



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 431**

51 Int. Cl.:  
**C12N 15/67** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **02763110 .0**

96 Fecha de presentación : **04.10.2002**

97 Número de publicación de la solicitud: **1432808**

97 Fecha de publicación de la solicitud: **30.06.2004**

54 Título: **Regulación de la traducción de genes expresados heterológamente.**

30 Prioridad: **05.10.2001 US 327003 P**  
**05.10.2001 EP 01203772**  
**19.04.2002 EP 02076593**

45 Fecha de publicación de la mención BOPI:  
**20.10.2011**

45 Fecha de la publicación del folleto de la patente:  
**20.10.2011**

73 Titular/es: **H1 H4 HOLDING B.V.**  
**Pandastraat 9**  
**6531 VC Nijmegen, NL**

72 Inventor/es:  
**Van Herpen, Marinus, Maria, Antonius;**  
**Hulzink, Jozef, Maurinus, Raymond y**  
**Croes, Anton, Felix**

74 Agente: **Tomás Gil, Tesifonte Enrique**

ES 2 366 431 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

Regulación de la traducción de genes expresados heterológamente

**Antecedentes de la invención**

[0001] La reproducción sexual en plantas y animales requiere la producción de gametos. Aunque hay muchas diferencias citológicas en las fases de desarrollo de los gametos entre el reino animal y el vegetal, se dan varias similitudes. Tanto en los vegetales como en los animales, la producción de gametos es un proceso muy ordenado que se caracteriza por las transiciones de las células madre de un estado fisiológico al siguiente mediante divisiones meióticas y mitóticas. Esta serie de transiciones está dirigida por cambios múltiples en la expresión genética. Varias fases del desarrollo de gametos en diferentes especies vegetales y animales se llevan a cabo casi sin actividad transcripcional. En estas fases, los ARNm previamente sintetizados se traducen en productos que son esenciales para que el desarrollo pueda continuar. Esto implica que, en estas especies, el control postranscripcional de la expresión genética es el principio fundamental del desarrollo de los gametos. Un ejemplo de regulación postranscripcional de expresión genética durante el desarrollo de gametos es la maduración y germinación de los gametofitos masculinos (polen) en las plantas angiospermas. El polen inmaduro consiste en una célula generativa pequeña y una célula vegetativa grande que se forman a partir de microesporas mediante una mitosis haploide asimétrica (mitosis de polen I). En las fases posteriores del desarrollo, algunas especies tienen una segunda mitosis haploide (mitosis de polen II) que da como resultado polen tricelular, un proceso ausente en la mayoría de especies (es decir, polen bicelular). La maduración de este polen bi y tricelular se completa mediante una serie de procesos de desarrollo que produce la deshidratación progresiva del polen y su transición a la dormancia. La maduración de polen de diferentes especies de vegetales viene acompañada de una acumulación de grandes cantidades de ARNr, ARNt, ARNm y ribosomas. En cuanto el polen aterriza en un estigma compatible, se da una rehidratación considerable del grano de polen. Esto lleva a una rápida reactivación de los mecanismos de traducción, que usan los ARNt, ARNr y ARNm previamente acumulados. Las proteínas que se sintetizan a partir de estos productos almacenados son necesarias para la fase progámica del polen, es decir, para la germinación del polen, el crecimiento posterior del tubo de polen y la segunda mitosis haploide en el caso del polen bicelular.

[0002] A pesar de la importancia de los procesos postranscripcionales que regulan la expresión genética del polen, se sabe poco sobre los mecanismos subyacentes a la regulación postranscripcional de la expresión genética del polen. Tan sólo hay un estudio en la técnica que se centra en la regulación traduccional de un gen expresado de polen (*lat52*) (Bate *et al.*, (1996) *Plant J.*, 10(4), 613-623). En este estudio, se demuestra la implicación del 5' UTR en la regulación de la traducción específica del polen. Más en general, en el ritmo de traducción influyen los elementos en *cis* de los ARNm. Por ejemplo, el nivel de traducción de especies de ARNm a partir de distintos sistemas eucarióticos es modulado por elementos en *cis* del 5' UTR, la secuencia codificante o el 3' UTR. Estos elementos en *cis* actúan influyendo en la estabilidad del ARNm, en la iniciación de la traducción o en su alargamiento.

[0003] La regulación de la síntesis de la proteína del polen del tabaco NTP303 se desarrolla a nivel postranscripcional. Las transcripciones del gen *npt303* se pueden detectar por primera vez tras la mitosis I del polen y continúan acumulándose durante la maduración del polen y el posterior crecimiento del tubo del polen (Weterings *et al.*, (1992) *Plant Mol. Biol.* 18(6), 1101-1111). En cambio, la proteína sólo aparece en cantidades detectables cuando comienza la rehidratación del polen (Wittink *et al.*, (2000) *Sex. Plant Reprod.* 12(5), 276-284). Así, a pesar de la acumulación de su ARNm, no se da una síntesis eficaz de la proteína NTP303 durante el desarrollo del polen, cuya restricción sólo se alivia al comenzar la germinación del polen. Sin embargo, no está claro qué elementos en *cis* del ARNm de *npt303*, si los hay, son los responsables de este mecanismo de regulación traduccional. En concreto, no está claro si cualquiera de esos elementos podría ser usado para regular la expresión de proteínas además de la NPT303, como, por ejemplo, las proteínas heterólogas.

**Descripción de la invención****Definiciones**

[0004] A continuación, se muestran las definiciones de los términos tal y como se usan en la invención.

Planta

[0005] Como se utiliza aquí, el término "planta" hace referencia bien a una planta entera, incluyendo en general la clase de plantas más altas susceptibles a técnicas de transformación, incluyendo tanto plantas monocotiledóneas como dicotiledóneas, o bien una parte de una planta, como, por ejemplo, raíces, tallos, hojas, pétalos, frutos, semillas, tubérculos, polen, meristemas, callos, sépalos, bulbos y flores. El término planta, como se utiliza aquí, también hace referencia, sin limitaciones, a células vegetales en semillas, cultivos de suspensión, embriones, regiones meristemáticas, tejidos de callo, hojas, raíces, brotes, tejido gametofítico y esporofítico, polen, protoplastos y microesporas. Además, todos los tejidos de las plantas en todos los órganos están incluidos en la definición del término planta como se utiliza aquí. Los tejidos de planta incluyen, pero no se limitan a, tejidos diferenciados e indiferenciados de una planta, incluyendo polen, tubos de polen, granos de polen, raíces, brotes, brotes de meristemas, nodos coleoptilares, napas, hojas, pétalos cotiledóneos, óvulos, tubérculos, semillas y granos. Los tejidos de las plantas pueden estar en la planta, o en un órgano, tejido o cultivo celular. Como se utiliza aquí, las plantas monocotiledóneas hacen referencia a una planta cuyas semillas tienen sólo un cotiledón, u órgano del embrión que almacena y absorbe el alimento. Como se utiliza en este caso, las plantas dicotiledóneas hacen

referencia a una planta cuyas semillas tienen dos cotiledones. Las plantas incluidas en la invención son todas plantas susceptibles de transformación.

#### Operativamente vinculado

5 [0006] Como se utiliza aquí, el término "operativamente vinculado" hace referencia a dos o más elementos de una secuencia de ácido nucleico que están físicamente vinculados y que tienen una relación funcional entre sí. Por ejemplo, un promotor está operativamente vinculado a una secuencia codificante si el promotor es capaz de iniciar o regular la transcripción o expresión de esa secuencia codificante, en cuyo caso, debería entenderse que la secuencia está "bajo el control" del promotor. Generalmente, cuando dos secuencias de ácidos nucleicos están operativamente vinculadas, éstas tendrán la misma orientación y normalmente también estarán en el mismo marco de lectura. Normalmente serán esencialmente contiguas, aunque podría no ser necesario.

#### Promotor

15 [0007] Como se utiliza aquí, el término "promotor" hace referencia a un fragmento de ácido nucleico que funciona para controlar la transcripción de uno o más genes, localizado hacia arriba con respecto a la dirección de transcripción del sitio de iniciación de la transcripción del gen, y se identifica estructuralmente por la presencia de un sitio de unión para la ARN-polimerasa dependiente de ADN, sitios de iniciación de la transcripción y otras secuencias de ADN, incluyendo, pero sin limitarse a, sitios de unión del factor de transcripción, sitios de unión de proteínas activadoras y represoras, así como otras secuencias de nucleótidos conocidas por un experto en la técnica, para actuar directa o indirectamente en la regulación de la cantidad de transcripción del promotor.

#### Hibridación de ortólogos de ácido nucleico e hibridación del 5' UTR

20 [0008] Toda secuencia de nucleótidos capaz de hibridar con las secuencias de nucleótidos de la SEC. ID. N°. 1 se define como parte del 5' UTR de la invención. Las condiciones de hibridación rigurosas se definen aquí como condiciones que permiten que una secuencia de ácidos nucleicos de al menos 25 nucleótidos, aunque preferiblemente de 50, 75 ó 100, y más preferiblemente de 150 o más, hibride a una temperatura de aproximadamente 65°C en una solución que contiene aproximadamente 1 M de sal, preferiblemente 6 x SSC o cualquier otra solución con una fuerza iónica comparable, y con un lavado a 65°C en una solución que contiene aproximadamente 0,1 M de sal o menos, preferible 0.2 x SSC o cualquier otra solución con una fuerza iónica comparable. Es preferible realizar la hibridación durante toda la noche, es decir, al menos durante 10 horas y es preferible realizar el lavado durante al menos una hora cambiando la solución de lavado al menos dos veces. Estas condiciones normalmente permiten que la hibridación específica de las secuencias tengan aproximadamente un 90% de identidad de secuencia o más. Las condiciones de hibridación de moderadas se definen aquí como condiciones que permiten que una secuencia de ácidos nucleicos de al menos 50 nucleótidos, preferiblemente 150 o más, hibride a una temperatura de aproximadamente 45°C en una solución que tiene aproximadamente 1 M de sal, preferiblemente 6 x SSC o cualquier otra solución con una fuerza iónica comparable, y con un lavado a temperatura ambiente en una solución que tiene aproximadamente 1 M de sal, preferiblemente 6 x SSC o cualquier otra solución con una fuerza iónica comparable. Es preferible realizar la hibridación durante toda la noche, es decir, al menos durante 10 horas, y es preferible realizar el lavado durante al menos una hora cambiando la solución de lavado al menos dos veces. Estas condiciones normalmente permiten que la hibridación específica de las secuencias tenga hasta un 50% de identidad de secuencia. El experto en la técnica será capaz de modificar estas condiciones de hibridación para identificar específicamente secuencias cuya identidad varía entre un 50% y un 90%.

#### Homólogos

40 [0009] Se entiende que el término "homólogo", cuando se usa para indicar la relación entre un ácido nucleico (recombinante) o una molécula polipeptídica dados y un organismo o célula huésped dados, significa que en la naturaleza el ácido nucleico o la molécula polipeptídica los produce una célula u organismo huésped de la misma especie, preferiblemente de la misma variedad o raza. Si es homóloga a una célula huésped, la secuencia de ácido nucleico que codifica un polipéptido estará típicamente vinculada operativamente a otra secuencia del promotor o, en su caso, a otra secuencia de señal secretora y/o secuencia de terminación diferente a la de su entorno natural.

50 [0010] Cuando se usa para indicar la relación de dos secuencias de ácido nucleico, el término "homólogo" significa que una secuencia monocatenaria de ácido nucleico puede hibridar con una secuencia monocatenaria de ácido nucleico complementaria. El grado de hibridación puede depender de varios factores, que incluyen la extensión de la identidad entre las secuencias y las condiciones de hibridación, como la temperatura y la concentración de sal, tal como se discutirá a continuación. Es preferible que la región de identidad sea mayor que 5 bp, y es aún más preferible que sea mayor que 10 bp.

#### Heterólogos

55 [0011] El término "heterólogo", cuando se usa respecto a un ácido nucleico o a una molécula polipeptídica se refiere a un ácido nucleico o a un polipéptido de una célula exógena que no aparece de forma natural como parte del organismo, célula, genoma o secuencia de ADN o ARN en el que está presente, o que se encuentra en una célula o lugar o lugares en el genoma o secuencias de ADN o ARN que difieren de los que se encuentran en la naturaleza. Los ácidos nucleicos heterólogos o las proteínas no son endógenas a la célula en la que se introducen, sino que se obtienen de otra célula producida sintéticamente o por recombinación. Generalmente, aunque no necesariamente,

tales ácidos nucleicos codifican proteínas que no son producidas normalmente por la célula en la que el ADN se transcribe o expresa. De forma similar, los códigos exógenos de ARN para las proteínas no se expresan normalmente en la célula en donde está presente el ARN exógeno. Además, se sabe que una proteína heteróloga o un polipéptido pueden estar compuestos por elementos homólogos dispuestos en un orden y/u orientación que no se encuentra normalmente en el organismo, tejido o célula huésped del mismo al que se transfiere, es decir, la secuencia de nucleótidos que codifica dicha proteína o dicho polipéptido se origina a partir de la misma especie, pero se modifica sustancialmente la composición y/o localización genómica a partir de su forma nativa mediante intervención humana deliberada. También se puede hacer referencia a los ácidos nucleicos heterólogos y a las proteínas como ácidos nucleicos exógenos o proteínas exógenas. Cualquier ácido nucleico o proteína que un experto en la técnica reconocería como exógeno o heterólogo a la célula en la que se expresa, en este documento, se engloba en el término ácido nucleico heterólogo o proteína. El término heterólogo también se aplica a combinaciones no naturales de secuencias de ácido nucleico o de aminoácidos, es decir, combinaciones donde al menos dos de las secuencias combinadas son exógenas entre sí.

