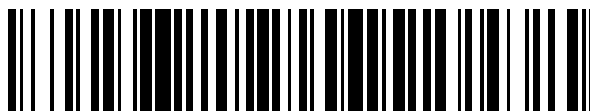


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 627 490**

51 Int. Cl.:

**C12N 15/82** (2006.01)

**C12N 15/113** (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.01.2014 PCT/IB2014/058289**

87 Fecha y número de publicación internacional: **24.07.2014 WO14111858**

96 Fecha de presentación y número de la solicitud europea: **15.01.2014 E 14707214 (4)**

97 Fecha y número de publicación de la concesión europea: **01.03.2017 EP 2946017**

54 Título: **Secuencia de ADN artificial con función de líder optimizada en 5' (5'-UTR) para la sobreexpresión de proteínas recombinantes en plantas y método para la producción de proteínas recombinantes en plantas**

30 Prioridad:

**16.01.2013 IT UD20130002**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**28.07.2017**

73 Titular/es:

**ADIENNE PHARMA & BIOTECH SA (100.0%)  
Via Zurigo 46  
6900 Lugano, CH**

72 Inventor/es:

**MARCHETTI, STEFANO;  
PATTI, TAMARA y  
SECCO, ERIKA**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 627 490 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**

Secuencia de ADN artificial con función de líder optimizada en 5' (5'-UTR) para la sobreexpresión de proteínas recombinantes en plantas y método para la producción de proteínas recombinantes en plantas

**Campo de la invención**

5 La presente invención se refiere a una secuencia de ADN artificial con función de líder optimizada en 5' (5'-UTR) para la sobreexpresión de proteínas recombinantes en plantas, y a un método para la producción de proteínas recombinantes en plantas.

**Antecedentes de la invención**

10 Existen muchas aproximaciones que se pueden adoptar para mejorar la expresión de genes heterólogos en plantas. De hecho, todos los elementos que constituyen un gen ejercen, o pueden ejercer, una función de control sobre la expresión del gen, modulando el proceso de transcripción y/o traducción. Las secuencias sin traducir presentes en los extremos 5' y 3' del mRNA (denominadas 5'-UTR y 3'-UTR, en las que UTR significa "región sin traducir") no son una excepción, y de hecho se deben considerar objetivos preferentes para modificaciones adecuadas, ya que, en gran medida, determinan la eficacia de la traducción y la renovación del propio mRNA. De hecho, numerosas pruebas demuestran que:

- la estructura m7Gppp (caperuza 5') presente en el extremo 5' terminal del mRNA es esencial para incorporar el complejo eIF4F, capaz de unirse a la subunidad ribosómica 40S (Franks y Likke-Andersen, 2008);

- por medio del componente eIF4G, el complejo eIF4F interacciona con la cola de poli(A) presente en el extremo 3' terminal del mRNA, lo que permite a este último adquirir una estructura circular (Franks y Likke-Andersen, 2008);

20 - la cola de poli(A) y el complejo eIF4F reducen la hidrólisis enzimática de la estructura de la caperuza 5', y por lo tanto previenen la degradación rápida del mRNA mediante exonucleasas citoplasmáticas activas en los extremos 5' terminales de monofosfato (Franks y Likke-Andersen, 2008);

25 - la secuencia 5'-UTR puede contener elementos capaces de influir en la formación de la estructura de la caperuza 5', la unión de esta última con el factor eIF4E, la incorporación de la subunidad ribosómica 40S, la constitución de polisomas, la velocidad de disociación espontánea del complejo 43S, el reconocimiento del auténtico codón de inicio de la traducción AUG;

- la secuencia 5'-UTR puede contener además secuencias que representan sitios de unión al ADN para factores de transcripción específicos, y por lo tanto puede modificar la actividad de transcripción de los promotores que están en posición anterior.

30 Por lo tanto, es evidente que es necesario considerar especialmente la 5'-UTR, también denominada región líder, en los programas de ingeniería vegetal para incrementar el nivel de expresión de las proteínas recombinantes.

Sin embargo, por diversas razones, no es sencillo en absoluto diseñar secuencias líder de eficacia elevada, incluso para una persona experta en la técnica. En primer lugar, se debe considerar la gran variabilidad de la secuencia observable entre regiones líder de diferentes genes que pertenecen al mismo genoma o a genomas relacionados.

35 Esta variabilidad hace muy difícil identificar tramos potenciales capaces de conferir una característica mejorada al líder, y prácticamente imposible predecir posibles interacciones con otros elementos o secuencias que constituyen la región 5'-UTR. En segundo lugar, la longitud total de la región líder debe estar contenida posiblemente en 100-120 pb, preferiblemente 80 pb, para no incrementar la frecuencia de la disociación espontánea del complejo 43S de la propia región. Esto impone una elección estricta de los componentes que se usarán realmente en la construcción del tramo líder, en detrimento de otros. En tercer lugar, la región líder no debería contener secuencias palindrómicas o una composición de nucleótidos rica en G/C, para impedir la formación de estructuras secundarias en el transcrito que no se pueden separar por medio de la intervención de eIF4A. Finalmente, una porción minoritaria, pero en cualquier caso significativa, de la secuencia (alrededor del 10%) no puede variar libremente, sino que debe contener elementos funcionales esenciales, tales como, de manera específica, el sitio iniciador *Inr* y el motivo Kozak o un motivo similar a Kozak equivalente.

45 La solicitud WO 2008/080954 describe la combinación de elementos CAA repetidos con elementos CT repetidos en secuencias 5'-UTR utilizables para incrementar la expresión de proteínas recombinantes en plantas. Además, también describe la co-presencia de poli(CAA) y poli(CT) con el sitio iniciador de la transcripción (*Inr*) del promotor 35S de CaMV, es decir, el virus del mosaico de la coliflor (Guilley et al., 1982) y/o con el octámero ACAATTAC del líder  $\Omega$  de TMV (Gallie y Walbot, 1992). De hecho, el documento WO 2008/080954 describe una secuencia líder denominada LLTCK que contiene, por ejemplo, todos los elementos citados anteriormente:

1. Sitio *Inr* del gen 35S de CaMV para la formación eficaz de la caperuza de mRNA;

2. Región poli(CAA) similar al "potenciador traduccional" presente en el líder  $\Omega$  de TMV (Gallie y Walbot, 1992);

3. Secuencia rica en elementos CT, similar a algunos líderes de plantas (Bolle et al., 1996);

4. Octámero del líder  $\Omega$  de TMV.

5 El efecto del líder LLTCK del documento WO 2008/080954 se estudió en tabaco, mediante el uso del líder del gen 35S de CaMV como comparación, que está presente en un gran número de vectores comerciales, determinando los niveles de expresión del gen indicador *uidA* (que codifica la enzima  $\beta$ -glucuronidasa, GUS) bajo control del promotor constitutivo 35S de CaMV. El líder LLTCK provocó un incremento de la concentración de la enzima GUS igual a 8-12 veces la del líder de control.

Sin embargo, existe la necesidad de incrementar adicionalmente la eficacia del tramo 5'-UTR para la expresión de transgenes, y por lo tanto de proteínas recombinantes, en plantas.

10 En particular, para incrementar adicionalmente la eficacia del tramo 5'-UTR para la expresión de transgenes en plantas comparado con el estado de la técnica, considerando que LLTCK es el único líder sintético de eficacia elevada del que se conocen los efectos sobre los procesos de transcripción y traducción de la información genética, puede ser útil considerar este líder como modelo o punto de partida para las intervenciones para mejorarlos.

15 Como se ha dicho, el documento WO 2008/080954 proporciona la posibilidad de combinar elementos CAA repetidos con elementos CT repetidos, e identifica una serie de factores capaces de hacer más evidente la ventaja de dicha combinación.

20 Hay asociada una aplicación preferente a cada factor. Se da una importancia particular a la presencia del motivo octamérico ACAATTAC albergado por el líder  $\Omega$  de TMV; de hecho, según el documento WO 2008/080954, se puede obtener un líder eficaz de la unión de tramos del líder  $\Omega$  de TMV con una región que alberga motivos CT repetidos.

Dentro del líder  $\Omega$  conocido del documento WO 2008/080954, se pueden observar secuencias repetidas de diferentes tipos: una secuencia se representa mediante el trinucleótido CAA repetido 11 veces, aunque no siempre de manera contigua; la otra secuencia se representa mediante el motivo octamérico ACAATTAC repetido 3 veces.

25 Se ha demostrado experimentalmente que ambas secuencias pueden provocar un gran incremento de la expresión génica, y actúan a nivel postranscripcional.

Aunque el octámero contiene un trinucleótido CAA, el aumento de la expresión génica está asociado a la presencia de la secuencia completa, y no de CAA solo.

Es importante subrayar que el octámero contiene un tramo rico en A/T, es decir, AATTA, que a su vez incluye el triplete ATT.

30 Como posible solución técnica preferente, los inventores del documento WO 2008/080954 indican mantener la secuencia octamérica ACAATTAC, incluso si ésta contiene la secuencia AATTA, y por lo tanto un sitio de inicio de la traducción ATT no canónico.

35 Obviamente, creen que la inclusión del motivo octamérico mencionado anteriormente es más importante, incluso si esto implica la introducción de una secuencia rica en A/T y con un supuesto codón de inicio de la traducción. Se debe subrayar que en la ID de secuencia n° 1 (LLTCK) del documento WO 2008/080954, se indican de manera específica otras secuencias ricas en A/T, colocadas respectivamente:

1. inmediatamente después del sitio iniciador (TATTTTTA);
2. dentro del tramo de poli(CAA) (AATA);
3. al final de este tramo, en un sitio que implica de nuevo al octámero (ATTA);
- 40 4. justo después del octámero (TATTT).

Tres de cuatro secuencias portan el triplete ATT, como el octámero.

A continuación se proporciona, como comparación, la secuencia líder LLTCK conocida, destacando las regiones ricas en A/T (subrayadas) y los tripletes ATT (caracteres mayores); el tramo ACAATTAC en negrita corresponde al motivo del octámero:

45 ACACGTATTTTTACAACAAATACCAACAACAACAACAACAACAACA  
TTACAATTACCGTATTTCTCTCTCTAG

También se subraya que esta secuencia LLTCK conocida no proporciona ninguna región poli(CAA) contigua a la región poli(CT).

En este caso también, aunque son conscientes de la presencia de sitios de inicio de la traducción no canónicos en las regiones ricas en A/T, los inventores del documento WO 2008/080954 han proporcionado el uso de dichas regiones en la construcción de un líder eficaz como LLTCK.

5 De hecho, las secuencias ricas en A/T, de manera específica los tipos 1 y 4 como se describió anteriormente, se hallan no solamente en el líder  $\Omega$  de TMV, sino también en el centro del líder de AMV usado habitualmente como potenciador de la traducción como alternativa a  $\Omega$ .

