la bibliografía. Ejemplos de textos particularmente útiles para consulta incluyen los siguientes: Sambrook Molecular Cloning; A Laboratory Manual, Segunda Edición (1989); DNA Cloning, volúmenes I y II (D. N. Glover ed. 1985); Oligonucleotide Synthesis (M. J. Gait ed. 1984); Nucleic Acid Hybridization (B. D. Names & S. J. Higgins eds. 1984); Transcription and Translation (B. D. Hames & S. J. Higgins eds. 1984); Animal Cell Culture (R. I. Freshney ed. 1986); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984); The Methods in Enzymology Series (Academic Press, Inc.), especialmente volúmenes 154 y 155; Gene Transfer Vectors for Mammalian Cells (J. H. Millar y M. P. Calos eds. 1987, Cold Spring Harbor Laboratory); Immunocytochemical Methods in Cells and Molecular Biology (Mayer y Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Segunda Edición (Springer Verlag, N.Y.); y Handbook of Experimental Immunology, volúmenes I-IV (D.M. Weir y C. C. Blackwell eds. 1986).

Como aquí se emplea, el término “polipéptido” incluye cualquier péptido o proteína que esté formada por dos o más aminoácidos unidos entre sí mediante enlaces péptidos o enlaces péptidos modificados, es decir, isómeros péptidos. Este término se refiere tanto a cadenas cortas (péptidos y oligopéptidos) como a cadenas largas (proteínas).

El polipéptido de la presente invención puede tener la forma de una proteína madura o puede ser una pre-, pro- o prepro-proteína que puede activarse mediante la división de la pre-, pro- o prepro- porción para producir un polipéptido maduro activo. En tales polipéptidos, la pre-, pro- o prepro-secuencia puede ser una secuencia líder o de segregación o puede ser una secuencia que se emplee para la purificación de la secuencia de polipéptido maduro.

El polipéptido del primer aspecto de la invención puede formar parte de una proteína de fusión. Por ejemplo, con frecuencia resulta ventajoso incluir una o más secuencias adicionales de aminoácido que puedan contener secuencias de secreción o secuencias líder, pro-secuencias, secuencias que pueden ayudar a la purificación, o secuencias que confieran elevada estabilidad proteica, por ejemplo durante la producción recombinante. De manera alternativa o adicionalmente, el polipéptido maduro o puede fusionarse con otro compuesto, como un compuesto para aumentar la vida media del polipéptido (por ejemplo, un glicol polietileno).

Los polipéptidos pueden contener aminoácidos diferentes a los 20 aminoácidos que codifican el gen, modificados bien mediante procesos naturales, como procesos post-translacionales o mediante técnicas de modificación química que son bien conocidas en la técnica. Entre las modificaciones bien conocidas que comúnmente pueden estar presentes en polipéptidos de la presente invención son la glicosilación, la unión de lípidos, la sulfación, carboxilación gamma, por ejemplo residuos ácidos glutámicos, hidroxilación y ribosilación ADP. Otras modificaciones potenciales incluyen acetilación, acilación, amidación, unión covalente de flavina, unión covalente de haeme, unión covalente de un nucleótido o derivado nucleótido, unión covalente de un derivado lípido, unión covalente de fosfatidilinositol, enlace cruzado, ciclización, formación de enlace disulfuro, demetilación, formación de enlaces cruzados covalentes, formación de cisteína, formación de piroglutamato, formilación, formación de ancla GPI, yodinación, mutilación, miristoilación, oxidación, procesos proteolíticos, fosforilación, prenilación, racemización, selenoilación, adición mediada por transferencia de RNA de aminoácidos a proteínas tales como arginilación y ubiquitinación.

Las modificaciones pueden darse en cualquier momento en un polipéptido, incluyendo el eje del péptido, las cadenas secundarias del aminoácido y las terminales del amino o carboxi. De hecho, el bloqueo de la terminal amino o carboxil en un polipéptido, o el de ambos, mediante una modificación covalente es común en polipéptidos que se dan de manera natural o en sintéticos y tales modificaciones pueden estar presentes en polipéptidos de la presente invención.

Las modificaciones que se dan en un polipéptido a menudo serán una función de cómo está hecho un polipéptido. Para polipéptidos que están hechos de manera recombinante, la naturaleza y extensión de las modificaciones en gran parte se determinarán mediante la capacidad de modificación post-translacional de la célula huésped particular y de la señales de modificación que están presentes en la secuencia de aminoácido del polipéptido en cuestión. Por ejemplo, los patrones de glicosilación varían entre diferentes tipos de células huésped.

Los polipéptidos de la presente invención pueden prepararse de cualquier modo apropiado. Tales polipéptidos incluyen polipéptidos aislados que ocurren de manera natural (por ejemplo, purificados de cultivos celulares) o polipéptidos producidos de manera recombinante (incluyendo proteínas de fusión), polipéptidos producidos sintéticamente o polipéptidos que se producen mediante combinación de estos métodos.

Los polipéptidos funcionalmente equivalentes del primer aspecto de la invención son homólogos a los polipéptidos INSP002. Se dice que dos polipéptidos son "homólogos", como aquí se emplea el término, si la secuencia de uno de los polipéptidos tiene un grado lo suficientemente alto de identidad o similitud con la secuencia del otro polipéptido. "Identidad" indica que en cualquier posición particular en las secuencias alineadas, el residuo de aminoácido es idéntico entre las secuencias. "Similitud" indica que, en cualquier posición particular en las secuencias alineadas, el residuo de aminoácido es de un tipo similar entre las secuencias. Los grados de identidad y similitud se pueden calcular fácilmente (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New Cork, 1988; Biocomputing. Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New Cork, 1993; Computer Analysis of Sequence Data, Parte 1, Griffin, A. M., y Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, New York, 1991).

Por lo tanto, los polipéptidos homólogos incluyen variantes biológicas naturales (por ejemplo, variantes alélicas o variaciones geográficas dentro de las especies a partir de las cuales se derivan los polipéptidos) y mutantes (como los mutantes que contienen sustituciones, inserciones o eliminaciones de aminoácidos) de los polipéptidos INSP002. Tales mutantes pueden incluir polipéptidos en los que uno o más de los residuos de aminoácido se sustituyen por un residuo de aminoácido conservado o no conservado (preferentemente un residuo de aminoácido conservado) y tal residuo de aminoácido sustituido puede ser o no codificado por el código genético. Tales sustituciones habituales se encuentran entre Ala, Val, Leu e Ile; entre Ser y Thr; entre los residuos ácidos Asp y Glu; entre Asn y Gln; entre los residuos básicos Lys y Arg; o entre los residuos aromáticos Phe y Tyr. En particular, las variantes preferentes son aquellas en las que varios aminoácidos, por ejemplo, entre 5 y 10, 1 y 5, 1 y 3, 1 y 2, o sólo 1 son sustituidos, eliminados o añadidos en cualquier combinación. Son especialmente preferentes las sustituciones, adiciones y eliminaciones silenciosas que no alteran las propiedades o actividades de la proteína. Son también especialmente preferentes en este aspecto las sustituciones conservadoras. Tales mutantes también incluyen polipéptidos en los que uno o más de los residuos de aminoácido incluyen un grupo sustituyente.

Normalmente, se considera más del 30% de identidad entre dos polipéptidos como indicación de equivalencia funcional. Polipéptidos funcionalmente equivalentes del primer aspecto de la invención tienen un grado de identidad secuencial con el polipéptido INSP002, o con los fragmentos activo del mismo, de más del 70%. Polipéptidos más preferentes tienen unos grados de identidad superiores al 80%, 90%, 95%, 98%, 99% o más, respectivamente.

Los polipéptidos del primer aspecto de la invención también incluyen fragmentos de los polipéptidos INSP002 y fragmentos de los equivalentes funcionales de los polipéptidos INSP002, siempre y cuando esos fragmentos retengan la actividad de un miembro de la subfamilia de pliegue con nudo cistina DAN.

Como aquí se emplea, el término "fragmento" se refiere a un polipéptido que tiene una secuencia de aminoácido que es igual en parte, pero no en total, a la secuencia de aminoácido de los polipéptidos INSP002 o uno de sus equivalentes funcionales. Los fragmentos deben estar formados al menos por aminoácidos consecutivos n de la secuencia

y, dependiendo de la secuencia particular, n es preferentemente 7 o más (por ejemplo, 8, 10, 12, 14, 16, 18, 20 o más). Pequeños fragmentos pueden formar un determinante antigénico.

Tales fragmentos pueden ser “independientes”, es decir, no son parte o no están fusionados con otros aminoácidos o polipéptidos, o pueden estar formados dentro de un polipéptido más grande del cual forman una parte o región. Cuando están constituidos en un polipéptido más grande, el fragmento de la invención forma más preferentemente una única región continua. Por ejemplo, ciertas realizaciones preferentes se relacionan con un fragmento que tiene una región pre- y/o pro- polipéptido fusionada con la terminal amino del fragmento y/o una región adicional fusionada con la terminal carboxil del fragmento. Sin embargo, varios fragmentos pueden estar constituidos dentro de un único polipéptido más grande.

Los polipéptidos de la presente invención o sus fragmentos inmunogénicos (comprendiendo al menos un determinante antigénico) pueden emplearse para generar ligandos, como los anticuerpos policlonales y monoclonales, que son inmunoespecíficos de los polipéptidos. Tales anticuerpos pueden emplearse para aislar o para identificar clones que expresan los polipéptidos de la invención o para purificar polipéptidos mediante cromatografía de afinidad. Los anticuerpos también pueden emplearse como ayudas terapéuticas o de diagnóstico, entre otras aflicciones, y resultarán obvios para un experto en la técnica.

El término “inmunoespecífico” significa que los anticuerpos tienen sustancialmente mayor afinidad para los polipéptidos de la invención que su afinidad para otros polipéptidos relacionados en la técnica anterior. Como aquí se emplea, el término “anticuerpo” se refiere a moléculas intactas así como a fragmentos de las mismas, como Fab, F(ab')₂, y Fv, que son capaces de enlazarse con el determinante antigénico en cuestión. Por lo tanto, tales anticuerpos pueden enlazarse con polipéptidos del primer aspecto de la invención.

Por “afinidad sustancialmente superior” se entiende que es un aumento apreciable en la afinidad para un polipéptido de la invención en comparación con la afinidad conocida para proteínas segregadas conocidas.

Preferentemente, la afinidad es al menos 1.5 veces, 2 veces, 5 veces, 10 veces, 100 veces, 10³ veces, 10⁴ veces, 10⁵ veces o 10⁶ veces mayor para un polipéptido de la invención que para proteínas segregadas conocidas tales como citoquinas de pliegue con nudo cistina y en particular tales como los miembros de la subfamilia DAN.

Si se desean anticuerpos policlonales, se puede inmunizar un mamífero seleccionado, como un ratón, conejo, cabra o caballo con un polipéptido del primer aspecto de la invención. El polipéptido empleado para inmunizar al animal puede derivarse mediante tecnología recombinante de DNA o puede sintetizarse químicamente. Si se desea, el polipéptido puede conjugarse con una proteína portadora. Los portadores comúnmente empleados con los cuales los polipéptidos pueden enlazarse químicamente incluyen albúmina de suero bovino, tiroglobulina y hemocianina de lapa. A continuación, el polipéptido enlazado se emplea para inmunizar al animal. El suero del animal inmunizado se recoge y se trata de acuerdo con los procedimientos conocidos por ejemplo por cromatografía de inmunoafinidad.

Los anticuerpos monoclonales para los polipéptidos del primer aspecto de la invención también pueden producirse fácilmente por parte de un experto en la técnica. La metodología general para hacer anticuerpos monoclonales usando la tecnología de hibridoma es bien conocida (ver, por ejemplo, Kohler, G. y Milstein, C. *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, 77-96 en *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985).

Los paneles de anticuerpos monoclonales creados contra los polipéptidos del primer aspecto de la invención pueden analizarse para varias propiedades, por ejemplo, para isotipo, epítipo, afinidad, etc. Los anticuerpos monoclonales resultan particularmente útiles en la purificación de polipéptidos individuales contra los que están dirigidos. De manera alternativa, los genes que codifican los anticuerpos monoclonales de interés pueden aislarse para hibridomas, por ejemplo, mediante técnicas PCR conocidas en la técnica, y pueden clonarse y expresarse en vectores apropiados.

Los anticuerpos quiméricos, en los cuales las regiones variables no humanas se unen o fusionan con regiones constantes humanas (ver, por ejemplo, Liu *et al.*, *Proc. Natl. Acad. Sci. USA*, 84, 3439-3443 (1987)), también pueden emplearse.

El anticuerpo puede modificarse para que sea menos inmunogénico en un individuo, por ejemplo, mediante humanización (ver Jones *et al.*, *Nature*, 321, 522 (1986); Verhoeven *et al.*, *Science*, 239, 1534 (1988); Kabat *et al.*, *J. Immunol.*, 147, 1709 (1991); Queen *et al.*, *Proc. Natl. Acad. Sci. USA*, 86, 10029 (1989); Gorman *et al.*, *Proc. Natl. Acad. Sci. USA*, 88, 34181 (1991); y Hodgson *et al.*, *Bio/Technology*, 9, 421 (1991)). El término “anticuerpo humanizado”, como aquí se emplea, se refiere a moléculas de anticuerpo en las cuales los aminoácidos CDR y otros aminoácidos seleccionados en los dominios variables de la cadenas pesada y/o ligera de un anticuerpo perteneciente a un donante no humano se han sustituido en lugar de los aminoácidos equivalentes en un anticuerpo humano. Por lo tanto, el anticuerpo humanizado se parece en gran medida a un anticuerpo humano pero tiene la habilidad de enlace del anticuerpo del donante.

En una alternativa adicional, el anticuerpo puede ser un anticuerpo “biespecífico”, es decir, que es un anticuerpo que tiene dos dominios de enlace antígeno diferentes, estando cada dominio dirigido contra un epítipo diferente.

La tecnología de expresión en fago puede utilizarse para seleccionar genes que codifican anticuerpos con actividades de enlace hacia los polipéptidos de la invención bien a partir de repertorios de V-genes amplificados PCR de linfocitos da partir de humanos analizados para la posesión de anticuerpos relevantes o a partir de bibliotecas simples (McCafferty, J. *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). La afinidad de estos anticuerpos puede también mejorarse mediante reestructuración de cadena (Clackson T. *et al.*, (1991) Nature 352, 624-628).

Los anticuerpos generados mediante las técnicas arriba mencionadas, bien policlonales o monoclonales, tienen una utilidad adicional basada en que pueden utilizarse como reagentes en inmunoensayos, radioinmunoensayos (RIA) o ensayos inmunoabsorbentes enlazados con enzimas (ELISA). En estas aplicaciones, los anticuerpos pueden etiquetarse con un reagente analíticamente detectable como un radioisótopo, una molécula fluorescente o una enzima.

Las moléculas de ácido nucleico preferentes de segundo y tercer aspecto de la invención son aquellas que codifican las secuencias de polipéptido señaladas en SEQ ID NO:6, SEQ ID NO:8 y SEQ ID NO:14 y polipéptidos funcionalmente equivalentes. Estas moléculas de ácido nucleico pueden emplearse en métodos y aplicaciones aquí descritas. Las moléculas de ácido nucleico de la invención comprenden preferentemente al menos nucleótidos consecutivos n a partir de las secuencias aquí descritas donde, dependiendo de la secuencia particular, n es 10 o más (por ejemplo, 12, 14, 15, 18, 20, 25, 30, 35, 40 o más).

Las moléculas de ácido nucleico de la invención también incluyen secuencias que son complementarias con las moléculas de ácido nucleico arriba descritas (por ejemplo, para fines de antisentido o sondeo).

Las moléculas de ácido nucleico de la presente invención pueden tener la forma de RNA, como mRNA, o la forma de DNA, incluyendo por ejemplo cDNA, DNA sintético o DNA genómico. Tales moléculas de ácido nucleico pueden obtenerse mediante clonación, mediante técnicas sintéticas químicas, mediante una combinación de las mismas. Las moléculas de ácido nucleico pueden prepararse, por ejemplo, mediante síntesis química usando técnicas tales como síntesis química con fosoramidita en fase sólida, a partir de bibliotecas genómicas o de cDNA o mediante separación de un organismo. Las moléculas RNA generalmente pueden generarse mediante la transcripción *in vitro* o *in vivo* de secuencias de DNA.

Las moléculas de ácido nucleico pueden ser de doble trenzado o de trenzado sencillo. El DNA de trenzado sencillo puede ser el trenzado codificador, también conocido como trenzado sentido, o puede ser el trenzado no codificador, también conocido como el trenzado anti-sentido.

El término “molécula de ácido nucleico” también incluye análogos de DNA y RNA, como los que contienen los ejes modificados, y ácidos nucleicos péptidos (PNA). El término “PNA”, como aquí se emplea, se refiere a una molécula antisentido o a un agente anti-gen que está constituido por un oligonucleótido de al menos cinco nucleótidos en longitud unido con un eje péptido de residuos de aminoácido, que preferiblemente termina en lisina. La lisina terminal confiere solubilidad a la composición. Los PNAs pueden estar pegilados para extender su periodo de vida en una célula, donde preferentemente se enlazan de manera complementaria con DNA y RNA de trenzado sencillo y frenan el alargamiento de la trascricpción (Nielsen, P. E. *et al.*, (1993) Anticancer Drugs Des. 8:53-63).

Una molécula de ácido nucleico que codifica el polipéptido de la SEQ ID NO:6 puede ser idéntica a la secuencia codificadora de la molécula de ácido nucleico mostrada en la SEQ ID NO:5 (nucleótidos 152 a 721), tal y como se indica en SEQ ID NO:11. Una molécula de ácido nucleico que codifica el polipéptido de la SEQ ID NO:8 puede ser idéntica a la secuencia codificadora de la molécula de ácido nucleico mostrada en la SEQ ID NO:5 (nucleótidos 218 a 721), tal y como se indica en SEQ ID NO:9. Una molécula de ácido nucleico que codifica el polipéptido de la SEQ ID NO: 14 puede ser idéntica a la secuencia codificadora de la molécula de ácido nucleico mostrada en la SEQ ID NO: 13 (nucleótidos 69 a 719), tal y como se indica en SEQ ID NO: 15.

Estas moléculas también pueden tener una secuencia diferente que, como resultado de la degeneración del código genético, codifica un polipéptido de la SEQ ID NO: 6, SEQ ID NO: 8, o SEQ ID NO: 14. Tales moléculas de ácido nucleico que codifican el polipéptido de la SEQ ID NO: 6, SEQ ID NO: 8, o SEQ ID NO: 14 pueden incluir, aunque no limitarse a, la secuencia codificadora para el polipéptido maduro por sí mismo; la secuencia codificadora para el polipéptido maduro y las secuencias codificadoras adicionales, como las que codifican una secuencia líder o secretora, como una secuencia pro-, pre- o prepro- polipéptido; la secuencia codificadora del polipéptido maduro, con o sin las secuencias codificadoras adicionales anteriormente mencionadas, junto con secuencias adicionales no codificadoras, incluyendo las secuencias 5' y 3' no codificadoras, como las secuencias transcritas y no traducidas que juegan un papel en la trascricpción (incluyendo señales de terminación), enlace de ribosoma y estabilidad mRNA. Las moléculas de ácido nucleico también pueden incluir secuencias adicionales que codifican aminoácidos adicionales, como aquellas que proporcionan funcionalidades adicionales.

Las moléculas de ácido nucleico del segundo y tercer aspecto de la invención a su vez pueden codificar los fragmentos de los equivalentes funcionales de los polipéptidos y fragmentos del primer aspecto de la invención. Tal molécula de ácido nucleico puede ser una variante que ocurra de manera natural tal como una variante alélica que ocurre de manera natural, o la molécula puede ser una variante que no se conozca que ocurra de manera natural. Tales variantes que no ocurren de manera natural de la molécula de ácido nucleico pueden realizarse mediante técnicas de mutagénesis, incluyendo aquellas aplicadas a las moléculas de ácido nucleico, células y organismos.

Entre las variantes en este aspecto, existen variantes que se diferencian de las moléculas de ácido nucleico anteriormente mencionadas por medio de sustituciones, eliminaciones o inserciones de nucleótidos. Las sustituciones, eliminaciones e inserciones pueden implicar uno o más nucleótidos. Las variantes pueden alterarse regiones codificadora o no codificadoras o en ambas. Las alteraciones en las regiones codificadoras pueden producir sustituciones, eliminaciones o inserciones conservadoras o no conservadoras.

Las moléculas de ácido nucleico de la invención también pueden crearse mediante ingeniería empleando métodos generalmente conocidos en la técnica, por una variedad de razones, incluyendo la modificación de la clonación, el proceso, y/o la expresión del producto gen (el polipéptido). El cambio de DNA mediante fragmentación arbitraria y re-montaje de PCR de fragmentos de gen y oligonucleótidos sintéticos están incluidos como técnicas que pueden emplearse para trabajar mediante ingeniería con secuencias de nucleótidos. La mutagénesis dirigida al punto puede utilizarse para insertar nuevos puntos de restricción, alterar los patrones de glicosilación, cambiar la preferencia del codón, producir variantes en los empalmes o uniones, introducir mutaciones y etcétera.

Las moléculas de ácido nucleico que codifican un polipéptido del primer aspecto de la invención puede ligarse a una secuencia heteróloga de modo que la molécula de ácido nucleico codifique una proteína de fusión. Tales moléculas de ácido nucleico se incluyen dentro del segundo o del tercer aspecto de la invención. Por ejemplo, para analizar bibliotecas de péptidos para inhibidores de la actividad del polipéptido, puede resultar útil expresar, usando dicha molécula de ácido nucleico combinada, una proteína de fusión que pueda reconocerse mediante un anticuerpo disponible en el mercado. Una proteína de fusión también puede crearse para que contenga un punto de división localizado entre la secuencia de polipéptido de la invención y la secuencia de una proteína heteróloga de modo que el polipéptido pueda dividirse y purificarse a partir de la proteína heteróloga.

Las moléculas de ácido nucleico de la invención también incluyen moléculas antisentido que son parcialmente complementarias con las moléculas de ácido nucleico que codifican polipéptidos de la presente invención y que por lo tanto se hibridizan con las moléculas de ácido nucleico codificadoras (hibridización). Tales moléculas antisentido, como los oligonucleótidos, pueden diseñarse para reconocer, enlazarlos específicamente y prevenir la transcripción de un ácido nucleico objetivo que codifica un polipéptido de la invención, como se conocerá por aquellos expertos en la técnica (ver, por ejemplo, Cohen, J. S., *Trends in Pharm. Sci.*, 10, 435 (1989), Okano, J. *Neurochem.* 56, 560 (1990); O'Connor, J. *Neurochem* 56, 560 (1991); Lee *et al.*, *Nucleic Acids Res* 6, 3073 (1979); Cooney *et al.*, *Science* 241, 456 (1988); Dervan *et al.*, *Science* 251, 1360 (1991).

El término "hibridización" tal y como aquí se emplea hace referencia a dos moléculas de ácido nucleico entre las cuales existe un enlace de hidrógeno. Normalmente, una molécula estará fijada a un soporte sólido y la otra estará libre en una solución. Por lo tanto, las dos moléculas pueden estar en contacto entre sí bajo condiciones que favorecen el enlace de hidrógeno. Factores que afectan este enlace incluyen: el tipo de volumen del disolvente; la temperatura de reacción; el tiempo de hibridización; la agitación; agentes que bloquean la unión no específica de la molécula en fase líquida con el soporte sólido (reagente de Denhardt o BLOTTO); la concentración molecular; el uso de compuestos para aumentar la velocidad de asociación molecular (sulfato de dextrano o glicol de polietileno); y el rigor de la condiciones de lavado tras la hibridización (ver Sambrook *et al.*, [*supra*]).

La inhibición de la hibridización de una molécula completamente complementaria con una molécula objetivo puede examinarse empleando un ensayo de hibridización, tal y como se conoce en la técnica (ver, por ejemplo Sambrook *et al.*, [*supra*]). Una molécula sustancialmente homóloga competirá a continuación e inhibirá el enlace de una molécula completamente homóloga con la molécula objetivo bajo varias condiciones de rigor, tal y como se describe en Whal, G. M., y S. L. Berger (1987; *Methods Enzymol.*, 152:399-407) y Kimmel, A. R., (1987; *Methods Enzymol.* 152:507-511).

"Rigor" hace referencia a las condiciones en una reacción de hibridización que favorecen la asociación de moléculas muy similares sobre asociación de moléculas que difieren. Condiciones de hibridización de elevado rigor se definen como la incubación nocturna a 42°C en una solución que comprende 50% formamida, 5XSSC (150mM NaCl, 15mM citrato de trisodio), 50mM fosfato sódico (pH 7.6), 5x solución Denhardt, 10% sulfato de dextrano, y 20 microgramos/ml desnaturalizados, DNA de esperma de salmón cortado, seguido del lavado de los filtros en 0.1X SSC a aproximadamente 65°C. Las condiciones de bajo rigor implican que la reacción de hibridización se lleve a cabo a 35°C (ver Sambrook *et al.*, [*supra*]). Preferiblemente, las condiciones empleadas para la hibridización son aquellas pertenecientes al rigor elevado.

Las realizaciones preferentes de este aspecto de la invención son las moléculas de ácido nucleico que son al menos 70% idénticas en su longitud completa con una molécula de ácido nucleico que codifica los polipéptidos ISNP002 (SEQ ID NO:6, SEQ ID NO:8 y SEQ ID NO:14) y las moléculas de ácido nucleico que son sustancialmente complementarias con tales moléculas de ácido nucleico. Preferentemente, una molécula de ácido nucleico de acuerdo con este aspecto de la invención comprende una región que es al menos 80% idéntica en su completa longitud con la secuencia codificadora para la SEQ ID NO:4 dada en la SEQ ID NO:3, las secuencia codificadora para las SEQ ID NO:4 y SEQ ID NO:8 dadas en SEQ ID NO:5, SEQ ID NO:9 y SEQ ID NO:11, o las secuencias codificadoras para SEQ ID NO:14 dada en SEQ ID NO:13 y SEQ ID NO:15, o es una molécula de ácido nucleico que es complementaria con las mencionadas. En este aspecto, son preferibles las moléculas de ácido nucleico que son al menos 90%, preferiblemente al menos 95%, más preferiblemente 98% o 99% idénticas en su completa longitud con las mismas. Las realizaciones preferentes en este aspecto son las moléculas de ácido nucleico que codifican polipéptidos que sustancialmente retienen la misma función o actividad biológica que los polipéptidos ISNP002.

La invención también proporciona un proceso para detectar una molécula de ácido nucleico de la invención, que comprende los pasos de: (a) contactar una sonda nucleica de acuerdo con la invención con una muestra biológica bajo condiciones hibridizadoras para formar elementos dúplex; y (b) detectar cualquiera de estos elementos dúplex que se formen.

Como se describe adicionalmente a continuación con ensayos que pueden utilizarse de acuerdo con la invención, una molécula de ácido nucleico tal y como se ha descrito anteriormente puede emplearse como una sonda hibridizadora para RNA, cDNA o DNA genómico, con el fin de aislar cDNAs de completa longitud y clones genómicos codificadores de polipéptidos INSP002 y para aislar cDNA y clones genómicos de genes homólogos o heterólogos que tienen una elevada similitud secuencial con el gen codificador de este polipéptido.

