



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 285 768**

(51) Int. Cl.:

C12N 15/12 (2006.01)

C07K 14/47 (2006.01)

C07K 16/18 (2006.01)

G01N 33/50 (2006.01)

G01N 33/53 (2006.01)

C12Q 1/68 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **98918453 .6**

(86) Fecha de presentación : **16.04.1998**

(87) Número de publicación de la solicitud: **0975755**

(87) Fecha de publicación de la solicitud: **02.02.2000**

(54)

Título: **Proteínas CRSP (proteínas segregadas ricas en cisteína), moléculas de ácido nucleico que codifican para las mismas y uso de las mismas.**

(30)

Prioridad: **16.04.1997 US 843704**
17.04.1997 US 842898
15.01.1998 US 71589 P
20.01.1998 US 9802

(73)

Titular/es: **MILLENNIUM PHARMACEUTICALS, Inc.**
40 Landsdowne Street
Cambridge, Massachusetts 02139, US

(45)

Fecha de publicación de la mención BOPI:
16.11.2007

(72)

Inventor/es: **McCarthy, Sean, A.**

(45)

Fecha de la publicación del folleto de la patente:
16.11.2007

(74)

Agente: **Toro Gordillo, Ignacio María**

ES 2 285 768 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

Proteínas CRSP (proteínas segregadas ricas en cisteína), moléculas de ácido nucleico que codifican para las mismas y usos de las mismas.

Antecedentes de la invención

Las proteínas segregadas desempeñan un papel íntegro en la formación, diferenciación y mantenimiento de células en organismos multicelulares. Por ejemplo, en la técnica se sabe que las proteínas secretoras están implicadas en la señalización entre células que están calientes en contacto directo. Tales moléculas de señalización segregadas son particularmente importantes en el desarrollo de tejido de vertebrados durante la embriogénesis así como en el mantenimiento del estado diferenciado de tejidos adultos. Por ejemplo, las interacciones inductivas que se producen entre capas celulares vecinas y los tejidos en el embrión en desarrollo son sumamente dependientes de la existencia y regulación de moléculas de señalización segregadas. En interacciones inductivas, las señales bioquímicas segregadas por una población celular influyen en el destino del desarrollo de una segunda población celular, normalmente alterando el destino de la segunda población celular. Por ejemplo, las proteínas Wnt se reconocen en la actualidad como una de las principales familias de moléculas de señalización importantes para el desarrollo en organismos que varían desde *Drosophila* hasta ratones. Para una revisión véase Cadigan, K. M. y R. Nusse (1997) *Genes & Development*, 11:3286-3305.

También se reconoce bien actualmente que las principales familias de moléculas de señalización tienen un papel doble que desempeñar tanto en el desarrollo de un organismo así como en la estimulación o mantenimiento del estado diferenciado de tejidos en el animal adulto. Además, las principales familias de moléculas de señalización se han implicado en el control de la proliferación de células en tejidos adultos maduros, por ejemplo, durante la renovación celular normal en el organismo adulto así como en la regeneración tisular activada como resultado del daño al tejido adulto.

Sumario de la invención

La presente invención se basa, al menos en parte, en el descubrimiento de moléculas de ácido nucleico que codifican para una familia novedosa de proteínas segregadas, denominadas en el presente documento proteínas segregadas ricas en cisteína ("CRSP" o "proteínas CRSP"). Las moléculas CRSP de la presente invención son útiles como agentes moduladores para regular una variedad de procesos celulares. Consecuentemente, en un aspecto, esta invención proporciona moléculas de ácido nucleico aisladas que codifican para proteínas CRSP o partes biológicamente activas de las mismas, así como fragmentos de ácido nucleico adecuados como cebadores o sondas de hibridación para la detección de ácidos nucleicos que codifican para CRSP.

En una realización, la invención proporciona una molécula de ácido nucleico de CRSP tal como se define en las reivindicaciones adjuntas.

Otro aspecto de la invención proporciona un vector que comprende una molécula de ácido nucleico de CRSP. En ciertas realizaciones, el vector es un vector de expresión recombinante. En otra realización, la invención proporciona una célula huésped que contiene un vector de la invención. La invención también proporciona un método para producir una proteína CRSP cultivando en un medio adecuado una célula huésped de la invención que contiene un vector de expresión recombinante de manera que se produce una proteína CRSP.

Otro aspecto de esta invención muestra proteínas CRSP aisladas o recombinantes y polipéptidos. En una realización, una proteína CRSP aislada tiene una secuencia señal y al menos un dominio rico en cisteína, preferiblemente una región rica en cisteína, y se segrega. En otra realización, una proteína CRSP aislada tiene una secuencia de aminoácidos tal como se define en las reivindicaciones adjuntas.

Las proteínas CRSP de la presente invención, o partes biológicamente activas de las mismas, pueden estar operativamente unidas a un polipéptido distinto de CRSP para formar proteínas de fusión de CRSP. La invención muestra adicionalmente anticuerpos que se unen específicamente a proteínas CRSP, tales como anticuerpos monoclonales o policlonales. Además, las proteínas CRSP o partes biológicamente activas de las mismas pueden incorporarse a composiciones farmacéuticas, que opcionalmente incluyen vehículos farmacéuticamente aceptables.

En otro aspecto, la presente invención proporciona un método para detectar la expresión de CRSP en una muestra biológica poniendo en contacto la muestra biológica con un agente que puede detectar un polipéptido, una proteína o una molécula de ácido nucleico de CRSP, de manera que se detecta la presencia de un polipéptido, una proteína o una molécula de ácido nucleico de CRSP en la muestra biológica.

En otro aspecto, la presente invención proporciona un método para detectar la presencia de actividad de CRSP en una muestra biológica poniendo en contacto la muestra biológica con un agente que puede detectar un indicador de actividad de CRSP de manera que se detecta la presencia de actividad de CRSP en la muestra biológica.

En otro aspecto, la invención proporciona un método para modular la actividad de CRSP que comprende poner en contacto la célula con un agente que modula la actividad de CRSP de tal manera que se modula la actividad de

CRSP en la célula. En una realización, el agente inhibe la actividad de CRSP. En otra realización, el agente estimula la actividad de CRSP. En una realización, el agente es un anticuerpo que se une específicamente a una proteína CRSP. En otra realización, el agente modula la expresión de CRSP modulando la transcripción de un gen de CRSP o la traducción de un ARNm de CRSP. Todavía en otra realización, el agente es una molécula de ácido nucleico que tiene una secuencia de nucleótidos que es antisentido con respecto a la cadena codificante de un ARNm de CRSP o un gen de CRSP.

Los métodos basados en la presente invención pueden usarse para tratar a un sujeto que tiene un trastorno caracterizado por actividad o expresión de ácido nucleico o proteína CRSP aberrante administrando un agente que es un modulador de CRSP al sujeto. En una realización, el modulador de CRSP es una proteína CRSP. En otra realización el modulador de CRSP es una molécula de ácido nucleico de CRSP. Todavía en otra realización, el modulador de CRSP es un péptido, molécula peptidomimética u otra molécula pequeña. En una realización preferida el trastorno se caracteriza por expresión de ácido nucleico o proteína CRSP aberrante es un trastorno del desarrollo, de diferenciación o proliferativo.

La presente invención también proporciona un ensayo diagnóstico para identificar la presencia o ausencia de una alteración genética caracterizada por al menos una de (i) mutación o modificación aberrante de un gen que codifica para una proteína CRSP; (ii) regulación errónea de dicho gen; y (iii) modificación posterior a la traducción aberrante de una proteína CRSP, en la que una forma de tipo natural de dicho gen codifica para una proteína con una actividad de CRSP.

En otro aspecto, la invención proporciona un método para identificar un compuesto que se une a, o modula la actividad de, una proteína CRSP, proporcionando una composición indicadora que comprende una proteína CRSP que tiene actividad de CRSP, poniendo en contacto la composición indicadora con un compuesto de prueba, y determinando el efecto del compuesto de prueba sobre la actividad de CRSP en la composición indicadora para identificar un compuesto que modula la actividad de una proteína CRSP.

Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada y reivindicaciones.

Breve descripción de los dibujos

La figura 1 representa la secuencia de ADNc y la secuencia de aminoácidos prevista de CRSP-1 humana. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 2479 de SEQ ID NO: 1. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 350 de SEQ ID NO: 2.

La figura 2 representa la secuencia de ADNc y secuencia de aminoácidos prevista de CRSP-2 humana. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 848 de SEQ ID NO: 4. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 224 de SEQ ID NO: 5.

La figura 3 representa la secuencia de ADNc y la secuencia de aminoácidos prevista de CRSP-3 humana. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 1529 de SEQ ID NO: 7. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 266 de SEQ ID NO: 8.

La figura 4 representa la secuencia de ADNc y la secuencia de aminoácidos prevista de CRSP-4 humana. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 702 de SEQ ID NO: 10. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 179 de SEQ ID NO: 11.

La figura 5 representa la secuencia de ADNc y la secuencia de aminoácidos prevista de N terminal de tipo CRSP humana. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 928 de SEQ ID NO: 13. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 242 de SEQ ID NO: 14.

La figura 6 representa las secuencias de aminoácidos de CRSP-1 humana, CRSP-2 humana, CRSP-3 humana, y CRSP-4 humana, así como la secuencia consenso generada a partir de la alineación de estas secuencias. La figura detalla la relación entre las proteínas CRSP de la presente invención.

La figura 7 es un diagrama esquemático que representa la relación entre las proteínas CRSP de la presente instantánea. La figura representa los dominios biológicos y funcionales de CRSP humana. Los péptidos señal se indican mediante cajas rellenas. Los dominios ricos en cisteína de una región rica en cisteína de CRSP se representan como CRD-1 y CRD-2.

La figura 8 representa la secuencia de ADNc y la secuencia de aminoácidos prevista de CRSP-1 murina. La secuencia de nucleótidos corresponde a los ácidos nucleicos 1 a 928 de SEQ ID NO: 16. La secuencia de aminoácidos corresponde a los aminoácidos 1 a 349 de SEQ ID NO: 17.

La figura 9 es un diagrama esquemático que representa la relación entre la secuencia de nucleótidos de CRSP-1 y las de RIG y 7-1 de tipo RIG (números de registro U32331 y AF034208, respectivamente).

Descripción detallada de la invención

La presente invención se basa en el descubrimiento de moléculas novedosas, denominadas en el presente documento como moléculas de ácido nucleico y proteínas CRSP, que comprenden una familia de moléculas que tienen ciertas características estructurales y funcionales conservadas. Se pretende que el término “familia” cuando se refiere a la proteína y moléculas de ácido nucleico de la invención signifique dos o más proteínas o moléculas de ácido nucleico que tienen un dominio estructural común y que tienen suficiente homología de secuencia de aminoácidos o nucleótidos tal como se define en el presente documento. Tales miembros de familia pueden producirse de manera natural y pueden ser o bien de la misma especie o bien de diferentes especies. Por ejemplo, una familia puede contener una primera proteína de origen humano, así como otras proteínas distintas de origen humano, o alternativamente pueden contener homólogos de origen no humano. Los miembros de una familia también pueden tener características funcionales comunes.

En una realización, se identifica un miembro de la familia de CRSP basándose en la presencia de al menos un “domino rico en cisteína” en la proteína o molécula de ácido nucleico correspondiente. Tal como se define en el presente documento, un “dominio rico en cisteína” se refiere a una parte de una proteína CRSP (por ejemplo, CRSP-1) que es rica en residuos de cisteína, es decir, al menos aproximadamente 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ó 20 aminoácidos son residuos de cisteína. Preferiblemente, un “dominio rico en cisteína” es un dominio proteico que tiene una secuencia de aminoácidos de aproximadamente 30-100 aminoácidos, de manera preferible de aproximadamente 35-95 aminoácidos, de manera más preferible de aproximadamente 40-90 aminoácidos, de manera más preferible de aproximadamente 45-85 aminoácidos, de manera incluso más preferible de aproximadamente 50-80 aminoácidos, y de manera incluso más preferible de aproximadamente 55-75, 60-70, ó 65 aminoácidos, de los que al menos aproximadamente 3-20, de manera preferible aproximadamente 5-15, o de manera más preferible aproximadamente 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ó 20 aminoácidos son residuos de cisteína.

Un dominio rico en cisteína preferido tiene al menos aproximadamente 10, al menos aproximadamente 15, 16, 17, 18, 19, o preferiblemente 20 residuos de cisteína ubicados en la misma posición de aminoácido relativa que los residuos de cisteína en CRSP-1 humana que tiene SEQ ID NO: 2. Por ejemplo, tal como se muestra en la figura 6, CRSP-2 tiene al menos aproximadamente 10 residuos de cisteína ubicados en la misma posición de aminoácido relativa que los residuos de cisteína en CRSP-1 humana que tiene SEQ ID NO: 2 (por ejemplo, cys151 en CRSP-2, SEQ ID NO: 5, está ubicada en la misma posición de aminoácido relativa que cys214 en CRSP-1, SEQ ID NO: 2; cys156 en CRSP-2, SEQ ID NO: 5, está ubicada en la misma posición de aminoácido relativa que cys219 en CRSP-1, SEQ ID NO: 2; y cys157 en CRSP-2, SEQ ID NO: 5, está ubicada en la misma posición de aminoácido que cys220 en CRSP-1, SEQ ID NO: 2). De manera similar, tal como se muestra en la figura 6, CRSP-3 tiene al menos aproximadamente 10 residuos de cisteína ubicados en la misma posición de aminoácido relativa que los residuos de cisteína en CRSP-1 humana que tiene SEQ ID NO: 2. Tal como también se muestra en la figura 6, CRSP-4 tiene al menos aproximadamente 10 residuos de cisteína ubicadas en la misma posición de aminoácido relativa que los residuos de cisteína en CRSP-1 humana que tiene SEQ ID NO: 2.

Una proteína CRSP preferida de la presente invención es una proteína humana (por ejemplo, codificada por una secuencia de nucleótidos que corresponde a un gen humano que se produce de manera natural). Por ejemplo, en una realización, una proteína CRSP-1 (CRSP-1 humana) contiene un primer dominio rico en cisteína que contiene aproximadamente los aminoácidos 147-195 de SEQ ID NO: 2, que tiene 10 residuos de cisteína, y un segundo dominio rico en cisteína que contiene aproximadamente los aminoácidos 201-284 de SEQ ID NO: 2, que tiene 10 residuos de cisteína (las posiciones de los residuos de cisteína se representan en la figura 6). En otra realización, una proteína CRSP-2 (CRSP-2 humana) contiene un primer dominio rico en cisteína que contiene aproximadamente los aminoácidos 41-90 de SEQ ID NO: 5, que tiene 10 residuos de cisteína, y un segundo dominio rico en cisteína que contiene aproximadamente los aminoácidos 41-218 de SEQ ID NO: 5, que tiene 10 residuos de cisteína (las posiciones de los residuos de cisteína se representan en la figura 6). En otra realización, una proteína CRSP-3 (proteína CRSP-3 humana) contiene un primer dominio rico en cisteína que contiene aproximadamente los aminoácidos 85-138 de SEQ ID NO: 8, que tiene 10 residuos de cisteína, y un segundo dominio rico en cisteína que contiene aproximadamente los aminoácidos 182-263 de SEQ ID NO: 8, que tiene 10 residuos de cisteína (las posiciones de los residuos de cisteína se representan en la figura 6). En otra realización, una proteína CRSP-4 (CRSP-4 humana) contiene un primer dominio rico en cisteína que contiene aproximadamente los aminoácidos 1-47 de SEQ ID NO: 11, que tiene 8 residuos de cisteína, y un segundo dominio rico en cisteína que contiene aproximadamente los aminoácidos 96-176 de SEQ ID NO: 11, que tiene 10 residuos de cisteína (las posiciones de los residuos de cisteína se representan en la figura 6).

Proteínas CRSP preferidas tienen más de un dominio rico en cisteína, más preferiblemente tienen al menos dos dominios ricos en cisteína y, por tanto, tienen una región rica en cisteína. Tal como se usa en el presente documento, el término “región rica en cisteína” se refiere a un dominio proteico que incluye al menos dos dominios ricos en cisteína y tiene una secuencia de aminoácidos de aproximadamente 120-200, de manera preferible de aproximadamente 130-190, de manera más preferible de aproximadamente 140-180 residuos de aminoácidos, y de manera incluso más preferible de al menos aproximadamente 135-175 aminoácidos de los que al menos aproximadamente 10-30, de manera preferible aproximadamente 15-20, y de manera más preferible aproximadamente 16, 17, 18, ó 19 de los aminoácidos son residuos de cisteína. En una realización preferida, una región rica en cisteína está ubicada en la región C-terminal de una proteína CRSP. Por ejemplo, en una realización, una proteína CRSP-1 contiene una región rica en cisteína que contiene aproximadamente los aminoácidos 147-284 de SEQ ID NO: 2, que tiene 20 residuos de cisteína en las posiciones indicadas en la figura 6. En otra realización, una proteína CRSP-2 contiene una región rica en cisteína

que contiene aproximadamente los aminoácidos 41-218 de SEQ ID NO: 5, que tiene 20 residuos de cisteína en las posiciones indicadas en la figura 6. En otra realización, una proteína CRSP-3 contiene una región rica en cisteína que contiene aproximadamente los aminoácidos 85-263 de SEQ ID NO: 8, que tiene 20 residuos de cisteína en las posiciones indicadas en la figura 6. En otra realización, una proteína CRSP-4 contiene una región rica en cisteína que contiene aproximadamente los aminoácidos 1-176 de SEQ ID NO: 2, que tiene 18 residuos de cisteína en las posiciones indicadas en la figura 6.

En otra realización, además de los dominios ricos en cisteína, la región rica en cisteína contiene una región espaciadora que separa el primer y segundo dominio rico en cisteína. Tal como se usa en el presente documento, la “región espaciadora” se refiere a residuos de aminoácidos que están ubicados entre los dominios ricos en cisteína primero y segundo de una región rica en cisteína e incluye residuos de aminoácidos ubicados en C-terminal con respecto al primer dominio rico en cisteína y en N-terminal con respecto al segundo dominio rico en cisteína. Tal como se define en el presente documento, una “región espaciadora” se refiere a un dominio proteico de aproximadamente 5-70, de manera preferible de aproximadamente 10-65 aminoácidos, de manera más preferible de aproximadamente 15-60 aminoácidos, de manera incluso más preferible de aproximadamente 20-55, y de manera incluso más preferible de aproximadamente 25-50, 30-45 ó 35-40 aminoácidos. Por ejemplo, la proteína CRSP-1 contiene una región espaciadora de aproximadamente los aminoácidos 196-200 de SEQ ID NO: 2; la proteína CRSP-2 contiene una región espaciadora de aproximadamente los aminoácidos 91-137 de SEQ ID NO: 5; la proteína CRSP-3 contiene una región espaciadora de aproximadamente los aminoácidos 139-181 de SEQ ID NO: 8; y la proteína CRSP-4 contiene una región espaciadora de aproximadamente los aminoácidos 48-95 de SEQ ID NO: 11.

En otra realización de la invención, la proteína CRSP tiene al menos un dominio rico en cisteína, preferiblemente una región rica en cisteína, y una secuencia señal. Tal como se usa en el presente documento, una “secuencia señal” se refiere a un péptido que contiene aproximadamente 20 aminoácidos que se produce en el extremo N-terminal de proteínas de membrana secretoras e integrales y que contiene un gran número de residuos de aminoácidos hidrófobos. Por ejemplo, una secuencia señal contiene al menos aproximadamente 14-28 residuos de aminoácidos, de manera preferible aproximadamente 16-26 residuos de aminoácidos, de manera mas preferible aproximadamente 18-24 residuos de aminoácidos, y de manera más preferible aproximadamente 20-22 residuos de aminoácidos, y tiene al menos aproximadamente el 40-70%, de manera preferible aproximadamente el 50-65%, y de manera más preferible aproximadamente el 55-60% de residuos de aminoácidos hidrófobos (por ejemplo, alanina, valina, leucina, isoleucina, fenilalanina, tirosina, triptófano, o prolina). Tal “secuencia señal”, también se denomina en la técnica como un “péptido señal”, sirve para dirigir una proteína que contiene una secuencia de este tipo hacia una bicapa lipídica. Por ejemplo, en una realización, una proteína CRSP-1 contiene una secuencia señal de aproximadamente los aminoácidos 1-23 de SEQ ID NO: 2. En otra realización, una proteína CRSP-2 contiene una secuencia señal de aproximadamente los aminoácidos 1-19 de SEQ ID NO: 5. En otra realización, una proteína CRSP-3 contiene una secuencia señal de aproximadamente los aminoácidos 1-20 de SEQ ID NO: 8.

Las proteínas CRSP de la presente invención pueden usarse para identificar miembros adicionales de la familia de CRSP. Por ejemplo, se identificó una proteína que tiene homología con CRSP-1 usando la secuencia que codifica para la región única N-terminal de CRSP-1 para buscar una base de datos de secuencias de proteínas. Una proteína de este tipo, denominada en el presente documento como “N terminal de tipo CRSP” se representa en la figura 5. La secuencia de nucleótidos de N terminal de tipo CRSP (SEQ ID NO: 13) codifica para una proteína que tiene 242 aminoácidos (SEQ ID NO: 14). La secuencia de nucleótidos de N terminal de tipo CRSP incluye una región no traducida en 5' que contiene los nucleótidos 1-74 de SEQ ID NO: 13, una región codificante que contiene los nucleótidos 75-800 de SEQ ID NO: 13 (que corresponde a los nucleótidos 1-726 de SEQ ID NO: 15), y una región no traducida en 3' que contiene los nucleótidos 801-928 de SEQ ID NO: 13.

Consecuentemente, una realización de la invención muestra una proteína CRSP que tiene al menos un dominio rico en cisteína, preferiblemente al menos una región rica en cisteína. Otra realización muestra una proteína CRSP que tiene al menos una región rica en cisteína, en la que la región rica en cisteína incluye al menos un dominio rico en cisteína. Otra realización muestra una proteína CRSP que tiene al menos una región rica en cisteína, en la que la región rica en cisteína incluye al menos dos dominios ricos en cisteína. Otra realización muestra una proteína que tiene homología del 20, 30, 40, 50, 60, 70, 80, 90, 95, ó 99% con respecto a un dominio rico en cisteína de una proteína CRSP de la invención (por ejemplo, CRSP-1, CRSP-2, CRSP-3, o CRSP-4).

Todavía otra realización de la invención muestra una proteína CRSP que tiene al menos un dominio rico en cisteína, preferiblemente al menos una región rica en cisteína y un péptido señal. Otra realización muestra una proteína CRSP que tiene al menos un dominio rico en cisteína, preferiblemente al menos una región rica en cisteína y un péptido señal, en la que la región rica en cisteína incluye al menos dos dominios ricos en cisteína. Otra realización muestra una proteína CRSP que tiene al menos un dominio rico en cisteína, preferiblemente al menos una región rica en cisteína y un péptido señal, en la que la región rica en cisteína incluye al menos dos dominios ricos en cisteína y un espaciador.

Moléculas CRSP preferidas de la presente invención se definen en las reivindicaciones adjuntas. Tal como se usa en el presente documento, el término “suficientemente homóloga” se refiere a una primera secuencia de aminoácidos o de nucleótidos que contiene un número suficiente o mínimo de residuos de aminoácidos o de nucleótidos idénticos o equivalentes (por ejemplo, un residuo de aminoácidos que tiene una cadena lateral similar) con respecto a una segunda secuencia de aminoácidos o nucleótidos de manera que las secuencias de aminoácidos o nucleótidos primera y segunda compartan dominios estructurales comunes y/o una actividad funcional común. Por ejemplo, las secuencias

de aminoácidos o nucleótidos que comparten dominios estructurales comunes tienen al menos aproximadamente un 40% de homología, preferiblemente un 50% de homología, más preferiblemente un 60%-70% de homología a lo largo de la secuencia de aminoácidos de los dominios y contienen al menos un, preferiblemente dos, más preferiblemente tres, e incluso más preferiblemente cuatro, cinco o seis dominios estructurales, se definen en el presente documento como suficientemente homólogas. Además, las secuencias de aminoácidos o nucleótidos que comparten al menos un 40%, preferiblemente un 50%, más preferiblemente un 60, 70, u 80% de homología y comparten una actividad funcional se definen en el presente documento como suficientemente homólogas.

Tal como se usa indistintamente en el presente documento, una “actividad de CRSP”, “actividad biológica de CRSP” o “actividad funcional de CRSP”, se refiere a una actividad ejercida por una molécula de ácido nucleico, polipéptido o proteína CRSP sobre una célula sensible a CRSP tal como se determina *in vivo*, o *in vivo*, según las técnicas habituales. En una realización, una actividad de CRSP es una actividad directa, tal como una asociación con una molécula diana de CRSP. Tal como se usa en el presente documento, una “molécula diana” es una molécula con la que se une o interactúa una proteína CRSP en la naturaleza, de manera que se logra función mediada por CRSP. Una molécula diana de CRSP puede ser una molécula distinta de CRSP o un polipéptido o proteína CRSP de la presente invención. En una realización a modo de ejemplo, una molécula diana de CRSP es una proteína unida a membrana (por ejemplo, un receptor de superficie celular o “receptor de CRSP”) o una forma modificada de una proteína de este tipo que se ha alterado de manera que la proteína es soluble (por ejemplo, producida recombinantemente de manera que la proteína no exprese un dominio de unión a membrana). En otra realización, una diana de CRSP es una segunda molécula de proteína soluble (por ejemplo, una “pareja de unión a CRSP” o “sustrato de CRSP”). En una realización de este tipo a modo de ejemplo, una pareja de unión a CRSP puede ser una segunda proteína distinta de CRSP soluble o una segunda molécula de proteína CRSP de la presente invención. Alternativamente, una actividad de CRSP es una actividad indirecta, tal como una actividad de señalización celular mediada por interacción de la proteína CRSP con una segunda proteína (por ejemplo, un receptor de CRSP). Tal como se usa en el presente documento, el término “receptor de CRSP” se refiere a una proteína o complejo de proteína, al que puede unirse una proteína CRSP, por ejemplo, CRSP humana. Un receptor puede ser un receptor de superficie celular, por ejemplo, un receptor de hormona nuclear. Los receptores de CRSP pueden aislarse mediante métodos conocidos en la técnica y descritos adicionalmente en el presente documento. La interacción de una proteína CRSP con un receptor de CRSP puede resultar en la transducción de una señal de la superficie celular con respecto al núcleo. La señal transducida puede ser, un aumento del calcio intracelular, un aumento de fosfatidilinositol u otra molécula, y puede dar como resultado, por ejemplo, la fosforilación de proteínas específicas, una modulación de transcripción génica y cualquiera de las otras actividades biológicas expuestas en el presente documento.

En una realización preferida, una actividad de CRSP es al menos una o más de las siguientes actividades: (i) interacción de una proteína CRSP con, y/o unión a, una segunda molécula, (por ejemplo, una proteína, tal como un receptor de CRSP, una forma soluble de un receptor de CRSP, un receptor para un miembro de la familia de *wnt* de las proteínas de señalización, o una molécula de señalización distinta de CRSP); (ii) interacción de una proteína CRSP con una proteína intracelular a través de un receptor de CRSP unido a membrana; (iii) formación de complejo entre una proteína CRSP soluble y una segunda pareja de unión a CRSP soluble (por ejemplo, una molécula de proteína distinta CRSP o una segunda molécula de proteína CRSP); (iv) interacción con otras proteínas extracelulares (por ejemplo, regulación de adhesión celular dependiente de *wnt* a componentes de matriz extracelular); (v) unión a y eliminación de una molécula no deseable (por ejemplo, una actividad desintoxicante o función de defensa); y/o (vi) una actividad enzimática. Todavía en otra realización preferida, una actividad de CRSP es al menos una o más de las siguientes actividades: (1) modulación de transducción de señales celulares, o bien *in vitro* o bien *in vivo* (por ejemplo, antagonismo de la actividad de miembros de la familia de *wnt* de proteínas segregadas o supresión de transducción de señales dependiente de *wnt*); (2) regulación de comunicación entre células (por ejemplo, regulación de interacciones célula-célula dependientes de *wnt*); (3) regulación de expresión de genes cuya expresión está modulada por la unión de CRSP (por ejemplo, CRSP-1) a un receptor; (4) regulación de transcripción génica en una célula implicada en el desarrollo o diferenciación, o bien *in vitro* o bien *in vivo* (por ejemplo, inducción de diferenciación celular); (5) regulación de transcripción génica en una célula implicada en el desarrollo o diferenciación, en la que al menos un gen codifica para proteína específica de diferenciación; (6) regulación de transcripción génica en una célula implicada en el desarrollo o diferenciación, en la que al menos un gen codifica para una segunda proteína segregada; (7) regulación de transcripción génica en una célula implicada en el desarrollo o diferenciación, en la que al menos un gen codifica para una molécula de transducción de señales; (8) regulación de proliferación celular, o bien *in vitro* o bien *in vivo* (por ejemplo, inducción de proliferación celular o inhibición de proliferación como en el caso de supresión de tumorigénesis (por ejemplo, supresión de formación de glioblastoma)); (9) formación y mantenimiento de disposiciones espaciales ordenadas de tejidos diferenciados en vertebrados, tanto adultos como embrionarios (por ejemplo, inducción de formación de cabeza durante el desarrollo del vertebrado o mantenimiento de células progenitoras hematopoyéticas); (10) modulación de muerte celular, tal como estimulación de supervivencia celular; (11) regulación de migración celular; y/o (12) modulación inmunitaria.

Tal como se denomina en el presente documento, “proteínas específicas de diferenciación” incluyen proteínas implicadas en la transición de una célula desde el fenotipo no diferenciado hasta el diferenciado. Por ejemplo, tales proteínas pueden ser proteínas estructurales específicas de diferenciación o factores de transcripción específicos de diferenciación. Tales proteínas específicas de diferenciación se expresan generalmente a niveles superiores en células que están haciendo la transición desde el fenotipo no diferenciado al fenotipo diferenciado (por ejemplo, durante el desarrollo embrionario o durante la regeneración de tejido maduro en el animal adulto), o se expresan a niveles superiores en células totalmente diferenciadas o diferenciadas de manera terminal comparadas con sus contrapartes

no diferenciadas. También, tal como se denomina en el presente documento, “genes específicos de diferenciación” incluyen moléculas de ácido nucleico que codifican para proteínas específicas de diferenciación.

Por consiguiente, otra realización de la invención muestra proteínas CRSP aisladas y polipéptidos que tienen una actividad de CRSP. Proteínas CRSP preferidas tienen al menos una región rica en cisteína y una actividad de CRSP. En otra realización preferida, la proteína CRSP tiene al menos una región rica en cisteína, en la que la región rica en cisteína comprende al menos un dominio rico en cisteína, y una actividad de CRSP. En otra realización preferida, la proteína CRSP tiene al menos una región rica en cisteína, en la que la región rica en cisteína comprende al menos dos dominios ricos en cisteína, y una actividad de CRSP. Todavía en otra realización preferida, una proteína CRSP comprende además una secuencia señal. Aún en otra realización preferida, una proteína CRSP tiene una región rica en cisteína, una actividad de CRSP, y una secuencia de aminoácidos tal como se define en las reivindicaciones adjuntas.

El ADNc de CRSP-1 humana, que tiene aproximadamente 2479 nucleótidos de longitud, codifica para una proteína que tiene aproximadamente 350 residuos de aminoácidos de longitud. La proteína CRSP humana contiene una secuencia señal N-terminal y una región rica en cisteína que comprende dos dominios ricos en cisteína. Puede encontrarse una región rica en cisteína de CRSP al menos, por ejemplo, desde aproximadamente los aminoácidos 147-284 de SEQ ID NO: 2. La región rica en cisteína de CRSP-1 comprende un primer dominio rico en cisteína desde aproximadamente los aminoácidos 147-195 de SEQ ID NO: 2 y un segundo dominio rico en cisteína desde aproximadamente los aminoácidos 201-284 de SEQ ID NO: 2. La proteína CRSP-1 humana es una proteína segregada que contiene adicionalmente una secuencia señal en aproximadamente los aminoácidos 1-23 de SEQ ID NO: 2. La predicción de un péptido señal de este tipo puede hacerse, por ejemplo, utilizando el algoritmo informático SIGNALP (Henrik, *et al.* (1997) Protein Engineering 10:1-6).

El ADNc de CRSP-2 humana, que tiene aproximadamente 848 nucleótidos de longitud, codifica para una proteína que tiene aproximadamente 224 residuos de aminoácidos de longitud. La proteína CRSP humana contiene una secuencia señal N-terminal y una región rica en cisteína que comprende dos dominios ricos en cisteína. Puede encontrarse una región rica en cisteína de CRSP al menos, por ejemplo, desde aproximadamente los aminoácidos 41-218 de SEQ ID NO: 5. La región rica en cisteína de CRSP-2 comprende un primer dominio rico en cisteína desde aproximadamente los aminoácidos 41-90 de SEQ ID NO: 5 y un segundo dominio rico en cisteína desde aproximadamente los aminoácidos 138-218 de SEQ ID NO: 5. La proteína CRSP-2 humana es una proteína segregada que contiene adicionalmente una secuencia señal en aproximadamente los aminoácidos 1-19 de SEQ ID NO: 5.

El ADNc de CRSP-3, que tiene aproximadamente 1529 nucleótidos de longitud, codifica para una proteína que tiene aproximadamente 266 residuos de aminoácidos de longitud. La proteína CRSP humana contiene una secuencia señal N-terminal y una región rica en cisteína que comprende dos dominios ricos en cisteína. Puede encontrarse una región rica en cisteína de CRSP al menos, por ejemplo, desde aproximadamente los aminoácidos 85-263 de SEQ ID NO: 8. La región rica en cisteína de CRSP-3 comprende un primer dominio rico en cisteína desde aproximadamente los aminoácidos 85-138 de SEQ ID NO: 8 y un segundo dominio rico en cisteína desde aproximadamente los aminoácidos 182-263 de SEQ ID NO: 8. La proteína CRSP-3 humana es una proteína segregada que contiene adicionalmente una secuencia señal en aproximadamente los aminoácidos 1-20 de SEQ ID NO: 8.

El ADNc de CRSP-4 humana, que tiene aproximadamente 702 nucleótidos de longitud, codifica para una proteína que tiene aproximadamente 179 residuos de aminoácidos de longitud. La proteína CRSP humana contiene una región rica en cisteína que comprende dos dominios ricos en cisteína. Puede encontrarse una región rica en cisteína de CRSP al menos, por ejemplo, desde aproximadamente los aminoácidos 1-176 de SEQ ID NO: 11. La región rica en cisteína de CRSP-4 comprende un primer dominio rico en cisteína desde aproximadamente los aminoácidos 1-47 de SEQ ID NO: 11 y un segundo dominio rico en cisteína desde aproximadamente los aminoácidos 96-176 de SEQ ID NO: 11.

Se describen diversos aspectos de la invención con más detalle en las siguientes subsecciones:

I. Moléculas de ácido nucleico aisladas

Un aspecto de la invención se refiere a moléculas de ácido nucleico aisladas que codifican para proteínas CRSP o partes biológicamente activas de las mismas, así como fragmentos de ácido nucleico suficientes para usarse como sondas de hibridación para identificar ácidos nucleicos que codifican para CRSP (por ejemplo, ARNm de CRSP) y fragmentos para usarse como cebadores de PCR para la amplificación o mutación de moléculas de ácido nucleico de CRSP. Se pretende que el término “molécula de ácido nucleico”, tal como se usa en el presente documento, incluya moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm) y análogos del ADN o ARN generado usando análogos de nucleótidos. La molécula de ácido nucleico puede ser de cadena sencilla o de cadena doble, pero preferiblemente es ADN de cadena doble.

Una molécula de ácido nucleico “aislada” es una que está separada de otras moléculas de ácido nucleico que están presentes en la fuente natural del ácido nucleico. Preferiblemente, un ácido nucleico “aislado” está libre de secuencias que flanquean de manera natural el ácido nucleico (es decir, secuencias ubicadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo a partir del que se deriva el ácido nucleico. Por ejemplo, en diversas realizaciones, la molécula de ácido nucleico de CRSP aislada puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencias de nucleótidos que flanquean de manera natural la molécula de

ácido nucleico en el ADN genómico de la célula de la que se deriva el ácido nucleico. Además, una molécula de ácido nucleico “aislada”, tal como una molécula de ADNc, puede estar sustancialmente libre de otro material celular, o medio de cultivo cuando se produce por técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente.

Una molécula de ácido nucleico de la presente invención, por ejemplo, una molécula de ácido nucleico que tiene la secuencia de nucleótidos tal como se define en las reivindicaciones adjuntas, puede aislarse usando técnicas de biología molecular habituales y la información de secuencia proporcionada en el presente documento. Usando todo o una parte de la secuencia de ácido nucleico de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, o SEQ ID NO: 10, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452 como una sonda de hibridación, pueden aislarse moléculas de ácido nucleico de CRSP usando técnicas de hibridación y clonación habituales (por ejemplo, tales como se describen en Sambrook, J., Fritsh, E. F., y Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2ª Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Además, una molécula de ácido nucleico que abarca toda o parte de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, o SEQ ID NO: 10, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452 puede aislarse mediante la reacción en cadena de la polimerasa (PCR) usando cebadores de oligonucleótidos diseñados basándose en la secuencia de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, o SEQ ID NO: 10, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452.

Un ácido nucleico de la invención puede amplificarse usando ADNc, ARNm o alternatively, ADN genómico, como un molde y cebadores de oligonucleótidos apropiados según técnicas de amplificación por PCR habituales. El ácido nucleico así amplificado puede clonarse en un vector apropiado y caracterizarse por análisis de secuencia de ADN. Además, los oligonucleótidos correspondientes a la secuencia de nucleótidos de CRSP pueden prepararse mediante técnicas sintéticas habituales, por ejemplo, usando un sintetizador de ADN automatizado.

En una realización preferida, una molécula de ácido nucleico aislada de la invención comprende una secuencia de nucleótidos tal como se define en las reivindicaciones.

La secuencia de SEQ ID NO: 1 corresponde al ADNc de CRSP-1 humana. Este ADNc comprende secuencias que codifican para la proteína CRSP-1 humana (es decir, “la región codificante”, desde los nucleótidos 38-1087), así como secuencias no traducidas en 5’ (nucleótidos 1 a 37) y secuencias no traducidas en 3’ (nucleótidos 1088 a 2479). Alternativamente, la molécula de ácido nucleico puede comprender sólo la región codificante de SEQ ID NO: 1, (por ejemplo, nucleótidos 38 a 1087, correspondientes a SEQ ID NO: 3). Se depositó un plásmido que contiene la secuencia de nucleótidos de longitud completa que codifica para CRSP-1 en la colección americana de cultivos tipo (ATCC), Manassas, Virginia, EE.UU. 20110-2209 el 11 de Junio de 1997 y se asignó el número de registro 98452. Se depositó un plásmido que contiene la secuencia de nucleótidos de longitud completa que codifica para CRSP-1 en la colección americana de cultivos tipo (ATCC), Manassas, Virginia, EE.UU. 20110-2209 el 16 de enero de 1998 y se asignó el número de registro 98634.

La secuencia de SEQ ID NO: 4 corresponde al ADNc de CRSP-2 humana. Este ADNc comprende secuencias que codifican para la proteína CRSP-2 humana (es decir, “la región codificante”, desde los nucleótidos 126-796), así como secuencias no traducidas en 5’ (nucleótidos 1 a 125) y secuencias no traducidas en 3’ (nucleótidos 797 a 848). Alternativamente, la molécula de ácido nucleico puede comprender sólo la región codificante de SEQ ID NO: 4 (por ejemplo, nucleótidos 126 a 796, correspondientes a SEQ ID NO: 6).

La SEQ ID NO: 7 corresponde al ADNc de CRSP-3 humana. Este ADNc comprende secuencias que codifican para la proteína CRSP-3 humana (es decir, “la región codificante”, desde los nucleótidos 93-890), así como las secuencias no traducidas en 5’ (nucleótidos 1 a 92) y secuencias no traducidas en 3’ (nucleótidos 891 a 1529). Alternativamente, la molécula de ácido nucleico puede comprender sólo la región codificante de SEQ ID NO: 7 (por ejemplo, nucleótidos 93 a 890 correspondientes a SEQ ID NO: 9). Se depositó un plásmido que contiene la secuencia de nucleótidos de longitud completa que codifica para CRSP-3 en la colección americana de cultivos tipo (ATCC), Manassas, Virginia, EE.UU. 20110-2209 el 16 de enero de 1998 y se asignó el número de registro 98633.

La secuencia de SEQ ID NO: 10 corresponde al ADNc de CRSP-4 humana. Este ADNc comprende secuencias que codifican para la proteína CRSP-4 humana (es decir, “la región codificante”, desde nucleótidos 1-537), así como las secuencias no traducidas en 3’ (nucleótidos 538 a 702). Alternativamente, la molécula de ácido nucleico puede comprender sólo la región codificante de SEQ ID NO: 10 (por ejemplo, nucleótidos 1 a 537, correspondientes a SEQ ID NO: 12).

En otra realización preferida, una molécula de ácido nucleico aislada de la invención comprende una molécula de ácido nucleico que es un complemento de la secuencia de nucleótidos reivindicada. Una molécula de ácido nucleico que es complementaria a la secuencia nucleica reivindicada es una que es suficientemente complementaria a la secuencia de nucleótidos reivindicada de tal manera que puede hibridarse con la secuencia de nucleótidos reivindicada formando así un dúplex estable.

Moléculas de ácido nucleico asiladas útiles comprenden una secuencia de nucleótidos que es homóloga en al menos aproximadamente un 30-35%, de manera preferible aproximadamente un 40-45%; de manera más preferible aproximadamente un 50-55%, de manera preferible incluso más aproximadamente un 60-65%, y de manera preferible incluso más al menos aproximadamente un 70-75%, un 80-85%, un 90-95% o más con la secuencia de nucleótidos reivindicada o una parte de cualquiera de estas secuencias de nucleótidos.

Además, una molécula de ácido nucleico útil puede comprender sólo una parte de la secuencia de ácido nucleico de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452, por ejemplo un fragmento que puede usarse como una sonda o cebador o un fragmento que codifica para una parte biológicamente activa de una proteína CRSP. La secuencia de nucleótidos determinada a partir de la clonación de genes de CRSP humana permite la generación de sondas y cebadores diseñados para usarse en la identificación y/o clonación de homólogos de CRSP en otros tipos de células, por ejemplo a partir de otros tejidos, así como homólogos de CRSP de otros mamíferos. La sonda/cebador comprende normalmente un oligonucleótido sustancialmente purificado. El oligonucleótido comprende normalmente una región de secuencia de nucleótidos que hibrida en condiciones rigurosas con al menos aproximadamente 12, de manera preferible de aproximadamente 25, de manera más preferible de aproximadamente 40, 50 ó 75 nucleótidos consecutivos de una secuencia sentido de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452, de una secuencia antisentido de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452, o de un mutante que se produce de manera natural de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98452. En una realización a modo de ejemplo, una molécula de ácido nucleico útil comprende una secuencia de nucleótidos que hibrida en condiciones de hibridación rigurosas con una molécula de ácido nucleico que comprende los nucleótidos 470-2479 de SEQ ID NO: 1 o a una molécula de ácido nucleico que comprende los nucleótidos 1-475 de SEQ ID NO: 4.

Sondas basadas en la secuencia de nucleótidos de CRSP humana pueden usarse para detectar transcritos o secuencias genómicas que codifican para proteínas idénticas u homólogas. Por ejemplo, cebadores basados en el ácido nucleico representado en SEQ ID NO: 1 ó 3 pueden usarse en reacciones de PCR para clonar homólogos de CRSP (por ejemplo, homólogos de CRSP-1). En una realización preferida de la invención, se clonan homólogos de CRSP mediante amplificación por PCR (por ejemplo, RT-PCR) usando cebadores que hibridan con una parte de la secuencia de nucleótidos que codifica para el dominio rico en cisteína de CRSP. De igual manera, pueden usarse sondas basadas en las secuencias de CRSP objeto para detectar transcritos o secuencias genómicas que codifican para proteínas idénticas u homólogas. En realizaciones preferidas, la sonda comprende además un grupo marcador unido a la misma, por ejemplo, el grupo marcador puede ser un radioisótopo, un compuesto fluorescente, una enzima, o un cofactor enzimático. Tales sondas pueden usarse como parte de un kit de prueba diagnóstica para identificar células o tejido que expresan erróneamente una proteína CRSP, tal como midiendo un nivel de un ácido nucleico que codifica para CRSP en una muestra de células de un sujeto por ejemplo, detectando niveles de ARNm de CRSP o determinando si un gen de CRSP genómico ha mutado o se ha delecionado.

Un fragmento de ácido nucleico que codifica para una “parte biológicamente activa de una proteína CRSP” puede prepararse aislando una parte de una secuencia definida en las reivindicaciones adjuntas, que codifica para un polipéptido que tiene una actividad biológica de CRSP (las actividades biológicas de las proteínas CRSP se han descrito previamente), expresando la parte codificada de la proteína CRSP (por ejemplo, mediante expresión recombinante *in vitro*) y evaluando la actividad de la parte codificada de la proteína CRSP.

La invención engloba además moléculas de ácido nucleico que difieren de la secuencia de nucleótidos mostrada en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452, debido a la degeneración del código genético y por tanto codifica para las mismas proteínas CRSP que las codificadas por la secuencia de nucleótidos definida en las reivindicaciones adjuntas. En otra realización, una molécula de ácido nucleico aislada de la invención

tiene una secuencia de nucleótidos que codifica para una proteína que tiene una secuencia de aminoácidos definida en las reivindicaciones adjuntas.

Además de la secuencia de nucleótidos de CRSP humana mostradas en las reivindicaciones, los expertos en la técnica apreciarán que los polimorfismos de la secuencia de ADN que dan lugar a cambios en la secuencia de aminoácidos de las proteínas CRSP pueden existir dentro de una población (por ejemplo, la población humana). Pueden existir tales polimorfismos genéticos en los genes de CRSP entre individuos dentro de una población debido a variación alélica natural. Tal como se usa en el presente documento, los términos “gen” y “gen recombinante” se refieren a moléculas de ácido nucleico que comprenden un marco de lectura abierto que codifica para una proteína CRSP, preferiblemente una proteína CRSP de mamífero. Tales variaciones alélicas naturales pueden dar normalmente como resultado una varianza del 1-5% en la secuencia de nucleótidos de un gen de CRSP. Se pretende que todas y cada una de tales variaciones nucleotídicas y polimorfismos de aminoácidos resultantes en los genes de CRSP que son el resultado de variación alélica natural y que no alteran la actividad funcional de una proteína CRSP estén dentro del alcance de la invención.

Además, se pretende que moléculas de ácido nucleico que codifican para proteínas CRSP de otras especies, y que por tanto tienen una secuencia de nucleótidos que difiere de la secuencia de ácido nucleico humana, estén dentro del alcance de la invención. Por ejemplo, se ha identificado un ADNc de CRSP-1 murina basándose en la secuencia de nucleótidos de CRSP-1 humana. La secuencia de nucleótidos de CRSP-1 murina (SEQ ID NO: 16) codifica para una proteína CRSP-1 que tiene 349 aminoácidos. Las secuencias de aminoácidos y nucleótidos de CRSP-1 murina se muestran en la figura 8. La región codificante de CRSP-1 murina está representada por SEQ ID NO: 18.

Moléculas de ácido nucleico correspondientes a variantes alélicas naturales y homólogas de los ADNc de CRSP de la invención pueden aislarse basándose en su homología con los ácidos nucleicos de CRSP humana descritos en el presente documento usando el ADNc humano, o una parte del mismo, como una sonda de hibridación según las técnicas de hibridación habituales bajo condiciones de hibridación rigurosas. Ejemplos de tejidos y/o bibliotecas adecuados para el aislamiento de ácidos nucleicos objeto incluyen timo, ganglios linfáticos y tejido inflamatorio. ADNc que codifica para una proteína CRSP (por ejemplo, una proteína CRSP-1) puede obtenerse aislando ARNm total de una célula, por ejemplo, una célula de vertebrado, una célula de mamífero, o una célula humana, incluyendo células embrionarias. Después pueden prepararse ADNc de cadena doble a partir de ARNm total y posteriormente insertarse en un vector bacteriófago o plásmido adecuado usando una cualquiera de una variedad de técnicas conocidas. El gen que codifica para una proteína CRSP-1 también puede clonarse usando técnicas de reacción en cadena de polimerasa establecidas según la información de la secuencia de nucleótidos proporcionada por la invención. El ácido nucleico de la invención puede ser ADN o ARN o análogos de los mismos.

Por consiguiente, una molécula de ácido nucleico aislada útil tiene al menos 15 nucleótidos de longitud y hibrida en condiciones rigurosas con la molécula de ácido nucleico que comprende la secuencia de nucleótidos de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452. En otra realización, el ácido nucleico tiene al menos 30, 50, 100, 250 ó 500 nucleótidos de longitud. Tal como se usa en el presente documento, se pretende que el término “hibrida en condiciones rigurosas” describa las condiciones de hibridación y lavado en las cuales la secuencia de nucleótidos homólogas en al menos un 60% entre sí permanecen normalmente hibridadas entre sí. Preferiblemente, las condiciones son tales que secuencias homólogas en al menos aproximadamente un 70%, de manera más preferible en al menos aproximadamente un 80%, de manera incluso más preferible en al menos aproximadamente un 85% ó 90% entre sí permanecen hibridadas entre sí. Los expertos en la técnica conocen tales condiciones rigurosas y pueden encontrarse en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Un ejemplo no limitante preferido de condiciones de hibridación rigurosas es hibridación en 6X cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45°C, seguido por uno o más lavados en 0,2 X SSC, SDS al 0,1% a 50-65°C. Preferiblemente, una molécula de ácido nucleico aislada de la invención que hibrida en condiciones rigurosas con la secuencia de SEQ ID NO: 1 corresponde a una molécula de ácido nucleico que se produce de manera natural. Tal como se usa en el presente documento, una molécula de ácido nucleico “que se produce de manera natural” se refiere a una molécula de ARN o ADN que tiene una secuencia de nucleótidos que se produce en la naturaleza (por ejemplo, codifica para una proteína natural).

Además de variantes alélicas que se producen de manera natural de las secuencias de CRSP que pueden existir en la población, el experto en la técnica apreciará adicionalmente que pueden introducirse cambios mediante mutación en las secuencias de nucleótidos de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452, conduciendo así a cambios en la secuencia de aminoácidos de las proteínas CRSP codificadas, sin alterar la capacidad funcional de las proteínas CRSP. Por ejemplo, pueden hacerse sustituciones de nucleótidos que conducen a sustituciones de aminoácidos (particularmente sustituciones de aminoácidos conservativas) en residuos de aminoácidos “no esenciales” en la secuencia de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos

del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452. Un residuo de aminoácido “no esencial” es un residuo que puede alterarse de la secuencia de tipo natural de CRSP (por ejemplo, la secuencia de SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11) sin alterar la actividad biológica, mientras que un residuo de aminoácido “esencial” se requiere para la actividad biológica. Por ejemplo, residuos de aminoácidos

5 que se conservan entre las proteínas CRSP de la presente invención (por ejemplo, residuos de cisteína dentro de dominios ricos en cisteína), se espera que sean particularmente poco susceptibles de alteración. Además, los residuos de aminoácidos que se conservan entre proteína CRSP y otras proteínas que tienen dominio rico en cisteína es poco probable que sean susceptibles de alteración.

10 Por consiguiente, la invención se refiere a moléculas de ácido nucleico que codifican para proteínas CRSP que contienen cambios en residuos de aminoácidos que no son esenciales para la actividad. Tales proteínas CRSP difieren en la secuencia de aminoácidos de SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11 aunque mantienen la actividad biológica. Una molécula de ácido nucleico aislada útil comprende una secuencia de nucleótidos que codifica para una proteína, en la que la proteína comprende una secuencia de aminoácidos homóloga en

15 al menos aproximadamente un 60% con la secuencia de aminoácidos de SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11. Una proteína útil codificada por la molécula de ácido nucleico es homóloga en al menos aproximadamente un 65-70% con la SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11, de manera más preferible homóloga en al menos aproximadamente un 75-80% con SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11, de manera incluso más preferible homóloga en al menos aproximadamente un 85-

20 90% con SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11, y de la manera más preferible homóloga en al menos aproximadamente un 95% con SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11.

Una molécula de ácido nucleico aislada que codifica para una proteína CRSP homóloga a la proteína de SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11 puede prepararse introduciendo una o más sustituciones, adiciones o delecciones de nucleótidos en la secuencia de nucleótidos de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452, de manera que se introducen una o más sustituciones, adiciones o delecciones de

30 aminoácidos en la proteína codificada. Pueden introducirse mutaciones en SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452, mediante técnicas habituales, tales como mutagénesis dirigida al sitio y mutagénesis mediada por PCR. Preferiblemente, las sustituciones de aminoácidos conservativas se hacen en uno o más residuos de aminoácidos no esenciales previstos. Una “sustitución de aminoácidos conservativa” es una en la que el residuo de aminoácido se sustituye por un residuo de aminoácido que tiene una cadena lateral similar. En la técnica se han

40 definido familias de residuos de aminoácidos que tienen cadenas laterales similares. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales no cargadas (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Por tanto, un residuo de aminoácido no esencial previsto en una proteína CRSP (por ejemplo, uno no ubicado en un dominio rico en cisteína) se

45 sustituye preferiblemente por otro residuo de aminoácido de la misma familia de cadena lateral. Alternativamente, en otra realización, pueden introducirse mutaciones aleatoriamente a lo largo de todo o parte de una secuencia que codifica para de CRSP, tal como mediante mutagénesis por saturación, y los mutantes resultantes pueden examinarse para seleccionar actividad biológica de CRSP para identificar mutantes que conservan actividad. Tras la mutagénesis de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633, o la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98452, puede expresarse recombinantemente la proteína codificada y puede determinarse la actividad de la proteína.

55 Una proteína CRSP mutante puede someterse a ensayo para determinar calcio intracelular, un aumento de fosfatidilinositol u otra molécula, y puede dar como resultado, por ejemplo, la fosforilación de proteínas específicas, una modulación de transcripción génica y cualquiera de las otras actividades biológicas expuestas en el presente documento.

60 Una proteína CRSP mutante también puede someterse a ensayo para determinar la capacidad de (1) modular la transducción de señal celular, o bien *in vitro* o *in vivo*; (2) regular la comunicación entre células; (3) regular la expresión de genes cuya expresión está modulada por la unión de CRSP (por ejemplo, CRSP-1) a un receptor; (4) regular la transcripción génica en una célula implicada en el desarrollo o diferenciación, o bien *in vitro* o bien *in vivo*;

65 (5) regular la proliferación celular, o bien *in vitro* o bien *in vivo*; (6) formar y/o mantener disposiciones espaciales ordenadas de tejidos diferenciados en vertebrados; (7) modular la muerte celular (por ejemplo supervivencia celular); (8) regular la migración celular; y/o (9) modular la función del sistema inmunitario.

Además de las moléculas de ácido nucleico que codifican para las proteínas CRSP descritas anteriormente, la invención se refiere a moléculas de ácido nucleico aisladas que son antisentido con respecto a las mismas. Un ácido nucleico “antisentido” comprende una secuencia de nucleótidos que es complementaria a un ácido nucleico “sentido” que codifica para una proteína, por ejemplo, complementario a la cadena codificante de una molécula de ADNc de cadena doble o complementario a una secuencia de ARNm. Por consiguiente, un ácido nucleico antisentido puede unirse mediante enlace de hidrógeno a un ácido nucleico sentido. El ácido nucleico antisentido puede ser complementario a una cadena que codifica para CRSP entera, o sólo a una parte de la misma. En una realización, una molécula de ácido nucleico antisentido es antisentido con respecto a una “región codificante” de la cadena codificante de una secuencia de nucleótidos que codifica para CRSP. El término “región codificante” se refiere a la región de la secuencia de nucleótidos que comprende codones que se traducen en residuos de aminoácidos (por ejemplo, la región codificante de CRSP-1 humana corresponde a SEQ ID NO: 3, la región codificante de CRSP-2 humana corresponde a SEQ ID NO: 6, la región codificante de CRSP-3 humana corresponde a SEQ ID NO: 9, la región codificante de CRSP-4 humana corresponde a SEQ ID NO: 12, y la región codificante de N terminal de tipo CRSP humana corresponde a SEQ ID NO: 15). En otra realización, la molécula de ácido nucleico antisentido es antisentido con respecto a una “región no codificante” de la cadena codificante de una secuencia de nucleótidos que codifica para CRSP. El término “región no codificante” se refiere a secuencias en 5' y 3' que flanquean la región codificante que no se traducen en aminoácidos (es decir, también denominadas regiones no traducidas en 5' y 3').

Dadas las secuencias de cadena codificante codifican para CRSP descritas en el presente documento (por ejemplo, SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 9, o SEQ ID NO: 12), pueden diseñarse ácidos nucleicos antisentido según las reglas de apareamiento de bases de Watson y Crick. La molécula de ácido nucleico antisentido puede ser complementaria a la región codificante entera de ARNm de CRSP, pero más preferiblemente es un oligonucleótido que es antisentido a sólo una parte de la región codificante o no codificante de ARNm de CRSP. Por ejemplo, el oligonucleótido antisentido puede ser complementario a la región que rodea el sitio de inicio de traducción de ARNm de CRSP. Un oligonucleótido antisentido puede tener, por ejemplo, aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50 nucleótidos de longitud. Un ácido nucleico antisentido puede construirse usando síntesis química y reacciones de ligamiento enzimático usando procedimientos conocidos en la técnica. Por ejemplo, un ácido nucleico antisentido (por ejemplo, un oligonucleótido antisentido) puede sintetizarse químicamente usando nucleótidos que se producen de manera natural o nucleótidos diversamente modificados diseñados para aumentar la estabilidad biológica de las moléculas o para aumentar la estabilidad física del dúplex formado entre los ácidos nucleico antisentido y sentido, por ejemplo, pueden usarse derivados de fosforotioato y nucleótidos sustituidos con acridina. Ejemplos de nucleótidos modificados que pueden usarse para generar el ácido nucleico antisentido incluyen 5-fluorouracilo, 5-bromouracilo, 5-clorouracilo, 5-yodouracilo, hipoxantina, xantina, 4-acetilcitosina, 5-(carboxihidroximetil)uracilo, 5-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluracilo, dihidrouracilo, beta-D-galactosilqueosina, inosina, N6-isopenteniladenina, 1-metilguanina, 1-metilinosina, 2,2-dimetilguanina, 2-metiladenina, 2-metilguanina, 3-metilcitosina, 5-metilcitosina, N6-adenina, 7-metilguanina, 5-metilaminometiluracilo, 5-metoxiaminometil-2-tiouracilo, beta-D-manosilqueosina, 5'-metoxycarboximetiluracilo, 5-metoxiuracilo, 2-metiltio-N6-isopenteniladenina, ácido uracil-5-oxiacético (v), wybutoxosina, pseudouracilo, queosina, 2-tiocitosina, 5-metil-2-tiouracilo, 2-tiouracilo, 4-tiouracilo, 5-metiluracilo, éster metílico del ácido uracil-5-oxiacético, ácido uracil-5-oxiacético (v), 5-metil-2-tiouracilo, 3-(3-amino-3-N-2-carboxipropil)uracilo, (acp3)w, y 2,6-diaminopurina. Alternativamente, puede producirse biológicamente el ácido nucleico antisentido usando un vector de expresión en el que se ha subclonado un ácido nucleico en una orientación antisentido (es decir, el ARN transcrito a partir del ácido nucleico insertado será de orientación antisentido con respecto a un ácido nucleico diana de interés, descrito en detalle a en la siguiente subsección).

Las moléculas de ácido nucleico antisentido se administran normalmente a un sujeto o se generan *in situ* de manera que hibridan con, o se unen a, ARNm celular y/o ADN genómico que codifica para una proteína CRSP para así inhibir la expresión de la proteína, por ejemplo, inhibiendo la transcripción y/o traducción. La hibridación puede ser mediante complementariedad de nucleótido convencional para formar un dúplex estable, o, por ejemplo, en el caso de una molécula de ácido nucleico antisentido que se une a dúplex de ADN, mediante interacciones específicas en la ranura principal de la doble hélice. Un ejemplo de vía de administración de moléculas de ácido nucleico antisentido incluye inyección directa en un sitio tisular. Alternativamente, pueden modificarse moléculas de ácido nucleico antisentido para dirigirse a células seleccionadas y luego administrarse sistemáticamente. Por ejemplo, para administración sistémica, pueden modificarse moléculas antisentido de manera que se unen específicamente a receptores o antígenos expresados sobre una superficie celular seleccionada, por ejemplo, uniendo las moléculas de ácido nucleico antisentido a péptidos o anticuerpos que se unen a receptores o antígenos de superficie celular. Las moléculas de ácido nucleico antisentido también pueden administrarse a células usando los vectores descritos en el presente documento. Para lograr concentraciones intracelulares suficientes de moléculas antisentido, se prefieren constructos de vectores en los que se coloca la molécula de ácido nucleico antisentido bajo el control de un promotor pol II o pol III fuerte.

La molécula de ácido nucleico antisentido puede ser una molécula de ácido nucleico α -anomérica. Una molécula de ácido nucleico α -anomérica forma híbridos de cadena doble específicos con ARN complementario en los que, al contrario que las unidades β habituales, las cadenas corren paralelas la una a la otra (Gaultier *et al.* (1987) Nucleic Acids. Res. 15:6625-6641). La molécula de ácido nucleico antisentido también puede comprender un 2'-o-metilribonucleótido (Inoue *et al.* (1987) Nucleic Acids Res. 15:6131-6148) o un análogo ARN-ADN quimérico (Inoue *et al.* (1987) FEBS Lett. 215:327-330).

En aún otra realización, un ácido nucleico antisentido de la invención es una ribozima. Las ribozimas son moléculas de ARN catalíticas con actividad ribonucleasa que pueden escindir un ácido nucleico de cadena sencilla, tal como

un ARNm, con el que tienen una región complementaria. Por tanto, pueden usarse ribozimas (por ejemplo, ribozimas cabeza de martillo (descritas en Haselhoff y Gerlach (1988) *Nature* 334:585-591)) para escindir catalíticamente transcritos de ARNm de CRSP para así inhibir la traducción de ARNm de CRSP. Puede diseñarse una ribozima que tiene especificidad por un ácido nucleico que codifica para CRSP basándose en la secuencia de nucleótidos de ADNc de CRSP descrita en el presente documento (es decir, SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98634, o la secuencia de nucleótidos del inserto de ADN del plásmido depositado en la ATCC con el número de registro 98633). Por ejemplo, puede construirse un derivado de un ARN de *Tetrahymena* L-19 IVS en el que la secuencia de nucleótidos del sitio activo es complementaria a la secuencia de nucleótidos que debe escindirse en un ARNm que codifica para CRSP. Véase, por ejemplo, Cech *et al.* patente estadounidense número 4.987.071; y Cech *et al.* patente estadounidense número 5.116.742. Alternativamente, puede usarse ARNm de CRSP para seleccionar un ARN catalítico que tiene una actividad ribonucleasa específica de una combinación de moléculas de ARN. Véase, por ejemplo, Bartel, D. y Szostak, J. W. (1993) *Science* 261:1411-1418.

Alternativamente, puede inhibirse la expresión del gen de CRSP dirigiendo secuencias de nucleótidos complementarias a la región reguladora de la CRSP (por ejemplo, el promotor y/o potenciador de CRSP) para formar estructuras de triple hélice que evitan la transcripción del gen de CRSP en células diana. Véase de manera general, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann N. Y. Acad. Sci.* 660:27-36; y Maher, L. J. (1992) *Bioassays* 14(12):807-15.

Las moléculas de ácido nucleico de CRSP de la presente invención pueden modificarse en el resto básico, resto de azúcar o esqueleto de fosfato para mejorar, por ejemplo, la estabilidad, hibridación o solubilidad de la molécula. Por ejemplo, el esqueleto fosfato desoxirribosa de las moléculas de ácido nucleico puede modificarse para generar ácido nucleicos peptídicos (véase Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). Tal como se usan en el presente documento, los términos “ácidos nucleicos peptídicos” o “PNA” se refieren a compuestos miméticos de ácido nucleico, por ejemplo, compuestos miméticos de ADN, en los que el esqueleto de fosfato desoxirribosa se sustituye por un esqueleto pseudopeptídico y sólo se conservan las cuatro nucleobases naturales. Se ha demostrado que el esqueleto neutral de los PNA permite hibridación específica con ADN y ARN en condiciones de baja concentración iónica. La síntesis de oligómeros de PNA puede realizarse usando protocolos de síntesis peptídica en fase sólida habituales tal como se describe en Hyrup B. *et al.* (1996) citado anteriormente; Perry-O’Keefe *et al.* *PNAS* 93: 14670-675.

Los PNA de moléculas de ácido nucleico de CRSP pueden usarse en aplicaciones terapéuticas y de diagnóstico. Por ejemplo, pueden usarse PNA como agentes antisentido o antígenos para la modulación específica de secuencia de la expresión génica mediante, por ejemplo, inducción de la parada de transcripción o traducción o inhibición de la replicación. Los PNA de moléculas de ácido nucleico de CRSP también pueden usarse en el análisis de mutaciones de un único par de bases en un gen, (por ejemplo, mediante adición de estructura de unión por PCR dirigida por PNA); como “enzimas de restricción artificiales” cuando se usan en combinación con otras enzimas, (por ejemplo, nucleasas S 1 (Hyrup B. (1996) citado anteriormente)); o como sondas o cebadores para la hibridación o secuenciación de ADN (Hyrup B. *et al.* (1996) citado anteriormente; Perry-O’Keefe citado anteriormente).