A continuación, como comparación, se proporcionan las secuencias del líder  $\Omega$  de TMV (a) y del líder de AMV (b), destacando las regiones ricas en A/T (subrayadas) y los tripletes ATT (caracteres mayores):

(a)

ACCTCGAGTATTTTTACAACAATTACCAACAACAACAACACAACAAA  
 10 CAACATTACAATTACTATTACAATTACACC

(b)

ACCTCGAGTTTTTTATTTTAATTTCTTTCAAATACTTCCATCCC

15 Con respecto a la importancia real de los tripletes ATT en la inducción del inicio del proceso de traducción en un punto indeseado del mRNA en el líder, se debe indicar en la presente memoria que el auténtico codón de inicio de la traducción (ATG) necesita una secuencia de contexto adecuada para ser reconocido como tal por el complejo de traducción; es muy probable para una persona experta en la técnica que igualmente deba existir un contexto adecuado para el reconocimiento de tripletes de inicio de la traducción no canónicos tales como ATT y CTG.

20 Sin embargo, los contextos de reconocimiento de los tripletes no se conocen por el momento, y por lo tanto la persona experta no puede establecer, estudiando el estado de la técnica, si un cierto triplete ATT (o CTG) representa realmente un sitio de inicio de la traducción no canónico, y en qué grado funciona como tal.

Frente a estas pruebas, lo importante al determinar la elección del uso de  $\Omega$ , AMV o líderes que proceden de ellos, es el efecto positivo, demostrado de manera experimental, de la inclusión del líder  $\Omega$  o líder de AMV sobre el nivel de expresión génica.

25 La persona experta sabe, sin embargo, que si un triplete ATT o CGT del líder se interpretase realmente como un codón de inicio de la traducción, se produciría una proteína diferente, no la programada, y esto podría provocar problemas de bioequivalencia funcional y estructural, especialmente críticos en el caso de las proteínas destinadas a una aplicación terapéutica.

30 Los inventores del documento WO 2008/080954, que trabajan principalmente en el campo farmacéutico, son conscientes de los riesgos potenciales y, prudentemente, construyen su secuencia 5'-UTR colocando todos los tripletes ATT a una distancia recíproca que es siempre un múltiplo de 3, y un codón de parada (TAG) en el marco de lectura con respecto a ellos, hacia el final de la secuencia líder. De manera aún más ingeniosa, el final de la secuencia LLTCK se representa mediante el sitio de restricción para *Xba* I (TCTAGA) que tiene la función triple de albergar el codón de parada (TAG), de contribuir a la formación de una región poli(CT), de producir un posible contexto favorable para el reconocimiento de un codón de inicio auténtico localizado inmediatamente después, así  
 35 como de constituir un sitio de clonación extremadamente útil en 5' de la secuencia codificante deseada.

Otras personas expertas se comportan de otra manera y simplemente dejan los tripletes ATT en las secuencias ricas en A/T relativas.

De hecho, es común hallar líderes sintéticos con una secuencia programada que alberga tripletes ATT incluso en una posición divergente del marco de lectura auténtico.

40 A partir de lo anterior se puede concluir que, como otras patentes y publicaciones que preceden a esta descripción, el documento WO 2008/080954:

1. no enseña a eliminar motivos ricos en A/T de las secuencias 5'-UTR, sino exactamente lo contrario;
2. no enseña a eliminar tripletes ATT de secuencias 5'-UTR derivadas de omega o derivadas de AMV, sino exactamente lo contrario;
- 45 3. no enseña cómo producir contextos favorables para la expresión génica en ausencia de motivos ricos en A/T, alberguen o no tripletes ATT;
4. no enseña cómo construir variantes más eficaces que el líder LLTCK usado en los ejemplos del documento WO

2008/080954.

Al considerar todo esto, no se sugiere ni se promueve de ninguna manera la necesidad de eliminar las secuencias ricas en A/T y los tripletes ATT de manera explícita o implícita por el estado de la técnica, y por lo tanto no es en absoluto evidente para una persona experta en la técnica.

- 5 Además, debido a que cada sustitución, delección o adición de nucleótidos es potencialmente capaz de generar líderes con un comportamiento inesperado, también el efecto de tal eliminación, como cualquier otra manipulación de la secuencia 5'-UTR, no es en absoluto evidente para una persona experta en la técnica.

10 Por lo tanto, la presente invención propone, de una manera nueva e inventiva, la síntesis de variantes de 5'-UTR dotadas de elementos nuevos o combinaciones nuevas de elementos, que constituyen una solución técnica ventajosa, capaces de modificar y mejorar significativamente el estado de la técnica. El solicitante ha ideado, ensayado y realizado la presente invención para obtener estos y otros propósitos y ventajas.

15 A menos que se definan de otra manera, todos los términos técnicos y científicos usados en la presente memoria y a continuación tienen el mismo significado que entiende habitualmente una persona de experiencia habitual en el campo de la técnica al que pertenece la presente invención. Aunque se pueden usar métodos y materiales similares o equivalentes a los descritos en la presente memoria en la práctica y en los ensayos de la presente invención, los métodos y materiales se describen a continuación como ejemplo. En caso de conflicto, prevalecerá la presente solicitud, lo que incluye sus definiciones. Los materiales, métodos y ejemplos tienen un propósito meramente ilustrativo, y no se deben entender de manera restrictiva.

### Sumario de la invención

20 La presente invención se expone y caracteriza en las reivindicaciones independientes, mientras las reivindicaciones dependientes describen otras características de la invención o variantes de la idea inventiva principal.

25 De acuerdo con el propósito anterior, la presente descripción se refiere al campo de la biotecnología de plantas, y en particular se ocupa de la elevación del nivel productivo de proteínas recombinantes en plantas modificadas genéticamente mediante el uso de líderes artificiales construidos de manera adecuada según la presente descripción, obtenidos por medio de síntesis artificial y producto del intelecto, ya que no se hallan en la naturaleza.

30 Ciertas formas de realización descritas en la presente memoria se refieren a un ADN artificial de una región líder 5'-UTR para la expresión de transgenes en plantas. El ADN artificial según las características de la presente descripción es eficaz para incrementar la expresión de transgenes en plantas y comprende, a lo largo de la dirección 5'→3', un sitio iniciador *Inr* y una secuencia consenso Kozak o similar a Kozak, respectivamente, en los extremos 5' y 3' terminales correspondientes. El ADN artificial según las características de la presente descripción también comprende, entre el sitio iniciador *Inr* y la secuencia consenso Kozak o similar a Kozak, una diversidad de regiones poli(CAA) o (CAA)<sub>n</sub>, cada una formada por un oligonucleótido que consiste en dos o más copias de un elemento CAA contiguas entre sí, y una diversidad de regiones poli(CT) o (CT)<sub>m</sub> en el mismo número que las regiones poli(CAA), y cada una formada mediante un oligonucleótido que consiste en dos o más copias de un elemento CT contiguas entre sí, en el que al menos una, opcionalmente cada, región poli(CAA), en la dirección 5'→3', está antes de una región poli(CT), es decir, en posición 5', y al menos una región poli(CAA), en la dirección 5'→3', está contigua con una región poli(CT).

En ciertas formas de realización, el ADN artificial proporciona la presencia de secuencias que no se pueden asociar con motivos ricos en A/T, es decir, proporciona la ausencia de motivos ricos en A/T.

40 En ciertas formas de realización, los motivos ricos en A/T que no están presentes en el ADN artificial según la presente descripción se pueden definir como tramos o secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos de adenina (A) y/o timina (T), en cualquier combinación entre sí.

En ciertas formas de realización, el ADN artificial proporciona la presencia de secuencias que no se pueden asociar con elementos de trinucleótidos ATT, es decir, proporciona la ausencia de elementos de trinucleótidos ATT.

45 En ciertas formas de realización, el ADN artificial proporciona la presencia de secuencias que no se pueden asociar con elementos de trinucleótidos CTG, es decir, proporciona la ausencia de elementos de trinucleótidos CTG.

En ciertas formas de realización, el ADN artificial proporciona la ausencia de tramos homopoliméricos, es decir, secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos idénticos.

50 En ciertas formas de realización, se puede elegir el mismo valor n para las regiones poli(CAA) o se puede elegir de manera autónoma para las diversas regiones poli(CAA), es decir, se puede seleccionar un valor n diferente para al menos una de las regiones poli(CAA) con respecto a otra u otras regiones poli(CAA).

En ciertas formas de realización, n es un número entero mayor o igual a 2, opcionalmente comprendido entre 3 y 9, opcionalmente entre 4 y 8, opcionalmente entre 5 y 7.

- En ciertas formas de realización, para al menos una región poli(CAA),  $n$  es igual a 7, por ejemplo para al menos dos regiones poli(CAA)  $n$  es igual a 7.
- 5 En ciertas formas de realización, se puede elegir el mismo valor  $m$  para las regiones poli(CT) o se puede elegir de manera autónoma para las diversas regiones poli(CT), es decir, se puede seleccionar un valor de  $m$  diferente para al menos una de las regiones poli(CT) con respecto al valor de  $m$  de otra u otras regiones poli(CT).
- En ciertas formas de realización,  $m$  puede ser un número entero mayor o igual a 2, opcionalmente comprendido entre 3 y 5. Según ciertos aspectos, para al menos una región poli(CT),  $m$  es igual a 5. Según otros aspectos, para al menos una región poli(CT),  $m$  es igual a 3. En posibles implementaciones, para una región poli(CT),  $m$  es igual a 5 y para otra región poli(CT),  $m$  es igual a 3.
- 10 En ciertas formas de realización, el ADN artificial contiene dos regiones poli(CAA) y dos regiones poli(CT), de las cuales una región poli(CAA) puede estar contigua a una región poli(CT) y posiblemente otra región poli(CAA) puede no estar contigua a otra región poli(CT).
- 15 En ciertas formas de realización, una primera región poli(CAA) está antes, es decir, en posición 5', de una primera región poli(CT) y una segunda región poli(CAA) está después de dicha primera región poli(CT) y antes, es decir, en posición 5', de una segunda región poli(CT).
- En ciertas formas de realización, la primera región poli(CAA) está contigua a la primera región poli(CT).
- En otras formas de realización, la primera región poli(CAA) no está contigua a la primera región poli(CT).
- En ciertas formas de realización, la segunda región poli(CAA) está contigua a la primera región poli(CT).
- En otras formas de realización, la segunda región poli(CAA) no está contigua a la primera región poli(CT).
- 20 En ciertas formas de realización, la segunda región poli(CAA) está contigua a la segunda región poli(CT).
- En otras formas de realización, la segunda región poli(CAA) no está contigua a la segunda región poli(CT).
- En ciertas formas de realización, para la primera región poli(CAA) el valor de  $n$  es igual a 7, es decir, comprende 7 copias del triplete CAA.
- 25 En ciertas formas de realización, para la segunda región poli(CAA) el valor de  $n$  es igual a 7, es decir, comprende 7 copias del triplete CAA.
- En ciertas formas de realización, para la primera región poli(CT) el valor de  $m$  es igual a 5, es decir, comprende 5 copias del dinucleótido CT.
- En ciertas formas de realización, para la segunda región poli(CT) el valor de  $m$  es igual a 3, es decir, comprende 3 copias del dinucleótido CT.
- 30 En ciertas formas de realización, entre la segunda región poli(CAA) y la segunda región poli(CT) hay una secuencia AG. En ciertas formas de realización, entre la segunda región poli(CAA) y la segunda región poli(CT) hay exclusivamente la secuencia AG.
- En ciertas formas de realización, el sitio iniciador *Inr* es el sitio de inicio de la transcripción de 35S de CaMV o es un sitio iniciador *Inr* con una secuencia consenso 5'-YYANWYY-3', en la que:
- 35 Y=C, T;  
N=A, C, G, T;  
W=A, T.
- En posibles formas de ejemplos de la realización, el sitio iniciador *Inr* es 5'-TCACATC-3'.
- 40 En ciertas formas de realización, entre el sitio iniciador *Inr* y la primera región poli(CAA) a lo largo de la dirección 5'→3' hay una secuencia AAGTTTC. En ciertas formas de realización, entre el sitio iniciador *Inr* y la primera región poli(CAA) a lo largo de la dirección 5'→3' hay exclusivamente la secuencia AAGTTTC.
- En ciertas formas de realización, el ADN artificial tiene una longitud comprendida entre 40 y 150 pb.
- En ciertas formas de realización, el ADN artificial tiene un contenido de GC menor del 50%.
- 45 En ciertas formas de realización, el ADN artificial comprende la secuencia mostrada en SEQ ID N°: 1, o la secuencia mostrada en SEQ ID N°: 2, ambas incluidas en la lista de secuencias adjunta.