En este aspecto, las siguientes técnicas, entre otras conocidas en la técnica, pueden utilizarse y se comentan a continuación con fines ilustrativos. Métodos para secuenciar DNA y análisis son bien conocidos y generalmente se encuentran disponibles en la técnica y de hecho pueden emplearse para practicar muchas de las realizaciones de la invención aquí descritas. Tales métodos pueden emplear enzimas como el fragmento Klenow de polimerasa DNA I, Secuensa (US Biochemical Corp, Cleveland, OH), polimerasa Taq (Perkin Elmer), polimerasa T7 termoestable (Amersham, Chicago, IL), o combinaciones de polimerasas y exonucleasas correctoras de pruebas como las encontradas en la Amplificación ELONGASE lanzadas al mercado por Gibco/BRL (Gaithersburg, MD). Preferentemente, el proceso secuencial puede automatizarse empleando máquinas como Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Ciclo Térmico Peltier (PTC200; MJ Research, Watertown, MA) y Catalizador ABI y Secuenciadores de DNA 373 y 377 (Perkin Elmer).

Un método para aislar una molécula de ácido nucleico codificadora de un polipéptido con una función equivalente con los polipéptidos INSP0002 es el consistente en sondear una biblioteca genómica o de cDNA con una sonda natural o artificialmente diseñada usando procedimientos estándar que se reconocen en la técnica (ver, por ejemplo, "Current Protocols in Molecular Biology", Ausubel *et al.*, (eds.), Green Publishing Association y John Wiley Interscience, New York, 1989, 1992). Las sondas consisten al menos en 15, preferiblemente al menos 30, y más preferiblemente al menos 50 bases contiguas que corresponden con o que son complementarias con secuencias de ácido nucleico a partir del gen codificador apropiado (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13 o SEQ ID NO:15), son sondas particularmente útiles. Tales sondas pueden etiquetarse con un reagente analíticamente detectable para facilitar su identificación. Reagentes útiles incluyen, aunque no se limitan a, radioisótopos, tintes fluorescentes y enzimas que son capaces de catalizar la formación de un producto detectable. Empleando tales sondas, el experto en la técnica será capaz de aislar copias complementarias de polinucleótidos DNA genómico, cDNA o RNA que codifican proteínas de interés a partir de fuentes humanas, mamíferas u otras fuentes animales y analizando dichas fuentes en busca de secuencias relacionadas, por ejemplo, miembros adicionales de la familia, tipo y/o subtipo.

En muchos casos, las secuencias aisladas de cDNA estarán incompletas, y la región codificadora del polipéptido será más corta, normalmente en el 5'. Se encuentran disponibles varios métodos para obtener cDNAs de longitud completa, o para extender cDNAs cortos. Tales secuencias pueden extenderse utilizando una secuencia parcial de nucleótido y mediante el empleo de varios métodos conocidos en la técnica para detectar secuencias dirigidas en dirección descendente como elementos impulsores y reguladores. Por ejemplo, un método que puede emplearse se abas en el método de Amplificación Rápida de Extremos de cDNA (RACE; ver, por ejemplo, Frohman *et al.*, PNAS USA 85, 8998-9002, 1988). Modificaciones recientes de esta técnica, ejemplificadas por la tecnología MarathonTM (Clontech Laboratorios Inc.), por ejemplo, han simplificado de manera significativa la búsqueda de cDNAs más largos. Una técnica ligeramente diferente, denominada PCR "en punto de restricción", emplea imprimadores universales para recuperar secuencias de ácido nucleico desconocidas adyacentes a dicho punto (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). PCR inverso también puede emplearse para amplificar o para extender secuencias empleando imprimadores divergentes en base a una región conocida (Triglia, T. *et al.*, (1988) Nucleic Acids Res. 16:8186). Otro método que puede emplearse es PCR de captura que implica la amplificación PCR de fragmentos de DNA adyacentes a la secuencia conocida en DNA de cromosoma humano o de levadura artificial (Lagerstrom, M. *et al* (1991) PCR Methods Applic., 1, 111-119). Otro método que puede emplearse para recuperar secuencias desconocidas es la de Parker, J. D. *et al.*, (1991); Nucleic Acids Res. 19:3055-3060). De manera adicional, se puede usar PCR, imprimadores anidados, y bibliotecas PromoterFinderTM para recorrer DNA genómico (Clontech, Palo Alto, CA). Este proceso evita la necesidad de analizar bibliotecas y es útil a la hora de buscar uniones intrón/exón.

Cuando se analizan cDNAs de longitud completa, es preferible emplear bibliotecas que han sido seleccionadas en base a su tamaño para incluir cADNs más grandes. Así mismo, son preferibles bibliotecas preparadas arbitrariamente, en el sentido de que contendrán más secuencias que las contenidas en las regiones 5' de los genes. El uso de una biblioteca arbitrariamente preparada puede resultar especialmente preferible en situaciones en las que una biblioteca oligo d(T) no produce cADNs de longitud completa. Las bibliotecas genómicas pueden ser útiles para la extensión de la secuencia en regiones reguladoras no transcritas 5'.

En una realización de la invención, las moléculas de ácido nucleico de la presente invención pueden emplearse para localización de cromosomas. En este técnica, una molécula de ácido nucleico se dirige específicamente a, o puede hibridarse con, una localización particular en un cromosoma humano individual. El mapeo de las secuencias relevantes a los cromosomas de acuerdo con la presente invención es un paso importante en la correlación confirmatoria de aquellas secuencias con la enfermedad asociada al gen. Una vez que la secuencia se ha mapeado hasta conseguir

una localización cromosómica precisa, la posición física de la secuencia en los cromosomas puede relacionarse con los datos del mapa genético. Estos datos se encuentran, por ejemplo, en V. McKusick, Mendelian Inheritance in Man (disponible en línea a través de la Biblioteca Welch Medical de la Universidad Johns Hopkins). La relación entre los genes y las enfermedades que se han mapeado con la misma región cromosómica se identifican a continuación por medio de un análisis de conexión o enlace (herencia conjunta de genes físicamente adyacentes). Esto proporciona información válida para investigadores que buscan genes de enfermedades usando clonación posicional u otras técnicas de descubrimiento genético. Una vez que la enfermedad o síndrome se ha localizado de un modo rudimentario mediante conexión genética con una región genómica particular, cualquiera de las secuencias que mapean tal área pueden representar genes asociados o reguladores para posteriores investigaciones. La molécula de ácido nucleico puede también emplearse para detectar diferencias en la localización cromosómica debido a la traslocalización, inversión, etc, entre individuos normales, portadores o afectados.

Las moléculas de ácido nucleico de la presente invención también son valiosas para localización de tejido. Tales técnicas permiten la determinación de patrones de expresión del polipéptido en tejidos mediante la detección de los mRNAs que los codifica. Estas técnicas incluyen técnicas de hibridización *in situ* y técnicas de amplificación de nucleótidos, tales como PCR. Los resultados de estos estudios proporcionan una indicación de las funciones normales del polipéptido en el organismo. Además, estudios comparativos del patrón de expresión normal de mRNAs con los de mRNAs codificados por un gen mutante proporcionan perspectivas en el papel de polipéptidos mutantes en la enfermedad. Tal expresión inapropiada puede ser de naturaleza temporal, espacial o cuantitativa.

Los enfoques de genes silenciadores también pueden considerarse para bajo-regular la expresión endógena de un gen codificador de un polipéptido de la invención. La interferencia RNA (RNAi) (Elbashir, SM *et al.*, Nature 2001, 411, 494-498) es un método de gen silenciador post-transcripcional específico de secuencia que puede emplearse. Oligonucleótidos cortos dsRNA se sintetizan *in vitro* y se introducen en una célula. El enlace específico secuencial de estos oligonucleótidos dsRNA provoca la degradación de mRNA objetivo, reduciendo o eliminando la expresión de proteína objetivo.

La eficacia de los enfoques silenciadores de gen establecidos anteriormente pueden analizarse a través de la medida de la expresión del polipéptido (por ejemplo, mediante transferencia de tipo Western), y en el nivel RNA se emplean metodologías basadas en Taqman.

Los vectores de la presente invención comprenden moléculas de ácido nucleico de la invención y pueden ser vectores de clonación o expresión. Las células huésped de la invención, que pueden transformarse, transfectarse o transducido con los vectores de la invención pueden ser procarióticos o eucarióticos.

Los polipéptidos de la invención pueden prepararse en forma recombinante mediante expresión de sus moléculas de ácido nucleico codificadoras en vectores que se encuentran contenidos en una célula huésped. Tales métodos de expresión son bien conocidos por aquellos expertos en la técnica y muchos se describen con detalle por parte de Sambrook *et al* (*supra*) y Fernandez & Hoeffler (1998, eds "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New Cork, Sydney, Tokyo, Toronto).

En general, puede emplearse cualquier sistema o vector que sea adecuado para mantener, propagar o expresar moléculas de ácido nucleico para producir polipéptidos en el huésped requerido. La secuencia de nucleótido apropiada puede insertarse en un sistema de expresión mediante cualquiera de las técnicas bien conocidas y rutinarias, por ejemplo, las descritas por Sambrook *et al.* (*supra*). En términos generales, el gen codificador puede situarse bajo el control de un elemento controlador como un impulsor, punto de enlace de ribosoma (para expresión bacteriana) y, opcionalmente, un operador, de modo que la secuencia DNA codificadora del polipéptido deseado se transcribe en RNA en la célula huésped transformada.

Los ejemplos de sistemas de expresión adecuados, por ejemplo, los sistemas cromosómicos, episomales y derivados de virus, por ejemplo, vectores derivados de: plásmidos bacterianos, bacteriófago, transposones, episomas de levadura, elementos de inserción, elementos cromosómicos de levadura, virus y baculovirus, virus papota como SV40, virus correspondientes a vacunas, adenovirus, virus de viruela aviar, virus pseudorabies y retrovirus, o combinaciones de los mismos, como los derivados de elementos plásmidos o bacteriófagos genéticos, incluyendo cósmidos y fagémidos. Los cromosomas artificiales humanos (HACs) pueden también emplearse para enviar fragmentos más grandes de DNA que pueden estar contenidos y expresarse en un plásmido.

En particular, los sistemas de expresión adecuados incluyen microorganismos como bacterias transformadas con bacteriófago recombinante, vectores de expresión DNA plásmidos o cósmicos; levadura transformada con vectores de expresión de levadura; sistemas celulares de insectos infectados con vectores de expresión víricos (por ejemplo, baculovirus); sistemas celulares de plantas transformados con vectores de expresión víricos (por ejemplo, virus mosaicos de coliflor, CaMV; virus mosaico de tabaco, TMV) o con vectores de expresión bacterianos (por ejemplo, plásmidos Ti o Pbr322); o sistemas celulares animales. Los sistemas de translación libres de células también pueden emplearse para producir los polipéptidos de la invención.

La introducción de moléculas de ácido nucleico codificadoras de un polipéptido de la presente invención en las células huésped puede efectuarse mediante métodos descritos en muchos manuales estándares de laboratorio, como Davis *et al.*, Basic Methods in Molecular Biology (1986) y Sambrook *et al.*, [*supra*]. Métodos particularmente adecua-

dos incluyen transfección de fosfato cálcico, transfección mediada por DEAE-dextrano, transvección, microinyección, transfección mediada por lípido catiónico, electroporación, transducción, carga de raspado, introducción balística o infección (ver Sambrook *et al.*, 1989 [*supra*]; Ausubel *et al.*, 1991 [*supra*]; Spector, Goldman & Leinwald, 1998). En células eucarióticas, los sistemas de expresión pueden ser bien pasajeros (por ejemplo, episomales) o permanentes (integración cromosómica) de acuerdo con las necesidades del sistema.

La molécula de ácido nucleico codificadora puede o no incluir una secuencia codificadora de una secuencia de control, como un péptido señalizador o una secuencia líder, tal y como se desee, por ejemplo para la secreción del polipéptido trasladado al lumen del retículo endoplásmico, en un espacio periplásmico en el medio extracelular. Estas señales pueden ser endógenas al polipéptido o pueden ser señales heterólogas. Las secuencias líderes pueden retirarse mediante el huésped bacteriano en un proceso post-translacional.

Además de las secuencias de control, puede resultar deseable para añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido relativo al crecimiento de la célula huésped. Ejemplos de secuencias reguladoras son aquellas que provocan que la expresión de un gen aumente o disminuya en respuesta a un estímulo químico o físico, incluyendo la presencia de un compuesto regulador o a varias condiciones de temperatura o metabólicas. Las secuencias reguladoras son aquellas regiones no traducidas del vector, como amplificadores, impulsores y regiones no traducidas 5' y 3'. Estas regiones interactúan con las proteínas celulares huésped para llevar a cabo la transcripción o translación. Tales secuencias reguladoras pueden variar en su fuerza y especificidad. Dependiendo del sistema de vector y del huésped utilizado, cualquier número de elementos adecuados de transcripción y traslación, incluyendo impulsores constitutivos e inducibles, pueden emplearse. Por ejemplo, cuando se clonan en sistemas bacteriano, los promotores inducibles tales como promotor híbrido lacZ del fagémido Bluescript (Stratagene, Lajilla, CA) o plásmido pSport1TM (Gibco BRL) y similares pueden usarse. El promotor baculovirus polyhedrin puede emplearse en células de insectos. Los promotores o impulsores derivados de los genomas de células de plantas (por ejemplo, shock de calor, RUBISCO y genes de proteínas de almacenaje) o a partir de virus de plantas (por ejemplo, promotores virales o secuencias líderes) pueden clonarse en el vector. En los sistemas celulares mamíferos, los promotores de genes mamíferos o de virus mamíferos son preferibles. Es necesario generar una línea celular que contenga múltiples copias de la secuencia, pueden emplearse vectores basados en SV40 o EBV con un marcador seleccionador apropiado.

Un vector de expresión se construye de modo que la secuencia codificadora del ácido nucleico particular se localice en el vector con las secuencias reguladoras apropiadas, estando el posicionamiento y la orientación de la secuencia codificadora con respecto a las secuencias reguladoras de tal modo que la secuencia codificadora se transcriba bajo el "control" de las secuencias reguladoras, es decir, la polimerasa RNA que se enlaza con la molécula DNA en las secuencias de control transcribe la secuencia codificadora. En algunos casos puede resultar necesario modificar la secuencia de modo que pueda unirse a las secuencias de control con la orientación apropiada, es decir, para mantener el marco de lectura.

Las secuencias de control y otras secuencias reguladoras pueden estar ligadas con la secuencia codificadora del ácido nucleico antes de la inserción en el vector. De manera alternativa, la secuencia codificadora puede clonarse directamente en un vector de expresión que ya contenga las secuencias de control y un punto de restricción apropiado.

A largo plazo, la producción de un polipéptido recombinante, es preferible la expresión estable. Por ejemplo, las líneas celulares que establemente expresan el polipéptido de interés pueden transformarse empleando vectores de expresión que pueden contener orígenes virales de réplica y/o elementos de expresión endógenos y un gen marcador elegible sobre el mismo vector o en un vector separado. Tras la introducción del vector, se deja que las células crezcan durante 1-2 días en un medio enriquecido ante de cambiarlas al medio selectivo. El fin del marcador elegible es conferir resistencia a la selección, y su presencia permite el crecimiento y la recuperación celular que con éxito expresan las secuencias introducidas. Los clones resistentes de las células establemente transformadas pueden proliferarse empleando técnicas de cultivo adecuadas para el tipo celular.

Las líneas celulares mamíferas disponibles como huéspedes para expresión son bien conocidas en la técnica e incluyen cualquiera de las líneas celulares inmortalizadas disponible en la Colección Americana de Cultivos Tipo (ATCC) incluyendo, aunque no limitando, ovario de hámster chino (CHO), HeLa, riñón de bebé hámster (BHK), riñón de mono (COS), C127, 3T3, BHK, HEK 293, melanoma Bowes y células de carcinoma hepatocelular humano (por ejemplo Hep G2) y un número de otras líneas celulares.

En el sistema baculovirus, los materiales de los sistemas de expresión celular baculovirus/insecto están comercialmente disponibles en forma de kit en, entre otros, Invitrogen, San Diego CA (el kit "MaxBac"). Estas técnicas son generalmente conocidas por aquellos expertos en la técnica y se describen por completo en Summers and Smith, Texas Agricultural Experiment Station Bulletin n° 1555 (1987). En particular, las células huésped adecuadas para uso en este sistema incluyen células de insectos como las células *Drosophila* S2 y *Spodoptera* Sf9.

Existen muchos sistemas de expresión relativos al cultivo de células de plantas y sistemas de expresión genética de plantas completas conocidos en la técnica. Ejemplos de sistemas de expresión genética celular de planta adecuados incluyen aquellos descritos en US 5,693,506; US 5,659,122; y US 5,608,143. Ejemplos adicionales de expresión genética en cultivos de células de plantas se han descrito por Zenk, *Phytochemistry* 30, 3861-3863 (1991).

En particular, se pueden utilizar todas las plantas de las cuales los protoplastos se pueden aislar y cultivar para dar plantas completas regeneradas, de modo que se recuperan plantas completas que contienen el gen transferido. Prácticamente todas las plantas pueden regenerarse a partir de células o tejidos cultivados, incluyendo aunque no limitando todas las especies principales de caña de azúcar, remolacha azucarera, algodón, fruta y otros árboles, legumbres y verduras.

Ejemplos de células huésped particularmente preferentes incluyen *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* y células de *Bacillus subtilis*.

Ejemplos de células huésped particularmente preferentes par expresión de hongos incluyen células de levadura (por ejemplo, *S. cerevisiae*) y células *Aspergillus*.

Cualquier número de sistemas selectivos conocidos en la técnica puede emplearse para recuperar las líneas celulares transformadas. Entre los ejemplos se incluyen virus herpes simple timidina quinasa (Wigler, M. *et al.* (1977) Cell 11:223-32) y genes de adenina fosforibosiltransferasa (Lowy, I., *et al.*, (1980) Cell 22:817-23) que pueden emplearse en células tk- o aprt±, respectivamente.

Así mismo, la resistencia de antimetabolito, antibiótico o herbicida puede emplearse como la base para la selección; por ejemplo, reductasa dihidrofolato (DHFR) que aporta resistencia a metotrexato (Wigler, M. *et al.* (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, que da resistencia a la neomicina de aminoglicosidas y G-418 (Colbere-Garapin, F. *et al.* (1981) J. Mol. Biol. 150:1-14) y als o pat, que aportan resistencia a clorosulfurón y acetiltransferasa fosfinotricina, respectivamente. Se han descrito genes adicionales elegibles, cuyos ejemplos resultarán evidentes para aquellos expertos en la técnica.

A pesar de que la presencia o ausencia de expresión de gen marcador sugiere que el gen de interés también se encuentra presente, su presencia o expresión pueden necesitar confirmación. Por ejemplo, si la secuencia relevante se inserta en una secuencia de gen marcador, las células transformadas que contienen las secuencias apropiadas pueden identificarse mediante la ausencia de función de gen marcador. De manera alternativa, un gen marcador puede colocarse en conjunto con una secuencia codificando un polipéptido de la invención bajo el control de un único promotor. La expresión de un gen marcado en respuesta a la inducción o selección normalmente también indica la expresión del gen en conjunto.

De modo alternativo, las células huésped que contienen una secuencia de ácido nucleico que codifica un polipéptido de la invención y que expresan dicho polipéptido pueden identificarse mediante un conjunto de procedimientos conocidos por aquellos expertos en la técnica. Estos procedimientos incluyen, aunque no se limitan a, hibridaciones DNA-DNA o DNA-RNA y bioensayos con proteínas, por ejemplo, clasificación celular activada por fluorescencia (FACS) o técnicas de inmunoensayo (como en ensayo inmunoabsorbente enlazado con enzima [ELISA] y radioinmunoensayo [RIA]), que incluyen membrana, solución, o tecnologías basadas en chips para la detección y/o cuantificación de ácido nucleico o proteínas (ver Hampton R. *et al.* (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul, MN) y Maddox, D. E. *et al.*, (1983) J. Exp. Med, 158, 1211-1216).

Una amplia variedad de técnicas con etiquetas y conjugaciones son conocidas por aquellos expertos en la técnica y pueden emplearse en varios ensayos con ácido nucleico y aminoácidos. Los medios para producir hibridación etiquetada o sondas de PCR para detectar secuencias relacionadas con moléculas de ácido nucleico codificadoras de polipéptidos de la presente invención incluyen el etiquetaje oligo, la translación con hendidura (nick), el etiquetaje final o la amplificación PCR empleando polinucleótido etiquetado.

De manera alternativa, las secuencias codificadoras del polipéptido de la invención pueden clonarse en un vector para la producción de un sonda mRNA. Tales vectores son conocidos en la técnica, están disponibles en el mercado y pueden emplearse para sintetizar sondas RNA *in vitro* mediante la adición de una polimerasa adecuada RNA como T7, T3 o SP6 y nucleótidos etiquetados. Estos procedimientos pueden llevarse a cabo empleando una variedad de kits disponibles en el mercado (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); y U.S. Biochemical Corp., (Cleveland, OH)).

Moléculas reporteras adecuadas o etiquetas, que pueden emplearse para facilitar la detección, incluyen radionúclidos, enzimas y agentes fluorescentes, quimioluminiscentes o cromogénicos así como sustratos, cofactores, inhibidores, partículas magnéticas y similares.

Las moléculas de ácido nucleico de acuerdo con la presente invención pueden también emplearse para crear animales transgénicos, en particular animales roedores. Tales animales transgénicos forman un aspecto adicional de la presente invención. Esto se puede realizar localmente mediante la modificación de células somáticas, o mediante terapia de línea de germen para incorporar modificaciones hereditables. Dichos animales transgénicos pueden resultar particularmente útiles en la generación de modelos de animales para moléculas farmacéuticas efectivas como moduladores de los polipéptidos de la presente invención.

El polipéptido puede recuperarse y purificarse a partir de cultivos celulares recombinantes y por medio de métodos bien conocidos incluyendo sulfato de amonio o precipitación de etanol, extracción ácida, cromatografía de intercambio de anión o catión, cromatografía de fosfocelulosa, cromatografía de interacción hidrofóbica, cromatografía de afini-

dad, cromatografía de hidroxilapatita y cromatografía de lectina. La cromatografía líquida de elevada actuación resulta particularmente útil para la purificación. Técnicas bien conocidas para replegar proteínas pueden emplearse para regenerar una conformación activa cuando el polipéptido se desnaturaliza durante el aislamiento y la purificación.

Construcciones de vectores especializados pueden también emplearse para facilitar la purificación de proteínas, tal y como se desee, uniendo secuencias que codifican los polipéptidos de la invención con una secuencia de nucleótido codificadora de un dominio polipéptido que facilitará la purificación de proteínas solubles. Ejemplos de tales dominios facilitadores de la purificación incluyen péptidos quelantes de metal como módulos histidina-triptofano que permiten la purificación en metales inmovilizados, los dominios de proteína A que permiten la purificación en inmunoglobulina inmovilizada, y el dominio empleado en el sistema de purificación FLAGS extensión/afinidad (Immunex Corp., Seattle, WA). La inclusión de secuencias conectoras rajables como aquellas específicas para Factor XA o enteroquinasa (Invitrogen, San Diego, CA) entre el dominio de purificación y el polipéptido de la invención puede emplearse para facilitar la purificación. Dicho vector de expresión proporciona una expresión de la proteína de fusión que contiene el polipéptido de la invención fusionados con varios residuos de histidina precedentes a un punto de división tioredoxina o enteroquinasa. Los residuos de histidina facilitan la purificación mediante IMAC (cromatografía de afinidad de ión metal inmovilizado tal y como se describe en Porath, J. *et al.* (1992), Prot. Exp. Purif. 3: 263-281) mientras que el punto de división de tioredoxina o enteroquinasa proporciona un medio para purificar el polipéptido a partir de la proteína de fusión. Una descripción de vectores que contienen proteínas de fusión está incluida en Kroll, D. J. *et al.*, (1993); DNA Cell Biol. 12:441-453).

Si el polipéptido se va a expresar para uso en ensayos de análisis, generalmente es preferible que se produzca en una superficie de la célula huésped en la cual se expresa. De esta forma, las células huésped pueden cultivarse antes de su uso en el ensayo de análisis, por ejemplo empleando técnicas tales como clasificación celular activada por fluorescencia (FACS) o técnicas de inmunoafinidad. Si el polipéptido se segrega en el medio, el medio puede recuperarse con el fin de recuperar y purificar el polipéptido expresado. Si el polipéptido se produce intracelularmente, en primer lugar las células deben lisarse antes de que el polipéptido se recupere.

El polipéptido de la invención puede usarse para analizar bibliotecas de compuestos en una gran variedad de técnicas de análisis de fármacos. Tales compuestos pueden activar (agonista) o inhibir (antagonista) el nivel de expresión del gen o de la actividad del polipéptido de la invención y formar un aspecto adicional de la presente invención. Los compuestos preferentes son efectivos para alterar la expresión de un gen natural que codifica un polipéptido del primer aspecto de la invención o para regular la actividad de un polipéptido del primer aspecto de la invención.

Los compuestos agonistas o antagonistas pueden aislarse de, por ejemplo, células, preparaciones libres de células, bibliotecas químicas o mezclas de productos naturales. Estos agonistas o antagonistas pueden ser sustratos naturales o modificados, ligandos, enzimas, receptores o miméticos estructurales o funcionales. Para una correcta revisión de tales técnicas analíticas ver Coligan *et al.*, Current Protocols in Immunology 1(2): Capítulo 5 (1991).

Los compuestos que tienen más posibilidades de ser buenos antagonistas son las moléculas que se enlazan con el polipéptido de la invención sin provocar efectos biológicos del polipéptido tras el enlace con el mismo. Los antagonistas potenciales incluyen pequeñas moléculas orgánicas, péptidos, polipéptidos y anticuerpos que se enlazan con el polipéptido de la invención y que por lo tanto inhiben o extinguen su actividad. De este modo, el enlace del polipéptido con moléculas de enlace celular normal puede inhibirse, de modo que se prevenga la actividad biológica normal del polipéptido.

El polipéptido de la invención que se emplea en dicha técnica de análisis puede estar libre en una solución, fijado a un soporte sólido, puede nacer en una superficie celular o puede estar localizado intracelularmente. En general, dichos procedimientos analíticos pueden implicar el uso de células apropiadas o membranas celulares que expresen el polipéptido y que estén en contacto con al menos un compuesto para observar el enlace, la estimulación o la inhibición de una respuesta funcional. La respuesta funcional de las células en contacto con el compuesto de la prueba se compara a continuación con las células control que no han estado en contacto con el compuesto de la prueba. Dicho ensayo puede calcular si el compuesto de la prueba da como resultado una señal generada por la activación del polipéptido, empleando un sistema apropiado de detección. Los inhibidores de la activación generalmente se analizan en la presencia de un agonista conocido y se observa el efecto sobre la activación por el agonista en la presencia del compuesto de la prueba.

Un método preferente para identificar un compuesto agonista o antagonista de un polipéptido de la presente invención comprende:

- (a) poner en contacto una célula que expresa en la superficie de la misma el polipéptido de acuerdo con el primer aspecto de la invención, estando el polipéptido asociado con un segundo componente capaz de proporcionar una señal detectable en respuesta al enlace de un compuesto con el polipéptido, con un compuesto a ser analizado bajo condiciones que permitan el enlace con el polipéptido; y
- (b) determinar si el compuesto se enlaza y activa o inhibe el polipéptido midiendo el nivel de una señal generada a partir de la interacción del compuesto con el polipéptido.