#### Identidad de secuencias

[0012] La "identidad de secuencias", como se conoce en la técnica, es una relación entre dos o más secuencias de aminoácidos (polipéptido o proteína) o dos o más secuencias de ácido nucleico (polinucleótido), que se determina al comparar las secuencias. En la técnica, el porcentaje de "identidad" indica el grado de parentesco de las secuencias entre secuencias de aminoácidos o ácido nucleico como determina la correspondencia entre las cadenas de dichas secuencias. Dos secuencias de aminoácidos se consideran "similares" si los polipéptidos sólo difieren en sustituciones de aminoácidos conservados. En la determinación del grado de similitud del aminoácido, el experto en la materia tiene en cuenta las sustituciones "conservadoras" de aminoácidos. Las sustituciones conservadoras de aminoácidos se refieren al intercambio de aminoácidos con cadenas laterales similares. Por ejemplo, un grupo de aminoácidos con cadenas laterales alifáticas es el de la glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos con cadenas laterales alifáticas con hidróxilo es el de la serina y la treonina; un grupo de aminoácidos con cadenas laterales que contienen amida es el de la asparagina y la glutamina; un grupo de aminoácidos con cadenas laterales aromáticas es el de la fenilalanina, tirosina y triptófano; un grupo de aminoácidos con cadenas laterales básicas es el de la lisina, arginina e histidina; y un grupo de aminoácidos con cadenas laterales que contienen azufre es el de la cisteína y la metionina. Los grupos de sustitución de aminoácidos conservadores que se prefieren son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, y asparagina-glutamina. Las variantes sustitutivas de las secuencias de aminoácidos que se divulgan en este documento son aquellas en las que se ha quitado al menos un residuo de las secuencias divulgadas y se ha insertado un residuo diferente en su lugar. Preferiblemente, el cambio del aminoácido es conservador. Las sustituciones conservadoras preferidas para cada uno de los aminoácidos de origen natural son de la siguiente manera: Ala por ser; Arg por lys; Asn por gln o his; Asp por glu; Cys por ser o ala; Gln por asn; Glu por asp; Gly por pro; His por asn o gln; Ile por leu o val; Leu por ile o val; Lys por arg; Asn por gln o glu; Met por leu o ile; Phe por met, leu o tyr; Ser por thr; Thr por ser; Trp por tyr; Tyr por trp o phe; y Val por ile o leu.

[0013] La "identidad" y la "similitud" pueden cacularse fácilmente mediante métodos conocidos, incluyendo pero sin limitarse sólo a, los descritos en Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genoma Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., y Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991; y Carillo, H., y Lipman, D., SIAM J. Applied Math., 48:1073 (1988).

[0014] Los métodos preferidos para determinar la identidad se diseñan para proporcionar la mayor correspondencia posible entre las secuencias evaluadas. Los métodos para determinar la identidad y la similitud están codificados en programas informáticos a disposición del público. Los métodos de programas de ordenador preferidos para determinar la identidad y la similitud entre dos secuencias incluyen, por ejemplo, el paquete de programas GCG (Devereux, J., *et al.*, Nucleic Acids Research 12 (1):387 (1984)), BestFit y FASTA (Altschul, S. F. *et al.*, J. Mol. Biol. 215:403-410 (1990)). La familia de programas BLAST 2.0, que se puede usar en la búsqueda de similitud entre bases de datos, incluye: BLASTN para secuencias de consultas de nucleótidos en comparación con secuencias de base de datos de nucleótidos; BLASTX para secuencias de consultas de nucleótidos en comparación con secuencias de base de datos de proteínas; BLASTP para secuencias de consultas de proteínas en comparación con secuencias de bases de datos de proteínas; TBLASTN para secuencias de preguntas de proteínas en comparación con secuencias de bases de datos de nucleótidos; y TBLASTX para secuencias de consultas de nucleótidos en comparación con secuencias de base de datos de nucleótidos. El programa BLASTX está a disposición del público en el Centro Nacional para la Información Biotecnológica de los Estados Unidos (NCBI) y en otras fuentes (BLAST Manual, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894 Altschul, S., *et al.*, J. Mol. Biol. 215: 403-410 (1990)). El conocido algoritmo de Smith-Waterman puede también utilizarse para determinar la identidad.

[0015] Los parámetros preferidos en la comparación de secuencias de polipéptidos incluyen los siguientes: algoritmo: Needleman y Wunsch, J. Mol. Biol. 48:443-453 (1970); matriz de comparación: BLOSSUM62 de Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. EE.UU. 89:10915-10919 (1992); penalización por *gap*: 12; y penalizaciones por extensión de *gap*: 4. Hay un programa útil con estos parámetros se encuentra a disposición del público conocido como el programa "Ogap" de Genetics Computer Group, situado en Madison, Wisconsin. Los parámetros

mencionados son los parámetros por defecto de las comparaciones de aminoácidos (así como la no penalización de *gaps* finales).

[0016] Los parámetros preferidos en la comparación de ácido nucleico incluyen los siguientes: algoritmo: Needleman y Wunsch, J. Mol. Biol. 48:443-453 (1970); matriz de comparación: coincidencias=+10; diferencias=0; penalización por creación de *gap*: 50; penalización por extensión de *gap*: 3. Disponible como el programa Gap de Genetics Computer Group, situado en Madison, Wisconsin. Más arriba se proporcionan los parámetros de fichero de las comparaciones de ácido nucleico.

#### ARN mensajero

[0017] El "ARN mensajero (ARNm)", como se utiliza en este documento, se refiere a una copia complementaria temporal del ARN de la cadena antisentido (cadena intrascrita o molde) de ADN codificador de proteínas. En eucariotas, normalmente se transcribe como un pre-ARNm relativamente largo (también llamado transcripción primaria o ARN<sub>nh</sub>) que se procesa después, todavía dentro del núcleo, para eliminar los intrones. También pueden darse modificaciones postranscripcionales. El ARNm maduro se transporta después al citoplasma, donde se traduce en proteína en el ribosoma. Además, el ARNm generalmente comprende una región que especifica la secuencia de proteínas, flanqueado a ambos lados por regiones no traducidas llamadas 5' y 3' "untranslated regions" (5'UTR y 3'UTR).

#### Ácido nucleico antisentido

[0018] El "ácido nucleico antisentido", como se utiliza en este documento, se refiere al ARN, ADN o molécula de ANP complementario al total o a una parte del objetivo de transcripción primaria o ARNm y que bloquea la traducción de la secuencia de nucleótidos que se tiene por objetivo.

#### **Descripción detallada de la invención**

[0019] En un primer aspecto, la invención se refiere a un método para la regulación de la traducción de la segunda secuencia de nucleótidos que codifica una proteína en una célula, donde el método es tal y como se define en reivindicación 1 o en la 2.

[0020] En otro aspecto, la invención se refiere a un método para la expresión de una proteína o polipéptido de interés en una planta, incluyendo los siguientes pasos:

- a) proporcionar un constructo de ácidos nucleicos, que comprende una primera secuencia de nucleótidos con al menos un 50% de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N° 1, vinculada operativamente a una segunda secuencia de nucleótidos que codifica una proteína o polipéptido de interés y operativamente vinculado también a un promotor heterólogo,
- b) poner una planta en contacto con dicho constructo de ácidos nucleicos para obtener una planta transformada, y
- c) someter dicha planta transformada a condiciones que llevan a la expresión de la proteína o polipéptido de interés, y recuperar opcionalmente dicha proteína o polipéptido.

[0021] Según la invención, el constructo de ácidos nucleicos comprende una primera secuencia de nucleótidos que tiene al menos un 50 % de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N° 1 (usando el algoritmo BLAST de BLASTN 22.1, 13-04-2001; penalizaciones de *gap*: existencia 5, extensión 2). El constructo de ácidos nucleicos, según la invención, comprende preferiblemente una primera secuencia de nucleótidos que tiene al menos un 36%, aunque es más preferible que tenga al menos un 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% ó 98 % de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N° 1. En una forma preferida de realización de la invención de manera particular, el constructo de ácidos nucleicos comprende una primera secuencia de nucleótidos que tiene el 100% de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N° 1. La secuencia de nucleótidos presentada en este documento como SBQ ID. N° 1 es la secuencia 5' UTR del gen *ntp303* de *Nicotiana tabacum*.

[0022] En otra forma de realización, la invención se refiere a un método para la expresión de una proteína de interés en una planta que incluye los siguientes pasos:

- a) proporcionar un constructo de ácidos nucleicos, que comprende una primera secuencia de nucleótidos que tiene una secuencia de nucleótidos con al menos un 50% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N° 1 o una secuencia de nucleótidos con al menos un 35% de identidad de secuencia de nucleótidos con los nucleótidos 4 - 76 de la secuencia de nucleótidos de la SEC. ID. N° 1 o una combinación de los mismos, vinculado operativamente a una segunda secuencia de nucleótidos que codifica una proteína o polipéptido de interés y vinculado operativamente también a un promotor heterólogo,
- b) poner una planta en contacto con dicho constructo de ácidos nucleicos para obtener una planta transformada, y
- c) someter dicha planta transformada a condiciones que conducen a la expresión de la proteína o polipéptido de interés, y la recuperación opcional de dicha proteína o polipéptido.

[0023] Preferiblemente, dicha primera secuencia de nucleótidos comprende una secuencia de nucleótidos que tiene, al menos, un 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% y particularmente un 100% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de SEC. ID. N°: 1 o una secuencia de nucleótidos que tiene al menos un 55%, aunque más preferible al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% y particularmente un 100% de identidad de secuencia de nucleótidos con los nucleótidos 4 - 76 de la secuencia de nucleótidos de la SEC. ID. N°: 1 o una combinación de la misma.

[0024] En otra forma de realización del método anteriormente mencionado, la primera secuencia de nucleótidos comprende una secuencia de nucleótidos que tiene, al menos, un 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% y particularmente un 100% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de SEC. ID. N°: 1 o una secuencia de nucleótidos que tiene el 100% de identidad de secuencia de nucleótidos con los nucleótidos 27 - 50 de la secuencia de nucleótidos de SEC. ID. N°: 1 o una combinación del mismo. Los nucleótidos 27 - 50 de la secuencia de nucleótidos de SEC. ID. N°: 1 consisten en una repetición de 8 unidades GAA. Preferiblemente, dicha primera secuencia de nucleótidos comprende una secuencia de nucleótidos que consiste en 7, 6, 5, 4 o 3 unidades GAA.

[0025] Una secuencia de nucleótidos según la invención puede presentarse en forma de ARN o en forma de ADN, incluyendo el ADN genómico, es decir, el ADN que incluye intrones, ADNc o ADN sintético. El ADN puede ser monocatenario o bicatenario y, si es monocatenario, puede ser la cadena transcrita o la cadena no transcrita (antisentido). El ADN o ARN con un esqueleto modificado por la estabilidad o por otras razones son otra parte de la invención. Por otra parte, el ADN o ARN que comprende bases poco habituales, como la inosina o bases modificadas como las bases tritiladas, son también una parte de la invención. La secuencia de nucleótidos puede también ser una variante alélica de la secuencia de nucleótidos según la invención. Si se desea, la secuencia de nucleótidos puede prepararse o alterarse sintéticamente, de modo que las preferencias conocidas del codón de la expresión del huésped deseado pueda usarse de manera ventajosa. Se ha demostrado, por ejemplo, que las preferencias del codón y las preferencias del contenido del GC de monocotiledones y dicotiledones difieren (Murray *et al.*, Nucl. Acids Res. 17: 477-498 (1989)).