En ciertas formas de realización, la secuencia consenso Kozak o similar a Kozak es una secuencia que requiere la presencia de un elemento R que es una purina en posición -3, es decir, localizada en la tercera posición antes del codón de inicio de la traducción.

En ciertas formas de realización, el ADN artificial según la presente invención no contiene el octámero ACAATTAC.

- 5 Ciertas formas de realización descritas en la presente memoria se refieren a un vector de expresión que comprende ADN artificial de una región líder 5'-UTR eficaz para incrementar la expresión de proteínas recombinantes en plantas, en particular, por ejemplo, proteínas humanas, según las formas de realización descritas en la presente memoria.

En ciertas formas de realización, el vector de expresión comprende:

- 10 i) un promotor específico de endospermo de origen natural o artificial antes, es decir, en posición 5', de una secuencia de nucleótidos de origen natural o artificial que codifica la forma madura de una proteína;
- ii) el ADN artificial de la región líder 5'-UTR eficaz para incrementar la expresión de proteínas recombinantes en plantas como se describe en la presente memoria;
- 15 iii) una secuencia de nucleótidos de origen natural o artificial que codifica un péptido señal para transportar la proteína recombinante dentro de la luz del retículo endoplásmico de las células que constituyen el tejido del endospermo, y así favorecer su acumulación tisular;
- iv) la secuencia de nucleótidos de origen natural o artificial que codifica la forma madura de la proteína de interés;
- v) una región 3'-UTR de origen natural o artificial.

En ciertas formas de realización, el promotor i) es el promotor del gen de glutelina 4 de arroz (GluB4).

- 20 En ciertas formas de realización, la secuencia de nucleótidos del elemento iii) es la secuencia PSGluB4 que codifica el péptido señal usado en arroz para transportar el precursor de glutelina 4 dentro del retículo endoplásmico.

En ciertas formas de realización, la secuencia de nucleótidos del elemento iv) es la secuencia que codifica la forma humana madura de la enzima beta-glucosidasa ácida.

- 25 En ciertas formas de realización, la región 3'-UTR del elemento v) es el terminador NOS o el terminador del gen GluB4.

Ciertas formas de realización descritas en la presente memoria se refieren a una cepa bacteriana que alberga un plásmido que contiene una secuencia de ADN artificial como se describe en la presente memoria, en particular, por ejemplo, elegida de un grupo que comprende la especie *Escherichia coli*, *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes*.

- 30 Ciertas formas de realización descritas en la presente memoria se refieren a una cepa bacteriana modificada que contiene una secuencia de ADN artificial según las formas de realización descritas en la presente memoria, independientemente del tipo de organismo hospedador.

- 35 Ciertas formas de realización descritas en la presente memoria se refieren a células vegetales transformadas con vectores de expresión que contienen la secuencia de ADN artificial como se describe en la presente memoria, bajo control de un promotor elegido de un grupo que comprende un promotor constitutivo, un promotor específico de tejido y en particular, por ejemplo, un promotor específico de semillas, un promotor inducible, un promotor con actividad transcripcional dependiente de la fase, un promotor activo en cloroplastos y un promotor activo en mitocondrias.

- 40 Ciertas formas de realización descritas en la presente memoria se refieren a plantas caracterizadas por la expresión transitoria de cualquier proteína cuyo ARN mensajero contiene la secuencia de ADN artificial descrita en la presente memoria; la expresión transitoria significa la producción de dicha proteína mediante vectores virales, agroinfiltración, bombardeo con micropartículas, electroporación.

- 45 Ciertas formas de realización descritas en la presente memoria se refieren a plantas dicotiledóneas transformadas de manera estable con vectores de expresión que contienen la secuencia de ADN artificial según las formas de realización descritas en la presente memoria.

En ciertas formas de realización, las plantas dicotiledóneas comprenden una o más especies que pertenecen a las familias *Solanaceae*, *Papilionaceae* y/o *Cruciferae*.

Ciertas formas de realización descritas en la presente memoria se refieren a la progenie de las plantas dicotiledóneas como se describió anteriormente.

- 50 Ciertas formas de realización descritas en la presente memoria se refieren a plantas monocotiledóneas

transformadas con vectores de expresión que contienen la secuencia de ADN artificial según las formas de realización descritas en la presente memoria.

5 En ciertas formas de realización, las plantas monocotiledóneas comprenden una o más especies que pertenecen a la familia *Graminaceae* (*Poaceae*), tales como, por ejemplo, arroz cultivado (*Oryza sativa* L.), maíz (*Zea mays* L.), cebada (*Hordeum vulgare* L.) y/o trigo (género *Triticum*).

Ciertas formas de realización descritas en la presente memoria se refieren a la progenie de las plantas monocotiledóneas como se describió anteriormente.

Ciertas formas de realización se refieren a la secuencia de ADN artificial según las formas de realización descritas en la presente memoria para un uso elegido de un grupo que comprende:

- 10 - el uso para la producción biotecnológica de moléculas;
- el uso para la síntesis de proteínas recombinantes, en particular, por ejemplo, destinadas a inducir resistencia a patógenos virales, bacterianos o fúngicos, o destinadas a inducir resistencia a herbicidas o para obtener una composición alterada de ácidos grasos en la materia prima y en los productos derivados de ella, o para obtener un valor nutricional alterado de la materia prima y en los productos derivados de ella, o para la producción de
- 15 combustibles, cauchos y/o bioplásticos;
- el uso para la síntesis de enzimas industriales y proteínas comerciales;
- el uso para la síntesis de proteínas farmacéuticas;
- el uso para la síntesis de vacunas elegidas de un grupo que comprende: vacunas administradas de manera oral destinadas a seres humanos o animales, vacunas inyectables destinadas a seres humanos o animales, vacunas inyectables específicas de paciente, preferiblemente específicas de idiotipo, para usarlas en el tratamiento de
- 20 tumores del sistema linfático;
- el uso para la síntesis de proteínas implicadas en la producción de metabolitos secundarios;
- el uso para la síntesis de proteínas utilizables directamente o indirectamente como factores en la identificación y/o selección de células transformadas.

25 Ciertas formas de realización descritas en la presente memoria se refieren a la semilla de una planta transformada para la expresión de una proteína humana, en particular, por ejemplo, una enzima lisosómica humana, que contiene un vector de expresión según las formas de realización descritas en la presente memoria.

Ciertas formas de realización descritas en la presente memoria se refieren a una semilla como se describió anteriormente, para el uso en el tratamiento terapéutico, en particular, por ejemplo, para el uso en la terapia de

30 sustitución enzimática, aún más en particular, por ejemplo, en las siguientes enfermedades: enfermedad de Gaucher, glucogenosis tipo II o enfermedad de Pompe, enfermedad de Fabry, enfermedad de Niemann-Pick tipo B, Mucopolisacaridosis I, II, IV.

Ciertas formas de realización se refieren a un método para la producción de proteínas recombinantes en plantas, que comprende la transformación de las plantas mediante el uso de un vector de expresión como se describe en la

35 presente memoria.

En ciertas formas de realización, la transformación de las plantas es eficaz para conseguir el confinamiento de la proteína en un endospermo no absorbido por el embrión, y para permitir que la presencia de cantidades elevadas de la proteína en el endospermo de la semilla no provoque efectos negativos sobre la viabilidad de la semilla y la velocidad de germinación.

40 En ciertas formas de realización, el método proporciona la acumulación de la proteína dentro del endospermo de la semilla de la planta, en particular, por ejemplo, la proteína se acumula en el endospermo dentro de las vacuolas de almacenamiento de proteínas (PSV) o cuerpos proteicos (PB).

En ciertas formas de realización, el vector de expresión se introduce en cepas bacterianas que se usan, directamente o indirectamente, para la transformación de las plantas, en las que la cepa bacteriana se puede elegir

45 de un grupo que comprende la especie *Escherichia coli*, *Agrobacterium tumefaciens* y *Agrobacterium rhizogenes*.

En ciertas formas de realización, las plantas transformadas son cereales.

En ciertas formas de realización, la cepa bacteriana se usa para la transformación de callos embriogénicos de arroz (*Oryza sativa* subespecie *japonica*).

50 En ciertas formas de realización, la proteína recombinante es una enzima lisosómica, en particular, por ejemplo, beta-glucosidasa ácida humana, o por ejemplo alfa-glucosidasa ácida humana.



En ciertas formas de realización, el método comprende el procesamiento industrial de la semilla de la planta.

En ciertas formas de realización, el procesamiento industrial de la semilla de la planta proporciona el descascarado y pulido de las semillas maduras recogidas de plantas de cereal transformadas para eliminar el componente fibroso, el germen, y la capa aleurónica que contiene contaminantes proteicos.

- 5 En ciertas formas de realización, el método comprende la purificación de la proteína obtenida.

En ciertas formas de realización, la purificación proporciona, por orden, una cromatografía con interacciones hidrófobas, una cromatografía con intercambio iónico y una filtración en gel.

- 10 En ciertas formas de realización, la purificación proporciona aplicar resinas cromatográficas similares en composición química y/o estructura y/o función, modificar parcialmente los parámetros de elución, y duplicar un paso para recargar la fracción eluida en la columna.