Los PNA de CRSP pueden modificarse, (por ejemplo, para potenciar su estabilidad o captación celular), uniendo grupos lipófilos u otros grupos cooperadores a PNA, mediante la formación de quimeras PNA-ADN, o mediante el uso de liposomas u otras técnicas de administración de fármacos conocidas en la técnica. Por ejemplo, pueden generarse quimeras PNA-ADN de moléculas de ácido nucleico de CRSP que pueden combinar las propiedades ventajosas de PNA y ADN. Tales quimeras permiten que enzimas de reconocimiento de ADN, (por ejemplo, ARNasa H y ADN polimerasas) interaccionen con la parte de ADN mientras la parte del PNA proporciona una alta afinidad de unión y especificidad. Las quimeras PNA-ADN pueden unirse usando ligadores de longitudes apropiadas seleccionados en cuando al apilamiento de bases, número de enlaces entre las nucleobases y orientación (Hyrup B. (1996) citado anteriormente). La síntesis de quimeras PNA-ADN puede realizarse tal como se describe en Hyrup B. (1996) citado anteriormente y Finn P. J. *et al.* (1996) *Nucleic Acids. Res.* 24 (17): 3357-63. Por ejemplo, puede sintetizarse una cadena de ADN sobre un soporte sólido usando la química de acoplamiento de fosforamidita habitual y análogos de nucleósidos modificados, por ejemplo, puede usarse 5’-(4-metoxitritil)amino-5’-desoxi-timidina-fosforamidita, como un entre el PNA y el extremo 5’ del ADN (Mag, M. *et al.* (1989) *Nucleic Acids Res.* 17: 5973-88). Después se acoplan los monómeros de PNA de una manera gradual para producir una molécula quimérica con un segmento de PNA en 5’ y un segmento de ADN en 3’ (Finn P.J. *et al.* (1996) citado anteriormente). Alternativamente, pueden sintetizarse moléculas quiméricas con un segmento de ADN en 5’ y un segmento de PNA en 3’ (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

El oligonucleótido puede incluir otros grupos unidos tales como péptidos (por ejemplo, para seleccionar como diana receptores de células huésped *in vivo*), o agentes que facilitan el transporte a través de la membrana celular (véase, por ejemplo, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; publicación PCT número WO88/09810, publicada el 15 de diciembre de 1988) o la barrera hematoencefálica (véase, por ejemplo, la publicación PCT número WO89/10134, publicada el 25 de abril de 1988). Además pueden modificarse los oligonucleótidos con agentes de escisión activada por la hibridación (véase, por ejemplo, Krol *et al.* (1988) *BioTechniques* 6:958-976) o agentes intercalables. (Véase, por ejemplo, Zon (1988) *Pharm. Res.* 5:539-549). Con este fin, el oligonucleótido puede conjugarse con otra molécula, (por ejemplo, un péptido, un agente de reticulación activada por la hibridación, un agente de transporte o agente de escisión activada por la hibridación).

II. Proteínas CRSP aisladas y anticuerpos anti-CRSP

Un aspecto de la invención se refiere a proteínas CRSP aisladas, y partes biológicamente activas de las mismas, así como fragmentos de polipéptidos adecuados para su uso como inmunógenos para crear anticuerpos anti-CRSP. En una realización, las proteínas CRSP nativas pueden aislarse a partir de fuentes celulares o tisulares mediante un esquema de purificación apropiado usando técnicas de purificación de proteínas habituales. En otra realización, las proteínas CRSP se producen mediante técnicas de ADN recombinante. Como alternativa a la expresión recombinante, puede sintetizarse químicamente un polipéptido o proteína CRSP usando técnicas de síntesis de péptidos habituales.

Una proteína “aislada” o “purificada” o una parte biológicamente activa de la misma está sustancialmente libre de material celular u otras proteínas contaminantes de la fuente celular o tisular de la que se deriva la proteína CRSP, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetizan químicamente. La expresión “sustancialmente libre de material celular” incluye preparaciones de proteína CRSP en las que la proteína está separada de componentes celulares de las células de las que se aísla o produce recombinantemente. En una realización, la expresión “sustancialmente libre de material celular” incluye preparaciones de proteína CRSP que tienen menos de aproximadamente un 30% (en peso seco) de proteína distinta de CRSP (también denominada en el presente documento como “proteína contaminante”), de manera más preferible menos de aproximadamente un 20% de proteína distinta de CRSP, de manera aún más preferible menos de aproximadamente un 10% de proteína distinta de CRSP, y de la manera más preferible menos de aproximadamente un 5% de proteína distinta de CRSP. Cuando se produce de manera recombinante la proteína CRSP o parte biológicamente activa de la misma, también es de manera preferible sustancialmente libre de medio de cultivo, es decir, el medio de cultivo representa menos de aproximadamente el 20%, de manera más preferible menos de aproximadamente el 10%, y de la manera más preferible menos de aproximadamente el 5% del volumen de la preparación de proteína.

La expresión “sustancialmente libre de precursores químicos u otros productos químicos” incluye preparaciones de proteína CRSP en las que la proteína se separa de los precursores químicos u otros productos químicos que están implicados en la síntesis de la proteína. En una realización, la expresión “sustancialmente libre de precursores químicos u otros productos químicos” incluye preparaciones de proteína CRSP que tienen menos de aproximadamente el 30% (en peso seco) de precursores químicos o productos químicos distintos de CRSP, de manera más preferible menos de aproximadamente el 20% de precursores químicos o productos químicos distintos de CRSP, de manera aún más preferible menos de aproximadamente el 10% de precursores químicos o productos químicos distintos de CRSP, y de la manera más preferible menos de aproximadamente el 5% de precursores químicos o productos químicos distintos de CRSP.

Partes biológicamente activas de una proteína CRSP incluyen péptidos que comprenden secuencias de aminoácidos suficientemente homólogas a, o derivadas de, la secuencia de aminoácidos de la proteína CRSP, por ejemplo, la secuencia de aminoácidos mostrada en SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 8, o SEQ ID NO: 11, que incluye menos aminoácidos que las proteínas CRSP de longitud completa, y muestran al menos una actividad de una proteína CRSP. Normalmente, las partes biológicamente activas comprenden un dominio o motivo con al menos una actividad de la proteína CRSP. Una parte biológicamente activa de una proteína CRSP puede ser un polipéptido que tiene, por ejemplo, 10, 25, 50, 100 o más aminoácidos de longitud.

En una realización, una parte biológicamente activa de una proteína CRSP comprende al menos una región rica en cisteína. En otra realización, una parte biológicamente activa de una proteína CRSP comprende al menos una región rica en cisteína, en la que la región rica en cisteína incluye al menos un dominio rico en cisteína. Todavía en otra realización, una parte biológicamente activa de una proteína CRSP comprende al menos una secuencia señal.

En una realización alternativa, una parte biológicamente activa de una proteína CRSP comprende una secuencia de aminoácidos de CRSP que carece de una secuencia señal.

Debe entenderse que una parte biológicamente activa de una proteína CRSP preferida de la presente invención puede contener al menos uno de los dominios estructurales anteriormente identificados. Una parte biológicamente activa más preferida de una proteína CRSP puede contener al menos dos de los dominios estructurales anteriormente identificados. Una parte biológicamente activa incluso más preferida de una proteína CRSP puede contener al menos tres de los dominios estructurales anteriormente identificados. Una parte biológicamente activa particularmente preferida de una proteína CRSP de la presente invención puede contener al menos cuatro de los dominios estructurales anteriormente identificados.

Además, otras partes biológicamente activas, en las que se delecionan otras regiones de la proteína, pueden prepararse mediante técnicas recombinantes y evaluarse para determinar una o más actividades funcionales de una proteína CRSP nativa.

La proteína CRSP de la invención tiene una secuencia de aminoácidos tal como se muestra en las reivindicaciones adjuntas.

Para determinar el porcentaje de homología de dos secuencias de aminoácidos o de dos ácidos nucleicos, se alinean las secuencias con fines de comparación óptima (por ejemplo, pueden introducirse huecos en la secuencia de una primera secuencia de aminoácidos o de ácido nucleico para la alineación óptima con una segunda secuencia de

aminoácidos o de ácido nucleico). Entonces se comparan los residuos de aminoácidos o nucleótidos en las posiciones de aminoácidos o posiciones de nucleótidos correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que en la posición correspondiente en la segunda secuencia, entonces las moléculas son homólogas en esa posición (es decir, tal como se usa en el presente documento “homología” de aminoácido o ácido nucleico es equivalente a “identidad” de aminoácido o ácido nucleico). El porcentaje de homología entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % homología = nº de posiciones idénticas / nº total de posiciones x 100). La determinación del porcentaje de homología entre dos secuencias puede lograrse usando un algoritmo matemático. Un ejemplo no limitante preferido de un algoritmo matemático utilizado para la comparación de dos secuencias es el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modificado como en Karlin y Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77. Un algoritmo de este tipo se incorpora en los programas NBLAST y XBLAST de Altschul, *et al.* (1990) J. Mol. Biol. 215:403-10. Pueden realizarse búsquedas BLAST de nucleótidos con el programa NBLAST, puntuación = 100, longitud de palabra = 12 para obtener secuencias de nucleótidos homólogas a moléculas de ácido nucleico de CRSP de la invención. Pueden realizarse búsquedas BLAST de proteínas con el programa XBLAST, puntuación = 50, longitud de palabra = 3 para obtener secuencias de aminoácidos homólogas a moléculas de proteína CRSP de la invención. Para obtener alineaciones con huecos con fines comparativos puede usarse Gapped BLAST tal como se describe en Altschul *et al.*, (1997) Nucleic Acids Research 25(17):3389-3402. Cuando se utilizan los programas BLAST y Gapped BLAST, pueden usarse los parámetros por defecto de los respectivos programas (por ejemplo, XBLAST y NBLAST). Véase <http://www.ncbi.nlm.nih.gov>. Otro ejemplo no limitante preferido de un algoritmo matemático utilizado para la comparación de secuencias es el algoritmo de Myers y Miller, CABIOS (1989). Un algoritmo de este tipo se incorpora en el programa ALIGN (versión 2.0) que es parte del paquete de software de alineación de secuencias de GCG. Cuando se utiliza el programa ALIGN para comparar secuencias de aminoácidos, pueden usarse una tabla de residuos de peso PAM120, una penalización de longitud de hueco de 12, y una penalización de hueco de 4.

La invención puede usarse para proporcionar proteínas quiméricas o de fusión de CRSP. Tal como se usa en el presente documento, una “proteína quimérica” o “proteína de fusión” de CRSP comprende un polipéptido de CRSP operativamente unido a un polipéptido distinto de CRSP. Un “polipéptido de CRSP” se refiere a un polipéptido que tiene una secuencia de aminoácidos correspondiente a CRSP, mientras que un “polipéptido distinto de CRSP” se refiere a un polipéptido que tiene una secuencia aminoácidos correspondiente a una proteína que no es sustancialmente homóloga a la proteína CRSP, por ejemplo, una proteína que es diferente de la proteína CRSP y que se deriva de un organismo idéntico o diferente. Dentro de una proteína de fusión de CRSP del polipéptido de CRSP puede corresponder a todo o una parte de una proteína CRSP. Una proteína de fusión de CRSP comprende al menos una parte biológicamente activa de una proteína CRSP. Otra proteína de fusión de CRSP comprende al menos dos partes biológicamente activas de una proteína CRSP. Otra proteína de fusión de CRSP comprende al menos tres partes biológicamente activas de una proteína CRSP. Dentro de la proteína de fusión, se pretende que el término “operativamente unido” indique que el polipéptido de CRSP y el polipéptido distinto de CRSP están fusionados en marco entre sí. El polipéptido distinto de CRSP puede fusionarse al extremo N-terminal o C-terminal del polipéptido de CRSP.

Por ejemplo, una proteína de fusión es una proteína de fusión GST-CRSP en la que las secuencias de CRSP están fusionadas al extremo C-terminal de las secuencias de GST. Tales proteínas de fusión pueden facilitar la purificación de CRSP recombinante.

Otra proteína de fusión es una proteína CRSP que contiene una secuencia señal heteróloga en su extremo N-terminal. Por ejemplo, la secuencia señal de CRSP nativa (es decir, aproximadamente los aminoácidos 1 a 23 de SEQ ID NO: 2) puede eliminarse y sustituirse por una secuencia señal de otra proteína. En ciertas células huésped (por ejemplo, células huésped de mamíferos), la expresión y/o secreción de CRSP puede aumentarse a través del uso de una secuencia señal heteróloga.

Todavía otra proteína de fusión es una proteína de fusión CRSP-inmunoglobulina en la que las secuencias de CRSP que comprenden principalmente las regiones ricas en cisteína de CRSP están fusionadas a secuencias derivadas de otro miembro de la familia de proteínas de inmunoglobulina. También se han preparado derivados solubles de glucoproteínas de superficie celular en la superfamilia de genes de inmunoglobulina que consisten en un dominio extracelular de la glucoproteína de superficie celular unido a una región constante (Fc) de inmunoglobulina (véase por ejemplo, Capon, D.J. *et al.* (1989) Nature 337:525-531 y Capon patentes estadounidenses números 5.116.964 y 5.428.130 [constructos CD4-IgGI]; Linsley, P.S. *et al.* (1991) J. Exp. Med. 173:721-730 [un constructo CD28-IgGI y un constructo B7-I-IgGI]; y Linsley, P.S. *et al.* (1991) J. Exp. Med. 174:561-569 y patente estadounidense número 5.434.131 [un CTLA4-IgG1]). Tales proteínas de fusión han resultado ser útiles para modular las interacciones receptor-ligando. Se han preparado derivados solubles de proteínas de superficie celular del receptor de las proteínas de la superfamilia del receptor de factor de necrosis tumoral (TNFR) que consisten en un dominio extracelular del receptor de superficie celular fusionado con una región constante (Fc) de inmunoglobulina (véase por ejemplo Moreland *et al.* (1997) N. Engl. J. Med. 337(3):141-147; van der Poll *et al.* (1997) Blood 89(10):3727-3734; y Ammann *et al.* (1997) J. Clin. Invest. 99 (7):1699-1703).

Las proteínas de fusión CRSP-inmunoglobulina pueden incorporarse en composiciones farmacéuticas y administrarse a un sujeto para inhibir una interacción entre un ligando de CRSP y un receptor de CRSP sobre la superficie de una célula, para así suprimir la transducción de señales mediada por CRSP *in vivo*. Las proteínas de fusión CRSP-inmunoglobulina pueden usarse para afectar a la biodisponibilidad de un receptor relacionado con CRSP. La inhibición de la interacción ligando de CRSP/CRSP puede ser terapéuticamente útil tanto para el tratamiento de trastornos de

diferenciación o proliferativos, así como para la modulación (por ejemplo, estimulación o inhibición) de respuestas de desarrollo, adhesión celular y/o destino celular. Además, las proteínas de fusión CRSP-inmunoglobulina de la invención pueden usarse como inmunógenos para producir anticuerpos anti-CRSP en un sujeto, para purificar ligandos de CRSP y en ensayos de selección para identificar moléculas que inhiben la interacción de CRSP con un ligando de CRSP.

Preferiblemente, una proteína de fusión o quimérica de CRSP se produce mediante técnicas de ADN recombinante habituales. Por ejemplo, fragmentos de ADN que codifican para diferentes secuencias polipeptídicas están ligados juntos en marco según técnicas habituales, por ejemplo empleando extremos como o en asta para ligamiento, digestión por enzimas de restricción para proporcionar los extremos apropiados, bloqueo de extremos cohesivos según conven-
ga, tratamiento con fosfatasa alcalina para evitar unión no deseable y ligamiento enzimático. El gen de fusión puede sintetizarse mediante técnicas convencionales incluyendo sintetizadores de ADN automatizados. Alternativamente, la amplificación por PCR de fragmentos génicos puede realizarse usando cebadores de anclaje que dan lugar a proyec-
ciones complementarias entre dos fragmentos génicos consecutivos que pueden aparearse posteriormente y volver a
amplificarse para generar una secuencia génica quimérica (véase, por ejemplo, *Current Protocols in Molecular Bio-*
logy, eds. Ausubel *et al.* John Wiley & Sons: 1992). Además, hay muchos vectores de expresión comercialmente
disponibles que ya codifican para un resto de fusión (por ejemplo, un polipéptido de GST). Un ácido nucleico que
codifica para CRSP puede clonarse en un vector de expresión de este tipo de manera que el resto de fusión esté unido
en marco a la proteína CRSP.

La presente invención también se refiere a variantes de las proteínas CRSP que funcionan o bien como agonistas de CRSP (compuestos miméticos) o bien como antagonistas de CRSP. Pueden generarse variantes de las proteínas CRSP mediante mutagénesis, por ejemplo, mutación puntual diferenciada o truncamiento de una proteína CRSP. Un agonista de las proteínas CRSP puede conservar sustancialmente la misma, o un subconjunto, de las actividades biológicas de la forma que se produce de manera natural de una proteína CRSP. Un antagonista de una proteína CRSP puede inhibir una o más actividades de la forma que se produce de manera natural de la proteína CRSP mediante, por ejemplo, unión competitiva a un elemento aguas abajo o aguas arriba de una cascada de señalización celular que incluye la proteína CRSP. Por tanto, puede provocarse efectos biológicos específicos mediante tratamiento con una variante de función limitada. En una realización, el tratamiento de un sujeto con una variante que tiene un subconjunto de las actividades biológicas de la forma que se produce de manera natural de la proteína tiene menos efectos secundarios en un sujeto con respecto al tratamiento con la forma que se produce de manera natural de la proteína CRSP.

Variantes de una proteína CRSP que funcionan o bien como agonistas de CRSP (compuestos miméticos) o bien como antagonistas de CRSP pueden identificarse examinando bibliotecas combinatorias de mutantes, por ejemplo, mutantes de truncamiento, de una proteína CRSP para seleccionar la actividad agonista o antagonista de una proteína CRSP. En una realización, una biblioteca diversa de variantes de CRSP se genera mediante mutagénesis combinatoria en el nivel de ácido nucleico y se codifica por una biblioteca génica diversa. Una biblioteca diversa de variantes de CRSP puede producirse mediante, por ejemplo, ligamiento enzimático de una mezcla de oligonucleótidos sintéticos en secuencias génicas de manera que un conjunto degenerado de secuencias de CRSP potenciales pueda expresarse como polipéptidos individuales, o alternativamente, como un conjunto de proteínas de fusión mayores (por ejemplo, para presentación en fagos) que contienen el conjunto de secuencias de CRSP en las mismas. Existe una variedad de métodos que pueden usarse para producir bibliotecas de variantes de CRSP potenciales a partir de una secuencia de oligonucleótidos degenerada. La síntesis química de una secuencia génica degenerada puede realizarse en un sintetizador de ADN automático, y entonces puede ligarse el gen sintético en un vector de expresión apropiado. El uso de un conjunto degenerado de genes permite proporcionar, en una mezcla, todas las secuencias que codifican para el conjunto deseado de secuencias de CRSP potenciales. En la técnica se conocen métodos para sintetizar oligonucleótidos degenerados (véase, por ejemplo, Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu. Rev. Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

Además, pueden usarse bibliotecas de fragmentos de una secuencia que codifica para una proteína CRSP para generar una población diversa de fragmentos de CRSP para examinar y posteriormente seleccionar variantes de una proteína CRSP. En una realización, puede generarse una biblioteca de fragmentos de secuencia codificante tratando un fragmento de PCR de doble cadena de una secuencia que codifica para CRSP con una nucleasa en condiciones en las que la creación de mellas se produce aproximadamente sólo una vez por molécula, desnaturalizando el ADN de cadena doble, volviendo a naturalizar el ADN para formar ADN de cadena doble que puede incluir pares sentido/antisentido de diferentes productos con mella, eliminando partes de cadena sencilla de los dúplex reformados mediante el tratamiento con nucleasa S I, y ligando la biblioteca de fragmentos resultante en un vector de expresión. Mediante este método, puede derivarse una biblioteca de expresión que codifica para fragmentos N-terminal, C-terminal e internos de diversos tamaños de la proteína CRSP.

En la técnica se conocen varias técnicas para examinar productos génicos de bibliotecas combinatorias preparados mediante mutaciones puntuales o truncamiento, y para examinar bibliotecas de ADNc para seleccionar productos génicos que tienen una propiedad seleccionada. Tales técnicas son adaptables para examinar rápidamente bibliotecas de genes generadas mediante la mutagénesis combinatoria de las proteínas CRSP. Las técnicas más ampliamente usadas, que son susceptibles de análisis riguroso, para examinar grandes bibliotecas génicas incluyen normalmente clonar la biblioteca génica en vectores de expresión replicables, transformar células apropiadas con la biblioteca de vectores resultante y expresar genes combinatorios en condiciones en las que la detección de una actividad deseada facilita el aislamiento del vector que codifica para el gen cuyo producto se detectó. La mutagénesis de conjunto

recursiva (REM - "Recursive Ensemble Mutagenesis"), una nueva técnica que potencia la frecuencia de los mutantes funcionales en las bibliotecas, puede usarse en combinación con los ensayos de selección para identificar variantes de CRSP (Arkin y Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6(3):327-331).

- 5 Los ensayos basados en células pueden explotarse para analizar una biblioteca de CRSP diversa. Por ejemplo, una biblioteca de vectores de expresión puede transfectarse en una línea celular que normalmente responde a un ligando particular de una manera dependiente de CRSP. Después se ponen en contacto las células transfectadas con el ligando y puede detectarse el efecto de la expresión del mutante sobre la señalización por el ligando, por ejemplo, midiendo cualquiera de varias respuestas celulares inmunitarias. El ADN del plásmido puede recuperarse entonces a
10 partir de las células que puntúan para la inhibición, o alternativamente, para la potenciación de inducción de ligando, y caracterizarse adicionalmente los clones individuales.

- Una proteína CRSP aislada, o una parte o fragmento de la misma, puede usarse como un inmunógeno para generar anticuerpos que se unen a CRSP usando técnicas habituales para la preparación de anticuerpo policlonal y monoclonal.
15 Puede usarse una proteína CRSP de longitud completa o, alternativamente, la invención proporciona fragmentos de péptidos antigénicos de CRSP para su uso como inmunógenos. El péptido antigénico de CRSP comprende al menos 8 residuos de aminoácidos y abarca un epítipo de CRSP de manera que un anticuerpo preparado contra el péptido forma un complejo inmunitario específico con CRSP. Preferiblemente, el péptido antigénico comprende al menos 10 residuos de aminoácidos, más preferiblemente al menos 15 residuos de aminoácidos, incluso más preferiblemente al
20 menos 20 residuos de aminoácidos, y lo más preferiblemente al menos 30 residuos de aminoácidos.

Epítopos preferidos abarcados por el péptido antigénico son regiones de CRSP que están ubicadas sobre la superficie de la proteína, por ejemplo, regiones hidrófilas.

- 25 Normalmente se usa un inmunógeno de CRSP para preparar anticuerpos mediante inmunización de un sujeto adecuado, (por ejemplo, conejo, cabra, ratón u otro mamífero) con el inmunógeno. Una preparación inmunogénica apropiada puede contener, por ejemplo, proteína CRSP expresada de manera recombinante o un polipéptido de CRSP químicamente sintetizado. La preparación puede incluir adicionalmente un adyuvante, tal como un adyuvante completo o incompleto de Freund, o agente inmunoestimulante similar. La inmunización de un sujeto adecuado con una
30 preparación de CRSP inmunogénica induce una respuesta de anticuerpo policlonal anti-CRSP.

- Por consiguiente, otro aspecto de la invención se refiere a anticuerpos anti-CRSP. El término "anticuerpo" tal como se usa en el presente documento se refiere a moléculas de inmunoglobulina y partes inmunológicamente activas de moléculas de inmunoglobulinas, es decir, moléculas que contienen un sitio de unión a antígeno que se une específicamente a (inmunorreacciona con) un antígeno, tal como CRSP. Ejemplos de partes inmunológicamente activas de moléculas de inmunoglobulinas incluyen fragmentos F(ab) y F(ab')₂ que pueden generarse tratando el anticuerpo con una enzima tal como pepsina. La invención proporciona anticuerpos policlonales y monoclonales que se unen a CRSP. El término "anticuerpo monoclonal" o "composición de anticuerpo monoclonal", tal como se usa en el presente documento, se refiere a una población de moléculas de anticuerpo que contienen sólo una especie de sitio de unión a
40 antígeno que puede inmunorreaccionar con un epítipo particular de CRSP. Por tanto, una composición de anticuerpo monoclonal presenta una afinidad de unión única para una proteína CRSP particular con la que inmunorreacciona.

- Pueden prepararse anticuerpos policlonales anti-CRSP tal como se describió anteriormente inmunizando un sujeto adecuado con un inmunógeno de CRSP. El título de anticuerpo anti-CRSP en el sujeto inmunizado puede monitorizarse
45 a lo largo del tiempo mediante técnicas habituales, tales como con un ensayo de inmunoabsorción unido a enzimas (ELISA) usando CRSP inmovilizada. Si se desea, las moléculas de anticuerpo dirigidas contra CRSP pueden aislarse del mamífero (por ejemplo, de la sangre) y purificarse adicionalmente mediante técnicas bien conocidas, tales como cromatografía de proteína A para obtener la fracción de IgG. En un momento apropiado tras la inmunización, por ejemplo, cuando los títulos de anticuerpo anti-CRSP son lo más altos, pueden obtenerse las células productoras de
50 anticuerpos a partir del sujeto y usarse para preparar anticuerpos mediante técnicas habituales, tales como la técnica del hibridoma originalmente descrita por Kohler y Milstein (1975) Nature 256:495-497 (véase también, Brown *et al.* (1981) J. Immunol. 127:539-46; Brown *et al.* (1980) J. Biol. Chem. 255:4980-83; Yeh *et al.* (1976) PNAS 76:2927-31; y Yeh *et al.* (1982) Int. J. Cancer 29:269-75), la técnica de hibridoma de célula B humana más reciente (Kozbor *et al.* (1983) Immunol Today 4:72), la técnica de hibridoma de EBV (Cole *et al.* (1985), Monoclonal Antibodies and Cancer
55 Therapy, Alan R. Liss, Inc., pp. 77-96) o técnicas de trioma. La tecnología para producir hibridomas de anticuerpo monoclonal se conoce bien (véase de manera general R. H. Kenneth, en Monoclonal Antibodies: A New Dimension In Biological Analyses, Plenum Publishing Corp., Nueva York, Nueva York (1980); E. A. Lerner (1981) Yale J. Biol. Med., 54:387-402; M. L. Gefter *et al.* (1977) Somatic Cell Genet. 3:231-36). En resumen, una línea celular inmortal (normalmente un mieloma) se fusiona con linfocitos (normalmente esplenocitos) de un mamífero inmunizado con un
60 inmunógeno de CRSP tal como se describió anteriormente, y los sobrenadantes del cultivo de las células de hibridoma resultantes se examinan para identificar un hibridoma que produce un anticuerpo monoclonal que se une a CRSP.

- Cualquiera de los muchos protocolos bien conocidos usados para fusionar linfocitos y líneas celulares inmortalizadas puede aplicarse con fines de generación de un anticuerpo monoclonal anti-CRSP (véase, por ejemplo, G. Galfre *et al.* (1977) Nature 266:55052; Gefter *et al.* Somatic Cell Genet., citado anteriormente; Lerner, Yale J Biol. Med., citado anteriormente; Kenneth, Monoclonal Antibodies, citado anteriormente). Además, el experto en la técnica apreciará que existen muchas variaciones de tales métodos que también pueden ser útiles. Normalmente, la línea celular inmortal (por ejemplo, una línea celular de mieloma) se deriva de la misma especie de mamífero que los linfocitos. Por ejemplo,

pueden prepararse hibridomas murinos fusionando linfocitos de un ratón inmunizado con una preparación inmunogénica de la presente invención con una línea celular de ratón inmortalizado. Líneas celulares inmortales preferidas son líneas celulares de mieloma de ratón que son sensibles a medio de cultivo que contiene hipoxantina, aminopterina y timidina ("medio HAT"). Puede usarse cualquiera de una variedad de líneas celulares de mieloma como una pareja de fusión según técnicas habituales, por ejemplo, las líneas de mieloma P3-NS1/1-Ag4-1, P3-x63-Ag8.653 o Sp2/O-Ag14. Estas líneas de mieloma están disponibles de la ATCC. Normalmente, las células de mieloma de ratón sensibles a HAT se fusionan con esplenocitos de ratón usando polietilenglicol ("PEG"). Entonces se seleccionan células de hibridoma resultantes de la fusión usando medio HAT, que mata células de mieloma no fusionadas y fusionadas de manera no productiva (los esplenocitos no fusionados mueren tras varios días porque no se transforman). Se detectan células de hibridoma que producen un anticuerpo monoclonal de la invención examinando los sobrenadantes de cultivo para seleccionar anticuerpos que se unen a CRSP, por ejemplo, usando un ensayo ELISA habitual.

Alternativamente a la preparación de hibridomas que segregan anticuerpo monoclonal, puede identificarse un anticuerpo anti-CRSP monoclonal y aislarse examinando una biblioteca de inmunoglobulina combinatoria recombinante (por ejemplo, una biblioteca de presentación en fagos de anticuerpo) con CRSP para así aislar miembros de la biblioteca de inmunoglobulina que se unen a CRSP. Existen kits disponibles comercialmente para generar y examinar bibliotecas de presentación de fagos (por ejemplo, el sistema de anticuerpo de fago recombinante ("Recombinant Phage Antibody System") de Pharmacia, número de catálogo 27-9400-0 1; y el kit de presentación de fago ("Phage Display Kit") SurfZAP™ de Stratagene, número de catálogo 240612). Adicionalmente, ejemplos de métodos y reactivos particularmente útiles para su uso para generar y examinar bibliotecas de presentación de anticuerpo pueden encontrarse en, por ejemplo, Ladner *et al.* patente estadounidense número 5.223.409; Kang *et al.* publicación PCT internacional número WO 92/18619; Dower *et al.* publicación PCT internacional número WO 91/17271; Winter *et al.* publicación PCT internacional WO 92/20791; Markland *et al.* publicación PCT internacional número WO 92/15679; Breitling *et al.* publicación PCT internacional WO 93/01288; McCafferty *et al.* publicación PCT internacional número WO 92/01047; Garrard *et al.* publicación PCT internacional número WO 92/09690; Ladner *et al.* publicación PCT internacional número WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *PNAS* 89:3576-3580; Garrard *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *PNAS* 88:7978-7982; y McCafferty *et al.* *Nature* (1990) 348:552-554.

Adicionalmente, anticuerpos anti-CRSP recombinantes, tal como anticuerpos monoclonales quiméricos y humanizados, que comprenden tanto partes humanas como no humanas, que pueden prepararse usando técnicas de ADN recombinante habituales, están dentro del alcance de la invención. Tales anticuerpos monoclonales quiméricos y humanizados pueden producirse mediante técnicas de ADN recombinante conocidas en la técnica, por ejemplo usando métodos descritos en Robinson *et al.* solicitud internacional número PCT/US86/02269; Akira, *et al.* solicitud de patente europea 184.187; Taniguchi, M., solicitud de patente europea 171.496; Morrison *et al.* solicitud de patente europea 173.494; Neuberger *et al.* publicación PCT internacional número WO 86/01533; Cabilly *et al.* patente estadounidense número 4.816.567; Cabilly *et al.* solicitud de patente europea 125.023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; y Shaw *et al.* (1988) *J Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter patente estadounidense número 5.225.539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; y Beidler *et al.* (1988) *J Immunol.* 141:4053-4060.

Puede usarse un anticuerpo anti-CRSP (por ejemplo, anticuerpo monoclonal) para aislar CRSP mediante técnicas habituales, tales como cromatografía de afinidad o inmunoprecipitación. Un anticuerpo anti-CRSP puede facilitar la purificación de CRSP natural a partir de células y de CRSP producidas de manera recombinante expresadas en células huésped. Además, puede usarse un anticuerpo anti-CRSP para detectar proteína CRSP (por ejemplo, en un lisado celular o sobrenadante celular) con el fin de evaluar la abundancia y patrón de expresión de la proteína CRSP. Pueden usarse anticuerpos anti-CRSP de manera diagnóstica para monitorizar niveles de proteína en tejido, por ejemplo, como parte de procedimiento de ensayo clínico, para, por ejemplo, determinar la eficacia de un régimen de tratamiento dado. Puede facilitarse la detección acoplando (es decir, uniéndolo físicamente) el anticuerpo a una sustancia detectable. Ejemplos de sustancias detectables incluyen diversas enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes, y materiales radiactivos. Ejemplos de enzimas adecuadas incluyen peroxidasa de rábano, fosfatasa alcalina, β -galactosidasa, o acetilcolinesterasa; ejemplos de complejos de grupo prostético incluyen estreptavidina/biotina y avidina/biotina; ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína, isocianato de fluoresceína, rodamina, diclorotriazinilamina fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de un material luminiscente incluye luminol; ejemplos de materiales bioluminiscentes incluyen luciferasa, luciferina, y aequorina, y ejemplos de materiales radiactivos adecuados incluyen ^{125}I , ^{131}I , ^{35}S o ^3H .

III. Vectores de expresión recombinantes y células huésped

Otro aspecto de la invención se refiere a vectores, preferiblemente vectores de expresión, que contienen un ácido nucleico que codifica para CRSP (o una parte de la misma). Tal como se usa en el presente documento, el término "vector" se refiere a una molécula de ácido nucleico que puede transportar otro ácido nucleico al que se ha unido. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de cadena doble circular en el que pueden ligarse segmentos de ADN adicionales. Otro tipo de vector es un vector viral, en el que pueden ligarse segmentos

de ADN adicionales en el genoma viral. Ciertos vectores pueden experimentar replicación autónoma en una célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano y vectores de mamíferos episomales). Otros vectores (por ejemplo, vectores de mamíferos no episomales) se integran en el genoma de una célula huésped tras la introducción en la célula huésped, y por tanto se replican junto con el genoma huésped. Además, ciertos vectores pueden dirigir la expresión de genes a los que están operativamente unidos. Tales vectores se denominan en el presente documento como “vectores de expresión”. En general, vectores de expresión de utilidad en técnicas de ADN recombinante a menudo están en forma de plásmidos. En la presente memoria descriptiva, “plásmido” y “vector” pueden usarse indistintamente ya que el plásmido es la forma más comúnmente usada de vector. Sin embargo, se pretende que la invención incluya otras formas de vectores de expresión, tales como vectores virales (por ejemplo, virus adenoasociados, adenovirus y retrovirus defectuosos para la replicación), que presentan funciones equivalentes.

Los vectores de expresión recombinantes de la invención comprenden un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen una o más secuencias reguladoras, seleccionadas basándose en las células huésped que van a usarse para la expresión, que están operativamente unidas a la secuencia de ácido nucleico secuencia que va a expresarse. Dentro de un vector de expresión recombinante, se pretende que “operativamente unido” signifique que la secuencia de nucleótidos de interés está unida a la(s) secuencia(s) reguladora(s) de una manera que permite la expresión de la secuencia de nucleótidos (por ejemplo, en un sistema transcripción/traducción *in vitro* o en una célula huésped cuando se introduce el vector en la célula huésped). Se pretende que el término “secuencia reguladora” incluya promotores, potenciadores y otros elementos de control de expresión (por ejemplo, señales de poliadenilación). Tales secuencias reguladoras se describen, por ejemplo, en Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Las secuencias reguladoras incluyen las que dirigen expresión constitutiva de una secuencia de nucleótidos en muchos tipos de célula huésped y las que dirigen la expresión de la secuencia de nucleótidos sólo en ciertas células huésped (por ejemplo, secuencias reguladoras específicas de tejido). Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va a transformarse, el nivel de expresión de proteína deseado, etc. Los vectores de expresión de la invención pueden introducirse en células huésped para así producir proteínas o péptidos, incluyendo péptidos o proteínas de fusión, codificadas por ácidos nucleicos tal como se describió en el presente documento (por ejemplo, proteínas CRSP, formas mutantes de CRSP, proteínas de fusión, etc.).

Los vectores de expresión recombinantes de la invención pueden diseñarse para expresar CRSP en células procariontas o eucariotas. Por ejemplo, puede expresarse CRSP en células bacterianas tal como *E. coli*, células de insecto (usando vectores de expresión de baculovirus), células de levadura o células de mamífero. Células huésped adecuadas se tratan adicionalmente en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Alternativamente, puede transcribirse y traducirse el vector de expresión recombinante *in vitro*, por ejemplo usando secuencias reguladoras de promotor de T7 y polimerasa de T7.

La expresión de proteínas en procariontas se realiza lo más a menudo en *E. coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas o bien de fusión o bien no de fusión. Los vectores de fusión añaden una variedad de aminoácidos a una proteína codificada en los mismos, normalmente en el extremo amino-terminal de la proteína recombinante. Tales vectores de fusión normalmente sirven para tres fines: 1) aumentar la expresión de proteína recombinante; 2) aumentar la solubilidad de la proteína recombinante; y 3) ayudar en la purificación de la proteína recombinante actuando como un ligando en purificación por afinidad. A menudo, en vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante del resto de fusión posterior a la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen factor Xa, trombina y enterocinasa. Los vectores de expresión de fusión típicos incluyen pGEX (Pharmacia Biotech Inc; Smith, D. B. y Johnson, K. S. (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) y pRIT5 (Pharmacia, Piscataway, NJ) que fusionan glutatión S-transferasa (GST), proteína de unión a maltosa E, o proteína A, respectivamente, a la proteína recombinante diana.

Pueden utilizarse proteínas de fusión purificadas en ensayos de actividad de CRSP, en unión a ligando de CRSP (por ejemplo, ensayos directos o ensayos competitivos descritos con detalle a continuación), para generar anticuerpos específicos para proteínas CRSP, como ejemplos. En una realización preferida, una fusión de CRSP expresada en un vector de expresión retroviral de la presente invención puede utilizarse para infectar células de la médula ósea que se transplantan posteriormente en los receptores irradiados. La patología del sujeto receptor se examina entonces tras haber transcurrido suficiente tiempo (por ejemplo seis (6) semanas).

Ejemplos de vectores de expresión de *E. coli* no de fusión incluyen pTrc (Amann *et al.*, (1988) Gene 69:301-315) y pET 11d (Studier *et al.*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). La expresión de gen diana a partir del vector pTrc se basa en la transcripción de ARN polimerasa del huésped a partir de un promotor de fusión híbrido *trp-lac*. La expresión del gen diana a partir del vector pET 11d se basa en la transcripción a partir de un promotor de fusión *gn10-lac* de T7 mediada por una ARN polimerasa viral coexpresada (*gn1* de T7). Esta polimerasa viral se suministra por cepas huésped BL21(DE3) o HMS174(DE3) de un profago λ residente que alberga un gen *gn1* de T7 bajo control transcripcional del promotor 5 *lacUV*.

Una estrategia para maximizar la expresión de proteína recombinante en *E. coli* es expresar la proteína en una bacteria huésped con una capacidad afectada para escindir proteolíticamente la proteína recombinante (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Otra estrategia es alterar la secuencia de ácido nucleico del ácido nucleico que va a insertarse en un vector de expresión de manera que los codones individuales para cada aminoácido son los utilizados preferentemente en *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Tal alteración de secuencias de ácido nucleico de la invención puede realizarse mediante técnicas de síntesis de ADN habituales.

En otra realización, el vector de expresión de CRSP es un vector de expresión de levadura. Ejemplos de vectores para la expresión en levadura *S. cerevisiae* incluyen pYepSec1 (Baldari, *et al.*, (1987) *Embo J* 6:229-234), pMFa (Kurjan y Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), y picZ (InVitrogen Corp, San Diego, CA).

Alternativamente, puede expresarse CRSP en células de insecto usando vectores de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (por ejemplo, células Sf 9) incluyen la serie pAc (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) y la serie pVL (Lucklow y Summers (1989) *Virology* 170:31-39).

Todavía en otra realización, un ácido nucleico de la invención se expresa en células de mamífero usando un vector de expresión de mamífero. Ejemplos de vectores de expresión de mamíferos incluyen pCDM8 (Seed, B. (1987) *Nature* 329:840) y pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). Cuando se usa en células de mamífero, las funciones de control del vector de expresión se proporcionan a menudo por elementos reguladores virales. Por ejemplo, promotores comúnmente usados derivan de poliovirus, adenovirus 2, citomegalovirus y virus de simio 40. Para otros sistemas de expresión adecuados tanto para células procariotas como eucariotas véanse los capítulos 16 y 17 de Sambrook, J., Fritsch, E. F., y Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

En otra realización, el vector de expresión de mamífero recombinante puede dirigir la expresión del ácido nucleico preferiblemente en un tipo de célula particular (por ejemplo, se usan elementos reguladores específicos de tejido para expresar el ácido nucleico). En la técnica se conocen elementos reguladores específicos de tejido. Ejemplos no limitantes de promotores específicos de tejido adecuados incluyen el promotor de albúmina (específico de hígado; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), promotores específicos de linfocitos (Calame y Eaton (1988) *Adv. Immunol.* 43:235-275), en particular promotores de receptores de células T (Winoto y Baltimore (1989) *EMBO J.* 8:729-733) e inmunoglobulinas (Banerji *et al.* (1983) *Cell* 33:729-740; Queen y Baltimore (1983) *Cell* 33:741-748), promotores específicos de neuronas (por ejemplo, el promotor de neurofilamento; Byrne y Ruddle (1989) *PNAS* 86:5473-5477), promotores específicos de páncreas (Edlund *et al.* (1985) *Science* 230:912-916), y promotores específicos de glándula mamaria (por ejemplo, promotor de suero de leche; patente estadounidense 4.873.316 y solicitud de patente europea número 264.166). También se abarcan promotores regulados por desarrollo, por ejemplo los promotores de hox (Kessel y Gruss (1990) *Science* 249:374-379) y el promotor de α -fetoproteína (Campes y Tilghman (1989) *Genes Dev.* 3:537-546).

La invención proporciona además un vector de expresión recombinante que comprende una molécula de ADN de la invención clonada en el vector de expresión en una orientación antisentido. Es decir, la molécula de ADN está operativamente unida a una secuencia reguladora de una manera que permite la expresión (mediante transcripción de la molécula de ADN) de una molécula de ARN que es antisentido con respecto a ARNm de CRSP. Pueden elegirse secuencias reguladoras operativamente unidas a un ácido nucleico clonado en la orientación antisentido que dirijan la expresión continua de una molécula de ARN antisentido en una variedad de tipos de células, por ejemplo potenciadores y/o promotores virales, o pueden elegirse secuencias reguladoras que dirijan la expresión específica de tejido o específica de tipo de célula constitutiva de ARN antisentido. El vector de expresión antisentido puede estar en forma de un plásmido recombinante, fagémido o virus atenuado en el que los ácidos nucleicos antisentido se producen bajo el control de una región reguladora de alta eficacia, cuya actividad puede determinarse por el tipo de célula en la que se introduce el vector. Para una evaluación de la regulación de expresión génica usando genes antisentido véase Weintraub, H. *et al.*, *Antisense ARN as a molecular tool for genetic analysis*, *Reviews - Trends in Genetics*, Vol. 1 (1) 1986.

Otro aspecto de la invención se refiere a células huésped en las que se ha introducido un vector de expresión recombinante de la invención. Los términos "célula huésped" y "célula huésped recombinante" se usan indistintamente en el presente documento. Se entiende que tales términos se refieren no sólo a la célula objeto particular sino a la progenie o progenie potencial de una célula de este tipo. Dado que pueden producirse ciertas modificaciones en generaciones posteriores debido o bien a mutación o bien a influencias ambientales, tal progenie puede, de hecho, no ser idéntica a la célula original, pero aun así se incluyen dentro del alcance del término tal como se usa en el presente documento.

Una célula huésped puede ser cualquier célula procariota o eucariota. Por ejemplo, puede expresarse una proteína CRSP en células bacterianas tales como *E. coli*, células de insecto, células de levadura o mamífero (tales como células de ovario de hámster chino (CHO) o células COS). Los expertos en la técnica conocen otras células huéspedes adecuadas.

El ADN del vector puede introducirse en células procariotas o eucariotas mediante técnicas de transfección o transformación convencionales. Tal como se usa en el presente documento, se pretende que los términos “transformación” y “transfección” se refieran a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula huésped, incluyendo coprecipitación de fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, lipofección, o electroporación. Métodos adecuados para transformar o transfectar células huésped pueden encontrarse en Sambrook, *et al.* (Molecular Cloning: A Laboratory Manual. 2ª, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), y otros manuales de laboratorio.

Para la transfección estable de células de mamífero, se sabe que, dependiendo del vector de expresión y técnica de transfección usada, sólo una pequeña fracción de células puede integrar el ADN foráneo en su genoma. Con el fin de identificar y seleccionar estos integrantes, un gen que codifica para un marcador seleccionable (por ejemplo, resistencia frente a antibióticos) se introduce generalmente en las células huésped junto con el gen de interés. Marcadores seleccionables preferidos incluyen los que confieren resistencia frente a fármacos, tales como G418, higromicina y metotrexato. Puede introducirse ácido nucleico que codifica para un marcador seleccionable en una célula huésped en el mismo vector que el que codifica para CRSP o puede introducirse en un vector separado. Pueden identificarse células transfectadas de manera estable con el ácido nucleico introducido mediante selección de fármaco (por ejemplo, las células que han incorporado el gen de marcador seleccionable sobrevivirán, mientras que las otras células morirán).

Una célula huésped de la invención, tal como una célula huésped procariota o eucariota en cultivo, puede usarse para producir (es decir, expresar) proteína CRSP. Por consiguiente, la invención proporciona adicionalmente métodos para producir proteína CRSP usando las células huésped de la invención. En una realización, el método comprende poner en cultivo la célula huésped de la invención (en la que se ha introducido un vector de expresión recombinante que codifica para CRSP) en un medio adecuado de manera que se produce la proteína CRSP. En otra realización, el método comprende adicionalmente aislar CRSP del medio o de la célula huésped.

Las células huésped de la invención también pueden usarse para producir animales transgénicos no humanos. Por ejemplo, en una realización, una célula huésped de la invención es un ovocito fertilizado no humano o una célula madre embrionaria no humana en la que se han introducido secuencias que codifican para CRSP. Tales células huésped pueden usarse entonces para crear animales transgénicos no humanos en los que se han introducido secuencias de CRSP exógenas en su genoma o animales recombinantes homólogos en los que se han alterado secuencias de CRSP endógenas. Tales animales son útiles para el estudio de la función y/o actividad de CRSP y para identificar y/o evaluar moduladores de actividad de CRSP. Tal como se usa en el presente documento, un “animal transgénico” es un animal no humano, preferiblemente un mamífero, más preferiblemente un roedor tal como una rata o ratón, en el que una o más de las células del animal incluyen un transgén. Otros ejemplos de animales transgénicos incluyen primates no humanos, ovejas, perros, vacas, cabras, pollos, anfibios, etc. Un transgén es ADN exógeno que se integra en el genoma de una célula a partir de la cual se desarrolla un animal transgénico y que permanece en el genoma del animal maduro, dirigiendo de esta manera la expresión de un producto génico codificado en uno o más tipos de células o tejidos del animal transgénico. Tal como se usa en el presente documento, un “animal recombinante homólogo” es un animal no humano, preferiblemente un mamífero, más preferiblemente un ratón, en el que se ha alterado un gen de CRSP endógeno mediante recombinación homóloga entre el gen endógeno y la molécula de ADN exógeno introducido en una célula del animal, por ejemplo, una célula embrionaria del animal, antes del desarrollo del animal.

Un animal transgénico no humano de la invención puede crearse introduciendo un ácido nucleico que codifica para CRSP en el pronúcleo masculino de un ovocito fertilizado, por ejemplo, mediante microinyección, infección retroviral, y permitiendo que el ovocito se desarrolle en un animal de acogida hembra pseudopreñada. La secuencia de ADNc de CRSP humana de la invención puede introducirse como un transgén en el genoma de un animal no humano. Alternativamente, un homólogo no humano de un gen de CRSP humana, tal como un gen de CRSP de ratón, puede aislarse basándose en la hibridación con el ADNc de CRSP humana (descrito adicionalmente en la subsección I anterior) y usarse como un transgén. También pueden incluirse secuencias intrónicas y señales de poliadenilación en el transgén para aumentar la eficacia de la expresión del transgén. Puede(n) unirse operativamente una(s) secuencia(s) reguladora(s) específica(s) de tejido al transgén de CRSP para dirigir la expresión de la proteína CRSP a células particulares. Los métodos para generar animales transgénicos no humanos mediante manipulación embrionaria y microinyección, particularmente en animales tales como ratones, se han convertido en convencionales en la técnica y se describen, por ejemplo, en las patentes estadounidenses números 4.736.866 y 4.870.009, ambas concedidas a Leder *et al.*, patente estadounidense número 4.873.191 concedida a Wagner *et al.*, y en Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Se usan métodos similares para la producción de otros animales transgénicos no humanos. Un animal fundador transgénico no humano puede identificarse basándose en la presencia del transgén de CRSP en su genoma y/o expresión de ARNm de CRSP en tejidos o células de los animales. Entonces puede usarse un animal fundador transgénico no humano para criar animales adicionales que llevan el transgén. Además, los animales transgénicos no humanos que llevan un transgén que codifica para CRSP pueden criarse adicionalmente para generar otros animales transgénicos no humanos que llevan otros transgenes.

Para crear un animal no humano recombinante homólogo, se prepara un vector que contiene al menos una parte de un gen de CRSP en el que se ha introducido una delección, adición o sustitución para así alterar, por ejemplo, afectar funcionalmente, el gen de CRSP. El gen de CRSP puede ser un gen humano (por ejemplo, el ADNc de SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 9 o SEQ ID NO: 12), pero más preferiblemente, es un homólogo no humano de un gen de CRSP humano. Por ejemplo, un gen de CRSP de ratón de SEQ ID NO: 16 puede usarse para construir un

vector de recombinación homólogo adecuado para alterar un gen de CRSP endógeno en el genoma del ratón. En una realización preferida, se diseña el vector de manera que, tras la recombinación homóloga, el gen de CRSP endógeno se altera funcionalmente (es decir, ya no codifica para una proteína funcional; también denominado como un vector “deficiente”). Alternativamente, puede diseñarse el vector de manera que, tras la recombinación homóloga, el gen de CRSP endógeno se muta o de otra manera se altera pero aún codificará para proteína funcional (por ejemplo, puede alterarse la región reguladora en sentido 5' para así alterar la expresión de la proteína CRSP endógena). En el vector de recombinación homólogo, la parte alterada del gen de CRSP está flanqueada en sus extremos 5' y 3' por ácido nucleico adicional del gen de CRSP para permitir que se produzca la recombinación homóloga entre el gen de CRSP exógeno llevado por el vector y un gen de CRSP endógeno en una célula madre embrionaria no humana. El ácido nucleico de CRSP flanqueante adicional es de suficiente longitud como para una recombinación homóloga satisfactoria con el gen endógeno. Normalmente, se incluyen varias kilobases de ADN flanqueante (tanto en el extremo 5' como 3') en el vector (véase por ejemplo, Thomas, K.R. y Capecchi, M.R. (1987) *Cell* 51:503 para una descripción de vectores de recombinación homólogos). Se introduce el vector en una línea de células madre embrionarias no humanas (por ejemplo, mediante electroporación) y se seleccionan células en las que el gen de CRSP introducido se ha recombinado de manera homóloga con el gen de CRSP endógeno (véase por ejemplo, Li, E. *et al.*, (1992) *Cell* 69:915). Las células seleccionadas se inyectan entonces en un blastocito de un animal no humano (por ejemplo, un ratón) para formar quimeras de agregación (véase por ejemplo, Bradley, A. en *Teratocarcinomas and Embryonic Stem Cells A Practical Approach*, E. J. Robertson, ed. (IRL, Oxford, 1987) págs. 113-152). Entonces puede implantarse un embrión quimérico no humano en un animal de acogida no humano hembra pseudopreñada adecuada y llevar el embrión a término. La progenie que alberga el ADN recombinado de manera homóloga en sus células germinales puede usarse para criar animales en los que todas las células del animal contienen el ADN recombinado de manera homóloga por la transmisión de línea germinal del transgén. Se describen adicionalmente métodos para construir vectores de recombinación homólogos y animales recombinantes homólogos en Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 y en las publicaciones internacionales PCT números: WO 90/11354 por Le Mouellec *et al.*; WO 91/01140 por Smithies *et al.*; WO 92/0968 por Zijlstra *et al.*; y WO 93/04169 por Berns *et al.*

En otra realización, pueden producirse animales no humanos transgénicos que contienen sistemas seleccionados que permiten expresión regulada del transgén. Un ejemplo de un sistema de este tipo es el sistema *cre/loxP* recombinasa del bacteriófago P1. para una descripción del sistema *cre/loxP* recombinasa, véase, por ejemplo, Lakso *et al.* (1992) *PNAS* 89: 6232-6236. Otro ejemplo de sistema recombinasa es el sistema FLP recombinasa de *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355). Si se usa un sistema *cre/loxP* recombinasa para regular la expresión del transgén, se requieren animales que contienen transgenes que codifican tanto para la *Cre* recombinasa como para una proteína seleccionada. Tales animales pueden proporcionarse mediante la construcción de animales transgénicos “dobles” no humanos, por ejemplo, mediante reproducción de dos animales transgénicos no humanos, conteniendo uno un transgén que codifica para una proteína seleccionada y conteniendo el otro un transgén que codifica para una recombinasa.

Clones de los animales transgénicos no humanos descritos en el presente documento también pueden producirse según los métodos descritos en Wilmut, I. *et al.* (1997) *Nature* 385:810-813. En resumen, una célula, por ejemplo, una célula somática, del animal transgénico puede aislarse e inducirse a salir del ciclo de crecimiento y entrar en la fase G₀. Después puede fusionarse la célula quiescente, por ejemplo, mediante el uso de pulsos eléctricos, a un ovocito endonucleado no humano de un animal de la misma especie a partir de la que se aísla la célula quiescente. El ovocito no humano reconstruido se cultiva entonces de manera que se desarrolle para dar una mórula o blastocito y luego se transfiere a un animal de acogida hembra pseudopreñada. Las crías nacidas de este animal de acogida hembra serán un clon del animal a partir del cual se aísla la célula, por ejemplo, la célula somática.

IV. Composiciones farmacéuticas

Las moléculas de ácido nucleico de CRSP, las proteínas CRSP, y los anticuerpos anti-CRSP (también denominados en el presente documento “principios activos”) de la invención pueden incorporarse en composiciones farmacéuticas adecuadas para su administración. Tales composiciones comprenden normalmente la molécula de ácido nucleico, proteína, o anticuerpo y un vehículo farmacéuticamente aceptable. Tal como se usa en el presente documento, la expresión “vehículo farmacéuticamente aceptable” pretende incluir cualquiera y todos los disolventes, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes retardantes de la absorción e isotónicos, y similares, compatibles con la administración farmacéutica. El uso de tales medios y agentes para los principios farmacéuticamente activos se conoce bien en la técnica. Excepto en la medida en que cualquier medio o agente convencional sea incompatible con el principio activo, se contempla el uso del mismo en las composiciones. También pueden incorporarse en las composiciones, principios activos complementarios.

Una composición farmacéutica de la invención se formula para que sea compatible con su vía de administración deseada. Ejemplos de vías de administración incluyen la administración parenteral, por ejemplo, vía intravenosa, intradérmica, subcutánea, oral (por ejemplo, inhalación), transdérmica (tópica), transmucosa, y rectal. Las disoluciones o suspensiones utilizadas para la aplicación parenteral, intradérmica, o subcutánea pueden incluir los siguientes componentes: un diluyente estéril, tal como agua para inyección, solución salina, aceites fijos, polietilenglicoles, glicerina, propilenglicol u otros disolventes sintéticos; agentes antibacterianos tales como alcohol benéfico o parabenos metílicos; antioxidantes tales como ácido ascórbico o bisulfito de sodio; agentes quelantes tales como ácido etilendiaminotetraacético; tampones tales como acetatos, citratos o fosfatos y agentes para el ajuste de la tonicidad tales como el cloruro de sodio o la dextrosa. El pH puede ajustarse con ácidos o bases, tales como ácido clorhídrico o hidróxido de

sodio. La preparación parenteral puede incluirse en ampollas, jeringas desechables o viales de dosis múltiples hechos de vidrio o plástico.

Las composiciones farmacéuticas adecuadas para el uso inyectable incluyen disoluciones acuosas estériles (que son solubles en agua) o dispersiones y polvos estériles para la preparación improvisada de la dispersión o las disoluciones inyectables estériles. Para la administración intravenosa, los vehículos adecuados incluyen solución salina fisiológica, agua bacteriostática, Cremophor EL™ (BASF, Parsippany, NJ) solución salina tamponada con fosfato (PBS). En todos los casos, la composición debe ser estéril y debe ser fluida hasta el grado de que exista una fácil capacidad de inyección. Debe ser estable en las condiciones de fabricación y almacenamiento y debe conservarse frente a la acción contaminante de microorganismos tales como bacterias y hongos. El vehículo puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol, y polietilenglicol líquido, y similares), y mezclas adecuadas de los mismos. La fluidez apropiada puede mantenerse, por ejemplo, mediante el uso de un recubrimiento tal como lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de la dispersión y mediante el uso de tensioactivos. La prevención de la acción de los microorganismos puede lograrse mediante diversos agentes antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido ascórbico, timerosal, y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol, cloruro de sodio, en la composición. Puede provocarse la absorción prolongada de las composiciones inyectables incluyendo en la composición un agente que retrasa la absorción, por ejemplo, monoestearato de aluminio y gelatina.

Las disoluciones inyectables estériles pueden prepararse mediante la incorporación del principio activo (por ejemplo, una proteína CRSP o anticuerpo anti-CRSP) en la cantidad requerida en un disolvente apropiado con uno o una combinación de los componentes enumerados anteriormente, según sea necesario, seguido por esterilización por filtración. Generalmente, las dispersiones se preparan mediante la incorporación del principio activo en un vehículo estéril que contiene un medio de dispersión básico y otros componentes de los enumerados anteriormente. En el caso de los polvos estériles para la preparación de las disoluciones inyectables estériles, los métodos de preparación preferidos son secado a vacío, y liofilización, lo que produce un polvo del principio activo más cualquier componente adicional deseado a partir de una disolución previamente filtrada para esterilidad del mismo.

Las composiciones orales generalmente incluyen un diluyente inerte o un vehículo comestible. Pueden incorporarse en cápsulas de gelatina o fabricarse por compresión para dar lugar a comprimidos. Para el fin de la administración terapéutica oral, el principio activo puede incorporarse con excipientes y utilizarse en la forma de comprimidos, trociscos o cápsulas. Las composiciones orales también pueden prepararse utilizando un vehículo fluido para su uso como un enjuague bucal, en el que el compuesto en el vehículo fluido se aplica por vía oral y se utiliza como enjuague y se expectora o se traga. Pueden incluirse agentes de unión y/o materiales adyuvantes farmacéuticamente compatibles como parte de la composición. Los comprimidos, píldoras, cápsulas, trociscos y similares pueden contener cualquiera de los componentes o compuestos siguientes de naturaleza similar: un aglutinante tal como celulosa microcristalina, goma de tragacanto o gelatina; un excipiente tal como almidón o lactosa; un agente discrepante tal como ácido algínico, Primogel o almidón de maíz; un lubricante tal como estearato de magnesio o Sterotes; un deslizante tal como dióxido de silicio coloidal; un agente edulcorante tal como sacarosa o sacarina; o un agente aromatizante tal como menta, salicilato de metilo o aromatizante de naranja.

Para la administración mediante inhalación, los compuestos se administran en la forma de un pulverizador de aerosol a partir de un recipiente o dispensador presurizado que contiene un propelente adecuado, por ejemplo, un gas tal como dióxido de carbono, o un nebulizador.

La administración sistémica también puede realizarse mediante medios transmucosos o transdérmicos. Para la administración transmucosa o transdérmica, se utilizan en la formulación agentes penetrantes apropiados para la permeación de la barrera. Tales agentes penetrantes se conocen generalmente en la técnica, e incluyen, por ejemplo, para la administración transmucosa, detergentes, sales biliares y derivados del ácido fusídico. La administración transmucosa puede llevarse a cabo a través del uso de aerosoles nasales o supositorios. Para la administración transdérmica, los principios activos se formulan en ungüentos, bálsamos, geles o cremas, tal como se conoce generalmente en la técnica.

Los compuestos también pueden prepararse en la forma de supositorios (por ejemplo, con bases de supositorio convencionales tales como mantequilla de cacao y otros glicéridos) o enemas de retención para la administración rectal.

En una realización, los principios activos se preparan con vehículos que protegerán al compuesto frente a la rápida eliminación del organismo, tales como una formulación de liberación controlada, incluyendo implantes y sistemas de administración microencapsulados. Pueden usarse polímeros biodegradables, biocompatibles, tales como acetato de etilvinilo, polianhídridos, ácido poliglicólico, colágeno, poliortoésteres y ácido poliláctico. Los métodos para la preparación de tales formulaciones serán evidentes para los expertos en la técnica. Los materiales también pueden obtenerse comercialmente de Alza Corporation y Nova Pharmaceuticals, Inc. También pueden utilizarse suspensiones de liposomas (incluyendo liposomas dirigidos para células infectadas con anticuerpos monoclonales frente a antígenos virales) como vehículos farmacéuticamente aceptables. Estos pueden prepararse según los métodos conocidos por los expertos en la técnica, por ejemplo, tal como se describe en la patente estadounidense número 4.522.811.

Resulta especialmente ventajoso formular las composiciones orales o parenterales en formas farmacéuticas unitarias para facilitar la administración y la uniformidad de la dosificación. La forma farmacéutica unitaria, tal como se usa en el presente documento, se refiere a unidades físicamente diferenciadas adecuadas como dosificaciones unitarias para el sujeto que va a tratarse; conteniendo cada unidad una cantidad predeterminada del principio activo calculada para producir el efecto terapéutico deseado en asociación con el vehículo farmacéutico requerido. La especificación para las formas farmacéuticas unitarias de la invención están dictadas por y dependen directamente de las características únicas del principio activo y del efecto terapéutico particular que va a lograrse, y de las limitaciones inherentes en la técnica de la formulación de un principio activo de este tipo para el tratamiento de los individuos.

La toxicidad y la eficacia terapéutica de los compuestos de este tipo pueden determinarse mediante procedimientos farmacéuticos habituales en cultivos celulares o animales de experimentación, por ejemplo, para determinar la DL50 (la dosis letal para el 50% de la población) y la DE50 (la dosis terapéuticamente eficaz en el 50% de la población). La razón de dosis entre los efectos tóxicos y los terapéuticos es el índice terapéutico y puede expresarse como la razón DL50/DE50. Se prefieren los compuestos que muestran índices terapéuticos elevados. Aunque pueden utilizarse compuestos que muestran efectos secundarios tóxicos, debe tenerse cuidado en diseñar un sistema de administración que dirija tales compuestos al sitio del tejido afectado con el fin de minimizar el posible daño a células no infectadas y, de esta manera, reducir los efectos secundarios.

Pueden utilizarse los datos obtenidos a partir de ensayos con cultivos celulares y estudios con animales para formular un intervalo de dosificación para su uso en seres humanos. La dosificación de tales compuestos se encuentra preferiblemente dentro de un intervalo de concentraciones circulantes que incluyen la DE50 con poca o ninguna toxicidad. La dosificación puede variar dentro de este intervalo, dependiendo de la forma farmacéutica empleada y de la vía de administración utilizada. Para cualquier compuesto usado en el método de la invención, la dosis terapéuticamente eficaz puede calcularse inicialmente a partir de los ensayos del cultivo celular. Puede formularse una dosis en modelos animales para lograr un intervalo de concentraciones plasmáticas circulantes que incluyen la CI50 (es decir, la concentración del compuesto de prueba que logra la mitad de la inhibición máxima de los síntomas) tal como se determina en el cultivo celular. Tal información puede utilizarse para determinar más exactamente las dosis útiles en humanos. Pueden medirse los niveles en plasma, por ejemplo, mediante cromatografía de líquidos de alta resolución.

Las moléculas de ácidos nucleicos de la invención pueden insertarse en vectores y utilizarse como vectores de terapia génica. Los vectores de terapia génica pueden administrarse a un sujeto mediante, por ejemplo, inyección intravenosa, administración local (véase la patente estadounidense 5.328.470) o mediante inyección esteotáctica (véase por ejemplo, Chen *et al.* (1994) PNAS 91:3054-3057). La preparación farmacéutica del vector de terapia génica puede incluir el vector de terapia génica en un diluyente aceptable, o puede comprender una matriz de liberación lenta en la que está incluido el vehículo de administración génica. Alternativamente, cuando puede producirse el vector de administración génica completo intacto a partir de células recombinantes, por ejemplo, los vectores retrovirales, la preparación farmacéutica puede incluir una o más células que producen el sistema de administración génica.

Las composiciones farmacéuticas pueden incluir un envase, paquete, o dispensador, junto con instrucciones para la administración.

V. Usos y métodos de la invención

Las moléculas de la presente invención (por ejemplo, moléculas de ácidos nucleicos, proteínas, homólogos de proteínas, y anticuerpos descritos en el presente documento) pueden utilizarse en uno o más de los siguientes métodos: a) ensayos de selección; b) medicina de diagnóstico (por ejemplo, ensayos de diagnósticos, ensayos de pronóstico, ensayos clínicos de monitorización, y ensayos farmacogenéticos); y c) métodos de tratamiento (por ejemplo, terapéuticos y profilácticos). Tal como se describe en el presente documento, una proteína CRSP de la invención tiene una o más de las siguientes actividades: calcio intracelular, un aumento en el fosfatidilinositol u otra molécula, y puede dar como resultado, por ejemplo, la fosforilación de proteínas específicas, una modulación de la transcripción génica y cualquiera de las otras actividades biológicas explicadas en el presente documento.

En una realización preferida, una actividad de CRSP es al menos una o más de las siguientes actividades: (i) interacción de una proteína CRSP con y/o unión a una segunda molécula, (por ejemplo, una proteína, tal como un receptor de CRSP (por ejemplo, CRSP-1), una forma soluble de un receptor de CRSP, un receptor para un miembro de la familia *wnt* de proteínas de señalización, o una molécula de señalización distinta de CRSP); (ii) interacción de una proteína CRSP con una proteína intracelular mediante un receptor de CRSP unido a la membrana; (iii) formación de complejo entre una proteína CRSP soluble y una segunda pareja de unión de CRSP soluble (por ejemplo, una molécula de proteína distinta de CRSP o una segunda molécula de proteína CRSP); (iv) interacción con otras proteínas extracelulares (por ejemplo, regulación de la adhesión celular dependiente de *wnt* a los componentes de la matriz extracelular); (v) unión a y eliminación de una molécula no deseada (por ejemplo, una función de actividad detoxificante o de defensa); y/o (vi) una actividad enzimática, y por tanto puede utilizarse en, por ejemplo, (1) modulación de transducción de señales celulares, ya sea *in vitro* o *in vivo* (por ejemplo, antagonismo de la actividad de los miembros de la familia *wnt* de proteínas segregadas o supresión de la transducción de señales dependientes de *wnt*); (2) regulación de la comunicación entre células (por ejemplo, regulación de interacciones célula-célula dependientes de *wnt*); (3) regulación de la expresión de genes cuya expresión está modulada mediante la unión de la CRSP (por ejemplo, CRSP-1) a un receptor; (4) regulación de la transcripción génica en una célula que participa en el desarrollo o la diferenciación, ya sea *in vitro* o *in vivo* (por ejemplo, inducción de diferenciación celular); (5) regulación de la transcripción génica en una célula

que participa en el desarrollo o la diferenciación, en la que al menos un gen codifica para una proteína específica de diferenciación; (6) regulación de la transcripción génica en una célula que participa en el desarrollo o la diferenciación, en la que al menos un gen codifica para una segunda proteína segregada; (7) regulación de la transcripción génica en una célula que participa en el desarrollo o la diferenciación, en la que al menos un gen codifica para una molécula de transducción de señales; (8) regulación de la proliferación celular, ya sea *in vitro* o *in vivo* (por ejemplo, inducción de la proliferación celular o inhibición de la proliferación como en el caso de la supresión de la tumorigénesis (por ejemplo, supresión de formación de glioblastoma)); (9) formación y mantenimiento de disposiciones espaciales ordenadas de tejidos diferenciados en vertebrados, tanto adultos como embrionarios (por ejemplo, inducción de la formación de la cabeza durante el desarrollo de vertebrados o mantenimiento de las células progenitoras hematopoyéticas); (10) modulación de la muerte celular, tal como estimulación de la supervivencia celular; (11) migración de las células de regulación; y/o (12) modulación inmunitaria.