[0026] En una forma de realización preferida del método según la invención, el constructo de ácidos nucleicos comprende una segunda secuencia de nucleótidos que codifica una proteína o polipéptido de interés que está vinculado operativamente a cualquiera de las primeras secuencias de nucleótidos tal como se ha definido anteriormente. La proteína o polipéptido de interés puede ser una proteína o polipéptido homólogo, pero en una forma de realización preferida de la invención la proteína o polipéptido de interés es una proteína heteróloga. Una segunda secuencia de nucleótidos que codifica una proteína heteróloga o polipéptido puede derivar completamente o en parte de cualquier fuente conocida por la técnica, incluyendo un genoma o episoma vírico o bacteriano, ADN nuclear eucariota o plásmido, ADNc o ADN sintetizado químicamente. La segunda secuencia de nucleótidos puede constituir una región codificante ininterrumpida o puede incluir uno o más intrones enlazados por empalmes de unión apropiados. También puede estar compuesta por segmentos derivados de distintas fuentes, de origen natural o sintético. La segunda secuencia de nucleótidos que codifica la proteína o polipéptido de interés según el método de la invención es preferiblemente una secuencia de nucleótidos en toda su longitud, pero puede también ser una parte funcionalmente activa u otra parte de dicha secuencia de nucleótidos en toda su longitud. La proteína o polipéptido de interés puede ser una proteína o un polipéptido que confiere, por ejemplo, resistencia a los insectos, resistencia a la sequía, resistencia a la enfermedad, resistencia a los herbicidas, inmunidad, una absorción mejorada de nutrientes, minerales o agua de la tierra, o un metabolismo modificado en la planta. En otra forma de realización, la planta se usa para la sobreproducción de la proteína o polipéptido de interés. La segunda secuencia de nucleótidos que codifica la proteína o polipéptido de interés puede también comprender secuencias de señal que dirigen la proteína o polipéptido de interés cuando se expresa en una ubicación específica de la célula o tejido. Tales secuencias de señal incluyen, aunque de forma no limitativa, secuencias que dirigen la proteína o polipéptido de interés a organelos, otras células vegetales o espacio intercelular. Además, la segunda secuencia de nucleótidos que codifica la proteína o polipéptido de interés puede también comprender secuencias que faciliten la purificación y detección de la proteína mediante, por ejemplo, la transferencia de Western y ELISA (p. ej. C-Myc o secuencias de polihistidina).

[0027] La proteína o polipéptido de interés puede tener aplicaciones industriales o medicinales (farmacéuticas). Los ejemplos de proteínas o polipéptidos con aplicaciones industriales incluyen enzimas como, por ejemplo, la lipasa (p. ej. usadas en la industria de detergentes), la proteasa (usada, entre otras cosas, en la industria de detergentes, en elaboración de cerveza y similares), enzimas de degradación de pared celular (como la celulasa, la pectinasa, las beta-1,3/4 y la beta-1-6-gluanasas, la rhamnogalacturonasa, la mananasa, la xilanasa, la pululanasa, la galactanasa, la esterasa y similares, usadas en el tratamiento de la fruta para la fabricación del vino y similares, o en pienso), la fitasa, la fosfolipasa, la glicosidasa (como la amilasa, la beta-glucosidasa, la arabinofuranosidasa, la rhamnosidasa, la apiosidasa y similares), enzimas para lácteos (p. ej. la quimosina). Las proteínas o los polipéptidos de los mamíferos, y preferiblemente de los humanos, y/o las enzimas con aplicaciones terapéuticas, cosméticas o diagnósticas incluyen, aunque de forma no limitativa, la insulina, la albúmina de suero (ASH), la lactoferrina, la hemoglobina  $\alpha$  y B, el activador del tejido plasminógeno (tPA), eritropoyetina (EPO), factores de necrosis tumoral (TNF), BMP (proteína morfogenética ósea), factores de crecimiento (G-CSF, GM-CSF, M-CSF, PDGF, EGF, y similares), las hormonas peptídicas (p. ej. la calcitonina, la somatomedina, la somatotropina, las hormonas del crecimiento, la hormona estimulante del folículo (FSH), las interleucinas (IL-x), y los interferones (IFN-y). También se

incluyen los antígenos víricos y bacterianos, p. ej. para su uso en vacunas, incluyendo, p. ej., la subunidad B de la toxina termolábil, la subunidad B de la toxina del cólera, la proteína de envoltura de la superficie del virus de hepatitis B, la proteína del cápside del virus de Norwalk, la glicoproteína B del citomegalovirus humano, la glicoproteína S, el interferón, y la corona de gastroenteritis transmisible de receptores virales y similares. Además, se incluyen genes que codifican mutantes o análogos de dichas proteínas.

[0028] En una forma de realización de la invención, el constructo de ácidos nucleicos comprende además un promotor para el control y la iniciación de la transcripción de la segunda secuencia de nucleótidos. Es preferible que el promotor sea capaz de causar la expresión de la segunda secuencia de nucleótidos en la célula huésped elegida. Dicho promotor, por ejemplo específico del polen o heterólogo, está operativamente vinculado a cualquiera de las secuencias de nucleótidos mencionadas anteriormente. En una forma de realización preferida de la invención, el promotor es un promotor vegetal, es decir, un promotor capaz de iniciar la transcripción en células vegetales. Los promotores de las plantas, como se utilizan en este documento, incluyen los promotores específicos de tejido, los preferidos de tejido, los específicos del tipo de célula, los inducibles y los constitutivos. Los promotores específicos de tejido son promotores que inician la transcripción sólo en determinados tejidos y se refieren a una secuencia de ADN que proporciona señales de reconocimiento para la ARN-polimerasa y/u otros factores necesarios para que se inicie la transcripción, y/o para controlar la expresión de la secuencia codificante con precisión en tejidos determinados o en algunas células de estos tejidos. La expresión realizada en la manera específica de tejido puede darse sólo en tejidos individuales o en combinaciones de tejidos. En una forma de realización preferida de la invención, la expresión es específica de polen o semilla, es decir, la expresión es específica sólo en polen o semillas. Los promotores específicos de polen y específicos de semilla incluyen, aunque de forma no limitativa, los promotores de los genes específicos de polen *ntp303* (*N. tabacum*) y *zm13* (*Z. mays*) y los genes específicos de semilla *dc8* (*D. carota*), *rab17* (*Z. mays*), *rab16b* (*O. sativa*) y *em* (*T. aestivum*). El grupo de promotores específicos de tejido es examinado por Edwards, J.W. & Coruzzi, G.M., Annu. Rev. Genet. 24, 275-303 (1990) e incluyen, aunque de forma no limitativa, los promotores específicos de embrión, como los promotores del gen de betaconglucina de las proteínas de almacenamiento embrionario de las semillas de soja, los genes de legúmina de judía común, el gen de betafaseolina y los genes de napina y cruciferina procedentes de la semilla de colza; los promotores específicos de endosperma, como los promotores de genes de zeína de maíz, los genes de glutenina de trigo y los genes de hordeína de cebada; los promotores específicos de fruta tales como el promotor del gen E8 del tomate sensible al etileno; los promotores específicos de tubérculos, como el promotor de la patatina de Clase 1 de la patata y los promotores específicos de hoja, como los promotores del gen de la subunidad pequeña de la ribulosa 1,5 bifosfato carboxilasa y el gen *a/b* de la proteína de enlace *a/b* de la clorofila.

[0029] Los promotores preferidos de tejido son promotores que preferentemente inician la transcripción en determinados tejidos, tales como hojas, raíces, tallos, flores o semillas.

[0030] Los promotores específicos de tipos de célula son promotores que principalmente dirigen la expresión en determinados tipos de células en uno o más órganos; por ejemplo, en las células vasculares de las raíces u hojas. Los promotores inducibles son promotores capaces de activar la transcripción de una o más secuencias de ADN o genes como respuesta a un inductor. Las secuencias de ADN o de genes no serán transcritas sin la presencia del inductor. Los inductores conocidos en la técnica incluyen concentraciones de sal altas, frío, calor o elementos tóxicos e incluyen agentes patógenos como los virus. Los inductores pueden ser agentes químicos, como los herbicidas, las proteínas, los reguladores de crecimiento, los metabolitos y los compuestos fenólicos. El inductor puede también ser un agente de iluminación, como la luz y la oscuridad en diferentes modalidades, que incluyen la longitud de onda, la intensidad, la fluencia, la dirección y la duración. La activación de un promotor inducible se establece mediante la aplicación del inductor. El grupo de promotores generalmente inducibles incluye, aunque no es limitativo, el promotor de choque térmico *hsp70* de *Drosophila melanogaster*, un promotor inducible por el frío de *Brassica napus* y un promotor de alcohol deshidrogenasa inducido por etanol. Los promotores vegetales específicos inducibles incluyen, aunque de forma no limitativa, el promotor inducible por tetraciclina y el promotor de la  $\alpha$ -amilasa.

[0031] Los promotores constitutivos son promotores que se activan bajo muchas condiciones medioambientales y en muchos tipos de tejido diferentes. El grupo de promotores constitutivos incluye, aunque de manera no limitativa, el promotor 35S o el promotor 19S del virus del mosaico de la coliflor (CaMV), el promotor de ubiquitina, el promotor de revestimiento del virus del mosaico del tabaco (TMV), los promotores de virus del mosaico de las nervaduras de la mandioca (CsVMV), el promotor actina 1 del arroz, y regiones reguladoras asociadas a los genes de *Agrobacterium*, como la nopalina sintasa (Nos), manopina sintasa (Mas) y la octopina sintasa (Ocs).

[0032] El constructo de ácidos nucleicos según la invención es preferiblemente un vector, en particular una secuencia de nucleótidos plásmidos, cósmidos o fagos, lineal o circular, de un ADN o ARN mono o bicatenario, derivado de cualquier fuente, en el que varias secuencias de nucleótidos se han unido o recombinado en una única construcción que es capaz de introducir cualquiera de las secuencias de nucleótidos de la invención en orientación de sentido o antisentido en una célula, en particular en una célula vegetal. La elección del vector depende de los procedimientos recombinantes llevados a cabo y la célula huésped utilizada. El vector puede ser un vector de replicación autónoma o puede replicarse junto con el cromosoma en el que ha sido integrado. Preferiblemente, el vector contiene un marcador de selección. Los marcadores útiles dependen de la célula huésped elegida y son bien conocidos por personas expertas en la técnica. En caso de que la proteína deba obtenerse a partir de hojas o raíces, la infección de células con un vector viral tiene la ventaja de que una gran proporción de las células seleccionadas

pueden recibir el ácido nucleico. Adicionalmente, las moléculas codificadas en el vector viral, por ejemplo, por un ADNc que contiene el vector viral, se expresan eficazmente en células que han absorbido ácido nucleico del vector viral. Los vectores adecuados que puede ser entregados mediante procedimientos actualmente conocidos incluyen, aunque de forma no limitativa, vectores de virus del herpes simplex, vectores de adenovirus, vectores de papovavirus (como los vectores del virus del papiloma humano, los vectores *polyomavirus* y los vectores SV40), los vectores de virus adenoasociados, los vectores retrovirales, los virus seudorrabia, los vectores de herpesvirus alfa y similares. Un estudio minucioso de los vectores virales, especialmente de vectores virales adecuados para modificar células que no se replican, y de cómo usar tales vectores conjuntamente con la expresión de polinucleótidos de interés se puede encontrar en el libro VIRAL VECTORS: GENE THERAPY AND NEUROSCIENCE APPLICATIONS (Ed. Caplitt y Loewy, 1995). Los vectores plásmidos basados en el género *Agrobacterium* son los preferidos para la transformación estable de constructos de ácido nucleico en el genoma de una planta. La elección del vector de transformación depende del procedimiento de transformación llevado a cabo y de la célula huésped utilizada. Los vectores binarios Ti que se pueden utilizar en la transferencia de genes mediada por *Agrobacterium* incluyen pBIN19; pC22; pGA482 y pPCV001.

[0033] Una célula huésped recombinante, como una célula de mamífero (con la excepción de las células humanas), vegetal, animal, de insectos, fúngica o bacteriana, que contiene una o más copias de un constructo de ácidos nucleicos de acuerdo con la invención es un tema adicional de la invención. Por célula huésped se entiende una célula que contiene un constructo de ácidos nucleicos, como un vector, y soporta la replicación y/o expresión del constructo de ácidos nucleicos. Ejemplos de bacterias adecuadas son las bacterias Gram positivas, como varias especies de los géneros *Bacillus*, *Streptomyces* y *Staphylococcus* y las bacterias Gram negativas, como las diferentes especies de los géneros *Escherichia* y *Pseudomonas*. En el grupo de células micóticas se utilizan preferiblemente células de levadura. La expresión en la levadura puede conseguirse utilizando cepas de levadura como la *Pichia pastoris*, la *Saccharomyces cerevisiae* y la *Hansenula polymorfa*. Asimismo, células de insecto, como las células de *Drosophila* y Sf9, pueden utilizarse como células huéspedes. De manera alternativa, un sistema de expresión adecuado puede ser también un sistema de baculovirus o sistemas de expresión que utilizan células mamíferas como CHO, Cos o células de melanoma de Bowes. Para procedimientos de transformación en plantas, las bacterias adecuadas incluyen *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes*.