### Breve descripción de los dibujos

Estas y otras características de la presente invención serán evidentes a partir de la siguiente descripción de ciertas formas de realización, proporcionadas como un ejemplo no limitante con referencia a los dibujos adjuntos, en los que:

- 15 - la Fig. 1 muestra la distribución de los valores obtenidos mediante el uso de un ensayo de 4-MUG en plantas de tabaco transformadas con pSTART y pSTART-STE;

- la Fig. 2 muestra la distribución del contenido de proteína GCasi, determinado mediante el uso de un ensayo de DAS-ELISA y expresado en µg de GCasi por gramo de harina de arroz, en plantas que albergan el líder LLTCK y el líder STE;

- 20 - la Fig. 3A muestra el diagrama del vector de expresión en tabaco pSTART-STE, en el que:

RBR: repetición del borde derecho

LBR: repetición del borde izquierdo

35S de CaMV: promotor 35S del virus del mosaico de la coliflor

GUS: proteína indicadora

- 25 NOS ter: terminador de la Nopalina sintasa de *Agrobacterium tumefaciens*

- La fig. 3B muestra el tramo sintetizado de manera artificial que contiene una parte del promotor 35S de CaMV (desde el sitio *Sca I*) y el líder STE;

- La Fig. 4A muestra el diagrama de los vectores de expresión en arroz pCAMBIA1300/PMI/GluB4-LLTCK/STE::GCasi::GluB4ter; en el que:

- 30 RBR: repetición del borde derecho

LBR: repetición del borde izquierdo

GluB4-LLTCK: promotor de glutelina 4 de arroz con el líder LLTCK

GluB4-STE: promotor de glutelina 4 de arroz con el líder STE

GCasi: gen que codifica la enzima humana beta-glucosidasa ácida (hGCasi)

- 35 GluB4 ter: terminador de glutelina 4 de arroz

35S pro: promotor 35S de CaMV

PMI: gen que codifica la fosfomanosa isomerasa (marcador de selección de las plantas transformadas)

35S ter: terminador 35S de CaMV

- 40 - las Figs. 4B y 4C representan respectivamente el tramo sintetizado de manera artificial que contiene la parte final del promotor GluB4 (desde el sitio *Bfr I*) y los líderes LLTCK y STE.

### Descripción detallada de ciertas formas de realización

Ahora se hace referencia con detalle a las diversas formas de realización de la presente invención, de las cuales se describen uno o más ejemplos a continuación en la presente memoria. Cada ejemplo se proporciona a modo de

ilustración de la invención, y no se debe entender como una limitación de la misma. Por ejemplo, las características mostradas o descritas en la medida en que son parte de una forma de realización se pueden adoptar en, o junto con, otras formas de realización para producir otra forma de realización. Se entiende que la presente invención incluirá tales modificaciones y variantes.

5 En un intento de incrementar adicionalmente la eficacia del tramo 5'-UTR para la expresión de transgenes en plantas comparado con el estado de la técnica, se ha ideado una secuencia de ADN artificial, denominada a continuación en la presente memoria STE, secuencia STE o líder STE, que contiene elementos de trinucleótidos CAA repetidos y elementos de dinucleótidos CT repetidos, como se describe en el documento WO 2008/080954, cuya secuencia STE se optimiza para la sobreexpresión de proteínas recombinantes en plantas.

10 Se debería indicar que esta secuencia STE nueva e inventiva ha proporcionado un incremento de la expresión génica en dos especies de plantas no relacionadas y junto con promotores, terminadores y secuencias codificantes diferentes.

Partiendo del estado de la técnica como se discutió anteriormente, el solicitante llevó a cabo otros experimentos destinados a desarrollar un nuevo tipo de líder según la presente descripción.

15 En particular, el solicitante consideró que muchos virus que atacan a las plantas producen mensajeros sin caperuza 5', y en muchos casos también sin la cola de poli(A). Estas pruebas permitieron al solicitante suponer que en estos virus las regiones sin traducir en 5' (5'-UTR) y 3' (3'-UTR) albergan secuencias capaces de sustituir funcionalmente la estructura de la caperuza 5' y la cola de poli(A), respectivamente. Estas secuencias, indispensables en el mensajero viral, podrían ser, sin embargo, menos importantes dentro de los líderes de genes eucarióticos, y de  
20 manera específica los genes vegetales, porque los mensajeros que producen tienen siempre la caperuza 5' y, excepto por escasas excepciones, también la cola de poli(A).

En particular, el solicitante realizó la hipótesis de que las secuencias esenciales para los líderes virales, pero no para los líderes de genes eucarióticos, corresponden a una o más secuencias ricas en A/T, tales como, por ejemplo, las indicadas previamente mediante los números 1-4 dentro de LLTCK o las porciones de la misma. Por lo tanto, se  
25 inició una actividad de diseño con la intención de obtener una secuencia líder sintética totalmente desprovista de secuencias ricas en A/T, y por lo tanto sin regiones de octámero  $\Omega$  ni tripletes ATT; en la formación del nuevo líder, también se decidió excluir el trinucleótido CTG y tramos de homopolímeros formados por la repetición de cualquier nucleótido. Para mantener la longitud del líder sustancialmente inalterada con respecto al documento WO 2008/080954, se sustituyeron las regiones ricas en A/T por motivos CAA y CT repetitivos. La secuencia resultante,  
30 denominada STE, se comparó con la SEQ. ID N°. 1 del documento WO 2008/080954 en diferentes contextos.

Los resultados obtenidos permitieron establecer que, de acuerdo con la hipótesis del solicitante, pero a diferencia de lo que se esperaba basándose en el estado de la técnica disponible para la persona experta, eliminar los elementos ricos en A/T y sustituirlos con elementos CAA y CT repetidos siempre provoca un incremento significativo de la  
35 expresión de los genes indicadores usados en los experimentos comparativos entre líderes, incluso si dichos elementos ricos en A/T se conservan dentro de los líderes virales usados habitualmente como potenciadores de la traducción. Comparado con LLTCK del documento WO 2008/080954, el nuevo tipo de 5'-UTR según la presente descripción representa una mejor solución técnica útil para resolver, en el campo industrial, los problemas asociados a la producción, extracción y purificación eficaz de proteínas heterólogas.

40 Por lo tanto, las formas de realización descritas en la presente memoria proporcionan ADN artificial de una región líder 5'-UTR eficaz para incrementar la expresión de proteínas recombinantes en plantas, que comprende a lo largo de la dirección 5'→3' una diversidad de regiones poli(CAA) o  $(CAA)_n$ , y una diversidad de regiones poli(CT) o  $(CT)_m$  en el mismo número que las regiones poli(CAA).

En ciertas formas de realización, cada región poli(CAA) está formada por un oligonucleótido que consiste en dos o más copias de un elemento CAA contiguas entre sí.

45 En ciertas formas de realización, cada región poli(CT) está formada por un oligonucleótido que consiste en dos o más copias de un elemento CT contiguas entre sí.

En ciertas formas de realización, al menos una, opcionalmente cada, región poli(CAA), en la dirección 5'→3', está antes de una región poli(CT), es decir, en posición 5'.

50 En ciertas formas de realización, al menos una región poli(CAA), en la dirección 5'→3', está contigua a una región poli(CT).

En ciertas formas de realización, n es un número entero, que se puede seleccionar para ser igual o diferente entre las regiones poli(CAA), mayor o igual a 2, opcionalmente comprendido entre 3 y 9, opcionalmente entre 4 y 8, opcionalmente entre 5 y 7. Por ejemplo, el valor n puede ser el mismo para las regiones poli(CAA) y puede ser, por ejemplo, igual a 7.

55 En ciertas formas de realización, m es un número entero, que se puede seleccionar para ser igual o diferente entre

las regiones poli(CT), mayor o igual a 2, opcionalmente comprendido entre 3 y 5. Por ejemplo, el valor *m* puede ser diferente para las regiones poli(CT) y puede ser, por ejemplo, igual a 3 o 5.

Aunque en general los valores de *n* y *m* se pueden seleccionar para ser diferentes entre sí, también se pueden proporcionar ciertas formas de realización en las que los valores de *n* y *m* se seleccionan para ser iguales entre sí.

5 En posibles implementaciones, se pueden proporcionar dos regiones poli(CAA) y dos regiones poli(CT). A lo largo de la dirección 5'→3' se puede proporcionar una primera región poli(CAA), una primera región poli(CT) posterior, contigua a la primera región poli(CAA) precedente, una segunda región poli(CAA) sucesiva, contigua a la primera región poli(CT) precedente y una segunda región poli(CT) sucesiva, no contigua a la segunda región poli(CAA).

10 En ciertas formas de realización, la secuencia STE se puede caracterizar por aspectos que son una mejora en comparación con el documento WO 2008/080954, y que se refieren a una o más de las siguientes características, destinadas a hacer que la secuencia STE sea más compatible con un sistema de expresión eucariótico:

1. Ausencia de motivos ricos en A/T, es decir, secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos de adenina (A) y/o timina (T), en cualquier combinación de los mismos;

2. Ausencia de elementos de trinucleótidos ATT;

15 3. Ausencia de elementos de trinucleótidos CTG;

4. Ausencia de tramos homopoliméricos, es decir, secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos idénticos.

20 En otras palabras, el ADN artificial según la presente descripción no contiene ninguno de los componentes siguientes: elementos ricos en A/T, elementos de trinucleótidos ATT, elementos de trinucleótidos CTG y tramos homopoliméricos, es decir, secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos idénticos.

25 En ciertas formas de realización, una secuencia de ADN artificial descrita en la presente memoria puede contener un sitio *Inr*. El sitio *Inr* puede tener una secuencia 5'-YYANWYY-3', con las limitaciones anteriores de los puntos 1 (ausencia de motivos ricos en A/T) y 2 (ausencia de elementos de trinucleótidos ATT), en la que Y=C, T; N=A, C, G, T; W=A, T; de manera alternativa, el sitio *Inr* puede tener una secuencia 5'-ACACG-3' (sitio de inicio de la transcripción para 35S de CaMV).

30 En ciertas formas de realización, en el extremo 3', la región líder puede contener también un contexto de nucleótidos favorable al reconocimiento del auténtico codón de inicio de la traducción ATG (motivo Kozak o similar a Kozak o secuencia consenso). Un motivo Kozak o similar a Kozak o secuencia consenso requiere la presencia de un elemento R que es una purina (adenina "A" o guanina "G") en posición 3 antes del codón de inicio de la traducción, para identificar el contexto adecuado para reconocer el auténtico codón de inicio de la traducción. La posición 3 antes del codón de inicio (o posición -3) significa una posición 3 nucleótidos antes del elemento "A" del codón ATG, para el que se asigna convencionalmente la posición +1. La secuencia Kozak o similar a Kozak puede ser sucesiva y contigua, por ejemplo, con la segunda región poli(CT) como se discutió anteriormente.

35 Además, en ciertas formas de realización, la secuencia líder STE descrita en la presente memoria puede tener una longitud comprendida entre 40 y 150 pb, y puede tener opcionalmente un contenido de GC menor del 50%.