Un método más preferente para la identificación de un agonista o antagonista de un polipéptido de la invención comprende:

- (a) poner en contacto una célula que expresa en la superficie de la misma el polipéptido, estando el polipéptido asociado con un segundo componente capaz de proporcionar una señal detectable en respuesta al enlace de un compuesto con el polipéptido, con un compuesto a ser analizado bajo condiciones que permitan el enlace con el polipéptido; y
- (b) determinar si el compuesto se enlaza y activa o inhibe el polipéptido comparando el nivel de una señal generada a partir de la interacción del compuesto con el polipéptido con el nivel de una señal en la ausencia del compuesto.

En realizaciones adicionales preferentes, los métodos generales que se acaban de describir pueden comprender la identificación de agonista o antagonista en la presencia de un ligando etiquetado o no etiquetado para el polipéptido.

En otra realización del método para identificar agonista o antagonista de un polipéptido de la presente invención comprende:

Determinar la inhibición o enlace de un ligando con células que poseen un polipéptido de la invención en la superficie de las mismas, o con membranas celulares que contienen dicho polipéptido, en la presencia de un compuesto candidato bajo condiciones que permitan el enlace con el polipéptido, y determinar la cantidad de enlace de ligando con el polipéptido. Un compuesto capaz de provocar la reducción de enlace de un ligando es considerado agonista o antagonista. De manera preferible el ligando se etiqueta.

Más en particular, un método para analizar un compuesto polipéptido agonista o antagonista comprende los siguientes pasos:

- (a) incubar un ligando etiquetado con una célula completa que exprese un polipéptido de acuerdo con la invención en la superficie celular, o una membrana celular que contenga un polipéptido de la invención;
- (b) medir la cantidad de ligando etiquetado enlazado con la célula completa con la membrana celular;
- (c) añadir un compuesto candidato a la mezcla del ligando etiquetado y la célula completa o la membrana celular del paso (a) y dejar que la mezcla consiga equilibrio;
- (d) medir la cantidad de ligando etiquetado enlazado a la célula completa o membrana celular tras el paso (c); y
- (e) comparar la diferencia en el ligando etiquetado enlazado en el paso (b) y (d), de tal modo que el compuesto que provoque la reducción en el enlace en el paso (d) se considere un agonista o un antagonista.

Pueden encontrarse polipéptidos para modular un conjunto de procesos fisiológicos y patológicos de un modo dependiente de dosis en los ensayos arriba descritos. Por lo tanto, los “equivalentes funcionales” de los polipéptidos de la invención incluyen polipéptidos que muestran alguna de las mismas actividades moduladoras en los ensayos arriba descritos de un modo dependiente de dosis. A pesar de que el grado de actividad dependiente de dosis no tiene por qué ser idéntico al de los polipéptidos de la invención, preferiblemente los “equivalentes funcionales” mostrarán una dependencia de dosis sustancialmente similar en un ensayo de una actividad determinada en comparación con los polipéptidos de la invención. En ciertas realizaciones de las descritas anteriormente, pueden emplearse ensayos sencillos de enlace, en los cuales la adherencia al compuesto de la prueba a una superficie que soporta el polipéptido se detecta por medio de una etiqueta directamente o indirectamente asociada con el compuesto de la prueba o en un ensayo que incorpora la competición con un competidor etiquetado. En otra realización, los ensayos analíticos de fármaco competitivo pueden emplearse, en los cuales los anticuerpos desnaturalizados que son capaces de enlazarse específicamente con el polipéptido compiten con un compuesto de la prueba para el enlace. De este modo, los anticuerpos pueden emplearse para detectar la presencia de algún compuesto de la prueba que contenga afinidad específica de enlace para el polipéptido.

Así mismo, también pueden diseñarse ensayos para detectar el efecto de compuestos de prueba añadidos en la producción de mRNA codificador del polipéptido en las células. Por ejemplo, se puede emplear un ELISA que mide los niveles segregados y asociados a las células del polipéptido empleando anticuerpos monoclonales o policlonales mediante métodos estándar conocidos en la técnica, y este ensayo puede emplearse para buscar compuestos que puedan inhibir o potenciar la producción del polipéptido a partir de células o tejidos adecuadamente manipulados. A continuación se puede medir la formación de complejos de enlace entre el polipéptido y el compuesto que está siendo analizado.

Métodos de ensayo que también se incluyen en los términos de la presente invención son aquellos que implican el uso de los genes y polipéptidos de la invención en ensayos de sobreexpresión o ablación. Dichos ensayos implican la manipulación de niveles de estos genes/polipéptidos en células y el cálculo del impacto de esta manipulación en la fisiología de las células manipuladas. Por ejemplo, tales experimentos revelan detalles de rutas señaladas y

metabólicas en las cuales los genes/polipéptidos particulares están implicados, generan información relativa a las identidades de polipéptidos con los cuales los polipéptidos estudiados interactúan y proporcionan pistas en relación con los métodos por medio de los cuales los genes y proteínas relacionadas se regulan.

- 5 Otra técnica para el análisis farmacéutico que puede emplearse proporciona un análisis de elevado rendimiento de compuestos que poseen una afinidad adecuada de enlace con el polipéptido de interés (ver la Solicitud de Patente Internacional WO84/03564). En este método, un gran número de diferentes compuestos de la prueba se sintetizan sobre un sustrato sólido, que a continuación puede reaccionar con el polipéptido de la invención y puede lavarse. Un modo de inmovilizar el polipéptido es el empleo de anticuerpos no neutralizadores. Posteriormente, el polipéptido
10 enlazado puede detectarse empleando métodos que son bien conocidos en la técnica. El polipéptido purificado también puede cubrirse directamente en placas para uso en las técnicas analíticas de fármacos anteriormente mencionadas.

- El polipéptido de la invención puede emplearse para identificar receptores enlazados con la membrana o solubles, por medio de técnicas de enlace de receptor estándar que son conocidas en la técnica, tales como el ensayo de enlace
15 de ligando el de enlace cruzado, en los cuales el polipéptido es etiquetado con un isótopo radioactivo, se modifica químicamente, o se fusiona con una secuencia de péptido que facilita su detección o purificación, y se incuba con una fuente del receptor putativo (por ejemplo, una composición de células, membranas celulares, sobrenadantes celulares, extractos de tejidos, o fluidos corporales). La eficacia del enlace puede medirse empleando técnicas biofísicas como la resonancia de plasmón superficial y espectroscopia. Los ensayos de enlace pueden usarse para la purificación y
20 clonación del receptor, pero también pueden identificar agonistas y antagonistas del polipéptido, que compitan con el enlace del polipéptido con su receptor. Métodos estándar para realizar ensayos analíticos son bien conocidos y comprendidos en la técnica.

- La invención incluye los ligandos que son anticuerpos que modulan la actividad o antigenicidad del polipéptido de la invención descubiertas por los métodos descritos con anterioridad.
25

- La invención también proporciona composiciones farmacéuticas que comprenden un polipéptido, ácido nucleico o ligando de la invención en combinación con un portador farmacéuticamente aceptable. Estas composiciones pueden ser adecuadas como reagentes terapéuticos o de diagnóstico, como vacunas, o como otras composiciones inmunogénicas, tal y como se expresa con más detalle a continuación.
30

- De acuerdo con la terminología aquí empleada, una composición que contiene un polipéptido, ácido nucleico, ligando o compuesto [X] está "sustancialmente libre de" impurezas [aquí, Y] cuando al menos el 85% por peso del total X+Y en la composición es X. Preferentemente, X comprende al menos aproximadamente el 90% por peso del total X+Y en la composición, más preferentemente al menos aproximadamente el 95%, 98% o incluso el 99% por peso.
35

- La composición farmacéutica debería preferentemente comprender una cantidad terapéuticamente efectiva del polipéptido, molécula de ácido nucleico o ligando de la invención. El término "cantidad terapéuticamente efectiva" como aquí se emplea hace referencia a una cantidad de un agente terapéutico necesario para tratar, mejorar o prevenir una enfermedad o condiciones objetivo, o para mostrar un efecto terapéutico o preventivo detectable. Para cualquier compuesto, la dosis terapéuticamente efectiva puede estimarse bien en ensayos de cultivo celular, por ejemplo, de células neoplásicas, o en modelos animales, normalmente ratones, conejos, perro o cerdos. El modelo animal también puede emplearse para determinar el rango de concentración apropiado y ruta de administración. A continuación, dicha
40 información puede emplearse para determinar dosis útiles y rutas de administración en humanos.

- La cantidad precisa efectiva para un sujeto humano dependerá de la severidad del estado de la enfermedad, de la salud general del sujeto, edad, peso, y género del sujeto, dieta, tiempo y frecuencia de administración, combinación o combinaciones de fármacos, sensibilidad a las reacciones, y la tolerancia/respuesta a la terapia. Esta cantidad puede determinarse mediante experimentación rutinaria y se encuentra dentro del juicio del doctor. En general, una dosis efectiva oscilará entre 0.01 mg/kg y 50 mg/kg, preferiblemente entre 0.05 mg/kg y 10 mg/kg. Las composiciones pueden administrarse individualmente a un paciente o pueden administrarse en combinación con otros agentes, fármacos u hormonas.
50

- Una composición farmacéutica también puede contener un portador farmacéuticamente aceptable para la administración de un agente terapéutico. Tales portadores incluyen anticuerpos y otros polipéptidos, genes y otros agentes terapéuticos como liposomas, siempre y cuando el propio portador no provoque la producción de anticuerpos perjudiciales para el individuo que está recibiendo la composición, y que se puedan administrarse sin demasiada toxicidad. Los portadores adecuados pueden ser macromoléculas grandes, lentamente metabolizadas como proteínas, polisacáridos, ácidos polilácticos, ácidos poliglicólicos, aminoácidos poliméricos, copolímeros de aminoácido y partículas inactivas de virus.
60

- Las sales farmacéuticamente aceptables también se pueden emplear en este paso, por ejemplo, sales ácidas minerales como hidroclouros, hidrobromuros, fosfatos, sulfatos, y similares; y las sales de ácidos orgánicos como acetatos, propionatos, malonatos, benzoatos, y similares. Una meticulosa descripción sobre portadores farmacéuticamente aceptables se encuentra disponible en Remington's Pharmaceutical Sciences (Mack Pub. Co., N. J. 1991).
65

Los portadores farmacéuticamente aceptables en composiciones terapéuticas pueden contener de manera adicional líquidos tales como el agua, salina, glicerol y etanol. Además, sustancias auxiliares como agentes humectantes o emulsionantes, sustancias búfer pH, y similares, pueden estar presentes en tales composiciones. Los mencionados portadores permiten que las composiciones farmacéuticas se puedan formular como pastillas, píldoras, grageas, cápsulas, líquidos, geles, jarabes, papillas o suspensiones, y similares, para la ingesta por parte del paciente.

Una vez formuladas, las composiciones de la invención pueden administrarse directamente al sujeto. Los sujetos a tratar pueden ser animales; en particular, se pueden tratar sujetos humanos.

Las composiciones farmacéuticas utilizadas en esta invención pueden administrarse por cualquiera de las rutas incluyendo, aunque no limitando, las aplicaciones orales, intravenosas, intramusculares, intraarteriales, intramedulares, intratecales, intraventriculares, transdérmicas o transcutáneas (por ejemplo, ver WO98/20734), medios subcutáneos, intraperitoneales, intranasales, enterales, tópicos, sublinguales, intravaginales o rectales. También pueden emplearse pistolas de gen o hidrosprays para la administración de las composiciones farmacéuticas de la invención. Normalmente, las composiciones terapéuticas pueden prepararse como inyectables, bien como soluciones líquidas o suspensiones; también pueden prepararse formas sólidas adecuadas para la solución en, o suspensión en, vehículos líquidos antes de la inyección.

El envío directo de las composiciones generalmente se llevará a cabo mediante inyección, subcutáneamente, intraperitonealmente, intravenosamente o intramuscularmente, o se enviará al espacio intersticial de un tejido. Las composiciones también pueden administrarse en una lesión. El tratamiento de dosis puede ser un programa de una única dosis o un programa con múltiples dosis.

Si la actividad del polipéptido de la invención es excesiva en un estado particular de la enfermedad, se encuentran disponibles varias propuestas. Una propuesta consiste en la administración a un sujeto de un compuesto inhibidor (antagonista) tal y como se ha descrito anteriormente, junto con un portador farmacéuticamente aceptable en una cantidad efectiva para inhibir la función del polipéptido, bloqueando el enlace de ligandos, sustratos, enzimas, receptores o inhibiendo una segunda señal, y por ello aliviando la condición anormal. Preferentemente, dichos antagonistas son anticuerpos. Más preferentemente, dichos anticuerpos son quiméricos y/o humanizados para minimizar su inmunogenicidad, tal y como se ha establecido anteriormente.

En otra propuesta, pueden administrarse las formas solubles del polipéptido que retienen la afinidad de enlace para el ligando, sustrato, enzima o receptor en cuestión. Normalmente, el polipéptido se administrará en la forma de fragmentos que retengan las partes relevantes.

En una propuesta alternativa, la expresión del gen codificador del polipéptido puede inhibirse usando técnicas de bloqueo de expresión, como el uso de moléculas de ácido nucleico antisentido (tal y como se ha descrito anteriormente), bien generadas internamente o administradas por separado.

Las modificaciones de expresión del gen pueden obtenerse diseñando secuencias complementarias o moléculas antisentido (DNA, RNA, o PNA) para las regiones de control, 5' o reguladores (secuencia señalizadora, promotores, amplidores o intrones) del gen codificador del polipéptido. Se manera similar, la inhibición puede conseguirse usando la metodología base-emparejamiento "triple hélice". El emparejamiento con triple hélice es útil porque provoca la inhibición de la habilidad de la doble hélice para abrirse lo suficiente para el enlace de polimerasas, factores de transcripción, o moléculas reguladoras. Los recientes avances terapéuticos que emplean el triple DNA se han descrito en varios libros (Gee, J. E. *et al.* (1994) In: Huber, B. E. y B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). La secuencia complementaria o molécula antisentido también puede diseñarse para bloquear la traslación de mRNA previniendo así la transcripción del enlace con los ribosomas. Dichos oligonucleótidos pueden administrarse o pueden generarse *in situ* a partir de expresión *in vivo*.

Además, la expresión del polipéptido de la invención puede evitarse usando ribozimas específicos a su secuencia codificadora mRNA. Las ribozimas son RNAs catalíticamente activos que pueden ser naturales o sintéticos (ver por ejemplo Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Las ribozimas sintéticas pueden diseñarse para dividir mRNAs específicamente en posiciones seleccionadas evitando de este modo la traslación de los mRNAs en el polipéptido funcional. Las ribozimas pueden sintetizarse con un eje o una columna natural de ribosa fosfato y con bases naturales, tal y como se encuentran normalmente en las moléculas RNA. De manera alternativa, las ribozimas pueden sintetizarse con ejes o columnas no naturales, por ejemplo, 2'-O-metil RNA, para proporcionar protección a partir de la degradación de ribonucleasa y pueden contener bases modificadas.

Las moléculas de RNA pueden modificarse para aumentar la estabilidad intracelular y la media de vida. Las posibles modificaciones incluyen, aunque no se limitan a, la adición de secuencias flanqueantes en los extremos 3' o 5' de la molécula o el uso de fosforotioato o 2' O-metil en lugar de enlaces de fosfodiesterasa en el interior de la columna o eje molecular. Este concepto es inherente en la producción de PNAs y puede extenderse en todas las moléculas mediante la inclusión de bases no tradicionales como inopina, queosina, butosina, así como acetil-, metil-, tio- y formas similarmente modificadas de adenina, histidina, guanina, timina y uridina que no son fácilmente reconocibles por endonucleasas endógenas.

Para tratar condiciones anormales relacionadas con una bajo-expresión del polipéptido de la invención y su actividad, se disponen de varias propuestas. Una propuesta consiste en administrar al sujeto una cantidad terapéuticamente efectiva de un compuesto que active el polipéptido, por ejemplo, un agonista como el descrito anteriormente, para aliviar la condición anormal. De manera alternativa, puede administrarse una cantidad terapéutica del polipéptido en combinación con un portador farmacéutico adecuado para restablecer el equilibrio fisiológico del polipéptido.

La terapia genética puede emplearse para efectuar la producción endógena del polipéptido mediante células relevantes en el sujeto. La terapia genética se emplea para tratar permanentemente la producción inapropiada del polipéptido reemplazando el gen defectuoso por un gen terapéutico correcto.

La terapia genética de la presente invención puede darse *in vivo* o *ex vivo*. La terapia de gen *ex vivo* requiere el aislamiento y purificación de células del paciente, la introducción de un gen terapéutico y la introducción de células genéticamente alteradas en el paciente. Si embargo, la terapia genética *in vivo* no requiere el aislamiento y la purificación de las células de un paciente.

El gen terapéutico normalmente se encuentra “empaquetado” para la administración a un paciente. Los vehículos encargados del envío del gen pueden ser no virales, como liposomas, o virus deficientes de réplica, como adenovirus tal y como se describe por Berkner, K. L., en Curr. Top. Microbiol. Immunol., 158, 39-66 (1992) o vectores de virus asociado a adeno (AAV) tal y como se describen por Muzyczka, N., en Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) y en la Patente U.S. N° 5,252,479. Por ejemplo, una molécula de ácido nucleico codificadora de un polipéptido de la invención puede producirse para su expresión en un vector retroviral deficiente de réplica. Este constructo de expresión puede a continuación aislarse e introducirse en una célula empaquetadora transducida con un vector plásmido retroviral que contiene RNA codificador del polipéptido, de tal modo que ahora la célula empaquetadora produce partículas virales infecciosas que contienen el gen de interés. Estas células productoras pueden administrarse a un sujeto para la creación de células *in vivo* y la expresión del polipéptido *in vivo* (ver Capítulo 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (y referencias citadas en el mismo) en Human Molecular Genetics (1996), T Strachan y A. P. Read, BIOS Scientific Publishers Ltd).

Otra propuesta es la administración de “DNA desnudo” en el cual el gen terapéutico se inyecta directamente en el flujo sanguíneo del tejido muscular.

En situaciones en las que el polipéptido o las moléculas de ácido nucleico de la invención son agentes causantes de enfermedad, la invención indica que pueden emplearse en vacunas para aumentar los anticuerpos contra el agente causante de la enfermedad. Mientras que el polipéptido o la molécula de ácido nucleico anteriormente mencionados son regulados por incremento, el desarrollo de la vacuna puede implicar el aumento de anticuerpos o células T contra tales agentes (tal y como se describe en WO00/29428).

Las vacunas de acuerdo con la invención puede ser bien profilácticas (es decir, para prevenir la infección) o terapéuticas (es decir, para tratar la enfermedad tras la infección). Dichas vacunas consisten en inmunizar antígeno(s), inmunógeno(s), polipéptido(s), proteína(s) o ácido nucleico, normalmente en combinación con portadores farmacéuticamente aceptables como los descritos anteriormente, que incluyen cualquier portador que por sí mismo no provoque la producción de anticuerpos perjudiciales para el individuo que está recibiendo la composición. Además, estos portadores pueden funcionar como agentes inmunoestimuladores (“adyuvantes”). Además, el antígeno o inmunógeno puede conjugarse con un toxoide bacteriano, como toxoide de difteria, tétano, cólera, *H. pylori*, y otros patógenos.

Debido a que los polipéptidos pueden romperse en el estómago, las vacunas que comprenden polipéptidos se administran preferentemente parenteralmente (por ejemplo, mediante inyección subcutánea, intramuscular, intravenosa o intradérmica). Las formulaciones adecuadas para administración parenteral incluyen soluciones por inyección estériles acuosas o no acuosas que pueden contener antioxidantes, búferes, bacteriostáticos, soluciones derretidas que hacen que la formulación se transforme en isotónica con la sangre del recipiente, y suspensiones estériles acuosas y no acuosas que pueden incluir agentes suspensores o agentes espesantes.

Las formulaciones de vacunas de la invención pueden presentarse en envases de única dosis o de múltiple dosis. Por ejemplo, las ampollas y viales sellados pueden almacenarse en condiciones de congelación seca tan sólo necesitando la adición de un portador líquido estéril inmediatamente antes del uso. La dosis dependerá de la actividad específica de la vacuna y puede determinarse de modo sencillo mediante experimentación rutinaria.

El envío genético de anticuerpos que se enlazan con los polipéptidos de acuerdo con la invención puede verse también afectados, por ejemplo como se ha descrito en la Solicitud de Patente Internacional WO98/55607.

La tecnología referida a la inyección reactiva (ver por ejemplo, www.powderjet.com) puede también resultar útil en la comulación de las composiciones para vacunas.

Varios métodos adecuados para la vacunación y sistemas de envíos para vacunas se describen en la Solicitud de Patente Internacional W000/29428.

Esta invención también se refiere al uso de moléculas de ácido nucleico de acuerdo con la presente invención como reagentes de diagnóstico. La detección de una forma mutada del gen caracterizada por las moléculas de ácido nucleico de la invención que se asocia con una disfunción proporcionará una herramienta de diagnóstico que puede añadir o definir un diagnóstico de una enfermedad, o la propensión a una enfermedad, que puede darse como resultado a partir de una infraexpresión, sobreexpresión, o expresión temporal o espacial alterada del gen. Los individuos que transportan mutaciones en el gen pueden detectarse en el nivel ADN mediante una gran variedad de técnicas.

Las moléculas de ácido nucleico pueden obtenerse a partir de células de un sujeto, así como a partir de sangre, orina, saliva, tejido de biopsia o material de autopsia. El DNA genómico puede usarse directamente para la detección o puede amplificarse enzimáticamente mediante el uso de PCR, reacción de cadena de la ligasa (LCR), amplificación por desplazamiento de la cadena (SDA), u otras técnicas de amplificación (ver Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer *et al.*, J. Virol. Meth., 35, 117-126 (1991); Van Brunt, J., Bio/Technology, 8, 291-294 (1990)) anteriores al análisis.

En una realización, este aspecto de la invención proporciona un método de diagnóstico de enfermedad en un paciente, que comprende el cálculo del nivel de expresión de un gen natural codificador de un polipéptido de acuerdo con la invención y comparando dicho nivel de expresión con un nivel de control, donde un nivel que es diferente a dicho nivel de control es un indicador de enfermedad. El método puede comprender los siguientes pasos:

a) poner en contacto una muestra de tejido de un paciente con una sonda de ácido nucleico bajo condiciones rigurosas que permiten la formación de un complejo híbrido entre una molécula de ácido nucleico de la invención y la sonda;

b) poner en contacto una muestra de control con dicha sonda bajo las mismas condiciones empleadas en el paso a);

c) y detectar la presencia de complejos híbridos en dichas muestras;

donde la detección de niveles del complejo híbrido en la muestra del paciente que difieren de los niveles del complejo híbrido en la muestra control es indicativo de enfermedad.

Un aspecto adicional de la invención comprende un método de diagnóstico que incluye los siguientes pasos:

a) obtener una muestra de tejido a partir de un paciente que está sometido a pruebas en busca de una enfermedad;

b) aislar una molécula de ácido nucleico de acuerdo con la invención a partir de dicha muestra de tejido; y

c) diagnosticar la enfermedad al paciente detectando la presencia de una mutación en la molécula de ácido nucleico que está asociada con la enfermedad.

Para ayudar a la detección de moléculas de ácido nucleico en los métodos anteriormente descritos, puede incluirse un paso de amplificación, por ejemplo, el empleo de PCR.

Las eliminaciones e inserciones pueden detectarse mediante un cambio en el tamaño del producto amplificado en comparación con el genotipo normal. Las mutaciones de punto pueden identificarse hibridizando DNA amplificado con RNA etiquetado de la invención o de manera alternativa, secuencias DNA antisentido etiquetadas de la invención. Las secuencias que combinan perfectamente pueden distinguirse a partir de parejas desiguales mediante digestión RNasa o mediante el cálculo de diferencias en las temperaturas de fundición. La presencia o ausencia de la mutación en el paciente puede detectarse mediante la puesta en contacto de DNA con una sonda de ácido nucleico que hibridice con ADN bajo condiciones rigurosas para formar una molécula híbrida de doble trenzado, teniendo la molécula híbrida de doble trenzado una porción no hibridizada de la cadena sonda de ácido nucleico en una parte correspondiente a una mutación asociada con la enfermedad; y la detección de presencia o ausencia de una parte no hibridizada de la cadena sonda como una indicación de la presencia o ausencia de una mutación asociada con la enfermedad en la parte correspondiente de la cadena de DNA.

Los citados diagnósticos son particularmente útiles para pruebas prenatales o incluso neonatales.

Las mutaciones de punto y otras diferencias secuenciales entre el gen referencia y los genes "mutantes" pueden identificarse mediante otras técnicas bien conocidas, como la secuenciación directa de DNA o el polimorfismo conformacional de único trenzado (ver Orita *et al.*, Genomics, 5, 874-879 (1989)). Por ejemplo, puede emplearse un imprimador secuencial con producto PCR de doble trenzado o con una molécula templada de único trenzado generada mediante un PCR modificado. La determinación secuencial se lleva a cabo mediante procedimientos convencionales con nucleótidos radioetiquetados o mediante procedimientos secuenciales automáticos con etiquetas fluorescentes. Los segmentos de DNA clonados pueden también emplearse como sondas para detectar segmentos específicos de DNA. La sensibilidad de este método aumenta en gran medida cuando se combina con PCR. Además, las mutaciones de punto y otras variaciones secuenciales, como los polimorfismos, pueden detectarse tal y como se ha descrito anteriormente, por ejemplo, mediante el uso de oligonucleótidos específicos de alegos para la amplificación PCR de secuencias que difieren mediante nucleótidos sencillos.

Las diferencias secuenciales de ADN también pueden detectarse por medio de alteraciones en la movilidad electroforética de fragmentos de ADN en geles, con o sin agentes desnaturalizadores, o mediante secuenciación directa de DNA (por ejemplo, Myers *et al.*, Science (1985) 230:1242). Los cambios secuenciales en localizaciones específicas también pueden revelarse mediante ensayos de protección de nucleasa, como RNasa y protección S1 o el método de división química (ver Cotton *et al.*, Proc. Natl. Acad. Sci. USA (1985) 85:4397-4401).

Además de electroforesis de gel convencional y secuencias de DNA, las mutaciones tales como microeliminaciones, aneuploidies, translocalizaciones, inversiones, también pueden detectarse por medio de análisis *in situ* (ver, por ejemplo, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), es decir, las secuencias de DNA o RNA en las células pueden analizarse en búsqueda de mutaciones sin la necesidad de aislamiento y/o inmovilización en una membrana. La hibridización de fluorescencia *in situ* (FISH) es en este momento el método más común aplicado y han aparecido numerosas referencias a FISH (ver, por ejemplo, Trachuck *et al.*, Science, 250, 559-562 (1990), y Trask *et al.*, Trenes, Genet., 7, 149-154 (1991)).

En otra realización de la invención, puede construirse una selección de sondas de oligonucleótidos comprendiendo una molécula de ácido nucleico de acuerdo con la invención para llevar a cabo un análisis eficiente de variantes genéticas, mutaciones y polimorfismos. Los métodos de tecnología con selecciones son bien conocidos en la técnica, poseen una aplicabilidad general y pueden emplearse para dirigir una gran variedad de cuestiones en genética molecular, incluyendo expresión genética, enlace genético y variabilidad genética (ver por ejemplo: M. Chee *et al.*, Science (1996), Vol. 274, pp 610-613).