[0034] Otro aspecto de la invención se refiere a una planta que es genéticamente modificada preferiblemente mediante el método de la invención, en el que la planta comprende un constructo de ácidos nucleicos como se define más arriba en el presente documento. El constructo de ácidos nucleicos es preferiblemente un constructo que contiene secuencias de ácidos nucleicos modificados o manipulados *in vitro* o, al menos, *ex planta*. Como tal, el constructo de ácidos nucleicos proporciona preferiblemente a la planta una combinación de secuencias de ácidos nucleicos que no se encuentran en la naturaleza. El constructo de ácidos nucleicos preferiblemente se mantiene estable, bien como un elemento que se replica de manera autónoma, o bien, más preferiblemente, el constructo de ácidos nucleicos se integra en el genoma de la planta, en cuyo caso el constructo se integra normalmente en posiciones aleatorias en el genoma de la planta, por ejemplo, mediante recombinación no homóloga. Las plantas que se prefieren en la invención incluyen el tabaco, la patata, el azúcar, la remolacha, la soja, el maíz, el arroz, el altramuz, la alfalfa, la *Arabidopsis* y la *Brassica*. Las plantas estables transformadas (transgénicas) o células vegetales se producen mediante métodos conocidos. El término transformación estable se refiere a la exposición de plantas, tejidos o células de las mismas a métodos para transferir e incorporar ADN exógeno al genoma de la planta. Estos métodos incluyen, aunque de forma no limitativa, la transformación de genes mediada por *Agrobacterium tumefaciens*, la transferencia de ADN purificado mediante bombardeo de micropartículas, la electroporación de protoplastos y la microinyección o uso de fibras de silicio para facilitar la penetración y la transferencia de ADN en la célula vegetal. Las plantas dicotiledóneas se transforman más frecuentemente mediante la transferencia de genes mediada por *Agrobacterium*, como, por ejemplo, mediante el cocultivo de protoplastos de regeneración de plantas o cultivos de células con *Agrobacterium tumefaciens*. En general, cuando el *Agrobacterium tumefaciens* se usa para la transformación, los vectores de transformación son preferiblemente vectores cointegrados o vectores binarios. Las plantas dicotiledóneas pueden transformarse además mediante la transformación de discos de hoja, mediante la transformación de protoplastos por transferencia de ADN inducida por polietileno glicol, electroporación, sonicación o microinyección, así como la transformación de células o tejidos intactos mediante micro o macroinyección en tejidos o embriones, la electroporación del tejido, incubación de embriones secos en una solución con ADN, la infiltración al vacío de semillas y la transferencia de genes biolística. Las plantas monocotiledóneas se transforman mediante, por ejemplo, el bombardeo de partículas, la incorporación de ADN inducido eléctrica o químicamente en la electroporación de protoplastos de células parcialmente permeabilizadas, la macroinyección de ADN en inflorescencias, la microinyección de ADN en microesporas y pro-embryones, la introducción de ADN en polen en germinación y la integración de ADN en embriones mediante hinchamiento.

[0035] Un método alternativo para expresar una proteína o polipéptido de interés en plantas cuenta con la expresión transitoria de vectores basados en virus. Se sabe que los virus se replican con una eficiencia muy alta y, en algunos casos, pueden infectar la planta huésped entera, creando el potencial para expresar una proteína o polipéptido de interés en cantidades grandes. Los vectores que se pueden usar en este método alternativo son tobamovirus, potexvirus y potivirus.

[0036] Como alternativa, junto a la expresión en células huéspedes tales como las células vegetales, la proteína o polipéptido de interés se puede producir en sistemas de traducción acelular usando ARN derivado de los constructos



de ácidos nucleicos de la presente invención.

[0037] En el método según la invención, el constructo de ácidos nucleicos puede además comprender opcionalmente otros elementos reguladores conocidos en la técnica que se adecuan a dicho método. Estos incluyen, aunque de forma no limitativa, elementos presentes en el 5' UTR, en el 3' UTR y en las secuencias de nucleótidos codificantes de secuencias de nucleótidos de homólogas y/o heterólogas, incluyendo el elemento de respuesta al hierro (IRE), el elemento traduccional regulador en cis (TLRE) o los uORFs de los 5' UTR y las poli(U)extensiones de los 3' UTRs. Preferiblemente, dichos elementos reguladores están vinculados operativamente a las secuencias de nucleótidos y a los promotores según la invención.

[0038] Cuando un tejido o célula transformados (p. ej., trozos de hoja, segmentos de vástago, raíces, pero también protoplastos o células vegetales cultivados en suspensión) se obtiene con el método según la invención, se pueden regenerar plantas enteras a partir de dicho tejido o célula transformada en un medio adecuado, que opcionalmente puede contener antibióticos o biocidas conocidos en la técnica para la selección de células transformadas.

[0039] Las plantas transformadas resultantes se identifican preferiblemente mediante selección. Por lo tanto, el constructo de ácidos nucleicos según la invención preferiblemente comprende también un gen marcador que puede proporcionar la capacidad de selección o análisis en la planta tratada. Los marcadores seleccionables se prefieren generalmente para eventos de transformación vegetal, pero no están disponibles para todas las especies vegetales. Los marcadores seleccionables adecuados pueden ser genes de resistencia a los antibióticos o a los herbicidas que, cuando se insertan en algunas células de la planta en cultivo, les confieren a estas células la capacidad de resistir la exposición a un antibiótico o a un herbicida. Otro tipo de gen marcador es el que se puede analizar mediante un ensayo bioquímico o histoquímico, aunque dicho gen no puede ser seleccionado. Un gen marcador adecuado que es muy útil para tal transformación vegetal es el gen GUS. Jefferson *et al.*, EMBO J., 6: 3901-3907 (1987) divulgan el protocolo general a seguir en un ensayo de GUS. El gen GUS codifica una enzima que cataliza la segmentación de 5-bromo-4-cloro-3-indolil glucurónido, un sustrato que tiene un color azulado en la segmentación. Así, el uso de un gen GUS proporciona un ensayo conveniente a la hora de detectar la expresión del ADN introducido en las plantas mediante análisis histoquímico de las mismas. En un ejemplo del proceso de transformación, el gen que se quería expresar en la planta podía engancharse en serie con el gen GUS. El constructo en serie se podía transformar en plantas, y las plantas resultantes podían ser analizadas en la expresión de la enzima GUS. Otro ejemplo de un gen marcador es la luciferasa. Una ventaja de este marcador es el procedimiento no destructivo de aplicación del sustrato y la detección posterior. Las plantas transformadas pueden identificarse también mediante la expresión del gen de interés.

[0040] En un siguiente paso, la planta transformada, parte o célula de la misma está sujeta a condiciones que conducen a la expresión de la proteína o polipéptido de interés, y que opcionalmente recuperan dicha proteína o polipéptido. Los pasos de recuperación dependen de la proteína o polipéptidos expresados y la célula huésped utilizada, pero pueden comprender el aislamiento de la proteína o polipéptido. Cuando se aplica a una proteína/polipéptido, el término "aislamiento" indica que la proteína se encuentra en condiciones que no son las de su entorno nativo. En una forma preferida, la proteína aislada se encuentra sustancialmente sin otras proteínas, particularmente otras proteínas homólogas. Se prefiere proporcionar la proteína en una forma pura de más del 40%, aunque más preferiblemente en forma pura de más del 60%. Incluso es más preferible proporcionar la proteína en una forma altamente purificada, es decir, con una pureza de más del 80%, más preferiblemente con una pureza de más del 95%, e incluso más preferiblemente con una pureza de más del 99%, como lo establece la SDS-PAGE. Si se desea, la segunda secuencia de nucleótidos puede estar ligada a una secuencia de nucleótidos heteróloga para codificar una proteína de fusión que facilita la purificación de las proteínas y la detección de las mismas, por ejemplo, en la transferencia Western y en la prueba ELISA. Las secuencias heterólogas adecuadas incluyen, aunque de forma no limitativa, las secuencias de nucleótidos que codifican proteínas como, por ejemplo, la glutatión S-transferasa, la proteína de enlace de maltosa, la polihistidina de unión de metal, la proteína verde fluorescente, la luciferasa y la beta-galactosidasa. La proteína puede también acoplarse a portadores, indetectivos o etiquetas no-peptídicos que facilitan el trazado de la proteína, tanto *in vivo* como *in vitro*, y que permiten la identificación y la cuantificación de unión de la proteína a los sustratos. Tales etiquetas, indetectivos o portadores son bien conocidos en la técnica e incluyen, aunque de forma no limitativa, biotina, etiquetas radiactivas y etiquetas fluorescentes.

[0041] En una forma de realización particularmente preferida del método según la invención, la planta que se usa para la expresión de la proteína o polipéptido de interés es una planta de *Nicotianan tabacum* transgénica (homocigota), doble haploide, silenciada para *Ntp303*. La planta transgénica tiene su origen en *Nicotiana tabacum* cv. *Petit Havana*, que fue transformada con *ntp303* antisentido y seleccionada para ello mediante el cultivo de antera. Cuarenta plántulas independientes transformadas se generaron a partir de anteras mediante el cultivo de la antera *in vitro*. Cuatro de ellas fueron dobles haploides y podían producir polen haploide. Los caracteres morfológicos y agronómicos de estos dobles haploides con gametos ln se compararon con los de su progenitor natural. La planta transgénica y las flores que esta produce son exactamente las mismas que las del tipo natural. Por otra parte, proporciona una cantidad normal de polen, pero los cruces no dan descendencia (es decir, la planta es masculina estéril) No obstante, el polen puede germinarse *in vitro* y el crecimiento de su tubo de polen es similar al del polen natural. La germinación *in planta* del polen es escasa y los poco tubos de polen que se forman dejan de crecer después 10 mm de crecimiento en el estilo.

Preferiblemente, la proteína o polipéptido de interés se expresa en tubos de polen o semillas de una planta utilizados en el método según la invención. Los cultivos masivos de tubos de polen in vitro se pueden establecer en una cámara de germinación según las especificaciones de Schrauwen y Linskens (1967) Acta Bot. Neerl. 16 (5), 177-179.

- 5 [0042] Además, una parte de la invención es una planta transgénica (homocigota) doble haploide de *Nicotiana tabacum* silenciada para *Ntp303*, comprendiendo cualquiera de los constructos de ácido nucleico anteriormente descritos. La propagación, la cosecha y el material tisular de dicha planta transgénica, incluyendo, aunque de manera no limitativa, hojas, raíces, brotes y flores, son también una parte de la invención.

### Ejemplos

#### 10 Materiales y métodos

##### Material vegetal

- 15 [0043] Se utilizaron plantas de *Nicotiana tabacum* L. cv. Petit Habana SR1 cultivadas en invernadero como fuente de polen y tejido de hoja para el bombardeo con microproyectiles. Para evaluar la expresión transitoria de los diferentes genes quiméricos durante el desarrollo del polen, el polen inmaduro que estaba en la fase bicelular tardía fue aislado asépticamente de capullos de flores de 35 mm longitud en un medio M1 tal y como se ha descrito anteriormente (Tupý *et al.* (1991) Sex. Plant Reprod. 4(4): 284-287 ). La expresión transitoria de diferentes constructos de fusión de genes durante el crecimiento del tubo de polen fue medida usando polen maduro que fue aislado de flores de tabaco deshidratadas (Herpen *et al.*, (1992) Sex. Plant Reprod. 5, 304-309).

- 20 [0044] Después del aislamiento, el gránulo del polen fue suspendido en 100 µl en el medio M1 a una densidad de 10<sup>8</sup> células/ml. Para fijar el polen para el bombardeo de partículas, la suspensión de polen fue pipeteada sobre la superficie de una membrana Hybond-N (Amersham) estéril que fue colocada en un 1% de un medio M1 solidificado de agar. Tras el bombardeo, la membrana que contiene el polen bicelular tardío o maduro se puso en remojo en 10 ml de medio M1 o medio Read (Read *et al.*, (1993) Protoplasma 174, 101-115.), respectivamente. El polen bicelular tardío se incubó a 25°C a oscuras con agitación vigorosa. Después del centrifugado, el polen maduro fue suspendido en un tubo de 10-ml que contenía 0,5 ml de medio Read seguido de una incubación a oscuras de 20 horas. El tratamiento del tejido de la hoja antes y después del bombardeo se realizó como lo describe Hamilton *et al.*, (1992) Plant Mol. Biol. 18, 211-218.

[0045] En todos los casos, los bombardeos se realizaron durante 60 min de colocación del material vegetal sobre el medio solidificado.