Un ejemplo de una secuencia líder, denominada SEQ ID N°: 1, según ciertas formas de realización, es:

(1) 5'-ACACGAAGTTTCCAACAACAACAACAACAACACTCTCTCTCT  
CAACAACAACAACAACAACAAGCTCTCTAGA-3'

Otro ejemplo de una secuencia líder, denominada SEQ ID N°: 2, según ciertas formas de realización, es:

(2) 5'-CACATCAAGTTTCCAACAACAACAACAACAACAACACTCTCTCTCT  
CAACAACAACAACAACAACAAGCTCTCTAGA-3'

40 En ciertas formas de realización, la secuencia líder descrita en la presente memoria, tal como por ejemplo SEQ ID N°: 1 y SEQ ID N°: 2, puede tener un sitio iniciador *Inr* en 5', tal como el sitio de inicio de la transcripción de 35S de CaMV (SEQ ID N°: 1, variante 1) o un sitio *Inr* con una secuencia consenso típica de los genes eucarióticos, 5'-YYA<sup>+</sup>NWYY-3', en la que A<sup>+</sup> representa el primer nucleótido transcrito, Y=C, T; N=A, C, G, T; W=A, T (TCACATC en SEQ ID N°: 2, variante 2). Después del sitio iniciador, hay bloques prolongados y alternativos de poli(CAA) y poli(CT), repetidos, por ejemplo, dos veces. Como se dijo, además, para estimular el reconocimiento del codón de inicio ATG, puede haber una secuencia Kozak o similar a Kozak en el extremo 3' terminal (por ejemplo en ambas variantes se puede incluir en TCTAGA, que corresponde al sitio de restricción para *Xba* I).

Comparado con el tipo de líder descrito en el documento WO 2008/080954, la secuencia de ADN artificial descrita en la presente memoria puede proporcionar, por lo tanto, nuevas especificaciones para producir 5'-UTRs artificiales. Tales especificaciones, no proporcionadas en el documento WO 2008/080954, pueden reflejarse en modificaciones precisas de composición y estructura de la región líder y en nuevas aplicaciones preferentes.

5 La entidad de las variaciones se puede deducir, a modo de ejemplo, comparando la secuencia LLTCK descrita en el documento WO 2008/080954 con las variantes de ejemplo del líder STE descritas anteriormente. Estas últimas no tienen ninguna secuencia definible como elementos ricos en A/T (elementos ricos en AU, AREs), ni tripletes ATT que están presentes por otra parte en LLTCK después del sitio *Inr*, internamente y a los lados del octámero del líder  $\Omega$  de TMV.

10 Para demostrar la mayor eficacia del nuevo líder STE, se hizo una comparación entre éste y la secuencia LLTCK descrita en el documento WO 2008/080954, analizando los niveles de expresión de dos genes indicadores en especies de plantas no relacionadas, tales como tabaco (*Nicotiana tabacum* L.) y arroz (*Oryza sativa* L.).

En tabaco, el gen considerado fue *uidA* (GUS) y las construcciones usadas para la transformación genética, es decir, 35S-LLTCK::*uidA*::NOS ter (pSTART) y 35S-STE::*uidA*::NOS ter (pSTART-STE), se obtuvieron sustituyendo la secuencia líder presente en pBI121 (Clontech) por LLTCK y STE, respectivamente (Ejemplo 1). Más exactamente, fue la secuencia pBI121 la que se sustituyó y se manipuló, comprendida entre la región *Inr* de 35S de CaMV (ACACG), mantenida en común para ambas construcciones, y el sitio de restricción *Xba* I (TCTAGA). Los niveles de expresión del gen indicador se estudiaron mediante el uso de un ensayo fluorimétrico con 4-MUG (Jefferson et al., 1987), caracterizado por una considerable sensibilidad, precisión, velocidad y facilidad de ejecución. En particular, para determinar cuantitativamente la expresión del gen GUS en las plantas transformadas, se llevaron a cabo ensayos fluorimétricos con extractos de proteínas en bruto obtenidos al prensar tres hojas jóvenes completamente distendidas. Los valores de actividad específica para  $\beta$ -glucuronidasa (GUS), expresados en milimoles de 4-MU producidos por mg de proteína, se normalizaron con respecto a la concentración de proteína total calculada mediante el uso de un ensayo de Bradford. La Fig. 1 muestra los datos obtenidos con las plantas transformadas con las dos construcciones (media de los valores registrados para los tres extractos de hojas) en orden decreciente.

Se llevó a cabo un análisis estadístico con los datos mostrados en la Fig. 1, en este caso un análisis unidireccional de la varianza (ANOVA), cuyos resultados se muestran en la Tabla 1 siguiente.

Tabla 1: ANOVA llevado a cabo con los datos medios obtenidos en plantas transformadas con las construcciones 35S-LLTCK::*uidA*::NOS ter y 35S-STE::*uidA*::NOS ter.

RESUMEN

Grupos	Cuenta	Suma	Media	Varianza
pSTART	20	3,985	0,199	0,151
STE	23	12,973	0,564	0,514

30

ANÁLISIS DE LA VARIANZA

Origen de la varianza	SQ	gdl	MQ	F	Valor de la significación
Entre grupos	1,424	1	1,424	4,119*	P < 0,05
Dentro de grupos	14,169	41	0,346		
Total	15,593	42			

El análisis mostró la existencia de diferencias estadísticamente significativas ( $P < 0,05$ ) entre las dos poblaciones analizadas (pSTART y pSTART-STE). A partir del examen conjunto de la Tabla 1 y la Fig. 1 es posible afirmar que el líder STE provoca incrementos no marginales en los niveles de expresión de *uidA*. En particular, si se consideran los valores medios obtenidos en las dos poblaciones de plantas transformadas, el líder STE conduce a un incremento de los niveles de expresión del gen indicador GUS de alrededor de 2,8 veces en comparación con LLTCK.

35

En confirmación de lo que se observó en tabaco (especie modelo para la clase de dicotiledóneas), también se llevaron a cabo experimentos con arroz, un cereal usado de manera generalizada en el campo biotecnológico. Como con la primera especie, la comparación se llevó a cabo con dos construcciones de expresión, exactamente como los otros elementos. En particular, se compararon los siguientes vectores:

40

pCAMBIA1300/PMI/GluB4-LLTCK::GCasi::GluB4ter;

pCAMBIA1300/PMI/GluB4-STE::GCasi::GluB4ter.

Sin embargo, se debe subrayar que en el arroz se ensayó el efecto de un tipo diferente de líder en un contexto de expresión específica de la semilla, mediante el uso de elementos de control diferentes. Más exactamente, se usó el promotor de glutelina 4 de arroz (GluB4), y el terminador correspondiente (GluB4ter). Como gen indicador se eligió hGCasi, es decir, la secuencia que codifica la enzima beta-glucosidasa ácida humana; la detección de la proteína recombinante se puede llevar a cabo con una sensibilidad y precisión considerables por medio de un ensayo inmunológico (DAS-ELISA). Con respecto a la secuencia líder, en ambos vectores se usó el sitio *Inr* de GluB4, debido a que se halla dentro de la secuencia consenso eucariótica YYANWYY. Además, el sitio de inicio de la transcripción del promotor 35S de CaMV pareció menos adecuado, ya que este virus ataca solamente a las plantas dicotiledóneas.

Cada vector se insertó en *Agrobacterium tumefaciens* mediante el uso de electroporación para la transformación de *Oryza sativa*, var. CR W3 (Hiei et al., 1994). Se obtuvieron dos poblaciones de plantas transgénicas, y cada una consistió en 50 plantas individuales. Se recogieron las semillas maduras de cada planta y se usaron para la extracción de las proteínas totales. Los extractos de proteínas obtenidos se analizaron en DAS-ELISA para determinar el contenido de GCasi. La Fig. 2 muestra la distribución de los datos obtenidos.

El análisis unidireccional de la varianza permitió establecer que las diferencias de expresión del gen indicador halladas entre las dos poblaciones de arroz consideradas son estadísticamente significativas (Tabla 2).

Tabla 2: ANOVA llevado a cabo con los datos obtenidos en plantas transformadas con las construcciones pCAMBIA1300/PMI/GluB4-LLTCK::GCasi::GluB4ter y pCAMBIA1300/PMI/GluB4-STE::GCasi::GluB4ter.

#### RESUMEN

Grupos	Cuenta	Suma	Media	Varianza
Líder STE	50	5052,015	101,040	1714,609
Líder LLTCK	50	1468,772	29,375	643,617

#### ANÁLISIS DE LA VARIANZA

Origen de la varianza	SQ	gdl	MQ	F	Valor de la significación
Entre grupos	128396,3296	1	128396,330	108,892*	P < 0,05
Dentro de grupos	115553,0807	98	1179,113		
Total	243949,4102	99			

A partir de la Tabla 2 y del gráfico de la Fig. 2 mostrado anteriormente, es evidente que el líder STE proporciona niveles de expresión definitivamente mayores que el líder LLTCK. En particular, el líder STE provoca un incremento de los niveles de expresión del gen indicador GCasi de alrededor de 3,5 veces los de LLTCK.

#### Ejemplos

Ejemplo 1: Producción del vector de expresión en tabaco pSTART-STE

El punto de partida para la producción del vector pSTART-STE fue el vector de expresión pSTART, obtenido en un trabajo previo (De Amicis et al., 2007). Este último vector, a su vez obtenido a partir de una modificación del vector original pBI121 (Clontech), tiene un casete de expresión que consiste en el promotor 35S de CaMV con el líder LLTCK, el gen indicador que codifica la proteína GUS y el terminador NOS. Para obtener pSTART-STE (Fig. 3A), se sustituyó el líder LLTCK en pSTART por el líder STE. Para este fin, se sintetizó de manera artificial la secuencia que correspondía a una parte del promotor 35S (desde el sitio *Sca* I), a la que se le añadió la secuencia del líder STE, en este caso del ejemplo de SEQ ID N°: 1. El tramo sintetizado (702 pb, Fig. 3B) se sustituyó en pSTART mediante digestión con las enzimas de restricción *Sca* I y *Xba* I, recuperación del vector y ligadura con ADN ligasa de la nueva secuencia sintetizada.

Ejemplo 2: Producción de vectores de expresión con el promotor GluB4 y los líderes LLTCK y STE

Las secuencias líder LLTCK y STE se sintetizaron de manera artificial. En particular, en ambos casos, el tramo sintetizado correspondió a la secuencia comprendida entre el sitio *Bfr* I, presente en la parte terminal del promotor de glutelina 4 de arroz (GluB4) y el sitio *Xba* I, presente en el extremo 3' terminal de los propios líderes (Figs. 4B y 4C). Más exactamente, este tramo resultó ser igual a 328 pb para LLTCK (Fig. 1B) y 315 pb para STE (Fig. 4C).