En consecuencia, una realización de la presente invención supone un método de uso (por ejemplo, un ensayo diagnóstico, un ensayo pronóstico, o un método profiláctico/terapéutico de tratamiento) en el que se usa una molécula de la presente invención (por ejemplo, una proteína CRSP, un ácido nucleico de CRSP, o un modulador de CRSP), por ejemplo, para diagnosticar, pronosticar y/o tratar una enfermedad y/o estado en el que está indicada cualquiera de las actividades anteriormente mencionadas (es decir, actividades (i) - (vi) y (1) - (12) en el párrafo anterior). En otra realización, la presente invención supone un método de uso (por ejemplo, un ensayo diagnóstico, un ensayo pronóstico, o la fabricación de un medicamento para un método de tratamiento profiláctico/terapéutico) en el que se usa una molécula de la presente invención (por ejemplo, una proteína CRSP, ácido nucleico de CRSP, o un modulador de CRSP), por ejemplo, en la fabricación de un medicamento para el diagnóstico, pronóstico, y/o tratamiento de sujetos, preferiblemente de un sujeto humano, en el que está alterada patológicamente cualquiera de las actividades anteriormente mencionadas. En una realización preferida, los métodos de uso (por ejemplo, ensayos de diagnósticos, ensayos de pronóstico, o fabricación de medicamentos para métodos de tratamiento profilácticos/terapéuticos) suponen la producción de un medicamento para su administración a un sujeto, preferiblemente un sujeto humano, de una molécula de la presente invención (por ejemplo, una proteína CRSP, ácido nucleico de CRSP, o un modulador de CRSP) para el diagnóstico, pronóstico, y/o tratamiento terapéutico. En otra realización, los métodos de uso (por ejemplo, ensayos de diagnósticos, ensayos de pronóstico, o fabricación de medicamentos para métodos de tratamiento profilácticos/terapéuticos tratamiento) suponen la producción de un medicamento para su administración a un sujeto humano de una molécula de la presente invención (por ejemplo, una proteína CRSP, ácido nucleico de CRSP, o un modulador de CRSP).

Otras realizaciones de la invención están relacionadas con el uso de moléculas de ácidos nucleicos aisladas de la invención que pueden usarse, por ejemplo, para expresar la proteína CRSP (por ejemplo, mediante un vector de expresión recombinante en una célula huésped en aplicaciones de terapia génica), para detectar ARNm de CRSP (por ejemplo, en una muestra biológica) o una alteración genética en un gen de CRSP, y para modular la actividad de CRSP, tal como se describe adicionalmente más adelante. Además, las proteínas CRSP pueden usarse como fármacos o compuestos de selección que modulan la actividad de CRSP, así como para tratar trastornos caracterizados por producción insuficiente o excesiva de proteína CRSP o producción de formas de proteína CRSP que tienen actividad disminuida o aberrante en comparación con la proteína CRSP de tipo natural (por ejemplo, trastornos del desarrollo, enfermedades proliferativas tales como cáncer así como enfermedades, estados o trastornos caracterizados por la diferenciación y/o supervivencia celular anómala, una estructura extracelular anómala, o una anomalía en un mecanismo de defensa). Además, los anticuerpos anti-CRSP de la invención pueden utilizarse para detectar y aislar proteínas CRSP, regular la biodisponibilidad de las proteínas CRSP, y modular la actividad de CRSP. La expresión “una actividad aberrante”, tal como se aplica a una actividad de una proteína tal como CRSP (por ejemplo, CRSP-1), se refiere a una actividad que difiere de la actividad de la proteína natural o nativa o que difiere de la actividad de la proteína en un sujeto sano. Una actividad de una proteína puede ser aberrante debido a que es más fuerte que la actividad de su homóloga nativa. Alternativamente, una actividad puede ser aberrante debido a que es más débil o a que está ausente con respecto a la actividad de su homóloga nativa. Una actividad aberrante también puede ser un cambio en una actividad. Por ejemplo una proteína aberrante puede interactuar con una proteína diferente con respecto a su homóloga nativa. Una célula puede tener una actividad de CRSP aberrante (por ejemplo, CRSP-1) debido a la expresión superior o inferior a lo normal del gen que codifica para CRSP.

A. Ensayos de selección

La invención proporciona un método (también denominado en el presente documento un “ensayo de selección”) para identificar moduladores, es decir, compuestos o agentes de prueba o candidatos (por ejemplo, péptidos, peptidomiméticos, pequeñas moléculas u otros fármacos) que se unen a las proteínas CRSP o tienen un efecto estimulante o inhibidor sobre, por ejemplo, la expresión de CRSP o la actividad de CRSP. Los moduladores pueden incluir, por ejemplo, agonistas de CRSP y/o antagonistas de CRSP. El término “agonista”, tal como se usa en el presente documento, pretende referirse a un agente que imita o regula por incremento (por ejemplo potencia o complementa) una bioactividad de CRSP (por ejemplo, CRSP-1). Un agonista de CRSP puede ser un compuesto que imita a una bioactividad de una proteína CRSP, tal como la transducción de una señal desde un receptor de CRSP, mediante por ejemplo, la interacción con un receptor de CRSP-1. Un agonista de CRSP también puede ser un compuesto que regula por incremento la expresión de un gen de CRSP. Un agonista de CRSP también puede ser un compuesto que modula la expresión o actividad de una proteína que está situada posteriormente a un receptor de CRSP, imitando o potenciando así el efecto de la unión de CRSP a un receptor de CRSP.

“Antagonista” tal como se usa en el presente documento, pretende referirse a un agente que inhibe, disminuye o suprime una bioactividad de CRSP (por ejemplo, CRSP-1). Un antagonista puede ser un compuesto que disminuye la señalización de una proteína CRSP, por ejemplo, un compuesto que puede producir la unión a CRSP-1 o a un receptor de CRSP-1. Un antagonista de CRSP preferido inhibe la interacción entre una proteína CRSP y otra molécula, tal como un receptor de CRSP. Alternativamente, un antagonista de CRSP puede ser un compuesto que regula por disminución la expresión de un gen de CRSP. Un antagonista de CRSP también puede ser un compuesto que modula la expresión o la actividad de una proteína que está situada posteriormente a un receptor de CRSP, antagonizando así con el efecto de unión de CRSP a un receptor de CRSP.

En una realización, la invención proporciona ensayos para seleccionar compuestos de prueba o candidatos que se unen a o que modulan la actividad de una proteína CRSP o polipéptido o parte biológicamente activa de los mismos. En otra realización, la invención proporciona ensayos para seleccionar compuestos de prueba o candidatos que se unen a o modulan la actividad de un receptor de CRSP. Los compuestos de prueba de la presente invención pueden obtenerse utilizando cualquiera de los numerosos enfoques en los métodos de biblioteca combinatoria conocidos en la técnica, incluyendo: bibliotecas biológicas; bibliotecas en fase de disolución o fase sólida paralelas espacialmente dirigibles; métodos de biblioteca sintética que requieren deconvolución; el método de biblioteca de “una perla un compuesto” (“one-bead one-compound”); y métodos de biblioteca sintéticos utilizando selección mediante cromatografía de afinidad. El enfoque de la biblioteca biológica se limita a las bibliotecas de péptidos, mientras que los otros cuatro enfoques son aplicables a las bibliotecas de compuestos de péptidos, oligómeros no peptídicos o pequeñas moléculas (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Pueden encontrarse en la técnica ejemplos de métodos para la síntesis de bibliotecas moleculares, por ejemplo en: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; y en Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Las bibliotecas de compuestos pueden presentarse en disolución (por ejemplo, Houghten (1992) *Biotechniques* 13:412-421), o en perlas (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacterias (patente estadounidense 5.223.409 de Ladner), esporas (patente estadounidense 5.409 de Ladner), plásmidos (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) o en fagos (Scott y Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner citado anteriormente).

En una realización, un ensayo es un ensayo basado en células en el que una célula que expresa un receptor de CRSP en la superficie celular está en contacto con un compuesto de prueba y se determina la capacidad del compuesto de prueba para unirse a un receptor de CRSP. La célula, por ejemplo, puede ser de origen mamífero o una célula de levadura. La determinación de la capacidad del compuesto de prueba para unirse a un receptor de CRSP puede llevarse a cabo, por ejemplo, mediante el acoplamiento del compuesto de prueba con un radioisótopo o un marcador enzimático, de manera que la unión del compuesto de prueba al receptor de CRSP puede determinarse detectando el compuesto marcado en un complejo. Por ejemplo, los compuestos de prueba pueden marcarse con ^{125}I , ^{35}S , ^{14}C , o ^3H , ya sea directa o indirectamente, y el radioisótopo puede detectarse mediante el recuento directo de radioemisión o mediante recuento por centelleo. Alternativamente, los compuestos de prueba pueden marcarse enzimáticamente con, por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina o luciferasa, y el marcador enzimático puede detectarse mediante la determinación de la conversión de un sustrato apropiado en producto.

También está dentro del alcance de esta invención determinar la capacidad de un compuesto de prueba para interaccionar con un receptor de CRSP sin marcar ninguno de los compuestos que interaccionan. Por ejemplo, puede utilizarse un microfisiómetro para detectar la interacción de un compuesto de prueba con un receptor de CRSP sin el marcaje del compuesto de prueba o el receptor. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. Tal como se usa en el presente documento, un “microfisiómetro” (por ejemplo, CytosensorTM) es un instrumento analítico que mide la velocidad a la que una célula acidifica su entorno utilizando un detector potenciométrico dirigible mediante luz (LAPS). Los cambios en esta velocidad de acidificación pueden utilizarse como indicadores de la interacción entre el ligando y el receptor.

En una realización preferida, el ensayo comprende poner en contacto una célula que expresa un receptor de CRSP en la superficie celular con una proteína CRSP o una parte biológicamente activa de la misma, para formar una mezcla de ensayo, poner en contacto la mezcla de ensayo con un compuesto de prueba, y determinar la capacidad del compuesto de prueba para interaccionar con un receptor de CRSP, en la que la determinación de la capacidad del compuesto de prueba para interaccionar con un receptor de CRSP comprende determinar la capacidad del compuesto de prueba unirse preferentemente al receptor de CRSP en comparación con la capacidad de CRSP, o de una parte biológicamente activa de la misma, para unirse al receptor.

En otra realización, un ensayo es un ensayo basado en células que comprende poner en contacto una célula que expresa una molécula diana de CRSP con un compuesto de prueba y determinar la capacidad del compuesto de prueba para modular (por ejemplo estimular o inhibir) la actividad de la molécula diana de CRSP. La determinación de la capacidad del compuesto de prueba para modular la actividad de una molécula diana de CRSP puede llevarse a cabo,

por ejemplo, mediante la determinación de la capacidad de la proteína CRSP para unirse a o para interactuar con la molécula diana de CRSP.

La determinación de la capacidad de la proteína CRSP para unirse o interactuar con una molécula diana de CRSP puede llevarse a cabo mediante uno de los métodos descritos anteriormente para determinar la unión directa. En una realización preferida, la determinación de la capacidad de la proteína CRSP para unirse o interactuar con una molécula diana de CRSP puede llevarse a cabo mediante la determinación de la actividad de la molécula diana. Por ejemplo, la actividad de la molécula diana puede determinarse mediante la detección de la inducción de un segundo mensajero celular de la molécula diana (es decir Ca^{2+} intracelular, diacilglicerol, IP_3 , etc.), detección de la actividad catalítica/enzimática de la molécula diana de un sustrato apropiado, detección de la inducción de un gen indicador (que comprende un elemento regulador sensible a CRSP unido operativamente a un ácido nucleico que codifica para un marcador detectable, por ejemplo, luciferasa), o detección de una respuesta celular, por ejemplo, desarrollo, diferenciación o velocidad de proliferación.

Todavía en otra realización, un ensayo de la presente invención es un ensayo libre de células en el que una proteína CRSP o parte biológicamente activa de la misma se pone en contacto con un compuesto de prueba y se determina la capacidad del compuesto de prueba para unirse a la proteína CRSP o parte biológicamente activa de la misma. La unión del compuesto de prueba a la proteína CRSP puede determinarse directa o indirectamente tal como se describió anteriormente. En una realización preferida, el ensayo incluye poner en contacto la proteína CRSP o parte biológicamente activa de la misma con un compuesto conocido que se une a CRSP para formar una mezcla de ensayo, poner en contacto el mezcla de ensayo con un compuesto de prueba, y determinar la capacidad del compuesto de prueba para interactuar con una proteína CRSP, en la que la determinación de la capacidad del compuesto de prueba para interactuar con una proteína CRSP comprende determinar la capacidad del compuesto de prueba para unirse preferentemente a la CRSP o parte biológicamente activa de la misma as en comparación con el compuesto conocido.

En otra realización, el ensayo es un ensayo libre de células en el que una proteína CRSP o parte biológicamente activa de la misma se pone en contacto con un compuesto de prueba y se determina la capacidad del compuesto de prueba para modular (por ejemplo, estimular o inhibir) la actividad de la proteína CRSP o parte biológicamente activa de la misma. La determinación de la capacidad del compuesto de prueba para modular la actividad de una proteína CRSP puede llevarse a cabo, por ejemplo, mediante la determinación de la capacidad de la proteína CRSP para unirse a una molécula diana de CRSP mediante uno de los métodos descritos anteriormente para determinar la unión directa. La determinación de la capacidad de la proteína CRSP para unirse a una molécula diana de CRSP también puede llevarse a cabo utilizando una tecnología tal como Análisis de Interacción Biomolecular (BIA) en tiempo real. Sjolander, S. y Urbaniczky, C. (1991) Anal. Chem. 63:2338-2345 y Szabo *et al.* (1995) Curr. Opin. Struct. Biol. 5:699-705. Tal como se usa en el presente documento, "BIA" es una tecnología para el estudio de las interacciones bioespecíficas en tiempo real, sin marcar ninguno de los compuestos que interactúan (por ejemplo, BIAcore™). Los cambios en el fenómeno óptico de resonancia de plasmón superficial (SPR) pueden usarse como indicación de las reacciones en tiempo real entre las moléculas biológicas.

En una realización alternativa, la determinación de la capacidad del compuesto de prueba para modular la actividad de una proteína CRSP puede llevarse a cabo mediante la determinación de la capacidad de la proteína CRSP para modular adicionalmente la actividad de una molécula diana de CRSP. Por ejemplo, puede determinarse la actividad catalítica/enzimática de la molécula diana en un sustrato apropiado, tal como se describió anteriormente.

Todavía en otra realización, el ensayo libre de células supone poner en contacto una proteína CRSP o parte biológicamente activa de la misma con un compuesto conocido que se une a la proteína CRSP para formar una mezcla de ensayo, poner en contacto el mezcla de ensayo con un compuesto de prueba, y determinar la capacidad del compuesto de prueba para interactuar con la proteína CRSP, en la que la determinación de la capacidad del compuesto de prueba para interactuar con la proteína CRSP comprende determinar la capacidad de la proteína CRSP para unirse preferentemente a o para modular la actividad de una molécula diana de CRSP.

En muchos programas de selección de fármacos que prueban las bibliotecas de compuestos y extractos naturales, son deseable ensayos de alto rendimiento con el fin de maximizar el número de compuestos examinados en un periodo de tiempo dado. A menudo se prefieren los ensayos que se realizan en sistemas libres de células, tales como los que pueden derivarse de proteínas purificadas o semipurificadas, como selecciones "primarias" porque pueden generarse para permitir el rápido desarrollo y la detección relativamente fácil de una alteración en una diana molecular que está mediada por un compuesto de prueba. Además, los efectos de la toxicidad y/o biodisponibilidad celular del compuesto de prueba pueden ignorarse generalmente en el sistema *in vitro*, enfocándose por el contrario el ensayo principalmente en el efecto del fármaco sobre la diana molecular, ya que puede manifestarse en una alteración en la afinidad de unión con elementos aguas arriba o aguas abajo. En consecuencia, en un ensayo de selección ejemplo de la presente invención, el compuesto de interés se pone en contacto con una proteína CRSP (por ejemplo, CRSP-1) o una pareja de unión de CRSP (por ejemplo, CRSP-1), por ejemplo, un receptor. El receptor puede ser soluble o el receptor puede estar presente en una superficie celular. Entonces se añade a la mezcla del compuesto y a la proteína CRSP o pareja de unión de CRSP una composición que contiene una pareja de unión de CRSP o una proteína CRSP, respectivamente. La detección y cuantificación de los complejos de proteínas CRSP y Parejas de unión de CRSP proporcionan un medio para determinar la eficacia de un compuesto para inhibir (o potencia) la formación del complejo entre CRSP y una pareja de unión. La eficacia del compuesto puede evaluarse generando curvas de respuesta a la dosis a partir de los

datos obtenidos utilizando diversas concentraciones del compuesto de prueba. Además, también puede realizarse un ensayo de control para proporcionar una línea base para comparación. En el ensayo de control, se añade el polipéptido o pareja de unión de CRSP aislado y purificado a una composición que contiene la CRSP pareja de unión o el polipéptido CRSP, y se cuantifica la formación de un complejo en ausencia del compuesto de prueba.

Los ensayos libres de células de la presente invención son susceptibles de usarse con formas solubles y/o unidas a la membrana de proteínas aisladas (por ejemplo proteínas CRSP o partes biológicamente activas de las mismas o moléculas diana de CRSP). En el caso de los ensayos libres de células en los que se usa una forma unida a la membrana de proteína aislada (por ejemplo, una molécula diana o receptor de CRSP) puede ser deseable utilizar un agente solubilizante de manera que la forma unida a la membrana de la proteína aislada se mantenga en disolución. Ejemplos de agentes solubilizantes de este tipo incluyen detergentes no iónicos tales como n-octilglucósido, n-dodecilglucósido, n-dodecilmaltósido, octanoil-N-metilglucamida, decanoil-N-metilglucamida, Triton® X-100, Triton® X-114, Thesit®, isotridecilo(etilenglicol éter)_n, sulfonato de 3-[(3-colamidopropil)dimetilaminio]-1-propano (CHAPS), sulfonato de 3-[(3-colamidopropil)dimetilaminio]-2-hidroxi-1-propano (CHAPSO), o N-dodecil-N, sulfonato de N-dimetil-3-amonio-1-propano.

En más de una realización de los métodos de ensayo anteriores de la presente invención, puede ser deseable inmovilizar o bien CRSP o bien su molécula diana para facilitar la separación de las formas complejadas de las no complejadas de una o ambas proteínas, así como adaptar la automatización del ensayo. La unión de un compuesto de prueba a una proteína CRSP, o la interacción de una proteína CRSP con una molécula diana en presencia y ausencia de un compuesto candidato, puede llevarse a cabo en cualquier recipiente adecuado para contener los reactivos. Ejemplos de tales recipientes incluyen placas de microtitulación, tubos de ensayo y tubos de microcentrífuga. En una realización, puede proporcionarse una proteína de fusión que añade un dominio que permite que una o ambas proteínas se unan a una matriz. Por ejemplo, las proteínas de fusión glutatión-S-transferasa/CRSP o las proteínas de fusión glutatión-S-transferasa/proteína diana pueden adsorberse en perlas de glutatión-Sepharosa (Sigma Chemical, St. Louis, MO) o placas de microtitulación derivatizadas con glutatión, que entonces se combinan con el compuesto de prueba o el compuesto de prueba y o bien la proteína diana no adsorbida o bien la proteína CRSP, y la mezcla se incuba en estados conductores para la formación de complejos (por ejemplo, en estados fisiológicos para sal y pH). Tras la incubación, las perlas o pocillos de las placas de microtitulación se lavan para eliminar cualquier componente no unido, la matriz se inmoviliza en el caso de las perlas, el complejo se determina directa o indirectamente, por ejemplo, tal como se describió anteriormente. Alternativamente, los complejos pueden disociarse de la matriz y puede determinarse el nivel de unión o actividad de la CRSP utilizando técnicas habituales.

En los ensayos de selección de la invención también pueden utilizarse otras técnicas para inmovilizar proteínas en matrices. Por ejemplo, puede inmovilizarse o bien una proteína CRSP o bien una molécula diana de CRSP utilizando conjugación de biotina y estreptavidina. La proteína CRSP o las moléculas diana biotiniladas pueden prepararse a partir de biotina-NHS (N-hidroxi-succinimida) usando técnicas bien conocidas en la técnica (por ejemplo, kit de biotinilización, Pierce Chemicals, Rockford, IL), e inmovilizarse en los pocillos de placas de 96 pocillos recubiertos con estreptavidina (Pierce Chemical). Alternativamente, pueden derivatizarse anticuerpos reactivos con proteína CRSP o moléculas diana pero que no interfieren con la unión de la proteína CRSP a su molécula diana, a los pocillos de la placa, y a la molécula diana no unida o a la proteína CRSP atrapada en los pocillos mediante conjugación de anticuerpos. Los métodos para detectar tales complejos, además de los descritos anteriormente para los complejos inmovilizados a GST, incluyen inmunodetección de complejos utilizando anticuerpos reactivos con la proteína CRSP o la molécula diana, así como ensayos ligados a enzimas que se basan en la detección de una actividad enzimática asociada con la proteína CRSP o la molécula diana.

En otra realización, se identifican los moduladores de la expresión de CRSP en un método en el que una célula se pone en contacto con un compuesto candidato y se determina la expresión del ARNm de CRSP o la proteína CRSP en la célula. El nivel de expresión del ARNm de CRSP o la proteína CRSP en presencia del compuesto candidato se compara con el nivel de expresión de del ARNm de CRSP o la proteína CRSP en ausencia del compuesto candidato. El compuesto candidato puede identificarse entonces como un modulador de la expresión de CRSP basándose en esta comparación. Por ejemplo, cuando la expresión del ARNm de CRSP o la proteína CRSP es mayor (mayor de manera estadísticamente significativa) en presencia del compuesto candidato que en su ausencia, el compuesto candidato se identifica como un estimulador de la expresión del ARNm de CRSP o la proteína CRSP. Alternativamente, cuando la expresión del ARNm de CRSP o la proteína CRSP es menor (menor de manera estadísticamente significativa) en presencia del compuesto candidato que en su ausencia, el compuesto candidato se identifica como un inhibidor de la expresión del ARNm de CRSP o la proteína CRSP. El nivel de expresión del ARNm de CRSP o la proteína CRSP en las células puede determinarse mediante los métodos descritos en el presente documento para detectar del ARNm de CRSP o la proteína CRSP.

Todavía en otro aspecto de la invención, las proteínas CRSP pueden usarse como “proteínas cebo” en un ensayo de dos híbridos o un ensayo de tres híbridos (véase, por ejemplo, la patente estadounidense número 5.283.317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J. Biol. Chem. 268:12046-12054; Bartel *et al.* (1993) Biotechniques 14:920-924; Iwabuchi *et al.* (1993) Oncogene 8: 1693-1696; y documento WO94/10300 de Brent), para identificar otras proteínas, que se unen a o interaccionan con CRSP (“proteínas de unión a CRSP” o “CRSP-pb”) y modulan la actividad de CRSP. Tales proteínas de unión a CRSP también es probable que participen en la propagación de las señales producidas por las proteínas CRSP as, por ejemplo, a elementos aguas abajo de una ruta de señalización mediada por CRSP. Alternativamente, es probable que tales proteínas de unión a CRSP sean moléculas de la superficie

celular asociadas con células que no expresan CRSP, en las que tales proteínas de unión a CRSP participan en la transducción de señales.

El sistema de dos híbridos se basa en la naturaleza modular de la mayor parte de los factores de transcripción, que consiste en dominios de activación y unión al ADN que pueden separarse. En resumen, el ensayo utiliza dos constructos de ADN diferentes. En un constructo, el gen que codifica para una proteína CRSP se fusiona con un gen que codifica para el dominio de unión al ADN de un factor de transcripción conocido (por ejemplo, GAL-4). En el otro constructo, una secuencia de ADN, procedente de una biblioteca de secuencias de ADN, que codifica para una proteína no identificada ("presa" o "muestra") se fusiona con un gen que codifica para el dominio de activación del factor de transcripción conocido. Si las proteínas "cebo" y "presa" pueden interactuar, *in vivo*, formando un complejo dependiente de CRSP, los dominios de activación y unión al ADN del factor de transcripción se transportan en proximidad cercana. Esta proximidad permite la transcripción de un gen indicador (por ejemplo, LacZ) que está operativamente unido a un sitio de regulación de la transcripción que responde al factor de transcripción. Puede detectarse la expresión del gen indicador y pueden aislarse las colonias de células que contienen el factor de transcripción funcional y utilizarse para obtener el gen clonado que codifica para la proteína que interactúa con la proteína CRSP.

Esta invención se refiere además a agentes novedosos que identificados mediante los ensayos de selección descritos anteriormente y a procedimientos para producir tales agentes mediante el uso de estos ensayos. Por consiguiente, en una realización, la presente invención incluye un compuesto o agente que puede obtenerse mediante un método que comprende las etapas de uno cualquiera de los ensayos de selección mencionados anteriormente (por ejemplo, ensayos basados en célula o ensayos libres de células). Por ejemplo, en una realización, la invención incluye un compuesto o agente que puede obtenerse mediante un método que comprende poner en contacto una célula que expresa una molécula diana de CRSP con un compuesto de prueba y determinar la capacidad del compuesto de prueba para unirse a, o modular la actividad de, la molécula diana de CRSP. En otra realización, la invención incluye un compuesto o agente que puede obtenerse mediante un método que comprende poner en contacto una célula que expresa una molécula diana de CRSP con una proteína CRSP o una parte biológicamente activa de la misma, para formar una mezcla de ensayo, poner en contacto la mezcla de ensayo con un compuesto de prueba y determinar la capacidad del compuesto de prueba para interactuar con, o modular la actividad de, la molécula diana de CRSP. En otra realización, la invención incluye un compuesto o agente que puede obtenerse mediante un método que comprende poner en contacto una proteína CRSP o una parte biológicamente activa de la misma con un compuesto de prueba y determinar la capacidad del compuesto de prueba para unirse a, o modular (por ejemplo, estimular o inhibir) la actividad de, la proteína CRSP o la parte biológicamente activa de la misma. En todavía otra realización, la presente invención incluye un compuesto o agente que puede obtenerse mediante un método que comprende poner en contacto una proteína CRSP o una parte biológicamente activa de la misma con un compuesto conocido que se une a la proteína CRSP para formar una mezcla de ensayo, poner en contacto la mezcla de ensayo con un compuesto de prueba, y determinar la capacidad del compuesto de prueba para interactuar con, o modular la actividad de la proteína CRSP.

Por consiguiente, está dentro del alcance de esta invención el uso adicional de un agente identificado tal como se describe en el presente documento en un modelo animal apropiado. Por ejemplo, puede usarse un agente identificado tal como se describe en el presente documento (por ejemplo, un agente modulador de CRSP, una molécula de ácido nucleico de CRSP antisentido, un anticuerpo específico para CRSP, o una pareja de unión a CRSP) en un modelo animal para determinar la eficacia, toxicidad o efectos secundarios del tratamiento con tal agente. Alternativamente, un agente identificado tal como se describe en el presente documento puede usarse en un modelo animal para determinar el mecanismo de acción de tal agente.

La presente invención también se refiere a los usos de agentes novedosos identificados mediante los ensayos de selección descritos anteriormente para diagnósticos, pronósticos y tratamientos, tal como se describe en el presente documento. Por consiguiente, está dentro del alcance de la presente invención el uso de tales agentes en el diseño, formulación, síntesis, fabricación, y/o producción de un fármaco o composición farmacéutica para usarse en el diagnóstico, pronóstico o tratamiento, tal como se describe en el presente documento. Por ejemplo, en una realización, la presente invención incluye un método para sintetizar o producir un fármaco o composición farmacéutica haciendo referencia a la estructura y/o las propiedades de un compuesto que puede obtenerse mediante uno de los ensayos de selección descritos anteriormente. Por ejemplo, un fármaco o composición farmacéutica puede sintetizarse basándose en la estructura y/o propiedades de un compuesto obtenido mediante un método en el que una célula que expresa una molécula diana de CRSP se pone en contacto con un compuesto de prueba y se determina la capacidad del compuesto de prueba para unirse a, o modular la actividad de, la molécula diana de CRSP. En otra realización a modo de ejemplo, la presente invención incluye un método para sintetizar o producir un fármaco o una composición farmacéutica basándose en la estructura y/o las propiedades de un compuesto que puede obtenerse mediante un método en el que una proteína CRSP o una parte biológicamente activa de la misma se pone en contacto con un compuesto de prueba y se determina la capacidad del compuesto de prueba para unirse a, o modular (por ejemplo, estimular o inhibir) la actividad de, la proteína CRSP o la parte biológicamente activa de la misma.

B. Ensayos de detección

Partes o fragmentos de las secuencias de ADNc identificadas en el presente documento (y las correspondientes secuencias génicas completas) pueden usarse de numerosas maneras como reactivos polinucleotídicos. Por ejemplo, estas secuencias pueden usarse para: (i) mapear sus respectivos genes en un cromosoma; y, por tanto, localizar regiones génicas asociadas con una enfermedad genética; (ii) identificar a un individuo de una muestra biológica mínima

(tipificación de tejido); y (iii) ayuda en identificación forense de una muestra biológica. Estas aplicaciones se describen en las subsecciones a continuación.

1. Mapeo de cromosomas

Una vez se ha aislado la secuencia (o una parte de la secuencia) de un gen, puede usarse esta secuencia para mapear la ubicación del gen en un cromosoma. Este procedimiento se denomina mapeo cromosómico. Por consiguiente, partes o fragmentos de la secuencia de nucleótidos de CRSP, descritos en el presente documento, pueden usarse para mapear la ubicación de los genes de CRSP en un cromosoma. El mapeo de las secuencias de CRSP en los cromosomas es una primera etapa importante para correlacionar estas secuencias con genes asociados con enfermedad.

En resumen, los genes de CRSP pueden mapearse en cromosomas para preparar cebadores de PCR (preferiblemente de 15-25 pb de longitud) a partir de la secuencia de nucleótidos de CRSP. Puede usarse un análisis informático de las secuencias de CRSP para predecir cebadores que no se expanden más de un exón en el ADN genómico, complicando así el procedimiento de amplificación. Después pueden usarse estos cebadores para examen por PCR de híbridos de células somáticas que contienen cromosomas humanos individuales. Sólo los híbridos que contienen el gen humano correspondiente a las secuencias de CRSP darán como resultado un fragmento amplificado.

Se preparan los híbridos de células somáticas mediante la fusión de células somáticas de distintos mamíferos (por ejemplo, células humanas y de ratón). A medida que crecen y se dividen los híbridos de células humanas y de ratón, pierden gradualmente cromosomas humanos de manera aleatoria, pero conservan los cromosomas de ratón. Mediante el uso de medios en los que las células de ratón no pueden crecer, porque carecen de una enzima particular, pero las células humanas sí pueden, se conservará el cromosoma que contiene el gen que codifica para la enzima necesaria. Mediante el uso de diversos medios, pueden establecerse paneles de líneas celulares híbridas. Cada línea celular en un panel contiene o bien un cromosoma humano sencillo o bien un pequeño número de cromosomas humanos, y un conjunto completo de cromosomas de ratón, permitiendo un fácil mapeo de genes individuales para cromosomas humanos específicos. (D'Eustachio P. *et al.* (1983) *Science* 220: 919-924). También pueden producirse híbridos de células somáticas que contienen sólo fragmentos de cromosomas humanos usando cromosomas humanos con translocaciones y deleciones.

El mapeo por PCR de híbridos de células somáticas es un procedimiento rápido para asignar una secuencia particular a un cromosoma particular. Pueden asignarse tres o más secuencias por día usando un ciclador térmico sencillo. Usando la secuencia de nucleótidos de CRSP para diseñar cebadores oligonucleotídicos, puede lograrse la sublocalización con paneles de fragmentos de cromosomas específicos. Otras estrategias de mapeo que pueden usarse de manera similar para mapear una secuencia 9o, 1p, o 1v para su cromosoma incluyen hibridación *in situ* (descrita en Fan, Y. *et al.* (1990) *PNAS*, 87:6223-27), preselección con cromosomas separados por citometría de flujo marcados y preselección mediante hibridación con bibliotecas de ADNc específicas de cromosomas.

La hibridación *in situ* fluorescente (FISH) de una secuencia de ADN con una metafase cromosómica extendida puede usarse adicionalmente para proporcionar una localización cromosómica precisa en una etapa. Las extensiones de cromosomas pueden prepararse usando células cuya división se ha bloqueado en metafase mediante un producto químico tal como colcemid que altera el huso acromático mitótico. Los cromosomas pueden tratarse brevemente con tripsina, y luego teñirse con Giemsa. En cada cromosoma se revela un patrón de bandas claras y oscuras, de manera que los cromosomas pueden identificarse individualmente. La técnica FISH puede usarse con una secuencia de ADN de tan solo 500 ó 600 bases. Sin embargo, los clones superiores a 1.000 bases tienen mayor probabilidad de unirse a una única ubicación cromosómica con suficiente intensidad de señal para la detección simple. Preferiblemente 1.000 bases, y más preferiblemente 2.000 bases serán suficientes para obtener buenos resultados en una cantidad razonable de tiempo. Para una revisión de esta técnica, véase Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, Nueva York 1988).

Pueden usarse reactivos para el mapeo de cromosomas individualmente para marcar un único cromosoma o un sitio único en ese cromosoma, o pueden usarse paneles de reactivos para marcar sitios múltiples y/o cromosomas múltiples. De hecho, se prefieren los reactivos correspondientes a regiones no codificantes de los genes con fines de mapeo. Es más probable que las secuencias codificantes se conserven dentro de las familias de genes, aumentando así la posibilidad de hibridaciones cruzadas durante el mapeo cromosómico.

Una vez se ha mapeado una secuencia en una ubicación cromosómica precisa, la posición física de la secuencia en el cromosoma puede correlacionarse con los datos del mapa genético. (Tales datos se encuentran, por ejemplo, en V. McKusick, *Mendelian Inheritance in Man*, disponible en línea a través de la biblioteca médica Welch de la Universidad Johns Hopkins (Johns Hopkins University Welch Medical Library)). La relación entre un gen y una enfermedad, mapeada en la misma región cromosómica, puede identificarse entonces a través del análisis de unión (co-herencia de genes físicamente adyacentes), descrito en, por ejemplo, Egeland, J. *et al.* (1987) *Nature*, 325: 783-787.

Además, pueden determinarse las diferencias en las secuencias de ADN entre individuos afectados y no afectados con una enfermedad asociada al gen de CRSP. Si se observa una mutación en alguno o todos los individuos afectados pero no en cualquier individuo no afectado, entonces es probable que la mutación sea el agente causante de la enfermedad particular. La comparación de individuos afectados y no afectados generalmente implica mirar en pri-

mer lugar para detectar alteraciones estructurales en los cromosomas, tales como deleciones, translocaciones que son visibles desde las extensiones de cromosomas o detectables usando PCR basándose en esa secuencia de ADN. Por último, puede realizarse la secuenciación completa de los genes de varios individuos para confirmar la presencia de una mutación para distinguir mutaciones de polimorfismos.

2. Tipificación tisular

También pueden usarse las secuencias de CRSP de la presente invención para identificar individuos de muestras biológicas mínimas. El ejército de los Estados Unidos, por ejemplo, está considerando el uso de polimorfismos en la longitud de los fragmentos de restricción (RFLP) para la identificación de su personal. En esta técnica, se digiere el ADN genómico de un individuo con una o más enzimas de restricción, y se examinan con sondas en una transferencia tipo Southern para dar lugar a bandas únicas para la identificación. Este método no adolece de las limitaciones actuales de las "Dog Tags" (placas identificativas del ejército de los EE.UU.) que pueden perderse, cambiarse, o robarse, dificultando la identificación positiva. Las secuencias de la presente invención son útiles como marcadores adicionales de ADN para RFLP (descritas en la patente estadounidense número 5.272.057).

Además, las secuencias de la presente invención pueden usarse para proporcionar una técnica alternativa que determina la secuencia de ADN real base por base de partes seleccionadas del genoma de un individuo. Por tanto, la secuencia de nucleótidos de CRSP descrita en el presente documento puede usarse para preparar dos cebadores de PCR a partir de los extremos 5' y 3' de las secuencias. Estos cebadores pueden usarse entonces para amplificar el ADN de un individuo y posteriormente secuenciarlo.

Los paneles de secuencias de ADN correspondientes de individuos, preparados de esta manera, pueden proporcionar identificaciones individuales únicas, dado que cada individuo tendrá un conjunto único de tales secuencias de ADN debido a diferencias alélicas. Las secuencias de la presente invención pueden usarse para obtener tales secuencias de identificación a partir de individuos y de tejidos. La secuencia de nucleótidos de CRSP de la invención únicamente representa partes del genoma humano. La variación alélica se produce hasta cierto grado en las regiones codificantes de estas secuencias, y en mayor grado en las regiones no codificantes. Se estima que la variación alélica entre individuos humanos se produce con una frecuencia de aproximadamente una vez por cada 500 bases. Cada una de las secuencias descritas en el presente documento puede usarse, hasta cierto grado, como un patrón frente al que puede compararse el ADN de un individuo con fines identificativos. Dado que se producen mayores números de polimorfismos en las regiones no codificantes, se necesitan menos secuencias para diferenciar individuos. Las secuencias no codificantes de SEQ ID NO: 1, SEQ ID NO: 4, SEQ ID NO: 7, o SEQ ID NO: 10, pueden proporcionar cómodamente identificación de individuo positiva con un panel de quizás 10 a 1.000 cebadores, cada uno de los cuales da como resultado una secuencia amplificada no codificante de 100 bases. Si se usan las secuencias codificantes previstas, tales como las de SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 9, o SEQ ID NO: 12, un número más apropiado de cebadores para la identificación de individuo positiva será de 500-2.000.

Si se usa un panel de reactivos de una secuencia de nucleótidos de CRSP descritos en el presente documento para generar una base de datos de identificación única, esos mismos reactivos pueden usarse más adelante para identificar el tejido de ese individuo. Usando la base de datos de identificación única, puede realizarse identificación positiva del individuo, vivo o muerto, a partir de muestras de tejido extremadamente pequeñas.

3. Uso de secuencias de CRSP en biología forense

También pueden usarse técnicas de identificación basadas en ADN en biología forense. La biología forense es un campo científico que emplea tipificación genética de evidencias biológicas encontradas en la escena de un crimen como medio para identificar positivamente, por ejemplo, al autor principal de un crimen. Para realizar tal identificación, puede usarse tecnología de PCR para amplificar secuencias de ADN tomadas a partir de muestras biológicas muy pequeñas tales como tejidos, por ejemplo, cabello o piel, o fluidos corporales, por ejemplo, sangre, saliva, o semen encontrados en la escena de un crimen. La secuencia amplificada puede compararse entonces con un patrón, permitiendo así la identificación del origen de la muestra biológica.

Las secuencias de la presente invención pueden usarse para proporcionar reactivos polinucleotídicos, por ejemplo, cebadores de PCR, dirigidos hacia loci específicos en el genoma humano, lo que puede potenciar la fiabilidad de las identificaciones forenses basadas en ADN, por ejemplo, proporcionando otro "marcador de identificación" (es decir otra secuencia de ADN que es única para un individuo particular). Tal como se mencionó anteriormente, la información básica real de secuencia puede usarse para la identificación como una alternativa precisa a los patrones formada por fragmentos generados por enzimas de restricción. Las secuencias dirigidas hacia regiones no codificantes de la secuencia de nucleótidos de la invención son particularmente apropiadas para este uso, dado que se produce un mayor número de polimorfismos en las regiones no codificantes, haciendo que sea más fácil diferenciar a los individuos usando esta técnica. Ejemplos de reactivos polinucleotídicos incluyen las secuencias de nucleótidos de CRSP o una parte de las mismas, por ejemplo, los fragmentos derivados de las regiones no codificantes de las secuencias de nucleótidos de la invención, que tienen una longitud de al menos 20 bases, preferiblemente de al menos 30 bases.

Las secuencias de nucleótidos de CRSP descrita en el presente documento puede usarse adicionalmente para proporcionar reactivos polinucleotídicos, por ejemplo, sondas marcadas o que se pueden marcar que pueden usarse en, por ejemplo, una técnica de hibridación *in situ*, para identificar un tejido específico, por ejemplo, tejido cerebral.

Esto puede ser muy útil en casos en los que un patólogo forense se presenta con un tejido de origen desconocido. Los paneles de tales sondas de CRSP pueden usarse para identificar tejidos mediante especies y/o por tipo de órgano.

De una manera similar, estos reactivos, por ejemplo, sondas o cebadores de CRSP, pueden usarse para examinar cultivo tisular para detectar contaminación (es decir, examinar para detectar la presencia de una mezcla de distintos tipos de células en un cultivo).

C. Medicina de diagnóstico

La presente invención también se refiere al campo de la medicina de diagnóstico, en la que se usan ensayos de diagnóstico, ensayos de pronóstico y ensayos clínicos de monitorización con fines de pronóstico, (predictivos) para tratar así a un individuo profilácticamente. Por consiguiente, un aspecto de la presente invención se refiere a ensayos de diagnóstico para determinar la proteína CRSP y/o la expresión del ácido nucleico, así como la actividad de CRSP, en el contexto de una muestra biológica (por ejemplo, sangre, suero, células, tejido) para determinar así si un individuo está afectado por una enfermedad o trastorno o está en riesgo de desarrollar un trastorno asociado con la expresión o actividad de CRSP aberrante, tal como la proliferación, diferenciación, y/o supervivencia celular aberrante, dando como resultado por ejemplo una enfermedad neurodegenerativa o cáncer. La invención también se refiere a ensayos de pronóstico (o predictivos) para determinar si un individuo está en riesgo de desarrollar un trastorno asociado con la actividad, expresión de ácido nucleico o proteína CRSP. Por ejemplo, pueden someterse a ensayo las mutaciones en un gen de CRSP en una muestra biológica. Tales ensayos pueden usarse con fines de pronóstico o predictivos para tratar así profilácticamente a un individuo antes del comienzo de un trastorno caracterizado por o asociado con la actividad, expresión de ácido nucleico o proteína CRSP.

Otro uso de la invención se refiere a la monitorización de la influencia de agentes (por ejemplo, fármacos, compuestos) sobre la expresión o actividad de CRSP en ensayos clínicos.

Estos y otros agentes se describen con detalle a continuación en las siguientes secciones.

1. Ensayos de diagnóstico

Un método a modo de ejemplo para detectar la presencia o ausencia de la proteína CRSP o el ácido nucleico en una muestra biológica implica obtener una muestra biológica de un sujeto de prueba y poner la muestra biológica con un compuesto o agente que puede detectar la proteína CRSP o el ácido nucleico (por ejemplo, ARNm, ADN genómico) que codifica para la proteína CRSP, de manera que se detecte la presencia de proteína CRSP o ácido nucleico en la muestra biológica. Un agente preferido para la detección de ADN genómico o ARNm de CRSP es una sonda de ácido nucleico marcada que puede hibridarse con ADN genómico o ARNm de CRSP. La sonda de ácido nucleico puede ser, por ejemplo, un ácido nucleico de CRSP de longitud completa, tal como un ácido nucleico de la invención, o una parte del mismo, tal como un oligonucleótido de al menos 15, 30, 50, 100, 250 ó 500 nucleótidos de longitud y suficiente para hibridarse específicamente en condiciones rigurosas con ADN genómico o ARNm de CRSP. En el presente documento se describen otras sondas adecuadas para su uso en los ensayos de diagnóstico de la invención.

Un agente preferido para detectar la proteína CRSP es un anticuerpo que puede unirse a la proteína CRSP, preferiblemente un anticuerpo con una marca detectable. Los anticuerpos pueden ser policlonales, o más preferiblemente, monoclonales. Puede usarse un anticuerpo intacto o un fragmento del mismo (por ejemplo, Fab o F(ab')₂). Se pretende que el término “marcado”, en relación con la sonda o anticuerpo, englobe el marcaje directo de la sonda o anticuerpo acoplado (es decir, uniéndolo físicamente) una sustancia detectable a la sonda o anticuerpo, así como el marcaje indirecto de la sonda o anticuerpo mediante la reacción con otro reactivo que está marcado directamente. Ejemplos de marcaje indirecto incluyen la detección de un anticuerpo primario usando un anticuerpo secundario marcado por fluorescencia y el marcaje de extremo de una sonda de ADN con biotina de manera que pueda detectarse con estreptavidina marcada por fluorescencia. Se pretende que el término “muestra biológica” incluya tejidos, células y fluidos biológicos aislados de un sujeto, así como tejidos, células y fluidos presentes en un sujeto. Es decir, el método de detección de la invención puede usarse para detectar ADN genómico, proteína o ARNm de CRSP en una muestra biológica *in vitro* así como *in vivo*. Por ejemplo, las técnicas *in vitro* para la detección de ARNm de CRSP incluyen hibridaciones de tipo Northern e hibridaciones *in situ*. Las técnicas *in vitro* para la detección de proteína CRSP incluyen ensayos de inmunoabsorción unido a enzimas (ELISA), inmunotransferencias de tipo Western, inmunoprecipitaciones e inmunofluorescencia. Las técnicas *in vitro* para detectar ADN genómico de CRSP incluyen hibridaciones de tipo Southern. Además, las técnicas *in vitro* para la detección de proteína CRSP incluyen introducir un anticuerpo anti-CRSP marcado en un sujeto. Por ejemplo, puede marcarse el anticuerpo con un marcador radiactivo cuya presencia y ubicación en un sujeto puede detectarse mediante técnicas de formación de imágenes habituales.

En una realización, la muestra biológica contiene moléculas de proteína del sujeto de prueba. Alternativamente, la muestra biológica puede contener moléculas de ARNm del sujeto de prueba o moléculas de ADN genómico del sujeto de prueba. Una muestra biológica preferida es una muestra de suero aislada mediante medios convencionales a partir de un sujeto.

En otra realización, los métodos implican adicionalmente obtener una muestra biológica control de un sujeto control, poner en contacto la muestra control con un compuesto o agente que puede detectar ADN genómico, ARNm o proteína CRSP, de manera que se detecte la presencia de ADN genómico, ARNm o proteína CRSP en la muestra

biológica y se compare la presencia de ADN genómico, ARNm o proteína CRSP en la muestra control con la presencia de ADN genómico, ARNm o proteína CRSP en la muestra de prueba.

La invención también engloba kits para detectar la presencia de CRSP en una muestra biológica. Por ejemplo, el kit puede comprender un agente o compuesto marcado que puede detectar el ARNm o proteína CRSP en una muestra biológica; medios para determinar la cantidad de CRSP en la muestra; y medios para comparar la cantidad de CRSP en la muestra con un patrón. El compuesto o agente puede estar envasado en un recipiente adecuado. Adicionalmente, el kit puede comprender instrucciones para usar el kit para detectar ácido nucleico o proteína CRSP.

2. Ensayos de pronóstico

Los métodos de diagnóstico descritos en el presente documento pueden utilizarse además para identificar sujetos que tienen o están en riesgo de desarrollar una enfermedad o trastorno asociado con actividad o expresión de CRSP aberrante. Por ejemplo, los ensayos descritos en el presente documento, tal como los anteriores ensayos de diagnósticos o los siguientes ensayos, pueden utilizarse para identificar a un sujeto que tiene o está en riesgo de desarrollar un trastorno asociado con la actividad, la expresión del ácido nucleico o la proteína CRSP, tal como un trastorno proliferativo, un trastorno del desarrollo o de la diferenciación, un trastorno así como enfermedades hematopoyéticas, estados o trastornos caracterizados por supervivencia celular anómala, estructura extracelular anómala o una anomalía en un mecanismo de defensa. Alternativamente, los ensayos de pronóstico pueden utilizarse para identificar a un sujeto que tiene o está en riesgo de desarrollar una enfermedad proliferativa o de diferenciación (por ejemplo, cáncer). Por tanto, la presente invención proporciona un método para identificar una enfermedad o trastorno asociado con la actividad o la expresión de CRSP aberrante, en el que la muestra de prueba se obtiene a partir de un sujeto y se detecta el ácido nucleico o la proteína CRSP (por ejemplo, ARNm, ADN genómico), en el que la presencia de ácido nucleico o proteína CRSP es diagnóstico para un sujeto que tiene o está en riesgo de desarrollar una enfermedad o trastorno asociado con la actividad o la expresión de CRSP aberrante. Tal como se usa en el presente documento, una "muestra de prueba" se refiere a una muestra biológica obtenida de un sujeto de interés. Por ejemplo, una muestra de prueba puede ser un fluido biológico (por ejemplo, suero), muestra celular o tejido.

Además, los ensayos de pronóstico descritos en el presente documento pueden usarse para determinar si se le puede o no administrar a un sujeto un agente (por ejemplo, un agonista, antagonista, peptidomimético, proteína, péptido, ácido nucleico, molécula pequeña, u otro candidato a fármaco) para tratar una enfermedad o trastorno asociado con actividad o expresión de CRSP aberrante. Por ejemplo, pueden usarse tales métodos para determinar si un sujeto puede tratarse eficazmente con un agente para un trastorno, tal como trastorno proliferativo, trastorno de diferenciación o del desarrollo, un trastorno hematopoyético, así como trastornos caracterizados por supervivencia celular anómala, estructura extracelular anómala o una anomalía en un mecanismo de defensa. Alternativamente, pueden usarse tales métodos para determinar si puede o no tratarse eficazmente a un sujeto con un agente para una enfermedad de diferenciación o proliferativa (por ejemplo, cáncer). Por tanto, la presente invención proporciona métodos para determinar si un sujeto puede tratarse eficazmente con un agente para un trastorno asociado con actividad o expresión de CRSP aberrante, en el que se obtiene una muestra de prueba y se detecta la actividad o expresión de ácido nucleico o proteína CRSP (por ejemplo, en el que la abundancia de actividad o expresión de ácido nucleico o proteína CRSP es diagnóstico para un sujeto al que se le pueden administrar el agente para tratar un trastorno asociado con actividad o expresión de CRSP aberrante).

Los métodos de la invención también pueden usarse para detectar alteraciones genéticas en un gen de CRSP, determinando así si un sujeto con el gen alterado está en riesgo de padecer un trastorno caracterizado por desarrollo aberrante, diferenciación celular aberrante, proliferación celular aberrante o una respuesta hematopoyética aberrante. En realizaciones preferidas, los métodos incluyen detectar, en una muestra de células del sujeto, la presencia o ausencia de una alteración genética mediante al menos una de una alteración que afecta a la integridad de un gen que codifica para una proteína CRSP, o la expresión fallida del gen de CRSP. Por ejemplo, pueden detectarse tales alteraciones genéticas determinando la existencia de al menos una de 1) una delección de uno o más nucleótidos de un gen de CRSP; 2) una adición de uno o más nucleótidos a un gen de CRSP; 3) una sustitución de uno o más nucleótidos de un gen de CRSP; 4) un reordenamiento cromosómico de un gen de CRSP; 5) una alteración en el nivel de un transcrito de ARN mensajero de un gen de CRSP; 6) modificación aberrante de un gen de CRSP, tal como el patrón de metilación del ADN genómico; 7) la presencia de un patrón de corte y empalme de tipo no natural de un transcrito de ARN mensajero de un gen de CRSP; 8) un nivel de tipo no natural de una proteína CRSP; 9) pérdida alélica de un gen de CRSP; y 10) modificación posterior a la traducción inapropiada de una proteína CRSP. Tal como se describe en el presente documento, hay un gran número de técnicas de ensayo conocidas en la técnica que pueden usarse para detectar alteraciones en un gen de CRSP. Una muestra biológica preferida es una muestra tisular o sérica aislada mediante medios convencionales de un sujeto.

En ciertas realizaciones, la detección de la alteración implica el uso de una sonda/cebador en una reacción en cadena de la polimerasa (PCR) (véanse, por ejemplo, las patentes estadounidenses números 4.683.195 y 4.683.202), tal como PCR de anclaje o PCR RACE, o, alternativamente, en una reacción en cadena de la ligasa (LCR) (véase, por ejemplo, Landegran *et al.* (1988) Science 241:1077-1080; y Nakazawa *et al.* (1994) PNAS 91:360-364), la última de las cuales puede ser particularmente útil para detectar mutaciones puntuales en el gen de CRSP (véase Abravaya *et al.* (1995) Nucleic Acids Res. 23:675-682). Este método puede incluir las etapas de tomar una muestra de células de un paciente, aislar el ácido nucleico (por ejemplo, genómico, ARNm o ambos) de las células de la muestra, poner en contacto la muestra de ácido nucleico con uno o más cebadores que hibridan específicamente con un gen de CRSP en

condiciones tales que se produzca la hibridación y la amplificación del gen de CRSP (si está presente), y detectar la presencia o ausencia de un producto de amplificación, o detectar el tamaño del producto de amplificación y comparar la longitud con una muestra control. Se prevé que puede ser deseable usar PCR y/o LCR como una etapa de amplificación preliminar junto con cualquiera de las técnicas usadas para detectar mutaciones descritas en el presente documento.

Métodos de amplificación alternativos incluyen: replicación de secuencia autosostenida (Guatelli, J.C. *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), sistema de amplificación de traducción (Kwoh, D.Y. *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), replicasa Q-Beta (Lizardi, P.M. *et al.*, 1988, Bio/Technology 6:1197), o cualquier otro método de amplificación de ácido nucleico, seguido por la detección de las moléculas amplificadas usando técnicas bien conocidas por los expertos en la técnica. Estos esquemas de detección son especialmente útiles para la detección de moléculas de ácido nucleico si tales moléculas se encuentran en muy pequeña cantidad.

En una realización alternativa, pueden identificarse mutaciones en un gen de CRSP de una célula de muestra mediante alteraciones en las rutas de escisión de las enzimas de restricción. Por ejemplo, se aísla ADN de muestra y control, se amplifica (opcionalmente), se digiere con una o más endonucleasas de restricción, y se determinan los tamaños de longitud de fragmento mediante electroforesis en gel y se comparan. Las diferencias en los tamaños de longitud de fragmento entre ADN de muestra y control indica mutaciones en el ADN de muestra. Además, el uso de ribozimas específicas de secuencia (véase, por ejemplo, patente estadounidense número 5.498.531) puede usarse para puntuar la presencia de mutaciones específicas por desarrollo o pérdida de un sitio de escisión de ribozima.

En otras realizaciones, pueden identificarse las mutaciones genéticas en CRSP hibridando ácidos nucleicos de muestra y control, por ejemplo, ADN o ARN, a ensayos de alta densidad que contienen cientos o miles de sondas oligonucleotídicas (Cronin, M.T. *et al.* (1996) Human Mutation 7: 244-255; Kozal, M.J. *et al.* (1996) Nature Medicine 2: 753-759). Por ejemplo, pueden identificarse mutaciones genéticas en CRSP en matrices bidimensionales que contienen sondas de ADN generadas por luz tal como se describe en Cronin, M.T. *et al.* citado anteriormente. En resumen, una primera matriz de hibridación de sondas puede usarse para explorar a través de extensiones largas de ADN en una muestra y control para identificar cambios de bases entre las secuencias realizando matrices lineales de sondas solapantes de manera secuencial. Esta etapa permite la identificación de mutaciones puntuales. Esta etapa va seguida por una segunda serie de hibridación que permite la caracterización de mutaciones específicas usando matrices de sondas especializadas más pequeñas complementarias a todas las variantes o mutaciones detectadas. Cada serie de mutación está compuesta por conjuntos de sondas paralelas, uno complementario al gen de tipo natural y el otro complementario al gen mutante.

En todavía otra realización, puede usarse cualquiera de una variedad de reacciones de secuenciación conocidas en la técnica para secuenciar directamente el gen de CRSP y detectar mutaciones comparando la secuencia de la CRSP de muestra con la correspondiente secuencia de tipo natural (control). Ejemplos de reacciones de secuenciación incluyen las basadas en técnicas desarrolladas por Maxim y Gilbert ((1977) PNAS 74:560) o Sanger ((1977) PNAS 74:5463). También se contempla que puede utilizarse cualquiera de una variedad de procedimientos de secuenciación automatizados cuando se realizan ensayos de diagnóstico ((1995) Biotechniques 19: 448), incluyendo secuenciación mediante espectrometría de masas (véase, por ejemplo, Publicación PCT Internacional número WO 94/16101; Cohen *et al.* (1996) Adv. Chromatogr. 36:127-162; y Griffin *et al.* (1993) Appl. Biochem. Biotechnol. 38:147-159).

Otros métodos para detectar mutaciones en el gen de CRSP incluyen métodos en los que se usa protección a partir de agentes de escisión para detectar bases con apareamiento erróneo en heterodúplex de ARN/ARN o ARN/ADN (Myers *et al.* (1985) Science 230:1242). En general, la técnica "escisión por apareamiento erróneo" comienza proporcionando heterodúplex de ARN o ADN (marcados) mediante hibridación, que contienen la secuencia de CRSP de tipo natural con ARN o ADN potencialmente mutante obtenido de una muestra de tejido. Los dúplex de cadena doble se tratan con un agente que escinde las regiones de cadena sencilla del dúplex tal como existirá debido a los apareamientos erróneos de pares de bases entre las cadenas control y de muestra. Por ejemplo, pueden tratarse los dúplex de ARN/ADN con ARNasa y pueden tratarse los híbridos ADN/ADN con nucleasa S1 para digerir enzimáticamente las regiones erróneamente apareadas. En otras realizaciones, pueden tratarse los dúplex o bien de ADN/ADN o bien de ARN/ADN con hidroxilamina o tetróxido de osmio y con piperidina con el fin de digerir las regiones erróneamente apareadas. Tras la digestión de las regiones erróneamente apareadas, el material resultante se separa entonces por tamaño sobre geles de poliacrilamida de desnaturalización para determinar el sitio de mutación. Véase, por ejemplo, Cotton *et al.* (1988) Proc. Natl Acad Sci USA 85:4397; Saleeba *et al.* (1992) Methods Enzymol. 217:286-295. En una realización preferida, el ADN o ARN control puede marcarse para su detección.

En todavía otra realización, la reacción de escisión por apareamiento erróneo emplea una o más proteínas que reconocen pares de bases erróneamente apareadas en ADN de cadena doble (denominadas enzimas de "reparación de apareamiento erróneo de ADN") en sistemas definidos para detectar y mapear mutaciones puntuales en ADN de CRSP obtenidos de muestras de células. Por ejemplo, la enzima mutY de *E. coli* escinde A en apareamiento erróneo G/A y la timidina ADN glucosilasa a partir de células HeLa escinde T en apareamientos erróneos G/T (Hsu *et al.* (1994) Carcinogenesis 15:1657-1662). Según una realización ejemplo, una sonda basada en una secuencia de CRSP, por ejemplo, una secuencia de CRSP de tipo natural, hibrida con un ADNc u otro producto de ADN de una(s) célula(s) de prueba. Se trata el dúplex con una enzima de reparación de apareamiento erróneo de ADN, y los productos de escisión, si hay alguno, pueden detectarse a partir de protocolos de electroforesis o similares. Véase, por ejemplo, la patente estadounidense número 5.459.039.

En otras realizaciones, se usarán alteraciones en la movilidad electroforética para identificar mutaciones en genes de CRSP. Por ejemplo, puede usarse polimorfismo de conformación de cadena sencilla (SSCP) para detectar diferencias en la movilidad electroforética entre ácidos nucleicos mutantes y de tipo natural (Orita *et al.* (1989) Proc Natl. Acad. Sci USA: 86:2766, véase también Cotton (1993) Mutat Res 285:125-144; y Hayashi (1992) Genet Anal Tech Appl 9:73-79). Fragmentos de ADN de cadena sencilla de ácidos nucleicos CRSP de muestra y control se desnaturalizarán y se permitirá su renaturalización. La estructura secundaria de los ácidos nucleicos de cadena sencilla varía según la secuencia, la alteración resultante en la movilidad electroforética permite la detección de incluso un único cambio de base. Pueden marcarse los fragmentos de ADN o detectarse con sondas marcadas. La sensibilidad del ensayo puede aumentarse usando ARN (en lugar de ADN), en el que la estructura secundaria es más sensible a un cambio en la secuencia. En una realización preferida, el método objeto utiliza análisis de heterodúplex para separar moléculas de heterodúplex de cadena doble basándose en cambios de movilidad electroforética (Keen *et al.* (1991) Trends Genet 7:5).

En todavía otra realización, se somete a ensayo el movimiento de fragmentos mutantes o de tipo natural en geles de poliacrilamida que contienen un gradiente de agente de desnaturalización usando electroforesis en gel de gradiente de desnaturalización (DGGE) (Myers *et al.* (1985) Nature 313:495). Cuando se usa DGGE como el método de análisis, se modificará el ADN para asegurar que no está completamente desnaturalizado, por ejemplo añadiendo una estructura de unión de GC ("GC clamp") de aproximadamente 40 pb de ADN rico en GC de alto punto de fusión mediante PCR. En una realización adicional, se usa un gradiente de temperatura en lugar de un gradiente de desnaturalización para identificar diferencias en la movilidad del ADN de muestra y control (Rosenbaum y Reissner (1987) Biophys Chem 265:12753).

Ejemplos de otras técnicas para detectar mutaciones puntuales incluyen, pero no se limitan a, hibridación de oligonucleótidos selectiva, amplificación selectiva, o extensión de cebador selectiva. Por ejemplo, pueden prepararse cebadores oligonucleotídicos en los que la mutación conocida se coloca en el centro y luego hibrida con ADN diana en condiciones que permiten la hibridación sólo si se encuentra un apareamiento perfecto (Saiki *et al.* (1986) Nature 324:163; Saiki *et al.* (1989) Proc. Natl Acad. Sci USA 86:6230). Tales oligonucleótidos específicos de alelo se hibridan con ADN diana amplificado por PCR o varias mutaciones diferentes cuando los oligonucleótidos están unidos a la membrana de hibridación e hibridados con el ADN diana marcado.

Alternativamente, junto con la presente invención puede usarse la tecnología de amplificación específica de alelo que depende de la amplificación por PCR selectiva. Los oligonucleótidos usados como cebadores para la amplificación específica pueden llevar la mutación de interés en el centro de la molécula (de manera que la amplificación depende de la hibridación diferencial) (Gibbs *et al.* (1989) Nucleic Acids Res. 17:2437-2448) o en el extremo 3' terminal de un cebador cuando, en condiciones apropiadas, el apareamiento erróneo puede evitar, o reducir la extensión de la polimerasa (Prossner (1993) Tibtech 11:238). Además, puede ser deseable introducir un sitio de restricción novedoso en la región de la mutación para crear detección basada en escisión (Gasparini *et al.* (1992) Mol. Cell Probes 6:1). Se prevé que en ciertas realizaciones también puede realizarse la amplificación usando Taq ligasa para la amplificación (Barany (1991) Proc. Natl. Acad. Sci USA 88:189). En tales casos, sólo se producirá la ligamentación sólo si hay un apareamiento perfecto en el extremo 3' de la secuencia en 5' haciendo que sea posible detectar la presencia de una mutación conocida en un sitio específico buscando la presencia o ausencia de amplificación.

Los métodos descritos en el presente documento pueden llevarse a cabo, por ejemplo, utilizando kits de diagnóstico envasados previamente que comprenden al menos un reactivo de anticuerpo o ácido nucleico de sonda descritos en el presente documento, que pueden usarse convenientemente, por ejemplo, en la práctica clínica para diagnosticar a pacientes que muestran síntomas o historial familiar de una enfermedad o afección que implica un gen de CRSP.

Además, cualquier tipo de célula o tejido en el que se expresa CRSP puede utilizarse en los ensayos de pronóstico descritos en el presente documento.

3. Monitorización de los efectos durante ensayos clínicos

La monitorización de la influencia de agentes (por ejemplo, fármacos, compuestos) sobre la expresión o la actividad de CRSP (por ejemplo, modulación de transducción de señal celular, regulación de transcripción génica en una célula implicada en el desarrollo o la diferenciación, regulación de la proliferación celular) puede aplicarse no sólo en la selección de fármaco básica, sino también en ensayos clínicos. Por ejemplo, la eficacia de un agente determinado mediante un ensayo de selección, tal como se describe en el presente documento, para aumentar la expresión de gen de CRSP, los niveles de proteína, o para regular por incremento la actividad de CRSP, puede monitorizarse en ensayos clínicos de sujetos que muestran disminución de la expresión del gen de CRSP, niveles de proteína o actividad de CRSP regulada por disminución. Alternativamente, la eficacia de un agente determinado mediante un ensayo de selección para disminuir la expresión de gen de CRSP, los niveles de proteína, o regular por disminución la actividad de CRSP, puede monitorizarse en ensayos clínicos de sujetos que muestran aumento de la expresión del gen de CRSP, niveles de proteína, o actividad de CRSP regulada por incremento. En tales ensayos clínicos, la expresión o actividad de CRSP y, preferiblemente, de otros genes que se han visto implicados en, por ejemplo, un trastorno proliferativo, pueden usarse como un "lector" o marcadores del fenotipo de una célula particular.

Por ejemplo, y no a modo de limitación, pueden identificarse genes, incluyendo CRSP, que se modulan en células mediante tratamiento con un agente (por ejemplo, compuesto, fármaco o molécula pequeña) que modula la actividad de CRSP (por ejemplo, identificada en un ensayo de selección tal como se describe en el presente documento). Por

tanto, para estudiar el efecto de los agentes sobre trastornos proliferativos, trastornos del desarrollo o de diferenciación, trastornos hematopoyéticos así como trastornos caracterizados por supervivencia y/o diferenciación celular anómala, una estructura extracelular anómala, o una anomalía en un mecanismo de defensa, por ejemplo, en un ensayo clínico, pueden aislarse células y prepararse ARN y analizarse para detectar los niveles de expresión de CRSP y otros genes implicados en el trastorno proliferativo, trastorno del desarrollo o de diferenciación, trastorno hematopoyético así como trastornos caracterizados por supervivencia y/o diferenciación celular anómala, una estructura extracelular anómala o una anomalía en un mecanismo de defensa, respectivamente. Pueden cuantificarse los niveles de expresión génica (es decir, un patrón de expresión génica) mediante análisis de transferencia tipo Northern o RT-PCR, tal como se describe en el presente documento, o alternativamente midiendo la cantidad de proteína producida, mediante uno de los métodos tal como se describe en el presente documento, o midiendo los niveles de actividad de CRSP u otros genes. De esta manera, el patrón de expresión génica puede servir como un marcador, indicativo de la respuesta fisiológica de las células frente al agente. Por consiguiente, este estado de respuesta puede determinarse antes, y en distintos puntos durante el tratamiento del individuo con el agente.