#### 30 Preparación de constructos de fusión de genes que contienen diferentes UTR

- [0046] En todos los constructos, bien en una versión modificada de la región codificante de photinus-luciferina-4-monooxigenasa, *luc+*, o ADNc luciferasa procedente de *Renilla reniformis* (*rluc*) se utilizó como el gen indicador. El ADNc *luc+* se amplificó mediante una reacción en cadena de polimerasa (RCP) en el vector pGL3 (Promega) usando un cebador específico de secuencia hacia adelante que introdujo un sitio *NcoI* en el extremo 5' (5'-ATATCCATGGAAGACGCC; sitio *NcoI* subrayado; SEC. ID. N°: 2) y un cebador específico de secuencia inverso que introdujo un sitio *BamHI* en el extremo 3' (5'-ATATGGATCCTTACACGGCGATC; sitio *BamHI* subrayado; SEC. ID. N°: 3). El ADNc *rluc* se amplificó mediante RCP en el vector pRL-SV40 (Promega) usando los siguientes cebadores específicos de secuencia: 5'-GTGTCCATGGATGACTTCGAAAG (sitio *NcoI* subrayado; SEC. ID. N°: 4) y 5'-GTGTGGATCCTTATTGTTTCATTTTTGAG (sitio *BamHI* subrayado; SEC. ID. N°: 5). En la construcción de <sup>35S</sup>syn44 5'/35S 3', el producto de RCP de *luc+* fue digerido con *NcoI* y *BamHI* y, tras la eliminación del gen de luciferasa, ligado en los sitios *NcoI* y *BamHI* en pRTS2LUC (Bate *et al.*, (1996) Plant J., 10(4), 613-623). El pRTS2LUC (gentilmente proporcionado por el Dr. Davide Twell) contenía el promotor *CaMV35S* (Topfer *et al.*, 1987), un poliligador sintético de 44 pares de bases de longitud (denominado syn44 5' en este artículo), el ADNc luciferasa (Ow *et al.*, 1986) y la región no traducida *CaMV 35S 3'* (Topfer *et al.*, (1987) Nucl. Acids. Res. 15, 5890).

- 45 [0047] Se construyó un constructo casi idéntico, <sup>35S</sup>Rsyn44 5'/35S 3', en el que el gen *luc+* fue sustituido por la región codificante *rluc*. Para obtener un constructo de fusión de genes que contuviera tanto *ntp303* UTR como el promotor *CaMV35S* (pRH1-1 <sup>35S</sup>303 5'/303 3'), el syn44 5' UTR se extrajo de pRH4-1 usando las enzimas de restricción *XhoI* y *NcoI*. El 5' UTR de *ntp303* se amplificó mediante RCP en el clon genómico *ntp303* (Weterings *et al.*, (1995) Sex. Plant Reprod. 8(1) 11-17 ) usando los siguientes cebadores con sitios de restricción incorporados en el extremo 5': 5'-GTGTCTCGAGCAAGCTCTAGCAGGAAG (sitio *XhoI* subrayado; SEC. ID. N°: 6) y 5'-GTGTCCATGGGACGTTGTTTTTTTATTTC (sitio *NcoI* subrayado; SEC. ID. N°: 7). A continuación del RCP, el 5' UTR de *ntp303* se obtuvo mediante digestión de *XhoI-NcoI* y se ligó al constructo <sup>35S</sup>syn44 5'/35S 3' que carecía de syn44 5' UTR para crear el plásmido <sup>35S</sup>303 5'/35S 3'. Los oligonucleótidos 5'-ATATGGATCCATTCTGTAATGATCAATCTG (sitio *BamHI* subrayado; SEC. ID. N°: 8) y 5' ATATGAGCTCATTTAATGTTTTGTCCTA (sitio *SacI* subrayado; SEC. ID. N°: 9) se usaron para generar el 3' UTR de *ntp303* usando el clon genómico *ntp303* como molde. Este fragmento de RCP fue digerido con *BamHI* y *SacI* y se clonó en <sup>35S</sup>303 5'/35S 3' para reemplazar el *CaMV 35S 3'* UTR y para crear <sup>35S</sup>303 5'/303 3'.

[0048] Los constructos de fusión de genes con el promotor *ntp303* se realizaron como se describe a continuación. Utilizando el clon genómico de *ntp303* como molde, se amplificó un fragmento de promotor de 578 pares de bases

de largo, incluyendo el sitio de iniciación de la transcripción, (Weterings *et al.*, 1995) usando los cebadores 5'-ATATAAGCTTGATACACTCGCAACGTGTGT (sitio *Hind*III subrayado; SEC. ID. N°: 10) y 5'-ATATCTCGAGGAGCTTGCCTATTACCAT (sitio *Xho*I subrayado; SEC. ID. N°: 11). El fragmento amplificado del promotor *ntp303*, que incluye la región de la que se ha demostrado que refleja la región mínima hacia arriba del gen *ntp303* capaz de realizar la expresión directa del polen, se digirió con *Hind*III y *Xho*I y, tras la eliminación del promotor *CaMV* 35S, se ligó al <sup>35S</sup>R<sub>syn44</sub> 5'/35S 3', al <sup>35S</sup>303 5'/35S 3' y al <sup>35S</sup>303 5'/303 3' para crear <sup>35S</sup>R<sub>syn44</sub> 5'/35S 3', 303 5'/35S 3' y 303 5'/303 3', respectivamente. Para obtener un constructo conteniendo el promotor *ntp303*, los UTR de *ntp303* y el ADNc luciferasa *Renilla* (<sup>R</sup>303 5'/303 3'), el gen *luc+* se digirió desde 303 5'/303 3' utilizando *Nco*I y *Bam*HI, tras lo cual la región codificante *rluc* se ligó a los sitios *Nco*I y *Bam*HI. Los constructos regulados mediante el promotor *ntp303* y que expresaban el gen *luc+* contenían una versión más larga del enlazador sintético que la que se usó en todos los demás constructos. Esta guía sintética de 99 pares de bases de largo se obtuvo mediante RCP utilizando el plásmido pNBL52-44 como molde (una amable aportación del Dr. David Twell). Este fragmento, denominado *syn99* 5' en este artículo, se amplificó usando los siguientes cebadores: 5'-GTGTCTCGAGTTGCAATTGGATCC (sitio *Xho*I subrayado; SEC. ID. N°: 12) y 5'-GTGTCCATGGCCGCGGG (sitio *Nco*I subrayado; SEC. ID. N°: 13). Tras la eliminación del 5' UTR de *ntp303* que se encontraba en 303 5'/35S 3' y en 303 5'/303 3', el 5' UTR de *syn99* fue clonado en los sitios *Xho*I y *Nco*I, creando los constructos *syn99* 5'/35S 3' y *syn99* 5'/303 3' respectivamente.

[0049] Los constructos que contienen modificaciones del 5' UTR de *ntp303* ( $\Delta$  5' UTR) se obtuvieron todos mediante RCP usando el 5' UTR de *ntp303* en el constructo 303 5'/35S 3' como material de inicio. Los fragmentos que se obtuvieron mediante RCP se ordenaron completamente para excluir desajustes en las secuencias. Todos los constructos usados para la expresión transitoria se encontraban en el plásmido pUC19.

### **Bombardeo de microproyectiles**

[0050] Se obtuvieron microportadores, discos de ruptura y macroportadores de Bio-Rad. La preparación y el recubrimiento de los microportadores se realizó según el manual del fabricante (Bio-Rad). Para la transformación biolística del polen bicelular tardío y el polen maduro, se utilizaron partículas de oro de 250  $\mu$ g por bombardeo con un tamaño de 1  $\mu$ m y 1,6  $\mu$ m, respectivamente. Los microsoportes se revistieron con una cantidad total de 1  $\mu$ g de ADN que contenía 0,7  $\mu$ g de ADN de plásmido de ensayo y 0,3  $\mu$ g de ADN de plásmido de normalización. Los plásmidos de ensayo que contenían el promotor *ntp303* y el gen *luc+*, el promotor *CaMV* 35S y el gen *luc+*, y el promotor *ntp303* y el gen *rluc*, fueron coprecipitados con los plásmidos de normalización <sup>R</sup>*syn44* 5'/35S 3', <sup>35S</sup>R<sub>syn44</sub> 5'/35S 3' y *syn44* 5'/35S 3', respectivamente. El *syn44* 5'/35S 3' es idéntico al *syn99* 5'/35S 3', con la excepción de que contiene la guía del SYN44 5' en lugar de la guía del SYN99 5'. El bombardeo de proyectiles se realizó usando el sistema PDS-1000/He operado por helio de Bio-Rad. Para la transformación biolística del polen y las hojas, se utilizaron los siguientes parámetros de bombardeo: una distancia objetivo de 6 cm, una distancia *gap* de ¼ pulgadas, una distancia de macroproyectil/pantalla de parada de 8 mm, una cámara de vacío de 28 mmHg, y una presión de ruptura de los discos de ruptura de 1100 psi.

### **Ensayos de luciferasa**

[0051] Tras el bombardeo de partículas y la incubación de tejidos, se realizó la determinación cuantitativa de la expresión transitoria de los diferentes constructos de fusión de genes usando el Dual-Luciferase™ Reporter Assay System (Promega) disponible en el mercado. En este ensayo, las actividades de las luciferasas LUC+ y RLUC se midieron secuencialmente a partir de un único extracto de muestra usando un luminómetro provisto de dos autoinyectores (Wallac 1420 VICTOR<sup>2</sup>™, EG&G Wallac). La preparación de los tampones usados en el ensayo se realizó según el manual del fabricante (Promega). Tras la incubación, el polen en desarrollo se transfirió a un tubo Greiner de 10 ml y se recogió mediante centrifugado durante 2 minutos a 2.500 rpm. El polen en germinación se recogió mediante centrifugado durante 5 minutos a 1.000 rpm. En todos los casos, el granulado de polen se resuspendió en 100  $\mu$ l 1x de tampón de lisis pasiva (Promega) y triturado en nitrógeno líquido. Los extractos de polen se almacenaron a -70°C hasta ser utilizado para la actividad con el ensayo de la luciferasa. Los extractos (10  $\mu$ l) se pipetearon en una placa de microtitulación y después se colocaron en el luminómetro. Se inyectaron automáticamente 100  $\mu$ l de reactivo de ensayo de luciferasa II (Promega) seguidos, tras 2 segundos, por una cuenta de los fotones durante 10 segundos. Inmediatamente después de la cuantificación de la luminiscencia del LUC+, se apagó la reacción, y la reacción del RLUC se inició automáticamente después de la adición de 100  $\mu$ l del reactivo Stop&Glo® (Promega). La luminiscencia del RLUC también se midió durante 10 segundos, 2 segundos después de la inyección. Para compensar la variabilidad de la expresión de un mismo gen indicador de ensayo entre experimentos independientes, se determinó la proporción del LUC+: RLUC. Para cada constructo, se realizaron al menos seis bombardeos independientes.

### **Ejemplo 1**

#### **Constructos de fusión de genes UTR**

[0052] Se construyeron diferentes constructos de fusión de genes con el promotor *ntp303*, el gen indicador de photinus-luciferina-4-monooxigenasa y diferentes combinaciones de 5'- y 3' UTRs para indagar en la capacidad de los UTRs del *ntp303* de modular la expresión genética durante desarrollo del polen y el crecimiento del tubo de polen (Figura 1). Los nombres de los constructos se refieren a sus 5' UTR y 3' UTR (5' UTR / 3' UTR). La abreviaturas '35S' y 'R' que aparecen en mayúsculas antes del nombre de un constructo indican que el constructo contiene el

promotor CaMV 35S o la región codificante de luciferasa *Renilla* respectivamente.

[0053] Los constructos de fusión de genes UTR se introdujeron mediante bombardeo de partículas dentro del polen en desarrollo o maduro. Su expresión transitoria se midió después de un periodo de 20 horas de desarrollo *in vitro* o tras la germinación mediante mediciones de luminiscencia. Para corregir las diferencias en los rendimientos del bombardeo, se cobombardeó un segundo constructo con el promotor *ntp303*, un 5' UTR sintético (*syn44 5'*), la secuencia de terminación del virus de mosaico de la coliflor (35S 3') y un gen indicador de luciferasa de *Renilla* (<sup>R</sup>*syn44 5'/35S 3'*). El valor de la expresión transitoria del constructo de photinus-luciferina-4-monooxigenasa se normalizó acorde al valor del constructo de luciferasa de *Renilla*.