Para producir los vectores de expresión finales, se llevó a cabo una serie de etapas de sub-clonación intermedias en paralelo para los dos líderes, lo que permitió el montaje final de los casetes de expresión. En la etapa inicial, el líder presente de manera nativa después del promotor GluB4 se sustituyó por los líderes sintéticos LLTCK y STE. El punto de partida fue el vector pGEM-T/GluB4-NAT, que contiene el promotor de glutelina 4 fusionado con el líder nativo (nº de acc. de GenBank AY427571). El tramo terminal del promotor GluB4 (desde el sitio *Bfr* I) y el líder nativo se eliminaron mediante digestión con las enzimas *Bfr* I y *Xba* I, y se sustituyeron por las nuevas secuencias sintetizadas. De esta manera se produjeron dos vectores intermedios, pGEM-T/GluB4-LLTCK y pGEM-T/GluB4-STE, y posteriormente se verificaron mediante análisis de PCR, digestión enzimática y secuenciación.

Los casetes de expresión finales se ensamblaron partiendo del vector pUC18/GluB4ter. Este vector se sometió a dos etapas de sub-clonación sucesivas para la inserción del complejo GluB4-LLTCK (o GluB4-STE) y el gen indicador, respectivamente. En particular, en la primera sub-clonación, se digirió pUC18/GluB4ter con las enzimas de restricción *Sph* I y *Xba* I para ligar los tramos GluB4-LLTCK y GluB4-STE, extraídos de los vectores pGEM-T/GluB4-LLTCK y pGEM-T/GluB4-STE, respectivamente. En la segunda sub-clonación, se abrieron los vectores intermedios pUC18/GluB4-LLTCK::GluB4ter y pUC18/GluB4-STE::GLUB4ter mediante digestión con *Xba* I y *Sac* I para insertar el gen indicador (hGCasi), a su vez extraído del vector pMS/hGCasi mediante el uso de las mismas enzimas. De esta manera se obtuvieron los dos vectores pUC18, que contenían los casetes de expresión completamente ensamblados, es decir, pUC18/GluB4-LLTCK::GCasi::GluB4ter y pUC18/GluB4-STE::GCasi::GluB4ter.

Para producir los vectores finales, los dos casetes de expresión GluB4-LLTCK::GCasi::GluB4ter y GluB4-STE::GCasi::GluB4ter se extrajeron individualmente, por ejemplo mediante una digestión doble con *Eco* RI del pUC18 respectivo y se clonaron en el vector de expresión final pCAMBIA1300/PMI para constituir (Fig. 4A):

pCAMBIA1300/PMI/GluB4-LLTCK::GCasi::GluB4ter y

pCAMBIA1300/PMI/GluB4-STE::GCasi::GluB4ter.

### Ejemplo 3: Transformación genética de *Nicotiana tabacum* mediada por *Agrobacterium tumefaciens*

Para la transformación genética de tabaco (*Nicotiana tabacum*, variedad Xanthi) mediada por *A. tumefaciens*, se usó el protocolo de Horsch et al. (1985). A continuación se describirán brevemente las etapas principales del procedimiento completo.

#### Desinfección de las semillas

Para la preparación de las semillas de tabaco para usarlas en la transformación, primero se llevó a cabo una desinfección según el siguiente protocolo:

Colocar una pequeña cantidad de semillas en un tubo de ensayo estéril de 2 mL. Añadir alrededor de 1 mL de etanol del 95%. Mantener durante 2 min y agitar enérgicamente. Eliminar el etanol mediante el uso de una pipeta. Añadir 1 mL de hipoclorito del 2%. Dejar incubar durante 20 min, agitar, eliminar y añadir 1 mL de agua estéril; lavar las semillas de esta manera 5 veces. Dejar el agua del último lavado en el tubo de ensayo. Retirar una cierta cantidad de semillas y agua, mediante el uso de una varilla de la que la punta se ha retirado en condiciones estériles, y colocarla en un sustrato MS10 en una placa o en un tarro.

Mediante el uso de un asa bacteriológica o una pipeta Pasteur doblada en forma de L, distribuir las semillas con delicadeza.

Colocar las placas a germinar con luz en una cámara climática a 28 °C.

#### Transformación con *A. tumefaciens*

La transformación del material de hojas de *N. tabacum* mediante el uso de *A. tumefaciens* se llevó a cabo en las siguientes etapas:

- bajo una campana, rellenar 4-5 tubos de ensayo de 2 mL con 1,8 mL de caldo LB estéril. Inocular el *A. tumefaciens*, recogiendo con un palillo estéril una cantidad pequeña pero visible de colonias bacterianas cultivadas sobre la placa, y diluirlas en un tubo de ensayo; agitar enérgicamente;

- tomar una hoja de tabaco (de plantas de una edad de alrededor de 1 mes) y, mediante el uso de un sacabocados estéril, hacer discos de un diámetro de 7 mm del borde de la hoja; mediante el uso de unas pinzas, colocar los discos de hojas en una placa de sustrato MS10; colocar 30 discos por placa. Para cada cepa bacteriana obtener un total de al menos 200 discos. Preparar dos placas de control en las que colocar discos que no se infectarán y que siempre permanecerán en el medio MS10;

- infectar los discos con *A. tumefaciens*; verter el contenido de un tubo de ensayo, inoculado justo antes con la bacteria, sobre la placa que contiene los discos. Agitar suavemente con un movimiento rotatorio, para humedecer todos los discos, después eliminar el exceso de líquido con una pipeta. Distribuir los discos regularmente, mediante el uso de las pinzas;

- incubar las placas durante una noche con luz constante a una temperatura de 28 °C en una cámara de cultivo;
- transferir los discos de hojas a un sustrato de MS10 Cefotaxima 500 mg/L;
- incubar las placas durante 6 días con luz constante y a una temperatura de 24 °C;
- 5 • 8 días tras el inicio de la transformación, transferir los discos de hojas a un sustrato de MS10 Cefotaxima 500 mg/L-Kanamicina 200 mg/L para la selección de los callos transformados. Incubar durante 14 días en las mismas condiciones;
- cortar los brotes que consisten en al menos dos hojas, no quiméricas y con un aspecto normal. Transferirlas a un sustrato para que arraiguen (MS 0 con Cefotaxima 500 mg/L - Kanamicina 200 mg/L - IBA 2 mg/L);
- 10 • transferir las plantas arraigadas a turba en macetas o a un sistema de cultivo hidropónico para el cultivo, limpiando con delicadeza el sustrato de las raíces con agua. Colocar las plantas en una cámara climática manteniendo una temperatura de 26-30 °C con un periodo de luz de 16 horas y 8 horas de oscuridad.

Ejemplo 4: Transformación genética de *Oryza sativa* mediada por *Agrobacterium tumefaciens*.

15 Para la transformación de arroz, variedad CR W3, se usó el protocolo de Hiei et al. (1994), tal como lo modificó Hoge (Grupo de Investigación del Arroz, Instituto de Ciencias Vegetales, Universidad de Leiden) y Guiderdoni (programa Biotrop, Cirad, Montpellier, Francia) hasta que se obtuvieron callos transformados. Para la etapa de selección posterior se aplicó el protocolo de Datta y Datta (2006). A continuación se describirán brevemente las etapas principales del procedimiento completo.

Preparación y desarrollo de callos embriogénicos a partir de escudete de arroz

La transformación del arroz se llevó a cabo mediante el uso de callos embriogénicos obtenidos a partir del escudete.

20 Para inducir la proliferación de los callos a partir del tejido de escudete, se usó el siguiente protocolo operativo:

- las semillas de arroz se descascararon (eliminación de la cáscara);
- para eliminar los patógenos y saprófitos contaminantes potenciales capaces de interferir con la producción de los callos, se desinfectaron las cariósides, sin la cáscara;
- 25     a. en el primer tratamiento de desinfección, las semillas descascaradas permanecieron durante 2 min en una disolución de etanol del 70%;
- b. tras el tratamiento con etanol, las semillas se transfirieron a una disolución del 5% de hipoclorito sódico con 2 gotas de detergente Tween20 y se mantuvieron así con agitación lenta durante 30 min;
- c. para eliminar cualquier rastro de hipoclorito sódico que pudiera inhibir la inducción de los callos en el escudete, se llevó a cabo una serie de lavados, en agua estéril, que duraron cada uno 15 min;
- 30 • tras el último lavado, las semillas se secaron sobre papel absorbente estéril;
- se colocaron 12 semillas por placa sobre la superficie del sustrato usado para inducir los callos (CIM, medio de inducción de callos), dispensado en un volumen de 25 mL en las placas Petri (Ø 90 mm);
- las placas así obtenidas se incubaron en la oscuridad, a una temperatura de 28 °C durante 21 días; tras 1 semana de incubación, se eliminaron el endospermo y las raicillas para estimular el desarrollo del callo a partir del escudete (el escudete se reconoce por su masa compacta, parcialmente incluida en el endospermo, de color amarillo);
- 35 • tras 3 semanas de inducción, el callo se transfirió a un sustrato CIM fresco, y después las masas de callos se disgregaron, sin usar escalpelos, siguiendo las líneas de fractura presentes de manera natural en el callo;
- el subcultivo continuó durante otros 10 días para desarrollar el callo embriogénico y hacerlo adecuado para la transformación.
- 40

Co-cultivo de los callos con *A. tumefaciens* EHA 105

1. Para obtener cantidades suficientes de *A. tumefaciens* para la transformación, las cepas que albergaban los vectores plasmídicos anteriormente mencionados (Ejemplo 2) se incubaron durante 3 días a 30 °C en agar LB;
- 45 2. cuando se cultivó *Agrobacterium*, las capas de células bacterianas se desprendieron y se suspendieron en el medio de co-cultivo líquido (CCML), hasta que se obtuvo una D.O.<sub>600</sub> de alrededor de 1,0, que correspondió a 3-5·10<sup>9</sup> células/mL;

3. los mejores callos, es decir, aquellos con un diámetro de alrededor de 2 mm, compactos y con un color blanquecino, se transfirieron a una placa Petri que contenía 35 mL de suspensión bacteriana y se dejaron sumergidos durante 15 min, con agitación;

4. después se secó el callo mediante el uso de papel absorbente estéril;

5. se transfirió un número máximo de 20 callos por placa Petri de borde elevado (Sarstedt) que contenía el sustrato semisólido para co-cultivo (CCMS, medio de co-cultivo solidificado);

6. los callos se incubaron después en un ambiente oscuro, a una temperatura de 25 °C durante 3 días.

Selección de los callos basada en el sistema de marcador de PMI

10 Tras el co-cultivo de los callos de arroz embriogénicos con *Agrobacterium*, se seleccionaron los tejidos transformados, mediante el uso del sistema de selección basado en PMI (fosfomanosa isomerasa) como marcador seleccionable y manosa como agente selectivo. Este método proporciona el uso de sustratos de cultivo que contienen concentraciones crecientes de manosa y concentraciones decrecientes de sacarosa.