La presente invención puede usarse como base de un método para monitorizar la eficacia del tratamiento de un sujeto con un agente (por ejemplo, un agonista, antagonista, peptidomimético, proteína, péptido, ácido nucleico, molécula pequeña u otro candidato a fármaco identificado por los ensayos de selección descritos en el presente documento) que comprende las etapas de (i) obtener una muestra antes de la administración de un sujeto antes de la administración del agente, (ii) detectar el nivel de expresión de un ADN genómico, ARNm o proteína CRSP en la muestra antes de la administración; (iii) obtener una o más muestras posteriores a la administración del sujeto; (iv) detectar el nivel de expresión o actividad del ADN genómico, ARNm o proteína CRSP, en las muestras posteriores a la administración; (v) comparar el nivel de expresión o actividad de ADN genómico, ARNm o proteína CRSP, en la muestra antes de la administración con ADN genómico, ARNm o proteína CRSP en la muestra o muestras posteriores a la administración; y (vi) modificar la administración del agente al sujeto en consecuencia. Por ejemplo, el aumento de la administración del agente puede ser deseable para aumentar la expresión o actividad de CRSP a niveles superiores a los detectados, es decir, para aumentar la eficacia del agente. Alternativamente, puede ser deseable una disminución de la administración del agente para disminuir la expresión o actividad de CRSP a niveles inferiores a los detectados, es decir para disminuir la eficacia del agente. Según tal realización, puede usarse la expresión o actividad de CRSP como un indicador de la eficacia de un agente, incluso en ausencia de una respuesta de fenotipo observable.

C. Métodos de tratamiento

La presente invención puede usarse como la base tanto para métodos profilácticos como terapéuticos para tratar a un sujeto en riesgo de (o propenso a) padecer un trastorno o que tiene un trastorno asociado con actividad o expresión de CRSP aberrante. En relación con ambos métodos de tratamiento profiláctico y terapéutico, tales tratamientos pueden diseñarse o modificarse específicamente, basándose en el conocimiento obtenido del campo de farmacogenómica. La "farmacogenómica", tal como se usa en el presente documento, se refiere a la aplicación de tecnologías genómicas tales como la secuenciación génica, la genética estadística y el análisis de la expresión génica para fármacos en desarrollo clínico y en el mercado. Más específicamente, el término se refiere al estudio de cómo los genes de un paciente determinan su respuesta frente a un fármaco (por ejemplo, el "fenotipo de respuesta a fármaco", o "genotipo de respuesta a fármaco" de un paciente). Por tanto, la invención proporciona la base de los métodos para diseñar un tratamiento terapéutico o profiláctico de un individuo con cualquiera de las moléculas de CRSP de la presente invención, o bien con moduladores de CRSP según el genotipo de respuesta a fármaco de ese individuo. La farmacogenómica permite que un internista o médico dirija tratamientos profilácticos o terapéuticos a los pacientes que más se beneficiarán del tratamiento y para evitar el tratamiento de pacientes que experimentarán efectos secundarios relacionados con fármacos tóxicos.

1. Métodos profilácticos

En un aspecto, la invención proporciona la fabricación de un medicamento para evitar en un sujeto, una enfermedad o estado asociado con una actividad o expresión de CRSP aberrante. El medicamento incluye un agente que modula la expresión de CRSP o al menos una actividad de CRSP. Los sujetos que están en riesgo de padecer una enfermedad que está producida por o a la que contribuye la actividad o expresión de CRSP aberrante pueden identificarse mediante, por ejemplo, cualquiera o una combinación de ensayos de diagnóstico o pronóstico tal como se describe en el presente documento. La administración de un agente profiláctico puede producirse antes de la manifestación de los síntomas característicos de la aberración de CRSP, de manera que se evite una enfermedad o trastorno o, alternativamente, se retrase su progresión. Dependiendo del tipo de aberración de CRSP, por ejemplo, puede usarse un agonista de CRSP o agente antagonista de CRSP para tratar al sujeto. Puede determinarse el agente apropiado basándose en ensayos de selección descritos en el presente documento. Los métodos profilácticos se tratan con más detalle en las siguientes subsecciones.

2. Métodos terapéuticos

Otro aspecto de la invención proporciona la fabricación de un medicamento para modular la actividad o expresión de CRSP con fines terapéuticos. El método modulador de la invención implica poner en contacto una célula con un agente que modula una o más de las actividades de la actividad de la proteína CRSP asociada con la célula. Un agente que modula la actividad de la proteína CRSP puede ser un agente tal como se describe en el presente documento, tal como un ácido nucleico o una proteína, una molécula diana que se produce de manera natural de una proteína

CRSP, un péptido, un peptidomimético de CRSP, u otra molécula pequeña. En una realización, el agente estimula una o más actividades de proteína CRSP. Ejemplos de tales agentes estimulantes incluyen proteína CRSP y una molécula de ácido nucleico que codifica para CRSP que se ha introducido en la célula. En otra realización, el agente inhibe una o más actividades de la proteína CRSP. Ejemplos de tales agentes inhibidores incluyen moléculas de ácido nucleico de CRSP antisentido y anticuerpos anti-CRSP. Estos métodos moduladores pueden realizarse *in vitro* (por ejemplo, mediante el cultivo de la célula con el agente) o, alternativamente, *in vivo*, (por ejemplo, administrando el agente a un sujeto). Como tal, la presente invención proporciona la fabricación de medicamentos para tratar a un individuo aquejado de una enfermedad o trastorno caracterizado por actividad o expresión de CRSP aberrante de una proteína CRSP o molécula de ácido nucleico. En una realización, el medicamento incluye un agente (por ejemplo, un agente identificado mediante un ensayo de selección descrito en el presente documento), o combinación de agentes que modulan (por ejemplo, regulan por incremento o regulan por disminución) la actividad o la expresión de CRSP. Puede administrarse una proteína CRSP o una molécula de ácido nucleico como un tratamiento para compensar la actividad o la expresión de CRSP aberrante o reducida.

La estimulación de la actividad de CRSP es deseable en situaciones en las que CRSP está anómalamente reducida por disminución y/o en las que es probable que el aumento de la actividad de CRSP tenga un efecto beneficioso. De igual manera, la inhibición de la actividad de CRSP es deseable en situaciones en las que CRSP está anómalamente regulada por incremento y/o en las que es probable que la disminución de la actividad de CRSP tenga efectos beneficiosos. Un ejemplo de tal situación es cuando un sujeto tiene un trastorno caracterizado por diferenciación celular o de desarrollo aberrante. Otro ejemplo de tal situación es cuando el sujeto tiene una enfermedad proliferativa (por ejemplo, cáncer) o un trastorno caracterizado por una respuesta hematopoyética aberrante. Todavía otro ejemplo de tal situación es cuando es deseable lograr regeneración tisular en un sujeto (por ejemplo, cuando un sujeto ha experimentado daño cerebral o de la médula espinal y es deseable regenerar tejido neuronal de una manera regulada).

Por consiguiente, en una realización, la enfermedad es una enfermedad caracterizada por una proliferación, diferenciación, y/o supervivencia celular anómalas. Por ejemplo, la enfermedad puede ser una enfermedad hiper o hipoproliferativa. La invención también se refiere a la fabricación de un medicamento para tratar enfermedades caracterizadas por una proliferación, diferenciación, y/o supervivencia celular anómalas en un sujeto, que no están caracterizadas por actividad de CRSP anómala (por ejemplo, actividad de CRSP-1). De hecho, dado que es probable que CRSP pueda modular el estado proliferativo de una célula (es decir, estado de proliferación, diferenciación, y/o supervivencia de una célula), CRSP puede regular una enfermedad en la que el estado proliferativo de una célula resulta de un defecto distinto de la actividad de CRSP anómala.

Las enfermedades hiperproliferativas pueden tratarse con agentes terapéuticos de CRSP (por ejemplo, CRSP-1) incluyendo enfermedades neoplásicas e hiperplásicas, tales como diversas formas de cáncer y leucemias y trastornos fibroproliferativos. Otras enfermedades hiperproliferativas que pueden tratarse o evitarse con los agentes terapéuticos de CRSP objeto (por ejemplo agentes terapéuticos de CRSP-1) incluyen estados cancerosos, estados precancerosos, estados benignos. El estado que va a tratarse o evitarse puede ser un tumor sólido, tal como un tumor que surge en un tejido epitelial. Por consiguiente, el tratamiento de tal cáncer podría comprender la administración al sujeto de un agente terapéutico de CRSP que disminuya la interacción de CRSP con un receptor de CRSP. Otros cánceres que pueden tratarse o evitarse con una proteína CRSP incluyen sarcomas y carcinomas, por ejemplo, cáncer de pulmón, cáncer de colon, cáncer de próstata, mama, ovario, esófago, pulmón, melanoma, seminoma y adenocarcinoma de células escamosas. Otros tumores sólidos dentro del alcance de la invención incluyen los que pueden encontrarse en un libro de texto médico.

El estado que va a tratarse o evitarse también puede ser un tumor soluble, tal como leucemia, o bien crónico o bien agudo, incluyendo leucemia mielógena crónica o aguda, leucemia linfocítica crónica o aguda, leucemia promielocítica, leucemia monocítica, leucemia mielomonocítica, y eritroleucemia. Todavía otros trastornos proliferativos que pueden tratarse con un agente terapéutico de CRSP de la invención incluyen enfermedad de cadena pesada, mieloma múltiple, linfoma, por ejemplo, linfoma de Hodgkin y linfoma no Hodgkin y macroglobulemia de Waldenstroem.

Enfermedades o estados caracterizados por un tumor sólido o soluble pueden tratarse administrando un agente terapéutico de CRSP o bien local o bien sistemáticamente, de manera que se inhibe o disminuye la proliferación celular aberrante. A continuación se describen con detalle métodos para administrar los compuestos de la invención.

La invención también proporciona la fabricación de un medicamento para evitar la formación y/o desarrollo de tumores. Por ejemplo, puede anticiparse el desarrollo de un tumor mediante la presencia de una lesión específica, tal como una lesión preneoplásica, por ejemplo, hiperplasia, metaplasia, y displasia, que pueden detectarse, por ejemplo, mediante métodos citológicos. Tales lesiones pueden encontrarse; por ejemplo, en el tejido epitelial. Por tanto, la invención se refiere a la fabricación de un medicamento para inhibir la progresión de tal lesión en una lesión neoplásica. Al sujeto que tiene una lesión preneoplásica se le puede administrar una cantidad de agente terapéutico de CRSP-1 suficiente para inhibir la progresión de la lesión preneoplásica en una lesión neoplásica.

La invención también se refiere a la fabricación de un medicamento para tratar o prevenir enfermedades en las que se desea la proliferación celular. Por ejemplo, pueden usarse agentes terapéuticos de CRSP para estimular la reparación tisular o la cicatrización de heridas, tal como tras la cirugía o para estimular la cicatrización tisular de quemaduras. Otras enfermedades en las que es deseable la proliferación de células son las enfermedades hipoproliferativas, es decir, enfermedades caracterizadas por una proliferación anómalamente baja de ciertas células.

En todavía otra realización, la invención proporciona la fabricación de un medicamento para tratar o prevenir enfermedades o estados caracterizados por diferenciación celular aberrante. Por consiguiente, métodos para estimular la diferenciación celular en estados caracterizados por una inhibición de diferenciación celular normal que pueden o no estar acompañadas por una proliferación excesiva. Alternativamente, pueden usarse agentes terapéuticos de CRSP para inhibir la diferenciación de células específicas.

En un método preferido, la célula que prolifera y/o se diferencia de manera aberrante es una célula presente en el sistema nervioso. Se sugiere un papel para CRSP en el sistema nervioso al menos en parte a partir del hecho de que CRSP-1 humana se expresa en cerebro fetal humano. Por consiguiente, la invención proporciona la fabricación de un medicamento para tratar enfermedades o estados asociados con el sistema nervioso central o periférico. Por ejemplo, la invención se refiere a la fabricación de un medicamento para tratar lesiones del sistema nervioso asociadas con una proliferación, diferenciación o supervivencia aberrante de cualquiera de las siguientes células: neuronas, células de Schwann, células gliales y otro tipo de células neurales. Los trastornos del sistema nervioso incluyen, pero no se limitan a: daños en la médula espinal, daños cerebrales, lesiones asociadas con cirugía, lesiones isquémicas, lesiones cancerosas, lesiones infecciosas, lesiones degenerativas (enfermedad de Parkinson, enfermedad de Alzheimer, corea de Huntington, esclerosis lateral amiotrófica), enfermedades desmielinizantes (esclerosis múltiple, mielopatía asociada a inmunodeficiencia humana, mielopatía transversa, leucoencefalopatía multifocal progresiva, mielinolisis pontina), daños en neurona motora, atrofia muscular espinal progresiva, parálisis bulbar progresiva, esclerosis lateral primaria, atrofia muscular infantil y juvenil, parálisis bulbar progresiva infantil (síndrome Fazio-Londe), poliomielitis, y neuropatía motora y sensitiva hereditaria (enfermedad de Charcot-Marie-Tooth).

En otra realización, la invención proporciona la fabricación de un medicamento para potenciar la supervivencia y/o estimular la proliferación y/o diferenciación de células y tejidos *in vitro*. En una realización preferida, se usan los agentes terapéuticos de CRSP para estimular la regeneración y/o reparación tisular (por ejemplo, para tratar lesión nerviosa). Por ejemplo, pueden obtenerse tejidos de un sujeto y hacerlos crecer *in vitro* en presencia de un agente terapéutico de CRSP, de manera que las células tisulares estimulen para proliferar y/o diferenciar. Después puede volver a administrarse el tejido al sujeto.

Entre los enfoques que pueden usarse para aliviar los síntomas que implican una actividad de CRSP aberrante y/o una proliferación, diferenciación, y/o supervivencia celular anómalas, se encuentran, por ejemplo, las moléculas antisentido, ribozimas y de triple hélice, descritas anteriormente. Ejemplos de compuestos adecuados incluyen los antagonistas, agonistas u homólogos descritos con detalle anteriormente.

Todavía otros agentes terapéuticos de CRSP consisten en un primer péptido que comprende un péptido de CRSP que puede unirse a un receptor de CRSP, y un segundo péptido que es citotóxico. Tales agentes terapéuticos pueden usarse para seleccionar específicamente como objetivo y lisar células que expresan o sobreexpresan un receptor para CRSP.

3. Farmacogenómica

Las moléculas de CRSP de la presente invención, así como los agentes, o moduladores que tienen un efecto estimulante o inhibidor sobre la actividad de CRSP (por ejemplo, expresión del gen de CRSP) tal como se identifica mediante un ensayo de selección descrito en el presente documento, pueden administrarse a individuos para tratar (profiláctica o terapéuticamente) trastornos (por ejemplo, trastornos proliferativos o del desarrollo) asociados con actividad de CRSP aberrante. Junto con tal tratamiento, puede considerarse la farmacogenómica (es decir, el estudio de la relación entre el genotipo de un individuo y la respuesta de este individuo a un compuesto extraño o fármaco). Las diferencias en el metabolismo de agentes terapéuticos pueden dar lugar a toxicidad grave o fallo terapéutico mediante la alteración de la relación entre la dosis y la concentración en la sangre del fármaco activo farmacológicamente. Por tanto, un médico o internista puede considerar aplicar el conocimiento obtenido en estudios de farmacogenómica relevantes para determinar si administrar o no una molécula de CRSP o modulador de CRSP, así como diseñar la dosificación y/o régimen terapéutico de tratamiento con una molécula de CRSP o modulador de CRSP.

La farmacogenómica trata con variaciones hereditarias clínicamente significativas en la respuesta frente a fármacos debido a disposición de fármaco alterada y a la acción anómala en personas afectadas. Véase por ejemplo, Eichelbaum, M., Clin Exp Pharmacol Physiol, 1996, 23(10-11):983-985 y Linder, M.W., Clin Chem, 1997, 43(2):254-266. En general, pueden diferenciarse dos tipos de estados farmacogenéticos. Estados genéticos transmitidos como un factor único que altera la forma en que los fármacos actúan en el cuerpo (acción de fármaco alterada) o estados genéticos transmitidos como factores únicos que alteran la forma en que el cuerpo actúa sobre los fármacos (metabolismo de fármaco alterado). Estos estados farmacogenéticos pueden producirse o bien como defectos genéticos raros o como polimorfismos que se producen de manera natural. Por ejemplo, la deficiencia de la glucosa-6-fosfato deshidrogenasa (G6PD) es una enzimopatía hereditaria común en la que la principal complicación clínica es la hemólisis tras la ingesta de fármacos oxidantes (antimalariales, sulfonamidas, analgésicos, nitrofuranos,) y consumo de habas.

Un enfoque farmacogenómico para identificar genes que predicen la respuesta al fármaco, conocidos como “una asociación amplia de genoma”, se basa en primer lugar en un mapa de alta resolución del genoma humano que consiste en marcadores relacionados con genes ya conocidos (por ejemplo, un mapa de marcador de gen “bialélico” que consiste en 60.000-100.000 sitios polimórficos o variables del genoma humano, cada uno de los cuales tiene dos variantes). Tal mapa genético de alta resolución puede compararse con un mapa del genoma de cada uno de un número

estadísticamente significativo de pacientes que toma parte en un ensayo de fármaco en fase II/III para identificar marcadores asociados con una respuesta al fármaco particular observada o efectos secundarios. Alternativamente, tal mapa de alta resolución puede generarse a partir de una combinación de alguno de los diez millones de polimorfismos de un único nucleótido (SNP) conocidos en el genoma humano. Tal como se usa en el presente documento, un “SNP” es una alteración común que se produce en una base nucleotídica sencilla en una extensión de ADN. Por ejemplo, puede producirse un SNP una vez cada 1000 bases de ADN. Un SNP puede estar implicado en un proceso de enfermedad, sin embargo, la inmensa mayoría no estarán asociados a enfermedad. Dado un mapa genético basado en la aparición de tales SNP, pueden agruparse los individuos en categorías genéticas dependiendo de un patrón de SNP particular en su genoma individual. De tal manera, los regímenes de tratamiento pueden diseñarse, para grupos de individuos genéticamente similares, teniendo en cuenta características que pueden ser comunes entre tales individuos genéticamente similares.

Alternativamente, puede utilizarse un método denominado el “enfoque de gen candidato”, para identificar genes que predican respuestas ante fármacos. Según este método, si se conoce un gen que codifica para una diana de fármacos (por ejemplo, una proteína CRSP o receptor de CRSP de la presente invención), todas las variantes comunes de ese gen pueden identificarse bastante fácilmente en la población y puede determinarse si tener una versión del gen frente a otro está asociado con una respuesta al fármaco particular.

Como una realización ilustrativa, la actividad de las enzimas que metabolizan el fármaco es un determinante principal tanto de la intensidad como de la duración de la acción del fármaco. El descubrimiento de los polimorfismos genéticos de las enzimas que metabolizan el fármaco (por ejemplo, N-acetiltransferasa 2 (NAT 2) y enzimas CYP2D6 y CYP2C 19 de citocromo P450) ha proporcionado una explicación de por qué algunos pacientes no obtienen los efectos del fármaco esperados o muestran respuesta al fármaco exagerada y toxicidad grave tras tomar la dosis habitual y segura de un fármaco. Estos polimorfismos se expresan en dos fenotipos en la población, el metabolizador rápido (EM, extensive metabolizer) y el metabolizador lento (PM, poor metabolizer). La prevalencia de PM es diferente entre poblaciones diferentes. Por ejemplo, el gen que codifica para CYP2D6 es sumamente polimórfico y se han identificado varias mutaciones en PM, que dan lugar todas a la ausencia de CYP2D6 funcional. Los metabolizadores lentos de CYP2D6 y CYP2C19 experimentan bastante a menudo efectos secundarios y respuesta al fármaco exagerada cuando reciben las dosis habituales. Si un metabolito es el resto terapéutico activo, los PM no muestran respuesta terapéutica, tal como se demuestra por el efecto analgésico de la codeína mediada por su metabolito de morfina formado por CYP2D6. En el otro extremo están los denominados metabolizadores ultrarrápidos que no responden a las dosis habituales. Recientemente, se ha identificado que la base molecular del metabolismo ultrarrápido se debe a la amplificación del gen de CYP2D6.

Alternativamente, un método denominado “elaboración del perfil de expresión génica”, puede utilizarse para identificar genes que predican la respuesta al fármaco. Por ejemplo, la expresión génica de un animal al que se dosifica un fármaco (por ejemplo, una molécula de CRSP o modulador de CRSP de la presente invención) puede dar una indicación de si las rutas génicas relacionadas con la toxicidad se han activado o no.

Puede usarse la información generada a partir de más de uno de los enfoques farmacogenómicos anteriores pueden usarse para determinar la dosificación y los regímenes de tratamiento adecuados para el tratamiento profiláctico o terapéutico de un individuo. Este conocimiento, cuando se aplica a la dosificación o la selección de fármaco, puede evitar reacciones adversas o fallos terapéuticos y por tanto, potenciar la eficacia terapéutica o profiláctica cuando se trata a un sujeto con una molécula de CRSP o un modulador de CRSP, tal como un modulador identificado por uno de los ensayos de selección a modo de ejemplo descritos en el presente documento.

Esta invención se ilustra adicionalmente mediante los siguientes ejemplos que no deben considerarse como limitantes.

Ejemplos

Ejemplo 1

Aislamiento y caracterización de ADNc de CRSP-1 humana

En este ejemplo, se describe el aislamiento y la caracterización del gen que codifica para CRSP- 1 (también denominada “CRISPY-1” o “TANGO 59”).

Aislamiento de ADNc de CRSP-1 humana

La invención se basa al menos en parte en el descubrimiento de un gen humano que codifica para una proteína segregada, denominada en el presente documento proteína segregada rica en cisteína-1 (CRSP-1). Se aisló un ADNc parcial usando un método para atrapar la secuencia señal. Esta metodología se aprovecha del hecho de que moléculas tales como CRSP tienen una secuencia señal amino terminal que dirige ciertas proteínas segregadas y unidas a membrana a través del aparato secretor celular.

En resumen, se preparó una biblioteca de ADNc con cebador al azar usando ARNm preparado a partir de tejido cerebral fetal humano (Clontech, Palo Alto CA) usando el kit Stratagene-ZAP-ADNc Synthesis™, (número de catá-

- logo 20041). Se ligó el ADNc en el vector de expresión de mamífero pTrap adyacente a un ADNc que codifica para la fosfatasa alcalina de placenta carece de señal secretora. Se transformaron los plásmidos en *E. coli* y se preparó ADN usando el kit de purificación de ADN Wizard™ (Promega). Se transfectó ADN en células COS-7 con Lipofectamine™ (Gibco-BRL). Tras una incubación de 48 horas se sometieron a ensayo los sobrenadantes de las células COS para detectar fosfatasa alcalina en un contador de centelleo Micro-Beta de Wallac usando el kit Phospha-Light™ (Tropix Inc., número de catálogo BP300). Los ADN de plásmido individuales que puntúan positivo en el ensayo de secreción de fosfatasa alcalina de célula COS se analizaron adicionalmente mediante secuenciación de ADN usando procedimientos habituales.
- Usando un ADNc parcial aislado mediante el método descrito anteriormente, se clonó un ADNc de longitud completa que codifica para CRSP-1 humana. La secuencia de nucleótidos que codifica para la proteína CRSP-1 humana de longitud completa se muestra en la figura 1 y se expone como SEQ ID NO: 1. La proteína de longitud completa codificada por este ácido nucleico está comprendida por aproximadamente 350 aminoácidos y tiene la secuencia de aminoácidos mostrada en la figura 1 y expuesta como SEQ ID NO: 2. La parte codificante (marco de lectura abierto) de SEQ ID NO: 1 se expone como SEQ ID NO: 3.

Análisis de CRSP-1 humana

- La determinación del perfil hidrófobo de CRSP-1 humana que tiene la secuencia de aminoácidos expuesta en SEQ ID NO: 2 indica la presencia de una región hidrófoba desde aproximadamente el aminoácido 1 hasta aproximadamente el aminoácido 23 de la SEQ ID NO: 2. El análisis adicional de la secuencia de aminoácidos SEQ ID NO: 2 usando programas de predicción de péptido señal predijo la presencia de un péptido señal desde aproximadamente el aminoácido 1 hasta aproximadamente el aminoácido 19, 21, ó 23 de la SEQ ID NO: 2. Por consiguiente, la proteína CRSP-1 madura incluye aproximadamente 327, 329, ó 331 aminoácidos que se extienden desde aproximadamente el aminoácido 20, 22, ó 24 hasta aproximadamente el aminoácido 350 de SEQ ID NO: 2. La presencia de la secuencia señal, además del hecho de que se identificó CRSP-1 usando un sistema de atrapamiento de secuencia señal, indica que CRSP-1 es una proteína segregada. Es más, puede realizarse la predicción de tal péptido señal y tal sitio de escisión de péptido señal, por ejemplo, utilizando el algoritmo informático SIGNALP (Henrik, *et al.* (1997) Protein Engineering 10:1-6).

- Además, cuando se expresó el ADNc que codifica para CRSP-1 humana en células bacterianas con un marcador FLAG C-terminal, se encontró que el producto de proteína CRSP-1 marcada con FLAG se segregaba al medio celular.

- El examen de la secuencia de ADNc representada en la figura I muestra que CRSP-1 humana es particularmente rica en residuos de cisteína. Tal como se muestra en la figura 1, CRSP-1 contiene 20 residuos de cisteína localizados entre el aminoácido 147 y aminoácido 284 de SEQ ID NO: 2. Estos residuos de cisteína pueden formar posiblemente 10 puentes disulfuro.

- Una búsqueda BLAST (Altschul *et al.* (1990) J. Mol. Biol. 215:403) de las secuencias de nucleótidos y aminoácidos de CRSP-1 ha revelado que CRSP-1 es significativamente similar a ADNc de pollo que codifica para una proteína de función desconocida, número de registro de GenBank D26311. Se aisló este ADNc de una biblioteca de ADNc del cristalino de pollo y mostró que se expresaba en fibras del cristalino y epitelio del cristalino, pero no en la retina neural ni en las células del hígado. (Sawada *et al.* (1996) Int. J. Dev. Biol. 40:531). CRSP-1 y la proteína de pollo tienen una identidad de secuencia de aminoácidos del 56% y una similitud de secuencia de aminoácidos del 72%. La similitud de secuencia de aminoácidos entre la proteína del pollo y CRSP-1 humana es particularmente alta en el dominio rico en cisteína de CRSP-1 que está ubicada entre los aminoácidos 147 y 284 de SEQ ID NO: 2. En particular, los 20 residuos de cisteína de CRSP-1 ubicados en esta región están presentes en la proteína del pollo (véase la figura 3).

- Dos genes recientemente identificados en un examen para detectar supresores de formación de glioblastoma (Ligon *et al.*, 1997 Oncogene 14 1075-1081) también muestran homología significativa con hCRSP-1. Estos genes, RIG ("Regulados en Glioblastoma") y 7-1 de tipo RIG (números de registro de GenBank U32331 y AF034208, respectivamente), se identificaron en un examen diferencial para detectar ARNm regulados por la introducción de una copia normal del cromosoma 10 en una línea celular de glioblastoma que alberga una delección en el cromosoma 10 que estimula tumorigénesis. Se presenta un diagrama esquemático que resume la relación entre las secuencias de los genes de CRSP-1 y RIG como la figura 9. La región indicada de identidad entre CRSP-1, RIG y 7-1 de tipo RIG comprende una región del UTR en 3' del ARNm de CRSP-1 humana. Inicialmente se aisló este ADNc en el examen diferencial descrito y posteriormente se usó para aislar mensajeros de RIG y 7-1 de tipo RIG de longitud completa tal como se describe. 7-1 de tipo RIG es extremadamente homólogo (similar en un 96,8% tal como se determina usando el programa de alineación de proteína Lipman-Pearson, Ktuple (Tupla K): 2; Gap penalty (penalización por hueco): 4, Gap length penalty (penalización por longitud de hueco): 12; y homólogo en un 64,7% usando el programa de alineación de ADN Wilbur Lipman, Ktuple: 3; Gap penalty: 3; Window (intervalo): 20) a CRSP-1, aunque la proteína codificada carece de la secuencia señal N-terminal y por tanto, no se prevé que sea una proteína segregada. El ARNm de RIG de longitud completa sólo presenta homología con CRSP-1 en la región del ADNc inicial. Estos datos asocian CRSP-1 con glioblastoma humano y sugieren que CRSP-1 puede ser importante en la supresión del fenotipo tumoral. También es coherente un papel en el glioblastoma con el alto nivel de expresión de ARNm de CRSP-1 observado en el tejido cerebral humano. Además, la co-localización de los genes de CRSP-1, RIG y de tipo RIG en una región de cromosoma 11 (11p15.1) implicada en el desarrollo de astrocitoma maligno humano (Ligon *et al.* 1997 Oncogene 14 1075-1081) indica adicionalmente un papel para estos genes en tumorigénesis.

La proteína CRSP-1 humana también tiene alguna similitud de secuencia aminoácidos con la metalotioneína, particularmente en el dominio rico en cisteínas.

Distribución tisular de ARNm de CRSP-1

Este ejemplo describe la distribución tisular de ARNm de CRSP-1, tal como se determina mediante hibridación por transferencia de tipo Northern.

Las hibridaciones por transferencia de tipo Northern con las diversas muestras de ARN se realizaron en condiciones habituales y se lavaron en condiciones rigurosas, es decir, 0,2 x SSC a 65°C. En cada muestra, la sonda se hibridó con un ARN sencillo de aproximadamente 2,5 kb. Los resultados de la hibridación de la sonda con diversas muestras de ARNm se describen a continuación.

La hibridación de una transferencia de tipo Northern de tejido múltiple (MTN) fetal de Clontech (Clontech, LaJolla, CA) que contiene ARN de riñón, hígado, pulmón y cerebro fetal indicó la presencia de altos niveles de ARNm de CRSP-1 en pulmón y cerebro fetal y niveles ligeramente inferiores de ARNm de GRSP-1 en riñón fetal. Sin embargo, no se encontró nivel significativo de ARNm de CRSP-1 en el hígado fetal.

La hibridación de una transferencia de tipo Northern de tejido múltiple (MTN) humano de Clontech (Clontech, LaJolla, CA) que contiene ARN de páncreas, riñón, músculo esquelético, hígado, pulmón, placenta, cerebro y corazón adulto con una sonda de CRSP-1 humana indicó la presencia de altos niveles de ARNm de CRSP-1 en el corazón, niveles ligeramente inferiores en el cerebro y niveles mucho más inferiores en la placenta y pulmón. También se encontró algo de ARNm de CRSP-1 en el músculo esquelético adulto. Sin embargo, no se observaron niveles significativos de ARNm de CRSP-1 en el páncreas, riñón o hígado adulto. De manera interesante, el gen de pollo que era homólogo a CRSP-1 no se expresó a niveles detectables en ninguno de los hígados (Sawada *et al.* (1996) *Int. J. Dev. Biol.* 40:531).

Hibridación adicional de una transferencia de tipo Northern de tejido múltiple (MTN) humano de Clontech (Clontech, LaJolla, CA) que incluye ARN de médula espinal adulta reveló niveles elevados de expresión de CRSP-1 en ARNm aislado de médula espinal adulta.

Un análisis *in situ* reveló adicionalmente los siguientes patrones de expresión cuando hibridan secciones tisulares con sondas de CRSP-1. Se expresó ARNm de CRSP-1 en el corazón, pulmón, aorta, ojo (retina y cristalino) y cerebro (corteza) en embriones de ratón de 14,5 días. Se expresó ARNm de CRSP-1 en el corazón, pulmón, ojo (retina y cristalino), cerebro (cortical), y cartílago (pie, traquea, laringe, cabeza, esternón y médula espinal) en tejidos de ratones post-natales de 1,5 días. Se expresó ARNm de CRSP-1 en el cerebro (corteza, hipocampo, tronco encefálico), corazón (sólo aurícula), ojo (retina y capa exterior del cristalino) en tejidos de ratones adultos.

Por tanto, CRSP-1 se expresa de manera específica de tejido, observándose la mayor expresión en el cerebro, corazón y médula espinal.

Datos de expresión de proteína CRSP-1

Constructos

Se prepararon constructos de expresión para dos formas de CRSP-1 usando el vector de expresión pMET-stop de mamífero. La forma 1 comprendía un ADNc que incorpora la secuencia codificante de proteína CRSP-1 de 350 aa (CRSP-1 flag.larga) y la forma 2 comprendía la secuencia codificante de CRSP-1 completa salvo por los últimos 18 aminoácidos (CRSP-1 flag.corta). Se añadió una secuencia C-terminal que codifica para el epítipo FLAG (DYKDDDDK) a ambas formas de CRSP-1 para facilitar la detección y purificación. Se generaron ADNc de CRSP-1 flag mediante PCR a partir de un molde de ADNc de CRSP-1 de longitud completa y se ligaron a pMET-stop usando sitios de restricción EcoRI y SalI.

Transfección de prueba - expresión a pequeña escala

Se transfectaron constructos de expresión para CRSP-1 flag.larga y CRSP-1 flag.corta en células 293T usando 10 µl de lipofectamina (GIBCO/BRL) y 2 µg de ADN por pocillo de una placa de 6 pocillos de células que eran confluentes en un 70-80%. Tras 5 horas a 37°C, se alimentaron las células con 1 ml de FCS/DMEM al 20%. Tras la incubación durante la noche a 37°C, se acondicionaron las células en 1 ml de OptiMEM durante 48 horas a 37°C. Se solubilizaron las muestras de sobrenadante y sedimentos celulares en tampón de gel SDS-PAGE hirviendo, se eluyeron en un gel SDS-PAGE al 4-20%, se transfirieron a una membrana de nylon y se exploraron con sonda con el anticuerpo anti-FLAG monoclonal M2. Las muestras tanto del sobrenadante como del sedimento mostraron inmunoreactividad significativa en un intervalo de peso molecular de 40-65 kDa sobre película autorradiográfica usando un anticuerpo secundario conjugado HRP y reactivos de detección de ECL. Por tanto, se segregan ambas formas de CRSP-1 sometidas a prueba a partir de células 293T confirmando así experimentalmente que CRSP-1 es una proteína segregada. Debe observarse que los pesos moleculares de ambas formas de CRSP-1 sometidas a prueba son superiores de lo previsto a partir de la secuencia de aminoácidos, lo que sugiere que las proteínas CRSP-1 segregadas por células 293T pueden glucosilarse. Esto concuerda con la presencia de cuatro sitios potenciales para glucosilación unida en N en la proteína CRSP-1.

Producción de proteína CRSP-1 a gran escala

Para la expresión de CRSP-1 flag.larga a gran escala, se transfectaron a 30 x 150 mM, placas de células 293T con una confluencia del 70-80% con 27 µg de ADN, 100 µl de lipofectamina en 18 ml de OptiMEM durante 5 horas a 37°C. Se añadieron 18 ml de FCS/DMEM al 10% a cada placa y se incubaron durante la noche a 37°C. Veinticuatro horas tras el comienzo de la transfección, se aspiró el sobrenadante de transfección y se añadieron 35 ml de OptiMEM a cada placa y se incubaron las placas a 37°C durante 72 horas. Se recogió el medio condicionado, se centrifugó a 4000 rpm durante 30 min a 4°C, y se filtró a través de una unidad de filtro de 0,45 micras. Se pasaron 1100 ml a través de una columna de afinidad anti-FLAG M2 de 1,6 x 10 cm preequilibrada en tampón PBS, pH 7,4 a una velocidad de flujo de 2,0 ml por minuto. Tras lavar con 200 ml de tampón PBS pH 7,4, se eluyó el material unido mediante una etapa de tampón Glicina 200 mM, pH 3,0 y se recogieron fracciones de 0,5 ml. Tras la elución, se detectó un pico de proteína significativo por absorbancia a 280 nm. Se analizaron las muestras correspondientes al medio condicionado, las fracciones de flujo a través y eluidas mediante SDS-PAGE teñidas con plata y azul de Coomassie y mediante análisis por inmunotransferencia de tipo western tal como se describió anteriormente. Se detectó inmunoreactividad significativa en un intervalo de peso molecular de 40-65 kDa en medio condicionado y en fracciones eluidas pero no en la muestra de flujo a través, lo que indica que la proteína CRSP-1 flag.larga segregada se unió específicamente a la columna de afinidad y se eluyó eficazmente mediante las condiciones descritas. La tinción de geles de SDS-PAGE con azul de Coomassie sugirió que la proteína inmunoreactiva predominante constituía >90% de la proteína presente en el pico de proteína unida y eluida. Se juntaron las fracciones de pico de la proteína eluida y se sometieron a diálisis frente a solución salina tamponada de fosfato dando como resultado un volumen de 4 ml de proteína CRSP-1 flag.larga a una concentración de aproximadamente 1 mg/ml.

Ejemplo 2

25 *Aislamiento y caracterización de ADNc de CRSP-2, CRSP-3, y CRSP-4 humanas*

Para identificar proteínas novedosas relacionadas con CRSP-1, se usó la secuencia de aminoácidos de CRSP-1 humana para buscar las bases de datos privadas y dbEST usando TBLASTN (WashUversion, 2.0, matriz de búsqueda BLOSUM62). A partir de esta búsqueda, se identificaron cuatro secuencias de proteínas parciales distintas que presentaban homología significativa con CRSP-1 humana. Se diseñaron estas secuencias como secuencias parciales para las CRSP-h novedosas; CRSP-2-h, CRSP-3-h, CRSP-4-h y N terminal de tipo h-CRSP.

Se derivó la secuencia de proteína parcial hCRSP-2 de un clon de dbEST sencillo con número de registro AA565546. Este clon se obtuvo posteriormente de la colección IMAGE y se secuenció totalmente para definir la secuencia de hCRSP-2 completa representada en la figura 2.

Se derivó la secuencia de proteína parcial hCRSP-3 de un clon de jthKb075a10 con un código de identificación de secuencia jthKb075a10. Este clon se secuenció adicionalmente y se combinó esta información de secuencia con secuencias adicionales de dbEST usando el programa de reunión Phrap (P. Green, U. Washington) para definir la secuencia de hCRSP-3 completa representada en la figura 3. Se depositó el ADN para el clon de jthKb075a10 en la ATCC con el número de registro 98633. El análisis de transferencia tipo Northern de diversos tejidos usando una sonda específica para hCRSP-3 reveló que el ARNm de CRSP-3 se limitaba a la placenta. La expresión de ARNm de CRSP-3 todavía no se ha demostrado en otros tejidos humanos adultos normales.

Recientemente se ha descrito un homólogo murino y de *Xenopus* de CRSP-3 denominado dkk-1 ("dickkopf", alemán, cabeza dura) que se propone que codifica para una molécula de señalización segregada (Nature, 391, 357-368). Se describe dkk-1 como un inductor potente de la formación de la cabeza durante el desarrollo temprano de *Xenopus*, suponiendo aparentemente su mecanismo de acción la inhibición de la familia *wnt* de factores segregados.

Dada la homología entre CRSP-3 y dkk-1, se sugiere que hCRSP-3 puede presentar profunda actividad de inducción de la cabeza durante el desarrollo embrionario humano temprano. Además, se piensa que los mecanismos de acción de dkk-1 implican antagonismo de miembros de la familia *wnt* de proteínas segregadas. Se ha asociado la sobreexpresión de proteínas *wnt* con ciertos tipos de cáncer. Por ejemplo, la activación de la expresión de *wnt-1* mediante inserción proviral de MMTV provoca cáncer de mama en ratón (véase Cadigan y Nusse, Genes y Dev., 1997 11 3286-3305). Por tanto, al menos CRSP-3, y potencialmente otros miembros de la familia de CRSP, también pueden tener un papel en la supresión de señalización de *wnt* en cáncer.

Se derivó la secuencia de proteína parcial CRSP-4-h de un clon de dbEST sencillo con número de registro W55979. Posteriormente se obtuvo este clon de la colección IMAGE y se secuenció totalmente para definir una secuencia de CRSP-4-h mostrada en la figura 4. El análisis por transferencia tipo Northern de diversos tejidos usando una sonda específica para CRSP-4-h reveló que el ARNm de CRSP-4 se expresaba en diversos tejidos humanos adultos. La expresión de ARNm de CRSP-4 fue superior en el corazón, cerebro, placenta, pulmón y músculo esquelético.

Se derivó la secuencia de proteína parcial N terminal de tipo CRSP humana de un clon de dbEST sencillo con número de registro AA397836. Este clon se obtuvo posteriormente de la colección IMAGE y se secuenció totalmente para definir la secuencia de N terminal de tipo CRSP mostrada en la figura 5.

Estructura de las proteínas de la familia CRSP

En la figura 6 se muestra una alineación de las secuencias de aminoácidos de CRSP-1 humana, CRSP-2 humana, CRSP-3 humana, y CRSP-4 humana. Los residuos de aminoácidos que se conservan entre los miembros de la familia de CRSP están recuadrados. Los 20 residuos de cisteína conservados están indicados mediante un asterisco. Los dominios ricos en cisteína están indicados como CRD-1 y CRD-2. La estructura del dominio de las proteínas CRSP de longitud completa de la presente invención se muestra en la figura 7. La homología de aminoácidos y nucleótidos entre los miembros de la familia de CRSP es tal como sigue en las tablas 1 y 2.

TABLA 1

	CRSP-1	CRSP-2	CRSP-3	CRSP-4	mdkk-1	xdkk-1	CLFEST
CRSP-1	100	16,0	18,6	15,1	18,5	16,5	53,0
CRSP-2		100	33,7	35,2	32,6	33,7	16,2
CRSP-3			100	33,1	80,2	53,5	17,4
CRSP-4				100	30,5	33,7	12,5

Los porcentajes de homología de aminoácidos se determinaron usando el programa ALIGN, (versión 2.0) (paquete de software GCG) usando una tabla de residuos en peso PAM120, gap length penalty (penalización de longitud de hueco) de 12, y gap penalty (penalización de hueco) de 4.

TABLA 2

	CRSP-1	CRSP-2	CRSP-3	CRSP-4	mdkk-1	xdkk-1	CLFEST
CRSP-1	100	30,0	37,2	34,7	31,5	45,4	58,8
CRSP-2		100	43,0	35,9	38,8	38,4	36,7
CRSP-3			100	59,3	66,4	53,7	32,1
CRSP-4				100	38,8	38,4	36,7

Los porcentajes de homología de nucleótidos se determinaron usando el programa de alineación de ADN Wilbur Lipman, Ktuple: 3; Gap Penalty (penalización de hueco): 3; Window (intervalo): 20.

REIVINDICACIONES

1. Molécula de ácido nucleico aislada seleccionado del grupo constituido por:

a) un ácido nucleico que comprende la SEQ ID NO: 3 o la SEQ ID NO: 9;

b) un ácido nucleico que codifica para un polipéptido que comprende la secuencia de aminoácidos de los residuos de aminoácidos 24-350 de SEQ ID NO: 2, o la secuencia de aminoácidos de los residuos de aminoácidos 21-266 de SEQ ID NO: 8;

c) un ácido nucleico que codifica para un polipéptido que comprende una variante alélica que se produce de manera natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5, SEQ ID NO: 11, o SEQ ID NO: 14, en el que dicha variante alélica que se produce de manera natural modula la transducción de la señal celular o la proliferación celular, y en el que (i) dicho ácido nucleico que codifica para una variante alélica que se produce de manera natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5 hibrida con el complemento de SEQ ID NO: 6 en condiciones de lavado rigurosas, (ii) dicho ácido nucleico que codifica para una variante alélica que se produce de manera natural de la secuencia de aminoácidos de SEQ ID NO: 11 hibrida con el complemento de SEQ ID NO: 12 en condiciones de lavado rigurosas, (iii) dicho ácido nucleico que codifica para una variante alélica que se produce de manera natural de la secuencia de aminoácidos de SEQ ID NO: 14 hibrida con el complemento de SEQ ID NO: 15 en condiciones rigurosas de lavado, en el que las condiciones rigurosas de lavado son 0,2 X SSC, SDS al 0,1% a 50-65°C;

d) un ácido nucleico que comprende una secuencia de nucleótidos que es el complemento de a), b), o c); y

e) un vector de expresión de célula eucariota recombinante que comprende el ácido nucleico de a), b), c) o d).

2. Molécula de ácido nucleico aislada según la reivindicación 1, en la que la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos seleccionada del grupo constituido por SEQ ID NO: 1, y SEQ ID NO: 3.

3. Molécula de ácido nucleico aislada según la reivindicación 1, en la que dicha molécula de ácido nucleico aislada comprende una secuencia de nucleótidos seleccionada del grupo constituido por SEQ ID NO: 7, SEQ ID NO: 9 y la secuencia de nucleótidos del inserto de ADN del plásmido depositada en la ATCC con el número de registro 98633.

4. Molécula de ácido nucleico aislada según la reivindicación 1, en la que dicha molécula de ácido nucleico aislada codifica para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 2, los residuos de aminoácidos 20-350 de SEQ ID NO: 2, los residuos de aminoácidos 22-350 de SEQ ID NO: 2 y los residuos de aminoácidos 24-350 de SEQ ID NO: 2.

5. Molécula de ácido nucleico aislada según la reivindicación 1, en la que dicha molécula de ácido nucleico aislada codifica para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 8 y los residuos de aminoácidos 21-266 de SEQ ID NO: 8.

6. Molécula de ácido nucleico aislada según la reivindicación 1a), 1b) o 1c), en la que dicha molécula de ácido nucleico aislada además comprende:

i) una secuencia de ácido nucleico vector; y/o

ii) una secuencia de ácido nucleico que codifica para un polipéptido heterólogo.

7. Molécula de ácido nucleico aislada según la reivindicación 1e), en la que dicha molécula de ácido nucleico aislada además comprende una secuencia de ácido nucleico que codifica para un polipéptido heterólogo.

8. Célula huésped que comprende el vector de expresión de célula eucariota recombinante según la reivindicación 1e).

9. Célula huésped que comprende un vector que comprende el ácido nucleico de una cualquiera de las reivindicaciones 1a), 1b), 2, 3, 4 ó 5 en la que dicho ácido nucleico es un ácido nucleico recombinante.

10. Polipéptido aislado seleccionado del grupo constituido por:

a) un polipéptido que comprende al menos 15 aminoácidos contiguos de las SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14, polipéptido que modula la transducción de señales celulares o la proliferación celular;

b) un polipéptido que comprende una variante alélica que se produce de manera natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5, los residuos de aminoácidos 21-266 de SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14, en el que dicha variante alélica que se produce de manera natural modula la transducción de señales celulares o la proliferación celular y en el que (i) dicha variante alélica que se produce de

manera natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 6 en condiciones de lavado rigurosas, (ii) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de SEQ ID NO: 8 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 9 en condiciones de lavado rigurosas, (iii) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de SEQ ID NO: 11 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 12 en condiciones de lavado rigurosas, y (iv) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de SEQ ID NO: 14 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 15 en condiciones de lavado rigurosas, y en el que las condiciones de lavado rigurosas son 0,2 X SSC, SDS al 0,1% a 50-65°C;

c) un polipéptido que comprende los residuos de aminoácidos 24-350 de SEQ ID NO: 2; y

d) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 80% de identidad de secuencia de aminoácidos con la secuencia de aminoácidos de SEQ ID NO: 5, SEQ ID NO: 8 o SEQ ID NO: 11 sobre la longitud completa de SEQ ID NO: 5, SEQ ID NO: 8 o SEQ ID NO: 11, respectivamente, y modula la transducción de señales celulares o la proliferación celular.

11. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 15 aminoácidos contiguos de SEQ ID NO: 8 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

12. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 20 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 11 o SEQ ID NO: 14 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

13. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 30 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 11 o SEQ ID NO: 14 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

14. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 50 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 11 o SEQ ID NO: 14 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

15. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 100 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 11 o SEQ ID NO: 14 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

16. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 20 aminoácidos contiguos de SEQ ID NO: 8 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

17. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 30 aminoácidos contiguos de SEQ ID NO: 8 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

18. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 50 aminoácidos contiguos de SEQ ID NO: 8 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

19. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido comprende al menos 100 aminoácidos contiguos de SEQ ID NO: 8 y en el que dicho polipéptido modula la transducción de señales celulares o la proliferación celular.

20. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido aislado comprende una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 8 y los residuos de aminoácidos 21-266 de SEQ ID NO: 8.

21. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido aislado comprende una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 2, los residuos de aminoácidos 20-350 de SEQ ID NO: 2, los residuos de aminoácidos 22-350 de SEQ ID NO: 2, y los residuos de aminoácidos 24-350 de SEQ ID NO: 2.

22. Polipéptido aislado según la reivindicación 10, en el que dicho polipéptido aislado comprende una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 5, los residuos de aminoácidos 20-224 de SEQ ID NO: 5, SEQ ID NO: 11 y SEQ ID NO: 14.

ES 2 285 768 T3

23. Polipéptido aislado según una cualquiera de las reivindicaciones 10-22, que además comprende un polipéptido heterólogo.

24. Anticuerpo o parte inmunológicamente activa del mismo que se une específicamente a un polipéptido seleccionado del grupo constituido por:

a) un polipéptido constituido por 15 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14; y

b) un polipéptido constituido por una variante alélica que se produce de manera natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5, los residuos de aminoácidos 21-266 de SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14, en el que dicha variante alélica que se produce de forma natural modula la transducción de señales celulares o la proliferación celular y en el que (i) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 6 en condiciones de lavado rigurosas, (ii) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de los residuos de aminoácidos 21-266 de SEQ ID NO: 8 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 9 en condiciones de lavado rigurosas, (iii) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de SEQ ID NO: 11 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 12 en condiciones de lavado rigurosas, (iv) dicha variante alélica que se produce de forma natural de la secuencia de aminoácidos de SEQ ID NO: 14 está codificada por una molécula de ácido nucleico que hibrida con el complemento de SEQ ID NO: 15 en condiciones de lavado rigurosas, y en el que las condiciones de lavado rigurosas son 0,2 X SSC, SDS al 0,1% a 50-65°C.

25. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 24, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido constituido por 15 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14.

26. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 24, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido constituido por 20 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14.

27. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 24, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido constituido por 30 aminoácidos contiguos de SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 o SEQ ID NO: 14.

28. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 24, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido que está constituido por una secuencia de aminoácidos seleccionada del grupo constituido por SEQ ID NO: 5, SEQ ID NO: 8, SEQ ID NO: 11 y SEQ ID NO: 14.

29. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 28, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido que está constituido por la secuencia de aminoácidos de los residuos de aminoácidos 21-266 de SEQ ID NO: 8.

30. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 28, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se une específicamente a un polipéptido que está constituido por la secuencia de aminoácidos de los residuos de aminoácidos 20-224 de SEQ ID NO: 5.

31. Anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-30, en el que dicho anticuerpo o parte inmunológicamente activa del mismo se selecciona del grupo constituido por un anticuerpo humano, un anticuerpo monoclonal humanizado, un anticuerpo monoclonal quimérico y una parte inmunológicamente activa de uno de los anteriores.

32. Anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-30, en el que dicho anticuerpo o parte inmunológicamente activa del mismo es un anticuerpo monoclonal o una parte inmunológicamente activa de un anticuerpo monoclonal.

33. Anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-30, en el que dicho anticuerpo o parte inmunológicamente activa del mismo es una parte inmunológicamente activa seleccionada del grupo constituido por un fragmento Fab y un fragmento F(ab')₂.

34. Anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-33, en el que dicho anticuerpo o parte inmunológicamente activa del mismo comprende un marcador detectable.

35. Anticuerpo o parte inmunológicamente activa del mismo según la reivindicación 34, en el que dicho marcador detectable es una enzima, un grupo prostético, un material fluorescente, un material luminiscente, un material bioluminiscente o un material radiactivo.

36. Método para producir un polipéptido según la reivindicación 10, que comprende poner en cultivo la célula huésped según la reivindicación 8 en condiciones en las que se expresa el vector de expresión recombinante y se produce dicho polipéptido.

5 37. Método para producir un polipéptido según la reivindicación 20 o la reivindicación 21, que comprende poner en cultivo la célula huésped según la reivindicación 9 en condiciones en las que se expresa el ácido nucleico recombinante, y se produce dicho polipéptido.

10 38. Método para detectar la presencia de un polipéptido según una cualquiera de las reivindicaciones 10-23, en una muestra que comprende:

a) poner en contacto la muestra con un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35; y

15 b) determinar si dicho anticuerpo se une a un polipéptido en la muestra para así detectar la presencia de dicho polipéptido según una cualquiera de las reivindicaciones 10-23 en la muestra.

20 39. Kit que comprende un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35 e instrucciones para su uso.

40 40. Uso de un kit para detectar un polipéptido según una cualquiera de las reivindicaciones 10-23 en una muestra, comprendiendo dicho kit un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35 e instrucciones para su uso.

25 41. Método para detectar la presencia de una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-5 en una muestra que comprende:

30 a) poner en contacto la muestra con un cebador o sonda de ácido nucleico que hibrida específicamente con una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-5; y

b) determinar si el cebador o sonda de ácido nucleico se une a una molécula de ácido nucleico en la muestra para así detectar la presencia de una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-5 en la muestra.

35 42. Método según la reivindicación 41, en el que la muestra comprende moléculas de ARNm y se pone en contacto con una sonda de ácido nucleico.

40 43. Uso de un kit para detectar una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-5 en una muestra, comprendiendo dicho kit un cebador o sonda de ácido nucleico que hibrida específicamente con una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-5 e instrucciones para su uso.

44. Método *in vitro* para identificar un compuesto que se une a un polipéptido según una cualquiera de las reivindicaciones 10-23 que comprende:

45 a) poner en contacto un polipéptido según una cualquiera de las reivindicaciones 10-23, o una célula que expresa dicho polipéptido con un compuesto de prueba; y

b) determinar si el polipéptido se une con el compuesto de prueba.

50 45. Método según la reivindicación 44, en el que se detecta la unión del compuesto de prueba con el polipéptido mediante un método seleccionado del grupo constituido por:

a) detección de unión mediante detección directa de la unión compuesto de prueba/polipéptido;

55 b) detección de unión usando un ensayo de unión por competencia; y

c) detección de unión usando un ensayo para detectar la transducción de señales celulares o la proliferación celular.

60 46. Método *in vitro* para modular la transducción de señales celulares o la actividad de proliferación celular de un polipéptido según una cualquiera de las reivindicaciones 10-23, que comprende poner en contacto un polipéptido según una cualquiera de las reivindicaciones 10-23 o una célula que expresa dicho polipéptido con un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35 en una concentración suficiente para modular la transducción de señales celulares o la actividad de proliferación celular de dicho polipéptido.

65 47. Método para identificar un compuesto que modula la transducción de señales celulares o la actividad de proliferación celular de un polipéptido según una cualquiera de las reivindicaciones 10-23 que comprende:

ES 2 285 768 T3

a) poner en contacto un polipéptido según una cualquiera de las reivindicaciones 10-23 con un compuesto de prueba; y

5 b) determinar el efecto del compuesto de prueba sobre la transducción de señales celulares o la actividad de proliferación celular de dicho polipéptido para identificar así un compuesto que modula la transducción de señales celulares o la actividad de proliferación celular de dicho polipéptido.

10 48. Anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35, para su uso en tratamiento o diagnóstico.

49. Uso de un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35, para la fabricación de un medicamento para el tratamiento de un trastorno del desarrollo, de diferenciación o proliferativo.

15 50. Polipéptido según una cualquiera de las reivindicaciones 10-23, para su uso en tratamiento o diagnóstico.

51. Uso de un polipéptido según una cualquiera de las reivindicaciones 10-23, para la fabricación de un medicamento para el tratamiento de un trastorno del desarrollo, de diferenciación o proliferativo.

20 52. Uso de un polipéptido según una cualquiera de las reivindicaciones 10-23 para la fabricación de un medicamento o diagnóstico para el tratamiento, profilaxis o diagnóstico del cáncer.

25 53. Molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-7, para su uso en tratamiento o diagnóstico.

54. Uso de una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-7 para la fabricación de un medicamento para el tratamiento de un trastorno del desarrollo, de diferenciación o proliferativo.

30 55. Uso de un anticuerpo o parte inmunológicamente activa del mismo según una cualquiera de las reivindicaciones 24-35, para la fabricación de un medicamento o diagnóstico para el tratamiento, profilaxis o diagnóstico del cáncer.

56. Uso de una molécula de ácido nucleico según una cualquiera de las reivindicaciones 1-7, para la fabricación de un medicamento o diagnóstico para el tratamiento, profilaxis o diagnóstico del cáncer.

35 57. Célula huésped no humana que comprende un transgén que comprende la molécula de ácido nucleico según una cualquiera de las reivindicaciones 1a), 1b), 2, 3, 4 ó 5.

40

45

50

55

60

65

Figura 1A

[illegible]

Figura 1 B

ATCTGGACCAAGCTGTGGTAGA TGTGCAATAGAAATAGCTAATTTATTTCCCCANGTGTGTCTTTAAGCGTGGGCTG 1169
ACCAGGCTTCTTCTACATCTTCTTCCCACTAAGTTTCCCTCTGGCTTGCACAGCATGAGGTCTTGTGCATTTGTTTCAG 1248
CTCCCCCAGGCTGTTCTCCAGGCTTCACAGTCTGGTGTCTGGGAGAGTTCAGGCAGGGTTTAACTGCAGGAGCAGTTTGC 1327
CACCCCTGTCAGATTATTGGCTGCTTTGCCCTCTACCAAGTTGGCAGACAGCCGTTTGTCTACATGGCTTTGATAATTG 1406
TTTGAGGGAGGAGATGGAACAATGTGCAGTCTCCCTCTGATTGGTTTGGGGAAATGTGGAGAGAGTCCCCTGCTT 1485
TGCAAAACATCAACCTGGCAAAAATGCNACAAATGAAATTTTCCAGGCAGTTCTTCCATGGGCATAGGTAAAGCTGTGCCT 1564
TCAGCTGTTGCAGATGAATGTTCTGTTTCCCTGCTTACATGTGTTTATTCATCCAGCAGTGTGCTCAGCTCCTAC 1643
CTCTGTGCCAGGGCAGCATTTTCATATCCAAGATCAATTCCTCTCTCAGCACAGCCCTGGGAGGGGTCATTGTTCTC 1722
CTCGTCCATCAGGCAATTCAGAGGCTCAGAGACTGCAAGCTGCTTGCCCCAAGTCAACACAGCTAGTGAAGACCAGAGCAG 1801
TTTCATCTGGTTGTGACTCTAAGCTCAGTGTCTCTCCACTACCCCAACACAGCCCTTGGTGCCACCAAAAGTGTCTCCCC 1880
AAAAGGAAGGAGAAATGGGATTTTCTTTTGAGGCATGCACATCTGGAAATTAAGGTCAAACTAAATTTCTACATCCCTCTA 1959
AAAGTAAACTACTGTTAGGAACACAGCAGTGTCTCAGAGTGTGGGGCAGCCGCTCTTAATGAAGACAAATGATATTGAC 2038
ACTGTCCCTCTTTGGCAGTTGCATTAGTAACCTTTGAAAGGTATATGACTGAGCGTAGCATACAGGTTAACCTGCAGAAA 2117
CAGTACTTAGGTAAATTGTAGGGGAGGATTATAAATGAAATTTGCCAAATCACCTTAGCAGCAACTGAAGACAAATTATCA 2196
ACCACGTGGAGAAAATCAAA CCGAGCAGGGCTGTGTGAACATGGTTGTAAATATGCGACTGCGAACACTGAACCTCTACG 2275
CCACTCCACAAATGATGTTTTTCAGGTGTCATGGAGCTGTTGCCACCATGTATTATCCAGAGTTCTTAAAGTTTAAAGTT 2354
GCACATGATTGTATAAGCATGCTTCTCTTTCAGCTTTTAAATTTATGTATAAACATAAGTTGCTCATTTAGAAATCAAGCATAA 2433
ATCACTTCACTGCTAA 2479

Figura 2

GAATTCGGCAGAGACGACGCTGCTGAGCTGCCAGCTTAGTGAAGCTCTGCTGGTGAGAGACGCTCGCTTTG 79
 M V A A V L L G 8
 GTGAGGCACAGTCTGGACCTCCAGGAGCCCGGCGATTGAGG ATG GTG GCG GCC GTC CTG GGG 148
 L S W L C S P L G A L V L D F H N I R S 28
 CTG AGC TGG CTC TCC TCT CCC CTG GGA GCT CTG GTC GAC TTC AAC AAC ATC AGG AGC 208
 S A D L H G A R K G S Q C L S D T D C N 48
 TCT GCT GAC CTG CAT GGG GCC CGG AAG GGC TCA CAG TCC CTG TCT GAC AGC GAC TGC AAT 268
 T R K F C L Q P R D E K P F C A T C R G 68
 ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT GCT ACA TGT CGT GGG 328
 L R R R C Q R D A M C C P G T L C V N D 88
 TTG CGG AGG AGG TGC CAG CGA GAT GCC ATG TGC TGC CCT GGG ACA CTC TGT GTG AAC GAT 388
 V C T T M E D A T P I L E R Q L D E Q D 108
 GTT TGT ACT ACG ATG GAA GAT GCA ACC CCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT 448
 G T H A E G T T G H P V Q E N Q P K R K 128
 GGC ACA CAT GCA GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG AAG 508
 P S I K K S Q G R K G Q E G E S C L R T 148
 CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA AGT TGT CTG AGA ACT 568
 F D C G P G L C C A R H F W T K I C K P 168
 TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CGT CAT TTT TGG AGG AAA ATT TGT AAG CCA 628
 V L L E G Q V C S R R G H K D T A Q A P 188
 GTC CTT TTG GAG GGA CAG GTC TGC TCC AGA AGA GGG CAT AAA GAC ACT GCT CAA GCT CCA 688
 E I F Q R C D C G P G L L C R S Q L T S 208
 GAA ATC TTC CAG CGT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC AGC 748
 N R Q H A R L R V C Q K I E K L * 225
 AAT CGG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG CTA TAA 799
 ATATTTCAAAATAAAGACGAATCCACATTTCACAAAAAANAAAAA 848

Figura 3A

```

CCGGACGCGTGGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTTCTCTTTCTTCTCCCTCTT 79
      M M A L G A A G A T R V F V A M 16
GAGTCCTTCTGAG ATG ATG GCT CTG GGC GCA GCG GGA GCT ACC CGG GTC TTT GTC GCG ATG 140
V A A A L G G H P L L G V S A T L N S V 36
GTA GCG GCG GCT CTC GGC GGC CAC CCT CTG CTG GGA GTG AGC GCC ACC TTG AAC TCG GTT 200
L N S N A I K N L P P P L G G A A G H P 56
CTC AAT TCC AAC GCT ATC AAG AAC CTG CCC CCA CCG CTG GGC GGC GCT GCG GGG CAC CCA 260
G S A V S A A P G I L Y P G G N K Y Q T 76
GGC TCT GCA GTC AGC GCC GCG CCG GGA ATC CTG TAC CCG GGC GGG AAT AAG TAC CAG ACC 320
I D N Y Q P Y P C A E D E E C G T D E Y 96
ATT GAC AAC TAC CAG CCG TAC CCG TGC GCA GAG GAC GAG GAG TGC GGC ACT GAT GAG TAC 380
C A S P T R G G D A G V Q I C L A C R K 116
TGC GCT AGT CCC ACC CCG GGA GGG GAC GCA GGC GTG CAA ATC TGT CTC GCC TGC AGG AAG 440
R R K R C M R H A M C C P G N Y C K N G 136
CGC CGA AAA CGC TGC ATG CGT CAC GCT ATG TGC TGC CCC GGG AAT TAC TGC AAA AAT GGA 500
I C V S S D Q N H F R G E I E E T I T E 156
ATA TGC GTG TCT TCT GAT CAA AAT CAT TTC CGA GGA GAA ATT GAG GAA ACC ATC ACT GAA 560
S F G N D H S T L D G Y S R R T T L S S 176
AGC TTT GGT AAT GAT CAT AGC ACC TTG GAT GGG TAT TCC AGA AGA ACC ACC TTG TCT TCA 620
K M Y H T K G Q E G S V C L R S S D C A 196
AAA ATG TAT CAC ACC AAA GGA CAA GAA GGT TCT GTT TGT CTC CGG TCA TCA GAC TGT GCC 680
S G L C C A R H F W S K I C K P V L K E 216
TCA GGA TTG TGT TGT GCT AGA CAC TTC TGG TCC AAG ATC TGT AAA CCT GTC CTG AAA GAA 740
G Q V C T K H R R K G S H G L E I F Q R 236
GGT CAA GTG TGT ACC AAG CAT AGG AGA AAA GGC TCT CAT GGA CTA GAA ATA TTC CAG CGT 800
C Y C G E G L S C R I Q K D H H Q A S N 256
TGT TAC TGT GGA GAA GGT CTG TCT TGC CGG ATA CAG AAA GAT CAC CAT CAA GCC AGT AAT 860
S S R L H T C Q R H * 267
TCT TCT AGG CTT CAC ACT TGT CAG AGA CAC TAA 893
ACCAGCTATCCAAATGCAGTGAACCTCTTTTATATAATAGATGCTATGAAAACCTTTTATGACCTTCATCAACTCAAT 972
CCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAACACCTTCCAAAAACCTGGAGTGTAAAGAGCTTTG 1051
TTTCTTTATGGAACCTCCCTGTGATTGCAGTAAATTAAGTATTGTAAATTCAGTGTGGCACTTACCTGTAAATGCA 1130
ATGAACTTTTAATTATTTTCTAAAGGTGCTGCACTGCCTATTTTCTCTGTTTGTAAATTTTGTACACATTGA 1209
TTGTTATCTTGACTGACAAATATTCTATATTGAAGTGAAGTAAATCATTTTCAGCTTATAGTTCTTAAAAGCATAACCTT 1288
TTACCCCATTTNATTCTAGAGTCNAGAACGCAAGGATCTCTTGGAAATGACAAATGATAGGTACCTAAAATGTAACATGA 1367

```

Figura 3B

```

AAATACTAGCTTATTTCTGAAATGTACTATCTTAAATGCTTAAATTATATTTCCCTTTAGGCTGTGATAGTTTTTGAAA 1446
TAAATTTAACATTTAATATCATGAAATGKKTATAAGTAGACATAAAAAAAAAAAAAAAAAAGGGCGCGCTAGA 1525
CTAG 1529

```

Figura 4

E F G T R V G R Y C H S P H Q G S S A C 20
GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA TCA TCG GCC TGC 60
M V C R R K K K R C H R D G M C C P S T 40
ATG GTG TGT CGG AGA AAA AAG AAG CGC TGC CAC CGA GAT GGC ATG TGC TGC CCC AGT ACC 120
R C N N G I C I P V T E S I L T P H I P 60
CGC TGC AAT AAT GGC ATC TGT ATC CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG 180
A L D G T R H R D R N H G H Y S N H D L 80
GCT CTG GAT GGT ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG 240
G W Q N L G R P H T K M S H I K G H E G 100
GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA GGG CAT GAA GGA 300
D P C L R S S D C I E G F C C A R H F W 120
GAC CCC TGC CTA CGA TCA TCA GAC TGC ATT GAA GGG TTT TGC TGT GCT CGT CAT TTC TGG 360
T K I C K P V L H Q G E V C T K Q R K K 140
ACC AAA ATC TGC AAA CCA GTG CTC CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG 420
G S H G L E I F Q R C D C A K G L S C K 160
GGT TCT CAT GGG CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA 480
V W K D A T Y S S K A R L H V C Q K I * 180
GTA TGG AAA GAT GCC ACC TAC TCC TCC AAA GCC AGA CTC CAT GTG TGT CAG AAA ATT TGA 540
TCACCATTGAGGAACATCATCAATTGCACACTGTGAAGTTGTGTATTTAATGCATTATAGCATGGTGGAAAAATAGGTT 619
CAGATGCAGAAAGATGGCTAAAAATAAGAAACGTGATAAGAATATAGATGATCACAAAAAATAAAAAAAGATGCGG 698
CCGC 702