En una realización, la selección se prepara y emplea de acuerdo con los métodos descritos en la Solicitud PCT WO95/11995 (Chee *et al.*); Lockhart, D. J., *et al.* (1996) Nat. Biotech. 14: 1675-1680); y Schena, M. *et al.*, (1996) Proc. Natl. Acad. Sci. 93: 10614-5 10619). Las parejas de oligonucleótidos pueden oscilar entre dos a más de un millón. Los oligómeros se sintetizan en zonas designadas sobre un sustrato usando un proceso químico dirigido a la luz. El sustrato puede ser papel, nylon, u otro tipo de membrana, filtro, chip, bandeja de cristal, o cualquier soporte sólido adecuado. En otro aspecto, un oligonucleótido puede sintetizarse sobre una superficie del sustrato usando un procedimiento de enlace químico y una aparato de aplicación de tinta por inyección, tal y como se describe en la solicitud PCT WO95/251116 (Balteschweiler *et al.*). En otro aspecto, una selección "con rejillas" análoga a una mancha con punto (u orificio) puede emplearse para ordenar y enlazar fragmentos cDNA u oligonucleótidos a la superficie de un sustrato usando un sistema de vaciado, procedimientos térmicos, UV, de enlace mecánico o químico. Una selección, como las descritas anteriormente, pueden producirse a mano o mediante el uso de 5 dispositivos disponibles (aparatos de mancha con orificio o mancha con punto), materiales (cualquier soporte sólido adecuado), y máquinas (incluyendo instrumentos de robótica), y pueden contener 8, 24, 96, 384, 1536 o 6144 oligonucleótidos o cualquier número entre dos y más de un millón que den como lugar un uso eficiente de instrumentación disponible en el mercado.

Además de los métodos descritos con anterioridad, las enfermedades pueden diagnosticarse mediante métodos que incluyen la determinación, a partir de una muestra derivada de un sujeto, de un nivel anormalmente alto o bajo de polipéptido o mRNA. La expresión aumentada o disminuida puede medirse en el nivel RNA usando cualquiera de los métodos bien conocidos en la técnica para la cuantificación de polinucleótidos, como por ejemplo, amplificación de ácido nucleico, citando como ejemplos PCR, RT-PCR, protección RNasa, transferencia de tipo Northern y otros métodos de hibridización.

Las técnicas para los ensayos pueden usarse para determinar los niveles de un polipéptido de la presente invención es una muestra derivada de un huésped son bien conocidas para aquellos expertos en la técnica y se describen con detalle a continuación (incluyendo inmunoensayos, ensayos enlace-competitivo, análisis de transferencia de tipo Western y ensayos ELISA). Este aspecto de la invención proporciona un método de diagnóstico que comprende los siguientes pasos: (a) poner en contacto un ligando como el descrito anteriormente con una muestra biológica bajo condiciones adecuadas para la formación de un complejo ligando-polipéptido; y (b) detectar dicho complejo.

Los protocolos tales como ELISA, RIA y FACS para medir niveles de polipéptido pueden proporcionar de manera adicional una base para diagnosticar niveles alterados o anormales de expresión de polipéptido. Los valores normales o estándar para la expresión de polipéptido se establecen mediante la combinación de fluidos corporales o extractos celulares tomados de sujetos mamíferos normales, preferiblemente humanos, con un anticuerpo al polipéptido bajo condiciones adecuadas para la formación del complejo. La cantidad de formación de complejo estándar puede cuantificarse por medio de varios métodos, como por ejemplo el medio fotométrico.

Los anticuerpos que específicamente se enlazan con un polipéptido de la invención pueden usarse para el diagnóstico de condiciones o enfermedades caracterizadas por la expresión del polipéptido, o en ensayos para controlar pacientes que están siendo tratados con polipéptidos, moléculas de ácido nucleico, ligandos y otros compuestos de la invención. Los anticuerpos útiles para fines de diagnóstico pueden prepararse del mismo modo que los descritos anteriormente para terapia. Los ensayos de diagnóstico para el polipéptido incluyen métodos que utilizan el anticuerpo y una etiqueta para detectar el polipéptido en fluidos corporales humanos o extractos de células o tejidos. Los anticuerpos pueden emplearse con o sin modificación, y pueden etiquetarse uniéndolos, bien covalentemente o no covalentemente, con una molécula reportera. Pueden emplearse una gran variedad de moléculas reporteras conocidas en la técnica, algunas de ellas descritas anteriormente.

Las cantidades de polipéptidos expresados en las muestras del sujeto, control y enfermedad a partir de tejidos biopsiados se comparan con los valores estándar. La desviación entre los valores estándar y los del sujeto establece los parámetros para el diagnóstico de la enfermedad. Los ensayos de diagnóstico pueden emplearse para distinguir entre ausencia, presencia y exceso de expresión de polipéptido y para controlar la regulación de niveles de polipéptido durante la intervención terapéutica. Dichos ensayos pueden también emplearse para evaluar la eficacia de un régimen tratamiento terapéutico en particular en estudios animales, en ensayos clínicos o en el control del tratamiento de un paciente individual.

Un kit de diagnóstico de la presente invención puede estar formado por:

- (a) una molécula de ácido nucleico de la presente invención;
- (b) un polipéptido de la presente invención; o
- (c) un ligando de la presente invención.

En un aspecto de la invención, un kit de diagnóstico puede estar constituido por un primer envase que contiene una sonda de ácido nucleico que se hibridiza bajo condiciones rigurosas con una molécula de ácido nucleico de acuerdo con la invención; un segundo envase que contiene imprimadores útiles para amplificar la molécula de ácido nucleico; y instrucciones para usar la sonda y los imprimadores para facilitar el diagnóstico de la enfermedad. El kit puede además contener un tercer envase con un agente para digerir el RNA no hibridizado.

En un aspecto alternativo de la invención, el kit de diagnóstico puede contener una selección de moléculas de ácido nucleico, y que al menos una de las cuales sea una molécula de ácido nucleico de acuerdo con la invención.

Para detectar un polipéptido de acuerdo con la invención, el kit de diagnóstico puede estar formado por uno o más anticuerpos que se enlazan con un polipéptido de acuerdo con la invención; y un reagente útil para la detección de una reacción de enlace entre el anticuerpo y el polipéptido.

Tales kits se emplearán para diagnosticar una enfermedad o propensión a una enfermedad, en particular desórdenes proliferativos celulares, desórdenes inmunes/inflamatorios, desórdenes cardiovasculares, desórdenes neurológicos, desórdenes de desarrollo, desórdenes metabólicos, infecciones y otras condiciones patológicas. La enfermedad es preferentemente una enfermedad en la que se implican niveles anómalos de una citoquina con pliegue y nudo de cistina, preferentemente de un miembro de la subfamilia DNA. La enfermedad o desorden pueden ser aquel en el que están implicados los niveles anómalos de un ligando de una citoquina con pliegue y nudo de cistina, preferentemente de un miembro de la subfamilia DNA. Por ejemplo, la enfermedad o desorden puede ser aquel en el que se implican niveles anómalos de un miembro de la superfamilia TGBeta. En particular, la enfermedad o desorden puede ser aquel en el que están implicados BMPs, tales como neuropatías, nefropatías como nefropatía diabética, cáncer, curación de heridas, fibrosis, osteopenia, osteoporosis, fracturas y esclerosteosis. A continuación se describirán con más detalle varios aspectos y realizaciones de la presente invención a modo de ejemplo, con referencia particular a los polipéptidos INSP002.

Se apreciará que se pueden realizar modificaciones de los detalles sin salir del alcance de la invención.

Breve descripción de las figuras

Figura 1: Resultados de BLAST contra base de datos No redundante NCBI usando SEQ ID NO:2 combinada y secuencia de polipéptido SEQ ID NO:4.

Figura 2: Alineación generada por BLAST entre SEQ ID NO:2 combinada y secuencia de polipéptido SEQ ID NO:4 y la secuencia relacionada, proteína de *Homo sapiens* cerberus-relacionado 1.

Figura 3: Base de datos BLAST Superior Cuatro NCBI-nr y NCBI-nt golpea contra INSP002 el 26 de noviembre del 2002.

Figura 4: Alineación de INSP002 con AK095926.1

Figura 5: Alineación de INSP002 con IMAGE: 4558384

Figura 6: Secuencia de nucleótido INSP002 con traslación

Figura 7: Secuencia clonada parcial INSP002 con traslación

Figura 8: Mapa de PCRII-TOPO-INSP002 parcial

Figura 9: Alineación de predicción (superior) INSP002 y secuencia clonada parcial (inferior).

ES 2 306 802 T3

Figura 10: Secuencia de nucleótido y traslación de inserción cADN en Image: 4558384

Figura 11: Alineación de secuencias de predicción (superior) INSP002 con IMAGE: 4558384 (BCO25333.1) (inferior).

Figura 12: Secuencia de nucleótido y traslación de INSP002V generada por PCR de IMAGE 4558384

Figura 13: Mapa de PCR4blunt-TOPO-INSP002V.

Figura 14: Comparación entre predicción (superior) INSP002 y secuencia variante INSP002V (inferior).

Figura 15: Mapa de vector expresión pEAK12d

Figura 16: Mapa vector puerta pDONR201

Figura 17: Mapa de pEAK12d-INSP002-V-6HIS

Figura 18: Secuencia de INSP002 de longitud completa clonado a partir de corazón

Figura 19: Mapa de PCR4blunt-TOPO-INSP002FL que codifica INSP002 de longitud completa de corazón se muestra en la Figura 19.

Ejemplos

Ejemplo 1

Comparación de proteína INSP002 con proteínas en la base de datos secuencial

La secuencia de polipéptido derivada de combinar SEQ ID NO:2 y SEQ ID NO:4, que representa la traslación de exones consecutivos de INSP002 se empleó como una consulta contra la base de datos secuencia no redundante NCBI. Los principales diez parejas incluyen secuencias anotadas como cerberus o relacionadas con cerberus, que son miembros de familia nudo de cistina, los cuales se alinean con la secuencia consulta con valores E altamente significativos ($2E^{-10}$ a $3E^{-06}$) (Figura 1). La Figura 2 muestra la alineación de la secuencia consulta INSP002 con la secuencia de proteína *Homos sapiens*- relacionada 1 (Feng *et al.*, 2001).

La secuencia de polipéptido derivada de combinar SEQ ID NO:2 y SEQ ID NO:4, que representa la traslación de exones consecutivos de INSP002 se introdujo en SignalP V2.0.b2 (Nielsen *et al.*, 1997 Protein Eng 1:1-6). El programa predijo que la secuencia de polipéptido tenía un péptido señalizador. El lugar con más posibilidades de división para el péptido señalizador es el encontrado entre los residuos 22 y 23 de la secuencia del polipéptido, INSP002, derivada de combinar SEQ ID NO:2 y SEQ ID NO:4.

La secuencia de nucleótido SEQ ID NO:1, que codifica el polipéptido SEQ ID NO:2 exón 1, comprende la región no traducida 5' (5' UTR) y la secuencia codificadora de proteína (CDS). El CDS comienza en el nucleótido 152.

Ejemplo 2

Repetición de búsquedas BLAST

Las búsquedas BLAST de las bases de datos NCBI-nr y NCBI-NT se llevaron a cabo el 26 de noviembre del 2002 empleando la secuencia de polipéptido de SEQ ID NO:6, derivada de la combinación entre SEQ ID NO:2 y SEQ ID NO:4. Los cuatro principales blancos identificados por estas búsquedas se muestran en la Figura 3.

Las búsquedas revelaron que el polipéptido INSP002 es idéntico a la proteína FLJ38607 en el nivel de aminoácido y a la secuencia correspondiente del nucleótido AK095926, clonada a partir del corazón y depositada el 16 de julio del 2002. La Figura 4 muestra la alineación de la secuencia consulta INSP002 con la proteína derivada del clon AK095926 cDNA.

La unión del exón 1 y el exón 2 pronosticada para INSP002 se prueba de manera experimental mediante la existencia de AK095926.

```

450      .      :      .      :      .      :      .      :      .      :
451  GATGTGTAAGGCTGTGCCCTTCGTTTCAG      GTGTTCTCCCGGC
5      |||||||||||||||||||||||||||||>>>...>>>|||||||||||||
10225  GATGTGTAAGGCTGTGCCCTTCGTTTCAGGTG...CAGGTGTTCTCCCGGC
      Exon1      Exon2
10      500      .      :      .      :      .      :      .      :      .      :
492  CCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGC
      |||||||||||||||||||||||||||||||||||||||||||||||||||
13670  CCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGC

```

Las búsquedas también revelaron que las partes del INSP002 son idénticas a IMAGE clon 4558384 (BCO25333.1) depositadas el 8 de marzo del 2002. La Figura 5 muestra la alineación de las parte de la secuencia consulta INSP002 con IMAGE clon 4558384.

Ejemplo 3

Clonación parcial de cDNA para INSP002

i) *Bibliotecas cADN*

Las bibliotecas humanas de cDNA (en vectores bacteriófagos lambda (λ) se compraron en Stratagene o Clotech y se prepararon en el Instituto de Investigación Farmacéutica Sero en vectores λ ZAP o en vectores λ GT10 de acuerdo con los protocolos del fabricante (Stratagene). El DNA bacteriófago A se preparó a partir de cultivos a pequeña escala del tipo de huésped *E. coli* usando el sistema de purificación Wizard Lambda Preps DNA de acuerdo con las instrucciones del fabricante (Promega, Corporation, Madison WI). La lista de bibliotecas y variedades de huésped empleados se muestra en la Tabla 1.

ii) *PCR de cDNAs virtuales de biblioteca fago DNA*

Un cDNA parcial codificador de INSP002 (Figura 6) se obtuvo como un producto de amplificación PCR de 159 bp (Figura 7) usando imprimadores de clonación de gen específicos (INSP002-CP1 e INSP002-CP2, Figura 6 y Tabla II). El PCR se llevó a cabo en un volumen final de 50 μ l conteniendo IX AmpliTaq™ búfer, 200 μ M dNTPs, 50 pmoles cada uno de ellos de imprimadores de clonación, 2,5 unidades de AmpliTaq™ (Perkin Elmer) y 100 ng de cada biblioteca fago DNA usando un MJ Research DNA Engine, programado de la siguiente forma: 94°C, 1 min; 40 ciclos de 94°C, 1 min, x°C, e y min y 72°C, (donde x es la Tm más inferior - 5°C e y = 1 min por kb de producto); seguido por 1 ciclo a 72°C durante 7 minutos y un ciclo de sujeción a 4°C.

Los productos de amplificación se visualizaron en 0.8% geles de azarosa en 1 X búfer TAE Invitrogen) y los productos PCR que migran en la masa molecular pronosticada se purificaron del gel usando el Sistema de Purificación Wizard PCR Peps DNA (Promega). Los productos PCR eluyeron en 50 μ l de agua estéril y bien se subclonaron directamente o se almacenaron a -20°C.

iii) *Imprimadores de clonación específicos de gen para PCR*

Las parejas de imprimadores PCR que tienen una longitud de entre 18 y 25 bases fueron diseñados para amplificar la secuencia de longitud completa del cDNA virtual usando Primer Designer Software (Scientific & Educational Software, PO Box 72045, Durham, NC 27722-2045, USA). Los imprimadores PCR se optimizaron para que tuvieran Tm cercana a $55 \pm 10^\circ\text{C}$ y contenido GC de 40-60%. Los imprimadores se seleccionaron en base a su elevada selectividad para la secuencia objetivo INSP002 (con poca preparación o con preparación no específica con otras muestras).

iv) *Subclonación de Productos PCR*

Los productos PCR se subclonaron en el vector de clonación topoisomerasa I modificado (pCR II TOPO) usando un kit de clonación TOPO TA adquirido en Invitrogen Corporation (cat. No. K4600-01 y K4575-01, respectivamente) usando las condiciones especificadas por el fabricante. Brevemente, la amplificación de 4 μ l de producto PCR de gel purificado procedente de una biblioteca de riñón fetal humano (biblioteca número 12) se incubó durante 15 minutos a temperatura ambiente con 1 μ l de vector TOPO y 1 μ l de solución salina. La mezcla de la reacción se transformó a continuación en la variedad *E. coli* TOP10 (Invitrogen) del siguiente modo: se fundieron en hielo 50 μ l de alícuota de células One Shot TOP10 y se añadieron 2 μ l de reacción TOPO. La mezcla se incubó durante 15 minutos en hielo y a continuación se amortiguó por calor mediante incubación a 42°C durante exactamente 30 segundos. Las muestras volvieron al hielo y se añadieron 250 μ l de medio SOC templado (a temperatura ambiente).

ES 2 306 802 T3

Las muestras se incubaron mediante agitación (220 rpm) durante 1 hora a 37°C. A continuación, la mezcla de la transformación se colocó en placas sobre placas L-broth (LB) que contenían ampicilina (100 µg/ml) y se incubaron durante la noche a 37°C. Las colonias resistentes a la ampicilina que contenían insertos de cDNA se identificaron mediante colonia PCR.

v) Colonia PCR

Las colonias se inocularon en 50 µl de agua estéril usando un palillo de dientes estéril. Una alícuota de 10 µl del inóculo se sometió a continuación al PCR con un volumen de reacción total de 20 µl tal y como se ha descrito anteriormente, excepto que las parejas de imprimadores empleadas fueron SP6 y T7. Las condiciones del ciclo fueron las siguientes: 94°C, 2 minutos; 30 ciclos de 94°C, 30 segundos, 47°C, 30 segundos y 72°C durante 1 minuto; 1 ciclo, 72°C, 7 minutos. Las muestras se mantuvieron a continuación a 4°C (ciclo de mantenimiento o sujeción) antes de los posteriores análisis.

Los productos de reacción PCR se analizaron sobre geles de 10/0 agarosa en búfer 1 X TAE. Las colonias que dieron el tamaño de producto PCR esperado (159 bp cDNA + 187 bp debido al punto de clonación múltiple o MCS) crecieron durante la noche a 37°C en 5 ml de L-Broth (LB) que contenía ampicilina (100 µg/ml), con un proceso de agitación a 220 rpm a 37°C.

vi) Preparación de plásmido DNA y Secuencias

La miniprep de plásmido DNA se preparó a partir de cultivos de 5 ml usando un sistema de robótica Qiaprep Turbo 9600 (Qiagen) o un kit Wizard Plus SV Minipreps (Promega cat. No. 1460) de acuerdo con las instrucciones del fabricante. El plásmido DNA se eluyó en 100 µl de agua estéril. La concentración de DNA se midió empleando un fotómetro Eppendorf BO. El plásmido DNA (200-500 ng) se sometió a las secuencias de DNA con el imprimador T7 y el imprimador SP6 usando el sistema BigDye Terminator (Applied Biosystems cat. No. 4390246) de acuerdo con las instrucciones del fabricante. Las reacciones secuenciales se purificaron usando columnas Dye-Ex (Qiagen) o placas de limpieza Montage SEQ 96 (Millipore cat. No. LSKS09624) y a continuación se analizaron en un secuenciador Applied Biosystems 3700.

vii) Identificación de bibliotecas cDNA que contienen INSP002

Los productos PCR obtenidos con INSP002-CP1 y INSP002-CP2 y la migración en el correcto tamaño (159 bp) se identificaron en las bibliotecas cDNA de la corteza, colon, pulmón fetal y riñón fetal (bibliotecas 8, 9, 11 y 12). La secuencia del producto PCR clonado en el vector pCRII-TOPO se muestra en la figura 7, y el mapa de plásmido (plásmido ID 13422) se muestra en la figura 8.

El cDNA parcial clonado es una parte de INSP002 exón 2, tal y como se muestra por la alineación de la secuencia de nucleótido INSP002 pronosticada y la secuencia de nucleótido parcial clonada en la figura 9a y la alineación de la secuencia proteica INSP002 pronosticada y la secuencia proteica parcial en la figura 9b.

Ejemplo 4

Generación de INSP002 ORF de Image: 4558384

El clon Image 4558384 (en plásmido pOTB7) procedente de retinoblastoma se adquirió en Resgen (Invitrogen Corp.). La penetración de *E. coli* se extendió sobre una placa LB que contenía ampicilina (100 µl/ml) y creció durante la noche a 37°C. Las únicas colonias resistentes a la ampicilina se inocularon en 5 ml de LB que contenía ampicilina (100 µl/ml) y se incubaron con agitación a 220 rpm durante la noche a 37°C. Se preparó el miniprep de plásmido DNA y se secuenció utilizando los imprimadores SP6, T7, M13F, INSP002-CP1 e INSP002-CP2 tal y como se ha descrito en el Ejemplo 3, vi).

La secuencia de la inserción se muestra en la figura 10. La alineación del nucleótido y de la secuencia putativa de aminoácido de Image 4558384 cADN con INSP002 se muestra en la figura 11. Image 4558384 cADN parece ser una variante de unión de INSP002. Contiene una inserción de 87 bp que introduce un desplazamiento y un codón de parada prematuro en comparación con la secuencia pronosticada de INSP002. Además, la secuencia no traducida 3' también contiene una repetición Alu indicativo de la contaminación genómica DNA del cDNA. Los exones 2 y 4 de Image 4558384 son equivalentes a los exones 1 y 2 de la predicción INSP002. Sin embargo, Image 4558384 incorpora un exón extra entre los exones 1 y 2 de la predicción INSP002. El exón extra codifica un codón de parada prematuro que evita la traslación del dominio de nudo de cistina. Los límites de unión en Image 4558384 cADN son los siguientes:

ES 2 306 802 T3

```

500      .      :      .      :      .      :      .      :      .      :
492 GGCTGTGCCCTTCGTTTCAG          ACACGGGAGTCTCGCTATGTTG
      |||||>>>...>>>|||||
5 10234 GGCTGTGCCCTTCGTTTCAGGTG...TAGACACGGGAGTCTCGCTATGTTG
      Exon2                      Exon3
550      .      :      .      :      .      :      .      :      .      :
10 533 CCCAAGCTAGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCACCTCAGCC
      |||||>>>...>>>|||||
15 10741 CCCAAGCTAGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCACCTCAGCC
      600      .      :      .      :      .      :      .      :      .      :
20 583 TCC[REDACTED]GTTCTAG          GTGTTCTCCCGGCCCGGCTGCTCAGCCA
      ||[REDACTED]>>>...>>>|||||
25 10791 TCC[REDACTED]GTTCTAGGTG...CAGGTGTTCTCCCGGCCCGGCTGCTCAGCCA
      Exon 4

```

Se creó una estrategia PCR para retirar el DNA genómico con el fin de generar cDNA de longitud completa co-
 30 dicador de INSP002 ORF. Los imprimadores PCR fueron diseñados para amplificar el extremo 5' (ascendente) y el
 extremo 3' (descendente) de la secuencia INSP002 que flanquea la inserción 87 bp en el clon Image. El imprimador
 hacia tras para la secuencia ascendente y el imprimador hacia delante para la secuencia descendente contenían secuen-
 cias complementarias en sus extremos 3' y 5' respectivamente para proporcionar extremos que se estaban montados
 35 unos sobre otros, de modo que los productos PCR de cada reacción podían mezclarse y templarse juntos, permitiendo
 la amplificación de cDNA de longitud completa en un tercera reacción PCR, usando un imprimador anidado que va
 en dirección delantera y que es ascendente y un imprimador anidado que va en dirección trasera y que es descendente.

La primera reacción PCR para amplificar el extremo 5' de INSP002 (ascendente de la inserción 87 bp) contenía
 en un volumen final de 50 µl: 5 µl 10X búfer Platinum Pfx, 1.5 µl dNTPs (10mM), 1 µl MgSO₄ (50 mM), 1.5 µl
 40 de INSP002V-5'-F (10 µM), 1.5 µl de INSP002V-5'-R (10 µM), 0.7501 Platinum Pfx y 135 ng IMAGE: 4558384
 plásmido cDNA. Las condiciones de amplificación fueron 1 ciclo de 94°C durante 2 minutos, 30 ciclos de 94°C, 15
 segundos y 68°C, 1 minuto; y 1 ciclo de 68°C durante 7 minutos. La segunda reacción de PCR para amplificar el
 extremo 3' de INSP002 (descendente de la inserción 87 bp) se llevó a cabo bajo las mismas condiciones excepto que
 los imprimadores fueron: INSP002V-3'-F y INSP002V-3'-R.

Los productos de amplificación se visualizaron en geles de agarosa 8% en búfer 1 X TAE (Invitrogen). Los pro-
 ductos PCR migratorios en la masa molecular pronosticada (520 bp y 448 bp, para PCR 1 y PCR 2, respectivamente)
 se purificaron a partir del gel que emplea el Sistema de Purificación DNA de Wizard PCR Preps (Promega). Los pro-
 ductos PCR se eluyeron en 50 µg de agua estéril y la concentración de DNA se midió usando un fotómetro Eppendorf
 BO. Cincuenta ng de cada producto PCR purificado se empleó a continuación como una plantilla para un PCR anidado
 50 en una reacción de 50 µl que contenía 5 µl 10X búfer Platinum Pfx, 1.5 µl dNTPs (10 mM), 1 µl MgSO₄ (50 mM), 1.5
 µl de INSP002V- nido 5'-F (10 µM) y 1.5 µl de INSP002V- nido 3'-R (10 µM). La mezcla de la reacción se calentó a
 95°C durante 3 minutos y se añadieron 0.75 µl de polimerasa Platinum Pfx. Las condiciones de amplificación fueron
 las siguientes: 1 ciclo de 94°C durante 2 minutos; 30 ciclos de 94°C, 15 segundos; 61°C, 30 segundos y 68°C, 1 minuto;
 y 1 ciclo de 68°C durante 7 minutos. Los productos PCR migratorios en la masa molecular pronosticada de 719 bp
 55 se purificaron a partir del gel que emplea el Sistema de Purificación DNA de Wizard PCR Preps y se eluyeron en
 50 µl de agua estéril. Cuatro µl del producto PCR purificado se ligó a continuación en un vector pCR4 blunt TOPO
 tal y como se ha descrito en la sección 1.4. Las colonias resistentes a la ampicilina fueron testadas para inserciones
 mediante colonia PCR usando los imprimadores T3 y T7 tal y como se ha descrito en la sección 1.5. Las colonias que
 dieron el tamaño de producto PCR esperado (719 bp + 106 bp debido al punto de clonación múltiple o MCS) crecieron
 60 durante la noche a 37°C en 5 ml de L-Broth (LB) que contenía ampicilina (100 µg/ml), con un proceso de agitación a
 220 rpm a 37°C. El plásmido miniprep DNA se preparó a partir de cultivos de 5 ml y se secuenció con imprimadores
 T3 y T7 tal y como se describe en la sección 1.6. La secuencia de uno de los clones resultantes y el mapa plásmido
 correspondiente (pCR4 blunt TOPO-INSP002V) se muestran en las figuras 12 y 13 respectivamente. La traslación
 de la secuencia clonada indica que INSP002V contiene dos eliminaciones de aminoácido (ΔV107 y ΔV108) y una
 65 única sustitución de aminoácido (F110L) en comparación con la secuencia INSP002 pronosticada. La alineación de
 las secuencias de nucleótido y aminoácido para la predicción INSP002 e INSP002V se muestran en la figura 14.

Cuando se compara la predicción INSP002, la unión entre los exones 1 y 2 punto donante ascendente 6bp. Esto da como resultado la eliminación de dos aminoácidos (ValGlu). La unión aceptante usada es la misma que la usada por la predicción INSP002, a pesar de que hay dos errores secuenciales tras el receptor. Esto da como resultado una sustitución de aminoácido (Phe->Leu).