[0054] El efecto de los UTRs de *ntp303* en la expresión transitoria durante desarrollo del polen y el crecimiento del tubo de polen se investigó comparando el nivel de expresión de un constructo que contenía los UTR de *ntp303* (303 5'/303 3') con aquellos constructos que contenían UTRs de control (el *syn99 5'* o *syn44 5'* UTR y el 35S 3' UTR). Después de 20 horas de crecimiento del tubo de polen, la expresión de 303 5'/303 3' era aproximadamente de 60 y 6 veces superior a la de *syn99 5'/35S 3'* y *syn44 5'/35S 3'* respectivamente (Figura 2b). Las diferencias en el nivel de expresión ya se observaron en los tubos de polen que se bombardearon 5 horas antes. Estas grandes diferencias en el nivel de expresión no se observaron en el polen en estado de desarrollo incubado durante 20 horas después del bombardeo del constructo *ntp303* UTR. Aquí, los UTRs de *ntp303* aumentaron a un nivel de expresión que fue aproximadamente 4 veces superior al de *syn99 5'/35S 3'* UTRs y ligeramente inferior al de *syn44 5'/35S 3'* (Figura 2a). Los niveles de expresión de los constructos que contenían UTRs de control fueron más o menos los mismos durante el desarrollo del polen y el crecimiento del tubo de polen (Figura 2a y 2b). Esto ilustra claramente que la expresión mediada por estos UTRs de control es independiente de la fase de desarrollo en la que se evaluaron.

[0055] Para examinar si el 5' UTR o el 3' UTR del ARNm *ntp303* determina el nivel de la expresión genética durante el desarrollo del polen y el crecimiento del tubo de polen, los niveles de expresión transitoria de los constructos de fusión de genes que contenían el *ntp303 5' / 35S 3'* o los *syn44 5' / ntp303 3'* UTR se compararon con los de *syn44 5' / 35S 3'*. Durante el crecimiento del tubo de polen, el 5' UTR de *ntp303* aumentó la expresión del gen de luciferasa a un nivel que fue casi 8 veces superior al del 5' UTR de control (Figura 3b). Este efecto de aumento no se dio en el constructo 3' UTR de *ntp303*. Durante el desarrollo del polen (Figura 3a), no se observaron diferencias significativas en el nivel de expresión de los UTR de control y el UTR de *ntp303* que contenía constructos .

[0056] Para excluir la posibilidad de que el aumento de la expresión mediado por el 5' UTR de *ntp303* en tubos de polen en crecimiento fuera el resultado de una interacción específica entre el 5' UTR y la región codificante de photinus-luciferina-4-monooxigenasa, la región codificante de photinus-luciferina-4-monooxigenasa se sustituyó por la región codificante de luciferasa *Renilla* en los constructos *syn44 5'/35S 3'* y *303 5'/303 3'*. La photinus-luciferina-4-monooxigenasa y los ARNm de luciferasa *Renilla* no mostraron ninguna identidad de secuencia significativa entre sí. La normalización de la expresión de estos constructos se estableció por cobombardeo con un constructo que contenía el *syn44 5'* UTR, la secuencia de terminación 35S y la región codificante de photinus-luciferina-4-monooxigenasa. Como se muestra en la figura 4, los UTRs de *ntp303* aumentaron a un nivel de expresión que fue aproximadamente 7 veces superior al de los UTR de control. Como este efecto de aumento del 5' UTR de *ntp303* también se dio en los ARNm de photinus-luciferina-4-monooxigenasa, esto excluye la interacción específica entre el 5' UTR de *ntp303* y la región codificante.

## Ejemplo 2

### Aumento traduccional durante el crecimiento del tubo del polen.

[0057] Para averiguar si el aumento mediado por el 5' UTR de *ntp303* durante el crecimiento del tubo de polen fue el resultado de un evento de regulación postranscripcional o de una ruptura en la transcripción (es decir, en regulación transcripcional), se determinó la expresión y los niveles de transcripción de *syn99 5'/35S 3'*, *303 5'/35S 3'* y *syn99 5'/303 3'* (tabla 1). Se hibridizaron 10 µg del ARN total aislado de un extracto de polen, que también se usó para determinar el nivel de la expresión como se midió en la actividad luciferasa, con un 32 sonda de luciferasa catalogada como <sup>32</sup>P. El nivel de transcripción se determinó mediante el cálculo de la proporción de la señal de hibridación del ARNm de photinus-luciferina-4-monooxigenasa con la señal de hibridación del ARNm de luciferasa *Renilla* del constructo cobombardeado <sup>R</sup>*syn44 5'/35S 3'*. Después de 20 horas de crecimiento del tubo de polen, el 5' UTR de *ntp303* mostró un nivel de transcripción relativa que fue aproximadamente 2 veces superior al del *syn99 5'* UTR. El constructo con el 5' UTR de *ntp303* aumentó la expresión relativa 50 veces en comparación con *syn99 5'/35S 3'*. Así, las transcripciones de luciferasa quimérica del crecimiento del tubo de polen con el 5' UTR de *ntp303* se traducen más eficazmente que los ARNm de luciferasa que contienen los 5' UTR de control.

[0058] Tabla 1: análisis de los niveles relativos de la transcripción (representado como abundancia relativa del ARNm luc) y la traducción (presentado como actividad LUC relativa) de diferentes constructos de fusión de genes UTR. Véase la sección de resultados para una descripción de la metodología seguida. Los valores entre paréntesis representan los niveles relativos de la transcripción y la traducción tras la normalización hacia los valores relativos del constructo *syn99 5'/35S 3'*. Se evaluaron las mediciones después de 20 horas del crecimiento del tubo de polen.

	abundancia relativa del ARNm luc	actividad LUC relativa(rlu/10 seg <sup>-1</sup> )
syn99 5'/35S 3'	1,36 SE±0,17 (1,00)	1,02 SE±0,21 (1,00)
303 5'/35S 3'	3,70 SE±0,22 (2,72)	50,60 SE±0,22 (49,61)
syn99 5'/35S 3'	1,81 SE±0,22 (1,33)	2,50 SE±0,42 (2,45)

### Ejemplo 3

El aumento de expresión genética mediado por 5' UTR también ocurre en otros tipos de células además de en los tubos de polen en crecimiento

5 [0059] Para evaluar si el aumento de expresión en tubos de polen en crecimiento mediado en el 5' UTR de *ntp303* estaba restringida a un entorno específico de polen, se reconstruyeron los constructos syn44 5'/35S 3' y 303 5'/303 3' sustituyendo su promotor *ntp303* por el promotor *CaMV 35S*. El promotor *CaMV 35S* es casi inactivo en los tubos de polen, pero es muy activo en tejidos esporofíticos. Después del bombardeo de partículas de estos constructos en polen maduro y hojas jóvenes seguido de 20 horas de incubación *in vitro*, se ensayó la expresión transitoria. La normalización de la expresión de estos constructos se hizo con el nivel de expresión de un constructo cobombardeado que contenía el promotor *CaMV 35S*, el 5' UTR de syn44, el gen indicador de luciferasa *Renilla* y el 3' UTR de 35S. En los tubos de polen en crecimiento, el 5' UTR de *ntp303* aumentó la expresión transitoria hacia un nivel que fue aproximadamente 5 veces superior a los UTR de control (Figura 5a). Las diferencias en el nivel de expresión se acercaron a las de los constructos que contenían las mismas combinaciones de UTR pero que estaban vinculadas al promotor *ntp303* (compare las Figuras 5a y 2b). En las hojas jóvenes, el 5' UTR de *ntp303* aumentó la expresión a un nivel que fue aproximadamente 2 veces superior al de los UTR de control (Figura 5b). Estos datos demuestran que el aumento de la expresión mediado por 5' UTR de *ntp303* también ocurre en otros tipos de célula además de tubos de polen en crecimiento, tales como las células esporofíticas. No obstante, el aumento de expresión mediado por 5' UTR de *ntp303* es máximo en tubos de polen en crecimiento.

### Ejemplo 4

El aumento de la expresión durante el crecimiento del tubo de polen es atribuible a regiones específicas dentro del 5' UTR de *ntp303*

[0060] La Figura 6 ilustra la estructura secundaria prevista del 5' UTR de *ntp303* tal como se analiza con el paquete de software RNAdraw (Hofacke *et al.*, (1994) Chem. Monthly 125, 167-188). Hay dos estructuras putativas en horquilla denominadas H-I (nucleótidos 4-76 de la SEC. ID. N°: 1) y H-II (nucleótidos 104-151 de la SEC. ID. N°: 1). La estructura en horquilla H-I se localiza en el 5' terminal y tiene un valor de energía calculado ( $\Delta G$ ) de -64 kJ/mol. Esta estructura contiene ocho repeticiones de un triplete de GAA (nucleótidos 27-50 de la SEC. ID. N°: 1) en el bucle externo y un ARN bicatenario en el vástago de la estructura prevista H-I, que consiste en GAAGAAGA (14-21) y en la cadena complementaria TCTTCTTC (59-66). La estructura H-II está localizada 22 nucleótidos hacia arriba del sitio de iniciación de la traducción y tiene un valor de energía calculado ( $\Delta G$ ) de -26 kJ/mol. El efecto de las secuencias que residen en las estructuras H-I y H-II en el aumento de la expresión durante el crecimiento del tubo de polen se analizó mediante una serie de constructos de delección del 5' UTR de *ntp303* (Figura 7a y c). Estos constructos fueron bombardeados dentro del polen maduro y su expresión fue evaluada después de 20 horas de crecimiento del tubo de polen (Figura 7b y d). Se consiguió una inactivación casi completa de la expresión del gen indicador tras la delección de los últimos 70 nucleótidos del 3' terminal del 5' UTR de *ntp303* que incluía la estructura completa de H-II ( $\Delta 70$  303 5'/35S 3') (Figura 7b). Lo mismo ocurrió después de la delección interna de tan sólo la estructura de H-II ( $\Delta$ H-II 303 5'/35S 3') (Figura 7d). En ambos casos, los valores de la expresión se encontraban en el mismo intervalo que los valores precedentes (es decir, la autoluminiscencia medida del sustrato de luciferina). La Figura 7b muestra la expresión transitoria de los constructos de fusión de genes con delecciones en la estructura de horquilla H-I. El nivel mínimo de expresión transitoria se encontró después de la delección interna de la repetición (GAA)<sub>8</sub> ( $\Delta$ GAA 303 5'/35S 3'). Este nivel de expresión era comparable al nivel de expresión del constructo de control que contenía la guía del syn99 (datos no mostrados). Se dio una reducción de la expresión transitoria de aproximadamente el 94% después de la delección de los primeros 55 nucleótidos ( $\Delta 55$  303 5'/35S 3') en el 5' terminal que incluía la repetición (GAA)<sub>8</sub>. La delección de los primeros 29 nucleótidos en el 5' terminal del 5' UTR de *ntp303* ( $\Delta 29$  303 5'/35S 3') provocó sólo una leve reducción en la expresión transitoria en comparación con aquel del 5' UTR de *ntp303* sin modificar. Estos resultados claramente demuestran que las transcripciones de luciferasa que contienen delecciones dentro del 5' UTR de *ntp303* se expresan en un nivel inferior al de las transcripciones del 5' UTR de *ntp303* sin modificar. Las delecciones que están dentro de la estructura H-I o H-II o que pertenecen a las mismas provocaron una reducción diferente en la expresión transitoria. Se observó la ausencia del efecto de aumento después de la delección de la repetición (GAA)<sub>8</sub>, mientras que la delección de la totalidad de la estructura H-II provocó un completo colapso en la expresión.

**Ejemplo 5**Las estructuras H-I y H-II del 5' UTR de ntp303 influyen en la eficiencia de la traducción

[0061] Mediante la medición de los niveles relativos de transcripción y traducción de algunos de los constructos de delección del 5' UTR de *ntp303* (Tabla 2), se investigó si la reducción en la expresión transitoria de los constructos de delección del 5' UTR fue resultado de un cambio en la eficiencia de transcripción o de traducción. Los niveles de transcripción relativos de los constructos que contienen delecciones de la estructura H-II completa ( $\Delta 70$  303 5'/35S 3' y  $\Delta$ H-II 303 5'/35S 3') se redujeron a un nivel inferior al de *syn99* 5'/35S 3'. La delección interna de la repetición (GAA)<sub>8</sub> ( $\Delta$ GAA 303 5'/35S 3') resultó en un nivel relativo de transcripción inferior al nivel de transcripción del constructo que contenía el 5' UTR de *ntp303* sin modificar, pero los niveles de transcripción relativos se mantuvieron superiores al del constructo que contenía el 5' UTR sintético. A diferencia de los efectos de la delección de la repetición (GAA)<sub>8</sub> o de la estructura H-II en el nivel de transcripción relativa, se observó un efecto más drástico en los niveles de traducción relativos. Se observó una reducción drástica en el nivel de traducción relativa después de la delección de la estructura H-II, los valores del nivel de traducción normalizado se encontraban dentro de la escala de valores de los antecedentes. La delección de la repetición (GAA)<sub>8</sub> reveló un nivel de traducción relativo 2 veces inferior en comparación con 303 5'/303 3'. De estos datos, se llega a la conclusión de que la reducción observada en la expresión después de la delección de las estructuras H-I y H-II es principalmente el resultado de una reducción en la eficiencia de traducción. El efecto más grave de la reducción de la eficiencia de traducción se encontró después de la delección de la estructura H-II.