El procedimiento usado fue el siguiente:

15 • transferir los callos del co-cultivo con *A. tumefaciens* a un sustrato PSM (medio de preselección) sin manosa y que contiene un 3% de sacarosa; incubación durante 1 semana en la oscuridad a una temperatura de 28 °C;

• transferir los callos a un sustrato SMI (medio de selección I) que contiene un 2% de sacarosa y 1,5% de manosa e incubación durante 2 semanas en la oscuridad a una temperatura de 28 °C;

• transferir los callos a un sustrato SMII (medio de selección II) que contiene un 1% de sacarosa y 2% de manosa e incubación durante 2 semanas en la oscuridad a una temperatura de 28 °C;

20 • a continuación se da la regeneración.

Regeneración de plantas de arroz a partir de los callos transformados

La regeneración de las plantas supuestamente transformadas se dio gracias a una estimulación hormonal adecuada del callo transformado siguiendo el procedimiento informado en la presente memoria:

25 1. los callos de arroz embriogénicos seleccionados se transfirieron a placas Petri de borde elevado que contenían el sustrato PRM (medio de pre-regeneración) que contenía un 0,5% de sacarosa y 2,5% de manosa, e incubación en la oscuridad durante 2 semanas a una temperatura de 28 °C;

2. tras el paso al sustrato PRM, los callos se transfirieron al sustrato RM (medio de regeneración), sin manosa, hasta un número máximo de 8-10 unidades por placa Petri de borde elevado. Las plantas se cultivaron con luz, a 28 °C durante 3-4 semanas.

30 3. cuando las plantas crecieron lo suficiente como para separarlas del callo (altura  $\geq$  3 cm), se transfirieron a tubos de cultivo que contenían 25 mL del medio de arraigo (rm);

4. El subcultivo en los tubos continuó durante alrededor de 3 semanas, siempre a alrededor de 28 °C, con luz;

5. al final del proceso de regeneración, las plantas se transfirieron a turba y se cultivaron en condiciones de invernadero.

35 Ejemplo 5: Extracción de las proteínas totales a partir del tejido de hojas de tabaco transformado mediante el ensayo de 4-MUG

El procedimiento descrito a continuación permite producir y conservar extractos de hojas de tabaco, conservando la actividad enzimática de la proteína GUS durante mucho tiempo.

40 • En un tubo de ensayo de 1,5 mL, pesar 15 mg de PVP (polivinilpirrolidona, PM  $>$ 40000 g/mol) y añadir 200  $\mu$ L de tampón de extracción (véase la Tabla 3), agitar en vórtex y dejar en incubación a 4 °C durante al menos 30 min;

• Extraer el jugo de las hojas mediante el uso de una prensa Meku Pollähne;

• Extraer 100  $\mu$ L de jugo y añadirlo a la mezcla de tampón-PVP, manteniendo todo en hielo;

• Centrifugar durante 15 min a 4 °C a 11500 xg;

• Retirar el sobrenadante (~200  $\mu$ L) y transferirlo muy rápidamente a un tubo de ensayo nuevo;

45 • Congelarlo inmediatamente mediante el uso de nitrógeno líquido y conservarlo a -80 °C.



Tabla 3: Composición del tampón de extracción

Componentes	Cantidades por 100 mL
NaHPO <sub>4</sub> pH 7,0	5 mL de reserva 1 M
DTT 5 mM	0,5 mL de reserva 1 M
Na <sub>2</sub> EDTA 1 mM	0,2 mL de reserva 0,5 M
Lauril Sarcosina sódica al 0,1%	1 mL de reserva al 10%
Triton X-100 al 0,1%	1 mL de reserva al 10%
En volumen con H <sub>2</sub> O	100 mL

5 El procedimiento se aplicó sin distinción a todas las muestras sometidas a análisis fluorimétrico. Cada planta transformada se analizó por triplicado mediante el uso de extractos tomados de 3 hojas (etapa de expansión avanzada) presentes en la parte apical de la planta.

#### Ejemplo 6: Ensayo fluorimétrico de 4-MUG

10 Para determinar el contenido de la enzima GUS en los extractos proteicos de hojas obtenidos de las plantas transformadas, se hizo un ensayo fluorimétrico específico. El sustrato usado fue 4-Metilumbeliferil-β-D-glucurónido (MUG), que genera el compuesto fluorescente 4-metilumbeliferona (4-MU) en presencia de la enzima GUS. El protocolo siguiente procedió del procedimiento estándar indicado por Jefferson (1987), y se adaptó para llevar a cabo el ensayo en placas.

- En una placa de 96 pocillos (baja adherencia, Sarstedt) añadir 10 µL de extracto de hojas a 130 µL de disolución de MUG (Tabla 4);
- Dejar en incubación durante 1 h a 37 °C;
- 15 • Retirar 20 µL de reacción y añadirlos rápidamente a 230 µL de Na<sub>2</sub>CO<sub>3</sub> 0,2 M (disolución de parada) en una placa opaca con 96 pocillos (repetir al menos dos veces por muestra);
- En la placa opaca, llevar a cabo una curva de calibración con 4-MU (1 mM y diluciones sucesivas 1:2 durante un total de 4-5 puntos);
- Leer los valores mediante el uso de un fluorímetro de placas;
- 20 • Procesar los resultados mediante el uso del programa informático Curve Fitting Data Analysis (Promega).

Tabla 4: Composición de la disolución de 4-MUG

Componentes	Cantidades por 100 mL
MUG (PM 352,3) 1,2 mM	0,042 g
Tampón de extracción GUS (Tabla 3)	100 mL

#### Ejemplo 7: Extracción de las proteínas totales a partir de semillas de arroz transformadas

25 Para obtener los extractos de proteínas totales a ensayar mediante el uso de DAS-ELISA, se desarrolló un protocolo de extracción que incluyó las siguientes etapas.

- Se tomaron las espigas maduras de cada individuo;
- Las espigas se secaron en un lugar seco y ventilado durante alrededor de 3 días hasta que se obtuvo una humedad relativa de la semilla del 14%;
- Toma de muestras aleatoria de 40 semillas para cada línea;
- 30 • Las semillas se descascararon con una descascaradora manual de arroz;
- La muestra se molió con un micro-molino de vibración MM2 (Retsch) a una velocidad de 20 Hz durante 2 minutos, y se retiraron 70 mg de la harina obtenida;
- La harina se homogeneizó en un mortero con 1 mL de tampón de extracción (Tris-HCl 50 mM, NaCl 0,5 M,

pH=7,0);

- Dilución posterior con otros 7 mL del mismo tampón;
  - Incubación mientras se agitaba continuamente a 4 °C durante 1 h;
  - 1 mL retirado y centrifugado a 20000 xg durante 40 min a 4 °C;
- 5 • Se recuperó la fase líquida que contenía las proteínas, y se conservó a -20 °C.

Ejemplo 8: Análisis mediante DAS-ELISA

El ensayo mediante DAS-ELISA, basado en un reconocimiento inmunológico doble, se usó para determinar el contenido de GCasi de los extractos proteicos individuales. Para el análisis, las muestras se diluyeron 1:30. A continuación se informa de las etapas principales del ensayo:

- 10 • Distribuir en cada pocillo 100 µL del anticuerpo policlonal sin conjugar anti-GCasi diluido a 2 ng/µL en una disolución de revestimiento (PBS diluido 1:5, azida sódica (0,01%);
- Incubar la placa durante la noche a 4 °C;
  - Retirar el anticuerpo;
  - Distribuir 250-300 µL de disolución de bloqueo (PBS + BSA 2,5% + azida sódica 0,01%) en cada pocillo;
- 15 • Incubar la placa a 25 °C durante 20 min;
- Retirar la disolución de bloqueo;
  - Distribuir 50 µL/pocillo de cada dilución del patrón (200, 100, 50 y 25 pg/µL de imiglucerasa comercial; Sanofi-Genzyme), de cada muestra a analizar y de la muestra de control, que consistió en la disolución de dilución (PBS + Tween20 0,1% +BSA 1%);
- 20 • Incubar la placa durante 30 min a 37 °C mientras se agita;
- Lavar los pocillos 3 veces con 300 µL/pocillo de disolución de lavado (PBS + Tween20 0,1 %);
  - Distribuir 50 µL/pocillo de anticuerpo policlonal anti-GCasi conjugado con peroxidasa de rábano, diluido a 0,4 ng/µL en disolución de dilución;
  - Incubar la placa durante 30 min a 37 °C mientras se agita;
- 25 • Lavar los pocillos 3 veces con 300 µL/pocillo de disolución de lavado (PBS + Tween20 0,1 %);
- Distribuir 100 µL/pocillo de disolución de TMB;
  - Incubar la placa durante alrededor de 10 min a 25 °C;
  - Parar la reacción con disolución de parada (ácido clorhídrico 1 M) 100 µL/pocillo;
  - Leer la placa a 450 nm con el lector de placas Modulus II (Promega);
- 30 • Procesar los datos mediante el uso del programa informático Curve Fitting Data Analysis (Promega), asignando los valores de concentración conocidos de los patrones. Los valores de concentración de las muestras se obtuvieron mediante el uso de una curva lineal con cuatro parámetros, considerando el factor de dilución adoptado para obtener las concentraciones reales de los extractos.

#### **Bibliografía**

- 35 Beerman RW, Jongens TA (2011) A non-canonical start codon in the *Drosophila* fragile X gene yields two functional isoforms. *Neuroscience* 181: 48-66
- Bradford M.M. (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254
- 40 Franks TM, Lykke-Andersen J (2008) The control of mRNA decapping and P-body formation. *Mol Cell.* 32(5): 605-615
- Gallie DR, Walbot V (1992) Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. *Nucleic Acids Res* 20: 4631-4638

Gerashchenko MV et al. (2010). CUG start codon generates thioredoxin/glutathione reductase isoforms in mouse testes. *The Journal of Biological Chemistry*, 285: 4595-4602

Guilley H et al. (1982) Transcription of Cauliflower Mosaic Virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* 30: 763-773

5 Jefferson RA, 1987. Assaying chimeric genes in plants: the GUS fusion gene system. *Plant Molecular Biology Reporter* 4: 387-405

Schmitz J et al. (1996) Non canonical translation mechanisms in plants: efficient in vitro and in planta initiation at AUU codons of the tobacco mosaic virus enhancer sequence. *Nucleic Acids Res* 24: 257-263

10 Simpson GG et al. (2010) Non-canonical translation initiation of the Arabidopsis flowering time and alternative polyadenylation regulator FCA. *The Plant Cell* 22: 3764-3777

Tyc K et al. (1984) Multiple ribosome binding to the 5'-terminal leader sequence of tobacco mosaic virus RNA. Assembly of an 80S ribosome X mRNA complex at the AUU codon. *Eur J Biochem.* 140(3): 503-511