Figura 5

CTCGAGGCCAAAAATTCGGCACGAGGCCGGGCTGTGGTCTAGCATAAAGGCGGAGCCCAGAAGAAGGGGCGGGGT	M	1
ATG		77
G E A S P P A P A R R H L L V L L L L L		21
GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG CAT CTG CTG GTC CTG CTG CTG CTC CTC		137
S T L V I P S A A A P I H D A D A Q E S		41
TCT ACC CTG GTG ATC CCC TCC GCT GCA GCT CCT ATC CAT GAT GCT GAC GCC CAA GAG AGC		197
S L G L T G L Q S L L Q G F S R L F L K		61
TCC TTG GGT CTC ACA GGC CTC CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG AAA		257
G N L L R G I D S L F S A P M D F R G L		81
GGT AAC CTG CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG GAC TTC CGG GGC CTC		317
P G N Y H K E E N Q E H Q L G N N T L S		101
CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG CAC CAG CTG GGG AAC AAC ACC CTC TCC		377
S H L Q I D K M T D N K T G E V L I S E		121
AGC CAC CTC CAG ATC GAC AAG ATG ACC GAC AAC AAG ACA GGA GAG GTG CTG ATC TCC GAG		437
N V V A S I Q P A E G S F E G D L K V P		141
AAT GTG GTG GCA TCC ATT CAA CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA CCC		497
R M E E K E A L V P I Q K A T D S F H T		161
AGG ATG GAG GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG GAC AGC TTC CAC ACA		557
E L H P R V A F W I I K L P R R R S H Q		181
GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT AAG CTG CCA CGG CGG AGG TCC CAC CAG		617
D A L E G G H W L S E K R H R L Q A I R		201
GAT GCC CTG GAG GGC GGC CAC TGG CTC AGC GAG AAG CGA CAC CGC CTG CAG GCC ATC CGG		677
D G L R K G T H K D V L E E G T E S S S		221
GAT GGA CTC CGC AAG GGG ACC CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC TCC		737
H S R L S P R K T H L L Y I L R P S R Q		241
CAC TCC AGG CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC AGG CCC TCT CGG CAG		797
L *		243
CTG TAG		803
GGGTGGGACCGGGGAGCACCTGCCTGTAGCCCCATCAGACCCTGCCCAAGCACCATATGGAATAAAGTTCTTTCT		882
TACATCTAAAAAAAAAAAAAAAAAAAAAAAAAATTGGCGGCCGC		928

FIGURA 6

crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	MQRLGATLLCLLLAAAVPTAPAPAPTATSAPVKFGPALS	40
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	PQEEATLNEMPREVEELNEDTQHKLRSAVEEHEAEAAAK	80
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	ASSEVNLANLFFSYHNEITNTDTNVGNTIKVHREIKKITS	120
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	PLGALVLDNFNNIRSADLHGARKOSQCLSDTDCMFRKFC	160
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	SPTRAGGDAQVQICLACRAKRRRCMRHAMCCPCNYCKNGIC	200
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	VTLE DATP LLE ROLD E Q DCT HAE GTT QN P V Q E N Q . . . P K R	240
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	KPSIKKISQGRKGQEGESCLRPFDCCGGLCCARHPWTK..I	280
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	CKPVLLEGQVCSRRGKRD...AQAFEIFQRCDCAAG	320
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	LLCRS QLT SNA..QNA RLHVCQKIEKL.....	360
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	VPDEYEVOSEPHVEVRQLEDLERSLTEEM..REPAAAAAA	400
crsp-2h		
crsp-3h		
crsp-4h		
tango59		
Consensus	LLGAEI	406

FIGURA 7

La familia de CRSP humana

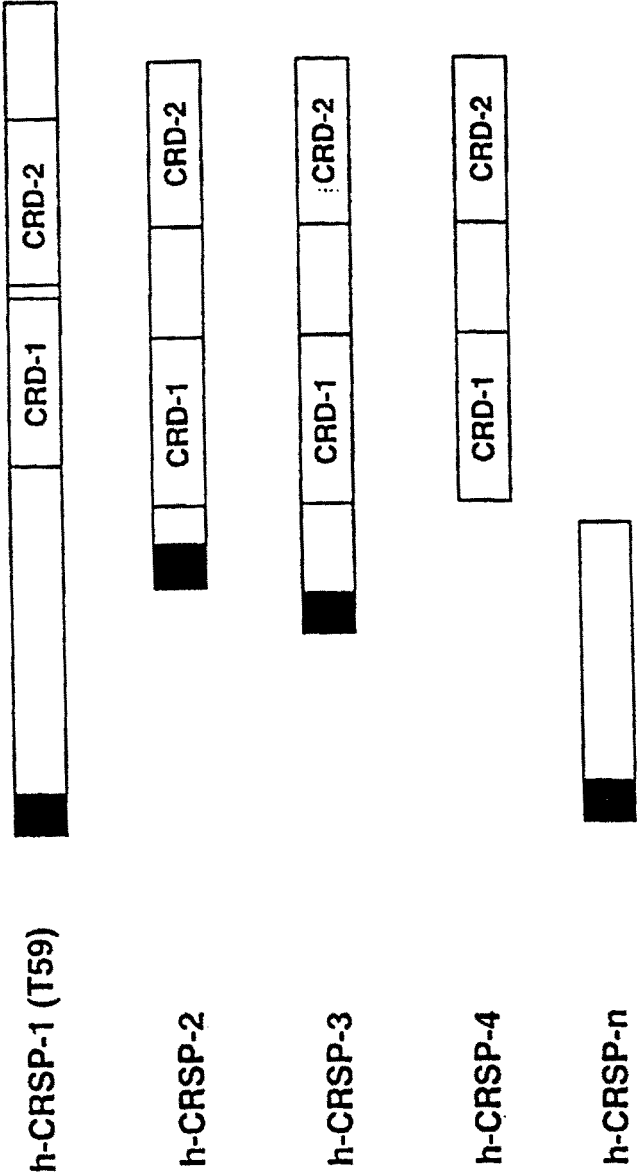


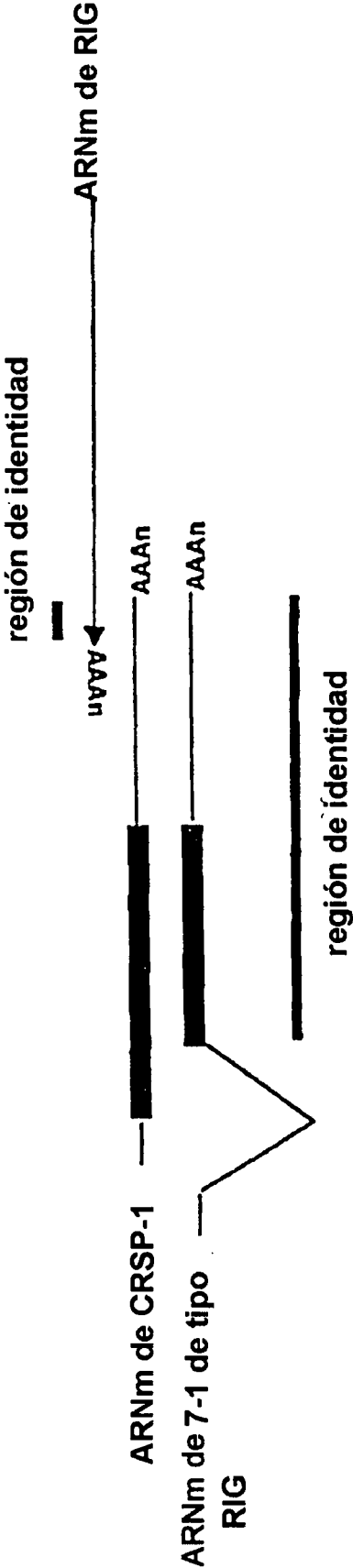
Figura 8A

FGTCGACCCACGCCTCCGCTGTGGCAGCCCCAGCTACCGGTCGTGACCAGATCCAGCTTGCAGCTCAGCTTTGTTTCATTC																										79
M Q R L G G I L L C T T T																										12
GAATTGGGCGGCCGCGCCAGCGCGGAACAAC ATG CAG CGG CTC GGG GGT ATT TTG CTG TGT ACA CTG																										145
L A A A V P T A P A P S P T V T W T P A	32																									
CTG GCG GCG GCG GTC CCC ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG	205																									
E P G P A L N Y P Q E E A T L N E M F R	5																									
GAG CCG GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT GAG ATG TTT CGA	265																									
E V E E L M E D T Q H K L R S A V E E M	72																									
GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA CTG CGC AGT GCC GTG GAG GAG ATG	125																									
E A E E A A A K T S S E V N L A S L P P	92																									
GAG GCG GAA GAA GCA GCT GCT AAA ACG TCC TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC	185																									
N Y H N E T S T E T R V G N N T V H V H	112																									
AAC TAT CAC AAT GAG ACC AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC	145																									
Q E V H K I T N N Q S G Q V V F S E T V	132																									
CAG GAA GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT TCT GAG ACA GTC	505																									
I T S V G D E E G K R S H E C I I D E D	152																									
ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC CAT GAA TGT ATC ATT GAT GAA GAC	565																									
C G P T R Y C Q F S S F K Y T C Q P C R	172																									
TGT GGG CCC ACC AGG TAC TGC CAG TTC TCC AGC TTC AAG TAC ACC TGC CAG CCA TGC CGG	625																									
D Q Q M L C T R D S E C C G D Q L C A W	192																									
GAC CAG CAG ATG CTA TGC ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG	685																									
G H C T Q K A T K G G N G T I C D N Q R	212																									
GGT CAC TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT GAC AAC CAG AGG	745																									
D C Q P G L C C A F Q R G L L F P V C T	232																									
GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA GGC CTG CTG TTC CCC GTG TGC ACA	805																									
P L P V E G E L C H D P T S Q L L D L I	252																									
CCC CTG CCC GTG GAG GGA GAG CTC TGC CAT GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC	865																									
T W E L E P E G A L D R C P C A S G L L	272																									
ACC TGG GAA CTG GAG CCT GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA	925																									
C O P H S H S L V Y M C K P A F V G S H	292																									
TGC CAG CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC GTG GGC AGC CAT	985																									
D H S E E S Q Y P R E A P D E Y E D V G	312																									
GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC CCG GAT GAG TAC GAA GAT GTT GGC	1045																									
F I G E V R Q E L E D L E R S L A Q E M	332																									
TTC ATA GGG GAA GTG CGC CAG GAG CTG GAA GAC CTG GAG CGG AGC CTA GCC CAG GAG ATG	1105																									

Figura 8B

A F E G P A P V E S L G G E E I * 350
 GCA TTT GAG GGG CCT GCC CCT GTG TCA CTA GGC GGA GAG GAG ATT TAG 1159
 GCCCAGACCCGCTGAGTCAGTCGTAGATGTGCAATAGAAATGGTAATTTATTTCCAGGAGTGTCCCAAGTGTGG 1238
 AATGGCGCAGCTCCTTCCCACTAGCTTTTCTCTGGCTTGACAAGGTACAGTGCAGTACATTTCTTCCAGCCGCCCTG 1317
 CTTCTCTGACTTGGGAAAGACAGGCATGGCCGGTAAGGCGAGGGTGAGTCGTCCCTCGCTGTTGCTAGAAACGCTGTC 1396
 TTGTTCTTCATGGATGGAAGATTTGTTTGAAGGAGAGGATGGGAAGGGTGAAGTCTGCTCATGATGGATTTGGGGGA 1475
 TACAGGAGGAGGATGCCTGCCTTGCAGACGTGGACTTGGCAAAATGTAACCTTTTGCTTTTGTCTTGGCGCGCTCCCAT 1554
 GGGCTGAGGCGAGTGCTACACAGAGCTATGCTGCTCTGTGGCCTCCACATATTCATCCCTGTGTTTCAGCTCCTACC 1633
 TCACCTGTCAGCACAGCCCTTCATAGCCACGCCCTCTTGCTCACCACAGCCTAGGAGGGACCAGAGGGGACCTTCTCT 1712
 CAGAGCCCCATGCTCTCTCTCAACCCCATACCCCATCTGTGCCAGCGACAGTCCTTCCAAATGGAGGAGTGAAT 1791
 CCTTTGGTTTAAATTAATTTTCTCCTTCAAGGCACGCCCTGCCACTAAGGTCAGGCTGACTTGCCCTCTTAACGTTG 1870
 TAGCAGTGTGTGGACACTGTCTTCCACCGACTGCTTCAATACCTCTGAAAGCCAGTGTCTGGAGTGCAGTTTCGTGTAA 1949
 ATTAATTTGCAGGAAGTATACTTGGCTAATTTGTAGGGCTAGGATTTGTAATGAATTTTGCAAAGTCCGCTTAGCAACAAT 2028
 GGAAAGCCTTTTCTCAGTCACACCGAGAAGTCAACAACCAAGCCAGGTTGTGTAGAGTACAGCTGTGACATACAGACAGAA 2107
 GAAGGCTGGGCTGGATGTCAGGCCCTCAGATGACGGTTTTCAGGTGCCAGGAACATTACCATTCTGTATCTATCCAGAGT 2186
 TATTAAAAATTGAAAGTTGCACACATTTGTATTAAGCATGCCTTTCTCCTCAGTTTAAATTTATATATACACAAACATG 2265
 TGGCCCTCAAAGATCATGCACAAACCACTACTCTTTGCTAAATTTCTTGGACTTTTCTCTTTGATTTTCAATAAATACAAA 2344
 TCCCTTTCATGCAAAAAAAGGGGCGCCGC 2381

Figura 9



LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE:
- (A) NOMBRE: MILLENNIUM BIOTHERAPEUTICS, INC.
 - (B) CALLE: 620 MEMORIAL DRIVE
 - 10 (C) CIUDAD: CAMBRIDGE
 - (D) ESTADO: MASSACHUSETTS
 - (E) PAÍS: EE.UU.
 - (F) CÓDIGO POSTAL: 02319
 - 15 (G) TELÉFONO:
 - (H) FAX
- (ii) TÍTULO DE LA INVENCION: PROTEÍNA CRSP NOVEDOSA Y MOLÉCULAS DE ACIDO NUCLEICO Y USOS DE LAS MISMAS
- 20 (iii) NÚMERO DE SECUENCIAS: 18
- (iv) DIRECCIÓN DE CORRESPONDENCIA
- 25 (A) DIRECCIÓN: LAHIVE & COCKFIELD, LLP
 - (B) CALLE: 25 STATE STREET
 - (C) CIUDAD: BOSTON
 - (D) ESTADO: MASSACHUSETTS
 - 30 (E) PAÍS: EE.UU.
 - (F) CÓDIGO POSTAL: 02109
- (v) FORMA LEGIBLE EN ORDENADOR
- 35 (A) TIPO DE MEDIO: DISQUETE
 - (B) ORDENADOR: IBM PC compatible
 - (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
 - 40 (D) SOFTWARE: PATENTIN Release # 1.0, Versión # 1.25
- (vi) DATOS ACTUALES DE SOLICITUD:
- (A) NÚMERO DE SOLICITUD: PCT/US98/
 - 45 (B) FECHA DE PRESENTACIÓN: 16 DE ABRIL 1998
 - (C) CLASIFICACIÓN:
- (vii) DATOS DE SOLICITUD ANTERIOR:
- 50 (A) NÚMERO DE SOLICITUD: US 08/843,704
 - (B) FECHA DE PRESENTACIÓN: 16 DE ABRIL 1997
 - (C) NÚMERO DE SOLICITUD: US 08/842,898
 - (D) FECHA DE PRESENTACIÓN: 17 DE ABRIL 1997
 - 55 (E) NÚMERO DE SOLICITUD: 60/071,589
 - (F) FECHA DE PRESENTACIÓN: 15 DE ENERO 1998
 - (G) NÚMERO DE SOLICITUD: 09/009,802
 - 60 (H) FECHA DE PRESENTACIÓN: 20 DE ENERO 1998
- (viii) ABOGADO/AGENTE DE INFORMACIÓN
- (A) NOMBRE: MANDRAGOURAS, AMY E
 - 65 (B) NÚMERO DE REGISTRO: 36.207
 - (C) REFERENCIA/NÚMERO DE DOCKET: MEI-008CPPC

ES 2 285 768 T3

(ix) INFORMACIÓN DE TELECOMUNICACIÓN

(A) TELÉFONO: (617)227-7400

(B) FAX:(617)742-4214

(2) INFORMACIÓN PARA SEQ ID NO: 1:

(i) CARACTERÍSTICAS DE SECUENCIA

(A) LONGITUD: 2479 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN_C

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 38..1087

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 1:

25	GGCACGAGGG GGCGGCGGCT GCGGGCGCAG AGCGGAG ATG CAG CGG CTT GGG GCC 55 Met Gln Arg Leu Gly Ala 1 5	
30	ACC CTG CTG TGC CTG CTG CTG GCG GCG GCG GTC CCC ACG GCC CCC GCG 103 Thr Leu Leu Cys Leu Leu Leu Ala Ala Ala Val Pro Thr Ala Pro Ala 10 15 20	
35	CCC GCT CCG ACG GCG ACC TCG GCT CCA GTC AAG CCC GGC CCG GCT CTC 151 Pro Ala Pro Thr Ala Thr Ser Ala Pro Val Lys Pro Gly Pro Ala Leu 25 30 35	
40	AGC TAC CCG CAG GAG GAG GCC ACC CTC AAT GAG ATG TTC CGC GAG GTT 199 Ser Tyr Pro Gln Glu Glu Ala Thr Leu Asn Glu Met Phe Arg Glu Val 40 45 50	
45	GAG GAA CTG ATG GAG GAC ACG CAG CAC AAA TTG CGC AGC GCG GTG GAA 247 Glu Glu Leu Met Glu Asp Thr Gln His Lys Leu Arg Ser Ala Val Glu 55 60 65 70	
50	GAG ATG GAG GCA GAA GAA GCT GCT GCT AAA GCA TCA TCA GAA GTG AAC 295 Glu Met Glu Ala Glu Glu Ala Ala Ala Lys Ala Ser Ser Glu Val Asn 75 80 85	
55	CTG GCA AAC TTA CCT CCC AGC TAT CAC AAT GAG ACC AAC ACA GAC ACG 343 Leu Ala Asn Leu Pro Pro Ser Tyr His Asn Glu Thr Asn Thr Asp Thr 90 95 100	
60	AAC GTT GGA AAT AAT ACC ATC CAT GTG CAC CGA GAA ATT CAC AAG ATA 391 Asn Val Gly Asn Asn Thr Ile His Val His Arg Glu Ile His Lys Ile 105 110 115	

ES 2 285 768 T3

	ACC AAC AAC CAG ACT GGA CAA ATG GTC TTT TCA GAG ACA GTT ATC ACA	439
	Thr Asn Asn Gln Thr Gly Gln Met Val Phe Ser Glu Thr Val Ile Thr	
	120 125 130	
5	TCT GTG GGA GAC GAA GAA GGC AGA AGG AGC CAC GAG TGC ATC ATC GAC	487
	Ser Val Gly Asp Glu Glu Gly Arg Arg Ser His Glu Cys Ile Ile Asp	
	135 140 145 150	
10	GAG GAC TGT GGG CCC AGC ATG TAC TGC CAG TTT GCC AGC TTC CAG TAC	535
	Glu Asp Cys Gly Pro Ser Met Tyr Cys Gln Phe Ala Ser Phe Gln Tyr	
	155 160 165	
15	ACC TGC CAG CCA TGC CGG GGC CAG AGG ATG CTC TGC ACC CGG GAC AGT	583
	Thr Cys Gln Pro Cys Arg Gly Gln Arg Met Leu Cys Thr Arg Asp Ser	
	170 175 180	
20	GAG TGC TGT GGA GAC CAG CTG TGT GTC TGG GGT CAC TGC ACC AAA ATG	631
	Glu Cys Cys Gly Asp Gln Leu Cys Val Trp Gly His Cys Thr Lys Met	
	185 190 195	
25	GCC ACC AGG GGC AGC AAT GGG ACC ATC TGT GAC AAC CAG AGG GAC TGC	679
	Ala Thr Arg Gly Ser Asn Gly Thr Ile Cys Asp Asn Gln Arg Asp Cys	
	200 205 210	
30	CAG CCG GGG CTG TGC TGT GCC TTC CAG AGA GGC CTG CTG TTC CCT GTG	727
	Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg Gly Leu Leu Phe Pro Val	
	215 220 225 230	
35	TGC ACA CCC CTG CCC GTG GAG GGC GAG CTT TGC CAT GAC CCC GCC AGC	775
	Cys Thr Pro Leu Pro Val Glu Gly Glu Leu Cys His Asp Pro Ala Ser	
	235 240 245	
40	CGG CTT CTG GAC CTC ATC ACC TGG GAG CTA GAG CCT GAT GGA GCC TTG	823
	Arg Leu Leu Asp Leu Ile Thr Trp Glu Leu Glu Pro Asp Gly Ala Leu	
	250 255 260	
45	GAC CGA TGC CCT TGT GCC AGT GGC CTC CTC TGC CAG CCC CAC AGC CAC	871
	Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu Cys Gln Pro His Ser His	
	265 270 275	
50	AGC CTG GTG TAT GTG TGC AAG CCG ACC TTC GTG GGG AGC CGT GAC CAA	919
	Ser Leu Val Tyr Val Cys Lys Pro Thr Phe Val Gly Ser Arg Asp Gln	
	280 285 290	
55	GAT GGG GAG ATC CTG CTG CCC AGA GAG GTC CCC GAT GAG TAT GAA GTT	967
	Asp Gly Glu Ile Leu Leu Pro Arg Glu Val Pro Asp Glu Tyr Glu Val	
	295 300 305 310	
60	GGC AGC TTC ATG GAG GAG GTG CGC CAG GAG CTG GAG GAC CTG GAG AGC	1015
	Gly Ser Phe Met Glu Glu Val Arg Gln Glu Leu Glu Asp Leu Glu Arg	
	315 320 325	
65	AGC CTG ACT GAA GAG ATG GCG CTG AGG GAG CCT GCG GCT GCC GCC GCT	1063
	Ser Leu Thr Glu Glu Met Ala Leu Arg Glu Pro Ala Ala Ala Ala Ala	
	330 335 340	

ES 2 285 768 T3

GCA CTG CTG GGA AGG GAA GAG ATT TAGATCTGGA CCAGGCTGTG GGTAGATGTG 1117
Ala Leu Leu Gly Arg Glu Glu Ile
345 150

5	CAATAGAAAT	AGCTAATTTA	TTTCCCCANG	TGTGTGCTTT	AAGCGTGGGC	TGACCAGGCT	1177
	TCTTCCTACA	TCTTCTTCCC	AGTAAGTTTC	CCCTCTGGCT	TGACAGCATG	AGGTGTTGTG	1237
	CATTTGTTCA	GCTCCCCCAG	GCTGTTCTCC	AGGCTTCACA	GTCTGCTGCT	TGGGAGAGTC	1297
10	AGGCAGGGTT	AAACTGCAGG	AGCAGTTTGC	CACCCCTGTC	CAGATTATTG	GCTGCTTTGC	1357
	CTCTACCAGT	TGGCAGACAG	CCGTTTGTTC	TACATGGCTT	TGATAATTGT	TTGAGGGGAG	1417
15	GAGATGGAAA	CAATGTGGAG	TCTCCCTCTG	ATTGGTTTTG	GGGAAATGTG	GAGAAGAGTG	1477
	CCCTGCTTTG	CAAACATCAA	CCTGGCAAAA	ATGCAACAAA	TGAATTTTCC	ACGCAGTTCT	1537
	TTCCATGGGC	ATAGGTAAGC	TGTGCCTTCA	GCTGTTGCAG	ATGAAATGTT	CTGTTCACCC	1597
20	TGCATTACAT	GTGTTTATT	ATCCAGCAGT	GTTGCTCAGC	TCCTACCTCT	GTGCCAGGGC	1657
	AGCATTITCA	TATCCAAGAT	CAATTCCCTC	TCTCAGCACA	GCCTGGGGAG	GGGGTCATTG	1717
25	TTCTCCTCGT	CCATCAGGGA	TTTCAGAGGC	TCAGAGACTG	CAAGCTGCTT	GCCCAAGTCA	1777
	CACAGCTAGT	GAAGACCAGA	GCAGTTTCAT	CTGTTGTGA	CTCTAAGCTC	AGTGCTCTCT	1837
	CCACTACCCC	ACACCAGCCT	TGGTGCCACC	AAAAGTGCTC	CCCAAAAGGA	AGGAGAATGG	1897
30	GATTTTCTT	TTGAGGCATG	CACATCTGGA	ATTAAGGTCA	AACTAATTCT	CACATCCCTC	1957
	TAAAAGTAAA	CTACTGTTAG	GAACAGCAGT	GTTCTCACAG	TGTGGGGCAG	CCGTCCTTCT	2017
35	AATGAAQACA	ATGATATTGA	CACTGTCCCT	CTTTGGCAGT	TGCATTAGTA	ACTTTGAAAG	2077
	GTATATGACT	GAGCGTAGCA	TACAGGTTAA	CCTGCAGAAA	CAGTACTTAG	GTAATTGTAG	2137
	GGCGAGGATT	ATAAATGAAA	TTTGCAAAAT	CACTTAGCAG	CAACTGAAGA	CAATTATCAA	2197
40	CCACGTGGAG	AAAATCAAAC	CGAGCAGGCC	TGTGTGAAAC	ATGGTTGTAA	TATGCCACTG	2257
	CGAACACTGA	ACTCTACGCC	ACTCCACAAA	TGATGTTTTC	AGGTGTCATG	GACTGTTGCC	2317
45	ACCATGTATT	CATCCAGAGT	TCTTAAAGTT	TAAAGTTGCA	CATGATTGTA	TAAGCATGCT	2377
	TTCTTTGAGT	TTTAAATTAT	GTATAAACAT	AAGTTGCATT	TAGAAATCAA	GCATAAATCA	2437
	CTTCAACTGC	TAAAAAATAA	AAAAAATAAA	AAAAAATAAA	AA		2497

(2) INFORMACIÓN PARA SEO ID NO: 2:

(i) CARACTERÍSTICAS DE SECUENCIA:

- (A) LONGITUD: 350 aminoácidos
(B) TIPO: aminoácido
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLECULA: proteína

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 2:

ES 2 285 768 T3

	Met	Gln	Arg	Leu	Gly	Ala	Thr	Leu	Leu	Cys	Leu	Leu	Leu	Ala	Ala	Ala	
	1				5					10					15		
5	Val	Pro	Thr	Ala	Pro	Ala	Pro	Ala	Pro	Thr	Ala	Thr	Ser	Ala	Pro	Val	
				20					25					30			
	Lys	Pro	Gly	Pro	Ala	Leu	Ser	Tyr	Pro	Gln	Glu	Glu	Ala	Thr	Leu	Asn	
			35					40					45				
10	Glu	Met	Phe	Arg	Glu	Val	Glu	Glu	Leu	Met	Glu	Asp	Thr	Gln	His	Lys	
		50					55					60					
	Leu	Arg	Ser	Ala	Val	Glu	Glu	Met	Glu	Ala	Glu	Glu	Ala	Ala	Ala	Lys	
15		65				70					75					80	
	Ala	Ser	Ser	Glu	Val	Asn	Leu	Ala	Asn	Leu	Pro	Pro	Ser	Tyr	His	Asn	
					85					90					95		
20	Glu	Thr	Asn	Thr	Asp	Thr	Asn	Val	Gly	Asn	Asn	Thr	Ile	His	Val	His	
				100					105					110			
	Arg	Glu	Ile	His	Lys	Ile	Thr	Asn	Asn	Gln	Thr	Gly	Gln	Met	Val	Phe	
25			115					120					125				
	Ser	Glu	Thr	Val	Ile	Thr	Ser	Val	Gly	Asp	Glu	Glu	Gly	Arg	Arg	Ser	
		130					135					140					
30	His	Glu	Cys	Ile	Ile	Asp	Glu	Asp	Cys	Gly	Pro	Ser	Met	Tyr	Cys	Gln	
		145				150					155					160	
	Phe	Ala	Ser	Phe	Gln	Tyr	Thr	Cys	Gln	Pro	Cys	Arg	Gly	Gln	Arg	Met	
35					165					170					175		
	Leu	Cys	Thr	Arg	Asp	Ser	Glu	Cys	Cys	Gly	Asp	Gln	Leu	Cys	Val	Trp	
				180					185					190			
40	Gly	His	Cys	Thr	Lys	Met	Ala	Thr	Arg	Gly	Ser	Asn	Gly	Thr	Ile	Cys	
			195					200					205				
	Asp	Asn	Gln	Arg	Asp	Cys	Gln	Pro	Gly	Leu	Cys	Cys	Ala	Phe	Gln	Arg	
45			210				215					220					
	Gly	Leu	Leu	Phe	Pro	Val	Cys	Thr	Pro	Leu	Pro	Val	Glu	Gly	Glu	Leu	
		225				230					235				240		

ES 2 285 768 T3

Cys His Asp Pro Ala Ser Arg Leu Leu Asp Leu Ile Thr Trp Glu Leu
245 250 255

Glu Pro Asp Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu
260 265 270

Cys Gln Pro His Ser His Ser Leu Val Tyr Val Cys Lys Pro Thr Phe
275 280 285

Val Gly Ser Arg Asp Gln Asp Gly Glu Ile Leu Leu Pro Arg Glu Val
290 295 300

Pro Asp Glu Tyr Glu Val Gly Ser Phe Met Glu Glu Val Arg Gln Glu
305 310 315 320

Leu Glu Asp Leu Glu Arg Ser Leu Thr Glu Glu Met Ala Leu Arg Glu
325 330 335

Pro Ala Ala Ala Ala Ala Ala Leu Leu Gly Arg Glu Glu Ile
340 345 350

(2) INFORMACIÓN PARA SEQ ID NO: 3:

(i) CARACTERÍSTICAS DE SECUENCIA

- (A) LONGITUD: 1050 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLECULA: ADN_c

(ix) CARACTERÍSTICA:

- (A) NOMBRE/CLAVE: CDS
- (B) LOCALIZACIÓN: 1..1050

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 3:

ATG CAG CGG CTT GGG GCC ACC CTG CTG TGC CTG CTG CTG GCG GCG GCG 48
Met Gln Arg Leu Gly Ala Thr Leu Leu Cys Leu Leu Leu Ala Ala Ala
1 5 10 15

GTC CCC ACG GCC CCC GCG CCC GCT CCG ACG GCG ACC TCG GCT CCA GTC 96
Val Pro Thr Ala Pro Ala Pro Ala Pro Thr Ala Thr Ser Ala Pro Val
20 25 30

AAG CCC GGC CCG GCT CTC AGC TAC CCG CAG GAG GAG GCC ACC CTC AAT 144
Lys Pro Gly Pro Ala Leu Ser Tyr Pro Gln Glu Glu Ala Thr Leu Asn
35 40 45

GAG ATG TTC CGC GAG GTT GAG GAA CTG ATG GAG GAC ACG CAG CAC AAA 192
Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys
50 55 60

ES 2 285 768 T3

	TTG	CGC	AGC	GCG	GTG	GAA	GAG	ATG	GAG	GCA	GAA	GAA	GCT	GCT	GCT	AAA	240
	Leu	Arg	Ser	Ala	Val	Glu	Glu	Met	Glu	Ala	Glu	Glu	Ala	Ala	Ala	Lys	
	65					70					75					80	
5	GCA	TCA	TCA	GAA	GTG	AAC	CTG	GCA	AAC	TTA	CCT	CCC	AGC	TAT	CAC	AAT	288
	Ala	Ser	Ser	Glu	Val	Asn	Leu	Ala	Asn	Leu	Pro	Pro	Ser	Tyr	His	Asn	
				85						90					95		
10	GAG	ACC	AAC	ACA	GAC	ACG	AAC	GTT	GGA	AAT	AAT	ACC	ATC	CAT	GTG	CAC	336
	Glu	Thr	Asn	Thr	Asp	Thr	Asn	Val	Gly	Asn	Asn	Thr	Ile	His	Val	His	
				100					105					110			
15	CGA	GAA	ATT	CAC	AAG	ATA	ACC	AAC	AAC	CAG	ACT	GGA	CAA	ATG	GTC	TTT	384
	Arg	Glu	Ile	His	Lys	Ile	Thr	Asn	Asn	Gln	Thr	Gly	Gln	Met	Val	Phe	
		115						120					125				
20	TCA	GAG	ACA	GTT	ATC	ACA	TCT	GTG	GGA	GAC	GAA	GAA	GGC	AGA	AGG	AGC	432
	Ser	Glu	Thr	Val	Ile	Thr	Ser	Val	Gly	Asp	Glu	Glu	Gly	Arg	Arg	Ser	
		130					135						140				
25	CAC	GAG	TGC	ATC	ATC	GAC	GAG	GAC	TGT	GGG	CCC	AGC	ATG	TAC	TGC	CAG	480
	His	Glu	Cys	Ile	Ile	Asp	Glu	Asp	Cys	Gly	Pro	Ser	Met	Tyr	Cys	Gln	
	145					150					155					160	
30	TTT	GCC	AGC	TTC	CAG	TAC	ACC	TGC	CAG	CCA	TGC	CGG	GGC	CAG	AGG	ATG	528
	Phe	Ala	Ser	Phe	Gln	Tyr	Thr	Cys	Gln	Pro	Cys	Arg	Gly	Gln	Arg	Met	
				165						170					175		
35	CTC	TGC	ACC	CGG	GAC	AGT	GAG	TGC	TGT	GGA	GAC	CAG	CTG	TGT	GTC	TGG	576
	Leu	Cys	Thr	Arg	Asp	Ser	Glu	Cys	Cys	Gly	Asp	Gln	Leu	Cys	Val	Trp	
				180						185				190			
40	GGT	CAC	TGC	ACC	AAA	ATG	GCC	ACC	AGG	GGC	AGC	AAT	GGG	ACC	ATC	TGT	624
	Gly	His	Cys	Thr	Lys	Met	Ala	Thr	Arg	Gly	Ser	Asn	Gly	Thr	Ile	Cys	
		195						200					205				
45	GAC	AAC	CAG	AGG	GAC	TGC	CAG	CCG	GGG	CTG	TGC	TGT	GCC	TTC	CAG	AGA	672
	Asp	Asn	Gln	Arg	Asp	Cys	Gln	Pro	Gly	Leu	Cys	Cys	Ala	Phe	Gln	Arg	
		210					215						220				
50	GGC	CTG	CTG	TTC	CCT	GTG	TGC	ACA	CCC	CTG	CCC	GTG	GAG	GGC	GAG	CTT	720
	Gly	Leu	Leu	Phe	Pro	Val	Cys	Thr	Pro	Leu	Pro	Val	Glu	Gly	Glu	Leu	
	225					230						235				240	
55	TGC	CAT	GAC	CCC	GCC	AGC	CGG	CTT	CTG	GAC	CTC	ATC	ACC	TGG	GAG	CTA	768
	Cys	His	Asp	Pro	Ala	Ser	Arg	Leu	Leu	Asp	Leu	Ile	Thr	Trp	Glu	Leu	
				245						250					255		
60	GAG	CCT	GAT	GGA	GCC	TTG	GAC	CGA	TGC	CCT	TGT	GCC	AGT	GGC	CTC	CTC	816
	Glu	Pro	Asp	Gly	Ala	Leu	Asp	Arg	Cys	Pro	Cys	Ala	Ser	Gly	Leu	Leu	
				260				265					270				
65	TGC	CAG	CCC	CAC	AGC	CAC	AGC	CTG	GTG	TAT	GTG	TGC	AAG	CCG	ACC	TTT	864
	Cys	Gln	Pro	His	Ser	His	Ser	Leu	Val	Tyr	Val	Cys	Lys	Pro	Thr	Phe	
		275						280					285				

ES 2 285 768 T3

	GTG GGG AGC CGT GAC CAA GAT GGG GAG ATC CTG CTG CCC AGA GAG GTC	912
	Val Gly Ser Arg Asp Gln Asp Gly Glu Ile Leu Leu Pro Arg Glu Val	
	290 295 300	
5	CCC GAT GAG TAT GAA GTT GGC AGC TTC ATG GAG GAG GTG CGC CAG GAG	960
	Pro Asp Glu Tyr Glu Val Gly Ser Phe Met Glu Glu Val Arg Gln Glu	
	305 310 315 320	
10	CTG GAG GAC CTG GAG AGG AGC CTG ACT GAA GAG ATG GCG CTG AGG GAG	1008
	Leu Glu Asp Leu Glu Arg Ser Leu Thr Glu Glu Met Ala Leu Arg Glu	
	325 330 335	
15	CCT GCG GCT GCC GCC GCT GCA CTG CTG GGA AGG GAA GAG ATT	1050
	Pro Ala Ala Ala Ala Ala Ala Leu Leu Gly Arg Glu Glu Ile	
	340 345 350	

(2) INFORMACIÓN PARA SEQ ID NO: 4:

- 20 (i) CARACTERÍSTICAS DE SECUENCIA:
- (A) LONGITUD: 848 pares de bases
 - (B) TIPO: ácido nucleico
 - 25 (C) TIPO DE CADENA: sencilla
 - (D) TOPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: ADN_c
- 30 (ix) CARACTERÍSTICA:
- (A) NOMBRE/CLAVE: CDS
 - (B) LOCALIZACIÓN: 125..796
- 35 (xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 4:

	GAATTCGGCA CGAGAGACGA CGTGCTGAGC TGCCAGCTTA GTGGAAGCTC TGCTCTGGGT	60
40	GGAGAGCAGC CTCGCTTTGG TGACGCACAG TGCTGGGACC CTCCAGGAGC CCCGGGATTG	120
	AAGG ATG GTG GCG GCC GTC CTG CTG GGG CTG AGC TGG CTC TGC TCT CCC	169
	Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro	
45	1 5 10 15	
	CTG GGA GCT CTG GTC CTG GAC TTC AAC AAC ATC AGG AGC TCT GCT GAC	217
	Leu Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp	
	20 25 30	
50	CTG CAT GGG GCC CGG AAG GGC TCA CAG TGC CTG TCT GAC ACG GAC TGC	265
	Leu His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys	
	35 40 45	
55	AAT ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT	313
	Asn Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys	
	50 55 60	

ES 2 285 768 T3

	GCT ACA TGT CGT GGG TTG CGG AGG AGG TGC CAG CGA GAT GCC ATG TGC	361
	Ala Thr Cys Arg Gly Leu Arg Arg Cys Gln Arg Asp Ala Met Cys	
	65 70 75	
5	TGC CCT GGG ACA CTC TGT GTG AAC GAT GTT TGT ACT ACG ATG GAA GAT	409
	Cys Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp	
	80 85 90 95	
10	GCA ACC CCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT GGC ACA CAT	457
	Ala Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His	
	100 105 110	
15	GCA GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG	505
	Ala Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg	
	115 120 125	
20	AAG CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA	553
	Lys Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu	
	130 135 140	
25	AGT TGT CTG AGA ACT TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CGT	601
	Ser Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg	
	145 150 155	
30	CAT TTT TGG ACG AAA ATT TGT AAG CCA GTC CTT TTG GAG GGA CAG GTC	649
	His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val	
	160 165 170 175	
35	TGC TCC AGA AGA GGG CAT AAA GAC ACT GCT CAA GCT CCA GAA ATC TTC	697
	Cys Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe	
	180 185 190	
40	CAG CGT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC	745
	Gln Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr	
	195 200 205	
45	AGC AAT CGG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG	793
	Ser Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys	

40 (2) INFORMACIÓN PARA SEQ ID NO: 5:

(i) CARACTERÍSTICAS DE SECUENCIA:

(A) LONGITUD: 224 aminoácidos

45 (B) TIPO: aminoácido

(C) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

50 (xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 5:

55

60

65

ES 2 285 768 T3

Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro Leu
 1 5 10 15
 Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp Leu
 5 20 25 30
 His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys Asn
 10 35 40 45
 Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys Ala
 50 55 60
 Thr Cys Arg Gly Leu Arg Arg Arg Cys Gln Arg Asp Ala Met Cys Cys
 15 65 70 75 80
 Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp Ala
 20 85 90 95
 Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His Ala
 100 105 110
 Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg Lys
 25 115 120 125
 Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu Ser
 30 130 135 140
 Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg His
 145 150 155 160
 Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val Cys
 35 165 170 175
 Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe Gln
 40 180 185 190
 Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr Ser
 195 200 205
 Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys Leu
 45 210 215 220

50 (2) INFORMACIÓN PARA SEQ ID NO: 6:

(i) CARACTERÍSTICAS DE SECUENCIA:

(A) LONGITUD: 672 pares de bases

55 (B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

60 (ii) TIPO DE MOLÉCULA: ADN_c

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

65 (B) LOCALIZACIÓN: 1..672

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 6:

ES 2 285 768 T3

	ATG GTG GCG GCC GTC CTG CTG GGG CTG AGC TGG CTC TGC TCT CCC CTG	48
	Met Val Ala Ala Val Leu Leu Gly Leu Ser Trp Leu Cys Ser Pro Leu	
	1 5 10 15	
5	GGA GCT CTG GTC CTG GAC TTC AAC AAC ATC AGG AGC TCT GCT GAC CTG	96
	Gly Ala Leu Val Leu Asp Phe Asn Asn Ile Arg Ser Ser Ala Asp Leu	
	20 25 30	
10	CAT GGG GCC CGG AAG GGC TCA CAG TGC CTG TCT GAC ACG GAC TGC AAT	144
	His Gly Ala Arg Lys Gly Ser Gln Cys Leu Ser Asp Thr Asp Cys Asn	
	35 40 45	
15	ACC AGA AAG TTC TGC CTC CAG CCC CGC GAT GAG AAG CCG TTC TGT GCT	192
	Thr Arg Lys Phe Cys Leu Gln Pro Arg Asp Glu Lys Pro Phe Cys Ala	
	50 55 60	
20	ACA TGT CGT GGG TTG CGG AGG AGG TGC CAG CGA GAT GCC ATG TGC TGC	240
	Thr Cys Arg Gly Leu Arg Arg Cys Gln Arg Asp Ala Met Cys Cys	
	65 70 75 80	
25	CCT GGG ACA CTC TGT GTG AAC GAT GTT TGT ACT ACG ATG GAA GAT GCA	288
	Pro Gly Thr Leu Cys Val Asn Asp Val Cys Thr Thr Met Glu Asp Ala	
	85 90 95	
30	ACC CCA ATA TTA GAA AGG CAG CTT GAT GAG CAA GAT GGC ACA CAT GCA	336
	Thr Pro Ile Leu Glu Arg Gln Leu Asp Glu Gln Asp Gly Thr His Ala	
	100 105 110	
35	GAA GGA ACA ACT GGG CAC CCA GTC CAG GAA AAC CAA CCC AAA AGG AAG	384
	Glu Gly Thr Thr Gly His Pro Val Gln Glu Asn Gln Pro Lys Arg Lys	
	115 120 125	
40	CCA AGT ATT AAG AAA TCA CAA GGC AGG AAG GGA CAA GAG GGA GAA AGT	432
	Pro Ser Ile Lys Lys Ser Gln Gly Arg Lys Gly Gln Glu Gly Glu Ser	
	130 135 140	
45	TGT CTG AGA ACT TTT GAC TGT GGC CCT GGA CTT TGC TGT GCT CGT CAT	480
	Cys Leu Arg Thr Phe Asp Cys Gly Pro Gly Leu Cys Cys Ala Arg His	
	145 150 155 160	
50	TTT TGG ACG AAA ATT TGT AAG CCA GTC CTT TTG GAG GGA CAG GTC TGC	528
	Phe Trp Thr Lys Ile Cys Lys Pro Val Leu Leu Glu Gly Gln Val Cys	
	165 170 175	
55	TCC AGA AGA GGG CAT AAA GAC ACT GCT CAA GCT CCA GAA ATC TTC CAG	576
	Ser Arg Arg Gly His Lys Asp Thr Ala Gln Ala Pro Glu Ile Phe Gln	
	180 185 190	
60	CGT TGC GAC TGT GGC CCT GGA CTA CTG TGT CGA AGC CAA TTG ACC AGC	624
	Arg Cys Asp Cys Gly Pro Gly Leu Leu Cys Arg Ser Gln Leu Thr Ser	
	195 200 205	
65	AAT CGG CAG CAT GCT CGA TTA AGA GTA TGC CAA AAA ATA GAA AAG CTA	672
	Asn Arg Gln His Ala Arg Leu Arg Val Cys Gln Lys Ile Glu Lys Leu	
	210 215 220	

(2) INFORMACIÓN PARA SEQ ID NO: 7:

(i) CARACTERÍSTICAS DE SEQUENCIA:

- (A) LONGITUD: 1529 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla

ES 2 285 768 T3

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLECULA: ADN_c

(ix) CARACTERISTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN:93..890

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 7:

```

CCGGACGCGT GGGCGGCACG GTTTCGTGGG GACCCAGGCT TGCAAAGTGA CGGTCATTTT    60
CTCTTTCTTT CTCCCTCTTG AGTCCTTCAG AG ATG ATG GCT CTG GGC GCA GCG    113
                                     Met Met Ala Leu Gly Ala Ala
                                     1      5
GGA GCT ACC CGG GTC TTT GTC GCG ATG GTA GCG GCG GCT CTC GGC GGC    161
Gly Ala Thr Arg Val Phe Val Ala Met Val Ala Ala Leu Gly Gly
      10      15      20
CAC CCT CTG CTG GGA GTG AGC GCC ACC TTG AAC TCG GTT CTC AAT TCC    209
His Pro Leu Leu Gly Val Ser Ala Thr Leu Asn Ser Val Leu Asn Ser
      25      30      35
AAC GCT ATC AAG AAC CTG CCC CCA CCG CTG GGC GGC GCT GCG GGG CAC    257
Asn Ala Ile Lys Asn Leu Pro Pro Pro Leu Gly Gly Ala Ala Gly His
      40      45      50      55
CCA GGC TCT GCA GTC AGC GCC GCG CCG GGA ATC CTG TAC CCG GGC GGC    305
Pro Gly Ser Ala Val Ser Ala Ala Pro Gly Ile Leu Tyr Pro Gly Gly
      60      65      70
AAT AAG TAC CAG ACC ATT GAC AAC TAC CAG CCG TAC CCG TGC GCA GAG    353
Asn Lys Tyr Gln Thr Ile Asp Asn Tyr Gln Pro Tyr Pro Cys Ala Glu
      75      80      85
GAC GAG GAG TGC GGC ACT GAT GAG TAC TGC GCT AGT CCC ACC CGC GGA    401
Asp Glu Glu Cys Gly Thr Asp Glu Tyr Cys Ala Ser Pro Thr Arg Gly
      90      95     100

```

ES 2 285 768 T3

	GGG GAC GCA GGC GTG CAA ATC TGT CTC GCC TGC AGG AAG CGC CGA AAA	449
	Gly Asp Ala Gly Val Gln Ile Cys Leu Ala Cys Arg Lys Arg Arg Lys	
	105 110 115	
5	CGC TGC ATG CGT CAC GCT ATG TGC TGC CCC GGG AAT TAC TGC AAA AAT	497
	Arg Cys Met Arg His Ala Met Cys Cys Pro Gly Asn Tyr Cys Lys Asn	
	120 125 130 135	
10	GGA ATA TGC GTG TCT TCT GAT CAA AAT CAT TTC CGA GGA GAA ATT GAG	545
	Gly Ile Cys Val Ser Ser Asp Gln Asn His Phe Arg Gly Glu Ile Glu	
	140 145 150	
15	GAA ACC ATC ACT GAA AGC TTT GGT AAT GAT CAT AGC ACC TTG GAT GGG	593
	Glu Thr Ile Thr Glu Ser Phe Gly Asn Asp His Ser Thr Leu Asp Gly	
	155 160 165	
20	TAT TCC AGA AGA ACC ACC TTG TCT TCA AAA ATG TAT CAC ACC AAA GGA	641
	Tyr Ser Arg Arg Thr Thr Leu Ser Ser Lys Met Tyr His Thr Lys Gly	
	170 175 180	
25	CAA GAA GGT TCT GTT TGT CTC CGG TCA TCA GAC TGT GCC TCA GGA TTG	689
	Gln Glu Gly Ser Val Cys Leu Arg Ser Ser Asp Cys Ala Ser Gly Leu	
	185 190 195	
30	TGT TGT GCT AGA CAC TTC TGG TCC AAG ATC TGT AAA CCT GTC CTG AAA	737
	Cys Cys Ala Arg His Phe Trp Ser Lys Ile Cys Lys Pro Val Leu Lys	
	200 205 210 215	
35	GAA GGT CAA GTG TGT ACC AAG CAT AGG AGA AAA GGC TCT CAT GGA CTA	785
	Glu Gly Gln Val Cys Thr Lys His Arg Arg Lys Gly Ser His Gly Leu	
	220 225 230	
40	GAA ATA TTC CAG CGT TGT TAC TGT GGA GAA GGT CTG TCT TGC CGG ATA	833
	Glu Ile Phe Gln Arg Cys Tyr Cys Gly Glu Gly Leu Ser Cys Arg Ile	
	235 240 245	
45	CAG AAA GAT CAC CAT CAA GCC AGT AAT TCT TCT AGG CTT CAC ACT TGT	881
	Gln Lys Asp His His Gln Ala Ser Asn Ser Ser Arg Leu His Thr Cys	
	250 255 260	
50	CAG AGA CAC TAAACCAGCT ATCCAAAATG CAGTGAAGTC CTTTTATATA	930
	Gln Arg His	
	265	
55	ATAGATGCTA TGAAAACCTT TTATGACCTT CATCAACTCA ATCCTAAGGA TATACAAGTT	990
	CTGTGGTTTC AGTTAAGCAT TCCAATAACA CCTTCCAAAA ACCTGGAGTG TAAGAGCTTT	1050
60	GTTTCCTTTAT GGAACCTCCCC TGTGATTGCA GTAAATTACT GTATTGTAAA TTCTCAGTGT	1110
	GGCACTTACC TGTAATGCA ATGAACTTT TAATTATTTT TCTAAAGGTG CTGCACTGCC	1170
65	TATTTTTCCT CTTGTTATGT AAATTTTGT ACACATTGAT TGTTATCTTG ACTGACAAAT	1230
	ATTCTATATT GAACTGAAGT AAATCATTTC AGCTTATAGT TCTTAAAAGC ATAACCCTTT	1290
	ACCCCATTTN ATTCTAGAGT CNAGAACGCA AGGATCTCTT GGAATGACAA ATGATAGGTA	1350
70	CCTAAAATGT AACATGAAAA TACTAGCTTA TTTTCTGAAA TGTACTATCT TAATGCTTAA	1410
	ATTATATTTT CTTTAGGCT GTGATAGTTT TTGAAATAAA ATTTAACATT TAATATCATG	1470
75	AAATGKTATA AGTAGACATA AAAAAAAAAA AAAAAAAAAA AGGGCGGCCG CTAGACTAG	1529

ES 2 285 768 T3

(2) INFORMACIÓN PARA SEQ. ID NO: 8:

(i) CARACTERÍSTICAS DE SECUENCIA

(A) LONGITUD: 266 aminoácidos

(B) TIPO: aminoácidos

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 8:

```

Met Met Ala Leu Gly Ala Ala Gly Ala Thr Arg Val Phe Val Ala Met
 1              5              10              15

Val Ala Ala Ala Leu Gly Gly His Pro Leu Leu Gly Val Ser Ala Thr
          20              25              30

Leu Asn Ser Val Leu Asn Ser Asn Ala Ile Lys Asn Leu Pro Pro Pro
          35              40              45

Leu Gly Gly Ala Ala Gly His Pro Gly Ser Ala Val Ser Ala Ala Pro
          50              55              60

Gly Ile Leu Tyr Pro Gly Gly Asn Lys Tyr Gln Thr Ile Asp Asn Tyr
 65              70              75              80

Gln Pro Tyr Pro Cys Ala Glu Asp Glu Glu Cys Gly Thr Asp Glu Tyr
          85              90              95

Cys Ala Ser Pro Thr Arg Gly Gly Asp Ala Gly Val Gln Ile Cys Leu
          100              105              110

Ala Cys Arg Lys Arg Arg Lys Arg Cys Met Arg His Ala Met Cys Cys
          115              120              125

Pro Gly Asn Tyr Cys Lys Asn Gly Ile Cys Val Ser Ser Asp Gln Asn
          130              135              140

His Phe Arg Gly Glu Ile Glu Glu Thr Ile Thr Glu Ser Phe Gly Asn
145              150              155              160

Asp His Ser Thr Leu Asp Gly Tyr Ser Arg Arg Thr Thr Leu Ser Ser
          165              170              175

```

ES 2 285 768 T3

Lys Met Tyr His Thr Lys Gly Gln Glu Gly Ser Val Cys Leu Arg Ser
 180 185 190
 5 Ser Asp Cys Ala Ser Gly Leu Cys Cys Ala Arg His Phe Trp Ser Lys
 195 200 205
 10 Ile Cys Lys Pro Val Leu Lys Glu Gly Gln Val Cys Thr Lys His Arg
 210 215 220
 Arg Lys Gly Ser His Gly Leu Glu Ile Phe Gln Arg Cys Tyr Cys Gly
 225 230 235 240
 15 Glu Gly Leu Ser Cys Arg Ile Gln Lys Asp His His Gln Ala Ser Asn
 245 250 255
 20 Ser Ser Arg Leu His Thr Cys Gln Arg His
 260 265

(2) INFORMACIÓN PARA SEQ ID NO: 9:

- 25 (i) CARACTERÍSTICAS DE SECUENCIA:
 (A) LONGITUD: 798 pares de bases
 (B) TIPO: ácido nucleico
 30 (C) TIPO DE CADENA: sencilla
 (D) TOPOLOGÍA: lineal
 (ii) TIPO DE MOLÉCULA: ADN_c
 35 (ix) CARACTERÍSTICA:
 (A) NOMBRE/CLAVE: CDS
 (B) LOCALIZACIÓN: 1..798
 40 (xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 9:

45 ATG ATG GCT CTG GGC GCA GCG GGA GCT ACC CGG GTC TTT GTC GCG ATG 48
 Met Met Ala Leu Gly Ala Ala Gly Ala Thr Arg Val Phe Val Ala Met
 1 5 10 15
 GTA GCG GCG GCT CTC GGC GGC CAC CCT CTG CTG GGA GTG AGC GCC ACC 96
 Val Ala Ala Ala Leu Gly Gly His Pro Leu Leu Gly Val Ser Ala Thr
 20 25 30
 50 TTG AAC TCG GTT CTC AAT TCC AAC GCT ATC AAG AAC CTG CCC CCA CCG 144
 Leu Asn Ser Val Leu Asn Ser Asn Ala Ile Lys Asn Leu Pro Pro Pro
 35 40 45
 55 CTG GGC GGC GCT GCG GGG CAC CCA GGC TCT GCA GTC AGC GCC GCG CCG 192
 Leu Gly Gly Ala Ala Gly His Pro Gly Ser Ala Val Ser Ala Ala Pro
 50 55 60
 60 GGA ATC CTG TAC CCG GGC GGG AAT AAG TAC CAG ACC ATT GAC AAC TAC 240
 Gly Ile Leu Tyr Pro Gly Gly Asn Lys Tyr Gln Thr Ile Asp Asn Tyr
 65 70 75 80

65

ES 2 285 768 T3

	CAG CCG TAC CCG TGC GCA GAG GAC GAG GAG TGC GGC ACT GAT GAG TAC	288
	Gln Pro Tyr Pro Cys Ala Glu Asp Glu Cys Gly Thr Asp Glu Tyr	
	85 90 95	
5	TGC GCT AGT CCC ACC CGC GGA GGG GAC GCA GGC GTG CAA ATC TGT CTC	336
	Cys Ala Ser Pro Thr Arg Gly Gly Asp Ala Gly Val Gln Ile Cys Leu	
	100 105 110	
10	GCC TGC AGG AAG CGC CGA AAA CGC TGC ATG CGT CAC GCT ATG TGC TGC	384
	Ala Cys Arg Lys Arg Arg Lys Arg Cys Met Arg His Ala Met Cys Cys	
	115 120 125	
15	CCC GGG AAT TAC TGC AAA AAT GGA ATA TGC GTG TCT TCT GAT CAA AAT	432
	Pro Gly Asn Tyr Cys Lys Asn Gly Ile Cys Val Ser Ser Asp Gln Asn	
	130 135 140	
20	CAT TTC CGA GGA GAA ATT GAG GAA ACC ATC ACT GAA AGC TTT GGT AAT	480
	His Phe Arg Gly Glu Ile Glu Glu Thr Ile Thr Glu Ser Phe Gly Asn	
	145 150 155 160	
25	GAT CAT AGC ACC TTG GAT GGG TAT TCC AGA AGA ACC ACC TTG TCT TCA	528
	Asp His Ser Thr Leu Asp Gly Tyr Ser Arg Arg Thr Thr Leu Ser Ser	
	165 170 175	
30	AAA ATG TAT CAC ACC AAA GGA CAA GAA GGT TCT GTT TGT CTC CGG TCA	576
	Lys Met Tyr His Thr Lys Gly Gln Glu Gly Ser Val Cys Leu Arg Ser	
	180 185 190	
35	TCA GAC TGT GCC TCA GGA TTG TGT TGT GCT AGA CAC TTC TGG TCC AAG	624
	Ser Asp Cys Ala Ser Gly Leu Cys Cys Ala Arg His Phe Trp Ser Lys	
	195 200 205	
40	ATC TGT AAA CCT GTC CTG AAA GAA GGT CAA GTG TGT ACC AAG CAT AGG	672
	Ile Cys Lys Pro Val Leu Lys Glu Gly Gln Val Cys Thr Lys His Arg	
	210 215 220	
45	AGA AAA GGC TCT CAT GGA CTA GAA ATA TTC CAG CGT TGT TAC TGT GGA	720
	Arg Lys Gly Ser His Gly Leu Glu Ile Phe Gln Arg Cys Tyr Cys Gly	
	225 230 235 240	
50	GAA GGT CTG TCT TGC CGG ATA CAG AAA GAT CAC CAT CAA GCC AGT AAT	768
	Glu Gly Leu Ser Cys Arg Ile Gln Lys Asp His His Gln Ala Ser Asn	
	245 250 255	
55	TCT TCT AGG CTT CAC ACT TGT CAG AGA CAC	798
	Ser Ser Arg Leu His Thr Cys Gln Arg His	
	260 265	

50 (2) INFORMACIÓN PARA SEQ ID NO: 10:

(i) CARACTERÍSTICAS DE SECUENCIA:

(A) LONGITUD: 702 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

60 (ii) TIPO DE MOLÉCULA: ADN_c

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1..537

65 (xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 10:

ES 2 285 768 T3

	GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA	48
	Glu Phe Gly Thr Arg Val Gly Arg Tyr Cys His Ser Pro His Gln Gly	
	1 5 10 15	
5	TCA TCG GCC TGC ATG GTG TGT CGG AGA AAA AAG AAG CGC TGC CAC CGA	96
	Ser Ser Ala Cys Met Val Cys Arg Arg Lys Lys Lys Arg Cys His Arg	
	20 25 30	
10	GAT GGC ATG TGC TGC CCC AGT ACC CGC TGC AAT AAT GGC ATC TGT ATC	144
	Asp Gly Met Cys Cys Pro Ser Thr Arg Cys Asn Asn Gly Ile Cys Ile	
	35 40 45	
15	CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG GCT CTG GAT GGT	192
	Pro Val Thr Glu Ser Ile Leu Thr Pro His Ile Pro Ala Leu Asp Gly	
	50 55 60	
20	ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG	240
	Thr Arg His Arg Asp Arg Asn His Gly His Tyr Ser Asn His Asp Leu	
	65 70 75 80	
25	GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA	288
	Gly Trp Gln Asn Leu Gly Arg Pro His Thr Lys Met Ser His Ile Lys	
	85 90 95	
30	GGG CAT GAA GGA GAC CCC TGC CTA CGA TCA TCA GAC TGC ATT GAA GGG	336
	Gly His Glu Gly Asp Pro Cys Leu Arg Ser Ser Asp Cys Ile Glu Gly	
	100 105 110	
35	TTT TGC TGT GCT CGT CAT TTC TGG ACC AAA ATC TGC AAA CCA GTG CTC	384
	Phe Cys Cys Ala Arg His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu	
	115 120 125	
40	CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG GGT TCT CAT GGG	432
	His Gln Gly Glu Val Cys Thr Lys Gln Arg Lys Lys Gly Ser His Gly	
	130 135 140	
45	CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA	480
	Leu Glu Ile Phe Gln Arg Cys Asp Cys Ala Lys Gly Leu Ser Cys Lys	
	145 150 155 160	
50	GTA TGG AAA GAT GCC ACC TAC TCC TCC AAA GCC AGA CTC CAT GTG TGT	528
	Val Trp Lys Asp Ala Thr Tyr Ser Ser Lys Ala Arg Leu His Val Cys	
	165 170 175	
55	CAG AAA ATT TGATCACCAT TGAGGAACAT CATCAATTGC AGACTGTGAA	577
	Gln Lys Ile	
60	GTGTGTATT TAATGCATTA TAGCATGGTG GAAAATAAGG TTCAGATGCA GAAGAATGGC	637
65	TAAAATAAGA AACGTGATAA GAATATAGAT GATCACAAAA AAAAAAAAAA AAAAGATGCG	697
	GCCGC	702

(2) INFORMACION PARA SEQ ID NO: 11:

(i) CARACTERÍSTICAS DE SECUENCIA

(A) LONGITUD: 179 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCION DE SECUENCIA: SEQ ID NO: 11:

ES 2 285 768 T3

[illegible]

40

(2) INFORMACIÓN PARA SEQ ID NO: 12:

(i) CARACTERISTICAS DE SECUENCIA:

(A) LONGITUD: 537 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN_C

(ix) CARATTERISTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1..537

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 12:

60

65

ES 2 285 768 T3

	GAA TTC GGC ACG AGG GTT GGG AGG TAT TGC CAC AGT CCC CAC CAA GGA	48
	Glu Phe Gly Thr Arg Val Gly Arg Tyr Cys His Ser Pro His Gln Gly	
	1 5 10 15	
5	TCA TCG GCC TGC ATG GTG TGT CGG AGA AAA AAG AAG CGC TGC CAC CGA	96
	Ser Ser Ala Cys Met Val Cys Arg Arg Lys Lys Lys Arg Cys His Arg	
	20 25 30	
10	GAT GGC ATG TGC TGC CCC AGT ACC CGC TGC AAT AAT GGC ATC TGT ATC	144
	Asp Gly Met Cys Cys Pro Ser Thr Arg Cys Asn Asn Gly Ile Cys Ile	
	35 40 45	
15	CCA GTT ACT GAA AGC ATC TTA ACC CCT CAC ATC CCG GCT CTG GAT GGT	192
	Pro Val Thr Glu Ser Ile Leu Thr Pro His Ile Pro Ala Leu Asp Gly	
	50 55 60	
20	ACT CGG CAC AGA GAT CGA AAC CAC GGT CAT TAC TCA AAC CAT GAC TTG	240
	Thr Arg His Arg Asp Arg Asn His Gly His Tyr Ser Asn His Asp Leu	
	65 70 75 80	
25	GGA TGG CAG AAT CTA GGA AGA CCA CAC ACT AAG ATG TCA CAT ATA AAA	288
	Gly Trp Gln Asn Leu Gly Arg Pro His Thr Lys Met Ser His Ile Lys	
	85 90 95	
30	GGG CAT GAA GGA GAC CCC TGC CTA CGA TCA TCA GAC TGC ATT GAA GGG	336
	Gly His Glu Gly Asp Pro Cys Leu Arg Ser Ser Asp Cys Ile Glu Gly	
	100 105 110	
35	TTT TGC TGT GCT CGT CAT TTC TGG ACC AAA ATC TGC AAA CCA GTG CTC	384
	Phe Cys Cys Ala Arg His Phe Trp Thr Lys Ile Cys Lys Pro Val Leu	
	115 120 125	
40	CAT CAG GGG GAA GTC TGT ACC AAA CAA CGC AAG AAG GGT TCT CAT GGG	432
	His Gln Gly Glu Val Cys Thr Lys Gln Arg Lys Lys Gly Ser His Gly	
	130 135 140	
45	CTG GAA ATT TTC CAG CGT TGC GAC TGT GCG AAG GGC CTG TCT TGC AAA	480
	Leu Glu Ile Phe Gln Arg Cys Asp Cys Ala Lys Gly Leu Ser Cys Lys	
	145 150 155 160	
50	GTA TGG AAA GAT GCC ACC TAC TCC TCC AAA GCC AGA CTC CAT GTG TGT	528
	Val Trp Lys Asp Ala Thr Tyr Ser Ser Lys Ala Arg Leu His Val Cys	
	165 170 175	
55	CAG AAA ATT	537
	Gln Lys Ile	

(2) INFORMACIÓN PARA SEQ ID NO: 13:

(i) CARACTERÍSTICAS DE SECUENCIA:

(A) LONGITUD: 928 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN_c

(ix) CARACTERÍSTICA:

ES 2 285 768 T3

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 75..800

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 13:

5

CTCGAGGCCA AAATTCGGCA CGAGGCCGGG CTGTGGTCTA GCATAAAGGC GGAGCCCAGA 60

10 AGAAGGGGCG GGGT ATG GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG 110
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg
1 5 10

15 CAT CTG CTG GTC CTG CTG CTG CTC CTC TCT ACC CTG GTG ATC CCC TCC 158
His Leu Leu Val Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser
15 20 25

20 GCT GCA GCT CCT ATC CAT GAT GCT GAC GCC CAA GAG AGC TCC TTG GGT 206
Ala Ala Ala Pro Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly
30 35 40

25 CTC ACA GGC CTC CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG 254
Leu Thr Gly Leu Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu
45 50 55 60

AAA GGT AAC CTG CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG 302
Lys Gly Asn Leu Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met
65 70 75

30 GAC TTC CGG GGC CTC CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG 350
Asp Phe Arg Gly Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu
80 85 90