[0062] Tabla 2: análisis de los niveles relativos de la transcripción (presentado como abundancia relativa del ARNm luc) y la traducción (presentado como actividad LUC relativa) de diferentes constructos de fusión de genes *ntp303* 5' UTR. Véase la sección de resultados para una descripción de la metodología seguida. Los valores entre paréntesis representan los niveles relativos de la transcripción y la traducción tras la normalización hacia los valores relativos del constructo *syn99* 5'/35S 3'. Se evaluaron las mediciones después 20 horas de crecimiento del tubo de polen.

	abundancia relativa del ARNm (cuentas)	actividad LUC relativa (url/10 seg <sup>-1</sup> )
303 5'/303 3'	2,97 SE±0,09 (2,18)	31,08SE±5,61 (30,47)
$\Delta$ AAG 303 5'/35S 3'	2,30 SE±0,51 (1,69)	17,49 SE±4,76 (17,15)
$\Delta$ 70 303 5'/35S 3'	1,13 SE±0,06 (0,83)	0,01 SE±0,00 (0,01)
$\Delta$ H-II 303 5'/35S 3'	0,82 SE±0,09 (0,60)	0,01 SE±0,00 (0,01)

**Descripción de las figuras**

[0063]

Figura 1: representación esquemática y nombres de los constructos de fusión de genes UTR usados en el presente estudio.

Figura 2: expresión transitoria de los constructos de fusión de genes UTR del polen en desarrollo (A) y tubos de polen en crecimiento (B). Url/10 seg<sup>-1</sup> son las unidades relativas de luz por cada 10 segundos de tiempo de medición según se determina mediante la normalización de la expresión absoluta del constructo de prueba con la del constructo de referencia.

Figura 3: expresión transitoria de los constructos de fusión de genes UTR de polen en desarrollo (A) y tubos de polen en crecimiento (B). Url/10 seg<sup>-1</sup> son las unidades relativas de luz por cada 10 segundos de medición de segundos según se determina mediante la normalización de la expresión absoluta del constructo de prueba con la del constructo de referencia.

Figura 4: expresión transitoria de constructos de fusión de genes UTR que contienen la región codificante de luciferasa *Renilla* en tubos de polen en crecimiento. Url/10 seg<sup>-1</sup> son las unidades relativas de luz por cada 10 segundos de tiempo de medición según se determina mediante la normalización de la expresión absoluta del constructo de prueba con la del constructo de referencia.

Figura 5: expresión transitoria de los constructos de fusión de genes UTR que contienen el promotor CaMV 35S de los tubos de polen en crecimiento (A) y hojas jóvenes (B). Url/10 seg<sup>-1</sup> son las unidades relativas de luz por cada 10 segundos de tiempo de medición según se determina mediante la normalización de la expresión absoluta del constructo de prueba con la del constructo de referencia.

Figura 6: estructura secundaria prevista del 5' UTR de *ntp303*. La predicción de la estructura y el cálculo del valor  $\Delta$ G se realizó usando el paquete de software de RNA draw. H-I y H-II representan dos estructuras en horquilla previstas. Véase la sección de resultados para la descripción del 5' UTR.

Figura 7: representaciones esquemáticas y efectos de la expresión de las mutaciones de *ntp303* 5' UTR H-I (A y C) y H-II (B y C). Véase la sección de resultados para una descripción detallada de los 5' UTR de *ntp303* modificados.  $\text{Url}/10 \text{ seg}^{-1}$  son las unidades relativas de luz por cada 10 segundos de tiempo de medición según lo determina la normalización de la expresión absoluta del constructo de prueba con la del constructo de referencia. Se evaluaron las mediciones después de 20 horas de crecimiento del tubo de polen.

## LISTADO DE SECUENCIAS

[0064]

- 5
- <110> Universidad católica de Nijmegen
- 10
- <120> Regulación de la traducción de genes heterológamente expresados
- <130> Regulación de la traducción
- 15
- <140>
- <141>
- <160> 13
- 20
- <170> Patente en Ver. 2.1
- <210> 1
- <211> 173
- <212> ADN
- 25
- <213> *Nicotiana tabacum*
- <220>
- <221> rasgos variados
- <222> (4) .. (76)
- 30
- <223> Elemento de regulación de la traducción
- <220>
- <221> rasgos variados
- <222> (104)..(151)
- 35
- <223> Elemento de regulación de la traducción
- <220>
- <221> región de repetición
- <222> (27) .. (50)
- 40
- <223> repetición-GAA
- <400> 1

```

caagctctag caggaagaag aaataagaag aagaagaaga agaagaagaa gcgtctcctc 60
ttcttcttgt gagagtaaaa aataaaactc ccaaaaaaaaa gaaatcatc aaaaaaaca 120
atttcaaaaa gagtttttgt gtttggggat taaagaataa aaaaaacaac gtc      173

```

```

5
    <210> 2
    <211> 18
    <212> ADN
    <213> Secuencia Artificial

    <220>
    <223> Descripción de la Secuencia Artificial: Cebador

10
    <400> 2
    atatccatgg aagacgcc      18

    <210> 3
15
    <211> 23
    <212> ADN
    <213> Secuencia Artificial

    <220>
20
    <223> Descripción de la Secuencia Artificial: Cebador

    <400> 3
    atatggatcc ttacacggcg atc      23

25
    <210> 4
    <211> 23
    <212> ADN
    <213> Secuencia Artificial

30
    <220>
    <223> Descripción de la Secuencia Artificial: Cebador

    <400> 4
35
    ggtccatgg atgacttcga aag      23

    <210> 5
    <211> 28
    <212> ADN
    <213> Secuencia Artificial