Ejemplo 5

Contrición de un plásmido para la expresión de INSP002V en células HEK293/EBNA

Un clon pCR4 blunt-TOPO que contiene la secuencia completamente codificadora (ORF) de INSP002V identificada por secuenciación de DNA (Figura 3) se empleó a continuación para subclonar la inserción en el vector de expresión celular del mamífero pEAK12d (figura 15) usando la metodología de clonación Gateway™ (Invitrogen).

i) Generación de Gateway compatible INSP002 ORF fusionada con una secuencia de etiqueta interior 6HIS

El primer paso del proceso de clonación Gateway implica una reacción PCR de dos pasos que genera ORF o INSP002 flanqueado en el extremo 5' por un punto de recombinación attB1 y la secuencia Kozak, y flanqueado en el extremo 3' por una secuencia que codifica una etiqueta interior 6 histidina, un codón de parada y el punto de recombinación attB2 (Ruta compatible cDNA9. La primera reacción PCR (en un volumen final de 50 µl) contiene: 24 ng de pCR4 blunt TOPO-INSP002V (plásmido 13075 y Figura 13), 1.5 µl dNTPs (10 mM), 5 µl de 10X búfer polimerasa Pfx, 1 µl MgSO₄ (50 mM), 0.5 µl de cada imprimador específico de gen (100 µM) (INSP002V-EX1 e INSP002V-EX2) y 0.5 µl polimerasa Platinum Pfx DNA (Invitrogen). La reacción PCR se llevó a cabo usando un paso inicial de desnaturalización a 95°C durante 2 minutos, y a continuación se realizaron 12 ciclos de 94°C, 15 segundos y 68°C durante 30 segundos. Los productos PCR se purificaron directamente a partir de la mezcla de la reacción usando el sistema de purificación Wizard PCR prep DNA (Promega) de acuerdo con las instrucciones del fabricante. La segunda reacción de PCR (en un volumen final de 50 µl) contenía 10 µl de producto PCR purificado, 1.5 µl dNTPs (10 mM), 5 µl de 10X búfer polimerasa Platinum Pfx, 0.5 µl de cada imprimador Gateway de conversión (100 µM) (GCP delantero y GCP trasero) y 0.5 µl de polimerasa Platinum Pfx DNA. Las condiciones para la segunda reacción fueron: 95°C durante 1 minuto; 4 ciclos de 94°C, 15 segundos; 45°C, 30 segundos y 68°C durante 3.5 minutos; 25 ciclos de 94°C, 15 segundos; 55°C, 30 segundos y 68°C, 3.5 minutos. Los productos PCR se purificaron tal y como se ha descrito anteriormente.

ii) Subclonación de Gateway compatible INSP002 ORF en un vector de entrada a Gateway pDONR201 y vector de expresión pEAK12d

El segundo paso del proceso de clonación Gateway implica la subclonación del producto PCR modificado Gateway en el vector pDONR201 de entrada Gateway (Invitrogen, figura 16) del siguiente modo: se incubaron 5 µl de producto purificado PCR con 1.5 µl de vector pDONR201 (0.1 µg/µl), 2 µl búfer BP y 1.5 µl de mezcla enzima clonasa BP (Invitrogen) en RT durante 1 hora. La reacción se frenó mediante la adición de proteínaasa K (2 µg) y se incubó a 37°C durante 10 minutos adicionales. Una alícuota de esta reacción (2p1) se transformó en células *E. coli* DH10B mediante electroporación usando un Pulsador Biorard Gene. Los transformantes se colocaron en placas LB-kanamicina. El DNA plásmido mini-prep se preparó a partir de 1-4 de las colonias resultantes usando Wizard Plus SV Minipreps kit (Promega), y 1.5 µl del eluato de plásmido y a continuación se usó en una reacción de recombinación empleando 1.5 µl vector pEAK12d (figura 9) (0.1 µg/µl), 2 µl búfer LR y 1.5 µl de clonasa RN (Invitrogen) en un volumen final de 10 µl. La mezcla se incubó en RT durante 1 hora, se frenó mediante la adición de proteínaasa K (2 µg) y se incubó a 37°C durante 10 minutos adicionales. Una alícuota de esta reacción (1 µl) se usó para transformar células *E. coli* DH10B mediante electroporación.

Los clones que contenían la correcta inserción se identificaron mediante la colonia PCR que actúa tal y como se ha descrito anteriormente excepto que los imprimadores pEAK12d (pEAK12d F y pEAK12d R) se emplearon para el PCR. El DNA plásmido mini prep se aisló de clones que contenían la correcta inserción usando un sistema de robótica Qiaprep Turbo 9600 (Qiagen) o manualmente usando un kit Wizard Plus SV Minipreps (Promega) y la secuencia se verificó empleando los imprimadores pEAK12d F y pEAK12d R.

Se preparó CsCl gradiente purificado maxi-prep DNA de plásmido pEAK12d-INSP002V-6HI(plásmido ID número 13227, figura 17) a partir de 500 ml de cultivo de clones verificados de secuencia (Sambrook J. *et al.*, en Molecular Cloning, A Laboratory Manual, 2ª edición, 1989, Cold Spring Harbor Laboratory Press), resuspendidos en una concentración de 1 µg/µl en agua estéril y se almacenó a -20°C.

iii) Construcción de vector de expresión pEAK12d

El vector pEAK12d es una versión compatible con el Sistema de Clonación Gateway del vector de expresión celular mamífero pEAK12 (adquirido en Edge Biosystems) en el cual el cDNA de interés se expresa bajo el control del promotor humano EF1α. pEAK12d se generó del siguiente modo:

pEAK12d se digirió con enzimas de restricción HindIII y NotI, cuyos extremos se suavizaron con Klenow (New England Biolabs) y desfosforilado usando fosfatasa alcalina de ternero fetal (Roche). Tras el proceso de desfosforilación, el vector fue ligado a la cinta de marco de lectura Gateway con punta no afilada C (sistema de conversión de

ES 2 306 802 T3

vector Gateway, Invitrogen cat no. 11828-019) que contiene puntos de recombinación flanqueantes del gen *ccdB* y con resistencia al cloranfenicol, y se transformaron en células *E. coli* DB3.1 (que permiten la propagación de vectores que contienen el gen *ccdB*). Mini prep DNA se aisló de varias de las colonias resultantes usando un kit Wizard Plus SV Miniprep (Promega) y se digirió con *AseI/EcoRI* para identificar los clones productores de un fragmento de 679 bp, indicando que la cinta había sido insertada en la orientación correcta. El plásmido resultante se llamó pEAK12d (figura 15).

Ejemplo 6

10 *Expresión de células mamíferas y purificación de INSP002-SV-6His V1 (plásmido #13227)*

293 células de riñón embrionario humano que expresan el antígeno nuclear de virus Epstein-Bar (HEK293-EBNA, Invitrogen) se mantuvieron en suspensión en medio ex-celular VPRO libre de suero (reserva de semillas, medio de mantenimiento, JRH). De 16 a 20 horas antes de la transfección (Día-1), las células se sembraron en frascos o matraces de 2 x T225 (50 ml por frasco en DMEM/F12 (1:1) que contenían 2% medio de sembrado FBS (JRH) en una densidad de 2×10^5 células/ml). Al día siguiente (día de la transfección) tuvo lugar la transfección usando un regánete JetPEITM (2 μ l/ μ g de plásmido DNA, PolyPlus-transfección). Para cada frasco, se co-transformaron 113 μ g de DNA (número #13227) con 2.3 μ g de GFP (gen reportero fluorescente). La mezcla de la transfección se añadió a continuación a los frascos de 2 x T225 y se incubó a 37°C (5%CO₂) durante 6 días. Con el fin de aumentar nuestras posibilidades de conseguir más material, repetimos este procedimiento en dos frascos adicionales para generar un total de 200 ml. la confirmación de una transfección positiva se realizó mediante el examen de fluorescencia cuantitativa el día 1 y el día 6 (Axiovert 10 Zeiss).

El día 6 (el día de la cosecha o recogida), los sobrenadantes (200 ml) de los cuatro frascos se reunieron y se centrifugaron (4°C, 400 g) y se colocaron en un recipiente con un único identificador.

Una alícuota (500 pi) se guardó para QC de la proteína etiquetada 6His (bioproceso interno QC).

Proceso de purificación

La muestra del medio de cultivo de 200 ml que contenía la proteína recombinante con una etiqueta 6His C-terminal se diluyó hasta lograr un volumen final de 400 ml con búfer frío (50 mM NaH₂PO₄, 600 mM NaCl; 8.7% (w/v) glicerol, pH 7.5). La muestra se filtró a través de un filtro estéril de 0.22 μ (Millipore, unidad de filtro de 500 ml) y se mantuvo a 4°C en un botella cuadrada con un medio estéril (Nalgene).

La purificación se llevó a cabo a 4°C en la terminal de trabajo VISION (Applied Biosystems) en conexión con un cargador de muestra automático (Labomatic). El procedimiento de purificación estuvo compuesto por dos pasos secuenciales, cromatografía de afinidad de metal en una columna Poros 20 MC (Applied Biosystems) cargada con iones Ni (4.6 x 50 mm, 0.83 ml), seguido de una filtración de gel sobre una columna (1,0 x 10 cm) en un medio Sephadex G-25 (Amersham Pharmacia).

Para el primer paso de cromatografía la columna de afinidad de metal se regeneró con 30 volúmenes de columna de solución EDTA (100 mM EDTA; 1 M NaCl; pH 8.0), se recargó con iones Ni por medio del lavado con 15 volúmenes de columna de una solución de 100 mM NiSO₄, se lavó con 10 volúmenes de columna de búfer A, seguido de 7 volúmenes de columna de búfer B (50 mM NaH₂PO₄; 600 mM NaCl; 8.7% (w/v) glicerol, 400 mM, imidazola, pH 7.5), y finalmente se equilibró con 15 volúmenes de columna de búfer A que contiene 15 mM de imidazola. La muestra se cargó en la columna de afinidad de metal Ni en tandas o grupos de 200 ml. El cargador de muestra Labomatic transfirió 200 ml de la muestra a un circuito cerrado para la muestra y esta muestra se cargó seguidamente en la columna de afinidad de metal Ni a una velocidad de flujo de 10 ml/min. El proceso de transferencia y carga se repitió una vez para cargar la muestra de 400 ml en la columna. Al final del proceso de carga la columna se lavó con volúmenes de 12 columnas de búfer A, seguido de 2 volúmenes de columna de búfer A que contiene 20 mM de imidazola. Durante el lavado de 20 mM de imidazola las proteínas contaminantes que estaban ligeramente unidas se eluyeron de la columna. La proteína recombinante etiquetada His finalmente se eluyó con 10 volúmenes de columna de búfer B a una velocidad de flujo de 2 ml/min, y la proteína eluida se recogió en una fracción de 1.6 ml.

Para el segundo paso de cromatografía, la columna de filtración de gel Sephadex G-25 se regeneró con 2 ml de búfer C (137 M NaCl; 2.7 mM KCl; 1.5 mM KH₂PO₄; 8 mM NaH₂PO₄; 20% (w/v) glicerol, pH 7.4). La fracción pico eluida de la columna Ni automáticamente, por medio del cargador de muestra integrado en el VISION, se cargó en la columna Sephadex G-25 y la proteína se eluyó con búfer C a un ritmo de flujo de 2 ml/min. La muestra desalada se recuperó en una fracción de 2.2. La fracción se filtró a través de un filtro de centrifugación estéril de 0.22 μ (Millipore), se alícuó, congeló y almacenó a -80°C. Una alícuota de la muestra se analizó en una transferencia del tipo Western sobre SDS-PAGE (2-14% gel NuPAGE; Novex) con anticuerpos anti-His.

Tras la electroforesis, las proteínas fueron electrotransferidas a del gel a una membrana de nitrocelulosa a 290 mA durante una hora a 4°C. La membrana se bloqueó con 5% de polvo de leche en búfer E (137 M NaCl; 2.7 mM KCl; 1.5 mM KH₂PO₄; 8 mM NaH₂PO₄; 0.1% Tween 20, pH 7.4) durante una hora a temperatura ambiente, y a continuación se incubó con una mezcla de dos anticuerpos anti-His policlonales de conejo (G-18 y H-15, 0.2 μ g/ml cada uno; Santa Cruz) en 2.5% de polvo de leche en búfer E durante toda la noche a 4°C. Tras una incubación adicional de una hora a

temperatura ambiente, la membrana se lavó con búfer E (3 x 10 minutos), y a continuación se incubó con un anticuerpo secundario conjugado HRP anti-conejo (DAKO, HRP 0399) diluido 1/300 en búfer E que contenía 2.5% de polvo de leche durante 2 horas a temperatura ambiente. Tras el lavado con búfer E (3 x 10 minutos), la membrana se desarrolló con el kit ECL (Amersham Pharmacia) durante un minuto. La membrana se expuso a continuación a un Hyperfilm (Amersham Pharmacia), la capa se desarrolló y la imagen de la transferencia de tipo Western se analizó visualmente.

Ejemplo 7

Clonación de INSP002 de longitud completa del corazón

La secuencia codificadora de longitud completa de INSP002 se clonó de cDNA del corazón del siguiente modo:

i) Generación de plantilla cDNA de corazón

Se compró RNA total del corazón humano en Clontech. La calidad y concentración de RNA se analizó usando un Bioanalizador Agilent 2100.

Para la síntesis de cDNA la reacción de la mezcla contenía: 1 μ l oligo (dT)15 imprimador (500 μ g/ml, Promega cat. No. C1101), 2 μ g total RNA, 1 μ l dNTPs (10 mM) en un volumen de 12 μ l. La mezcla se calentó a 65°C durante 5 minutos y a continuación se enfrió en hielo. A continuación se añadieron los siguientes agentes: 4 μ l 5X búfer primera cadena, 2 μ l DTT (0.1 M), 1 μ l inhibidor de ribonucleasa recombinante RNaseOut (40 unidades/ μ l, Promega, cat. No. N 2511) y se incubaron a 42°C durante 2 minutos antes de la adición de 1 μ l (200 unidades) de Superscript II (Invitrogen cat. No. 18064-014). La mezcla se incubó a 42°C durante 50 minutos y a continuación se calentó a 70°C durante 15 minutos. Para retirar la plantilla de RNA, se añadió 1 μ l (2 unidades) de *E. coli* RNasa H (Invitrogen cat. No. 18021-014) y la mezcla de la reacción se incubó de nuevo a 37°C durante 20 minutos. La mezcla con la reacción final se diluyó en 200 μ l con agua estéril y se almacenó a -80°C.

ii) Clonación de la secuencia codificadora de longitud completa de INSP002 mediante PCR

La secuencia codificadora de longitud completa de INSP002 se clonó a partir de cDNA de corazón humano mediante PCR en una mezcla de reacción PCR de 50 μ l que contenía 5 μ l de cDNA de corazón, 5 μ l 10X búfer Pfx, 1.5 μ l dNTPs (10mM), 1 μ l MgSO₄ (50 mM), 1.5 μ l de imprimador delantero específico de gen INSP002-FL-F (10 μ M), 1.5 μ l de imprimador trasero específico de gen INSP002-FL-R (10 μ M) y 0.5 μ l de polimerasa Platinum Pfx DNA (Invitrogen). Las condiciones del ciclo fueron 1 ciclo de 94°C, 4 minutos; 35 ciclos de 94°C, 15 segundos; 55°C, 30 segundos; 68°C, 1 minuto; 1 ciclo de 68°C, 10 minutos seguido de un ciclo de mantenimiento a 4°C.

Los productos de amplificación se visualizaron en geles de azarosa 0.8% en búfer 1 X TAE (Invitrogen) y los productos PCR migratorios en la masa molecular pronosticada (589 bp) se purificaron del gel utilizando el kit de Extracción de Gel Qiagen MinElute (Qiagen). Los productos PCR se eluyeron en 50 μ l de 10 mM Tris-HCl pH 8.5 y se subclonaron en vector pCR4 blunt TOPO tal y como se ha descrito anteriormente (sección 1.4). Varias colonias resistentes a la ampicilina se sometieron a PCR de colonia tal y como se ha descrito en la sección 1.5. Las colonias que contenían la inserción con el correcto tamaño (589 bp + 106 bp debido a MCS) crecieron durante la noche a 37°C en 5 ml L-Broth (LB) que contenía ampicilina (100 μ g/ml), con un proceso de agitación a 220 rpm a 37°C. Se preparó Miniprep plásmido DNA a partir de cultivos de 5 ml usando un sistema de robótica Qiaprep Turbo 9600 (Qiagen) o el kit Wizard Plus SV Miniprep (Promega cat. No. 1460) de acuerdo con las instrucciones del fabricante y se secuenciaron 200-500 ng de mini-prep DNA tal y como se describe en la sección 1.6 con los imprimadores T3 y T7 (Tabla III). La secuencia clonada se da en la figura 18. El mapa del plásmido resultante, pCR4-blunt TOPO-INSP002FL (plásmido ID. No. 13514) se muestra en la Figura 19.

TABLA I

Bibliotecas de cDNA humano

Biblioteca	Fuente de tejido/célula	Vector	Variedad de huésped	Suministrador	Nº cat.
1	Cerebro fetal humano	Zap II	XL1-Blue MRF'	Stratagene	936206
2	Ovario humano	GT10	LE392	Clontech	HL1098a
3	Pituitaria humana	GT10	LE392	Clontech	HL1097a
4	Placenta humana	GT11	LE392	Clontech	HL1075b
5	Testículo humano	GT11	LE392	Clontech	HL1010b
6	Sustancia negra humana	GT10	LE392	in house	
7	Cerebro fetal humano	GT10	LE392	in house	
8	Corteza cerebral humana	GT10	LE392	in house	
9	Colon humano	GT10	LE392	Clontech	HL1034a
10	Cerebro fetal humano	GT10	LE392	Clontech	HL1065a

ES 2 306 802 T3

TABLA I (continuación)

Biblioteca	Fuente de tejido/célula	Vector	Variedad de huésped	Suministrador	Nº cat.
11	Pulmón fetal humano	GT10	LE392	Clontech	HL1072a
12	Riñón fetal humano	GT10	LE392	Clontech	HL1071a
13	Hígado fetal humano	GT10	LE392	Clontech	HL1064a
14	Médula humana	GT10	LE392	Clontech	HL1058a
15	Monocitos sanguíneos periféricos humanos	GT10	LE392	Clontech	HL1050a
16	Placenta humana	GT10	LE392	in house	
17	SHSY5Y humano	GT10	LE392	in house	
18	Línea celular U373 humana	GT10	LE392	in house	
19	Línea celular CFPoc-1 humana	Uni Zap	XL1-Blue MRF'	Stratagene	936206
20	Retina humana	GT10	LE392	Clontech	HL1132a
21	Vejiga urinaria humana	GT10	LE392	in house	
22	Plaquetas humanas	Uni Zap	XL1-Blue MRF'	in house	
23	Neuroblastoma humano Kan + TS	GT10	LE392	in house	
24	Músculo suave bronquial humano	GT10	LE392	in house	
25	Músculo suave bronquial humano	GT10	LE392	in house	
26	Timo humano	GT10	LE392	Clontech	HL1127a
27	Bazo humano tramo 5'	GT11	LE392	Clontech	HL1134b
28	Monocitos sanguíneos periféricos humanos	GT10	LE392	Clontech	HL1050a
29	Testículo humano	GT10	LE392	Clontech	HL1065a
30	Cerebro fetal humano	GT10	LE392	Clontech	HL1065a
31	Sustancia negra humana	GT10	LE392	Clontech	HL1093a
32	Placenta humana #11	GT11	LE392	Clontech	HL1075b
33	Cerebro fetal humano	GT10	LE392	Clontech	custom
34	Placenta humana #59	GT10	LE392	Clontech	HL5014a
35	Pituitaria humana	GT10	LE392	Clontech	HL1097a
36	Páncreas humano #19	Uni Zap XR	XL1-Blue MRF'	Stratagene	937208
37	Placenta humana #19	GT11	LE392	Clontech	HL1008
38	Hígado humano tramo 5'	GT11	LE392	Clontech	HL1115b
39	Útero humano	Zap-CMV XR	XL1-Blue MRF'	Stratagene	980207
40	Biblioteca cDNA con gran inserción de riñón humano	TriplEx2	XLI-Blue	Clontech	HL5507u

TABLA II

Imprimadores de Clonación INSP002

Imprimador	Secuencia (5'-3')
INSP002-CP	CTC AGC CAT ACG CCT CCG AA
INSP002-CP2	GCT GAG CTG CCA GTG AGA CA
INSP002V-5'-F	ACC TGG AAG GAA GCG ACT GCA CTG A
INSP002V-5'-R	GCA GCC GGG CCG GGA GAG AAC GAA GGG CAC AGC CTT A
INSP002V-3'-F	AGG CTG TGC CCT TCG TTC TCT CCC GGC CCG GCT GCT C

ES 2 306 802 T3

TABLA II (continuación)

Imprimador	Secuencia (5'-3')
INSP002V-3'-R	ACT CCA GGA CGG GCA CTG TGT CTA C
INSP002V-5'nest-F	GTC GAC TGC TAG TGA CCT TGA G
INSP002V-3'nest-R	ACA TCA TCC AGG TCC ACG TCT T

TABLA III

Imprimadores para INSP002 subclonación y secuencias

Imprimador	Secuencia (5'-3')
SP6	ATT TAG GTG ACA CTA TAG
T7	TAA TAC GAC TCA CTA TAG GG
T3	ATT AAC CCT CAC TAA AGG GA
M13F	TGT AAA ACG ACG GCC AGT
pEAK 12-F	GCC AGC TTG GCA CTT GAT GT
pEAK 12-R	GAT GGA GGT GGA CGT GTC AG
MSP002V-EX1	AA GCA GGC TTC <u>GCC ACC</u> ATG CTC CTT GGC CAG CTA TC
INSP002V-EX2	<i>GTG ATG GTG ATG GTG</i> TGC TTT TGG GCT GCA GTG AC
GCP Forward	G GGG ACA AGT TTG TAC AAA AAA GCA GGC TTC <u>GCC ACC</u>
GCP Reverse	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT TCA <i>ATG GTG ATG GTG ATG GTG</i>

TABLA IV

Imprimadores PCR para clonación de NPS002 de longitud completa

Imprimador	Secuencia (5'-3')
INSP002-FL-F	GAT GCT CCT TGG CCA GCT AT
INSP002-FL-R	CCA TCC ACG ATG CTC AGT TC

Secuencia subrayada = Secuencia Kozak

Negrita = Codón de parada

Secuencia en *cursiva* = Su etiqueta

ES 2 306 802 T3

Listado secuencial

Nota: para los aminoácidos codificados por la unión exón-exón, el aminoácido se asignará al exón de más de 5'.

5 SEQ ID NO:1 (INSP002 - Exón de secuencia de nucleótido 1)

```

1 AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTC
51 AGAAGGCTCA GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC
101 TTGAGCCCAG TCCGGACAGA CAGACAGGCA GACAGACGCA CGGACAAGCA
151 GATGCTCCTT GGCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCC
201 TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC
251 TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCC
301 ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCCTGC
351 AGAAAGCCAG GCGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG
401 GTGGCTGCTG TCACTCTGCC GCTGAACCCT CAGGAAGTGA TCCAGGGGAT
451 GTGTAAGGCT GTGCCCTTCG TTCAG

```

20 SEQ ID NO:2 (INSP002 - Exón de secuencia de proteína 1)

```

1 MLLGQLSTLL CLLSGALPTG SGRPEPQSPR PQSWAAANQT WALGPGALPP
51 LVPASALGSW KMFGLQKAR QLGMRRLQRG QDEVAAVTLP LNPQEVIOGM
101 CKAVPFVQ

```

SEQ ID NO:3 (INSP002 - Exón de secuencia de nucleótido 2)

```

1 GTGTTCTCCC GCGCCGGCTG CTCAGCCATA CGCCTCCGAA ATCATCTGTG
51 CTTTGGTCAT TGTCTCTCTC TCTACATCCC TGGCTCGGAC CCCACCCAC
101 TAGTCCTGTG CAACAGCTGT ATGCCTGCTC GCAAGCGTTG GGCACCCGTG
151 GTCCTGTGGT GTCTCACTGG CAGCTCAGCC TCCCGTCGAC GGGTGAAGAT
201 ATCCACCATG CTGATCGAGG GGTGTCAGT CAGCCCAAAA GCATGA

```

SEQ ID NO:4 (INSP002 - Exón de secuencia de proteína 2)

```

1 VFSRPGCSAI RLNNHLFCFH CSSLYIPGSD PTPLVLCNSC MPARKRWAPV
51 VLWCLTGSSA SRMRVKISTM LIEGCHCSPK A

```

SEQ ID NO:5 (INSP002 - Secuencia de nucleótido)

```

1 AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTC
51 AGAAGGCTCA GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC
101 TTGAGCCCAG TCCGGACAGA CAGACAGGCA GACAGACGCA CGGACAAGCA
151 GATGCTCCTT GGCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCC
201 TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC
251 TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCC
301 ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCCTGC
351 AGAAAGCCAG GCGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG
401 GTGGCTGCTG TCACTCTGCC GCTGAACCCT CAGGAAGTGA TCCAGGGGAT
451 GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCGGCCC GGCTGCTCAG
501 CCATACGCCT CCAAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTCTAC
551 ATCCCTGGCT CGAACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCC
601 TGCTCGCAAG CGTGGGCAC CCGTGGTCTT GTGGTGTCTC ACTGGCAGCT
651 CAGCCTCCCG TCCAGGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGT
701 CACTGCAGCC CAAAGCATG A

```


ES 2 306 802 T3

SEQ ID NO:6 (INSP002 - Secuencia de proteína)

```

1  MLLGQLSTLL CILSGALPTG SGRPEPQSPR PQSWAAANQT WALGPGALPP
5  51  LVPASALGSW KAFLGLQKAR QLGMGRLQRG QDEVAAVTLP LNPQEVIOGM
101 CKAVPFVQVF SRPGCSAIRL RNHLCFGHCS SLYIPGSDPT PLVLCNSCMP
151 ARKRWAPVVL WCLTGSSASR RRVKISTMLI EGCHCSPKA

```

SEC ID NO:7 (INSP002 - Exón de secuencia de proteína I sin péptido señalizador)

```

1  RPEPQSPRPQ SWAAANQTWA LGPGALPPLV PASALGSWKA FLGLQKARQL
15  51  GMGRLQRGQD EVAAVTLPLN PQEVIQGMCK AVPFVQ

```

SEC ID NO:8 (INSP002 - Secuencia de proteína sin señal señalizadora)

```

1  RPEPQSPRPQ SWAAANQTWA LGPGALPPLV PASALGSWKA FLGLQKARQL
20  51  GMGRLQRGQD EVAAVTLPLN PQEVIQGMCK AVPFVQVFSR PGCSAIRLRN
101 HLCFGHCSSL YIPGSDPTPL VLCNSCMPAR KRWAPVVLWC LTGSSASRRR
151 VKISTMLIEG CHCSPKA

```

SEC ID NO:9 (Secuencia de nucleótido codificadora de secuencia de proteína INSP002 sin péptido señalizador - polipéptido maduro INSP002)

```

1  AGGCCTGAAC CCCAGTCTCC TCGACCTCAG TCCTGGGCTG CAGCCAATCA GACCTGGGCT
30  61  CTGGGCCAG GGGCCCTGCC CCCACTGGTG CCAGCTTCTG CCCTTGGGAG CTGGAAGGCC
121 TTCTTGGGCC TGCAGAAAGC CAGGCAGCTG GGGATGGGCA GGCTGCAGCG TGGGCAAGAC
181 GAGGTGGCTG CTGTGACTCT GCCGCTGAAC CCTCAGGAAG TGATCCAGGG GATGTGTAAG
241 GCTGTGCCCT TCGTTCAGGT GTTCTCCCGG CCCGGCTGCT CAGCCATACG CCTCCGAAAT
301 CATCTGTGCT TTGGTCATTG CTCCTCTCTC TACATCCCTG GCTCGGACCC CACCCCACTA
35  361 GTCCTGTGCA ACAGCTGTAT GCCTGCTCGC AAGCGTTGGG CACCCGTGGT CCTGTGGTGT
421 CTCACCTGGCA GCTCAGCCTC CCGTCGACGG GTGAAGATAT CCACCATGCT GATCGAGGGG
461 TGTCACCTGCA GCCCAAAGC ATGA