35

40

45

50

55

60

65

ES 2 285 768 T3

	CAC CAG CTG GGG AAC AAC ACC CTC TCC AGC CAC CTC CAG ATC GAC AAG	398
	His Gln Leu Gly Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys	
	95 100 105	
5	ATG ACC GAC AAC AAG ACA GGA GAG GTG CTG ATC TCC GAG AAT GTG GTG	446
	Met Thr Asp Asn Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val	
	110 115 120	
10	GCA TCC ATT CAA CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA	494
	Ala Ser Ile Gln Pro Ala Glu Gly Ser Phe Glu Gly Asp Leu Lys Val	
	125 130 135 140	
15	CCC AGG ATG GAG GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG	542
	Pro Arg Met Glu Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr	
	145 150 155	
20	GAC AGC TTC CAC ACA GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT	590
	Asp Ser Phe His Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile	
	160 165 170	
25	AAG CTG CCA CGG CGG AGG TCC CAC CAG GAT GCC CTG GAG GGC GCC CAC	638
	Lys Leu Pro Arg Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His	
	175 180 185	
30	TGG CTC AGC GAG AAG CGA CAC CGC CTG CAG GCC ATC CGG GAT GGA CTC	686
	Trp Leu Ser Glu Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu	
	190 195 200	
35	CGC AAG GGG ACC CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC	734
	Arg Lys Gly Thr His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser	
	205 210 215 220	
40	TCC CAC TCC AGG CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC	782
	Ser His Ser Arg Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu	
	225 230 235	
45	AGG CCC TCT CGG CAG CTG TAGGGGTGGG GACCGGGGAG CACCTGCCTG	830
	Arg Pro Ser Arg Gln Leu	
	240	
50	TAGCCCCCAT CAGACCCTGC CCCAAGCACC ATATGGAAAT AAAGTTCTTT CTTACATCTA	890
	AAAAAAAAAA AAAAAAAAAA AAAAAAATTG GCGGCCGC	928

(2) INFORMACIÓN PARA SEQ ID NO: 14:

(i) CARACTERÍSTICAS DE SECUENCIA:

(A) LONGITUD: 242 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 14:

ES 2 285 768 T3

Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His Leu Leu Val
 1 5 10 15
 5 Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala Ala Ala Pro
 20 25 30
 10 Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu Thr Gly Leu
 35 40 45
 Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys Gly Asn Leu
 50 55 60
 15 Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp Phe Arg Gly
 65 70 75 80
 20 Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His Gln Leu Gly
 85 90 95
 Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Met Thr Asp Asn
 100 105 110
 25 Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val Ala Ser Ile Gln
 115 120 125
 30 Pro Ala Glu Gly Ser Phe Glu Gly Asp Leu Lys Val Pro Arg Met Glu
 130 135 140
 Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr Asp Ser Phe His
 145 150 155 160
 35 Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile Lys Leu Pro Arg
 165 170 175
 40 Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His Trp Leu Ser Glu
 180 185 190
 Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu Arg Lys Gly Thr
 195 200 205
 45 His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser Ser His Ser Arg
 210 215 220
 50 Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu Arg Pro Ser Arg
 225 230 235 240
 Gln Leu

55

(2) INFORMACIÓN PARA SEQ ID NO: 15:

(i) CARACTERÍSTICAS DE SEQUENCIA:

60

(A) LONGITUD: 726 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

65

(ii) TIPO DE MOLÉCULA: ADN_C

(ix) CARACTERÍSTICA:

ES 2 285 768 T3

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1..726

(xi) DESCRIPCIÓN DE SECUENCIA :SEQ ID NO: 15:

5

10

15

20

25

30

35

40

45

50

55

60

65

ATG GGA GAA GCC TCC CCA CCT GCC CCC GCA AGG CGG CAT CTG CTG GTC	48
Met Gly Glu Ala Ser Pro Pro Ala Pro Ala Arg Arg His Leu Leu Val	
1 5 10 15	
CTG CTG CTG CTC CTC TCT ACC CTG GTG ATC CCC TCC GCT GCA GCT CCT	96
Leu Leu Leu Leu Leu Ser Thr Leu Val Ile Pro Ser Ala Ala Pro	
20 25 30	
ATC CAT GAT GCT GAC GCC CAA GAG AGC TCC TTG GGT CTC ACA GGC CTC	144
Ile His Asp Ala Asp Ala Gln Glu Ser Ser Leu Gly Leu Thr Gly Leu	
35 40 45	
CAG AGC CTA CTC CAA GGC TTC AGC CGA CTT TTC CTG AAA GGT AAC CTG	192
Gln Ser Leu Leu Gln Gly Phe Ser Arg Leu Phe Leu Lys Gly Asn Leu	
50 55 60	
CTT CGG GGC ATA GAC AGC TTA TTC TCT GCC CCC ATG GAC TTC CGG GGC	240
Leu Arg Gly Ile Asp Ser Leu Phe Ser Ala Pro Met Asp Phe Arg Gly	
65 70 75 80	
CTC CCT GGG AAC TAC CAC AAA GAG GAG AAC CAG GAG CAC CAG CTG GGG	288
Leu Pro Gly Asn Tyr His Lys Glu Glu Asn Gln Glu His Gln Leu Gly	
85 90 95	
AAC AAC ACC CTC TCC AGC CAC CTC CAG ATC GAC AAG ATG ACC GAC AAC	336
Asn Asn Thr Leu Ser Ser His Leu Gln Ile Asp Lys Met Thr Asp Asn	
100 105 110	
AAG ACA GGA GAG GTG CTG ATC TCC GAG AAT GTG GTG GCA TCC ATT CAA	384
Lys Thr Gly Glu Val Leu Ile Ser Glu Asn Val Val Ala Ser Ile Gln	
115 120 125	
CCA GCG GAG GGG AGC TTC GAG GGT GAT TTG AAG GTA CCC AGG ATG GAG	432
Pro Ala Glu Gly Ser Phe Glu Gly Asp Leu Lys Val Pro Arg Met Glu	
130 135 140	
GAG AAG GAG GCC CTG GTA CCC ATC CAG AAG GCC ACG GAC AGC TTC CAC	480
Glu Lys Glu Ala Leu Val Pro Ile Gln Lys Ala Thr Asp Ser Phe His	
145 150 155 160	
ACA GAA CTC CAT CCC CGG GTG GCC TTC TGG ATC ATT AAG CTG CCA CGG	528
Thr Glu Leu His Pro Arg Val Ala Phe Trp Ile Ile Lys Leu Pro Arg	
165 170 175	

ES 2 285 768 T3

	CGG AGG TCC CAC CAG GAT GCC CTG GAG GGC GGC CAC TGG CTC AGC GAG	576
	Arg Arg Ser His Gln Asp Ala Leu Glu Gly Gly His Trp Leu Ser Glu	
	180 185 190	
5	AAG CGA CAC CGC CTG CAG GCC ATC CGG GAT GGA CTC CGC AAG GGG ACC	624
	Lys Arg His Arg Leu Gln Ala Ile Arg Asp Gly Leu Arg Lys Gly Thr	
	195 200 205	
10	CAC AAG GAC GTC CTA GAA GAG GGG ACC GAG AGC TCC TCC CAC TCC AGG	672
	His Lys Asp Val Leu Glu Glu Gly Thr Glu Ser Ser Ser His Ser Arg	
	210 215 220	
15	CTG TCC CCC CGA AAG ACC CAC TTA CTG TAC ATC CTC AGG CCC TCT CGG	720
	Leu Ser Pro Arg Lys Thr His Leu Leu Tyr Ile Leu Arg Pro Ser Arg	
	225 230 235 240	
	CAG CTG	726
	Gln Leu	

20

(2) INFORMACIÓN PARA SEQ. ID NO: 16:

- (i) CARACTERÍSTICAS DE SECUENCIA:
- (A) LONGITUD: 2381 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

30

(ii) TIPO DE MOLÉCULA: ADN_c

(ix) CARACTERÍSTICA:

- (A) NOMBRE/CLAVE: CDS
- (B) UBICACIÓN: 110..1156

35

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 16:

40

	FGTCGACCCA CGCGTCCGCT GTGGCAGCCC AGCTACCGGT CGTGACCAGA TCCAGCTTGC	60
	AGCTCAGCTT TGTTTCATTCG AATTGGGCGG CGGCCAGCGC GGAACAAC ATG CAG	115
	Met Gln	
	1	
	CGG CTC GGG GGT ATT TTG CTG TGT ACA CTG CTG GCG GCG GCG GTC CCC	163
	Arg Leu Gly Gly Ile Leu Leu Cys Thr Leu Leu Ala Ala Ala Val Pro	
	5 10 15	
	ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG GAG CCG	211
	Thr Ala Pro Ala Pro Ser Pro Thr Val Thr Trp Thr Pro Ala Glu Pro	
	20 25 30	
	GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT GAG ATG	259
	Gly Pro Ala Leu Asn Tyr Pro Gln Glu Glu Ala Thr Leu Asn Glu Met	
	35 40 45 50	

60

65

ES 2 285 768 T3

	TTT CGA GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA CTG CGC	307
	Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys Leu Arg	
	55 60 65	
5	AGT GCC GTG GAG GAG ATG GAG GCG GAA GAA GCA GCT GCT AAA ACG TCC	355
	Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Lys Thr Ser	
	70 75 80	
10	TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC AAC TAT CAC AAT GAG ACC	403
	Ser Glu Val Asn Leu Ala Ser Leu Pro Pro Asn Tyr His Asn Glu Thr	
	85 90 95	
15	AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC CAG GAA	451
	Ser Thr Glu Thr Arg Val Gly Asn Asn Thr Val His Val His Gln Glu	
	100 105 110	
20	GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT TCT GAG	499
	Val His Lys Ile Thr Asn Asn Gln Ser Gly Gln Val Val Phe Ser Glu	
	115 120 125 130	
	ACA GTC ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC CAT GAA	547
	Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Lys Arg Ser His Glu	
	135 140 145	
25	TGT ATC ATT GAT GAA GAC TGT GGG CCC ACC AGG TAC TGC CAG TTC TCC	595
	Cys Ile Ile Asp Glu Asp Cys Gly Pro Thr Arg Tyr Cys Gln Phe Ser	
	150 155 160	
30	AGC TTC AAG TAC ACC TGC CAG CCA TGC CGG GAC CAG CAG ATG CTA TGC	643
	Ser Phe Lys Tyr Thr Cys Gln Pro Cys Arg Asp Gln Gln Met Leu Cys	
	165 170 175	
35	ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG GGT CAC	691
	Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Ala Trp Gly His	
	180 185 190	
40	TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT GAC AAC	739
	Cys Thr Gln Lys Ala Thr Lys Gly Gly Asn Gly Thr Ile Cys Asp Asn	
	195 200 205 210	
	CAG AGG GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA GGC CTG	787
	Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg Gly Leu	
	215 220 225	
45	CTG TTC CCC GTG TGC ACA CCC CTG CCC GTG GAG GGA GAG CTC TGC CAT	835
	Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu Cys His	
	230 235 240	
50	GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC ACC TGG GAA CTG GAG CCT	883
	Asp Pro Thr Ser Gln Leu Leu Asp Leu Ile Thr Trp Glu Leu Glu Pro	
	245 250 255	
55	GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA TGC CAG	931
	Glu Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu Cys Gln	
	260 265 270	

ES 2 285 768 T3

	CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC GTG GGC	979
	Pro His Ser His Ser Leu Val Tyr Met Cys Lys Pro Ala Phe Val Gly	
	275 280 285 290	
5	AGC CAT GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC CCG GAT	1027
	Ser His Asp His Ser Glu Glu Ser Gln Leu Pro Arg Glu Ala Pro Asp	
	295 300 305	
10	GAG TAC GAA GAT GTT GGC TTC ATA GGG GAA GTG CGC CAG GAG CTG GAA	1075
	Glu Tyr Glu Asp Val Gly Phe Ile Gly Glu Val Arg Gln Glu Leu Glu	
	310 315 320	
15	GAC CTG GAG CGG AGC CTA GCC CAG GAG ATG GCA TTT GAG GGG CCT GCC	1123
	Asp Leu Glu Arg Ser Leu Ala Gln Glu Met Ala Phe Glu Gly Pro Ala	
	325 330 335	
20	CCT GTG GAG TCA CTA GGC GGA GAG GAG GAG ATT TAGGCCCCAGA CCCAGCTGAG	1176
	Pro Val Glu Ser Leu Gly Gly Glu Glu Ile	
	340 345	
	TCACTGGTAG ATGTGCAATA GAAATGGCTA ATTTATTTTC CCAGGAGTGT CCCCAGTGT	1236
	GGAATGGCCG CAGCTCCTTC CCAGTAGCTT TTCCTCTGGC TTGACAAGGT ACAGTGCAGT	1296
25	ACATTTCTTC CAGCCGCCCT GCTTCTCTGA CTGGGAAAG ACAGGCATGG CGGGTAAGGG	1356
	CAGCGGTGAG TCGTCCCTCG CTGTTGCTAG AAACGCTGTC TTGTTCTTCA TGGATGGAAG	1416
30	ATTTGTTTGA AGGGAGAGGA TGGAAGGGG TGAAGTCTGC TCATGATGGA TTTGGGGGAT	1476
	ACAGGGAGGA GGATGCCTGC CTTGCAGACG TGGACTTGGC AAAATGTAAC CTTTGCTTTT	1536
35	GTCTTGCGCC GCTCCCATGG GCTGAGGCAG TGGCTACACA AGAGCTATGC TGCTCTGTGG	1596
	CCTCCCATCAT ATTCATCCCT GTGTTTCAGC TCCTACCTCA CTGTCAGCAC AGCCCTTCAT	1656
	AGCCACGCCC CCTCTTGCTC ACCACAGCCT AGGAGGGGAC CAGAGGGGAC TTCTCTCAGA	1716
40	GCCCCATGCT CTCTCTCTCA ACCCCATACC AGCCTCTGTG CCAGCGACAG TCCTTCCAAA	1776
	TGGAGGGAGT GAAATCCTTT GGTTAATTA TTTTCTCCTT CAAGGCACGC CTGCCACTAA	1836
45	GGTCAGGCTG ACTTGCATGT CCCTCTAACG TTCGTAGCAG TGTGGTGGAC ACTGTCTTCC	1896
	ACCGACTGCT TCAATACCTC TGAAAGCCAG TGCTCGGAGT GCAGTTCGTG TAAATTAATT	1956
	TGCAGGAAGT ATACTTGGCT AATTGTAGGG CTAGGATTGT GAATGAAATT TGCAAAGTCC	2016
50	CTTAGCAACA ATGGAAAGCC TTTCTCAGTC ACACCGAGAA GTCACAACCA AGCCAGGTTG	2076
	TGTAGAGTAC AGCTGTGACA TACAGACAGA AGAAGGCTGG GCTGGATGTC AGGCCTCAGA	2136
55	TGACGGTTTC AGGTGCCAGG AACTATTACC ATTCTGTATC TATCCAGACT TATTAAAATT	2196
	GAAAGTTGCA CACATTTGTA TAAGCATGCC TTTCTCCTGA GTTTTAAATT ATATGTATAC	2256
60	ACAAACATGT GGCCCTCAAA GATCATGCAC AAACCACTAC TCTTTGCTAA TTCTTGGACT	2316
	TTTCTCTTTG ATTTTCAATA AATACAAATC CCCTTCATGC AAAAAAAAAA AAAAAGGGCG	2376
	GCCGC	2381

65

(2) INFORMACIÓN PARA SEQ ID NO: 17:

ES 2 285 768 T3

(i) CARACTERISTICAS DE SECUENCIA

(A) LONGITUD: 349 aminoácidos

(B) TIPO: aminoácido

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 17:

```

Met Gln Arg Leu Gly Gly Ile Leu Leu Cys Thr Leu Leu Ala Ala Ala
 1           5           10           15

Val Pro Thr Ala Pro Ala Pro Ser Pro Thr Val Thr Trp Thr Pro Ala
          20           25           30

Glu Pro Gly Pro Ala Leu Asn Tyr Pro Gln Glu Glu Ala Thr Leu Asn
          35           40           45

Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys
 50           55           60

Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Ala Lys
 65           70           75           80

Thr Ser Ser Glu Val Asn Leu Ala Ser Leu Pro Pro Asn Tyr His Asn
          85           90           95

Glu Thr Ser Thr Glu Thr Arg Val Gly Asn Asn Thr Val His Val His
          100          105          110

Gln Glu Val His Lys Ile Thr Asn Asn Gln Ser Gly Gln Val Val Phe
          115          120          125

Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Lys Arg Ser
          130          135          140

His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Thr Arg Tyr Cys Gln
          145          150          155          160

Phe Ser Ser Phe Lys Tyr Thr Cys Gln Pro Cys Arg Asp Gln Gln Met
          165          170          175

Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Ala Trp
          180          185          190

```

ES 2 285 768 T3

Gly His Cys Thr Gln Lys Ala Thr Lys Gly Gly Asn Gly Thr Ile Cys
 195 200 205
 5 Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg
 210 215 220
 Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu
 225 230 235 240
 10 Cys His Asp Pro Thr Ser Gln Leu Leu Asp Leu Ile Thr Trp Glu Leu
 245 250 255
 15 Glu Pro Glu Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu
 260 265 270
 Cys Gln Pro His Ser His Ser Leu Val Tyr Met Cys Lys Pro Ala Phe
 275 280 285
 20 Val Gly Ser His Asp His Ser Glu Glu Ser Gln Leu Pro Arg Glu Ala
 290 295 300
 25 Pro Asp Glu Tyr Glu Asp Val Gly Phe Ile Gly Glu Val Arg Gln Glu
 305 310 315 320
 Leu Glu Asp Leu Glu Arg Ser Leu Ala Gln Glu Met Ala Phe Glu Gly
 325 330 335
 30 Pro Ala Pro Val Glu Ser Leu Gly Gly Glu Glu Glu Ile
 340 345
 35

(2) INFORMACIÓN PARA SEQ ID NO: 18:

(i) CARACTERÍSTICAS DE SECUENCIA:

- 40 (A) LONGITUD: 1047 pares de bases
 (B) TIPO: ácido nucleico
 (C) TIPO DE CADENA: sencilla
 45 (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN_C

(ix) CARACTERÍSTICA:

- 50 (A) NOMBRE/CLAVE: CDS
 (B) LOCALIZACIÓN: 1..1047

(xi) DESCRIPCIÓN DE SECUENCIA: SEQ ID NO: 18:

55 ATG CAG CCG CTC GGG GGT ATT TTG CTG TGT ACA CTG CTG GCG GCG GCG 48
 Met Gln Arg Leu Gly Gly Ile Leu Leu Cys Thr Leu Leu Ala Ala Ala
 1 5 10 15
 60 GTC CCC ACT GCT CCT GCT CCT TCC CCG ACG GTC ACT TGG ACT CCG GCG 96
 Val Pro Thr Ala Pro Ala Pro Ser Pro Thr Val Thr Trp Thr Pro Ala
 20 25 30
 65

ES 2 285 768 T3

	GAG CCG GGC CCA GCT CTC AAC TAC CCT CAG GAG GAA GCT ACG CTC AAT	144
	Glu Pro Gly Pro Ala Leu Asn Tyr Pro Gln Glu Glu Ala Thr Leu Asn	
	35 40 45	
5	GAG ATG TTT CGA GAG GTG GAG GAG CTG ATG GAA GAC ACT CAG CAC AAA	192
	Glu Met Phe Arg Glu Val Glu Glu Leu Met Glu Asp Thr Gln His Lys	
	50 55 60	
10	CTG CGC AGT GCC GTG GAG GAG ATG GAG GCG GAA GAA GCA GCT GCT AAA	240
	Leu Arg Ser Ala Val Glu Glu Met Glu Ala Glu Glu Ala Ala Ala Lys	
	65 70 75 80	
15	ACG TCC TCT GAG GTG AAC CTG GCA AGC TTA CCT CCC AAC TAT CAC AAT	288
	Thr Ser Ser Glu Val Asn Leu Ala Ser Leu Pro Pro Asn Tyr His Asn	
	85 90 95	
20	GAG ACC AGC ACG GAG ACC AGG GTG GGA AAT AAC ACA GTC CAT GTG CAC	336
	Glu Thr Ser Thr Glu Thr Arg Val Gly Asn Asn Thr Val His Val His	
	100 105 110	
25	CAG GAA GTT CAC AAG ATA ACC AAC AAC CAG AGT GGA CAG GTG GTC TTT	384
	Gln Glu Val His Lys Ile Thr Asn Asn Gln Ser Gly Gln Val Val Phe	
	115 120 125	
30	TCT GAG ACA GTC ATT ACA TCT GTA GGG GAT GAA GAA GGC AAG AGG AGC	432
	Ser Glu Thr Val Ile Thr Ser Val Gly Asp Glu Glu Gly Lys Arg Ser	
	130 135 140	
35	CAT GAA TGT ATC ATT GAT GAA GAC TGT GGG CCC ACC AGG TAC TGC CAG	480
	His Glu Cys Ile Ile Asp Glu Asp Cys Gly Pro Thr Arg Tyr Cys Gln	
	145 150 155 160	
40	TTC TCC AGC TTC AAG TAC ACC TGC CAG CCA TGC CGG GAC CAG CAG ATG	528
	Phe Ser Ser Phe Lys Tyr Thr Cys Gln Pro Cys Arg Asp Gln Gln Met	
	165 170 175	
45	CTA TGC ACC CGA GAC AGT GAG TGC TGT GGA GAC CAG CTG TGT GCC TGG	576
	Leu Cys Thr Arg Asp Ser Glu Cys Cys Gly Asp Gln Leu Cys Ala Trp	
	180 185 190	
50	GGT CAC TGC ACC CAA AAG GCC ACC AAA GGT GGC AAT GGG ACC ATC TGT	624
	Gly His Cys Thr Gln Lys Ala Thr Lys Gly Gly Asn Gly Thr Ile Cys	
	195 200 205	
55	GAC AAC CAG AGG GAT TGC CAG CCT GGC CTG TGT TGT GCC TTC CAA AGA	672
	Asp Asn Gln Arg Asp Cys Gln Pro Gly Leu Cys Cys Ala Phe Gln Arg	
	210 215 220	
60	GGC CTG CTG TTC CCC GTG TGC ACA CCC CTG CCC GTG GAG GGA GAG CTC	720
	Gly Leu Leu Phe Pro Val Cys Thr Pro Leu Pro Val Glu Gly Glu Leu	
	225 230 235 240	
65	TGC CAT GAC CCC ACC AGC CAG CTG CTG GAT CTC ATC ACC TGG GAA CTG	768
	Cys His Asp Pro Thr Ser Gln Leu Leu Asp Leu Ile Thr Trp Glu Leu	
	245 250 255	

ES 2 285 768 T3

	GAG CCT GAA GGA GCT TTG GAC CGA TGC CCC TGC GCC AGT GGC CTC CTA	816
	Glu Pro Glu Gly Ala Leu Asp Arg Cys Pro Cys Ala Ser Gly Leu Leu	
	260 265 270	
5	TGC CAG CCA CAC AGC CAC AGT CTG GTG TAC ATG TGC AAG CCA GCC TTC	864
	Cys Gln Pro His Ser His Ser Leu Val Tyr Met Cys Lys Pro Ala Phe	
	275 280 285	
10	GTG GGC AGC CAT GAC CAC AGT GAG GAG AGC CAG CTG CCC AGG GAG GCC	912
	Val Gly Ser His Asp His Ser Glu Glu Ser Gln Leu Pro Arg Glu Ala	
	290 295 300	
15	CCG GAT GAG TAC GAA GAT GTT GGC TTC ATA GGG GAA GTG CGC CAG GAG	960
	Pro Asp Glu Tyr Glu Asp Val Gly Phe Ile Gly Glu Val Arg Gln Glu	
	305 310 315 320	
20	CTG GAA GAC CTG GAG CGG AGC CTA GCC CAG GAG ATG GCA TTT GAG GGG	1008
	Leu Glu Asp Leu Glu Arg Ser Leu Ala Gln Glu Met Ala Phe Glu Gly	
	325 330 335	
25	CCT GCC CCT GTG GAG TCA CTA GGC GGA GAG GAG GAG ATT	1047
	Pro Ala Pro Val Glu Ser Leu Gly Gly Glu Glu Glu Ile	
	340 345	

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 1 2479 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 18,7%; puntuación de alineación global: -5757

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GGCACGAGGGGGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCT
5 D1 -----

      70      80      90     100     110     120
10 ./wwwt GCTGTGCCTGCTGCTGGCGGGCGGTCCTCCACGGCCCCCGCGCCCOCTCCGACGGCGAC
D1 -----

      130     140     150     160     170     180
15 ./wwwt CTCGGCTCCAGTCAAGCCCCGGCCCGGCTCTCAGCTACCCGCAGGAGGAGCCACCCTCAA
D1 -----

      190     200     210     220     230     240
20 ./wwwt TGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGACACGCAGCACAAATTGCGCAGCGC
25 D1 -----

      250     260     270     280     290     300
30 ./wwwt GGTGGAAGAGATGGAGGCAGAAGAAGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
D1 -----

      310     320     330     340     350     360
35 ./wwwt AAACCTTACCTCCCAGCTATCACAAATGAGACCAACACAGACACGAACGTTGGAAATAATAC
D1 -----

      370     380     390     400     410     420
40 ./wwwt CATCCATGTGCACCGAGAAATTCACAAGATAACCAACAACCAGACTGGACAAATGGTCTT
45 D1 -----

      430     440     450     460     470     480
50 ./wwwt TTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCAT
D1 -----

```

ES 2 285 768 T3

```

      490      500      510      520      530      540
./wwwt CATCGACGAGGACTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
5  D1 -----

      550      560      570      580      590      600
10 ./wwwt CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACCA
    D1 -----

      610      620      630      640      650      660
15 ./wwwt GCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAATGGGACCATCTG
    D1 -----

      670      680      690      700      710      720
20 ./wwwt TGACAACCAGAGGGAAGTCCAGCCGGGGCTGTGCTGTGCCTTCCAGAGAGGCCTGCTGTT
    D1 -----

      730      740      750      760      770      780
25 ./wwwt CCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTTTGCCATGACCCCGCCAGCCGGCT
    D1 -----

      790      800      810      820      830      840
30 ./wwwt TCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGC
    D1 -----

      850      860      870      880      890      900
35 ./wwwt CAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTGGTGTATGTGTGCAAGCCGACCTT
    D1 -----

      910      920      930      940      950      960
40 ./wwwt CGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCAGAGAGGTCCCCGATGAGTA
    D1 -----

      970      980      990      1000      1010      1020
45 ./wwwt TGAAGTTGGCAGCTTCATGGAGGAGGTGCCAGGAGCTGGAGGACCTGGAGAGGAGCCT
50
55
60
65

```

ES 2 285 768 T3

D1 -----

5
1030 1040 1050 1060 1070 1080
./wwwt GACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGA

D1 -----

10
1090 1100 1110 1120 1130 1140
./wwwt AGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTTATTT

15
D1 -----

20
1150 1160 1170 1180 1190 1200
./wwwt CCCCANGTGTGTGCTTTAAGCGTGGGCTGACCAGGCTTCTTCCTACATCTTCTTCCCAGT

D1 -----

25
1210 1220 1230 1240 1250 1260
./wwwt AAGTTTCCCCTCTGGCTTGACAGCATGAGGTGTTGTGCATTTGTTTCAGCTCCCCCAGGCT

D1 -----

30
1270 1280 1290 1300 1310 1320
./wwwt GTTCTCCAGGCTTCACAGTCTGGTGCTTGGGAGAGTCAGGCAGGGTTAAACTGCAGGAGC

35
D1 -----

40
1330 1340 1350 1360 1370 1380
./wwwt AGTTTGGCACCCCTGTCCAGATTATTGGCTGCTTTGCCTCTACCAGTTGGCAGACAGCCG

D1 -----

45
1390 1400 1410 1420 1430 1440
./wwwt TTTGTTCTACATGGCTTTGATAATTGTTTGAGGGGAGGAGATGGAACAATGTGGAGTCT

D1 -----

50
1450 1460 1470 1480 1490 1500
./wwwt CCCTCTGATTGGTTTPTGGGGAAATGTGGAGAAGAGTGCCCTGCTTTGCAAACATCAACCT

55
D1 -----

60
1510 1520 1530 1540 1550 1560
./wwwt GGCAAAAATGCAACAAATGAATTTTCCACGCAGTTCTTTCCATGGGCATAGGTAAGCTGT

65

ES 2 285 768 T3

D1 -----

5
1570 1580 1590 1600 1610 1620
./wwwt GCCTTCAGCTGTTGCAGATGAAATGTTCTGTTACCCTGCATTACATGTGTTTATTCATC

D1 -----

10
1630 1640 1650 1660 1670 1680
./wwwt CAGCAGTGTTGCTCAGCTCCTACCTCTGTGCCAGGGCAGCATTTCATATCCAAGATCAA

15
D1 -----

20
1690 1700 1710 1720 1730 1740
./wwwt TTCCCTCTCTCAGCACAGCCTGGGGAGGGGGTCATTGTTCTCTCTCGTCCATCAGGGATT

D1 -----

25
1750 1760 1770 1780 1790 1800
./wwwt CAGAGGCTCAGAGACTGCAAGCTGCTTGCCCAAGTCACACAGCTAGTGAAGACCAGAGCA

D1 -----

30
1810 1820 1830 1840 1850 1860
./wwwt GTTTCATCTGGTTGTGACTCTAAGCTCAGTGCTCTCTCCACTACCCACACCAGCCTTGG

D1 -----

35
1870 1880 1890 1900 1910 1920
./wwwt TGCCACCAAAGTGCTCCCCAAAAGGAAGGAGAATGGGATTTTTCTTTTGAGGCATGCAC

40
D1 -----

45
1930 1940 1950 1960 1970 1980
./wwwt ATCTGGAATTAAGGTCAAATAATTCTCACATCCCTCTAAAAGTAACTACTGTTAGGAA

D1 -----

50
1990 2000 2010 2020 2030 2040
./wwwt CAGCAGTGTTCTCACAGTGTTGGGGCAGCCGTCCTTCTAATGAAGACAATGATATTGACAC

.....

D1 ---CAGTGTTCTCACAGTGTTGGGGCAGCCGTCCTTCTAATGAAGACAATGATATTGACAC

55
10 20 30 40 50

2050 2060 2070 2080 2090 2100

60

65

ES 2 285 768 T3

```

./wwwT   TGTCCCTCTTTGGCAGTTGCATTAGTAACTTTGAAAGGTATATGACTGAGCGTAGCATAC
D1       TGTCCCTCTTTGGCAGTTGCATTAGTAACTTTGAAAGGTATATGACTGAGCGTAGCATAC
          60          70          80          90         100         110

          2110         2120         2130         2140         2150         2160
./wwwT   AGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAATGAAATTT
D1       AGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAATGAAATTT
          120         130         140         150         160         170

          2170         2180         2190         2200         2210         2220
./wwwT   GCAAAATCACTTAGCAGCAACTGAAGACAATTATCAACCACGTGGAGAAAATCAAACCGA
D1       GCAAAATCACTTAGCAGCAACTGAAGNCAATTATCAACCACGTGGAGAAAATCAAACCGA
          180         190         200         210         220         230

          2230         2240         2250         2260         2270         2280
./wwwT   GCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCACT
D1       GCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCACT
          240         250         260         270         280         290

          2290         2300         2310         2320         2330         2340
./wwwT   CCACAAATGATGTTTTTCAGGTGTCATGGACTGTTGCCACCATGTATTCATCCAGAGTTCT
D1       CCACAAATGATGTTTTTCAGGTGTCATGGACTGTTGCCACCATGTATTCATCCAGAGTTCT
          300         310         320         330         340         350

          2350         2360         2370         2380         2390         2400
./wwwT   TAAAGTTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTTAAATTATGTA
D1       TAAAGTTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTTAAATTATGTA
          360         370         380         390         400         410

          2410         2420         2430         2440         2450         2460
./wwwT   TAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCTAAAAAAAAAAAAA
D1       TAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCT-----
          420         430         440         450         460

          2470
./wwwT   AAAAAAAAAAAAAAAAAAAAAA
D1       -----

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 1 2479 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 14,3%; puntuación de alineación global: -7400

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GGCACGAGGGGGCGGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCT
      :: :
5 D2    A-----ACTCC

      70      80      90      100     110     120
./wwwt GCTGTGCCTGCTGCTGGCGGGCGGCTCCCCACGGCCCCCGCGCCCGCTCCGACGGCGAC
      :::: :
10 D2    CCTGTGATTGC-----
      10

      130     140     150     160     170     180
./wwwt CTCGGCTCCAGTCAAGCCCCGGCCCCGGCTCTCAGCTACCCGAGGAGGAGGCCACCCTCAA
      :: : : : :
15 D2    -----AGTAAAT-----TAC-----
      20

      190     200     210     220     230     240
./wwwt TGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGACACGCAGCACAAATTGCGCAGCCG
      :: : : : :
20 D2    -----TGT-----ATTG-----
      30

      250     260     270     280     290     300
./wwwt GGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
      :: : : : :
30 D2    -----TAAATTCTCA-----GTG-----TGGC
      40      50

      310     320     330     340     350     360
./wwwt AAACCTTACCTCCAGCTATCACAATGAGACCAACACAGACACGAACGTTGGAAATAATAC
      : : : : : : : : : : : : : : : :
35 D2    A--CTTACCT--GTAAATGCAATGA-----AACTTT--TAATTAT--
      60      70      80

      370     380     390     400     410     420
./wwwt CATCCATGTGCACCGAGAAATTACAAGATAACCAACAACCAGACTGGACAAATGGTCTT
      : :
40 D2    -----TT

      430     440     450     460     470     480
./wwwt TTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCAT
      :: : : : :
50 D2    TTC-----TA-----AAGGTGC-----

```

ES 2 285 768 T3

90

```

5      490      500      510      520      530      540
./wwwt CATCGACGAGGACTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
      : : : : : : : : : : : : : : : : : :
D2      -----TGCACTGCCTATTT-----TTCC-----
      100      110

10     550      560      570      580      590      600
./wwwt CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACCA
      : : : : : : : : : : : : : : : : : :
D2      -----TCT-----TGT-----
      120

15     610      620      630      640      650      660
./wwwt GCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAATGGGACCATCTG
      : : : : : : : : : : : : : : : : : :
D2      -----TATGT-----AAATT-----
      130

20     670      680      690      700      710      720
./wwwt TGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGTGCCCTTCCAGAGAGGCCTGCTGTT
D2      -----

25     730      740      750      760      770      780
./wwwt CCCTGTGTGCACACCCCTGCCCCGTGGAGGGCGAGCTTTGCCATGACCCCGCCAGCCGGCT
      : : : : : : : : : : : : : : : : : :
D2      ---TTTGTACACA-----TTGA-----TTGTTAT-----
      140      150

30     790      800      810      820      830      840
./wwwt TCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGC
      : : : : : : : : : : : : : : : : : :
D2      -CTTGAC-----
      160

35     850      860      870      880      890      900
./wwwt CAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTGGTGTATGTGTGCAAGCCGACCTT
D2      -----

40     910      920      930      940      950      960
./wwwt CGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCCAGAGAGGTCCCCGATGAGTA
      : : : : : : : : : : : : : : : : : :
D2      -----TGACAAATAT-----TC-----TA
      170

45     970      980      990      1000      1010      1020
./wwwt TGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCT
      : : : : : : : : : : : : : : : : : :

```

ES 2 285 768 T3

```

D2      T-----ATTGA-----
              180

5          1030      1040      1050      1060      1070      1080
./wwwt GACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGA
      : : : : :
D2      -ACTGAAG-----

10          1090      1100      1110      1120      1130      1140
./wwwt AGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTTATTT
      : : : : :
D2      -----TAAATCATTT
              190

15          1150      1160      1170      1180      1190      1200
./wwwt CCCCANGTGTGTGCTTTAAGCGTGGGCTGACCAGGCTTCTTCCTACATCTTCTTCCCAGT
      : : : : :
D2      CA-----GCTTAT-----
              200

20          1210      1220      1230      1240      1250      1260
./wwwt AAGTTTCCCCCTCTGGCTTGACAGCATGAGGTGTTGTGCATTTGTTTCAGCTCCCCCAGGCT
      : : : : :
D2      -AGTT-----CTTAAAAGCAT-----
              210      220

25          1270      1280      1290      1300      1310      1320
./wwwt GTTCTCCAGGCTTCACAGTCTGCTGCTTGGGAGAGTCAGGCAGGGTTAAACTGCAGGAGC
      : : :
D2      -----AAC-----

30          1330      1340      1350      1360      1370      1380
./wwwt AGTTTGCCACCCCTGTCCAGATTATTGGCTGCTTTGCCTCTACCAGTTGCCAGACAGCCG
      : : : : :
D2      -----CCTTTACC-----CCA
              230

35          1390      1400      1410      1420      1430      1440
./wwwt TTTGTTCTACATGGCTTTGATAATTGTTTGAGGGGAGGAGATGGAACAATGTGGAGTCT
      : : : : :
D2      TTT-----AATTCT-----AGAGTCT
              240      250

40          1450      1460      1470      1480      1490      1500
./wwwt CCCTCTGATTGGTTTTTGGGGAAATGTGGAGAAGAGTGCCCTGCTTTGCAAACATCAACCT
      : : : : :
D2      -----AGAAC-----GCAAGGATCT-CTT
              260

45          1510      1520      1530      1540      1550      1560
./wwwt GGCAAAAATGCAACAAATGAATTTTCCACGCAGTTCTTTCCATGGGCATAGGTAAGCTGT

```

ES 2 285 768 T3

```

5      ::      ::::      :::::
D2      GG-----AATG--ACAAATGA-----TAGGTA-----
      270              280

      1570      1580      1590      1600      1610      1620
./wwwt GCCTTCAGCTGTTGCAGATGAAATGTTCTGTTCCCTGCATTACATGTGTTTATTCATC
      ::      :::::      ::::
10      D2      -CCTA-----AAATGTA-----ACATG-----
      290              300

      1630      1640      1650      1660      1670      1680
./wwwt CAGCAGTGTGCTCAGCTCCTACCTCTGTGCCAGGGCAGCATTTCATATCCAAGATCAA
15      D2      -----AAAAT-----
      :: ::

      1690      1700      1710      1720      1730      1740
./wwwt TTCCCTCTCTCAGCACAGCCTGGGGAGGGGTCATTGTTCTCCTCGTCCATCAGGATTT
20      D2      -----

25      1750      1760      1770      1780      1790      1800
./wwwt CAGAGGCTCAGAGACTGCAAGCTGCTTGCCCAAGTCACACAGCTAGTGAAGACCAGAGCA
      ::      ::::      ::
30      D2      -----ACT--AGCT-----TA-----
      310

      1810      1820      1830      1840      1850      1860
./wwwt GTTTCATCTGTTGTGACTCTAAGCTCAGTGCTCTCTCCACTACCCACACCAGCCTTGG
35      D2      -TTT--TCTG-----
      320

      1870      1880      1890      1900      1910      1920
./wwwt TGCCACCAAAAGTGCTCCCCAAAAGGAAGGAGAATGGGATTTTCTTTTGAGGCATGCAC
40      D2      -----AAATGTACT-----
      330

      1930      1940      1950      1960      1970      1980
./wwwt ATCTGGAATTAAGGTCAAATAATTCTCACATCCCTCTAAAAGTAACTACTGTTAGGAA
45      D2      ATCT-----TAATGCTTA-----
      340

      1990      2000      2010      2020      2030      2040
./wwwt CAGCAGTGTTCACAGTGTGGGGCAGCCGTCCTTCTAATGAAGACAATGATATTGACAC
55      D2      -----AATTATATT-----
      350

      2050      2060      2070      2080      2090      2100
60      D2      -----

65

```

ES 2 285 768 T3

```

./wwwT TGTCCCTCTTTGGCAGTTGCATTAGTAACCTTTGAAAGGTATATGACTGAGCGTAGCATAC
      : : : :
D2  --TCCCT-----
      360

      2110      2120      2130      2140      2150      2160
./wwwT AGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAATGAAATTT
      : : : :
D2  -----TTAGG-----

      2170      2180      2190      2200      2210      2220
./wwwT GCAAAATCACTTAGCAGCAACTGAAGACAATTATCAACCACGTGGAGAAAAATCAAACCGA
D2  -----

      2230      2240      2250      2260      2270      2280
./wwwT GCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCACT
      : : : :      : : : :      : : : :
D2  -----CTGTG-----ATAGTTT-----TGAAAT-----
      370                                380

      2290      2300      2310      2320      2330      2340
./wwwT CCACAAATGATGTTTTTCAGGTGTCATGGACTGTTGCCACCATGTATTTCATCCAGAGTTCT
D2  -----

      2350      2360      2370      2380      2390      2400
./wwwT TAAAGTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTAAATTATGTA
      : : : : :
D2  -AAAATTTAA-----
      390

      2410      2420      2430      2440      2450      2460
./wwwT TAAACATRAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCTAAAAA
      : : : : : : : :
D2  -----CATTTA-ATATCAA-----
      400

      2470
./wwwT AAAAAAAAAAAAAAAAAA
      : : : : :
D2  -----AAAAA
      410

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 1 2479 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 16,0%; puntuación de alineación global: -7115

ES 2 285 768 T3

```

      10          20          30          40          50          60
./wwwt GGCACGAGGGGGCGGCGCTGCGGGCGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCT
      :: ::
5 D3 -----TTNGG-----

      70          80          90          100         110         120
./wwwt GCTGTGCCTGCTGCTGGCGGCGGGCGGTCCCCACGGCCCCCGCGCCCGCTCCGACGGCGAC
10 D3 -----

      130         140         150         160         170         180
./wwwt CTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTACCCGCAGGAGGAGGCCACCCTCAA
15 D3 -----

      190         200         210         220         230         240
./wwwt TGAGATGTTCCGCGAGGTTGAGGAAGTGATGGAGGACACGCAGCACAAATTGCGCAGCGC
      : : : : :           : :
25 D3 -----ATGGAG-----TT-----
              10

      250         260         270         280         290         300
./wwwt GGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAAGTGAACCTGGC
      : : : : :
30 D3 -----GATGAAG-----
              20

      310         320         330         340         350         360
./wwwt AAAC TTACCTCCCAGCTATCACAAATGAGACCAACACAGACACGAACGTTGGAAATAATAC
35 D3 -----

      370         380         390         400         410         420
./wwwt CATCCATGTGCACCGAGAAATTCACAAGATAACCAACAACCAGACTGGACAAATGGTCTT
      : : : : :           : : : : :
45 D3 --TCCAT-----AAAAGGT--T
                      30

      430         440         450         460         470         480
./wwwt TTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 D3 TTCATAG-CATCTATTATAT-----AAAGGAG-----

```

ES 2 285 768 T3

```

      40      50      60
      490      500      510      520      530      540
5  ./wwwt CATCGACGAGGACTGTGGGGCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
      : : : :
D3  -----TTCCA-----

      550      560      570      580      590      600
10 ./wwwt CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACCA
      : : : : : : :
D3  -----CTGCA-----TPTGGATA--A
      70      80

      610      620      630      640      650      660
15 ./wwwt GCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAATGGGACCATCTG
      : : : :
D3  GCTG-----

      670      680      690      700      710      720
20 ./wwwt TGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGTGCCTTCCAGAGAGGCCTGCTGTT
      : : :
D3  -----GTT

      730      740      750      760      770      780
25 ./wwwt CCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTTTGCCATGACCCCGCCAGCCGGCT
      : : :
D3  -----

      790      800      810      820      830      840
30 ./wwwt TCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGC
      : : : : : :
D3  -----TAG-----TGT--
      90

      850      860      870      880      890      900
35 ./wwwt CAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTGGTGTATGTCTGCAAGCCGACCTT
      : : : : :
D3  -----CTCTGACA-----
      100

      910      920      930      940      950      960
40 ./wwwt CGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCAGAGAGGTCCCCGATGAGTA
      : : :
D3  -----AGTG

      970      980      990      1000      1010      1020
45 ./wwwt TGAAGTTGGCAGCTTCATGGAAGAGGTGCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCT
      : : : : : : :
D3  -----

```

ES 2 285 768 T3

```

D3      TGAAG-----CCTAGA-----
        110

5          1030      1040      1050      1060      1070      1080
./wwwt GACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGA

D3      -----

10          1090      1100      1110      1120      1130      1140
./wwwt AGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTTATTT
        : : : : : : :
D3      --AGAAATTA---CTGG-----
        120

        1150      1160      1170      1180      1190      1200
./wwwt CCCCANGTGTGTGCTTTAAGCGTGGGCTGACCAGGCTTCTTCCTACATCTCTTCCCAGT
        : : : : : : :
D3      -----CTTGA---TGG---TGA---TCTTTCT-----GT
        130          140

        1210      1220      1230      1240      1250      1260
./wwwt AAGTTTCCCCTCTGGCTTGACAGCATGAGGTGTTGTGCATTGTTCAGCTCCCCAGGCT
        : : : : : : :
D3      A-----TCCGGCAAGACAGA-----CC-----
        150          160

30          1270      1280      1290      1300      1310      1320
./wwwt GTTCTCCAGGCTTCACAGTCTGGTGCTTGGGAGAGTCAGGCAGGTTAAACTGCAGGAGC
        : : : : : : :
D3      -TCTCCA-----CAGT-----AAC-----
        170

        1330      1340      1350      1360      1370
./wwwt AGTTTGCCACCCCTGTCCAGATTATTGGCTGCTTTGCCTCTACCAGTTGGCA-GACAGCC
        : : : : : : :
D3      -----AACGCTG---GAATATT-----TCTA---GT---CCATGAGAGCC
        180          190          200

        1380      1390      1400      1410      1420      1430
./wwwt GTTGTCTACATGGCTTTGATAATTGTTTGAGGGGAGGAGATGGAACAATGTGGAGTC
        : : : : : : :
D3      TTTTCTCCTA-----
        210

50          1440      1450      1460      1470      1480      1490
./wwwt TCCCTCTGATTGGTTTTTGGGGAAATGTGGAGAAGAGTGCCCTGCTTTGCAAACATCAACC
        : : : : : : :
D3      -----TGCTTGGT-----ACA-----
        220          230

        1500      1510      1520      1530      1540      1550
./wwwt TGGCAAAAATGCAACAAATGAATTTCCACGCAGTTCTTTCCATGGGCATAGCTAAGCTG

```

ES 2 285 768 T3

65

ES 2 285 768 T3

```

5      ./wwwt CTGTCCCTCTTTGGCAGTTGCATTAGTAACCTTTGAAAGGTATATGACTGAGCGTAGCATA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -TGTC---CTTTGG-----TGTGA-----TA
      340                               350

      2100      2110      2120      2130      2140      2150
10     ./wwwt CAGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAAATGAAATT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      CATT---TGAAGACA-----AGGTGGTT-----
      360                               370

      2160      2170      2180      2190      2200      2210
15     ./wwwt TGCAAAATCACCTTAGCAGCAACTGAAGACAATTATCAACCACGTGGAGAAAAATCAAACCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -----CTT-----CTG-----GAATACCCATCCA
      380                               390

      2220      2230      2240      2250      2260      2270
20     ./wwwt AGCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      AG--GTGCTATG--ATCAT-----TAC-----
      400                               410

      2280      2290      2300      2310      2320      2330
30     ./wwwt TCCACAAATGATGTTTTCAGGTGTCATGGACTGTTGCCACCATGTATTCATCCAGAGTTC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      ---CAAA---GCTTTCAG--TGATGG---TTCCTCAAT--TTC-TCC-----
      420                               430                               440

      2340      2350      2360      2370      2380      2390
35     ./wwwt TTAAAGTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTAAATTATGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -----TCGGAATGATT-----TTG-----
      450                               460

      2400      2410      2420      2430      2440      2450
40     ./wwwt ATAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCTAAAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      ---ATCAGAAG-----

      2460      2470
45     ./wwwt AAAAAAAAAAAAAAAAAA
D3      -----
50

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 1 2479 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 14,6%; puntuación de alineación global: -7433

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GGCACGAGGGGGCGGCGCTGCGGGCGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCT
5 D4      -----CTTG-----CGTT

      70      80      90      100     110     120
./wwwt GCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCCCCCGCGCCGCTCCGACGGCGAC
10 : :      : :
D4      GTT-----TGG-----
10

      130     140     150     160     170     180
./wwwt CTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTACCCGCAGGAGGAGGCCACCCTCAA
15 : : : :
D4      -----TACAGAC-----
20      20

      190     200     210     220     230     240
./wwwt TGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGACACGCAGCACAAATTGCGCAGCGC
25 : : : :      : : : : : :
D4      -----TTCCC-----CTGATGGAGCACT-----
30      30      40

      250     260     270     280     290     300
./wwwt GGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
30 : : : : : : : :
D4      GGT-----TTGC

      310     320     330     340     350     360
./wwwt AAACCTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAACGTTGGAAATAATAC
35 : : : :      : :
D4      AGATTT-----TGG-----
40      50

      370     380     390     400     410     420
./wwwt CATCCATGTGCACCGAGAAATTCAACAAGATAACCAACAACCAGACTGGACAAATGGTCTT
45 : : : :      : : : :      : :
D4      --TCCA-----GAAAT-----GAC-----
60      60

      430     440     450     460     470     480
./wwwt TTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCAT
50 : : : :      : : : :      : : : :
D4      ----GAG-----CACA-----GCAAAA-----
55

60

65

```

ES 2 285 768 T3

```

70                                80

5      490      500      510      520      530      540
./wwwt CATCGACGAGGACTGTGGGCCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
D4      -----CCC-----TTCAA-----

10     550      560      570      580      590      600
./wwwt CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAGTGCTGTGGAGACCA
D4      -----TGCA-----

15                                90

20     610      620      630      640      650      660
./wwwt GCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAATGGGACCATCTG
D4      -----GTCTG-----ATG-----ATC-G
100

25     670      680      690      700      710      720
./wwwt TGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGTGCCTTCCAGAGAGGCCTGCTGT
D4      T-----AGG-----CAG--GGGTCT-----CCTTC-----
110                                120

30     730      740      750      760      770      780
./wwwt CCCTGTGTGCACACCCCTGCCCCTGGAGGGCGAGCTTTGCCATGACCCCGCCAGCCGGCT
D4      -----

35

40     790      800      810      820      830      840
./wwwt TCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGC
D4      -----ATGCCCTTTTAT
130

45     850      860      870      880      890      900
./wwwt CAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTGGTGTATGTGTGCAAGCCGACCTT
D4      ATGTG-----ACATCTTAGTGTGTG-GT-----CTT
140                                150                                160

50     910      920      930      940      950      960
./wwwt CGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCAGAGAGGTCCCCGATGAGTA
D4      CCT-----AGATTCTG-----CCA-----TCCC-----
170

55                                170

60     970      980      990      1000      1010      1020
./wwwt TGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCT
65

```

ES 2 285 768 T3

D4 --AAGT-----CATGGT-----TTGAGTAA-----T
 180 190

```

          1030          1040          1050          1060          1070          1080
./wwwt GACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGA
      ::: :          :::

```

D4 GACCG-----TGG-----
200

```

      1090      1100      1110      1120      1130      1140
./wwwt AGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTTATTT
      ::: ::::      ::::

```

D4 -----TTTCGATCT-----CTGTG-----
 210

```

      1150      1160      1170      1180      1190      1200
./www: CCCANGTGTGTGCTTTAAGCGTGGGCTGACCAGGCTTCTTCTCTACATCTTCTTCCAGT
                                     :: ::

```

D4 -----CCGAGT
220

```

      1210      1220      1230      1240      1250      1260
./wwwt AAGTTTCCCTCTGGCTTGACAGCATGAGGTGTTGTGCATTTGTTTCAGCTCCCCCAGGCT
      :
                                     :

```

D4 A-----CCA-----

```

      1270      1280      1290      1300      1310      1320
./wwwt GTTCTCCAGGCTTCACAGTCTGGTGCCTTGGGAGAGTCAGGCAGGGTTAAACTGCAGGAGC
      ::::      : : : :      : : : :      : : : :      : : : :

```

D4 ---TCCAG---AGCCGGG---ATGTGAG---GGGTAA---
 230 240 250

```

      1330      1340      1350      1360      1370      1380
./wwwt AGTTTGCCACCCCTGTCCAGATTATTGGCTGCTTTGCCCTCTACCAAGTTGGCAGACAGCCG
              : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

D4 -----GATGCTTT-----CAGT-----
260

1390 1400 1410 1420 1430 1440
./wwwt TTTGTTCTACATGGCTTTGATAATTGTTTGAGGGGAGGAGATGGAAACAATGTGGAGTCT
:::

D4 AAC-
270

```

      1450      1460      1470      1480      1490      1500
./wwwt CCCTCTGATTGGTTTTGGGGAAATGTGGAGAAGAGTGCCCTGCTTTGCAACATCAACCT
      :
      :
      :

```

D4 -----TGGG-----

1510 1520 1530 1540 1550 1560
./wwwt GGCAAAAATGCAACAAATGAATTTTCCACGCAGTCTCTTCCATGGGCATAGGTAAGCTGT

ES 2 285 768 T3

```

D4      -----ATA-----
5
      1570      1580      1590      1600      1610      1620
./wwwt GCCTTCAGCTGTTGCAGATGAAATGTTCTGTTACCCCTGCATTACATGTGTTTATTCATC
      : : : : :
D4      -----CAGATGC-----CATTA-----
10
      1630      1640      1650      1660      1670      1680
./wwwt CAGCAGTGTTGCTCAGCTCCTACCTCTGTGCCAGGGCAGCATTTTCATATCCAAGATCAA
      : : : : :
D4      -----TTG--CAGC-----GGGTA-----
15
      1690      1700      1710      1720      1730      1740
./wwwt TTCCCTCTCTCAGCACAGCCTGGGGAGGGGGTCAATTGTTCTCCTCGTCCATCAGGGATT
      : : : : :
D4      -----CTGGGG-----
20
      1750      1760      1770      1780      1790      1800
./wwwt CAGAGGCTCAGAGACTGCAAGCTGCTTGCCCAAGTCACACAGCTAGTGAAGACCAGAGCA
      : : : : :
D4      CAG-----CACA-----
30
      1810      1820      1830      1840      1850      1860
./wwwt GTTTCATCTGGTTGTGACTCTAAGCTCAGTGCTCTCTCCACTACCCACACACCAGCCTTGG
      : : : : :
D4      -----TGCCATCTC-----GG
35
      1870      1880      1890      1900      1910      1920
./wwwt TGCCACCAAAAGTGCTCCCCAAAAGGAAGGAGAAATGGGATTTTCTTTTGAGGCATGCAC
      : : : : :
D4      TGGCAGC-----GCT-----TCTTTT-----
40
      1930      1940      1950      1960      1970      1980
./wwwt ATCTGGAATTAAGGTCAAACCTAATTCTCACATCCCTCTAAAAGTAAACTACTGTTAGGAA
      : : : : :
D4      -----TTCTC-----CGAC-----A
45
      1990      2000      2010      2020      2030      2040
./wwwt CAGCAGTGTTCTCACAGTGTTGGGGCAGCCGTCCTTCTAATGAAGACAATGATATTGACAC
      : : : : :
D4      CACCA-TG-----CAG-----GCCG-----ATGAT-----
50
      2050      2060      2070      2080      2090      2100
      360
      370
60
65

```

ES 2 285 768 T3

```
./wwwt TGTCCCTCTTTGGCAGTTGCATTAGTAACTTTGAAAGGTATATGACTGAGCGTAGCATAC
      :  :  :  :
D4      -----CCTTGG-----
```

```
      2110      2120      2130      2140      2150      2160
./wwwt AGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAATGAAATTT
```

```
D4      -----
```

```
      2170      2180      2190      2200      2210      2220
./wwwt GCAAAATCACTTAGCAGCAACTGAAGACAATTATCAACCACGTGGAGAAAATCAAACCGA
```

```
D4      -----TGGGGA-----
                        380
```

```
      2230      2240      2250      2260      2270      2280
./wwwt GCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCACT
```

```
D4      -----CTGTG-----GCAATAC-----CT
                        390
```

```
      2290      2300      2310      2320      2330      2340
./wwwt CCACAAATGATGTTTTCAGGTGTTCATGGACTGTTGCCACCATGTATTCATCCAGAGTTCT
```

```
D4      CC-CAACNGCTGT-----GCC-----
      400                        410
```

```
      2350      2360      2370      2380      2390      2400
./wwwt TAAAGTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTTAAATTATGTA
```

```
D4      -----
```

```
      2410      2420      2430      2440      2450      2460
./wwwt TAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCTAAAAAAAAAAAAA
```

```
D4      -----GAANT-----
```

```
      2470
./wwwt AAAAAAAAAAAAAAAAAAAAAA
```

```
D4      -----
```

ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 1 2479 nt frente a D8 rev 355 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 Identidad del 12,7%; puntuación de alineación global: -7722

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GGCACGAGGGGGCGGCGGCTGCGGGCGCAGAGCGGAGATGCAGCGGCTTGGGGCCACCCT
5 D8      -----
                                     : : : :
      70      80      90      100     110     120
10 ./wwwt GCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCCCCCGCGCCCGCTCCGACGGCGAC
D8      -----
      130     140     150     160     170     180
15 ./wwwt CTCGGCTCCAGTCAAGCCCCGGCCCGGCTCTCAGCTACCCGCGAGGAGGAGGCCACCCTCAA
      : : : :
20 D8      -----AGTC-----CAGGAAAA--CCAACC-----
                                     10      20
      190     200     210     220     230     240
25 ./wwwt TGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGACACGCAGCACAAATTGCGCAGCGC
      : : : :
D8      -----CAAA-----
      250     260     270     280     290     300
30 ./wwwt GGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAAGCATCATCAGAAGTGAACCTGGC
      : : : : : : : : : : : :
D8      -----AGGAAGCC-----AAGTAT--TAAGAA-----
                                     30      40
      310     320     330     340     350     360
35 ./wwwt AAACCTTACCTCCCAGCTATCACAATGAGACCAACACAGACACGAACGTTGGAATAATAC
      : : : : : : : : : : : :
40 D8      -----ATCACAAGG-----CAGGAAGGG-----
                                     50      60
      370     380     390     400     410     420
45 ./wwwt CATCCATGTGCACCGAGAAATTCACAAGATACCAACAACCAGACTGGACAAATGGTCTT
      : : : : :
D8      -----ACAAGA-----
                                     70
      430     440     450     460     470     480
50 ./wwwt TTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAAGGCAGAAGGAGCCACGAGTGCAT
      : : : : : : : :
55 D8      -----GGGAGA-----AAG-----
      60
65

```

ES 2 285 768 T3

80

```

5      490      500      510      520      530      540
./wwwt CATCGACGAGGACTGTGGGGCCAGCATGTACTGCCAGTTTGCCAGCTTCCAGTACACCTG
      : : :
D8      -----TTG-----

10     550      560      570      580      590      600
./wwwt CCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGGGACAGTGAAGTGCTGTGGAGACCA
      : : : : : : :
D8      -----TCTG-----AGA-----
15                                           90

      610      620      630      640      650      660
./wwwt GCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACCAGGGGCAGCAATGGGACCATCTG
      : : : : : : :
D8      -----ACT-----TT

20     670      680      690      700      710      720
./wwwt TGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGTGCCTTCCAGAGAGGCCCTGCTGTT
      : : : : : : :
D8      TGAC-----TGTG-----
25                                           100

      730      740      750      760      770      780
./wwwt CCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTTTGCCATGACCCGCCAGCCGGCT
      : : : : : : :
D8      -----GCCC-TGGA-----CTTTGC-----
35                                           110

      790      800      810      820      830      840
./wwwt TCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGAGCCTTGGACCGATGCCCTTGTGC
      : : : : : : :
D8      -----TGTGC-----
40                                           120

      850      860      870      880      890      900
./wwwt CAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTGGTGTATGTGTGCAAGCCGACCTT
      : : : : : : :
D8      -----T

50     910      920      930      940      950      960
./wwwt CGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTGCCAGAGAGGTCCCCGATGAGTA
      : : : : : : :
D8      CGT-----

55     970      980      990      1000      1010      1020
./wwwt TGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAGCTGGAGGACCTGGAGAGGAGCCT
      : : : : : : :
60
65

```

ES 2 285 768 T3

```

D8 -----CAT-----

5      1030      1040      1050      1060      1070      1080
./wwwt GACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCCGCCGCTGCACTGCTGGGAAGGGA

D8 -----

10     1090      1100      1110      1120      1130      1140
./wwwt AGAGATTTAGATCTGGACCAGGCTGTGGGTAGATGTGCAATAGAAATAGCTAATTTATTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15     D8 -----TTT-----TGGAC-----GAA-----AATTT-----
      130                                140

      1150      1160      1170      1180      1190      1200
./wwwt CCCANGTGTGTGCTTTAAGCGTGGGCTGACCAGGCTTCTTCCTACATCTTCTTCCCAGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 -----GT

25     1210      1220      1230      1240      1250      1260
./wwwt AAGTTTCCCCTCTGGCTTGACAGCATGAGGTGTTGTGCATTTGTTTCAGCTCCCCCAGGCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8     AAG-----CCAGTC-
30     150

      1270      1280      1290      1300      1310      1320
./wwwt GTTCTCCAGGCTTCACAGTCTGGTGCTTGGGAGAGTCAGGCAGGGTTAAACTGCAGGAGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35     D8 -----CTT-----TTGG-----AGG-----
      160

      1330      1340      1350      1360      1370      1380
./wwwt AGTTTGCCACCCCTGTCCAGATTATTGGCTGCTTTGCCTCTACCAGTTGGCAGACAGCCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8 -----GACAG--G
40     170

      1390      1400      1410      1420      1430      1440
./wwwt TTTGTTCTACATGGCTTTGATAATTGTTTGAGGGGAGGAGATGGAAACAATGTGGAGTCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50     D8     TCTG-----CT

      1450      1460      1470      1480      1490      1500
./wwwt CCCTCTGATTGGTTTTGGGGAAATGTGGAGAAGAGTGCCCTGCTTTGCAAACATCAACCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
55     D8     CC-----AGAAGAG-----
      180

      1510      1520      1530      1540      1550      1560
./wwwt GGCAAAAATGCAACAAATGAATTTTCCACGCAGTTCTTTCCATGGGCATAGGTAAGCTGT

```

ES 2 285 768 T3

```

D8      -----GGCATA---AAG---
                                190

5      1570      1580      1590      1600      1610      1620
./wwwt GCCTTCAGCTGTTGCAGATGAAATGTTCTGTTACCCCTGCATTACATGTGTTTATTCATC
                                :: ::::
D8      -----ACACTGC-----TCA--
                                200

10     1630      1640      1650      1660      1670      1680
./wwwt CAGCAGTGTGCTCAGCTCCTACCTCTGTGCCAGGGCAGCATTTCATATCCAAGATCAA
                                :::::
D8      -----AGCTCC-----AGA--AA
                                210

15     1690      1700      1710      1720      1730      1740
./wwwt TTCCCTCTCTCAGCACAGCCTGGGGAGGGGGTCATTGTTCTCCTCGTCCATCAGGGATTT
:      :: ::::
D8      T-----CTTCCAGC-----
                                220

20     1750      1760      1770      1780      1790      1800
./wwwt CAGAGGCTCAGAGACTGCAAGCTGCTTGCCCAAGTCACACAGCTAGTGAAGACCAGAGCA
D8      -----

25     1810      1820      1830      1840      1850      1860
./wwwt GTTTCATCTGGTTGTGACTCTAAGCTCAGTGCTCTCTCCACTACCCACACCCAGCCTTGG
                                :::: ::::
D8      -----GTTGCGACT-----GTG-----GCCCTGG
                                230      240

30     1870      1880      1890      1900      1910      1920
./wwwt TGCCACCAAAGTGCTCCCCAAAAGGAAGGAGAATGGGATTTTCTTTTGAGGCATGCAC
D8      -----

35     1930      1940      1950      1960      1970      1980
./wwwt ATCTGGAATTAAGGTCAAACATAATTCTCACATCCCTCTAAAAGTAACTACTGTTAGGAA
                                :::::
D8      -----ACTACTGT-----
                                250

40     1990      2000      2010      2020      2030      2040
./wwwt CAGCAGTGTCTCACAGTGTGGGGCAGCCGTCCTTCTAATGAAGACAATGATATTGACAC
                                :::: ::::
D8      -----GTC-----GAAGCCA-----ATTGAC--
                                260

45     2050      2060      2070      2080      2090      2100
./wwwt
D8

```

ES 2 285 768 T3

```

5  ./wwwt TGTCCCTCTTTGGCAGTTGCATTAGTAACCTTTGAAAAGGTATATGACTGAGCGTAGCATAC
D8  -----

10  2110      2120      2130      2140      2150      2160
./wwwt AGGTTAACCTGCAGAAACAGTACTTAGGTAATTGTAGGGCGAGGATTATAAATGAAATTT
D8  -----

15  2170      2180      2190      2200      2210      2220
./wwwt GCAAAATCACTTAGCAGCAACTGAAGACAAATTATCAACCACGTGGAGAAAATCAAACCGA
      : : : : : :
D8  -----CAGCAATCG-----
      270

20  2230      2240      2250      2260      2270      2280
./wwwt GCAGGGCTGTGTGAAACATGGTTGTAATATGCGACTGCGAACACTGAACTCTACGCCACT'
      : : : : : : : : : : : : : : : :
D8  GCAG-----CATGCTCG-ATTAAGAG-----TATGCCA--
      280      290      300

25  2290      2300      2310      2320      2330      2340
./wwwt CCACAAATGATGTTTTTCAGGTGTCATGGACTGTTGCCACCATGTATTCATCCAGAGTTCT
      : : : : : :
D8  --AAAAAT-----AG-----
      310

30  2350      2360      2370      2380      2390      2400
./wwwt TAAAGTTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTTTAAATTATGTA
      : : : : : :
D8  -----AAAAAGC-----TA
      320

35  2410      2420      2430      2440      2450      2460
./wwwt TAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACTTCAACTGCTAAAAAAAAAAAAA
      : : : : : : : : : : : : : : : :
D8  TAAATAT-----TTCAAAATAAAGAAGAATCCACAT-----TGC-----
      330      340      350

40  2470
./wwwt AAAAAAAAAAAAAAAAAAAAAA
D8  -----

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 3 1050 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 30,3%; puntuación de alineación global: -1869

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGGCGGCGGTCCCCACGGCC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 D1    ---CAG-----TGTTCTCACAGTG-----TGG-GGCAGCCGTCCT-----
      10      20      30

      70      80      90      100     110     120
./wwwt CCGCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 D1    -----TCTAATGAAGACAATGA-TATTGACA---CTGTCCC---TCTTTGGCAG
      40      50      60      70

      130     140     150     160     170     180
./wwwt CCGCAGGAGGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAAC TGATCGAGGAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 D1    TTGCATTAGTA---ACTTTGAA-----AGGTATATGA-CTGA-----
      80      90      100

      190     200     210     220     230     240
./wwwt ACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
25 D1    -----GCGTAGC-----ATACAGGTTAACCTGCAG--AAA
      110     120     130

      250     260     270     280     290     300
./wwwt GCATCATCAGAAGTGAACCTGGCAAACCTACCTCCCAGCTATCACAA TGAGACCAACACA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 D1    -----CAGTACTTA---GGTAATTGTA-----GGG-----
      140     150

      310     320     330     340     350     360
./wwwt GACACGAACGTTGGAAATAATACCATCCATGTGCACCGAGAAAATCACAAGATAACCAAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 D1    ---CGAGGATTATAAAT-----GAAATTGCAAAATCACTTAG
      160     170      180     190

      370     380     390     400     410     420
./wwwt AACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
45 D1    CAGCA-ACTGAA-----GNCAATTATCAACCACGTGG-AGA-----AA
      200     210     220

      430     440     450     460     470     480
./wwwt GGCAGTAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGGCCAGCATGTACTGCCAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 D1    ATCAAACCGAGC-----AGGCTGTGTGA--AACATG-----G