```



<220>  
 <223> Descripción de la Secuencia Artificial: Cebador

5 <400> 5  
 gtgtgatcc ttattgtca ttttgag 28

<210> 6  
 <211> 27

10 <212> ADN  
 <213> Secuencia Artificial

<220>  
 <223> Descripción de la Secuencia Artificial: Cebador

15 <400> 6  
 gtgtctcgag caagctctag caggaag 27

<210> 7  
 <211> 29

20 <212> ADN  
 <213> Secuencia Artificial

<220>  
 <223> Descripción de la Secuencia Artificial: Cebador

25 <400> 7  
 gtgtccatgg gacgttgttt ttttattc 29

30 <210> 8  
 <211> 30  
 <212> ADN  
 <213> Secuencia Artificial

35 <220>  
 <223> Descripción de la Secuencia Artificial: Cebador

<400> 8  
 atatgatcc attctgtaat gatcaatctg 30

40 <210> 9  
 <211> 28

<212> ADN  
 <213> Secuencia Artificial  
  
 <220>  
 5 <223> Descripción de la Secuencia Artificial: Cebador  
  
 <400> 9  
 atatgagctc attaatggt ttgccta 28  
  
 10 <210> 10  
 <211> 30  
 <212> ADN  
 <213> Secuencia Artificial  
  
 15 <220>  
 <223> Descripción de la Secuencia Artificial: Cebador  
  
 <400> 10  
 atataagctt gatacactcg caacgtgtgt 30  
  
 20 <210> 11  
 <211> 30  
 <212> ADN  
 <213> Secuencia Artificial  
  
 25 <220>  
 <223> Descripción de la Secuencia Artificial: Cebador  
  
 <400> 11  
 30 atatctcgag gagcttgac tattcacat 30  
  
 <210> 12  
 <211> 24  
 <212> ADN  
 35 <213> Secuencia Artificial  
  
 <220>  
 <223> Descripción de la Secuencia Artificial: Cebador  
  
 <400> 12  
 40 gtgtctcgag ttgcaattgg atcc 24

<210> 13

<211> 17

<212> ADN

<213> Secuencia Artificial

5

<220>

<223> Descripción de la Secuencia Artificial: Cebador

<400> 13

10

gtgtccatgg ccgcggg

17

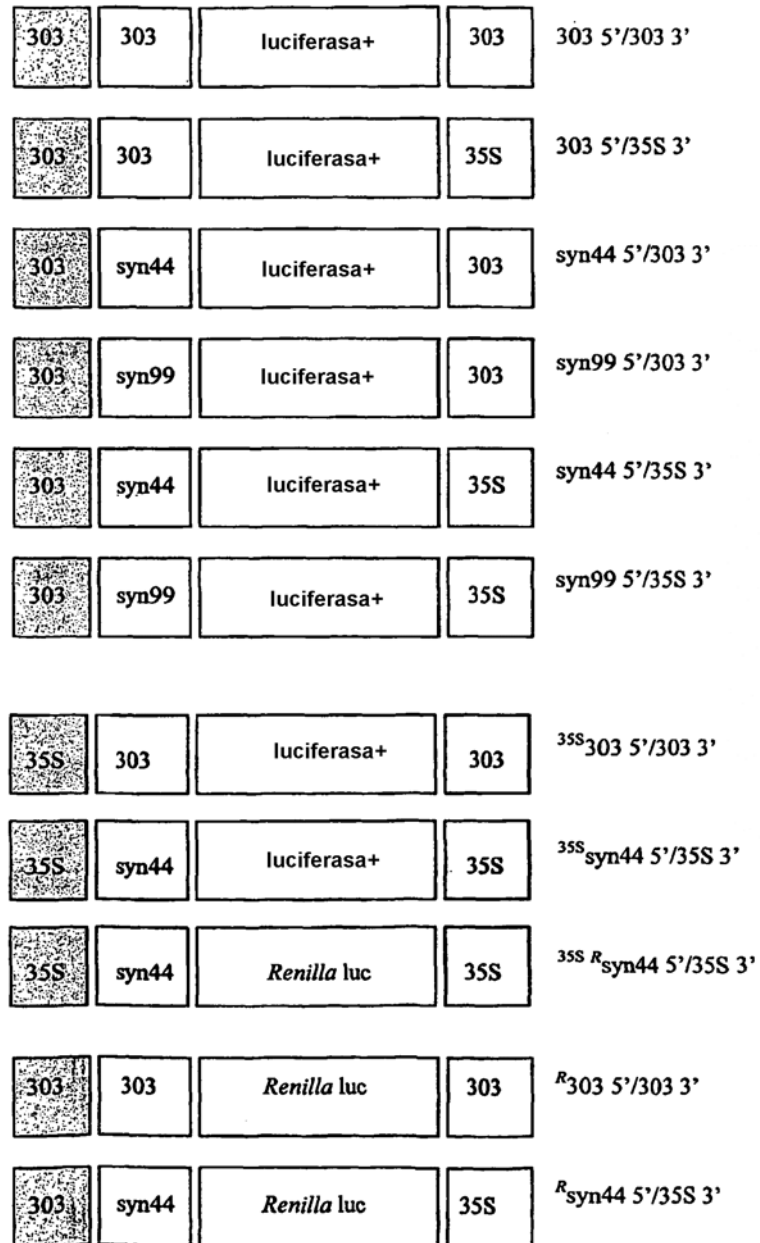
## REIVINDICACIONES

1. Método para la regulación de la traducción de una segunda secuencia de nucleótidos que codifica una proteína en una célula, en el que la célula es una célula vegetal, fúngica, bacteriana, animal o de mamífero no humana, y que incluye las etapas de:
  - 5 a) proporcionar un constructo de ácidos nucleicos con una primera secuencia de nucleótidos que tiene al menos un 50% de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N°: 1, operativamente vinculado a dicha segunda secuencia de nucleótidos que codifica una proteína y, además, está operativamente vinculado a un promotor,
  - 10 b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada, y
  - c) someter dicha célula transformada a condiciones que llevan a la expresión de la proteína.
2. Método para la regulación de la traducción de una segunda secuencia de nucleótidos que codifica una proteína en una célula, en el que la célula es una célula vegetal, fúngica, bacteriana, animal o de mamífero no humana, y que incluye las etapas de:
  - 15 a) proporcionar un constructo de ácidos nucleicos con una primera secuencia de nucleótidos que comprende una secuencia de nucleótidos que tiene al menos un 50% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N°: 1, o una secuencia de nucleótidos que tiene al menos un 55 % de identidad de secuencia de nucleótidos con los nucleótidos 4 - 76 de la secuencia de nucleótidos de la SEC. ID. N°: 1, o una combinación de los
  - 20 mismos, operativamente vinculado a dicha segunda secuencia de nucleótidos que codifica una proteína, y además está operativamente vinculado a un promotor,
  - b) poner en contacto una célula con dicho constructo de ácidos nucleicos para obtener una célula transformada, y
  - c) someter dicha célula transformada a condiciones que llevan a la expresión de la proteína.
- 25 3. Método según la reivindicación 2, en el que la primera secuencia de nucleótidos comprende una secuencia de nucleótidos que tiene al menos un 50% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N°: 1 o una secuencia de nucleótidos que tiene un 100% de identidad de secuencia de nucleótidos con los nucleótidos 27 - 50 de la secuencia de nucleótidos de la SEC. ID. N°: 1 o una combinación de la misma.
- 30 4. Método según cualquiera de las reivindicaciones 1 a 3, en el que la proteína es una proteína heteróloga.
5. Método según cualquiera de las reivindicaciones 1 a 4, en el que la proteína se expresa en una célula vegetal transgénica homocigota doble haploide de *Nicotiana tabacum* silenciada para *ntp303*.
6. Método según cualquiera de las reivindicaciones 1 a 5, en el que la proteína se expresa en el polen o la semilla de dicha planta.
- 35 7. Método según cualquiera de las reivindicaciones 1 a 6, en el que la primera secuencia de nucleótidos comprende la secuencia de nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N°: 1.
8. Método según cualquiera de las reivindicaciones 1 a 7, en el que la primera secuencia de nucleótidos comprende la secuencia de nucleótidos 4 - 76 de la secuencia de nucleótidos de la SEC. ID. N°: 1.
9. Método según cualquiera de las reivindicaciones 1 a 8, en el que la secuencia de nucleótidos comprende la
- 40 secuencia de nucleótidos 27 - 50 de la secuencia de nucleótidos de la SEC. ID. N°: 1.
10. Método según cualquiera de las reivindicaciones 1 a 9, en el que la primera secuencia de nucleótidos comprende la secuencia de la SEC. ID. N°:1.
11. Planta que es una planta de *Nicotiana tabacum* transgénica, homocigótica, doble haploide y silenciada para *ntp303*, en donde dicha planta comprende un constructo de ácidos nucleicos con una primera secuencia de
- 45 nucleótidos que tiene al menos un 50% de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N°: 1, operativamente vinculada a una segunda secuencia de nucleótidos que codifica una proteína.
12. Planta que es una planta de *Nicotiana tabacum* transgénica, homocigótica, doble haploide y silenciada para *ntp303*, en donde dicha planta comprende un constructo de ácidos nucleicos con una primera secuencia de
- 50 nucleótidos que tiene al menos un 50% de identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N°: 1, o una secuencia de nucleótidos que tiene al menos un 55% de identidad de secuencia de nucleótidos con los nucleótidos 4 - 76 de la secuencia de nucleótidos de la SEC. ID. N°: 1, o una combinación de las mismas, operativamente vinculada a una segunda secuencia de nucleótidos que codifica una proteína.
- 55 13. Planta según la reivindicación 12, donde la primera secuencia de nucleótidos comprende una secuencia de

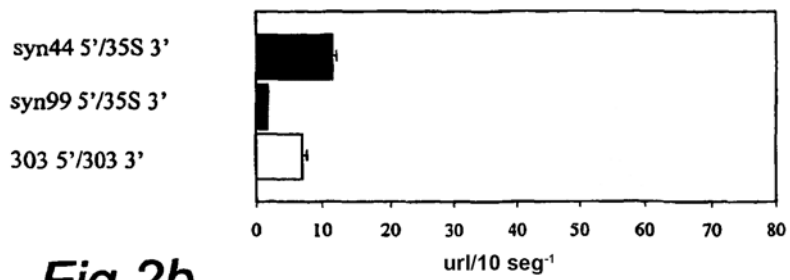
nucleótidos que tiene al menos un 50% identidad de secuencia de nucleótidos con los nucleótidos 104 - 151 de la secuencia de nucleótidos de la SEC. ID. N°: 1, o una secuencia de nucleótidos que tiene un 100% de identidad de secuencia de nucleótidos con los nucleótidos 27 - 50 de la secuencia de nucleótidos de la SEC. ID. N°: 1 o una combinación de las mismas.

- 5     **14.** Planta según cualquiera de las reivindicaciones 11 a 13, en donde la segunda secuencia de nucleótidos codifica una proteína heteróloga.
- 15.** Material de propagación de una planta según cualquiera de las reivindicaciones 11 a 14.
- 16.** Material de cosecha de una planta según cualquiera de las reivindicaciones 11 a 14.
- 17.** Tejido de una planta según cualquiera de las reivindicaciones 11 a 14.
- 10    **18.** Constructo de ácido nucleico con una primera secuencia de nucleótidos que tiene al menos un 50 % de identidad de secuencia de nucleótidos con la secuencia de nucleótidos de la SEC. ID. N°: 1, operativamente vinculado a una segunda secuencia de nucleótidos que codifica una proteína.
- 19.** Constructo de ácido nucleico según la reivindicación 18, en el que la secuencia de nucleótidos que codifica la proteína es posteriormente se vincula operativamente a un promotor.
- 15    **20.** Constructo de ácido nucleico según las reivindicaciones 18 y 19, en el que la primera secuencia de nucleótidos comprende la secuencia de SEC. ID. N°:1.
- 21.** Célula huésped recombinante animal no humano, mamífero no humano, vegetal, bacteriana o fúngica con una o más copias del constructo de ácidos nucleicos tal y como se define en cualquiera de las reivindicaciones 18 a 20.
- 20    **22.** Uso de la primera secuencia de nucleótidos tal y como se define en cualquiera de las reivindicaciones 1 a 10 para la expresión de una proteína y, opcionalmente, para la recuperación de la proteína expresada.
- 23.** Uso según la reivindicación 22, en el que la primera secuencia de nucleótidos se usa en una célula vegetal, fúngica, bacteriana, animal o de mamífero no humana.
- 25    **24.** Uso según la reivindicación 23, en el que una primera secuencia de nucleótidos se usa en un sistema de traducción acelular utilizando ARNs derivados de un constructo de ácido nucleico según cualquiera de las reivindicaciones 18-20.

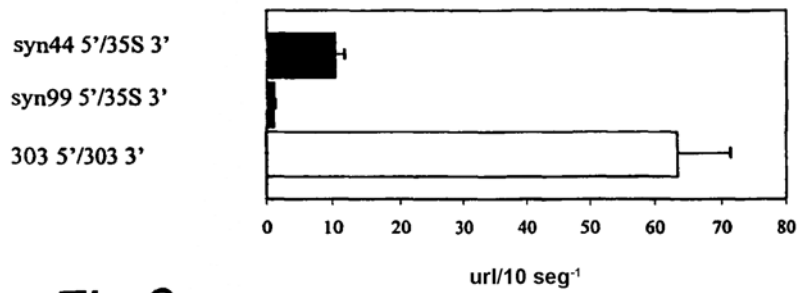
**Fig 1**



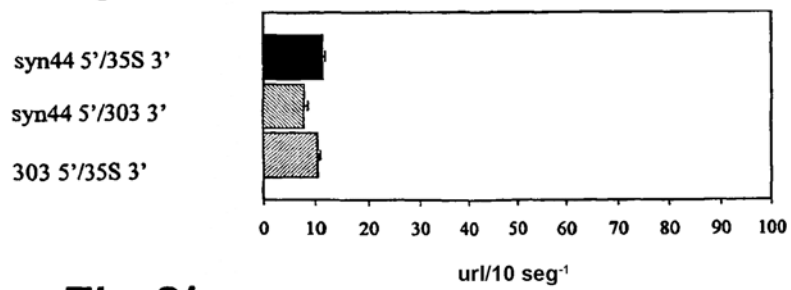
**Fig 2a**



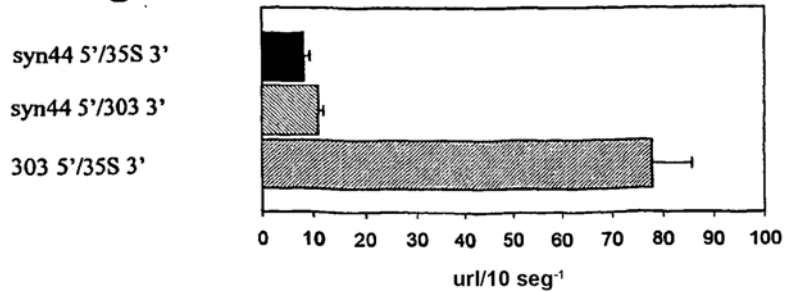
**Fig 2b**



**Fig 3a**

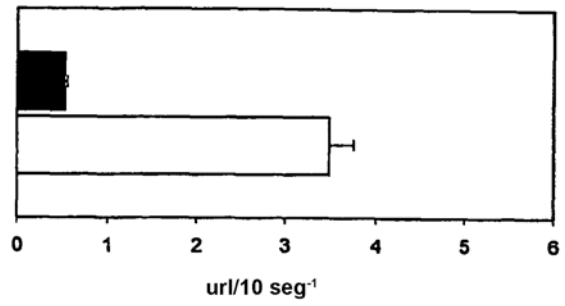


**Fig 3b**



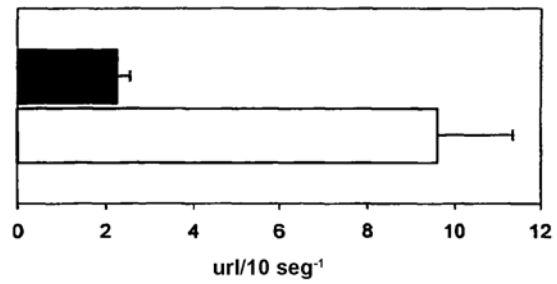
**Fig 4**

$R_{\text{syn44 5'}/35\text{S 3'}}$   
 $R_{\text{303 5'}/303 3'}$



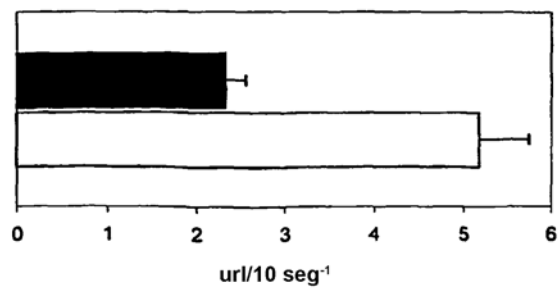
**Fig 5a**

$^{35}\text{S}_{\text{syn44 5'}/35\text{S 3'}}$   
 $^{35}\text{S}_{\text{303 5'}/303 3'}$



**Fig 5b**

$^{35}\text{S}_{\text{syn44 5'}/35\text{S 3'}}$   
 $^{35}\text{S}_{\text{303 5'}/303 3'}$

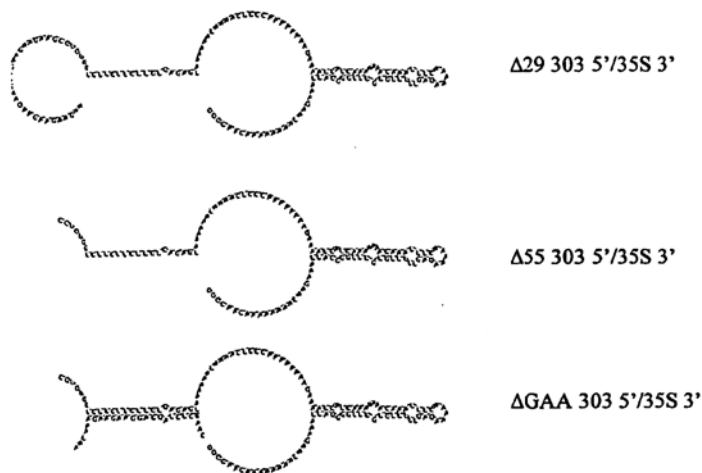




**Fig 6**

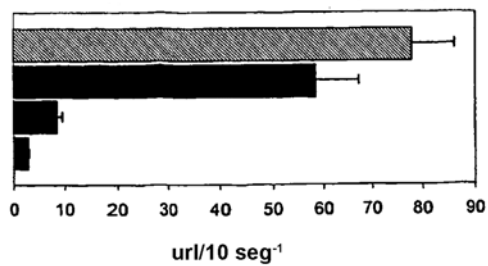


**Fig 7a**

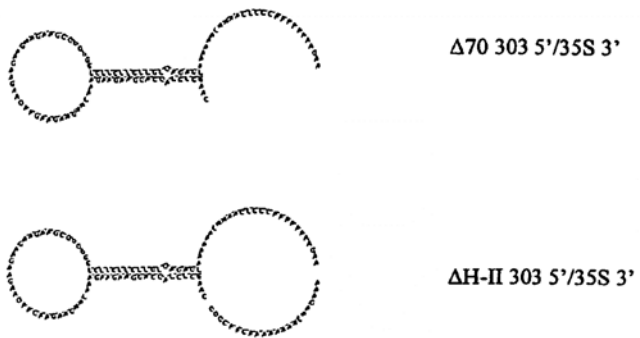


**Fig 7b**

303 5'/35S 3'  
 Δ29 303 5'/35S 3'  
 Δ55 303 5'/35S 3'  
 ΔGAA 303 5'/35S 3'



*Fig 7c*



*Fig 7d*

antecedentes  
 $\Delta 70$  303 5'/35S 3'  
 $\Delta H-II$  303 5'/35S 3'