```

SEC ID NO:10 (Secuencia de nucleótido codificadora de proteína exón 1 INSP002 sin péptido señalizador - polipéptido exón 1 maduro INSP002)

```

1  AGGCCTGAAC CCCAGTCTCC TCGACCTCAG TCCTGGGCTG CAGCCAATCA GACCTGGGCT
45  61  CTGGGCCAG GGGCCCTGCC CCCACTGGTG CCAGCTTCTG CCCTTGGGAG CTGGAAGGCC
121 TTCTTGGGCC TGCAGAAAGC CAGGCAGCTG GGGATGGGCA GGCTGCAGCG TGGGCAAGAC
181 GAGGTGGCTG CTGTGACTCT GCCGCTGAAC CCTCAGGAAG TGATCCAGGG GATGTGTAAG
241 GCTGTGCCCT TCGTTCAG

```

SEC ID NO:11 (Secuencia de nucleótido codificadora de proteína INSP002 sin la región no traducida 5')

```

1  ATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCCT
55  51  GCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCTT
101 GGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCCA
151 CTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCTGCA
201 GAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAGG
251 TGGCTGCTG TGACTCTGCC GCTGAACCC CAGGAAGTGA TCCAGGGGATG
301 TGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCGGCCG GGCTGCTCAGC
60  351 CATACGCCT CCGAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTCTACA
401 TCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCCT
451 GCTCGCAAG CGTTGGGCAC CCGTGGTCTT GTGGTGTCTC ACTGGCAGCTC
501 AGCCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGTC
551 ACTGCAGCC CAAAAGCATG A

```

ES 2 306 802 T3

SEQ ID NO:12 (Secuencia de nucleótido codificadora de proteína exón 1 INSP002 sin la región no traducida 5')

```

1   ATGCTCCTT GCCAGCTAT CCACTCTTCT GTGCTGCTT AGCGGGGCCCT
51  GCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCCT
101 GGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCAGGGG CCCTGCCCCCA
151 CTGGTGCCA GCTTCTGCC TTGGGAGCTG GAAGGCCCTC TTGGGCTGCA

201   GAAAGCCAG GCAGCTGGG ATGGGCAGGC TGCAGCGTG GCAAGACGAGG
251   TGGCTCTG TGACTCTGCC GCTGAACCT CAGGAAGTGA TCCAGGGGATG
301   TGTAAGGCT GTGCCCTTCG TTCAG

```

SEQ ID NO:13 (Secuencia de nucleótido codificadora de la variante polipéptido INSP002)

```

1   GTCGACTGCT AGTGACCTTG AGCCAGTCC GGACAGACAG ACAGGCAGAC AGACGCACGG
61  ACAAGCAGAT GCTCCTTGGC CAGCTATCCA CTCTTCTGTG CTTGCTTAGC GGGGCCCTGC
121 CTACAGGCTC AGGGAGGCCT GAACCCAGT CTCTCGACC TCAGTCTGG GCTGCAGCCA
181 ATCAGACCTG GGCTCTGGG CCAGGGGGCC TGCCCCACT GGTGCCAGCT TCTGCCCTTG
241 GGAGCTGGAA GGCTTCTTG GGCCTGCAGA AAGCCAGGCA GCTGGGGATG GGCAGGCTGC
301 AGCGTGGGCA AGACGAGGTG GCTGCTGTGA CTCTGCCGT GAACCTCAG GAAGTGATCC
361 AGGGGATGTG TAAGGCTGTG CCTTCTGTC TCTCCCGCC CGGTGCTCA GCCATACGCC
421 TCCGAAATCA TCTGTGCTTT GGTCAATTGCT CCTCTCTCTA CATCCCTGGC TCGGACCCCA
481 CCCCACTAGT CCTGTGCAC AGCTGTATGC CTGCTGCAA GCGTTGGGCA CCCGTGGTCC
541 TGTGGTGTCT CACTGGCAGC TCAGCTCCC GTGACGGGT GAAGATATCC ACCATGCTGA
601 TCGAGGGGTG TCAGTGCAGC CAAAAGCAT GAAGTGAAG TCTGGATGGG TGCACGGAGA
661 CACGCACCTT GGAGAAATGA GGGGAGATGG ACCAAGAAAG ACGTGGACCT GGATGATGT

```

SEQ ID NO:14 (Variante polipéptido INSP002)

```

1   MLLGQLSTLL CLLSGALPTG SGRPEQSPR PQSWAANQT WALGPGALPP LVPASALGSW
61  KAFGLQKAR QLGMGRLQRG QDEVAATLP LNPQEVIGM CKAVPFVLSR PGCSAIRLRN
121 HLCFGHCSSL YIPGSDPTPL VLCNSCMPAR KRWAPVVLWC LTGSSASRRR VKISTMLIEG
161 CHCSPKA

```

SEQ ID NO:15 (Secuencia de nucleótido codificadora de la variante polipéptido INSP002 sin la región no traducida 5')

```

1   ATGCTCCTTG GCCAGCTATC CACTCTTCTG TGCTGCTTA GCGGGGCCCT GCCTACAGGC
61  TCAGGGAGGC CTGAACCCCA GTCTCCTCGA CCTCAGTCCT GGGCTGCAGC CAATCAGACC
121 TGGGCTCTGG GCCAGGGGC CTGCCCCCA CTGGTGCCAG CTTCTGCCCT TGGGAGCTGG
181 AAGGCCTTCT TGGGCTGCA GAAAGCCAGG CAGCTGGGGA TGGGAGGCT GCAGCGTGGG
241 CAAGACGAGG TGGCTGCTGT GACTCTGCCG CTGAACCTC AGGAAGTAT CCAGGGGATG
301 TGTAAGGCTG TGCCCTTCGT TCTCTCCCG CCCGGTGTCT CAGCCATACG CCTCCGAAAT
361 CATCTGTGCT TTGGTCATTG CTCTCTCTC TACATCCCTG GCTCGGACCC CACCCCACTA
421 GTCTGTGCA ACAGCTGTAT GCCTGTCTGC AAGCGTTGGG CACCCGTGGT CCTGTGGTGT
481 CTCAGTGGA GCTCAGCCTC CCGTCGACGG GTGAAGATAT CCACCATGCT GATCGAGGGG
541 TGTCACTGCA GCCCAAAGC A

```

REIVINDICACIONES

1. Un polipéptido que tiene la función de un miembro de la subfamilia DAN de citoquinas de pliegue con nudo de cistina, donde dicho polipéptido:
 - (i) comprende o consiste en la secuencia de aminoácido tal y como se establece en SEQ ID NO:6, SEQ ID NO:8 o SEQ ID NO:14;
 - (ii) es un fragmento del polipéptido de (i); o
 - (iii) tiene una identidad secuencial superior al 70% con los polipéptidos de (i) y (ii).
2. Un polipéptido como el establecido en la reivindicación 1, que consiste en 7 o más, por ejemplo 8, 10, 12, 14, 16, 18, 20, o más residuos consecutivos de aminoácidos procedentes de la secuencia de aminoácido establecida en SEQ ID NO:6, SEQ ID NO:8 o SEQ ID NO:14.
3. Una molécula de ácido nucleico purificada que codifica un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2.
4. Una molécula de ácido nucleico purificada de acuerdo con la reivindicación 3, que comprende o consiste en la secuencia de ácido nucleico tal y como se establece en SEQ ID NO:5, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13 o SEQ ID NO:15.
5. Una molécula de ácido nucleico purificada de acuerdo con el polipéptido que tiene la función de un miembro de la subfamilia DAN de citoquinas de pliegue con nudo de cistina, que se hibridiza bajo condiciones elevadamente rigurosas con una molécula de ácido nucleico de acuerdo con la reivindicación 3 o reivindicación 4.
6. Un vector que comprende una molécula de ácido nucleico tal y como se ha establecido en las reivindicaciones 3 a 5.
7. Una célula huésped transformada con un vector de acuerdo con la reivindicación 6.
8. Un anticuerpo que se enlaza específicamente, y que preferiblemente inhibe la actividad de la subfamilia DAN, con un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2.
9. Un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5, un vector de acuerdo con la reivindicación 6, o un anticuerpo de acuerdo con la reivindicación 8, para uso en terapia o diagnóstico de enfermedad.
10. Un método de diagnosticar una enfermedad en un paciente, que consiste en analizar *in vitro* el nivel de expresión de un gen natural codificador de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, o analizar la actividad de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, en tejido procedente de dicho paciente y comparar dicho nivel de expresión o actividad con un nivel de control, donde un nivel que es diferente a dicho nivel de control es indicativo de enfermedad.
11. Un método de acuerdo con la reivindicación 10, que comprende los pasos de: (a) poner en contacto un anticuerpo de acuerdo con la reivindicación 8 con una muestra biológica bajo condiciones adecuadas para la formación de un complejo ligando- polipéptido; y (b) detectar dicho complejo.
12. Un método de acuerdo con la reivindicación 10, que comprende los siguientes pasos:
 - (a) poner en contacto una muestra de tejido del paciente con una sonda de ácido nucleico bajo condiciones rigurosas que permitan la formación de un complejo híbrido entre una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5 y la sonda;
 - (b) poner en contacto una muestra de control con dicha sonda bajo las mismas condiciones empleadas en el paso a); y
 - (c) detectar la presencia de complejos híbridos en dichas muestras; donde la detección de niveles del complejo híbrido en la muestra del paciente que difieren de los niveles del complejo híbrido en la muestra control es indicativo de enfermedad.
13. Un método de acuerdo con la reivindicación 10, que comprende:
 - a) poner en contacto una muestra de tejido del paciente con un imprimador de ácido nucleico bajo condiciones rigurosas que permitan la formación de un complejo híbrido entre una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5 y el imprimador;

b) poner en contacto una muestra de control con dicho imprimador bajo las mismas condiciones empleadas en el paso a); y

c) amplificar el ácido nucleico que sirve de muestra; y

d) detectar el nivel de ácido nucleico amplificado tanto de las muestras del paciente como de las muestras control; donde la detección de niveles de ácido nucleico amplificado en la muestra del paciente que difieren de manera significativa de los niveles de ácido nucleico en la muestra de control es indicativo de enfermedad.

14. Un método de acuerdo con las reivindicaciones 10 a 13, donde dicha enfermedad es un desorden proliferativo celular, desorden inmune/inflamatorio, desorden cardiovascular, desorden neurológico, desorden de desarrollo, desorden metabólico, infección u otra condición patológica.

15. Uso de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2 como un miembro de la subfamilia DAN de citoquinas de pliegue con nudo de cistina.

16. Una composición farmacéutica que comprende un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5, un vector de acuerdo con la reivindicación 6, o un anticuerpo de acuerdo con la reivindicación 8.

17. Una composición para una vacuna consistente en un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2 o una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5.

18. Uso de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5, un vector de acuerdo con la reivindicación 6, un anticuerpo de acuerdo con la reivindicación 8, o una composición farmacéutica de acuerdo con la reivindicación 16, en la fabricación de un medicamento para el tratamiento de desórdenes proliferativos celulares, desórdenes inmunes/inflamatorios, desórdenes cardiovasculares, desórdenes neurológicos, desórdenes de desarrollo, desórdenes metabólicos, infecciones y otras condiciones patológicas.

19. Un método para controlar el tratamiento terapéutico de enfermedad en un paciente *ex vivo*, que consiste en controlar durante un periodo de tiempo el nivel de expresión o actividad de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, o el nivel de expresión de molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5 en tejido de dicho paciente, donde la alteración de dicho nivel de expresión o actividad durante el periodo de tiempo hacia un nivel de control es indicativo de regresión de dicha enfermedad.

20. Un método para identificar un compuesto que es efectivo o eficaz en el tratamiento y/o diagnóstico de enfermedad, que consiste en contactar un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2, o una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5 con uno o más compuestos sospechosos de poseer afinidad de enlace para dicho polipéptido o molécula de ácido nucleico, y seleccionar un compuesto que se enlace específicamente con dicha molécula de ácido nucleico o polipéptido.

21. Un kit útil para diagnosticar una enfermedad y que consiste en un primer envase que contiene una sonda de ácido nucleico que se hibridiza bajo condiciones rigurosas con una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5; un segundo envase que contiene imprimadores útiles para amplificar dicha molécula de ácido nucleico; e instrucciones para usar la sonda y los imprimadores con el fin de facilitar el diagnóstico de la enfermedad.

22. El kit de la reivindicación 21, que además consiste en un tercer envase que contiene un agente para digerir RNA no hibridizado.

23. Un kit que comprende una selección de moléculas de ácido nucleico, donde al menos una de las cuales es una molécula de ácido nucleico de acuerdo con las reivindicaciones 3 a 5.

24. Un kit que comprende uno o más anticuerpos que se enlazan con un polipéptido tal y como se establece en la reivindicación 1 o reivindicación 2; y un reagente útil para la detección de una reacción de enlace entre el anticuerpo y dicho polipéptido.

25. Un animal no humano transgénico o knockout que se ha transformado para expresar niveles superiores, inferiores o niveles ausentes de un polipéptido de acuerdo con la reivindicación 1 o reivindicación 2.

26. Un método para analizar un compuesto efectivo para tratar una enfermedad, contactando un animal no humano transgénico de acuerdo con la reivindicación 25 con un compuesto candidato y determinando el efecto del compuesto en la enfermedad del animal.

Figura 1: Base de datos BLAST no redundante Top Diez NCBI golpea contra INSP002

Consulta= INSP002

(189 letras)

Base de datos: ncbi-nt

1,021,646 secuencias; 4,527,726,597 letras totales

Búsquedahecho

		Registro	z
Secuencias que producen alineaciones importantes:		(bits)	Valor
gi 15217064 gb AF400435.1 AF400435	Relacionada con el cerberus de Homo sapiens	69	2e-10
gi 11427432 ref XM_005320.1	Cerberus 1 de Homo sapiens (Xenopus la...	69	2e-10
gi 4885134 ref NM_005454.1	Cerberus 1 de Homo sapiens (Xenopus lae...	69	2e-10
gi 5802091 gb AF139721.1 AF139721	homólogo de cerberus de gallus gallus ..	68	3e-10
gi 5902633 gb AF179484.1 AF179484	Caronte de gallus gallus (CAR) mR...	67	5e-10
gi 2853615 gb AF035579.1 AF035579	homólogo de cerberus mus musculus ...	61	5e-08
gi 6753409 ref NM_009887.1	Homólogo de cerberus 1 mus musculus (Xen...	61	5e-08
gi 2654026 gb AF031896.1 AF031896	Relacionado con cerberus mus musculus ...	61	5e-08
gi 1513087 gb U64831.1 XLU64831	fac. que induce xenopus laevis de cabeza ...	57	9e-07
gi 6679017 ref NM_008675.1	neuroblastoma mus musculus, supres ...	55	3e-06

Figura 2: Alineación con el golpe superior, Cerebro de Homo sapiens relacionado 1

```
>gi|15217064|gb|AF400435.1|AF400435 Homo sapiens cerberus-relacionado 1
      (CER1) ARNm, cds completo
      Longitud = 804

      Registro = 68.9 bits (167), Esperado = 2e-10
      Identidades = 36/102 (35%), Positivos = 53/102 (51%)
      Estructura = +1

Consulta: 87 VTLPLNPQEVIOGMCKAVPPVQVFSRPGCSAIRLRNHLCPGHCSLYIPGSDPTPLVLCN 146
      V LP+ EV C+ VPF Q + GC + ++N+LCPG C S++ PG+ C+

Sbjct: 442 VILPIKSHEVHWETCRTVPFSQITIEGCEKVVVQNNLCFGKCGSVHFPGAQAQHSHTCS 621

Consulta: 147 SCMPARKRNAPVVLWCLTGSSAERRRVKISTMLIEGCHCSFK 188
      C+PA+ + L C SS +K+ NL+E C C K

Sbjct: 622 HCLPAKFTTMHLPLNCTELSSV----IKV-VNLVRECCQCKVK 732
```

Figura 3: Base de datos BLAST Top Cuatro NCBI-nr y NCBI-nt golpea contra INSP002

i) ncbi-nr

Base de datos: Todos no redundantes GenBank CDS
 traducciones +PDB+SwissProt+PIR+PRF
 1,242,768 secuencias ; 395,571,179 letras totales

Búsquedahecho

Secuencia que produce alineaciones importantes:			Registro (bits)	E Valor
ref NP_689867.1	proteína hipotética	FLJ38607 [Homo sapiens] >g...	394	e-109
ref XP_113990.1	proteína hipotética	XP_113990 [Homo sapiens]	221	5e-57
ref XP_164658.1	similar a la proteína hipotética	FLJ38607 [Homo ...	70	1e-11
ref NP_005445.1	cerberus 1; cerberus-relacionado 1; cerberus 1 (Xen...		69	3e-11
gb AAD51610.1	AP139721_1 cerberus homólogo [Gallus gallus]		68	6e-11

ii) ncbi-nt

Base de datos: Todas secuencias GenBank+EMBL+DDBJ+PDB (pero no EST, STS, GSS,
 o fase 0, 1 ó 2 HTGS secuencias)
 1,469,831 secuencias ; 7,238,625,236 letras totales

Búsquedahecho

Secuencia que produce alineaciones importantes:			Registro (bits)	E Valor
ref NM_152654.1	Homo sapiens proteína hipotética	FLJ38607 (FLJ...	394	e-108
dbj AK095926.1	Homo sapiens ADNc	FLJ38607 fis, clón HEART2004941	394	e-108
gb BC025333.1	Homo sapiens, clón	IMAGE:4558384, ARNm, partial cds	221	e-103
ref XM_113990.1	Homo sapiens LOC199699 (LOC199699), ARNm		221	e-103
gb AC092069.2	Homo sapiens cromosoma 19 clón	CTC-425F1, compl...	222	1e-56

Figura 4: Alineación de INSP002 con AK095926.1

```

>dbj|AK095926.1| Homo sapiens ADNg FLJ38607 fis, clón HEART2004941
      Longitud = 1731

  Registro= 394 bits (1013), Esperado = e-108
  Identidades = 189/189 (100%), Positivo = 189/189 (100%)
  Estructura = +2

Consulta: 1  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW 60
             MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW
Sbjct: 44  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW 223

Consulta: 61  KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQVFSRPGCSAIRL 120
             KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQVFSRPGCSAIRL
Sbjct: 224  KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQVFSRPGCSAIRL 403

Consulta: 121 RNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI 180
             RNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
Sbjct: 404  RNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI 583

Consulta: 181 EGCHCSPKA 189
             EGCHCSPKA
Sbjct: 584  EGCHCSPKA 610

```

Figura 5: Alineación de INSP002 con IMAGE: 4558384

```

>gb|BC025333.1| Homo sapiens, clón IMAGE:4558384, ARNm, cds parcial
      Longitud = 1746

  Registro= 221 bits (562), Esperado(2) = e-103
  Identidades = 108/108 (100%), Positivo = 108/108 (100%)
  Estructura = +1

Consulta: 1  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW 60
             MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW
Sbjct: 187  MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTNALGPQALPPLVPASALGSW 366

Consulta: 61  KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQ 108
             KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQ
Sbjct: 367  KAFLGLQKARQLGMRQLRGQDEVAAVTLPLNPQEVIGMCKAVPFVQ 510

  Registro= 178 bits (451), Esperado(2) = e-103
  Identidades = 81/81 (100%), Positivo = 81/81 (100%)
  Estructura = +2

Consulta: 109 VFSRPGCSAIRLRNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA 168
             VFSRPGCSAIRLRNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA
Sbjct: 596  VFSRPGCSAIRLRNHLCPGHCSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSA 775

Consulta: 169 SRRRVKISTMLIEGCHCSPKA 189
             SRRRVKISTMLIEGCHCSPKA
Sbjct: 776  SRRRVKISTMLIEGCHCSPKA 838

```


Figura 6: Secuencia de nucleótido INSP002 con traslación

```

1  AAATGCCTCC CAGGCTATCC AGGAGGGGCC AAGAGATTAA AAGCAGGTTT AGAAGGCTCA
61  GATGCCACTC ACCAGACAGC AGGGTCGACT GCTAGTGACC TTGAGCCCAG TCCGGACAGA
121 CAGACAGGCA GACAGACGCA CGGACAAGCA GATGCTCCTT GGCCAGCTAT CCACTCTTCT
      m l l g q l s t l

181 GTGCCTGCTT AGCGGGGCCC TGCCTACAGG CTCAGGGAGG CCTGAACCCC AGTCTCCTCG
      l c l l s g a l p t g s g r p e p q s p

241 ACCTCAGTCC TGGGCTGCAG CCAATCAGAC CTGGGCTCTG GGCCCAGGGG CCCTGCCCCC
      r p q s w a a a n q t w a l g p g a l p

301 ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG GAAGGCCTTC TTGGGCCTGC AGAAAGCCAG
      p l v p a s a l g s w k a f l g l q k a

361 GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG GCAAGACGAG GTGGCTGCTG TGA CTCTGCC
      r q l g m g r l q r g q d e v a a v t l

421 GCTGAACCCCT CAGGAAGTGA TCCAGGGGAT GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT
      p l n p q e v i q g m c k a v p f v q v

481 CTCCCCGGCCC GGCTGCTCAG CCATACGCCT CCGAATCAT CTGTGCTTTG GTCATTGCTC
      f s r p g c s a i r l r n h l c f g h c
      INSP002-CP1

541 CTCTCTCTAC ATCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA GCTGTATGCC
      s s l y i p g s d p t p l v l c n s c m

601 TGCTCGCAAG CGTTGGGCAC CCGTGGTCCT GTGGTGCTC ACTGGCAGCT CAGCCTCCCG
      p a r k r w a p v v l w c l t g s s a s
      INSP002-CP2

661 TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGT CACTGCAGCC CAAAAGCATG
      r r r v k i s t m l i e g c h c s p k a

721 AACTGAGCAT CGTGGATGGG TGCACGGAGA CACGCACCTT GGAGAAATGA GGGGAGAT
      -

```

Figura 7: Secuencia clonada parcial INSP002 con traslación

```

1  CTCAGCCATA CGCCTCCGAA ATCATCTGTG CTTTGGTCAT TGCTCCTCTC TCTACATCCC
   s a i r l r n h l c f g h c s s l y i

61  TGGCTCGGAC CCCACCCAC TAGTCCTGTG CAACAGCTGT ATGCTTGCTC GCAAGCGTTG
   p g s d p t p l v l c n s c m p a r k r

121 GGCACCCGTG GTCCTGTGGT GTCTCACTGG CAGCTCAGC
    w a p v v l w c l t g s s
  
```

Figura 8: Mapa de PCRII-TOPO-INSP002 parcial

Molécula: **product1, 4109 bps DNA Circular**
 Nombre de Archivo: **-7879480.cm5**

Descripción: **Ligado de Sin Nombre en PCRII-TOPO-abierto**

Características de la Molécula:

Tipo	Inicio	Final	Nombre	Descripción
REGION	1	336		LacZa'
MARCADOR	239		SP6	
GEN	337	495	INSP002	cds parcial Producto PCR insertado
REGION	496	747		'LacZa
MARCADOR	584		C T7	
REGION	749	1163		f1 ori
GEN	1497	2291	KanR	Gen resistente a la canamicina
GEN	2309	3169	AmpR	Gen resistente a la ampicilina
GEN	3314	3987		pUC ori

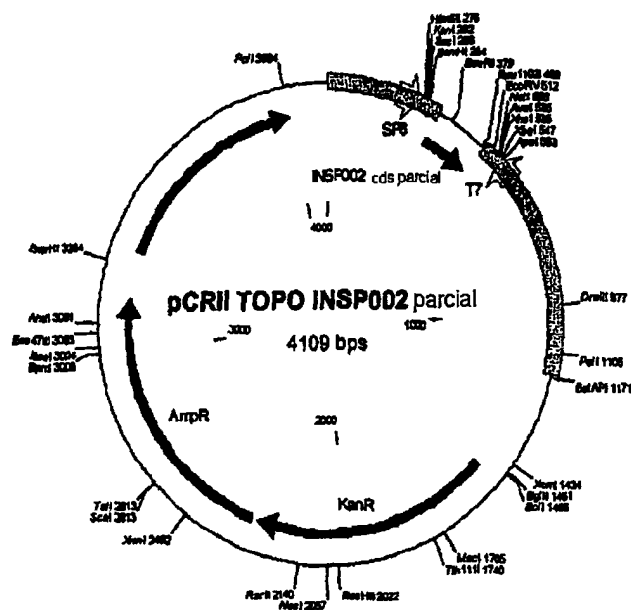


Figura 9: Alineación de predicción (superior) INSP002 y secuencia clonada parcial (inferior)

Figura 9a: Alineación de secuencias de nucleótidos

```

      10      20      30      40      50      60
INSPpr CAGAAATGCTCCAGGCTATCCAGGAGGGGCCAAGAGATTAAAGCAAGCTTCAGAAAGGC
Serono C-----
      70      80      90     100     110     120
INSPpr TCAGATGCCACTCACCAGACAGCAGGGTCGACTGCTAGTGACCTTAGGCCAGTCCGGAC
Serono -----
     130     140     150     160     170     180
INSPpr AGACAGACAGGCAGACAGACGCCACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
Serono -----
     190     200     210     220     230     240
INSPpr TCTGTGCTGCTTAGCGGGGCCCTGCTTACAGGCTCAGGGAGGCTGAACCCAGTCTCC
Serono -----
     250     260     270     280     290     300
INSPpr TCGACCTCAGTCTTGGGCTGCAGCCAAATCAGACCTGGGCTCTGGGCCAGGGGCCCTGCC
Serono -----
     310     320     330     340     350     360
INSPpr CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCCTTCTGGGCCTGCAGAAAGC
Serono -----
     370     380     390     400     410     420
INSPpr CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACAGGTGGCTGCTGTGACTCT
Serono -----
     430     440     450     460     470     480
INSPpr GCCGCTGAACCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTCAAGT
Serono -----
     490     500     510     520     530     540
INSPpr GTTCTCCCGGCCCGCTGCTCAGCCATACGCTCCGAAATCATCTGTGCTTTGGTCATTG
Serono -----TCAGCCATACGCTCCGAAATCATCTGTGCTTTGGTCATTG
                        10      20      30      40
     550     560     570     580     590     600
INSPpr CTCCTCTCTTACATCCCTGGCTGGACCCACCCCACTAGTCCCTGTGCAACAGCTGTAT
Serono CTCCTCTCTTACATCCCTGGCTGGACCCACCCCACTAGTCCCTGTGCAACAGCTGTAT
                        50      60      70      80      90     100
     610     620     630     640     650     660
INSPpr GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCTGTGGTGTCTCACTGGCAGCTCAGCCTC
Serono GCCTGCTCGCAAGCGTTGGGCACCCGTGGTCTGTGGTGTCTCACTGGCAGCTCAGC---
                        110     120     130     140     150

```

```

          670      680      690      700      710      720
INSPpr CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCACTGCAGCCCAAAAGC
Serono -----