```

ES 2 285 768 T3

```

230                240                250

          490          500          510          520          530          540
5  ./wwwt TTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGG
   ::                : : : : :                : : : : :
D1  TT-----GTA-ATATGCGA-----CTGC-----G
   260                270

          550          560          570          580          590          600
10 ./wwwt GACAGTGAGTGCTGTGGAGACCAGCTGTGTCTGGGGTCACTGCACCAAAATGGCCACC
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  AACACTGAACTCT-----AC-----GCCACTCCACAAA-----
   280                290                300

          610          620          630          640          650          660
15 ./wwwt AGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGT
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  -----TGATGT
   310

          670          680          690          700          710          720
25 ./wwwt GCCTTCCAGAGAGGCCTGTGTTCCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTT
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  ---TTTCAG-----GTGT-----CATGGACTG-----T
   320                330

          730          740          750          760          770          780
30 ./wwwt TGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  TGCCA-----CCA-----TGTAT-TCATCC-----AGAG-----
   340                350

          790          800          810          820          830          840
35 ./wwwt GCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  -----TTCTTAAAGT-----TTAAAGTTGCACA-----
   360                370

          850          860          870          880          890          900
45 ./wwwt GTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATUGGGAGATCCTGCTG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  -TGATTGTAT--AAGCATGCTTTCTT-----
   380                390

          910          920          930          940          950          960
50 ./wwwt CCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCCCCAGGAG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  -----TGAGTTTAAATT-----ATGTATAAA-----CATAAG
   400                410                420

          970          980          990          1000          1010
55 ./wwwt CTGGAGGACCTGGAGAGGAGCCTGACTGAAG-AGATGGCGCTGAGGGAGCCTGCGGCTGC
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  TTG-----CATTTAGA-----AATCAAGCATAAATCACT-----
   430                440                450

          1020          1030          1040          1050
65 ./wwwt CGCCGCTGCACTGCTGGGAAGGGAAGAGATT
   : : : : :
D1  -----TCAACTGCT-----
   460

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 3 1050 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 26,1%; puntuación de alineación global: -2234

```

10      10      20      30      40      50      60
./wwwt ATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCC
      :          : : : : : : : : : : : :
D2      A-----ACTCCCCTGTGATTGCAGTAAATTACTGTATT-----
      10      20      30

15      70      80      90      100     110     120
./wwwt CCCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
      : : : : : : : : : : : : : : : :
D2      -----GTAAATTCT-----CAGTGT-----GGCACT-----TAC
      40      50

20      130     140     150     160     170     180
./wwwt CCGCAGGAGGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAACTGATGGAGGAC
      : : : : : : : : : : : : : : : :
D2      CTGTAAATG-----CAATGA-----AACTTTT-----
      60      70

25      190     200     210     220     230     240
./wwwt ACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAA
      : : : : : : : : : : : : : : : :
D2      -----AATTA-----TTTTTCTAAA
      80      90

30      250     260     270     280     290     300
./wwwt GCATCATCAGAAGTGAACCTGGCAAACCTTACCTCCCAGCTATCACAAATGAGACCAACACA
      : : : : : : : : : : : : : : : :
D2      GGTGCTGCA-----CTGCCTATTTTCTCTT-GTTAT-----
      100     110     120

35      310     320     330     340     350     360
./wwwt GACACGAACGTTGGAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAAC
      : : : : : : : : : : : : : : : :
D2      -----GTAAATTT-----TGTACAC-----ATT-----GAT-----
      130     140

40      370     380     390     400     410     420
./wwwt AACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAA
      : : : : : : : : : : : : : : : :
D2      -----TGTTATCTTGACTGACAAATATTCTATATT-----GAA---
      150     160     170     180

45      430     440     450     460     470     480
./wwwt GGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCAGCATGTACTGCCAG
      : : : : : : : : : : : : : : : :
D2      --CTGAAGTAA-----ATCATTT-----CAGCTTATAGTTC--

```

ES 2 285 768 T3

```

190                                200                                210

      490      500      510      520      530      540
./wwwt TTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGG
5      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      -TTAAAAGCAT-----AACC-----CTTTACCC--
      220                                230

      550      560      570      580      590      600
./wwwt GACAGTGAGTGCTGTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAATGGCCACC
10      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      --CATTTAATTCT---AGA-----GTCTAG-----AA-----CGCA
      240                                250

      610      620      630      640      650      660
./wwwt AGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGT
15      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      AGG---ATCTCTTGGA-----TGACAAA-----TGATAG-----GT
      260      270                                280

      670      680      690      700      710      720
./wwwt GCCTTCCAGAGAGGCCTGCTGTTCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTT
20      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      ACCTAAAA-----TGTAACATG-----AAAATACT-----AGCTT
      290                                300      310

      730      740      750      760      770      780
./wwwt TGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGA
25      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      ---ATT-----TTCTGAA---ATCTACT-----ATCTTAATG--
      320                                330      340

      790      800      810      820      830      840
./wwwt GCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTG
30      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      --CTTAAAT-----
      350

      850      860      870      880      890      900
./wwwt GTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTG
35      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      ---TATATTTC-----CTTTAG-----GCTGTGA--TAGTTTTTGAATAA-----
      360      370      380

      910      920      930      940      950      960
./wwwt CCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAG
40      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      -----AATTTAACAT-TTAAT-----
      390      400

      970      980      990      1000      1010      1020
./wwwt CTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCC
45      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      -----AT-----

      1030      1040      1050
./wwwt GCCGCTGCACTGCTGGGAAGGGAAGAGATT
50      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D2      -----CA-----AA-----AAAAAAA
      410

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 3 1050 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 31,2%; puntuación de alineación global: -1835

```

10      10      20      30      40      50      60
./wwwt ATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGCGGGTCCCCACGGCC
      :: ::      :: ::      :: ::      ::
D3      -----TTNGGA-----TGGAGT--TGATGAAGG-----
15      10      20

      70      80      90      100     110     120
./wwwt CCCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
      ::::      ::      : : ::      ::
D3      -----TCCATAAAAG-----GTTTTC--TA-
      30      40

      130     140     150     160     170     180
./wwwt CCGCAGGAGGAGGCCACCCTCAATGAGATGTTCCGCGAGGTTGAGGAAGTGTGATGGAGGAC
      ::      :: :: : ::      ::      : ::::
D3      --GCA-----TCTATTATAT-----AA--AAGGAG--
      50      60

      190     200     210     220     230     240
./wwwt ACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAA
      :: : : ::      : ::::      ::::: : ::
D3      -----TTCCACTGC-----ATTTGGAT-----AAGCTGGTT--TAGT
      70      80      90

      250     260     270     280     290     300
./wwwt GCATCATCAGAAAGTGAACCTGGCAAACCTTACCTCCCAGCTATCACAATGAGACCAACACA
      : : : : ::::      : : :      :: : : : : :
D3      GTCTC-TGACAAAGTG---TG--AAGC-----CTAGAAGAATTA-----CT
      100     110     120

      310     320     330     340     350     360
./wwwt GACACGAACGTTGGAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAAC
      : : : : :      : : : : :      : : : : :      : : : : :
D3      GGCTTGATGGTG-----ATCTTCTGTATCCGG-----CAAGA-----
      130     140     150

      370     380     390     400     410     420
./wwwt AACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAA
      :::::      ::: :      ::::: :      : : :      : : :
D3      ---CAGAC-----CTTCTCC---ACAGTAA---CAACGCTGG-----AATAT
      160     170     180     190

      430     440     450     460     470     480
./wwwt GGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCCCAGCATGTACTGCCAG
      : : : : : : : : : : : : : : : : : : : : :
D3      TTCT--AG--TCCATGAGAGCCTTTTCTCCTATGCTTG-----GTAC--ACAC

```

ES 2 285 768 T3

```

                200      210      220
                490      500      510      520      530      540
5  ./wwwt TTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCGG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  TTGACCTTCTTTTCAGGACA-----GGTT-----
   240      250

                550      560      570      580      590      600
10 ./wwwt GACAGTGAGTGCTGTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAAATGGCCACC
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  TACAGATCTTG-----GACCAGAAGTGT--CTAG---CAC---AACACAAT-----CC
   260      270      280      290

                610      620      630      640      650      660
15 ./wwwt AGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGACTGCCAGCCGGGGCTGTGCTGT
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  TGAGGCA-CAGTCTGATGACCGGAGACAAACAGA---AC-----CT-TCTTGT
   300      310      320      330

                670      680      690      700      710      720
20 ./wwwt GCCTTCCAGAGAGGCCTGTGTTCCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAGCTT
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  CCTTT-----GG--TG-TGATACATTTTGAAGA-----CAAGGTGG-----T
   340      350      360      370

                730      740      750      760      770      780
25 ./wwwt TGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  T-----CTTCTGGAATACCCATCCAAGGTGCTA-----TGAT---
   380      390      400

                790      800      810      820      830      840
30 ./wwwt GCCTTGGACCGATGCCCTTGTGCCAGTGGCCCTCCTCTGCCAGCCCCACAGCCACAGCCTG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -CAATT--ACCAAAGC-----TTTCAGTGA-----TC
   410      420

                850      860      870      880      890      900
35 ./wwwt GTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCTG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  GTTT-----CCT-----CAATTT-----CTC
   430      440

                910      920      930      940      950      960
40 ./wwwt CCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGAG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  CTCGGAAA-----TGATTTTGA-----TCA-----
   450      460

                970      980      990      1000      1010      1020
45 ./wwwt CTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCTGCC
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -----GAAG-----

                1030      1040      1050
50 ./wwwt GCCGCTGCACTGCTGGGAAGGGAAGAGATT
D3  -----

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 3 1050 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 28,9%; puntuación de alineación global: -2000

```

10          10          20          30          40          50          60
./wwwt ATGCAGCGGCTTGGGGCCACCCCTGCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCC
          : : : : :          : : : : :
D4      -----CTTGCGT-----TGTT--TGG-----
          10

15          70          80          90          100          110          120
./wwwt CCCGCGCCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
          : : : : :
D4      -----TACAGAC-----
          20

20          130          140          150          160          170          180
./wwwt CCGCAGGAGGAGGCCACCCCTCAATGAGATGTTCCGCGAGGTTGAGGAACCTGATCGAGGAC
          : : : : :          : : : : : : :
D4      -----TTCCCC-----CTGATGGAGCAC
          30

25          190          200          210          220          230          240
./wwwt ACGCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAA
          :          : : : :          : : : : : : :
D4      TGGT-----TTGC-----AGATTTTGGTCCAGAA-----
          40          50          60

30          250          260          270          280          290          300
./wwwt GCATCATCAGAAAGTGAACCTGGCAAACCTTACCTCCCAGCTATCACAATGAGACCAACACA
          : : : : : : : : : : : : : : : : :
D4      -----AT--GACGAGCAC--AGCAAA--ACC-----CTT--CAATG-----CA
          70          80          90

35          310          320          330          340          350
./wwwt GAC--ACGAACGTTGGAAATAATACCATCCATGTG--CACCGAGAAATTCACAAGATAACC
          : : : : : : : : : : : : : : : : :
D4      GTCTGATGATCGTAGG-----CAGGGGTCTCC-----TTCA-----
          100          110          120

40          360          370          380          390          400          410
./wwwt AACAAACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAA
          : : : : : : : : : : : : : : : : :
D4      -----TGCCCTTTT-----ATATGTGACATCTT-----
          130          140

45          420          430          440          450          460          470
./wwwt GAAGGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCCCAGCATGTACTGC
          : : : : :          : : : : :
D4      -----AGTG-----TGTGGTC-----

```

ES 2 285 768 T3

150

```

5      480      490      500      510      520      530
./wwwt CAGTTTGGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACC
      :: : : : : : : : : : : : : : : : : : : : : : :
D4      ---TTCCTAGATTTC-----TGCCATCC-----CAAGTCATGGTTTG----
      160      170      180      190

10     540      550      560      570      580      590
./wwwt CGGGACAGTGAGTGCTGTGGAGACCAGCTGTGTGTCTGGGGTCACTGCACCAAATGGCC
      : : : : : : : : : : : : : : : : : : : : : :
D4      -----AGTAATGACCGTGGTTTCGATCTCTGTGCC--GAGT-----ACCAT-----
      200      210      220      230

15     600      610      620      630      640      650
./wwwt ACCAGGGGCGAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGC
      : : : : : : : : : : : : : : : : : : : : : :
D4      -CCAGAGCC-----GGGA-----TGTGA-----GGGGTTAAGA
      240      250

20     660      670      680      690      700      710
./wwwt TGTGCCTTCCAGAGAGCGCTGCTGTTCCCTGTGTGCACACCCCTGCCCGTGGAGGGCGAG
      : : : : : : : : : : : : : : : : : : : : : :
D4      TG---CTTTCAGTAA-----CTG-----GGATACAGA-
      260      270      280

25     720      730      740      750      760      770
./wwwt CTTTGGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGAT
      : : : : : : : : : : : : : : : : : : : : : :
D4      ---TGCCATTATTGC---AGC-GGGTACTGGGGC--AGCACATGCCATCTCG-----
      290      300      310      320

30     780      790      800      810      820      830
./wwwt GGAGCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGC
      : : : : : : : : : : : : : : : : : : : : : :
D4      -----GTGGCAGCG-CTTCTTTTCTCCGACACA-CCA-----
      330      340      350

35     840      850      860      870      880      890
./wwwt CTGGTGTATGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGGGAGATCCTG
      : : : : : : : : : : : : : : : : : : : : : :
D4      -----TGCAGGCCGATGATCCTTGG-----TGGGGA-----CTG
      360      370      380

40     900      910      920      930      940      950
./wwwt CTGCCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCCCCAG
      : : : : : : : : : : : : : : : : : : : : : :
D4      -----TGGCA-----AT-----
      390

45     960      970      980      990      1000      1010
./wwwt GAGCTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGAGGGAGCCTGCGGCT
      : : : : : : : : : : : : : : : : : : : : : :
D4      -----ACCT-----CCCAACNG-----CTG-----T
      400

50     1020      1030      1040      1050
./wwwt GCCGCCGCTGCACTGCTGGGAAGGGAAGAGATT
      : : : : : : : : : : : : : : : : : : : : : :
D4      GCCGA-----ANT
      410

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 3 1050 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 25,2%; puntuación de alineación global: -2269

```

10      10      20      30      40      50      60
./wwwt ATGCAGCGGCTTGGGGCCACCCTGCTGTGCCTGCTGCTGGCGGCGGCGGTCCCCACGGCC
      : : : : :
D8      -----CACCCAG-----

15      70      80      90      100     110     120
./wwwt CCGCGCGCCGCTCCGACGGCGACCTCGGCTCCAGTCAAGCCCGGCCCGGCTCTCAGCTAC
      : : : : : : : : :
20      D8      -----TCCAGGAAA-----ACCAAC
      10      20

25      130     140     150     160     170     180
./wwwt CCGCAGGAGGAGGCCACCCTCAATGAGATGTTCGCGGAGGTTGAGGAAC TGATGGAGGAC
      : : : : : : : : :
D8      CC--AAAAGGAAGCCA-----AGTATTAAGAAAT-----
      30      40

30      190     200     210     220     230     240
./wwwt ACCCAGCACAAATTGCGCAGCGCGGTGGAAGAGATGGAGGCAGAAGAAGCTGCTGCTAAA
      : : : : : : : : :
D8      -----CACAA-----GGCAG-----
      50

35      250     260     270     280     290     300
./wwwt GCATCATCAGAAGTGAACCTGGCAAACCTACCTCCCAGCTATCACAATGAGACCAACACA
      : : : : : : : : :
40      D8      -----GAAGGGA-----CAAGAGG
      60      70

45      310     320     330     340     350     360
./wwwt GACACGAACGTTGGAAATAATACCATCCATGTGCACCGAGAAATTCACAAGATAACCAAC
      : : : : : : : : :
D8      GA--GAAAGTTGT-----CTGAGAA-----
      80      90

50      370     380     390     400     410     420
./wwwt AACCAGACTGGACAAATGGTCTTTTCAGAGACAGTTATCACATCTGTGGGAGACGAAGAA
      : : : : :
55      D8      -----CTTTT-----

60      430     440     450     460     470     480
./wwwt GGCAGAAGGAGCCACGAGTGCATCATCGACGAGGACTGTGGGCCCAGCATGTACTGCCAG
      : : : : : : : : :
D8      -----GACTGTGCCCC-----TGGACT-----

```

ES 2 285 768 T3

```

                    100          110
                490      500      510      520      530      540
5  ./wwwt TTTGCCAGCTTCCAGTACACCTGCCAGCCATGCCGGGGCCAGAGGATGCTCTGCACCCCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -TTGC-----TG-----TGCTCGTCA-----
                                120

                550      560      570      580      590      600
10 ./wwwt GACAGTGAGTGCTGTGGAGACCAGCTGTGTCTGGGGTCACTGCACCAAAATGGCCACC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -----TTTTTGG-----ACGAAAATTGTGTA--
                                130          140

                610      620      630      640      650      660
15 ./wwwt AGGGGCAGCAATGGGACCATCTGTGACAACCAGAGGGACTGCCAGCCGGGGCTGTGCTGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -----AGCCA---GTCCTTTTG-----GAGGGAC-----
                                150          160

                670      680      690      700      710      720
20 ./wwwt GCCTTCCAGAGAGCCCTGCTGTTCCTGTGTGCACACCCTGCCCGTGGAGGGCGAGCTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -----AGGTCCTGCT-----CCAGAAGAGGG-----
                                170          180

                730      740      750      760      770      780
25 ./wwwt TGCCATGACCCCGCCAGCCGGCTTCTGGACCTCATCACCTGGGAGCTAGAGCCTGATGGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  ---CATAA-----AGACA---CTG---CTCA-----AGCT---CCAGA---A
                                190          200          210

                790      800      810      820      830      840
30 ./wwwt GCCTTGGACCGATGCCCTTGTGCCAGTGGCCTCCTCTGCCAGCCCCACAGCCACAGCCTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  ATCTTCCAGCGTTGC-----GACTGTGGCC-----CTG
                                220          230          240

                850      860      870      880      890
35 ./wwwt GTGTA-TGTGTGCAAGCCGACCTTCGTGGGGAGCCGTGACCAAGATGGGGAGATCCTGCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  GACTACTGTGTGCAAGCCAA--TT-----GACC---AGCA--ATCGGCAG--CATGCT
                                250          260          270          280

                900      910      920      930      940      950
40 ./wwwt GCCCAGAGAGGTCCCCGATGAGTATGAAGTTGGCAGCTTCATGGAGGAGGTGCGCCAGGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -----CGAT---TAAGA-----GTATGCCAAAA
                                290          300

                960      970      980      990      1000      1010
45 ./wwwt GCTGGAGGACCTGGAGAGGAGCCTGACTGAAGAGATGGCGCTGAGGGAGCCTGCCGCTGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  AATAGAAAAGCTATAAATATTTCAAAATAAAGAA-----GAATCCA-----
                                310          320          330          340

                1020      1030      1040      1050
50 ./wwwt CGCCGCTGCACTGCTGGGAAGGGAAGAGATT
      : : : : :
D8  -----CATTGC-----
                                350

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 4 848 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 36,2%; puntuación de alineación global: -1075

```

10      10      20      30      40      50      60
./wwwt GAATTCCGGCACGAGAGACGACGTGCTGAGCTGCCAGCTTAGTGGAAGCTCTGCTCTGGGT
      ; :      :: : ;      : ::      : ::
D1      CAGT-----GTTCTCA-----CAGTG-----TGGG-
15                      10                      20

      70      80      90      100     110
./wwwt GGAGAGCAGCCTCGCTT-TGGTGACGCACAGTGCTGGGACCCTCCAGGAGCCCCGGGATT
      : : : : : : : : : : : : : : : : : :
D1      -----GCAGCCGTCCTTCTAATGAAG-ACAATGAT-----ATT-
20                      30      40                      50

      120     130     140     150     160     170
./wwwt GAAGGATGGTGGCGGCCGTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCCCTGGGAGCTC
      : :      : : : : : : : : : : : : : : :
D1      GA-----CACTGTCCCTCTTTGGC--AGTTG-----
25                      60      70

      180     190     200     210     220     230
./wwwt TGGTCCTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCT
      : : : : : : : : : : : : : : : : : :
D1      -----CATTAGTAACCTTTG-----AAAGGTA
30                      80      90

      240     250     260     270     280     290
./wwwt CACAGTGCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATG
      :      : : : : : : : : : : : : : : :
D1      TA-----TGA CTGA-----GCG-TAGCATA CAGGTTAACCTGCAG-----
40                      100     110     120     130

      300     310     320     330     340     350
./wwwt AGAAGCCGTTCTGTGCTACATGTCGTGGCTTGC GGAGGAGGTGCCAGCGAGATGCCATGT
      : : : : : : : : : : : : : : : : : :
D1      --AAACAGTACTTAGGTAATTGT-----AGG-----GCGAG-----
45                      140     150                      160

      360     370     380     390     400     410
./wwwt GCTGCCCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAA
      : : : : : : : : : : : : : : : : : :
D1      -----GATTATAA-----ATGAAATTTGCAAAATCAC
50                      170     180

      420     430     440     450     460     470
./wwwt TATTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAAC TGGGCACC
      : : : : : : : : : : : : : : : : : :
D1      T--TAGCA--GCAACT-----GAAGNCA-----ATTATCAAC----CAC-
60

65

```

ES 2 285 768 T3

```

190                                200                                210

      480      490      500      510      520      530
5  ./wwwt CAGTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGG
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D1  --GTGGAGAAAATCAAACCGAGCAGG--GCT--GTGTGAA-----ACATGG-----
      220      230      240      250

10      540      550      560      570      580      590
./wwwt GACAAGAGGGAGAAAGTTGTCTGAGAACTTTTGACTGTGGCC-CTGGACTTTGCTGTGCT
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D1  -----TTGT-----AATATGCGACTGCGAACACTGAACTCTACGCCACT
      260      270      280      290

15      600      610      620      630      640      650
./wwwt CGTCATTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGA
      :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :
20 D1  CCACA-----AATGATGTTTT-----CAGGTGTCATGGA-----CTGTT-----
      300      310      320      330

25      660      670      680      690      700      710
./wwwt AGAGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCT
      :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :  ::  :
D1  ----GCCACCATG----TATTCA---TCCAGAGTTCTTAAAGTTT---AAAGTTGCACA
      340      350      360      370

30      720      730      740      750      760      770
./wwwt GGACTACTGTGTCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAAGAGTATGC
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D1  TGAT-----TGTATAAGC--AT-----GCTTTC-----TTTGAGTTTTAAATTATGT
35      380      390      400      410

      780      790      800      810      820      830
40 ./wwwt CAAAAAATAGAAAAGCTATAAATATTTCAAAATAAAGAAGAATCCACATTGCAAAAAAAA
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
D1  ATAAACATA----AGTTGC----ATTTAGAAATCAAGCATAAATCACTTC-----
      420      430      440      450

      840
45 ./wwwt AAAAAAAAAA
      ::
D1  ---AACTGCT
      460

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 4 848 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 33,2%; puntuación de alineación global: -1352

ES 2 285 768 T3

```

      10      20      30      40      50
./wwwt GAATTCGGCACGAGAGACGACGTGCTG-AGCTGCCAGCTTAGTGGAAGCTCTGCTCTGGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 D2    AACTCC----CCTGTGAT----TGCAGTAAATTACTG--TATTGTAAATT---CTCAGTG
      10      20      30      40

      60      70      80      90      100     110
./wwwt TGGAGAGCAGCCTCGCTTTGGTGACGCACAGTGCTGGGACCCTCCAGGAGCCCCGGGATT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 D2    TGG-----CACCTACCTGTA---AATGCAATGAAACT-----TT
      50      60      70

      120     130     140     150     160     170
./wwwt GAAGGATGGTGGCGGCCGCTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCTGGGAGCTC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 D2    TAATTATTTT-----CTAAAGGTGCTGCACTGCCTATTTT-TCC-----TC
      80      90      100     110

      180     190     200     210     220     230
./wwwt TGGTCCTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
25 D2    TTGTTATGTA-----AATTTTGTGA-----
      120     130

      240     250     260     270     280     290
./wwwt CACAGTGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 D2    CACATTGATTGT-----TA-----TCT-----TG
      140     150

      300     310     320     330     340     350
./wwwt AGAAGCCGTTCTGTGCTACATGTCGTGGGTTGCGGAGGAGGTGCCAGCGAGATGCCATGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 D2    A-----CTG---ACAAAT-----ATTCTATAT
      160     170

      360     370     380     390     400     410
./wwwt GCTGCCCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
45 D2    -----TGAAC-----TGAAGTA-----AATC-----
      180     190

      420     430     440     450     460     470
./wwwt TATTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 D2    -ATTT-----CAGCTT-AT-----AGTT--CTTAAAAGCATA-----ACC

```

ES 2 285 768 T3

```

                200                210                220
      480      490      500      510      520      530
5  ./wwwt CAGTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAAGGCAGGAAGG
   :  :                : : : :                : : : : : : : : : :
D2  CTTT-----ACCCCAT-----TTAA--TTCTAGAGTCTAGAACG
                230                240                250

      540      550      560      570      580      590
10 ./wwwt GACAAGAGGGAGAAAGTTGTCTGAGAACTTTGACTGTGGCCCTGGACTTTGCTGTGCTC
   : : : : : : : : : : : : : : : : : : : : : :
D2  --CAAG-----GATCTCTTGGAA---TGACAAATGATAGGTACCT-----
                260                270                280                290

      600      610      620      630      640      650
15 ./wwwt GTCATTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAA
   : : : : : : : :
D2  -----AAAAT--GTAA-----
                300

      660      670      680      690      700      710
20 ./wwwt GAGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTG
   : : : : : : : : : : : : : : : : : : : :
D2  -----CATGAAAATACT-----AGCTT-----ATTTTC-----TGAAA-----TG
                310                320                330

      720      730      740      750      760      770
25 ./wwwt GACTACTGTGTCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAAGAGTATGCC
   : : : : : : : : : : : : : : : : : : : :
D2  TACTA-----TCTTA-----ATGCTTAAATTATA-TTTCCC
                340                350                360

      780      790      800      810      820      830
30 ./wwwt AAAAAATAGAAAAGCTATAA-ATATTTCAAAATAAAGAAGAATCCACATTGCAAAAAAAA
   : : : : : : : : : : : : : : : : : : : :
D2  TT----TAG---GCTGTGATAGTTTTTGAAATAAA---ATTTAACATT--TAATATCA
                370                380                390                400

      840
35 ./wwwt AAAAAAAAAA
   : : : : : : : :
D2  AAAAAAAAAA
                410
40
45

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 4 848 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 36,9%; puntuación de alineación global: -1046

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGACGACGCTGCTGAGCTGCCAGCTTAGTGGAAGCTCTGCTCTGGGT
      :: ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
5 D3    ---TTGGGATGGAGT-----TGATGA-----AGGTCCATAAAAGGTTTTTC-----
      10      20      30

      70      80      90      100     110     120
./wwwt GGAGAGCAGCCTCGCTTTGGTGACGCACAGTGCTGGGACCCTCCAGGAGCCCCGGGATTG
      : ::::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
10 D3    --ATAGCA---TCTATTATATAAA-----AGGAGTTCC---ACTG
      40      50      60

      130     140     150     160     170     180
./wwwt AAGGATGGTGGCGGCCGCTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCCTGGGAGCTCT
      :      :      :      :      :      :      :      :      :      :      :
15 D3    CA-----TTTGGATAAGCTGGTTTAG-TGTCTC-TGACAAGTGT
      70      80      90      100

      190     200     210     220     230     240
./wwwt GGTCTTGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGCTC
      :  :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
20 D3    GAAGC---CTAGAAGAATTACTGGCTTGATGGTGATCTTCTGTATCCGGCAAG-----
      110     120     130     140     150

      250     260     270     280     290     300
./wwwt ACAGTGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGGGATGA
      ::::  : :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
30 D3    ACAGACCTTCTC---CACAGTAA-CAACGCTGGAAATATT-----TCTAGTCC---ATGA
      160     170     180     190     200

      310     320     330     340     350     360
./wwwt GAAGCCGTTCTGTGCTACATGTCGTGGGTTGCGGAGGAGGTGCCAGCGAGATGCCATGTG
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
35 D3    GA-GCC-TTTTCCTA-----TG
      210      220

      370     380     390     400     410     420
./wwwt CTGCCCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAAT
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
40 D3    CT---TGGTACACACT-----TGACCTTC-----T
      230      240

      430     440     450     460     470     480
./wwwt ATTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCC
      :  :  :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
50 D3    TTCAGGA---CAGGTTTAC-----AGATCTTGGAC---CAGAAGT-----

```

ES 2 285 768 T3

```

      250      260      270
      490      500      510      520      530
5  ./wwwt AGTCCAGGAAAAC-CAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -GTCTAGCACAAACACAATCC---TGAGGCACAGTCTGATGA-----CCGGA---
      280      290      300      310

      540      550      560      570      580      590
10 ./wwwt GACAAGAGGGAGAAAAGTTGTCTGAGAACTTTTGGCTGTGGCCCTGGACTTTGCTGTGCTC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  GACAAAC-----AGAACCCTT-----CTTGTCCCTTTGGTGTGATA
      320      330      340      350

      600      610      620      630      640      650
15 ./wwwt GTCATTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  --CATTTTGG-AAGACAAGGTG-----GTTCTTCTGGA---ATA---CCCATCCA---
      360      370      380      390

      660      670      680      690      700      710
20 ./wwwt GAGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -AGGTGCTA-----TGATCA--TTACCAAAGCTTTCAGTGATG-----GTTTCCT--
      400      410      420      430

      720      730      740      750      760      770
25 ./wwwt GACTACTGTGTCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAGAGTATGCC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -----CAATT-----TC-----TCCTCG-----
      440

      780      790      800      810      820      830
30 ./wwwt AAAAAATAGAAAAGCTATAAATATTTCAAATAAAGAAGAATCCACATTGCAAAAAAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -----GAAATGATTTTGATC-----AGAAG-----
      450      460

      840
35 ./wwwt AAAAAAAAAA
D3  -----

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 4 848 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 33,0%; puntuación de alineación global: -1306

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGAGACGACGTGCTGAGCTGCCAGCTTAGTGGAAGCTCTGCTCTGGGT
      :  :  :      :  :  :  :  :  :  :  :  :  :  :  :  :
5 D4      -----CTTGC GTTGT TTTGGTACAGA-CTTCCCCCTGA-TGGA-----GCACTGGTT
      10      20      30      40

      70      80      90      100     110     120
./wwwt GGAGAGCAGCCTCGCTTTGGTGACGCACAGTGCTGGGACCCTCCAGGAGCCCCGGGATTG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
10 D4      TG-----CAGATT-----TTGGTCCAGAAA--TGACGAG-CACAGCAAAA--CCCTTCAATG
      50      60      70      80      90

      130     140     150     160     170     180
./wwwt AAGGATGGTGGCGGCCGCTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCCCTGGGAGCTCT
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
15 D4      CAGTCTGATGATCGTAG-----GCAGGGG-----TCTCCT-----
      100     110     120

      190     200     210     220     230     240
./wwwt GGTCCCTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
20 D4      --TCATGCCCTTTTATA-----TG-TGACA-----TC
      130     140

      250     260     270     280     290     300
./wwwt ACAGTGCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATGA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
25 D4      TTAGTG--TGT-----GGTCTTC-----CTAGA--TTCTGCC--ATCCC-----
      150     160     170

      310     320     330     340     350     360
./wwwt GAAGCCGTTCTGTGCTACATGTCGTGGGTTGCGGAGGAGGTGCCAGCGAGATGCCATGTG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
30 D4      -AAGTCATGGTTTGAGTAATGACCGTGGTTTCGATCTCTGTGCC---GAG-TACCATC--
      180     190     200     210     220     230

      370     380     390     400     410     420
./wwwt CTGCCCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAAT
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
35 D4      CAGAGCCGGA-----TGTGAGGG--GTT-----AAGATGC-----T
      240     250     260

      430     440     450     460     470     480
./wwwt ATTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
40 D4      TTCAGTAA-----CTGG--GATACAGATGCCATTATTGCAGCGGGT--ACTGGGG--C
      430     440     450     460     470     480

```

ES 2 285 768 T3

```

                270          280          290          300
          490      500      510      520      530      540
5  ./wwwt AGTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGG
   : :
D4  AG-----CACATGCCA-----
   310

          550      560      570      580      590      600
10 ./wwwt ACAAGAGGGAGAAAGTTGTCTGAGAACTTTTGACTGTGGCCCTGGACTTTGCTGTGCTCG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  -----TCT-----CGGTGGCA-----GCGCT
   320          330

          610      620      630      640      650      660
15 ./wwwt TCATTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAAG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  TCTTTTCTCCGACA-----CACCA-----TGCAGG-----CCGATG
   340      350          360

          670      680      690      700      710      720
20 ./wwwt AGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTGG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  A-----TCCTTGGTGGGGACTGTGGC-----
   370      380

          730      740      750      760      770      780
30 ./wwwt ACTACTGTGTCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAGAGTATGCCA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  -----AAT--ACCTCCCAACNGCTG-----TGCCG
   390          400          410

          790      800      810      820      830      840
35 ./wwwt AAAAAATAGAAAAGCTATAAATATTTCAAAATAAAGAAGAATCCACATTGCAAAAAAAAAA
   : :
D4  AANT-----

40

./wwwt AAAAAAA
D4  -----
45

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 4 848 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 41,9%; puntuación de alineación global: -217

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGAGACGACGTGCTGAGCTGCCAGCTTAGTGGAAGCTCTGCTCTGGGT
5 D8 -----

      70      80      90     100     110     120
./wwwt GGAGAGCAGCCTCGCTTTGGTGACGCACAGTGCTGGGACCCTCCAGGAGCCCCGGGATTG
10 D8 -----

     130     140     150     160     170     180
./wwwt AAGGATGGTGGCGGCCGTCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCTGGGAGCTCT
15 D8 -----

     190     200     210     220     230     240
./wwwt GGTCCCTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTC
20 D8 -----

     250     260     270     280     290     300
./wwwt ACAGTGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATGA
25 D8 -----

     310     320     330     340     350     360
./wwwt GAAGCCGTTCTGTGCTACATGTCGTGGGTTCGGGAGGAGGTGCCAGCGAGATGCCATGTG
30 D8 -----

     370     380     390     400     410     420
./wwwt CTGCCCTGGGACACTCTGTGTGAACGATGTTGTACTACGATGGAAGATGCAACCCCAAT
35 D8 -----

     430     440     450     460     470     480
./wwwt ATTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCC
40 D8 -----
                                     : : : :
45 D8 -----CACCC
50
55
60
65

```

ES 2 285 768 T3

```

      490      500      510      520      530      540
./wwwt AGTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGG
      .....
D8      AGTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGG
      10      20      30      40      50      60

      550      560      570      580      590      600
./wwwt ACAAGAGGGAGAAAGTTGTCTGAGAACTTTTGACTGTGGCCCTGGACTTTGCTGTGCTCG
      .....
D8      ACAAGAGGGAGAAAGTTGTCTGAGAACTTTTGACTGTGGCCCTGGACTTTGCTGTGCTCG
      70      80      90      100      110      120

      610      620      630      640      650      660
./wwwt TCATTTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAAG
      .....
D8      TCATTTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAAG
      130      140      150      160      170      180

      670      680      690      700      710      720
./wwwt AGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTGG
      .....
D8      AGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTGG
      190      200      210      220      230      240

      730      740      750      760      770      780
./wwwt ACTACTGTGTGCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAGAGTATGCCA
      .....
D8      ACTACTGTGTGCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAGAGTATGCCA
      250      260      270      280      290      300

      790      800      810      820      830      840
./wwwt AAAAAATAGAAAAGCTATAAATATTTCAAATAAAGAAGAATCCACATTGCAAAAAAAAAA
      .....
D8      AAAAAATAGAAAAGCTATAAATATTTCAAATAAAGAAGAATCCACATTGC-----
      310      320      330      340      350

./wwwt AAAAAAAA
D8      -----

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 6 672 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 42,1%; puntuación de alineación global: -579

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGGTGGCGGCGCTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCTGGGAGCTCTGGTC
      :          : : : : : : : : : : : : : : : : : : : : : : : :
5 D1    CAG- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
      10      20

      70      80      90      100      110      120
./wwwt CTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTCACAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
10 D1    CT- - - - - TCTA- - - - - ATGAAGA-CAATGAT- - - - - ATTGA- - - - - CACTG
      30      40      50

      130      140      150      160      170      180
./wwwt TGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCGCGATGAGAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 D1    TCCCTCTTTGGCAGT- - - - - TGCATTA- - - - - GTAAC TT- - - - - TGAAA- - -
      60      70      80      90

      190      200      210      220      230
./wwwt CCGTTCTGTGCTACATGTC- GTGGGTTGCGGAGGAGGTG- - - - - CCAGCGAGATGCCATGTGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
20 D1    - - - - - GGTATATGACTGAGCGTAGCATAC-AGGTTAACCTGC-AGAAACA- - - - - GTAC
      100      110      120      130      140

      240      250      260      270      280      290
./wwwt TGCCCTGGGACACTCTGTGTGAACGATGTTGTACTACGATGGAAGATGCAACCCCAATA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 D1    T- - - - TAGGTAATTGTAGGG- - - - - CGAGGATTATAA- - - - - ATGAAATTTGCAA- - - - - AAT-
      150      160      170      180

      300      310      320      330      340      350
./wwwt TTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAAC TGGGCACCCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 D1    - - - - - CA-CTT-AGCAGCAA- - - - - CTGAAGN- - - - - CAATTATCAACCAC
      190      200      210

      360      370      380      390      400      410
./wwwt GTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
40 D1    GTGGAGAAAATCAAACCGAGCAGG- - - - - GCTGTGT- - - - - GAAA- - - - - CATGGTTGTAAT- - - - - A
      220      230      240      250      260

      420      430      440      450      460      470
./wwwt CAAGAGGGAGAAAGTTGTCTGAGAACTTTTGA CTGTGGCCCTGGACTT- TGCTGTGCTCG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 D1    TGCGACTGCGAACACTG- - - - - AACT- - - - - CTACGCCACTCCACAAATGATGTTTTCA

```

ES 2 285 768 T3

```

      270      280      290      300      310
      480      490      500      510      520      530
5  ./wwwt TCATTTTGGACGAAAATTGTAAGCCAGTCCTTTGGAGGGACAGGTCTGC-TCCAGAA
      :  : : : :      : : : :      : : : : : : : : : : : : : : :
D1  GGTGTCATGGAC-----TGTT-GCCAC-----CATGTATTCATCCA---
      320      330      340      350

      540      550      560      570      580      590
10 ./wwwt GAGGGCATAAAGACACTGCTCAAGCTCCAGA--AATCTTCCAGCGTTGCGACTGTGGCCC
      : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  GAGTTCCTTAAAGTT-----TAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAGTT
      360      370      380      390      400

      600      610      620      630      640      650
15 ./wwwt TGGACTACTGTGTGCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAGAGTATG
      : : : : : : : : : : : : : : : : : : : : : : : : : :
D1  TTAAATTATGTATAAACATAAGTTG-----CATTTAGAA-----ATCAAGCATA--
      410      420      430      440

      660      670
25 ./wwwt CCAAAAAATAGAAAAGCTA
      : : : : : : : :
D1  --AATCACTTCAACTGCT-
      450      460

```

ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 6 672 nt frente a D2 rev 416 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 Identidad del 38,6%; puntuación de alineación global: -742

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGGTGGCGGCCGTCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCTGGGAGCTCTGGTC
:      :      :      :      :      :
5 D2    AA-----CTCCCCCTGTGA-----
      10

      70      80      90      100     110     120
./wwwt CTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTCACAG
: : : : : : : : : : : : : : : :
10 D2    -----TTGCAGTAAATTA-----CTGTATTGT-----AAATTCTCA-----
      20      30      40

      130     140     150     160     170     180
./wwwt TGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATGAGAAG
: : : : : : : : : : : : : : : :
15 D2    -----GTGTGGCACT-----TACCTGTAAAT-----GCAATGAAA-----
      50      60      70

      190     200     210     220     230     240
./wwwt CCGTTCTGTGCTACATGTCTGGGTTGCGGAGGAGGTGCCAGCGAGATGCCATGTGCTGC
: : : : : : : : : : : : : : : :
25 D2    -----CTTT--TAATTATT-----TTTCTAA--AGGTGCT--GCACTGCCTATTTTTC
      80      90      100     110

      250     260     270     280     290     300
./wwwt CCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAATATTA
: : : : : : : : : : : : : : : :
30 D2    TCTTGT-----TATGTAA--ATTTTGTAC-ACATTGATTGTT-----ATCTT-
      120     130     140     150

      310     320     330     340     350     360
./wwwt GAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCCAGTC
: : : : : : : : : : : : : : : :
35 D2    GACTGACAA----ATATTCTATATTGAAC--TGAAGTAA-ATCATTTTCAGCTTATAGTT
      160     170     180     190     200

      370     380     390     400     410
./wwwt CAGGAAAACC-AACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGGACA
: : : : : : : : : : : : : : : :
45 D2    CTTAAAAGCATAACCCCTTTA-----CCCCATTT-----AATTCTAGAGTCTAGAACG--CA
      210     220     230     240     250

      420     430     440     450     460     470
./wwwt AGAGGGAGAAAAGTTGTCTGAGAACTTTTGTGCTGGCCCTGGACTTTGCTGTGCTCGTCA
: : : : : : : : : : : : : : : :
50 D2    AG-----GATCTCTTGGAA-----TGACAAATGATAGGTACCTAA-----A
      55

```

ES 2 285 768 T3

```

      260      270      280      290
      480      490      500      510      520      530
5  ./wwwt TTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGCTGCTCCAGAAGAGG
      : : : : : : : : : : : : : : : : :
D2  ATGTAACATGAAAATA-----CTAG--CTTATTTCTGAAATGTACTATC-----
      300      310      320      330

      540      550      560      570      580      590
10 ./wwwt GCATAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTGGACT
      : : : : : : : : : : : : : : : : :
D2  ---TTAA---TGCTTAAATT-----ATATTTC--CCTTTAGGCTGTGA-----T
      340      350      360      370

      600      610      620      630      640      650
15 ./wwwt ACTGTGTCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAAGAGTATGCCAAAA
      : : : : : : : : : : : : : : : : :
D2  AGTTTTTGAATAAAATTTAACAT-----TTAA---TAT---CAAAA
      380      390      400

      660      670
20 ./wwwt AATAGAAAAGCTA
      : : : : :
D2  AA---AAAA---A
      410

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 6 672 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 43,7%; puntuación de alineación global: -481

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGGTGGCGGCCGTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCCTGGGAGCTCTGGTC
      : :      : : : : : :      : :      : : : : : :
5 D3 TTNG-----GATGGAGTTGA-----TGAAGGTCCATAAAAGGTTT--TC
      10      20      30

      70      80      90      100      110
./wwwt CTGGACTTCAACAACATCA--GGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTCAC
      : : : : : : : : : : : : : : : : : : : : : :
10 D3 ATAG-CATCTATTATATAAAAGGAGTTCCA----CTGCATTTGGATAAGCTGGTTT---
      40      50      60      70      80

      120      130      140      150      160      170
./wwwt AGTGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATGAGA
      : : : : : : : : : : : : : : : : : : : : : :
15 D3 AGTGTCT--CTGACAAG--TGT-----GAA-----GCCT-----AGA
      90      100      110

      180      190      200      210      220      230
./wwwt AGCCGTTCTGTGCTACATGTCGTGGGTTGCCGAGGAGGTGCCAGCGAGATGCCATGTGCT
      : : : : : : : : : : : : : : : : : : : : : :
25 D3 AGAATTACTG-GCTTGATGGTGATCTTTCTGTATCCG--GCAAGACAGA--CCTT--CT
      120      130      140      150      160

      240      250      260      270      280      290
./wwwt GCCCTGGGACA-CTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAATA
      : : : : : : : : : : : : : : : : : : : : : :
30 D3 CCACAGTAACAACGCTG---GA---ATATTTCTAGTCC-ATGAGAGCCTTTCTCCTAT-
      170      180      190      200      210      220

      300      310      320      330      340      350
./wwwt TTAGAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCCA
      : : : : : : : : : : : : : : : : : : : : : :
35 D3 -----GCTTGGTACACACT-TGACCTTCTTCA----GGACA----GGTTTACA
      230      240      250      260

      360      370      380      390      400      410
./wwwt G-TCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGG
      : : : : : : : : : : : : : : : : : : : : : :
45 D3 GATCTTGGA-----CCAGAAGT--GTCTAGCACAAACACAATCCTGAGGCACAGTCTG
      270      280      290      300      310

      420      430      440      450      460      470
./wwwt ACAAGAGG-GAGAAAGTTGTCTGAQAACCTTTTGACTGTGGCCCTGGACTTTGCTGTGCTC
      : : : : : : : : : : : : : : : : : : : : : :
50 D3 ATGACCGGAGACAAAC-----AGAACCTT-----CTTGTCTTTGGTGTGATA

```

ES 2 285 768 T3

```

      320      330      340      350
      480      490      500      510      520      530
5  ./wwwt GTCATTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  --CATTTTGG-AAGACAAGGTG-----GTTCTTCTGGA--ATA--CCCATCCA---
      360      370      380      390

      540      550      560      570      580      590
10 ./wwwt GAGGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -AGGTGCTA-----TGATCATTACC--AAAGCTTTCAGTGATG-----GTTTCCT--
      400      410      420      430

      600      610      620      630      640      650
15 ./wwwt GACTACTGTGTCGAAGCCAAATTGACCAGCAATCGGCAGCATGCTCGATTAAAGAGTATGCC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3  -----CAATT-----TC-----TCCTCGGAAATGATTTTG--
      440      450      460

      660      670
25 ./wwwt AAAAAATAGAAAAGCTA
      : : : :
D3  ----ATCAGAAG-----

```

30 ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 6 672 nt frente a D4 rev 416 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 35 Identidad del 39,4%; puntuación de alineación global: -670

40

45

50

55

60

65

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGGTGGCGGCCGTCCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCTGGGAGCTCTGGTC
      :: :: :: :      ::::
5 D4 -----CTTGCGTTGTT-----TGGTA
      10

      70      80      90      100      110      120
./wwwt CTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGAAGGGCTCACAG
      : : :::: : : :::: : : : : : : : : : : : : : :
10 D4 CAG-ACTTCCCCCTGAT--GGAGCACTGGTTTGCAGATTTTGGTCCAGAAA-----
      20      30      40      50      60

      130      140      150      160      170      180
./wwwt TGCCTGTCTGACACGGACTGCAATACCAGAAAGTTCTGCCCTCCAGCCCCCGCATGAGAAG
      : : : : : : : : : : : : : : : : : : : : : :
15 D4 -----TGACGAGCACAGCAAAACC-----CTTCA-----ATG
      70      80      90

      190      200      210      220      230
./wwwt CCGTTCTGTGCTACATG-TCGTGGGTGCGGAGGAGGTGCCAGCGAGATGCCATGTGCTG
      : : : : : : : : : : : : : : : : : : : : : :
25 D4 CAGT-CTG-----ATGATCGTAGGC-----AGGGGTCTCCTTCATGCCCTTTTAT-
      100      110      120      130

      240      250      260      270      280      290
./wwwt CCCTGGGACA-CTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAATAT
      : : : : : : : : : : : : : : : : : : : : : :
30 D4 --ATGTGACATCTTAGTGTG-----TGGTCTTCCTA-----GATTCTGCCATCCCAA---
      140      150      160      170      180

      300      310      320      330      340      350
./wwwt TAGAAAGGCA-GCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCCA
      : : : : : : : : : : : : : : : : : : : : : :
35 D4 -----GTCATGGTT--TGAGTA--ATGAC---CGTGGTTTCG----ATCTCTGTGCCGA
      190      200      210      220

      360      370      380      390      400      410
./wwwt GTCCAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGGA
      : : : : : : : : : : : : : : : : : : : : : :
40 D4 GT-----ACCATCC---AG--AGCCGGGATGTGAGGGGTTAAGATGCTTTTCAG---
      230      240      250      260

      420      430      440      450      460      470
./wwwt CAAGAGGGGAGAAAGTTGTCTGAGAACTTTTGACTGTGGCCCTGGACTTTGCTGTGCTCGT
      : : : : : : : : : : : : : : : : : : : : : :
50 D4 TAACTGGGATACAGATGCC-----ATTATTGCAGCGGGTACTGGG-----GCAGCACATGC

```

ES 2 285 768 T3

```

      270      280      290      300      310
5      480      490      500      510      520      530
./wwwt CATTTTTGGACGAAAATTTGTAAGCCAGTCCTTTTGGAGGGACAGGTCTGCTCCAGAAGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4      CATCTCGG-----TGGCAGCGCTTCTTTTT-----TCTCCGACACAC--
      320      330      340      350

10     540      550      560      570      580      590
./wwwt GGGCATAAAGACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGTGGCCCTGGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4      ---CATGCAG---GCCGATGATCC-----TTG---GTGGGGACTGTGGC-----
      360      370      380

15     600      610      620      630      640      650
./wwwt CTACTGTGTGCGAAGCCAATTGACCAGCAATCGGCAGCATGCTCGATTAAAGATATGCCAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4      -----AAT--ACCTCCCAACNGCTG-----TGCCGA
      390      400      410

20     660      670
./wwwt AAAATAGAAAAGCTA
      :
D4      ANT-----

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 6 672 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 45,5%; puntuación de alineación global: 45

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGGTGGCGGCCGTCTCTGCTGGGGCTGAGCTGGCTCTGCTCTCCCCTGGGAGCTCTGGTC
5 D8 -----

      70      80      90      100      110      120
10 ./wwwt CTGGACTTCAACAACATCAGGAGCTCTGCTGACCTGCATGGGGCCCGGAAGGGCTCACAG
D8 -----

      130      140      150      160      170      180
15 ./wwwt TGCCTGTCTGACACGGAAGTCAATACCAGAAAGTTCTGCCTCCAGCCCCGCGATGAGAAG
D8 -----

      190      200      210      220      230      240
20 ./wwwt CCGTTCTGTGCTACATGTCGTGGGTTGCGGAGGAGGTGCCAGCGAGATGCCATGTGCTGC
25 D8 -----

      250      260      270      280      290      300
30 ./wwwt CCTGGGACACTCTGTGTGAACGATGTTTGTACTACGATGGAAGATGCAACCCCAATATTA
D8 -----

      310      320      330      340      350      360
35 ./wwwt GAAAGGCAGCTTGATGAGCAAGATGGCACACATGCAGAAGGAACAACCTGGGCACCCAGTC
D8 -----CAGTC

      370      380      390      400      410      420
40 ./wwwt CAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGGACAA
D8 CAGGAAAACCAACCCAAAAGGAAGCCAAGTATTAAGAAATCACAAGGCAGGAAGGGACAA
      10      20      30      40      50      60

      430      440      450      460      470      480
50 ./wwwt GAGGGAGAAAAGTTGTCTGAGAACTTTTGGCTGTGGCCCTGGACTTTGCTGTGCTCGTCAT
D8 GAGGGAGAAAAGTTGTCTGAGAACTTTTGGCTGTGGCCCTGGACTTTGCTGTGCTCGTCAT

```

5
10
15
20
25
30
35
40
45
50
55
60
65

ALIGN calcula una alineación global de dos secuencias
Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
Seq 7 1529 nt frente a D1 rev 465 nt
Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
Identidad del 24,4%; puntuación de alineación global: -3426

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt CCGGACGCGTGGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTT
      : :      : :      : :      : :      : :
5 D1    CAG-----TGTTCCT-CACAGT---GTGGGG---CAGCC-----
      10      20

      70      80      90      100     110     120
./wwwt CTCTTTCTTTCTCCCTCTTGAGTCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTA
      : :      : :      : :      : :      : :
10 D1    -----GTCCTTCTA--ATGAAGACA-----
      30      40

      130     140     150     160     170     180
./wwwt CCGGGGTCITTTGTCGCGATGGTAGCGGCGGCTCTCGGCGGCCACCCCTCTGCTGGGAGTGA
      : :      : :      : :      : :      : :
15 D1    -----ATGATATTGACA-CTGTC-----CCTCTT-TGGCAGT--
      50      60      70

      190     200     210     220     230     240
./wwwt GCGCCACCTTGAACTCGGTTCTCAATTCCAACGCTATCAAGAACCTGCCCCACCGCTGG
      : :      : :      : :      : :      : :
20 D1    -----TGCATTAG---TAACTTTGAAAGGTAT-ATGA-----CTGA
      80      90      100

      250     260     270     280     290     300
./wwwt GCGGCGCTGCGGGGCACCCAGGCTCTGCAGTCAGCGCCGCGCGGGAATCCTGTACCCGG
      : :      : :      : :      : :      : :
25 D1    -----GCGTAGCATACAGGTT-----AACCTG-----
      110     120

      310     320     330     340     350     360
./wwwt GCGGGAATAAGTACCAGACCATTGACAACCTACCAGCCGTACCCGTGCGCAGAGGACGAGG
      : :      : :      : :      : :      : :
30 D1    -----CAGA-----AAC-----
      130

      370     380     390     400     410     420
./wwwt AGTGCGGCACTGATGAGTACTGCGCTAGTCCCACCCGCGGAGGGGACGCAGGCGTGCAAA
      : :      : :      : :      : :      : :
35 D1    -----AGTACT---TAG-----GT---AA
      140

      430     440     450     460     470     480
./wwwt TCTGTCTCGCCTGCAGGAAGCGCCGAAAACGCTGCATGCGTCACGCTATGTGCTGCCCG
      : :      : :      : :      : :      : :
40 D1    T-----TGTAGGG-----CG

```

ES 2 285 768 T3

150

```

5      490      500      510      520      530      540
./wwwt GGAATTACTGCAAAAATGGAATATGCGTGTCTTCTGATCAAAATCATTTCGAGGAGAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      AGGATTAT-----AAATGAAATTTGC-----AAAATCACTT---AGCAGCAA
      160      170      180      190

10     550      560      570      580      590      600
./wwwt TTGAGGAAACCATCACTGAAAGCTTTGGTAATGATCATAGCACCTTGGATGGGTATTCCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----CTGAA-----GNCAATTATCA-----
      200      210

15     610      620      630      640      650      660
./wwwt GAAGAACCACCTTGTCTTCAAAAATGTATCACACCAAGGACAAGAAGTTCTGTTTGTCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----ACCACGTGG-----AGAAA-----
      220

20     670      680      690      700      710      720
./wwwt TCCGGTTCATCAGACTGTGCCTCAGGATTGTGTTGTGCTAGACACTTCTGCTCCAAGATCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----ATCAAACCGAG-----CAGGGCTGTGT-----
      230      240

25     730      740      750      760      770      780
./wwwt GTAAACCTGTCTCTGAAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----GAAACATGGT-----TGTA-----
      250      260

30     790      800      810      820      830      840
./wwwt GACTAGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----ATAT-----GCG-----ACTG-----CGAACACTGAAC-
      270      280

35     850      860      870      880      890      900
./wwwt ATCACCATCAAGCCAGTAATTCTTCTAGGCTTCACACTTGTCTAGAGACACTAAACCAGCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -----TCTACGC-----CACT-----
      290

40     910      920      930      940      950      960
./wwwt ATCCAAAATGCAGTGAACCTCTTTTATATAATAGATGCTATGAAAACCTTTTATGACCTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      --CCACAA-----ATGA-----
      300

45     970      980      990      1000      1010      1020
./wwwt CATCAACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCTCAATAACA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

ES 2 285 768 T3

```

D1 -----TGTTTTTCAG-----
                      310

5      1030      1040      1050      1060      1070      1080
./wwwt CCTTCCAAAAACCTGGAGTGTAAGAGCTTTGTTTCTTTATGGAACCTCCCTGTGATTGCA
                      :: ::      ::::      :::
D1      -----GTGTC---ATGGA---CTGT-----
                      320                      330

10     1090      1100      1110      1120      1130      1140
./wwwt GTAAATTACTGTATTGTAAATTCTCAGTGTGGCACTTACCTGTAAATGCAATGAAACTTT
                      :: ::      : ::::
D1      -----TGCCAC---CATGTA-----
                      340

15     1150      1160      1170      1180      1190      1200
./wwwt TAATTATTTTTCTAAAGGTGCTGCACTGCCTATTTTTCCTCTTGTTATGTAAATTTTTGT
                      ::: ::
D1      -----TTCATC-----

25     1210      1220      1230      1240      1250      1260
./wwwt ACACATTGATTGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCATTTC
:
D1      -----C

30     1270      1280      1290      1300      1310      1320
./wwwt AGCTTATAGTTCTTAAAAGCATAACCCTTTACCCCATTTNATTCTAGAGTCNAGAACGCA
: :      ::::: : :      :::
D1      AG-----AGTTCTTAAAG-----TTTA-----A
350      360

40     1330      1340      1350      1360      1370      1380
./wwwt AGGATCTCTTGGAATGACAAATGATAGGTACCTAAAATGTAACATGAAAATACTAGCTTA
: :      : :      : : ::::      : :      : : : :
D1      AG-----TTG-----CACATGAT-----TGTA-----TA -AGCATG
370      380      390

45     1390      1400      1410      1420      1430      1440
./wwwt TTTTCTGAAATGTACTATCTTAATGCTTAAATTATATTTCCCTTTAGGCTGTGATAGTTT
: : : :      : : :      : : : : : :
D1      CTTTCT-----TTGAGTTTTAAATTATGTAT-----
50      400      410

55     1450      1460      1470      1480      1490      1500
./wwwt TTGAAATAAAATTTAACATTTAATATCATGAAATGKTATAAGTAGACATAAAAAAAAAA
: : : : : : : : : : : : : : : : : :
D1      ---AAACATAAGTTG-CATTTA-----GAAATC---AAG---CATAAATCACTTCA
420      430      440      450

60     1510      1520
./wwwt AAAAAAAAAAAGGGCGGCCGCTAGACTAG

:
D1      A-----CTGCT-----
65      460

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 7 1529 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 27,1%; puntuación de alineación global: -2414

```

10      10      20      30      40      50      60
./wwwt CCGGACGCGTGCGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTT
D2      -----

15      70      80      90      100     110     120
./wwwt CTCTTTCTTTTCTCCCTCTTGAGTCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTA
D2      -----

20     130     140     150     160     170     180
./wwwt CCCGGGTCCTTTGTTCGCGATGGTAGCGGCGGCTCTCGGCGGCCACCCCTCTGCTGGGAGTGA
D2      -----

25     190     200     210     220     230     240
./wwwt GCGCCACCTTGAACCTCGGTTCTCAATCCAACGCTATCAAGAACCTGCCCCACCGCTGG
D2      -----

30     250     260     270     280     290     300
./wwwt GCGGCGCTGCGGGGCACCCAGGCTCTGCAGTCAGCGCCGCGCCGGAATCCTGTACCCGG
D2      -----

35     310     320     330     340     350     360
./wwwt GCGGGAATAAGTACCAGACCATTGACAACTACCAGCCGTACCCGTGCGCAGAGGACGAGG
D2      -----

40     370     380     390     400     410     420
./wwwt AGTGCGGCACTGATGAGTACTGCGCTAGTCCCACCCGCGGAGGGGACGCAGGCGTGCAA
D2      -----

45     430     440     450     460     470     480
./wwwt TCTGTCTCGCCTGCAGGAAGCGCCGAAAACGCTGCATGCGTCACGCTATGTGCTGCCCCG
D2      -----

```

ES 2 285 768 T3

```

      490      500      510      520      530      540
./wwwt GGAATTACTGCAAAAATGGAATATGCGTGTCTTCTGATCAAAATCATTTCGAGGAGAAA
5 D2 -----

      550      560      570      580      590      600
10 ./wwwt TTGAGGAAACCATCACTGAAAGCTTTGGTAATGATCATAGCACCTTGGATGGGTATTCCA
D2 -----

      610      620      630      640      650      660
15 ./wwwt GAAGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTCTGTTTGTCT
D2 -----

      670      680      690      700      710      720
20 ./wwwt TCCGGTCATCAGACTGTGCCTCAGGATTGTGTTGTGCTAGACACTTCTGGTCCAAGATCT
D2 -----

      730      740      750      760      770      780
30 ./wwwt GTAAACCTGTCTCTGAAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATG
D2 -----

      790      800      810      820      830      840
35 ./wwwt GACTAGAAATATTCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAAAG
D2 -----

      850      860      870      880      890      900
40 ./wwwt ATCACCATCAAGCCAGTAATTCTTCTAGGCTTCACACTTGTGAGAGACACTAAACCAGCT
D2 -----

      910      920      930      940      950      960
50 ./wwwt ATCCAAATGTCAGTGAACCTCTTTATATAATAGATGCTATGAAAACCTTTTATGACCTT
D2 -----

      970      980      990      1000      1010      1020
55 ./wwwt CATCAACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAACA
60
65

```

ES 2 285 768 T3

[illegible]

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 7 1529 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 24,0%; puntuación de alineación global: -3522

```

10      10      20      30      40      50      60
./wwwt CCGGACGCGTGGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTT
      :  ::
D3      -----TTNGGA-----TG-----

15      70      80      90      100     110     120
./wwwt CTCCTTCTTTCTCCCTCTTGAGTCCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTA
      :  ::  ::::
D3      -----GAGTTGATG-----
      10

20      130     140     150     160     170     180
./wwwt CCCGGGTCTTTGTTCGCGATGGTAGCGGCGGCTCTCGGCGGCCACCTCTGCTGGGAGTGA
      :  ::
D3      -----AAGGT-----
      20

25      190     200     210     220     230     240
./wwwt GCGCCACCTTGAACCTCGGTTCTCAATTCCAACGCTATCAAGAACCTGCCCCACCGCTGG
      :  :  :  :  :  :  :  :
D3      -----CCATAAA--AGGTTTTC--TAGCA-----
      30      40

30      250     260     270     280     290     300
./wwwt GCGGCGCTGCGGGGCACCCAGGCTCTGCAGTCAGCGCCGCGCGGGAATCCTGTACCCGG
      :  ::
D3      -----TCT-----

35      310     320     330     340     350     360
./wwwt GCGGGAATAAGTACCAGACCATTTGACAACTACCAGCCGTACCCGTGCGCAGAGGACGAGG
      :  ::  :  :  :  :
D3      -----ATTATATA-----AA-----ACG
      50

40      370     380     390     400     410     420
./wwwt AGTGCGGCACTGATGAGTACTGCGCTAGTCCCACCCGCGGAGGGGACGCAGGCGTGCAAA
      :  :  :  :  :  :
D3      AGTTCC-----ACTGC-----A
      60      70

45      430     440     450     460     470     480
./wwwt TCTGTCTCGCCTGCAGGAAGCGCCGAAACGCTGCATGCGTCACGCTATGTGCTGCCCGG
      :  :  :  :  :  :
D3      TTTG-----GATAA-GCTG-----GT-----

```

ES 2 285 768 T3

80

```

5      490      500      510      520      530      540
./wwwt GGAATTACTGCAAAAATGGAATATGCGTGTCTTCTGATCAAAATCATTTCGAGGAGAAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      ----TTA-----GTGTCT-CTGAC-----AA
                        90      100

10     550      560      570      580      590      600
./wwwt TTGAGGAAACCATCACTGAAAGCTTTGGTAATGATCATAGCACCTTGGATGGGTATTCCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      GTGTG-----AAGCCT-----A
                        110

15     610      620      630      640      650
./wwwt GAAGAACCACCTTGTCTTCAAAAATGTATCACACCAAGGACAAGAAGGT--TCTGTTTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      GAAGAATTAC--TGGCTT-----GATGGTGATCTTCTG
      120      130      140

20     660      670      680      690      700      710
./wwwt TCTCCGGTCATCAGACTGTGCCTCAGGATTGTGTTGTGCTAGACACTTCTGGTCCAAGAT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      TATCCGG-CA--AGAC-----AGAC-CTTCT---CCA---
      150      160      170

25     720      730      740      750      760      770
./wwwt CTGTAAACCTGTCTTGAAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      CAGTAA-----CAACGC-----
                        180

30     780      790      800      810      820      830
./wwwt TGGACTAGAAATATTCAGCGTTGTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      TGGAATATTTCTAGTCCAT-----GAGA-----GCCTT-----
      190      200      210

35     840      850      860      870      880      890
./wwwt AGATCACCATCAAGCCAGTAATTCTTAGGCTTCACACTTGTCAGAGACACTAAACCAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -----TTCTCTATGCTTGGTAC-----ACACT-----
                        220      230

40     900      910      920      930      940      950
./wwwt CTATCCAAAATGCAGTGAACCTCTTTTATATAATAGATGCTATGAAAACCTTTTATGACC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -----TGA--CCTT-----CTTTTCAGGAC-
                        240      250

45     960      970      980      990      1000      1010
./wwwt TTCATCAACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

ES 2 285 768 T3

```

D3 -----AGG-TTTACA-GATCT-TGG-----
      260

5      1020      1030      1040      1050      1060      1070
./wwwt CACCTTCCAAAAACCTGGAGTGTAAAGAGCTTTGTTTCTTTATGGAACCCCTGTGATTG
      : : : : : : : :
D3 -----ACCAGAAGTGT-----CT-----
      270      280

10     1080     1090     1100     1110     1120     1130
./wwwt CAGTAAATTACTGTATTGTAAATTCTCAGTGTGGCACTTACCTGTAAATGCAATGAAACT
      : : : : : : : : : : : : : :
D3 -----AGCACAAAC-----AATCCTGA-----GGCACAGTC-TG-----ATGACCGAGAC-
      290      300      310      320

15     1140     1150     1160     1170     1180     1190
./wwwt TTAAATTATTTTCTAAAGGTGCTGCAC'TGCC'TATTTTCCTCTTGTTATG-TAAATTTT
      : : : : : : : : : : : : : :
D3 -----AAA-----CAGAACCTTCTTGTCTT-TTGGTGTGATACATTTT
      330      340      350

20     1200     1210     1220     1230     1240     1250
./wwwt TGTACACATTGATTGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCAT
      : : : : : : : : : : : : : :
D3 -----TGAAGACAAGG--TGGT-TCTT-----CTGGAATA-----
      360      370      380

25     1260     1270     1280     1290     1300     1310
./wwwt TTCAGCTTATAGTTCTTAAAGCATAACCCTTTACCCCATTTNA'TCTAGAGTCNAGAAC
      : : : : : : : : : : : : : :
D3 -----CCCATC-----
      390

30     1320     1330     1340     1350     1360     1370
./wwwt GCAAGGATCTCTTGGAATGACAAATGATAGGTACCTAAAATGTAACATGAAAATACTAGC
      : : : : : : : : : : : : : :
D3 -----CAAGGTGCT-----ATGATCATTACCAAA-----GC
      400      410

35     1380     1390     1400     1410     1420     1430
./wwwt TTATTTTCTGAAATGTACTATCTTAATGCTTAAAT'TATATTCCCTTTAGGCTTGTATAG
      : : : : : : : : : : : : : :
D3 -----TT-----TCAG-----TGATGGTTTCCTCA-ATTTCCTC- GGAAATGAT--
      420      430      440      450

40     1440     1450     1460     1470     1480     1490
./wwwt TTTT'TGAAATAAAATTTAACATTTAATATCATGAAATGKTATAAGTAGACATAAAAAAAA
      : : : : : : : : : : : : : :
D3 -----T'TTGA-----TCA-GAA-----
      460

45     1500     1510     1520
./wwwt AAAAAAAAAAAAAAGGGCGGCCGCTAGACTAG
      :
D3 -----G

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 7 1529 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 21,1%; puntuación de alineación global: -3797