```

Figura 9b: Alineación de secuencias de proteínas

```

Serono -----
INSP00 MLLGQLSTLLCLLSCALPTGSCRPEPQSPPQSWAAANQTWALQPGALPPLVPASALGSW
          10      20      30      40      50      60

Serono -----SAIRL
          : : : :
INSP00 KAFLLGLQKARQLGMGRLQRQDEVAAVTLPLNPQEVIGMCKAVPFVQVFSRPGCSAIRL
          70      80      90      100     110     120

          10      20      30      40      50
Serono RNHLCPGHCSsLYIPGSDFTPLVLCNSCMPARKRWAPVVLNCLTGSSA-----
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
INSP00 RNHLCPGHCSsLYIPGSDFTPLVLCNSCMPARKRWAPVVLNCLTGSSASRRRVKISTMLI
          130     140     150     160     170     180

Serono -----
INSP00 EGCHCSPKA

```

Figura 10: Secuencia de nucleótido y traslación de inserción sDNA en Image:4558384

```

1  CGGCACGAGG GGGAGACCTG GAAGGAAGCG ACTGCACTGA TCCAGAAATG CCTCCCAGGC
    INSP002V-5'-F
61  TATCCAGGAG GGGCCAAGAG ATTTAAAGCA GGTTCAGAAG GCTCAGATGC CACTCACCAG
121  ACAGCAGGGT CGACTGCTAG TGACCTTGAG CCCAGTCCGG ACAGACAGAC AGGCAGACAG
    INSP002V-5'nest-F
181  ACGCACGGAC AAGCAGATGC TCCTTGGCCA GCTATCCACT CTTCTGTGCC TGCTTAGCGG
    m l l g q l s t l l c l l s
241  GGCCCTGCCT ACAGGCTCAG GGAGGCCTGA ACCCCAGTCT CCTCGACCTC AGTCCTGGGC
    g a l p t g s g r p e p q s p r p q s w
301  TGCAGCCAAT CAGACCTGGG CTCTGGGCCC AGGGGCCCTG CCCCCACTGG TGCCAGCTTC
    a a a n q t w a l g p g a l p p l v p a
361  TGCCCTTGGG AGCTGGAAGG CTTTCTTGGG CCTGCAGAAA GCCAGGCAGC TGGGGATGGG
    s a l g s w k a f l g l q k a r q l g m
421  CAGGCTGCAG CGTGGGCAAG ACGAGGTGGC TGCTGTGACT CTGCCGCTGA ACCCTCAGGA
    g r l q r g q d e v a a v t l p l n p q
481  AGTGATCCAG GGGATGTGTA AGGCTGTGCC CTTGCTTCAG ACCGGCACT CAGCTATAT
    e v i q g m c k a v p f v q t r e s r y
    INSP002V-5'-R
541  TCGCGGAGCT AATCTTGAGC TTCTGGGCTT AAGCAATGCT CCGACCTGAG CCGCTCTAGC
    v a q a s l e l l a s s n p p t s a s
601  TCGAGGTCTT CTCCCGGCCC GGCTGCTCAG CCATACGCCT CCGAAATCAT CTGTGCTTTG
    INSP002V-3'-F
661  GTCATTGCTC CTCTCTCTAC ATCCCTGGCT CGGACCCAC CCCACTAGTC CTGTGCAACA
721  GCTGTATGCC TGCTCGCAAG CGTTGGGCAC CCGTGGTCCT GTGGTGTCTC ACTGGCAGCT
781  CAGCCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT CGAGGGGTGT CACTGCAGCC
841  CAAAAGCATG AACTGAGCAT CTGGATGGGT GCACGGAGAC ACGCACCTTG GAGAAATGAG
901  GGGAGATGGA CCAAGAAAGA CGTGGACCTG GATGATGTAC TCTGGGTCAA GAGACCAGGG
    INSP002V-3'nest-R
961  ATGCAGGGTT AGGCAGACAG GTCCCAGAG TCCTCACCTT GCTCCCCAGA CAGTAGACAC
1021  AGTGCCCGTC CTGGAGTTGC ACCACTGATA GTCACAGCAC ACAATGATTG ACAACTCACT
    INSP002V-3'-R
1081  TTTTTTTTTT TTTTGTAGAT GGAGTCTCGC TCTGTGCCCC AGGCTGGAGT GCAGTGGCGC
1141  AATCTCAGCT CACTGCAAGC TCCACCTCCC GGGTTTATGC CATTCTCCTG TCTCAGCCTC
1201  CCGAGTAGCT GGGACTACAG GCACCCGCCA ACACGCCCGG STAATTTTTT GTATTTTTTAG
1261  TAAAGACAGG GTTTCACCGT GTTAGCCAGG ATGGTCTCTA TCTCCTGACC TCGTGATCTG
1321  CCTGCCTTGG CCTTATTATT TTTTTTTTAA GGACAGAGTC TCTCTCTGTC ACCCAGGCTG
1381  GAGTGCAATG GCGCGATCTT GGCTCACTGT AACTTCCACT TGCCAGGCTC AAGCAGTTCT

```

1441 CCTGCCTCAG CTCCTGAGT AGCTGGGACT ACAGGCACCC GCCACCATGC CCAGCTAATT
1501 TTTGTATTTT TAGTAGAGAC AGAGTTTCAC CATATTAGCC TGGCTGGTCT CAAACTCCTG
1561 GCCTCAGGTG ATCTGCCCAC CTCGGCCTCC CAAAGTGCTG GGATCAAATC CACTGTTAAT
1621 CATTAGGCTG AACTGTCTCT TATAGAATGA GGTCAAAGAC ACTCCCAGTT GCAGGGAGGG
1681 TAGATGGCCC CACCCAGACC GAGAGACACA GTGATGACCT CAGCCTAGGG ACMCCAAAAA
1741 AAAAAAAAAA AAAAAA

La posición de la inserción 87 bp en comparación con la predicción de INSP002 está **resaltada**

La secuencia ALu en la secuencia 3' no codificadora está *en cursiva*.

Figura 11: Alineación de secuencias de predicción (superior) INSP002 con IMAGE: 4558384 (BC025333.1) (inferior).

Figura 11a: Secuencias de nucleótido

```

INSP00 -----CAGAAATGCCTCCCAGGCTATCCAGGAG
          ::::::::::::::::::::::::::::
BC0253 GGGAGACCTGGAAGGAAGCGACTGCACTGATCCAGAAATGCCTCCCAGGCTATCCAGGAG
          10      20      30      40      50      60

          30      40      50      60      70      80
INSP00 GGGCCAAGAGATTAAAAGCAGGTTTCAGAAGGCTCAGATGCCACTCACCAGACAGCAGGGT
          ::::::::::::::::::::::::::::::::::::
BC0253 GGGCCAAGAGATTAAAAGCAGGTTTCAGAAGGCTCAGATGCCACTCACCAGACAGCAGGGT
          70      80      90      100     110     120

          90      100     110     120     130     140
INSP00 CGACTGCTAGTGACCTTGAGCCCAGTCCGGACAGACAGACAGGCAGACAGACGCACGGAC
          ::::::::::::::::::::::::::::::::::::
BC0253 CGACTGCTAGTGACCTTGAGCCCAGTCCGGACAGACAGACAGGCAGACAGACGCACGGAC
          130     140     150     160     170     180

          150     160     170     180     190     200
INSP00 AAGCAGATGCTCCTTGCGCCAGCTATCCACTCTTCTGTGCCTGCTTAGCGGGGCCCTGCCT
          ::::::::::::::::::::::::::::::::::::
BC0253 AAGCAGATGCTCCTTGCGCCAGCTATCCACTCTTCTGTGCCTGCTTAGCGGGGCCCTGCCT
          190     200     210     220     230     240

          210     220     230     240     250     260
INSP00 ACAGGCTCAGGGAGGCTGAACCCAGTCTCCTCGACCTCAGTCTGGGCTGCAGCCAAT
          ::::::::::::::::::::::::::::::::::::
BC0253 ACAGGCTCAGGGAGGCTGAACCCAGTCTCCTCGACCTCAGTCTGGGCTGCAGCCAAT
          250     260     270     280     290     300

          270     280     290     300     310     320
INSP00 CAGACCTGGGCTCTGGGCCAGGGGCCCTGCCCCACTGGTGCCAGCTTCTGCCCTTGGG
          ::::::::::::::::::::::::::::::::::::
BC0253 CAGACCTGGGCTCTGGGCCAGGGGCCCTGCCCCACTGGTGCCAGCTTCTGCCCTTGGG
          310     320     330     340     350     360

          330     340     350     360     370     380
INSP00 AGCTGGAAGGCCCTTCTTGGGCCTGCAGAAAGCCAGGCAGCTGGGGATGGGCAGGCTGCAG
          ::::::::::::::::::::::::::::::::::::
BC0253 AGCTGGAAGGCCCTTCTTGGGCCTGCAGAAAGCCAGGCAGCTGGGGATGGGCAGGCTGCAG
          370     380     390     400     410     420

          390     400     410     420     430     440
INSP00 CGTGGGCAAGACGAGGTGGCTGCTGTGACTCTGCCGCTGAACCCCTCAGGAAGTGATCCAG
          ::::::::::::::::::::::::::::::::::::
BC0253 CGTGGGCAAGACGAGGTGGCTGCTGTGACTCTGCCGCTGAACCCCTCAGGAAGTGATCCAG
          430     440     450     460     470     480

          450     460     470
INSP00 GGGATGTGTAAGGCTGTGCCCTTCGTTTCAG-----
          ::::::::::::::::::::::::::::
BC0253 GGGATGTGTAAGGCTGTGCCCTTCGTTTCAGACACGGGAGTCTCGCTATGTTGCCCAAGCT
          490     500     510     520     530     540

          480
INSP00 -----GTGTT
          ::::
BC0253 AGTCTTGAGCTTCTGGCCTCAAGCAATCCTCCACCTCAGCCTCCTGAGTTCTAGGTGTT
          550     560     570     580     590     600

          490     500     510     520     530     540
INSP00 CTCCCGCCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGCTC
          ::::::::::::::::::::::::::::::::::::
BC0253 CTCCCGCCCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTGCTC
          610     620     630     640     650     660

```


ES 2 306 802 T3

```

      550      560      570      580      590      600
INSP00 CTCTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTATGCC
      :
BC0253 CTCTCTCTACATCCCTGGCTCGGACCCACCCCACTAGTCCTGTGCAACAGCTGTATGCC
      670      680      690      700      710      720

      610      620      630      640      650      660
INSP00 TGCTCGCAAGCGTTGGGCACCCGTGGTCCTGTGGTGTCTCACTGGCAGCTCAGCCTCCCG
      :
BC0253 TGCTCGCAAGCGTTGGGCACCCGTGGTCCTGTGGTGTCTCACTGGCAGCTCAGCCTCCCG
      730      740      750      760      770      780

      670      680      690      700      710      720
INSP00 TCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCAGTGCAGCCCAAAGCATG
      :
BC0253 TCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCAGTGCAGCCCAAAGCATG
      790      800      810      820      830      840

INSP00 A-----
      :
BC0253 AACTGAGCATCTGGATGGGTGCACGGAGACACGCACCTTGGAGAAATGAGGGGAGATGGA
      850      860      870      880      890      900

INSP00 -----
BC0253 CCAAGAAAGACGTGGACCTGGATGATGTACTCTGGGTCAAGAGACCAGGGATGCAGGGTT
      910      920      930      940      950      960

INSP00 -----
BC0253 AGGCAGACAGGTCCCCAGAGTCCTCACCTGCTCCCCAGACAGTAGACACAGTGCCCGTC
      970      980      990      1000      1010      1020

INSP00 -----
BC0253 CTGGAGTTGCACCACTGATAGTCACAGCACACAATGATTGACAACTCACTTTTTTTTTTT
      1030      1040      1050      1060      1070      1080

INSP00 -----
BC0253 TTTTGTAGATGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGCAATCTCAGCT
      1090      1100      1110      1120      1130      1140

INSP00 -----
BC0253 CACTGCAAGCTCCACCTCCCGGGTTTATGCCATTCTCCTGTCTCAGCCTCCCGAGTAGCT
      1150      1160      1170      1180      1190      1200

INSP00 -----
BC0253 GGGACTACAGGCACCCGCCAACACGCCCGGCTAATTTTTTGTATTTTGTAAAGACAGG
      1210      1220      1230      1240      1250      1260

INSP00 -----
BC0253 GTTTCACCGTGTAGCCAGGATGGTCTCTATCTCCTGACCTCGTGATCTGCCTGCCTTGG
      1270      1280      1290      1300      1310      1320

INSP00 -----
BC0253 CCTTATTATTTTTTTTAAAGGACAGAGTCTCTCTGTACCCAGGCTGGAGTGCAATG
      1330      1340      1350      1360      1370      1380

```

ES 2 306 802 T3

```

INSP00 -----
BC0253 GCGCGATCTTGGCTCACTGTAACCTTCCACTTGCCAGGCTCAAGCAGTTCTCCTGCCTCAG
          1390      1400      1410      1420      1430      1440

INSP00 -----
BC0253 CCTCCTGAGTAGCTGGGACTACAGGCACCCGCCACCATGCCCAGCTAATTTTGTATTTT
          1450      1460      1470      1480      1490      1500

INSP00 -----
BC0253 TAGTAGAGACAGAGTTTCACCATATTAGCCTGGCTGGTCTCAAACCTCCTGGCCTCAGGTG
          1510      1520      1530      1540      1550      1560

INSP00 -----
BC0253 ATCTGCCACCTCGGCCTCCCAAAGTGCTGGGATCAAATCCACTGTTAATCATTAGGCTG
          1570      1580      1590      1600      1610      1620

INSP00 -----
BC0253 AACTGTCTCTTATAGAATGAGGTCAAAGACACTCCAGTTGCAGGGAGGGTAGATGCCCC
          1630      1640      1650      1660      1670      1680

INSP00 -----
BC0253 CACCCAGACCGAGAGACACAGTGATGACCTCAGCCTAGGGACACCAAAAAAAAAAAAAA
          1690      1700      1710      1720      1730      1740

INSP00 -----
BC0253 AAAAAA

```

Figura 11b: Secuencias de proteínas

```

4558384_OR      MLLGQLSTLLCLLSGALPTGSGRPEPQS PRPQSWAAANQTWALGPGALPPLVPASALGSW
INSP002_PR      MLLGQLSTLLCLLSGALPTGSGRPEPQS PRPQSWAAANQTWALGPGALPPLVPASALGSW
                  *****

4558384_OR      KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIOGMCKAVPFVQTRE----
INSP002_PR      KAFLGLQKARQLGMGRLQRGQDEVAAVTLPLNPQEVIOGMCKAVPFVQVFSRPGCSAIRL
                  *****

4558384_OR      -----SRYVAQAS---LELLASSNPP-----TSAS-----
INSP002_PR      RNHLFCFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
                  * *. . * * *. * .***

4558384_OR      -----
INSP002_PR      EGCHCSPKA

```

Figura 12: Secuencia de nucleótido y traslación de INSP002V generada por PCR de IMAGE 4558384

```

1  GTCGACTGCT AGTGACCTTG AGCCCAGTCC GGACAGACAG ACAGGCAGAC AGACGCACGG
61  ACAAGCAGAT GCTCCTTGGC CAGCTATCCA CTCTTCTGTG CCTGCTTAGC GGGGCCCTGC
    m l l g q l s t l l c l l s g a l

121 CTACAGGCTC AGGGAGGCCT GAACCCAGT CTCCTCGACC TCAGTCCTGG GCTGCAGCCA
    p t g s g r p e p q s p r p q s w a a a

181 ATCAGACCTG GGCTCTGGGC CCAGGGGCCC TGCCCCCACT GGTGCCAGCT TCTGCCCTTG
    n q t w a l g p g a l p p l v p a s a l

241 GGAGCTGGAA GGCCTTCTTG GGCCTGCAGA AAGCCAGGCA GCTGGGGATG GGCAGGCTGC
    g s w k a f l g l q k a r q l g m g r l

301 AGCGTGGGCA AGACGAGGTG GCTGCTGTGA CTCTGCCGCT GAACCCCTCAG GAAGTGATCC
    q r g q d e v a a v t l p l n p q e v i

361 AGGGGATGTG TAAGGCTGTG CCCTTCGTTC TCTCCCGGCC CGGCTGCTCA GCCATACGCC
    q g m c k a v p f v l s r p g c s a i r

421 TCCGAAATCA TCTGTGCTTT GGTCATTGCT CCTCTCTCTA CATCCCTGGC TCGGACCCCA
    l r n h l c f g h c s s l y i p g s d p

481 CCCCCTAGT CCTGTGCAAC AGCTGTATGC CTGCTCGCAA GCGTTGGGCA CCCGTGGTCC
    t p l v l c n s c m p a r k r w a p v v

541 TGTGGTGTCT CACTGGCAGC TCAGCCTCCC GTCGACGGGT GAAGATATCC ACCATGCTGA
    l w c l t g s s a s r r r v k i s t m l

601 TCGAGGGGTG TCACTGCAGC CCAAAGCAT GAACTGAGCA TCTGGATGGG TGCACGGAGA
    i e g c h c s p k a

661 CACGCACCTT GGAGAAATGA GGGGAGATGG ACCAAGAAAG ACGTGGACCT GGATGATGT

```

Figura 13: Mapa de pCR4blunt-TOPO-INSP002V

Molécula: **pCR4** completa en sus pares base **TOPO INSP002V**, 4676 bps ADN Circular
 Nombre del Archivo: **pCR4**completo en sus pares base **TOPO INSP002V.cm5**
 Descripción: Ligado de producto FR completo en sus pares base **INSP002** pCR4 completo en sus pares base **pCR4-TOPO**
 Notas: plásmido ID 13075
 Características de Molécula:

Tipo	Inicio	Final	Nombre	Descripción
REGION	2	216		región de promotor lac
REGION	205	221		sitio de cebado inverso M13
REGION	217	294		fusión del gen LacZa-ccdB
MARCADOR	243		T3	
REGION	262	294		Polienucleótido
REGION	294	294		sitio de clonación TOPO
REGION	295	1013		Producto PCR INSP002V
GEN	363	923	INSP002V ORF	
REGION	1014	1529		Fusión del gen LacZa-ccdB
REGION	1014	1031		Polienucleótido
REGION	1014	1014		sitio de clonación TOPO
MARKER	1066		C T7	
REGION	1089	1074	C	-20M13 sitio de cebado delantero
GEN	1878	2672	KanR	
REGION	2864	2868		sitio de enlace de ribosoma
GEN	2876	3736	AmpR	
REGION	3881	4554		Origen pUC

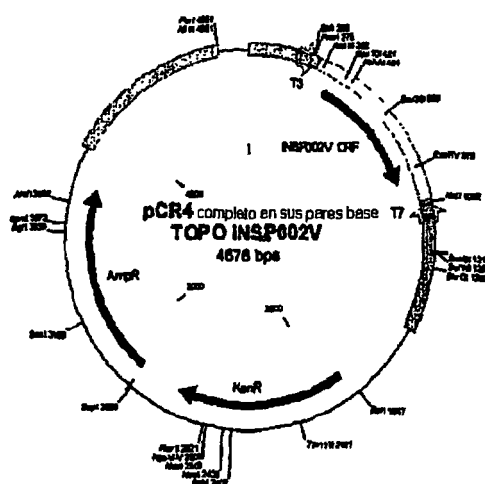


Figura 14: Comparación entre predicción (superior) INSP002 y secuencia variante INSP002V (inferior).

Figura 14a: Secuencias de nucleótido

```

INSPpr CAGAAATGCCTCCCAGGCTATCCAGGAGGGGCCAAGAGATTAAAAGCAGGTTCAGAAGGC

INSPen -----

              70      80      90      100     110     120
INSPpr TCAGATGCCACTCACCAGACACAGGGTCGACTGCTAGTGACCTTGAGCCCAGTCCGGAC
              .....
INSPen -----GTCGACTGCTAGTGACCTTGAGCCCAGTCCGGAC
                      10      20      30

              130     140     150     160     170     180
INSPpr AGACAGACAGGCAGACAGACGACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
              .....
INSPen AGACAGACAGGCAGACAGACGACGGACAAGCAGATGCTCCTTGGCCAGCTATCCACTCT
              40      50      60      70      80      90

              190     200     210     220     230     240
INSPpr TCTGTGCCTGCTTAGCGGGGCCCTGCCTACAGGCTCAGGGAGGCCTGAACCCAGTCTCC
              .....
INSPen TCTGTGCCTGCTTAGCGGGGCCCTGCCTACAGGCTCAGGGAGGCCTGAACCCAGTCTCC
              100     110     120     130     140     150

              250     260     270     280     290     300
INSPpr TCGACCTCAGTCCTGGGCTGCAGCCAATCAGACCTGGGCTCTGGGCCAGGGGCCCTGCC
              .....
INSPen TCGACCTCAGTCCTGGGCTGCAGCCAATCAGACCTGGGCTCTGGGCCAGGGGCCCTGCC
              160     170     180     190     200     210

              310     320     330     340     350     360
INSPpr CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCTTCTTGGGCCTGCAGAAAGC
              .....
INSPen CCCACTGGTGCCAGCTTCTGCCCTTGGGAGCTGGAAGGCCTTCTTGGGCCTGCAGAAAGC
              220     230     240     250     260     270

              370     380     390     400     410     420
INSPpr CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACGAGGTGGCTGCTGTGACTCT
              .....
INSPen CAGGCAGCTGGGGATGGGCAGGCTGCAGCGTGGGCAAGACGAGGTGGCTGCTGTGACTCT
              280     290     300     310     320     330

```

ES 2 306 802 T3

```

          430      440      450      460      470      480
INSPpr GCCGCTGAACCCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTTCAGGT
      :
INSPen GCCGCTGAACCCTCAGGAAGTGATCCAGGGGATGTGTAAGGCTGTGCCCTTCGTTTC----
          340      350      360      370      380      390

          490      500      510      520      530      540
INSPpr GTTCTCCCGGGCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
      :
INSPen --TCTCCCGGGCCGGCTGCTCAGCCATACGCCTCCGAAATCATCTGTGCTTTGGTCATTG
          400      410      420      430      440

          550      560      570      580      590      600
INSPpr CTCCTCTCTTACATCCCTGGCTCGGACCCACCCACTAGTCCTGTGCAACAGCTGTAT
      :
INSPen CTCCTCTCTTACATCCCTGGCTCGGACCCACCCACTAGTCCTGTGCAACAGCTGTAT
          450      460      470      480      490      500

          610      620      630      640      650      660
INSPpr GCCTGCTCGCAAGCGTTGGGCACCCGTCCTGTGGTGTCTCACTGGCAGCTCAGCCTC
      :
INSPen GCCTGCTCGCAAGCGTTGGGCACCCGTCCTGTGGTGTCTCACTGGCAGCTCAGCCTC
          510      520      530      540      550      560

          670      680      690      700      710      720
INSPpr CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCACTGCAGCCCCAAAAGC
      :
INSPen CCGTCGACGGGTGAAGATATCCACCATGCTGATCGAGGGGTGTCACTGCAGCCCCAAAAGC
          570      580      590      600      610      620

INSPpr ATGA-----
      :
INSPen ATGAAGTGAAGCATCTGGATGGGTGCACGGAGACACGCACCTTGGAGAAATGAGGGGAGAT
          630      640      650      660      670      680

INSPpr -----

INSPen GGACCAAGAAAGACGTGGACCTGGATGATGT
          690      700      710

```

Figura 14b: Secuencias de proteína

INSP002V	23	MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW
INSP002	1	MLLGQLSTLLCLLSGALPTGSGRPEPQSPRPQSWAAANQTWALGPGALPPLVPASALGSW

INSP002V	83	KAFGLQLKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPF--VLSRPGCSAIRL
INSP002	61	KAFGLQLKARQLGMGRLQRGQDEVAAVTLPLNPQEVIQGMCKAVPFVQVFSRPGCSAIRL
		***** * *****
INSP002V	141	RNHL CFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI
INSP002	121	RNHL CFGHCSSLYIPGSDPTPLVLCNSCMPARKRWAPVVLWCLTGSSASRRRVKISTMLI

INSP002V	201	EGCHCSPKA
INSP002	181	EGCHCSPKA

Figura 16: Mapa vector puerta pDONR201

Molécula: pDONR201, 4470 bps ADN Circular
 Nombre del Archivo: pDONR201.cm5

Descripción: Vector de entrada Gateway (Invitrogen) - plásmido ID# 13309

Características de la Molécula:

Tipo	Inicio	Final	Nombre
REGION	332	563	attP1
GEN	959	1264	ccdB
REGION	2513	2744	attP2
GEN	2868	3677	KanR
REGION	3794	4467	pUC ori

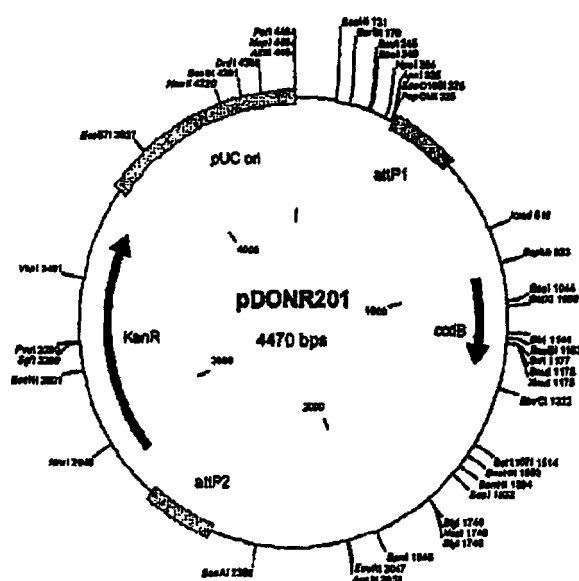


Figura 17: Mapa de pEAK12d-INSP002-V-6HIS

Molécula: pEAK12d-INSP002-V-6HIS, 7528 bps ADN Circular
Nombre de Archivo: pEAK12d-INSP002V-6HIS.cm5
Descripción: el ADNc se inserta entre los sitios attR1 y attR2
Notas: plásmido ID 13227

Características de Molécula:

Tipo	Inicio	Final	Nombre	Descripción
REGION	2	595		pmb-ori
GEN	596	1519	AmpR	
REGION	1690	2795	EF-1alfa	promotor
REGION	2703	2722		pEAK12F
REGION	2796	2845		MCS''
REGION	2855	2874		attB1
GEN	2888	3469	INSP002-V	
REGION	3474	3495		attB2
REGION	3501	3501		'MCS
REGION	3502	3930		poli A/empalme
REGION	3616	3597 C		pEAK12R
GEN	4549	3931 C		PUROMICINA R
REGION	4773	4550 C		promotor tK
REGION	5268	4774 C	Ori p	
GEN	7320	5268 C	EBNA-1	
REGION	7321	7520	sv40	

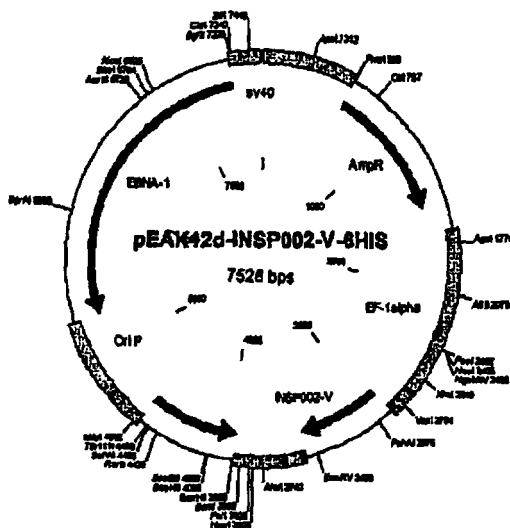


Figura 18: INSP002 de longitud completa clonado a partir de cDNA de cortazón humano

```

1  GATGCTCCTT GGCCAGCTAT CCACTCTTCT GTGCCTGCTT AGCGGGGCCC TGCCCTACAGG
   m l l g q l s t l l c l l s g a l p t
   INSP002-FL-F
61  CTCAGGGAGG CCTGAACCCC AGTCTCCTCG ACCTCAGTCC TGGGCTGCAG CCAATCAGAC
   g s g r p e p q s p r p q s w a a a n q
121 CTGGGCTCTG GGCCCAGGGG CCCTGCCCCC ACTGGTGCCA GCTTCTGCCC TTGGGAGCTG
   t w a l g p g a l p p l v p a s a l g s
181 GAAGGCCTTC TTGGGCCTGC AGAAAGCCAG GCAGCTGGGG ATGGGCAGGC TGCAGCGTGG
   w k a f l g l q k a r q l g m g r l q r
241 GCAAGACGAG GTGGCTGCTG TGA CTCTGCC GCTGAACCC T CAGGAAGTGA TCCAGGGGAT
   g q d e v a a v t l p l n p q e v i q g
301 GTGTAAGGCT GTGCCCTTCG TTCAGGTGTT CTCCCGGCCC GGCTGCTCAG CCATACGCCT
   m c k a v p f v q v f s r p g c s a i r
361 CCGAAATCAT CTGTGCTTTG GTCATTGCTC CTCTCTCTAC ATCCCTGGCT CGGACCCCAC
   l r n h l c f g h c s s l y i p g s d p
421 CCCACTAGTC CTGTGCAACA GCTGTATGCC TGCTCGCAAG CGTTGGGCAC CCGTGGTTCCT
   t p l v l c n s c m p a r k r w a p v v
481 GTGGTGTCTC ACTGGCAGCT CAGCCTCCCG TCGACGGGTG AAGATATCCA CCATGCTGAT
   l w c l t g s s a s r r r v k i s t m l
541 CGAGGGGTGT CACTGCAGCC CAAAAGCATG AACTGAGCAT CGTGGATGG
   i e g c h c s p k a
                                   INSP002-FL-R

```

Figura 19: Mapa de PCR4blunt-TOPO-INSP002FL

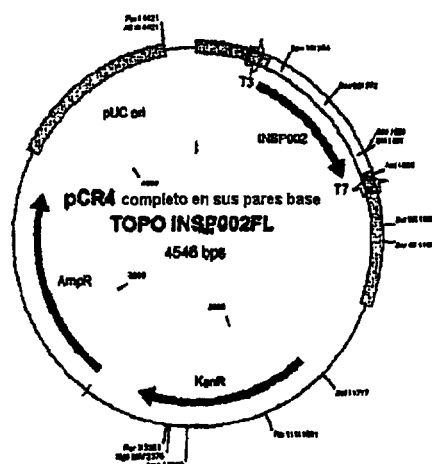
Molécula: **pCR4 completa en sus pares base TOPO INSP002FL, 4546 bps ADN Circular**

Nombre de Archivo: **pCR4 completo en sus pares base TOPO INSP002 FL.cm5,**

Descripción: **Ligado del sin nombre en pCR4 completo en sus pares base TOPO abierto***

Características de la Molécula:

Tipo	Inicio	Final	Nombre	Descripción
REGION	2	216		región del promotor lac
REGION	205	221		sitio de cebado inverso M13
REGION	217	294		fusión del gen LacZa-ccdB
MARCADOR	243		T3	
REGION	262	294		Polienlazador
REGION	294	294		sitio de clonado TOPO
REGION	295	883		producto PCR INSP002-F3R3
GEN	296	865	INSP002	INSP002 ORF
REGION	884	1399		fusión del gen 'LacZa-ccdB
REGION	884	901		Polienlazador
REGION	884	884		sitio de clonación 'TOPO
MARCADOR	936		C T7	promotor T7
REGION	959	944	C	sitio de cebado delantero -20M13
GEN	1748	2542	KanR	gen resistente a la canamicina ORF
REGION	2734	2738		sitio de enlace de ribosoma
GEN	2746	3606	AmpR	gen resistente a la ampicilina ORF
REGION	3751	4424	pUC ori	origen pUC



ES 2 306 802 T3

LISTA DE SECUENCIAS

<110> ARES TRADING S.A.

5 <120> PROTEÍNA DE PLIEGUE CON NUDO DE CISTINA

<130> P029147WO

10 <140> PCT/GB02/05865

<141> 2002-12-20

15 <150> GB 0130738.8

<151> 2001-12-21

<160> 38

20 <170> SeqWin99, version 1.02

<210> 1

<211> 475

25 <212> DNA

<213> *Homo sapiens*

<220>

30 <221> Secuencia de nucleótido INSP002 - Exon 1

<400> 1

35 **aaatgcctcc caggctatcc aggaggggcc aagagattaa aagcaggttc agaaggctca 60**
 gatgccactc accagacagc agggctcgact gctagtgacc ttgagcccag tccggacaga 120
 cagacaggca gacagacgca cggacaagca gatgctcctt ggccagctat ccactcttct 180
 gtgcctgctt agcggggccc tgcctacagg ctcaggaggg cctgaacccc agtctcctcg 240
 acctcagtc tgggctgcag ccaatcagac ctgggctctg ggcccagggg ccctgcccc 300
40 **actggtgcca gcttctgccc ttgggagctg gaaggccttc ttgggcctgc agaaagccag 360**
 gcagctgggg atgggcaggc tgcagcgtgg gcaagacgag gtggctgctg tgactctgcc 420
 gctgaaccct caggaagtga tccaggggat gtgtaaggct gtgcccttcg ttcag 475

45 <210> 2

<211> 108

<212> PRT

<213> *Homo sapiens*

50 <220>

<221> Secuencia de proteína INSP002 - Exon 1

55

60

65

ES 2 306 802 T3

<400> 2

```

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

```

<210> 3

25 <211> 246

<212> DNA

<213> *Homo sapiens*

30 <220>

<221> Secuencia de nucleótido INSP002 - Exon 2

<400> 3

```

35     gtgttctccc ggcccggctg ctcagccata cgcctccgaa atcatctgtg ctttggtcat 60
      tgctcctctc tctacatccc tggctcggac cccacccac tagtcctgtg caacagctgt 120
      atgcctgtc gcaagcgttg ggcacccgtg gtcctgtggt gtctcactg cagctcagcc 180
      tcccgtcgac ggggtgaagat atccaccatg ctgacgcagg ggtgtcactg cagcccaaaa 240
40     gcatga 246

```

<210> 4

<211> 81

<212> PRT

45 <213> *Homo sapiens*

<220>

50 <221> Secuencia de proteína INSP002 - Exon 2

<400> 4

```

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

```

ES 2 306 802 T3

<210> 5

<211> 721

<212> DNA

5 <213> *Homo sapiens*

<220>

<221> Secuencia de Nucleótido INSP002

10

<400> 5

```

aaatgcccc caggetatcc aggaggggcc aagagattaa aagcagggtc agaaggctca 60
gatgccactc accagacagc aggggtcgact gctagtgacc ttgagcccag tccggacaga 120
cagacaggca gacagacgca cggacaagca gatgctcctt ggccagctat ccactcttct 180
gtgcctgctt agcggggccc tgcctacagg ctcaggaggg cctgaacccc agtctcctcg 240
acctcagtc tgggctgcag ccaatcagac ctgggctctg ggcccagggg ccctgcccc 300
actggtgcca gcttctgccc ttgggagctg gaaggccttc ttgggcctgc agaaagccag 360
gcagctgggg atgggcaggc tgcagcgtgg gcaagacgag gtggctgctg tgactctgcc 420
gctgaaccct caggaagtga tccaggggat gtgtaaggct gtgcccttcg ttcaggtgtt 480
ctcccgccc ggctgctcag ccatacgctt ccgaaatcat ctgtgctttg gtcattgtct 540
ctctctctac atccctggct cggacccac cccactagtc ctgtgcaaca gctgtatgcc 600
tgctcgcaag cgttgggcac ccgtggctct gtggtgtctc actggcagct cagcctcccc 660
tcgacgggtg aagatatcca ccattgctgat cgaggggtgt cactgcagcc caaaagcatg 720
a 721

```

25 <210> 6

<211> 189

<212> PRT

<213> *Homo sapiens*

30

<220>

<221> Secuencia de proteína INSP002

35 <400> 6

```

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

```

ES 2 306 802 T3

<210> 7

<211> 86

<212> PRT

5 <213> *Homo sapiens*

<220>

<221> Secuencia de proteína INSP002 - Exon 1 sin péptido señalizador

10

<400> 7

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

30

<210> 8

<211> 167

<212> PRT

35 <213> *Homo sapiens*

<220>

<221> Secuencia de proteína INSP002 sin péptido señalizador

40

<400> 8

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

ES 2 306 802 T3

<210> 9

<211> 504

<212> DNA

5 <213> *Homo sapiens*

<220>

<221> Secuencia de nucleótido INSP002 - sin péptido señalizador

10

<400> 9

15

```

aggcctgaac ccagctctcc tcgacctcag tcctgggctg cagccaatca gacctgggct 60
ctgggcccag gggccctgcc cccactggtg ccagcttctg cccttgggag ctggaaggcc 120
ttcttgggcc tgcagaaagc caggcagctg gggatgggca ggctgcagcg tgggcaagac 180
gaggtggctg ctgtgactct gccgctgaac cctcaggaag tgatccaggg gatgtgtaag 240
gctgtgccct tcgttcaggt gttctcccg cccggtgct cagccatacg cctccgaaat 300
catctgtgct ttggtcattg ctctctctc tacatccctg gctcggaccc caccocacta 360
gtcctgtgca acagctgtat gctgctcgc aagcgttggg caccctgggt cctgtgggtg 420
ctcactggca gctcagctcc ccgtcgacgg gtgaagatat ccaccatgct gatcgagggg 480
tgtcactgca gcccaaaagc atga 504

```

20

<210> 10

<211> 258

<212> DNA

25 <213> *Homo sapiens*

<220>

<221> Secuencia de nucleótido INSP002 - sin Exon 1 sin péptido señalizador

30

<400> 10

35

```

aggcctgaac ccagctctcc tcgacctcag tcctgggctg cagccaatca gacctgggct 60
ctgggcccag gggccctgcc cccactggtg ccagcttctg cccttgggag ctggaaggcc 120
ttcttgggcc tgcagaaagc caggcagctg gggatgggca ggctgcagcg tgggcaagac 180
gaggtggctg ctgtgactct gccgctgaac cctcaggaag tgatccaggg gatgtgtaag 240
gctgtgccct tcgttcag 258

```

<210> 11

40

<211> 570

<212> DNA

<213> *Homo sapiens*

<220>

45

<221> Secuencia de nucleótido INSP002 sin 5'UTR

<400> 11

50

```

atgctccttg gccagctatc cactcttctg tgcctgctta gcggggccct gcctacaggg 60
tcagggaggc ctgaacccca gtctcctcga cctcagtcct gggtgcagc caatcagacc 120
tgggctctgg gccaggggc cctgccccca ctggtgccag cttctgccct tgggagctgg 180
aaggccttct tgggcctgca gaaagccagg cagctgggga tgggcaggct gcagcgtggg 240
caagacgagg tggctgctgt gactctgccg ctgaaccctc aggaagtgat ccaggggatg 300
tgtaaggctg tgcccttcgt tcaggtgttc tcccggcccg gctgctcagc catacgctc 360
cgaaatcatc tgtgctttgg tcattgctcc tctctctaca tccctggctc ggacccacc 420
ccactagtc tgtgcaacag ctgtatgcct gctcgcaagc gttgggcacc cgtggctctg 480
tggtgtctca ctggcagctc agcctcccg tgcaggggtga agatatccac catgctgatc 540
gaggggtgtc actgcagccc aaaagcatga 570

```

55

60

<210> 12

<211> 324

<212> DNA

<213> *Homo sapiens*

65

<220>

<221> Secuencia nucleótido INSP002 - exon 1 sin 5'UTR

ES 2 306 802 T3

<400> 12

```

5      atgctccttg gccagctatc cactcttctg tgcctgctta gcggggccct gcctacaggc   60
      tcagggaggg ctgaacccca gtctcctcga cctcagtcct gggctgcagc caatcagacc   120
      tgggctcttg gcccaggggc cctgccccca ctggtgccag cttctgccct tgggagctgg   180
      aaggccttct tgggectgca gaaagccagg cagctgggga tgggcaggct gcagcgtggg   240
      caagacgagg tggctgctgt gactctgccg ctgaaccctc aggaagtgat ccaggggatg   300
      tgtaaggctg tgcccttcgt tcag                                     324

```

10 <210> 13

<211> 719

<212> DNA

<213> *Homo sapiens*

15

<220>

<221> Secuencia de nucleótido INSP002 - polipéptido variante INSP002

20 <400> 13

```

25      gtcgactgct agtgacctg agcccagtc ggacagacag acaggcagac agacgcacgg   60
      acaagcagat gctccttggc cagctatcca ctcttctgtg cctgcttagc ggggccctgc   120
      ctacaggctc agggaggcct gaaccccagt ctctcgcacc tcagtcctgg gctgcagcca   180
      atcagacctg ggctctgggc ccaggggccc tgccccactt ggtgccagct tctgcccttg   240
      ggagctggaa ggccttcttg ggcctgcaga aagccaggca gctggggatg ggcaggctgc   300
      agcgtgggca agacgaggtg gctgctgtga ctctgccgct gaaccctcag gaagtgatcc   360
      aggggatgtg taaggctgtg cccttcgttc tctcccgccc cggctgctca gccatacgcc   420
      tccgaaatca tctgtgcttt ggtcattgct cctctctcta catccctggc tcggacccca   480
      cccactagt cctgtgcaac agctgtatgc ctgctcgcaa gcgttgggca cccgtggtcc   540
30      tgtggtgtct cactggcagc tcagcctccc gtcgacgggt gaagatatcc accatgctga   600
      tcgaggggtg tctactgcagc ccaaaagcat gaactgagca tctggatggg tgcacggaga   660
      cacgcacctt ggagaaatga ggggagatgg accaagaaag acgtggacct ggatgatgt   719

```

<210> 14

35 <211> 187

<212> PRT

<213> *Homo sapiens*

40 <220>

<221> Secuencia de polipéptido INSP002 - polipéptido variante INSP002

<400> 14

45

```

      Met Leu Leu Gly Gln Leu Ser Thr Leu Leu Cys Leu Leu Ser Gly Ala
      1           5           10          15

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

      Ser Trp Ala Ala Ala Asn Gln Thr Trp Ala Leu Gly Pro Gly Ala Leu
      35          40          45

      Pro Pro Leu Val Pro Ala Ser Ala Leu Gly Ser Trp Lys Ala Phe Leu
      50          55          60

      Gly Leu Gln Lys Ala Arg Gln Leu Gly Met Gly Arg Leu Gln Arg Gly
      65          70          75          80

      Gln Asp Glu Val Ala Ala Val Thr Leu Pro Leu Asn Pro Gln Glu Val
      85          90          95

      Ile Gln Gly Met Cys Lys Ala Val Pro Phe Val Leu Ser Arg Pro Gly
      100         105         110

```

65

ES 2 306 802 T3

Cys Ser Ala Ile Arg Leu Arg Asn His Leu Cys Phe Gly His Cys Ser
115 120 125

Ser Leu Tyr Ile Pro Gly Ser Asp Pro Thr Pro Leu Val Leu Cys Asn
130 135 140

Ser Cys Met Pro Ala Arg Lys Arg Trp Ala Pro Val Val Leu Trp Cys
145 150 155 160

Leu Thr Gly Ser Ser Ala Ser Arg Arg Arg Val Lys Ile Ser Thr Met
165 170 175

Leu Ile Glu Gly Cys His Cys Ser Pro Lys Ala
180 185

<210> 15

<211> 561

<212> DNA

<213> *Homo sapiens*

<220>

<221> Secuencia de nucleótido INSP002 - polipéptido variante INSP002 sin 5' UTR

<400> 15

atgctccttg	gccagctatc	cactcttctg	tgcttgctta	gcggggccct	gcctacaggc	60
tcagggaggg	ctgaacccca	gtctcctcga	cctcagtcct	gggctgcagc	caatcagacc	120
tgggctctgg	gcccaggggc	cctgccccca	ctggtgccag	cttctgccct	tgggagctgg	180
aaggccttct	tgggcctgca	gaaagccagg	cagctgggga	tgggcaggct	gcagcgtggg	240
caagacgagg	tggtgctgt	gactctgccg	ctgaaccctc	aggaagtgat	ccaggggatg	300
tgtaaggctg	tgcccttcgt	tctctcccg	cccggctgct	cagccatacg	cctccgaaat	360
catctgtgct	ttggtcattg	ctcctctctc	tacatccctg	gctcggaccc	cacccacta	420
gtcctgtgca	acagctgtat	gcctgctcgc	aagcgttggg	caccctgggt	cctgtggtgt	480
ctcactggca	gctcagcctc	ccgtcgacgg	gtgaagatat	ccaccatgct	gacgagggg	540
tgtcactgca	gcccaaaagc	a				561

<210> 16

<211> 20

<212> DNA

<213> Secuencia artificial

<220>

<223> Imprimador PCR

<400> 16

ctcagccata cgcctccgaa 20

<210> 17

<211> 20

<212> DNA

<213> Secuencia artificial

<220>

<223> Imprimador PCR

<400> 17

gctgagctgc cagttagaca 20

ES 2 306 802 T3

<210> 18
 <211> 25
 <212> DNA
 5 <213> Secuencia artificial

 <220>
 <223> Imprimador PCR
 10
 <400> 18

 accctggaagg aagcgactgc actga 25
 15

 <210> 19
 <211> 37
 20 <212> DNA
 <213> Secuencia artificial

 <220>
 25 <223> Imprimador PCR

 <400> 19

 gcagccgggc cgggagagaa cgaagggcac agcccta 37
 30

 <210> 20
 <211> 37
 35 <212> DNA
 <213> Secuencia artificial

 <220>
 40 <223> Imprimador PCR

 <400> 20
 45
 aggdgtgcc ctctgtctc tcccggcccg gctgctc 37

 <210> 21
 50 <211> 25
 <212> DNA
 <213> Secuencia artificial

 <220>
 55 <223> Imprimador PCR

 <400> 21
 60
 actccaggac gggcactgtg tctac 25

 <210> 22
 65 <211> 22
 <212> DNA

ES 2 306 802 T3

<213> Secuencia artificial

<220>

5 <223> Imprimador PCR

<400> 22

10 **gtogactgct agtgaccttg ag** 22

<210> 23

<211> 22

15 <212> DNA

<213> Secuencia artificial

<220>

20 <223> Imprimador PCR

<400> 23

25 **acatcatcca ggtccacgtc tt** 22

<210> 24

30 <211> 18

<212> DNA

<213> Secuencia artificial

<220>

35 <223> Imprimador PCR

<400> 24

40 **atttaggtga cactatag** 18

<210> 25

45 <211> 20

<212> DNA

<213> Secuencia artificial

<220>

50 <223> Imprimador PCR

<400> 25

55 **taatacgact cactataggg** 20

<210> 26

60 <211> 20

<212> DNA

<213> Secuencia artificial

65 <220>

<223> Imprimador PCR

ES 2 306 802 T3

<400> 26
 attaacccctc actaaaggga 20
 5
 <210> 27
 <211> 18
 <212> DNA
 10 <213> Secuencia artificial
 <220>
 <223> Imprimador PCR
 15
 <400> 27
 tgtaaacga cggccagt 18
 20
 <210> 28
 <211> 20
 <212> DNA
 25 <213> Secuencia artificial
 <220>
 <223> Imprimador PCR
 30
 <400> 28
 gccagcttgg cacttgatg 20
 35
 <210> 29
 <211> 20
 40 <212> DNA
 <213> Secuencia artificial
 <220>
 45 <223> Imprimador PCR
 <400> 29
 50 gatggagggtg gacgtgtcag 20
 <210> 30
 <211> 37
 55 <212> DNA
 <213> Secuencia artificial
 <220>
 60 <223> Imprimador PCR
 <400> 30
 65 aagcagggtt cgccaccatg ctcttggcc agctatc 37

ES 2 306 802 T3

<210> 31
 <211> 35
 <212> DNA
 5 <213> Secuencia artificial

 <220>
 <223> Imprimador PCR
 10
 <400> 31

 gtgatggtga tgggtgctt ttggctgca gtgac 35
 15
 <210> 32
 <211> 37
 <212> DNA
 20 <213> Secuencia artificial

 <220>
 25 <223> Imprimador PCR

 <400> 32

 ggggacaagt ttgtacaaa aagcaggctt cgccacc 37
 30
 <210> 33
 <211> 51
 35 <212> DNA
 <213> Secuencia artificial

 <220>
 40 <223> Imprimador PCR

 <400> 33

 ggggaccact ttgtacaaga aagctgggtt tcaatggtga tggatggt g 51
 45
 <210> 34
 50 <211> 20
 <212> DNA
 <213> Secuencia artificial

 <220>
 55 <223> Imprimador PCR

 <400> 34
 60
 gatgctcctt ggccagctat 20

 <210> 35
 65 <211> 20
 <212> DNA

ES 2 306 802 T3

<213> Secuencia artificial

<220>

5 <223> Imprimador PCR

<400> 35

10 ccatccacga tgctcagttc 20

<210> 36

<211> 1756

15 <212> DNA

<213> Secuencia artificial

<220>

20 <223> Inserción de Vector

<220>

<221> Secuencia de inserción de vector que contiene Image 4558384 (Figura 10)

25 <400> 36

	cggcacgagg	gggagacctg	gaaggaagcg	actgcaactga	tccagaaatg	cctcccaggc	60
	tatccaggag	gggccaagag	attaaaagca	ggttcagaag	gctcagatgc	cactcaccag	120
30	acagcagggt	cgactgctag	tgaccttgag	cccagtcagg	acagacagac	aggcagacag	180
	acgcacggac	aagcagatgc	tccttggcca	gctatccact	cttctgtgcc	tgcttagcgg	240
	ggccctgcct	acaggctcag	ggaggcctga	accccagtc	cctcgacctc	agtcctgggc	300
	tgagccaat	cagacctggg	ctctgggccc	aggggacctg	ccccactgg	tgccagcttc	360
	tgcccttggg	agctggaagg	ccttcttggg	cctgcagaaa	gccaggcagc	tggggatggg	420
	caggctgcag	cgtgggcaag	acgaggtggc	tgctgtgact	ctgccgctga	accctcagga	480
35	agtgatccag	gggatgtgta	aggctgtgcc	cttcgttcag	acacgggagt	ctcgctatgt	540
	tgcccaagct	agtcttgagc	ttctggcctc	aagcaatcct	cccacctcag	cctcctgagt	600
	tctagggtgt	ctcccgccgc	ggctgctcag	ccatacgcc	ccgaaatcat	ctgtgctttg	660
	gtcattgctc	ctctctctac	atccctggct	cggacccccc	cccactagtc	ctgtgcaaca	720
	gctgtatgcc	tgctcgcaag	cgttgggcac	ccgtggctct	gtgggtgtct	actggcagct	780
	cagcctcccg	tcgacgggtg	aagatatcca	ccatgctgat	cgagggggtg	cactgcagcc	840
40	caaaagcatg	aactgagcat	ctggatgggt	gcacggagac	acgcaccttg	gagaaatgag	900
	gggagatgga	ccaagaaaga	cgtggacctg	gatgatgtac	tctgggtcaa	gagaccaggg	960
	atgcagggtt	aggcagacag	gtccccagag	tcctcaccct	gctccccaga	cagtagacac	1020
	agtgcccgct	ctggagttgc	accactgata	gtcacagcac	acaatgattg	acaactcact	1080
	tttttttttt	tttttgagat	ggagtctcgc	tctgtcgccc	aggctggagt	gcagtggcgc	1140
45	aatctcagct	cactgcaagc	tccacctccc	gggtttatgc	cattctcctg	tctcagcctc	1200
	ccgagtagct	gggactacag	gcacccgcca	acacgcccgg	staatttttt	gtatttttag	1260
	taaagacagg	gtttcacctg	gttagccagg	atggctctcta	tctcctgacc	tcgtgatctg	1320
	cctgccttgg	ccttattatt	ttttttttta	ggacagagtc	tctctctgtc	accaggctg	1380
	gagtgcaatg	gcgcgatctt	ggctcactgt	aacttccact	tgccaggctc	aagcagttct	1440
	cctgcctcag	cctcctgagt	agctgggact	acaggcaccc	gccaccatgc	ccagctaatt	1500
50	tttgatattt	tagtagagac	agagtttcac	catattagcc	tggtgggtct	caaactcctg	1560
	gcctcagggt	atctgcccac	ctcggcctcc	caaagtgtcg	ggatcaaatc	cactgttaat	1620
	cattaggctg	aactgtctct	tatagaatga	ggtcaaaagc	actcccagtt	gcagggaggg	1680
	tagatggccc	cacccagacc	gagagacaca	gtgatgacct	cagcctaggg	acmccaaaaa	1740
	aaaaaaaaaa	aaaaaa					1756

55 <210> 37

<211> 133

<212> PRT

<213> Secuencia artificial

60 <220>

<223> *Homo sapiens*

65 <220>

<221> Producto de traslación pronosticado de SEQ ID:36

ES 2 306 802 T3

<400> 37

```

5      Met Leu Leu Gly Gln Leu Ser Thr Leu Leu Cys Leu Leu Ser Gly Ala
      1          5          10          15

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

10     Ser Trp Ala Ala Asn Gln Thr Trp Ala Leu Gly Pro Gly Ala Leu
          35          40          45

      Pro Pro Leu Val Pro Ala Ser Ala Leu Gly Ser Trp Lys Ala Phe Leu
          50          55          60

15     Gly Leu Gln Lys Ala Arg Gln Leu Gly Met Gly Arg Leu Gln Arg Gly
      65          70          75          80

      Gln Asp Glu Val Ala Ala Val Thr Leu Pro Leu Asn Pro Gln Glu Val
      85          90          95

      Ile Gln Gly Met Cys Lys Ala Val Pro Phe Val Gln Thr Arg Glu Ser
          100          105          110

25     Arg Tyr Val Ala Gln Ala Ser Leu Glu Leu Leu Ala Ser Ser Asn Pro
          115          120          125

      Pro Thr Ser Ala Ser
          130

```

<210> 38

<211> 589

<212> DNA

<213> *Homo sapiens*

<220>

<221> INSP002 de longitud completa donado a partir de cDNA de corazón humano

<400> 38

```

45     gatgctcctt ggccagctat ccactcttct gtgcctgctt agcggggccc tgcctacagg      60
      ctcagggagg cctgaacccc agtctcctcg acctcagtc tgggctgcag ccaatcagac      120
      ctgggctctg ggcccagggg ccctgcccc actggtgcc gcttctgccc ttgggagctg      180
      gaaggccttc ttgggcctgc agaaagccag gcagctgggg atgggcaggc tgcagcgtgg      240
      gcaagacgag gtggtctgctg tgactctgcc gctgaaccct caggaagtga tccaggggat      300
50     gtgtaaggct gtgcccttcg ttcagggtgt ctcccggccc ggctgctcag ccatacgcct      360
      ccgaaatcat ctgtgctttg gtcattgctc ctctctctac atccctggct cggaccccac      420
      ccactagtc ctgtgcaaca gctgtatgcc tgctcgcaag cgttgggcac ccgtggtcct      480
      gtggtgtctc actggcagct cagcctcccg tcgacgggtg aagatatcca ccatgctgat      540
      cgaggggtgt cactgcagcc caaaagcatg aactgagcat cgtggatgg      589

```