```

10      10      20      30      40      50      60
./wwwt CCGGACGCGTGCGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTT
      :      : : : :      : : : :      : :
15 D4    CT---TGC GTTG-----TTTGGTACAGAC-----
      10      20

      70      80      90      100     110     120
./wwwt CTCTTTCTTTCTCCCTCTTGAGTCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTA
      : : : : :      : : : :      : : : :
20 D4    -----TTCCCCCT-----GATG-----GAGC--
      30

      130     140     150     160     170     180
./wwwt CCCGGGTCTTTGTGCGGATGGTAGCGCGGCTCTCGGCGGCCACCCTCTGCTGGGAGTGA
25 D4    -----
      190     200     210     220     230     240
./wwwt GCGCCACCTTGAACCTCGGTTCTCAATTCCAACGCTATCAAGAACCTGCCCCACCGCTGG
      : : : : :
30 D4    -----ACT-GGTT-----
      40

      250     260     270     280     290     300
./wwwt GCGGCGCTGCGGGGCACCCAGGCTCTGCAGTCAGCGCCGCGCGGAATCCTGTACCCGG
      : : : : :      : : : :
40 D4    -----TGCAG-----ATTTTG-----
      50

      310     320     330     340     350     360
./wwwt GCGGGAATAAGTACCAGACCATTGACAACTACCAGCCGTACCCGTGCGCAGAGGACGAGG
45 D4    -----GT-CCAGA-----
      60

      370     380     390     400     410     420
./wwwt AGTGCGGCACTGATGAGTACTGCGCTAGTCCCACCCGCGGAGGGGACGCAGGCGTGCAAA
      : : : : : : : :
50 D4    -----AATGACGAGCACAG-----
      70

      430     440     450     460     470     480
./wwwt TCTGTCTCGCCTGCAGGAAGCGCCGAAAACGCTGCATGCGTCACGCTATGTGCTGCCCGG
      : : : : : : : :
60 D4    -----AAAACCCCTTCA-----

```

ES 2 285 768 T3

80

```

5      490      500      510      520      530      540
./wwwt GGAATTACTGCAAAAATGGAATATGCGTGTCTTCTGATCAAAATCATTTCGAGGAGAAA
      : : : : :
D4      -----ATGCAGTCTG-----
      90

10     550      560      570      580      590      600
./wwwt TTGAGGAAACCATCACTGAAAGCTTTGGTAATGATCATAGCACCTTGGATGGGTATTCCA
      : : : : : : : : : : :
D4      -----ATGATCGTAG-----GCAGGGGTCT-----
      100      110

15     610      620      630      640      650      660
./wwwt GAAGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTCTGTTTGTCT
      : : : : : : : : : : :
D4      -----CCTTCATGCCCTTTTATATGT-----GACA-----TCTTAGTG
      120      130      140      150

20     670      680      690      700      710      720
./wwwt TCCGGTCATCAGACTGTGCCTCAGGATTTGTGTTGTGCTAGACACTTCTGGTCCAAGATCT
      : : : : : : : : : : :
D4      TGTGGTCTTC-----CTAGA--TTCTG--CCA-----
      160      170

25     730      740      750      760      770      780
./wwwt GTAAACCTGTCTTGAAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATG
      : : : : : : : : : : :
D4      -----TCC-----CAAGT-----CAT-----GCTT-----TG
      180      190

30     790      800      810      820      830      840
./wwwt GACTAGAAATATTCCAGCGTTGTTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAAAG
      : : : : : : : : : : :
D4      -----AGTAATG-----ACCGTGGTTTCGATC-----TCTGT--GCCGAGTAC-----
      200      210      220

35     850      860      870      880      890      900
./wwwt ATCACCATCAAGCCAGTAATTCTTCTAGGCTTCACACTTGTTCAGAGACACTAAACCAGCT
      : : : : : : : : : : :
D4      . --CATCCAG-----AGCCGGA-----TGTGAGGG-----GTT
      230      240      250

40     910      920      930      940      950      960
./wwwt ATCCAAAATGCAGTGAACTCCTTTTATATAATAGATGCTATGAAAACCTTTTATGACCTT
      : : : : : : : : : : :
D4      A-----AGATGCT-----TT
      260

45     970      980      990      1000      1010      1020
./wwwt CATCAACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAACA
      : : : : : : : : : : :

```

ES 2 285 768 T3

```

D4      CAGTAACTG-----GGATA-----CAG-----ATGCCATTA---
          270                                280

5      1030      1040      1050      1060      1070      1080
./wwwt CCTTCCAAAACCTGGAGTGTAAGAGCTTTGTTTCTTTATGGAACCTCCCCTGTGATTGCA
          ::::
D4      -----TTGCA
          290

10     1090      1100      1110      1120      1130      1140
./wwwt GTAAATTACTGTATTGTAAATTCTCAGTGTGCCACTTACCTGTAAATGCAATGAACTTT
          :   :   :   :   :   :   :   :   :   :   :   :   :
15     D4      G-----CGG-----GTAC-----TGGGGCA-----GCACATGCCAT---CTC-
          300                                310      320

20     1150      1160      1170      1180      1190      1200
./wwwt TAATTATTTTCTAAAGGTGCTGCACTGCCTATTTTCTCTTGTATGTAAATTTTGT
          :::: : : : : : : : : : :
D4      -----GGTG--GCAGCGCTTCTTTTCTCCG-----
          330      340

25     1210      1220      1230      1240      1250      1260
./wwwt ACACATTGATTGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCATTC
          ::::
D4      ACACA-----
          350

30     1270      1280      1290      1300      1310      1320
./wwwt AGCTTATAGTTCTTAAAGCATAACCCCTTTACCCCATTTNATTCTAGAGTCNAGAACGCA
          :::: : : : :
35     D4      -----CCATG-----CAGGCCG--
          360

40     1330      1340      1350      1360      1370      1380
./wwwt AGGATCTCTTGGAATGACAAATGATAGGTACCTAAAATGTAACATGAAAATACTAGCTTA
          :   :   :   :   :   :   :   :   :   :
D4      ATGATC-CTTGG-----TGGGGAC-----TGTGGCA-----
          370                                380      390

45     1390      1400      1410      1420      1430      1440
./wwwt TTTTCTGAAATGTACTATCTTAATGCTTAAATTATATTTCCCTTTAGGCTGTGATAGTTT
          : : : : : :
D4      -----ATACCTCCC-----
          400

50     1450      1460      1470      1480      1490      1500
./wwwt TTGAAATAAAATTTAACATTTAATATCATGAAATGKTATAAGTAGACATAAAAAAAAAA
          : : : :
55     D4      -----AACNGCTGT-----
          410

60     1510      1520
./wwwt AAAAAAAAAAAGGGCGCCGCTAGACTAG
          : : : :
D4      -----GCCG---AANT--
          410

65

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 7 1529 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 18,5%; puntuación de alineación global: -4009

```

10      10      20      30      40      50      60
./wwwt CCGGACGCGTGGGCGGCACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTT
      :                               : : : : : : : : :
D8      C-----ACCCAGTCCAGGAAA-----
                        10

15      70      80      90     100     110     120
./wwwt CTCTTTCTTTCTCCCTCTTGAGTCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTA
      :
D8      -----A

20      130     140     150     160     170     180
./wwwt CCGGGTCTTTGTGCGGATGGTAGCGGCGGCTCTCGGCGGCCACCCTCTGCTGGGAGTGA
      : :
D8      CC-----

25      190     200     210     220     230     240
./wwwt GCGCCACCTTGAACCTCGGTTCTCAATTCGAACGCTATCAAGAACCTGCCCCACCGCTGG
      : : : : : : :
D8      -----AAC-----CAA-----AAG-----
                        20

30      250     260     270     280     290     300
./wwwt GCGGCGCTGCGGGGCACCCAGGCTCTGCAGTCAGCGCCGCGCCGGAATCCTGTACCCGG
      : : : : :
D8      -----GAAGCC-----
                        30

35      310     320     330     340     350     360
./wwwt GCGGGAATAAGTACCAGACCATTGACAACCTACCAGCCGTACCCGTGCGCAGAGGACGAGG
      : : : : : : :
D8      -----AAGTA-----TTAAGAAAT-----
                        40

40      370     380     390     400     410     420
./wwwt AGTGCGGCACTGATGAGTACTGCGCTAGTCCCACCCGCGGAGGGGACGCAGGCGTGCAA
      : : : : :
D8      -----CACAAG-----
                        50

45      430     440     450     460     470     480
./wwwt TCTGTCTCGCCTGCAGGAAGCGCCGAAAACGCTGCATGCGTCACGCTATGTGCTGCCCG
      : : : : : : :
D8      -----GCAGGAAG-----

```

ES 2 285 768 T3

60

```

5      490      500      510      520      530      540
./wwwt GGAATTACTGCAAAAATGGAATATGCGTGTCTTCTGATCAAAATCATTTCGGAGGAGAAA
      : : : : :
D8      -----GGACAA-----

10     550      560      570      580      590      600
./wwwt TTGAGGAAACCATCACTGAAAGCTTTGGTAATGATCATAGCACCTTGGATGGGTATTCCA
      : : : : : : : : :
D8      --GAGGGA-----GAAAG-----
      70      80

15     610      620      630      640      650      660
./wwwt GAAGAACCACCTTGTCTTCAAAAATGTATCACACCAAAGGACAAGAAGGTTCTGTTTGTC
      : : : : : : : : :
D8      -----TTGTCTG-----AGAA-----CTTT-----
      90

20     670      680      690      700      710      720
./wwwt TCCGGTCATCAGACTGTGCCTCAGGATTTGTGTTGTGCTAGACACTTCTGGTCCAAGATCT
      : : : : : : : : :
D8      -----GACTGTGGCCCTGGACTTTGCTGTGCTCGTCATTTTGGACGAAAATTT
      100      110      120      130      140

25     730      740      750      760      770      780
./wwwt GTAAACCTGTCTCTGAAAGAAGGTCAAGTGTGTACCAAGCATAGGAGAAAAGGCTCTCATG
      : : : : : : : : :
D8      GTAAGCCAGTCCTT-----TTGGAG-----G
      150      160

30     790      800      810      820      830      840
./wwwt GACTAGAAATATTCACGCGTTGTACTGTGGAGAAGGTCTGTCTTGCCGGATACAGAAAG
      : : : : : : : : :
D8      GAC-----AGGTCTG-CT--CCAGA--AGAGGG
      170      180

35     850      860      870      880      890      900
./wwwt ATCACCATCAAGCCAGTAATTCTTCTAGGCTTCACACTTGTGAGAGACACTAAACCAGCT
      : : : : : : : : :
D8      -----CATAAAG-----ACACTGCTCA-----AGCT
      190      200      210

40     910      920      930      940      950      960
./wwwt ATCCAAAATGCAGTGAACCTCTTTTATATAATAGATGCTATGAAAACCTTTTATGACCTT
      : : : : :
D8      --CCAGAA-----

45     970      980      990      1000      1010      1020
./wwwt CATCAACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCATTCCAATAACA
      :

```

ES 2 285 768 T3

D8 -----A

5
 ./wwwt 1030 1040 1050 1060 1070 1080
 CCTTCCAAAAACCTGGAGTGTAAGAGCTTTGTTTCTTTATGGAACTCCCCTGTGATTGCA
 : : : : : : : : : : : : : : : : : :
 D8 TCTTCCA-----GCGTTG-----CGACTGTG--GCC
 220 230 240

10
 1090 1100 1110 1120 1130 1140
 ./wwwt GTAAATTACTGTATTGTAAATTCTCAGTGTGGCACTTACCTGTAAATGCAATGAAACTTT
 : : : : : : : : : : : : : : : : : :
 D8 CTGGACTACTGTGTCG-----AAGCCAAT-----
 250 260

15
 1150 1160 1170 1180 1190 1200
 ./wwwt TAATTATTTTTCTAAAGGTGCTGCACTGCCTATTTTCTCTTGTATGTAAATTTTTGT
 D8 -----

25
 1210 1220 1230 1240 1250 1260
 ./wwwt ACACATTGATTGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCATTTTC
 : : : : : : : : : : : : : : : : : :
 D8 -----TGACCAGCAA-----TC
 270

30
 1270 1280 1290 1300 1310 1320
 ./wwwt AGCTTATAGTTCTTAAAAGCATAACCCCTTTACCCCATTTNATTCTAGAGTCNAGAACGCA
 : : : : : : : : : : : : : : : : : :
 D8 GGC-----AGCATGCTCGATTA-----AGAGT-----
 280 290

35
 1330 1340 1350 1360 1370 1380
 ./wwwt AGGATCTCTTGGAATGACAAATGATAGGTACCTAAAATGTAAACATGAAAATACTAGCTTA
 : : : : : : : : : : : : : : : : : :
 D8 -----ATGCCAAA-----AAATA-----
 300 310

40
 1390 1400 1410 1420 1430 1440
 ./wwwt TTTTCTGAAATGTACTATCTTAATGCTTAAATTATATTTCCCTTTAGGCTGTGATAGTTT
 : : : : : : : : : : : : : : : : : :
 D8 -----GAAAAG--CTAT-----AA--ATATTTTC-----
 320 330

45
 1450 1460 1470 1480 1490 1500
 ./wwwt TTGAAATAAAATTTAACATTTAATATCATGAAATGKTATAAGTAGACATAAAAAAAAAAAAA
 : : : : : : : : : : : : : : : : : :
 D8 -----AAAATAAAGA
 340

50
 1510 1520
 ./wwwt AAAAAAAAAAGGGCGGCCGCTAGACTAG

55
 : : : : : : : : : : : : : : : : : :
 D8 AGAAT-----CCACATTGC

60
 65

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 9 798 nt frente a D1 rev 465 nt

5 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 40,2%; puntuación de alineación global: -879

```

10      10      20      30      40      50      60
./wwwt ATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTCGCGATGGTAGCGGCGGCT
      : : : : : : : : : : : : : : : : : : : : : :
D1      --CAGTGTTC--CACAGTGTGGGCGAGCC--GTCCTTCT---AATGAAGACAATGAT
15      10      20      30      40

      70      80      90      100      110      120
./wwwt CTCGGCGGCCACCCTCTGCTGGGAGTGAGCGCCACCTTGAACCTCGGTTCTCAATTCCAAC
      : : : : : : : : : : : : : : : : : : : : : :
D1      ATTGACACTGTCCCTCTT-TGGCAGT-----TGCATTAG---TAACTTTGAAA
20      50      60      70      80      90

      130      140      150      160      170      180
./wwwt GCTATCAAGAACCTGCCCCACCGCTGGCGGCGCTGCGGGGCACCCAGGCTCTGCAGTC
      : : : : : : : : : : : : : : : : : : : : : :
D1      GGTAT-ATGA-----CTGAGCGTA-----GCATACAGGTT-----
25      100      110      120

      190      200      210      220      230      240
./wwwt AGCGCCGCGCGCGGAATCCTGTACCCGGGCGGGAATAAGTACCAGACCATTGACAACACTAC
      : : : : : : : : : : : : : : : : : : : : : :
D1      -----AACCTGCA-----GAAACAGTAC-----TT-----
30      130      140

      250      260      270      280      290      300
./wwwt CAGCCGTACCCGTGCGCAGAGGACGAGGAGTCCGGCACTGATGAGTACTGCGCTAGTCCC
      : : : : : : : : : : : : : : : : : : : : : :
D1      -AG--GTAATTGT-----AGGGCGAGGATT-----ATAAAT-----GAAAT'T---
40      150      160      170

      310      320      330      340      350
./wwwt ACCCGCGGAGGGGACGCAGGCGTGCAAAATCTGTCTCGCCT-GCAGGAAGCGCCGAAAACG
      : : : : : : : : : : : : : : : : : : : : : :
D1      -----TGCAAAA-----TCACTTAGCAGCAA---CTGAAGNCA
45      180      190      200

      360      370      380      390      400      410
./wwwt CTGCATGCGTCACGCTATGTGCTGCCCCGGAATTACTGCAAAAATGGAATATGCGTGTC
      : : : : : : : : : : : : : : : : : : : : : :
D1      AT-----TAT---CAACCACG-----TGGAGAAA-----
50      210      220

      420      430      440      450      460      470
./wwwt TTCTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAA
      : : : : : : : : : : : : : : : : : : : : : :
D1      -----ATCAAA-----CCGAGCAGGGCTGTGTGAAAC-----ATGGTT--GTAA
60

```

65

ES 2 285 768 T3

```

      230      240      250      260
    480      490      500      510      520      530
5  ./wwwt TGATCATAGCACCTTGGATGGGTATTCCAGAAGAACCACCTTGTCTTCAAAAATGTAICA
    :
D1      T-----ATGCGACTGC--GAA---CAC--TG-----AACTCTA---
              270              280              290

    540      550      560      570      580      590
10 ./wwwt CACCAAAAGGACAAGAAGGTTCTGTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTG
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      CGCCACTCCACAAA---TGATGTTT-TCAGGTGTCATG-GACTGTTGC-CACCAT-GTA
              300              310              320              330              340

    600      610      620      630      640      650
15 ./wwwt TTGTGCTAGACACTTCTGGTCCA-AGATCTGTAAACCTGTCCTGAAAGAAGGTCAAGTGT
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      TT-----CA-----TCCAGAGTTCT-TAAA---GT--TTAAAGTTGCACATGATT
              350              360              370              380

    660      670      680      690      700      710
20 ./wwwt GTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAAATATTCCAGCGTTGTTACTGTG
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      GTAT-AAGCAT-----GCTTCTTTGAGTTTAA--TTA-TGTA
              390              400              410

    720      730      740      750      760      770
25 ./wwwt GAGAAGGTCTGTCTTGCCGGATACAGAAAGATCACCATCAAGCCAGTAATTCTTCTAGGC
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      TA-AACATAAG--TTGC--ATTAGAAA-----TCAAGC--ATAAATCA-----C
              420              430              440              450

    780      790
30 ./wwwt TTCACACTTGTCAGAGACAC
    : : : : : : :
D1      TTCA-ACTGCT-----
              460

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 9 798 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 34,1%; puntuación de alineación global: -1153

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTGCGGATGGTAGCGGCGGCT
      :      :
5 D2    AA-----CTC-----

      70      80      90      100     110     120
./wwwt CTCGGCGGCCACCCCTCTGCTGGGAGTGAGCGCCACCTTGAACTCGGTTCTCAATTCCAAC
      : : : : : : : : : : : : : : : : : : : : : :
10 D2    -----CCCTGTGATTGCAGTAAA-----TT--ACTGTATTGTAAATTC-----
      10      20      30      40

      130     140     150     160     170     180
./wwwt GCTATCAAGAACCTGCCCCACCGCTGGGCGGCGCTGCGGGGCACCCAGGCTCTGCAGTC
      : : : : : : : : : : : : : : : : : : : : : :
15 D2    -----TCA-----GTGTGGCACTTA-----
      50

      190     200     210     220     230     240
./wwwt AGCGCCGCGCCGGGAATCCTGTACCCGGGCGGGAATAAGTACCAGACCATTGACAACCTAC
      : : : : : : : : : : : : : : : : : : : : : :
25 D2    -----CCTGTA-----AATG-----CAATGA-AACTTT
      60      70

      250     260     270     280     290     300
./wwwt CAGCCGTACCCCTGCGCAGAGGACGAGGAGTGCGGCACTGATGAGTACTGCGCTAGTCCC
      : : : : : : : : : : : : : : : : : : : : : :
30 D2    TA-----ATTATTTTCTAAAG-----GTGCTGCACTGCCTATTTTCTCTTGTTA-
      80      90      100     110     120

      310     320     330     340     350     360
./wwwt ACCCGCGGAGGGGACGCGAGGCGTGCAAAATCTGTCTCGCCTGCAGGAAGCGCCGAAAACGC
      : : : : : : : : : : : : : : : : : : : : : :
35 D2    -----TGTAATTTTGT-----ACACAT
      130      140

      370     380     390     400     410
./wwwt TGCATGCGTCACGCTATGTGCTGCCCCGGGAATTACTG-CAAAAATGGAATATGCGTGTC
      : : : : : : : : : : : : : : : : : : : : : :
45 D2    TGATTGT-----TATCT-----TGACTGACAAATATTCTATAT-----TG
      150      160     170

      420     430     440     450     460     470
./wwwt TTCTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAA
      : : : : : : : : : : : : : : : : : : : : : :
50 D2    AACTGAAGTAAATCATTTTCTAG-----CTTATAG--TTCTTAA

```

ES 2 285 768 T3

```

180      190      200                                210
      480      490      500      510      520      530
5  ./wwwt TGATCATAGCACCTTGGATGGGTATTCAGAGAACCACCTTGTCTTCAAAAATGTATCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  A-AGCATAACCCCTT-----ACCCCATTTAATTCTAGAGTCTAGAA
      220                                230      240      250

      540      550      560      570      580      590
10 ./wwwt CACCAAAGGACAAGAAGGTTCTGTTTGTCTCCGGTCATCAGACTG-TGCCTCAGGATTGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  CGC-----AAGGATCTCTTGAAT---GACAAATGATAGGTACCTAA--AATGT
      260      270      280      290

      600      610      620      630      640      650
15 ./wwwt GTTGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCTGAAAGAAGGTCAAGTGT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  AACATG---AAAATACT-----AGCT---TA---TTTTCTGAAA-----T
      300      310                                320

      660      670      680      690      700      710
20 ./wwwt GTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAAATATTCCAGCGTTGTTACTGTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  GTACTAT-CTTA-----ATGCTTAAATT-----ATATTTC--CCTTTAGGCTGTG
      330      340      350      360      370

      720      730      740      750      760      770
25 ./wwwt GAGAAGGTCTGTCTTGCCGGATACAGAAAGATCACCATCAAGCCAGTAATTCTTCTAGGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  -- ATAGTTT---TTG-----AAATAAAATTTAACATT-----TAATA-----
      380      390      400

      780      790
30 ./wwwt TTCACACTTGTCAGAGACAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D2  -TCAAA-----AAAAAAA
      410

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 9 798 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 38,0%; puntuación de alineación global: -987

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTCCGATGGTAGCGGGCGGCT
      :   : :   : :   : :   : :   : :   : :   : :   : :
5 D3    TTN--GGAT---GGAGTTGATGAAG-----GTCCATAA---AAGCTTT-----T
      10      20      30

      70      80      90      100     110     120
./wwwt CTCGGCGGCCACCCCTCTGCTGGGAGTGAGCGCCACCTTGAACTCGGTTCTCAATTCCAAC
      :   : :   : :   : :   : :   : :   : :   : :   : :
10 D3   CATAGCATCTATTATATAAAAGGAGTT---CCAC--TGCATTGGAT-----
      40      50      60      70

      130     140     150     160     170     180
./wwwt GCTATCAAGAACCTGCCCCACCGCTGGGCGGCGCTGCGGGCACCAGGCTCTGCAGTC
      :   : :   : :   : :   : :   : :   : :   : :   : :
15 D3   -----AAGCTG-----GTTTAGTGCTCTCTGA-----CAAG-TGTGAAG--
      80      90      100     110

      190     200     210     220     230     240
./wwwt AGCGCCGCGCCGGGAATCCTGTACCCGGCGGGAATAAGTACCAGACCATTTGACAACCTAC
      :   : :   : :   : :   : :   : :   : :   : :   : :
20 D3   -----CCTAGAAGAAT-TACT--GGCTTGAT--GGT----GATCTTT-----
      120     130     140

      250     260     270     280     290     300
./wwwt CAGCCGTACCCGTGCGCAGAGGACGAGGAGTGCGGCACCTGATGAGTACTGCGCTAGTCCC
      :   : :   : :   : :   : :   : :   : :   : :   : :
25 D3   ---CTGTATCCG-GC---AAGAC-AGACCTTCTCCAC-----AGTAAC
      150     160     170

      310     320     330     340     350     360
./wwwt ACCCGCGGAGGGGACGCGAGGCGTGCAAACTCTGTCTCGCCTGCAGGAAGCGCCGAAAACGC
      :   : :   : :   : :   : :   : :   : :   : :   : :
30 D3   A-----ACGCTGG-----AATATTTCTAG--TCCATGA-GAGCC-----
      180     190     200

      370     380     390     400     410     420
./wwwt TGCATGCGTCAAGCTATGTGCTGCCCGGGAATTACTGCAAAAATGGAATATGCGTGTCT
      :   : :   : :   : :   : :   : :   : :   : :   : :
35 D3   -----TTTTCTCC-----TA-TGCT---TGGTACACACTTGACC
      210     220     230

      430     440     450     460     470     480
./wwwt TCTGATCAAAATCATTTCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAAT
      :   : :   : :   : :   : :   : :   : :   : :   : :
40 D3   T---TC-----TTTC--AGGACAGGTT-----TACA

```

ES 2 285 768 T3

```

240                                250                                260

          490          500          510          520          530          540
5  ./wwwt GATCATAGCACCTTGGATGGGTATTCCAGAAGAACCACCTTGTCTTCAAAAATGTATCAC
      : : : :          : : : :          : : : : :          : : : :          : : :
D3      GATC-----TTGGA-----CCAGAAG-----TGTCT-----AGCAC
          270                                280

          550          560          570          580          590
10 ./wwwt ACCAAAGGACAAGAAGGTTCTGTTG-TCTCCGGTCATCAGACTGTGCCTCAGGATTGTG
      : : : : :          : : : : : : : : : : : : : : : : : : : : : : : :
D3      AACACA--ATCCTGAGGCACAGTCTGATGACCGGAGA-CAAACAGAACCTTCTTGTCTCCTT
          290          300          310          320          330          340

          600          610          620          630          640          650
15 ./wwwt TTGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAGAAGGTCAAGTGTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      TGGTGTGATACATTTTTG---AAGAC---AAGGTGGTTCTTCTGGAA-----
          350          360                                370          380

          660          670          680          690          700          710
20 ./wwwt TACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAAATATTCCAGCGTTGTTACTGTGG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      TACCCATCCAAGG-----TGCTATGATC-ATTACCAA-----AGC-----TTTCAGT--
          390          400          410                                420

          720          730          740          750          760          770
25 ./wwwt AGAAGGTCTGTCTTGCCGGATACAGAAAGATCACCATCAAGCCAGTAATTCTTCTAGGCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      -GATGGTTT-----CCTCAATTCTCC-----TCGGAAATGATT-----T
          430                                440                                450

          780          790
30 ./wwwt TCACACTTGTTCAGAGACAC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D3      TGA-----TCAGA---AG
          460

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 9 798 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 34,3%; puntuación de alineación global: -1204

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTCCGATGGTAGCGGCGGCT
      :      : : :      : : :      : : : : :
5 D4      CT-----TGC-----TTGTT-----TGGTACAGACTTC-
                        10      20

      70      80      90      100      110      120
./wwwt CTCGGCGGCCACCCCTCTGCTGGGAGTGAGCGCCACCTTGAACTCGGTTCTCAATTCCAAC
      : : : : : : : : : : : : : : :
10 D4      -----CCCCTGATGG-----AGC-----ACT-GGTT-----
                        30      40

      130      140      150      160      170      180
./wwwt GCTATCAAGAACCTGCCCCACCGCTGGGCGGCGCTGCCGGGCACCCAGGCTCTGCAGTC
      : : : : : : : : : : : : : : :
15 D4      -----TGCAG--

      190      200      210      220      230      240
./wwwt AGCGCCGCGCCGGGAATCCTGTACCCGGGCGGAATAAGTACCAGACCATTGACAACACTAC
      : : : : : : : : : : : : : : :
20 D4      -----ATTTTG-----GT-CCAGA--AATGACGAGCAC
                        50      60      70

      250      260      270      280      290      300
./wwwt CAGCCGTACCCGTGCGCAGAGGACGAGGAGTGCGGCACTGATGAGTACTGCGCTAGTCCC
      : : : : : : : : : : : : : : :
30 D4      -AGCAAAACCCTTC-----AATGCAGT-CTGATGAT-----CG-TAGGC--
                        80      90      100      110

      310      320      330      340      350      360
./wwwt ACCCGCGGAGGGGACGCGAGGCGTGCAAATCTGTCTCGCCTGCAGGAAGCGCCGAAACGC
      : : : : : : : : : : : : : : :
35 D4      -----AGGGGTCTCCTTCATGCCCTTTTATAT-----GTGAC-ATCTTAG
                        120      130      140

      370      380      390      400      410      420
./wwwt TGCATGCGTCACGCTATGTGCTGCCCCGGGAATTACTGCAAAAATGGAATATGCGTGTCT
      : : : : : : : : : : : : : : :
40 D4      TGTGTG-GTCTTCCTAGATTCTGCCATCCCAAGTCATGGTTGA-GTAATGACCGTGGTT
      150      160      170      180      190      200

      430      440      450      460      470
./wwwt TCTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCAC-TGAAAGCTTTGGTAA
      : : : : : : : : : : : : : : :
50 D4      TC-GATC---TCTGTGCCGAGTACCA-TCCAG--AGCCGGGATGTGAGGGGTTAAG--A

```

ES 2 285 768 T3

```

      210      220      230      240      250
5      480      490      500      510      520      530
./wwwt TGATCATAGCACCTTGGATGGGTATTCCAGAAGAACCACCTTGTCTTCAAAAATGTATCA
      :: :   :: :   :: :   :: :   :: :   :: :
D4      TGCTTTCAGTAACTGGGATA-----CAGA-----TGCCAT-----TATTG
      260      270      280      290

      540      550      560      570      580      590
10     ./wwwt CACCAAAGGACAAGAAGGTTCTGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTG
      :: :   :: :   :: :   :: :   :: :   :: :
D4      CAGCG-----GGTACTG-----GGGCAGCACA--TGCC--ATCTCG
      300      310      320

      600      610      620      630      640      650
15     ./wwwt TTGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAGAAGGTCAAGTGTG
      :: :   :: :   :: :   :: :   :: :   :: :
D4      --GTGGCAG-CGCTTCTTTT-----TCTCC-GACACA-----
      330      340      350

      660      670      680      690      700      710
20     ./wwwt TACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAAATATTCCAGCGTTGTTACTGTGG
      :: :   :: :   :: :   :: :   :: :   :: :
D4      --CCATGCA-----GGC-----CGATGATCCT-TGG
      360      370

      720      730      740      750      760      770
25     ./wwwt AGAAGGTCTGTCTTGCCGGATACAGAAAGATCACCATCAAGCCAGTAATTCTTCTAGGCT
      :   :: :   :   :   :   :   :   :   :   :   :   :
D4      TGG-GGACTGTG-----GCAATACCTCC--CAA-----CNGCTGTGCC
      380      390      400      410

      780      790
30     ./wwwt TCACACTTGTCAGAGACAC
      :
D4      -----GAANT

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 9 798 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 31,8%; puntuación de alineación global: -1309

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt ATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTTCGGATGGTAGCGGGCGGCT
5  D8  -----

      70      80      90      100     110     120
./wwwt CTCGGCGGGCCACCTCTGCTGGGAGTGAGCGCCACCTTGAATCGGTTCTCAATTCCAAC
10 D8  -----CAGGCGGGAGT-----CCAGGAA-----AA-----CCAAC
      : : : : :      : :      : :      : :      : :      : :
      10      20

      130     140     150     160     170     180
./wwwt GCTATCAAGAACCTGCCCCACCGCTGGGCGGCGCTGCGGGGCACCCAGGCTCTGCAGTC
15 D8  CCAA--AAG-----
      : :      : :

      190     200     210     220     230     240
./wwwt AGCGCCGCGCCGGGAATCCTGTACCCGGGCGGGAATAAGTACCAGACCATTGACAACCTAC
20 D8  -----GAAGCC-----AAGTA-----TTAAGAAATCA
      : :      : :      : : : :      : :      : :
      30      40      50

      250     260     270     280     290     300
./wwwt CAGCCGTACCCGTGCGCAGAGGACGAGGAGTGCGGCACTGATGAGTACTGCGCTAGTCCC
30 D8  CA-----AGG--CAGGAA--GGGAC--AAGAG-----
      : :      : :      : :      : :      : :
      60      70

      310     320     330     340     350     360
./wwwt ACCCGCGGAGGGGACGCGAGGCGTGCAAATCTGTCTCGCCTGCAGGAAGCGCCGAAAACGC
35 D8  -----GGAG-----AAAGTTGTCT-----GAGAACTT
      : : : :      : :      : : : :      : :      : :
      80      90

      370     380     390     400     410     420
./wwwt TGCATGCGTCACGCTATGTGCTGCCCCGGGAATTACTGCAAAAATGGAATATGCGTGTCT
40 D8  T-----TGAC-----TGTG--GCCCTGGACTTTGCTG-----TGCTCGTCA
      : :      : :      : : : :      : :      : :
      100     110     120

      430     440     450     460     470     480
./wwwt TCTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAAT
50 D8  T-----TTTGGACGAAAATTTGT--AAGCCAGTCCT-----TTTGG---
      : :      : :      : :      : :      : :      : :

```

ES 2 285 768 T3

```

130      140      150      160
      490      500      510      520      530
5  ./wwwt GATCATAGCACCTTGGATGGGTAT--TCCAGAAGAACCACCTTGTCTTCAAAAATGTATC
      ::      ::      ::      ::      ::      ::      ::      ::      ::
D8  -----AG-----GGACAGGTCTGTCTCCAGAAGAG-----GGCATAAAGA-----C
      170      180      190

      540      550      560      570      580      590
10 ./wwwt ACACCA-AAGGACAAGAAGGTTCTGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTG
      ::      ::      ::      ::      ::      ::      ::      ::      ::
D8  ACTGCTCAAGCTCCAGAAA--TCT-----TCCAGCGTTGCGACTGTGGCCCTGGACTA
      200      210      220      230      240

      600      610      620      630      640      650
15 ./wwwt TGTGTGTGCTAGACACTTCTGGTCCAAGATCTGTAAACCTGTCCTGAAAGAAGGTCAAGTG
      ::      ::      ::      ::      ::      ::      ::      ::      ::
D8  CTGTGTCTGAAGCCAATT--GACCAGCAATCGGCAGCATGCTC-----GATTAAGAG
      250      260      270      280      290

      660      670      680      690      700      710
20 ./wwwt TGTACCAAGCATAGGAGAAAAGGCTCTCATGGACTAGAAATATTCCAGCGTTGTTACTGT
      ::      ::      ::      ::      ::      ::      ::      ::      ::
D8  TATGCCAA--AAAATAGAAAAG-----CTATAAATATTTC-----
      300      310      320      330

      720      730      740      750      760      770
30 ./wwwt GGAGAAGGTCTGTCTTGCCGGATACAGAAAGATCACCATCAAGCCAGTAATTCTTCTAGG
      ::      ::      ::      ::      ::      ::      ::      ::      ::
D8  ---AA-----ATAAAGAAGAATC-----
      340

      780      790
35 ./wwwt CTTACACTTGTCTCAGAGACAC
      ::      ::      ::
D8  ---CACA-TTG-----C
      350
40

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 10 702 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 42,5%; puntuación de alineación global: -581

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCACCAAGGATCATCGGCCTGC
      :          : : : : : : : : : : : : : : : : : : : : : :
5 D1    CA-----GTGTTCTCACAGT-----GTGGGGCA---GCCGTC
      10      20

      70      80      90      100     110     120
./wwwt ATGGTGTGTCTGGAGAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      : :          : : : : : : : : : : : : : : : : : : : : : :
10 D1    CTTCT-----AATGAAGA-----CAATGATAT-----TGACACTGT-CC
      30      40      50      60

      130     140     150     160     170
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTACT--GAAAGCATCTTAACCCCTCACATCC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
15 D1    CTCTT-----TGGCAGTTGCATT--AGTAACTTTGAAAGGT-----ATAT--
      70      80      90      100

      180     190     200     210     220     230
./wwwt CGGCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
20 D1    -GACT--GAGCGTA---GCATACAGGT--TAACC-----TGACAGAAAC-AGTACT
      110     120     130     140

      240     250     260     270     280     290
./wwwt TCGGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
30 D1    TAGGTA---ATTGTAGGGCGAGGA-----TTATAAATGAAATTGCAAAATCACTTAG
      150     160     170     180     190

      300     310     320     330     340     350
./wwwt GAGACCCCTGCCACGATCATCAGACTGCATTGAAGGGTTTTGCTGTGCTCGTCATTTCT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 D1    CAG-CAACTGAAGNCAATTATCA-AC-----CACGT-----
      200     210     220

      360     370     380     390     400     410
./wwwt GGACCAAATCTGCAAACCAAGTGCTCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
40 D1    GGAG-AAAATC---AAACCG-----AGCAGGG-----CTGTGTGAAACAT-----
      230     240     250

      420     430     440     450     460     470
./wwwt AGGGTTCTCATGGGCTGGAAATTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGCA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 D1    --GGTTGTAATA-----TGCGACTG--CGAACA--CTG-----

```

ES 2 285 768 T3

```

                260                270                280
      480      490      500      510      520      530
./wwwt AAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAATTT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      AACTCT-----ACGCCAC---TCCAC-AAATGATGTTTTTCAGGTGT--CATGGACTGT
                290                300                310                320                330

      540      550      560      570      580      590
./wwwt GATCACCATTGAGGAACATCATCAATTGC-AGACTGTGAAGTTGTGTATTTAATGCATTA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      TGCCACCATGTAT--TCATCCAGAGTTCTTAAAGTTTAAAGTTGCACATG-ATTGTATAA
                340                350                360                370                380

      600      610      620      630      640      650
./wwwt TAGCATGGTGGAAAATAAGGTTTCAGATGCAGAAGAATGGCTAAA-ATAAGAAACGTGATA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      --GCATGCTT-----TCTTTGAGTTTTA-AATTATGTATAAACATAAGTT-----
      390                400                410                420

      660      670      680      690      700
./wwwt AGAATATAGATGATCACAAAAAAAAAAAAAAAAAAGATGCGGCCGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D1      -GCATTTAGAA-ATCAAGCATAAATCACTTCAACTGCT-----
      430                440                450                460

```

30 ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 10 702 nt frente a D2 rev 416 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 35 Identidad del 40,2%; puntuación de alineación global: -780

40

45

50

55

60

65

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :: :: :      :: ::      ::      ::      :: :: :      ::
5 D2    -AACTCCCC-----TGTGA---TTG---CAGT-----AA--ATTA-----CTGT
      10      20      30
      70      80      90      100      110      120
./wwwt ATGGTGTGTCCGAGAAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :: ::      ::      :  ::      ::      ::      ::
10 D2    ATTGT-----AAATTCTCAGTG-----TGGCA---CT-----TACC
      40      50
      130      140      150      160      170      180
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTACTGAAAGCATCTTAACCCCTCACATCCC
      :      ::      ::      ::      ::      :  :
15 D2    TG-----TAAATGCAA-----TGAAAC--TTTTAA---TTATTTT---
      60      70      80
      190      200      210      220      230      240
./wwwt GCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTTG
      :: :  :: :  :: :  :: :  :      :  :  :: :  :: :
25 D2    --TCTAAA-GGTGCT--GCACTGC-----CTATTTTTCCTC-----TTG
      90      100      110      120
      250      260      270      280      290      300
./wwwt GGATGGCAGAATCTAGGAAGACCACACACTAAGATGTACATATAAAAGGGCATGAAGGA
      ::      :: :  :  :      :: :  :  :  ::      ::      :  :: :
30 D2    TTAATGTA--AATTTTGT----ACACATTGAT-TGT----TAT-----CTTGACTGA
      130      140      150      160
      310      320      330      340      350      360
./wwwt GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTGTCTGTGCTCGTCATTCTGG
      :      :  :  :  :  :      :: :  :  :  :  ::      ::      :  :: :
35 D2    CAAATATTC-TATATTGA----ACTGAAGTAAATCATTTCA---GCTTAT-AGTTCTT-
      170      180      190      200      210
      370      380      390      400      410      420
./wwwt ACCAAAATCTGCAAACAGTGCTCCATCAGGGGGAAGTCTGTACCAAACAACGCAAGAAG
      :: :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
45 D2    ---AAAAGCAT-AACCCCTTACCCCATTT-----AATTCTAGAGTCTAGAACGCAAG---
      220      230      240      250      260
      430      440      450      460      470      480
./wwwt GGTTCCTCATGGGCTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGCAAA
      :  :: :  :: :  :: :  :  :      :  :  :  :  :  :
50 D2    -GATCTCTTGAATGACAAAT-----GATAGGTACCT-----AAA

```

ES 2 285 768 T3

```

      270      280      290
      490      500      510      520      530      540
5  ./wwwt GTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAATTTGA
      : :  : : : :      : : : : : : : : : : : :
D2  ATGT---AACATG-----AAA---ATACTAGCTTATTTTCTGAAATGT---
      300      310      320      330

      550      560      570      580      590      600
10 ./wwwt TCACCATTGAGGAACATCATCAATTGCAGACTGTGAAGTTGTGTATTTAATGCATTATAG
      : : : : : : : : : : : : : : : : : :
D2  -----ACTATCTT--AATGC-----TTAAATTA--TATT-----CCCTTTAG
      340      350      360

      610      620      630      640      650      660
15 ./wwwt CATGGTGGAAAATAAGGTTTCAGATGCAGAAGAATGGCTAAAATAAGAAACGTGATAAGAA
      : : : : : : : : : : : : : : : : : :
D2  GCTG-TG---ATA-GTTTTTGA-----AAT---AAAAT-----TTAACA
      370      380      390

      670      680      690      700
20 ./wwwt TATAGATGATCACAAAAAAAAAAAAAAAAAAGATGCGGCCGC
      : : : : : : : : : : : : : : : : : :
D2  TTTA----ATATCAAAAAAAAAAA-----
      400      410

```

30 ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 10 702 nt frente a D3 rev 469 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 35 Identidad del 41,9%; puntuación de alineación global: -598

40

45

50

55

60

65

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
5 D3    ---TTNGGATGGAGTTGATGAAGGT---CCATAA-----AAGGTT-----TTC
      10      20      30

      70      80      90      100     110     120
./wwwt ATGGTGTGTCTCGGAGAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
10 D3    ATAGCATCTATTATATAAAAGGAC---TTCCACTGCATTTGGATAAGCTGGTTTGTAGTGC
      40      50      60      70      80      90

      130     140     150     160     170     180
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTACTGAAAGCATCTTAACCCCTCACATCCCG
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
15 D3    T-CTG--ACAAG---TGTGAAGCCTAG-----AAGAAT-----TACTG
      100     110     120

      190     200     210     220     230     240
./wwwt GCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCA/TGACTTG
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
25 D3    GCT--TGATGGT-----GATC-----TTTCTGTATCC-----G
      130     140     150

      250     260     270     280     290
./wwwt GGATGGCAGAATCTAGGAAGAC/CACAC--ACTAAGATGTCACATATAAAAGGGCATGAAG
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
30 D3    GCAAGACAGACCTTCT-----CCACAGTAACAACGCTGGAATATTTCTA-GTCCATGA--
      160     170     180     190     200

      300     310     320     330     340     350
./wwwt GAGACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGCTGTGCTCGTCATTCT
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
35 D3    --GAGCCTTTTCTCCTAT-----GCTTGGTACACACTTGA-----CCTTCTTTCA
      210     220     230     240

      360     370     380     390     400     410
./wwwt GGACCAAAATCTGCAAACAGTGCTCCATCAGGGGGAAGTCTGTACCAACAACGCAAGA
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
45 D3    GGAC--AGGTTTACAGATCT-TGGACCA-----GAAGTGTCTAGC--ACAACACAA--
      250     260     270     280     290

      420     430     440     450     460     470
./wwwt AGGGTTTCTCATGGGCTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGCA
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
50 D3    -----TCCTGAG--GCA-----CAGTCTGATGACCG-GAGA-----CA

```

ES 2 285 768 T3

```

      300      310      320
    480      490      500      510      520      530
./wwwt AAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAATTT
5      ::      :::      :::: : : : : :      :::: : : : :
D3      AAC-----AGA-----ACCTTCTTGTCCCTTTG-----GTGTGATACATT-TTT
      330      340      350      360

    540      550      560      570      580      590
./wwwt GATCACCATTGAGGAACATCATCAATTGCAGACTGTGAAGTTGTGTATTTAATGCATTAT
10      :: : : : : : : : : : : : : : : : : : : : : : :
D3      GAAGACAAG-GTGGTTCTTCTGGAAT--ACCCATCCAAGGTGC---TATGAT-CATTA-
      370      380      390      400      410

    600      610      620      630      640      650
./wwwt AGCATGGTGGAAAAATAAGGTTTCAGATGCAGAAGAATGGCTAAAAATAAGAAACGTGATAAG
15      ::      : : : : : : : : : : : : : : : :
D3      -CCA-----AAGCT---TTCAGT---GATGGTTTCCTCAATTTCTCCTCG-GAAATG
20      420      430      440      450

    660      670      680      690      700
./wwwt AATATAGATGATCACAAAAAAAAAAAAAAAAAAGATGCGGCCGC
25      : : : : : : : :
D3      ATTTT---GATCAGAAG-----
      460

```

30 ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 10 702 nt frente a D4 rev 416 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 35 Identidad del 37,7%; puntuación de alineación global: -817

40

45

50

55

60

65

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
5 D4      -----CTTGCGTTGTTTGGTA-----CAGACTTCCC-----CCTG-
      10      20      30

      70      80      90      100      110      120
./wwwt ATGGTGTGTGTCGGAGAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
10 D4      ATGG-----AGCACTG-----GTTTGCAGATTTTGGTCCAGAA--
      40      50      60

      130      140      150      160      170      180
./wwwt CGCTGCAATAATGGCATCTGTATCCCACTTACTGAAAGCATCTTAACCCCTCACATCCCCG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
15 D4      -----ATGAC-----GAGCACAGCAAAACCCTTCA-ATGCAG
      70      80      90

      190      200      210      220      230      240
./wwwt GCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTTG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
20 D4      TCTGATGATCGTA---GGC---AGGGGTC-----TCCTT-----CATGCCCTT
      100      110      120      130

      250      260      270      280      290      300
./wwwt GGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAGGA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
25 D4      TTAT-----ATGTGACATCTT--AGTGTGTGGT---
      140      150

      310      320      330      340      350      360
./wwwt GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGCTGTGCTCGTCATTTCTGG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
30 D4      -----CTTCCTAGATTCTGCCATCCCAAGTCATGGTTTGAGTAATGACCGTGGTTTCG--
      160      170      180      190      200      210

      370      380      390      400      410      420
./wwwt ACCAAAATCTGCAAAACAGTGCTCCATCAGGGGGAAGTCTGTACCAACAACGCAAGAAG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
35 D4      -----ATCT-----CTGTGCC-----GA-----GTACCATCCAG-----AGCCG
      220      230

      430      440      450      460      470      480
./wwwt GGTTCTCATGGGCTGGAAATTTCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGCAAA
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
40 D4      GGATGTGAGGGGTT---AAGAT---GCTTT---CAGT---AA---CTG-----

```

ES 2 285 768 T3

```

      240      250      260      270
      490      500      510      520      530      540
5  ./wwwt GTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAATTGGA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  GGAT--ACAGATGCCAT-TATTGC-----AGCGGG--TAC---TGGGGCAGCACATGCCA
      280      290      300      310

      550      560      570      580      590      600
10 ./wwwt TCACCATTGAGGAACATCATCAATTGCAGACTGTGAAGTTGTGTATTTAATGCATTATAG
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  TC-----TCGGTGGCAGC-----GCT-TCTTTTTTCTCCGACACAC
      320      330      340      350

      610      620      630      640      650      660
15 ./wwwt CATGGTGGAAAATAAGGTTTCAGATGCAGAAGAATGGCTAAAATAAGAAACGTGATAAGAA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  CATGCAGGCCGATGATCCTTGGTGG--GGACTGTGGC---AATA-----CCTCCCAAC--
      360      370      380      390      400

      670      680      690      700
20 ./wwwt TATAGATGATCACAAAAAAAAAAAAAAAAAAGATGCGGCCGC
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D4  ---NGCTG-----TGCCGAANT
      410

```

ALIGN calcula una alineación global de dos secuencias
 Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17
 Seq 10 702 nt frente a D8 rev 355 nt
 Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4
 Identidad del 36,3%; puntuación de alineación global: -736

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :          : : : : : : : : : : : : : : : :
5 D8     CA-----CCCAGTCC-----AGGAAAACCAACCC--
                        10      20

      70      80      90      100     110     120
./wwwt ATGGTGTGTCTGGAGAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      : : : : : : : : : : : : : : : : : : : : : :
10 D8     -----AAAAGGAAGC-----CAAGTA--
                        30      40

      130     140     150     160     170     180
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTACTGAAAGCATCTTAACCCCTCACATCCCG
      : : : : : : : : : : : : : : : : : : : : : :
15 D8     -----TTAA-----

      190     200     210     220     230     240
./wwwt GCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTTG
      : : : : : : : : : : : : : : : : : : : : : :
20 D8     -----GAAATCAC-----
                        50

      250     260     270     280     290     300
./wwwt GGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAGGA
      : : : : : : : : : : : : : : : : : : : : : :
30 D8     --AAGGCAG-----GAAG-----GGACAAGAGGGA
                        60      70

      310     320     330     340     350     360
./wwwt GACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTGTCTGTGCTCGTCATTTCTGG
      : : : : : : : : : : : : : : : : : : : : : :
35 D8     GAAAGTTGTCTGAGAACTTTTGACTGTGGCCCTGGACTTTGCTGTGCTCGTCATTTTGG
      80      90      100     110     120     130

      370     380     390     400     410
./wwwt ACCAAAATCTGCAAACAGTGTCTCCATCAGGGGGAAGTCTGTACCA-AACAACGCAAGAA
      : : : : : : : : : : : : : : : : : : : : : :
45 D8     ACGAAAATTTGTAAGCCAGTCCCTTTTGGAGGGACAGGTCTGCTCCAGAAGAGGGCATAAA
      140     150     160     170     180     190

      420     430     440     450     460     470
./wwwt GGG--TTCTCATGGGCTGGAAATT'TCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC
      : : : : : : : : : : : : : : : : : : : : : :
50 D8     GACACTGCTCAAGCTCCAGAAATCTTCCAGCGTTGCGACTGT-----GGCCCTG-----

```

ES 2 285 768 T3

```

      200      210      220      230      240
      480      490      500      510      520      530
5  ./wwwt AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAATT
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  -----GACTAC-----TGTGTC-----
      250

10      540      550      560      570      580      590
      540      550      560      570      580      590
./wwwt TGATCACCATTGAGGAACATCATCAATTGCAGACTGTGAAGTTGTGTATTTAATGCATTA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  --GAAGCCAATTGAC-----CAGCAATCG-----GCA-----
      260      270      280

15      600      610      620      630      640      650
      600      610      620      630      640      650
./wwwt TAGCATGGTGGAAAATAAGGTTTCAGATGCAGAAGAATGGCTAAAATAAGAAACGTGATAA
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  --GCATGCTCGA--TTAAG-----AGTATGCCAAAAATAGAAAAGCTATAA
      290      300      310      320

20      660      670      680      690      700
      660      670      680      690      700
./wwwt GAATATAGATGATCACAAAAAAAAAAAAAAAAAAGATGCGGCCGC
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D8  ATATTT-----CAAAATAAAGAAGAATCCACATT-----GC
      330      340      350

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 12 537 nt frente a D1 rev 465 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 45,4%; puntuación de alineación global: -260

ES 2 285 768 T3

```

      10      20      30      40      50
./wwwt GAAT-TCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTG
      : : : : : : : : : : : : : : : : : : : : : :
5 D1 CAGTGTTCAC-AGTGTGGGGCAG-----CCGTCCTT-CTAATGAAGA-----C
      10      20      30      40

      60      70      80      90      100      110
./wwwt CATGGTGTGTCGGAGAGAAAAAGAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTAC
      : : : : : : : : : : : : : : : : : : : : : :
10 D1 AATGATATT-----GACA-----CTGTC-CCTCTTTGGCA--GTTCATTAGTAA
      50      60      70      80

      120      130      140      150      160      170
./wwwt CCGCTGCAATAATGGCATCTGTATCCAGTTACTGAAAGCATCTTAACCCCTCACATCCC
      : : : : : : : : : : : : : : : : : : : : : :
15 D1 CTT-TG---AAAGGTATATGACTGAGCGTAGCATAAGG---TTAACCTG-CAGAAACA
      90      100      110      120      130

      180      190      200      210      220      230
./wwwt GGCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTT
      : : : : : : : : : : : : : : : : : : : : : :
20 D1 GTACTTAGGTAATTGTAGGGCGAGGATTATAAATGA----AATTGCAAA--ATCACTT
      140      150      160      170      180

      240      250      260      270      280      290
./wwwt GGGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAGGGCATGAAGG
      : : : : : : : : : : : : : : : : : : : : : :
25 D1 AG-----CAGCAACT--GAAGNCAATTATCAACCACGT---GGAGAAAATCAAACCGAGC
      190      200      210      220      230

      300      310      320      330      340      350
./wwwt AGACCCCTGCCCTACGATCATCAGACTGCATTGAAGGGTTTGTCTGTGCTCGTCATTTCTG
      : : : : : : : : : : : : : : : : : : : : : :
35 D1 AGGGCTGTGTGAAACATGGT-----TGTAAT--ATGCGACTGC-GAACAC-TGAACCTTA
      240      250      260      270      280      290

      360      370      380      390      400      410
./wwwt GACCAAAATCTGCAAACCAAGTCTCCATCAGGGGGAA--GTCTGTACCAACA--ACGC
      : : : : : : : : : : : : : : : : : : : : : :
40 D1 CGCCAC--TCCACAAATGA-TGTT--TTCAGGTGTCATGGACTGTTGCCACCATGTATTC
      300      310      320      330      340

      420      430      440      450      460      470
./wwwt AAGAAGGGTTCTCATGGGCTGGAAATTTTC-CAGCGTTGCGACTGTGCGAAGGGCCTGTC
      : : : : : : : : : : : : : : : : : : : : : :
45 D1 ATCCAGAGTTCTTAAAG--TTTAAAGTTGCACATGATTGTATAAGCATGCTTTCTTTGAG
      420      430      440      450      460      470

      350      360      370      380      390      400

      480      490      500      510      520      530
./wwwt TTGCAAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAA
      : : : : : : : : : : : : : : : : : : : : : :
55 D1 TTTTAAATTATGTATAAACATAAGTTGCATTTAGAAATCAAGCATAAATCACT-TCA-AC
      410      420      430      440      450      460

./wwwt AATT
      :
D1 TGCT

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 12 537 nt frente a D2 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 42,8%; puntuación de alineación global: -429

```

10      10      20      30      40      50      60
./wwwt GAATTCGGCAGGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :: :: :      :: :      :::: :      ::      :: :: :      ::
D2      -AACTCCCC-----TGTG---ATTGC---AGT-----AA--ATTA-----CTGT
15      10      20      30

      70      80      90      100      110      120
./wwwt ATGGTGTGTGCGGAGAAAAAGAAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :: ::      :: :      : : :      :::: :      : :      : ::
D2      ATTGT-----AAATTCTCAGTG-----TGGCACTTAC-----CTGTA--
20      40      50      60

      130      140      150      160      170      180
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTACTGAAAGCATCTTAACCCCTCACATCCCG
      ::::: :      : : : :      : : :      :::: :      : : : :      : : : :
D2      -AATGCAAT---GAAACTTTTAATTATTTTCTAAAGGTGCTGCACTGCCATTATTTTC-
25      70      80      90      100      110

      190      200      210      220      230      240
./wwwt GCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTTG
      :::: : : : :      : : : :      : : :      :::: : : : :      : : : ::
D2      -CTCT---TGTTA---TGTAATTTTGTGTA---CACATTGATTGTT---ATCTTGACTGA
30      120      130      140      150      160

      250      260      270      280      290
./wwwt GGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAGG-
      : : : : :      : : : : :      : : : : :      : : : : :      : : : :
D2      CAA---ATATTCTATATTGAACTGA-AGTAAATCATTTAGCTTATAGTTCTTAAAGC
40      170      180      190      200      210

      300      310      320      330      340      350
./wwwt AGACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGCTGTGCTCGTCATTTCTG
      : : : : :      : : : : :      : : : : :      : : : : :      : : : :
D2      ATAACCCT-----TTACCCCATTTAATTCTAGAGTCTAGAAC-GCAAGG-ATCTCTT
45      220      230      240      250      260

      360      370      380      390      400      410
./wwwt GACCAAAATCTGCAAACCAGTGCTCCATCAGGGGGAAGTCTGTACCAA-ACAACGCAAGA
      :      :: :      :::: :      : : : : :      : :      :::: : : : : : :
D2      GG----AATGA-CAAATGATAGGTACCTAA-----AA---TGTAACATGAAAATACTAG-
50      270      280      290      300      310

      420      430      440      450      460      470
./wwwt AGGGTTCTCATGGGCTGGAAATTTTCCAGCGT--TGCGACTGTGCGAAGGGCCTGT---C
      : : :      : : : : :      : : : : :      : : :      : : : : :
D2      -----CTTATTTTCTG-AAATGTACTATCTTAATGCTTAAATTATATTTCCCTTTAGGC
60

```

ES 2 285 768 T3

```

          320          330          340          350          360
      480      490      500      510      520      530
5  ./wwwt TTGCAAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAA
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
D2  TGTGATAGTTTTTGAA-ATAAAATTAA-CATTTAATATCAAA-----AAAA
      370      380      390      400      410

10

      ./wwwt AATT
      :
D2  AAAA

15

```

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 12 537 nt frente a D3 rev 469 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 46,5%; puntuación de alineación global: -203

ES 2 285 768 T3

```

      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :: ::      :::: : : : : : : : : : : : : : : : :
5 D3    ---TTNGGAT-----GGAGTTGATG--AAGGTCC-----ATAAAAGGTTTTC
      10      20      30

      70      80      90      100     110     120
./wwwt ATGGTGTGTCGGAGAAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :: : : : : : : : : : : : : : : : : : : : : : :
10 D3    ATAGCATCTATTATATAAAAGGAG---TTCCACTGCATTTGGATAAGCTGGTTTAGTGTC
      40      50      60      70      80      90

      130     140     150     160     170     180
./wwwt CGCTGCAATAATGGCATCTGTATCCCACTTACTGAAAGCATCTTAACCCCTCACATCCCG
      :: : : : : : : : : : : : : : : : : : : : : :
15 D3    T-CTG-ACAAGTGTGAAGCCTAGAAGAATTACTGGCTTGATGGTGATCTTTCTGTATCCG
      100     110     120     130     140     150

      190     200     210     220     230
./wwwt GCTCTGGATGGTACT-CGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGA---
      :: : : : : : : : : : : : : : : : : : : : : :
20 D3    GCAA--GACAGACCTTCTCCACAGTAAC--AACGCTGGAATATTTCT-AGTCCATGAGAG
      160     170     180     190     200

      240     250     260     270     280     290
./wwwt -CTTGGGATGGCAGAATCTAGGAAGACC---ACACACTAAGATGTCACATATAAAAGGG
      :: : : : : : : : : : : : : : : : : : : : : :
25 D3    CCTT---TTCTCCTATGCTTGGTACACACTTGACC'TTCTTTCAGGACAGGTTTACA-GAT
      210     220     230     240     250     260

      300     310     320     330     340     350
./wwwt CATGAAGGAGACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTGCTGTGCTCGT
      : : : : : : : : : : : : : : : : : : : : : :
30 D3    CTTGGACCAGAA--GTGTCTA-GCACAACACAATCC--TGAGGCA-----CAGTC---T
      270     280     290     300

      360     370     380     390     400     410
./wwwt CATTTTCTGGACCAAAATCTGCAAACCACTGCTCCATCAGGGGGAAGTCTGTACCAA-ACA
      : : : : : : : : : : : : : : : : : : : : : :
35 D3    GATGACCGGAG-ACAAACAG-AACCTTCTTGTCCTTTGGTGTGATACATTTTGAAGACA
      310     320     330     340     350     360

      420     430     440     450     460     470
./wwwt ACGCAAGAAGGGTTCTCATGGGCTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCT
      : : : : : : : : : : : : : : : : : : : : : :
40 D3    AGGT-----GGTTCTT-----CTGGAA--TACCCATCCAAG-----GTGCTATGATCAT
      370     380     390     400

      480     490     500     510     520
./wwwt GTCTTGCAAAGTATGGAAAGATGC---CACCTACTCCTCCAAAGCCAGACTCCATGTGTG
      : : : : : : : : : : : : : : : : : : : : : :
45 D3    TAC---CAAAGCTTTCAGTGATGGTTTCTCAATTTCTCC---TCGGAAATGATTTTGA
      410     420     430     440     450     460

      530
./wwwt TCAGAAAATT
      : : : : :
50 D3    TCAGAAG---

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 12 537 nt frente a D4 rev 416 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 45,0%; puntuación de alineación global: -297

```

10      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      -----CTTGCGTTGTTTGGTACAGACT---TCCCCCTGATGGAGCACTGGTTTGC
      10      20      30      40

15      70      80      90      100     110     120
./wwwt ATGGTGTGTCCGAGAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      AGA'TTTTGGTCCAGAAATGACGAGCACAGCAAAACCCCTTCA-ATG-----CAGT--C
      50      60      70      80      90

20      130     140     150     160     170
./wwwt CGCTGCAATAAT-GGCATCTGTATCCAGTTACTGAAAGCATCTTAACCCCTCACATCCC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      TGATG--ATCGTAGGCAGGGGTCTCC-----TTCATGCCCTTTTATATGTGACATCTT
      100     110     120     130     140

25      180     190     200     210     220     230
./wwwt GGCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTT
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      AG---TGTTGTTCTTTC-----CTAGATT-----CTGCCAT--CCCAAGTCATGGTTT
      150     160     170     180     190

30      240     250     260     270     280     290
./wwwt GGGATGGCAGAATCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAGG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      G-----AGTA-----ATGACCGTGGTTTC-GATCTCT-----GTGCC-----G
      200     210     220

35      300     310     320     330     340     350
./wwwt AGACCCCTGCCTACGATCATCAGACTGCAT-TGAAGGGTTTTGCTGTGCTC-GTCATTC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      AG-----TACCATCCAGAGCCGGGATGTGAGGGGTTAAGATGCTTTCAGTAACT -
      230     240     250     260     270

40      360     370     380     390     400     410
./wwwt TGGACCAAAATCTGCAAACCAGTGTCCATCAGGGGGAAGTCTGTACCAAACAACGCAAG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      GGGATACAGATG-----CCATTATTGCA--GCGGGTA--CTGGGGCAGCACATGCCAT
      280     290     300     310     320

45      420     430     440     450     460     470
./wwwt AAGGGTTCTCATGGGCTGGAAATTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      CTCGGTGG-CA-GCGCTTCT--TTTTCTCCGACAC-ACCATGC--AGGCCGATGATCC

50      330     340     350     360     370

55      480     490     500     510     520     530
./wwwt AAAGTATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTTCAGAAAATT
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
D4      TTGGTGGGGACTG-TGGCA---ATACCTCCCAA-----CNGCTGTGCC---GAANT
      380     390     400     410

```

ES 2 285 768 T3

ALIGN calcula una alineación global de dos secuencias

Versión 2.0u, cítese por favor: Myers y Miller, CABIOS (1989) 4:11-17

Seq 12 537 nt frente a D8 rev 355 nt

Matriz de puntuación: ADN, penalizaciones de hueco: -14/-4

Identidad del 40,7%; puntuación de alineación global: -504

```

10      10      20      30      40      50      60
./wwwt GAATTCGGCACGAGGGTTGGGAGGTATTGCCACAGTCCCCACCAAGGATCATCGGCCTGC
      :
D8      CA-----CCCAGTCC-----AGGAAAACCAACCCAA
      10      20

15      70      80      90      100     110     120
./wwwt ATGGTGTGTCTCGGAGAAAAAAGAAGCGCTGCCACCGAGATGGCATGTGCTGCCCCAGTACC
      : : : : : : : : : : : : : : : : : : : : : :
D8      AAGGAA--GCCAAGTATTAAGAA-----ATCACAAGG-----CAGGA--
      30      40      50      60

20      130     140     150     160     170
./wwwt CGCTGCAATAATGGCATCTGTATCCCAGTTA-CTGAAAGCATCTTAACCCCTCACATCCC
      : : : : : : : : : : : : : : : : : : : : : :
D8      AGGGACAAGA--GGGA--GAA--AGTTGTCTGAGAACTT-TTGACTGT-----
      70      80      90      100

25      180     190     200     210     220     230
./wwwt GGCTCTGGATGGTACTCGGCACAGAGATCGAAACCACGGTCATTACTCAAACCATGACTT
      : : : : : : : : : : : : : : : : : : : : : :
D8      GGCCCTGGACTTTGCTGTGCTC-----GTCATITTTGGACGAA--AATT
      110     120      130     140

30      240     250     260     270     280     290
./wwwt GGGATGGCAGAA-TCTAGGAAGACCACACACTAAGATGTCACATATAAAAGGGCATGAAG
      : : : : : : : : : : : : : : : : : : : : : :
D8      TGTAAGCCAGTCCTTTTGGAGGGACA-----GGTCTGCTCCAGAAGAGGGCATAAAG
      150     160     170      180     190

35      300     310     320     330     340     350
./wwwt GAGACCCCTGCCTACGATCATCAGACTGCATTGAAGGGTTTTG-CTGTGCTCGTCATTTC
      : : : : : : : : : : : : : : : : : : : : : :
D8      ----ACACTGCTCAAGCTC--CAGAAATC-TTCCAGCGTTGCGACTGTGGCC-----C
      200     210     220     230     240

40      360     370     380     390     400     410
./wwwt TGGACCAAAATCTGCAAACAGTGCTCCATCAGGGGGAAGTCTGTACCAAACAACGCAAG
      : : : : : : : : : : : : : : : : : : : : : :
D8      TGGACTACTGTGTCGAAGCCAATTGACCAGCAATCGGCAG-----
      250     260     270     280

45      420     430     440     450     460     470
./wwwt AAGGGTTCTCATGGGCTGGAAATTTTCCAGCGTTGCGACTGTGCGAAGGGCCTGTCTTGC
      : : : : : : : : : : : : : : : : : : : : : :
D8      -----CAT--GCTCGA-----TTAAGAGTATGCCAAA-----AAATAGA
      420     430     440     450     460     470

```

ES 2 285 768 T3

		290		300	310		
		480	490	500	510	520	530
5	./www	AAAG-TATGGAAAGATGCCACCTACTCCTCCAAAGCCAGACTCCATGTGTGTCAGAAAAT					
		::: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::					
	D8	AAAGCTAT--AAATATTTCA--AAAT-----AAAGA-AGAATCCACAT-TGC-----					
		320	330	340	350		
10	./www	T					
	D8						
15							
20							
25							
30							
35							
40							
45							
50							
55							
60							
65							