**LISTA DE SECUENCIAS**

- <110> TRANSACTIVA S.R.L.
- 5 <120> Secuencia de ADN artificial con función de líder optimizada en 5' (5'-utr) para la sobreexpresión de proteínas recombinantes en plantas y método para la producción de proteínas recombinantes en plantas.
- <130> D4-3998
- 10 <150> UD2013A000002  
<151> 16-01-2013
- <160> 2
- 15 <170> PatentIn versión 3.5
- <210> 1  
<211> 75  
<212> DNA
- 20 <213> Artificial
- <220>  
<223> ADN artificial de una región 5'-UTR
- 25 <400> 1  
acacgaagtt tccaacaaca acaacaacaa caactctctc tctcaacaac aacaacaaca 60  
acaaagctct ctaga 75
- <210> 2  
<211> 77  
<212> ADN
- 30 <213> Artificial
- <220>  
<223> ADN artificial de una región 5'-UTR
- 35 <400> 2  
tcacatcaag ttccaacaa caacaacaac aacaactctc tctctcaaca aacaacaaca 60  
caacaaagct ctctaga 77

## REIVINDICACIONES

1. Un ADN artificial de una región líder 5'-UTR,  
dicho ADN artificial es eficaz para incrementar la expresión de proteínas recombinantes en plantas,  
dicho ADN artificial comprende, a lo largo de la dirección 5'→3', un sitio iniciador *Inr* y una secuencia consenso Kozak o similar a Kozak,  
5 dicho ADN artificial comprende además, entre el sitio iniciador *Inr* y la secuencia consenso Kozak o similar a Kozak:  
una diversidad de regiones poli(CAA) o (CAA)<sub>n</sub>, cada una formada por un oligonucleótido que consiste en dos o más copias de un elemento CAA contiguas entre sí,  
10 y una diversidad de regiones poli(CT) o (CT)<sub>m</sub> en el mismo número que las regiones poli(CAA), y cada una formada por un oligonucleótido que consiste en dos o más copias de un elemento CT contiguas entre sí,  
en el que al menos una, opcionalmente cada, región poli(CAA), en la dirección 5'→3', está antes de la región poli(CT), y al menos una región poli(CAA), en la dirección 5'→3', está contigua a una región poli(CT),  
con la condición de que dicho ADN artificial no contenga ninguno de los componentes siguientes: motivos ricos en A/T, elementos de trinucleótidos ATT, elementos de trinucleótidos CTG y tramos homopoliméricos, es decir,  
15 secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos idénticos.
2. El ADN artificial según la reivindicación 1, en el que n es un número entero, que se puede seleccionar igual o diferente entre las regiones poli(CAA), mayor o igual a 2, opcionalmente comprendido entre 3 y 9, opcionalmente entre 4 y 8, opcionalmente entre 5 y 7.
3. El ADN artificial según la reivindicación 1 o 2, en el que m es un número entero, que se puede seleccionar igual o diferente entre las regiones poli(CT), mayor o igual a 2, opcionalmente comprendido entre 3 y 5.  
20
4. El ADN artificial según cualquier reivindicación precedente, que contiene dos regiones poli(CAA) y dos regiones poli(CT), en el que una primera región poli(CAA) está antes de una primera región poli(CT) y una segunda región poli(CAA) está después de dicha primera región poli(CT) y antes de una segunda región poli(CT).
5. Un ADN artificial según cualquier reivindicación precedente, en el que el sitio iniciador *Inr* es el sitio de inicio de la transcripción 5'-ACACG-3' de 35S de CaMV o es un sitio iniciador *Inr* con la secuencia consenso 5'-YYANWYY-3', en la que:  
25  
Y=C, T;  
N=A, C, G, T;  
W=A, T.
6. Un ADN artificial según cualquier reivindicación precedente, que comprende la secuencia mostrada en SEQ ID N°: 1.  
30
7. Un ADN artificial según cualquier reivindicación de 1 a 5, que comprende la secuencia mostrada en SEQ ID N°: 2.
8. Un ADN artificial según cualquier reivindicación precedente, en el que la secuencia consenso Kozak o similar a Kozak es una secuencia que requiere la presencia de un elemento R que es una purina en posición 3 antes del codón de inicio de la traducción.  
35
9. El ADN artificial según cualquier reivindicación precedente, en el que los motivos ricos en A/T se definen como tramos o secuencias que consisten en más de 3, opcionalmente más de 4, nucleótidos de adenina (A) y/o timina (T), en cualquier combinación de los mismos.
- 40 10. El ADN artificial según cualquier reivindicación precedente, en el que dicho ADN artificial no contiene el octámero ACAATTAC.
11. Un vector de expresión que comprende un ADN artificial de una región líder 5'-UTR eficaz para incrementar la expresión de proteínas recombinantes en plantas según cualquier reivindicación de 1 a 10.
- 45 12. Un método para la producción de proteínas recombinantes en plantas, que comprende la transformación de las plantas mediante el uso de un vector de expresión según la reivindicación 11.

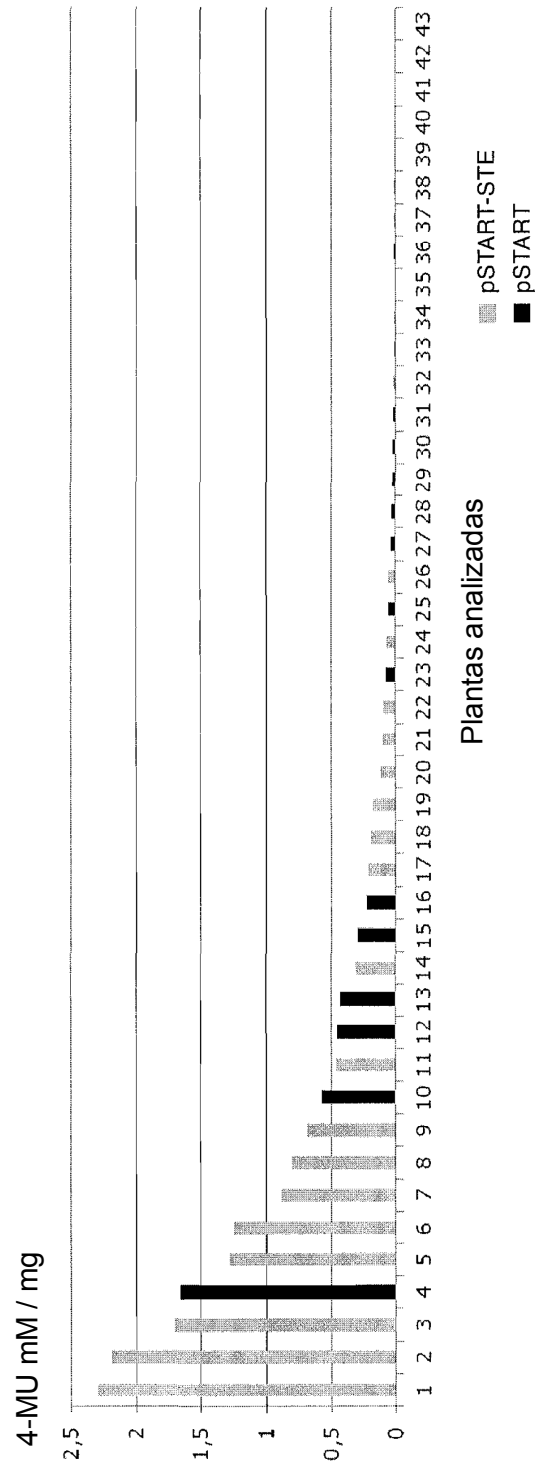


fig. 1

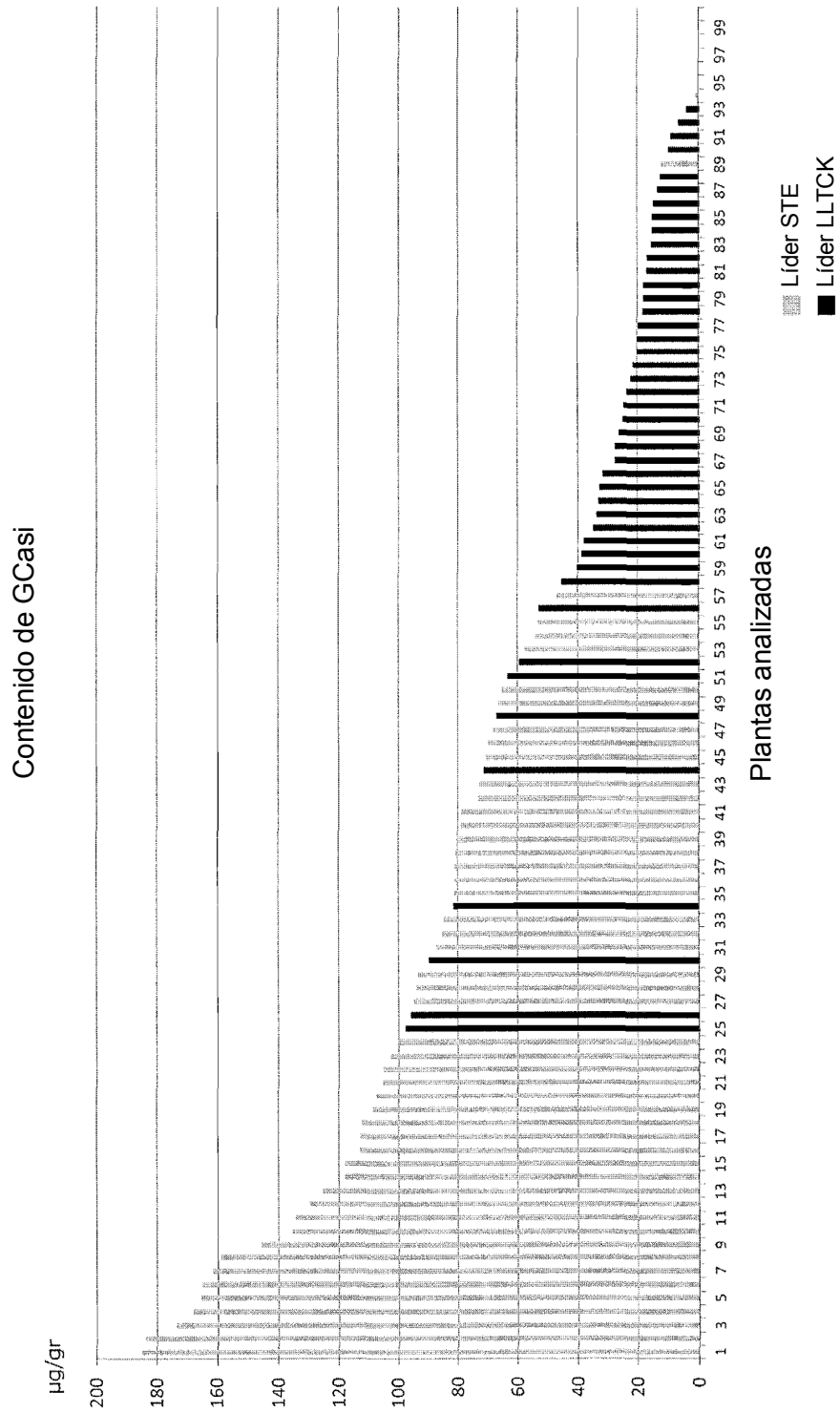


fig. 2

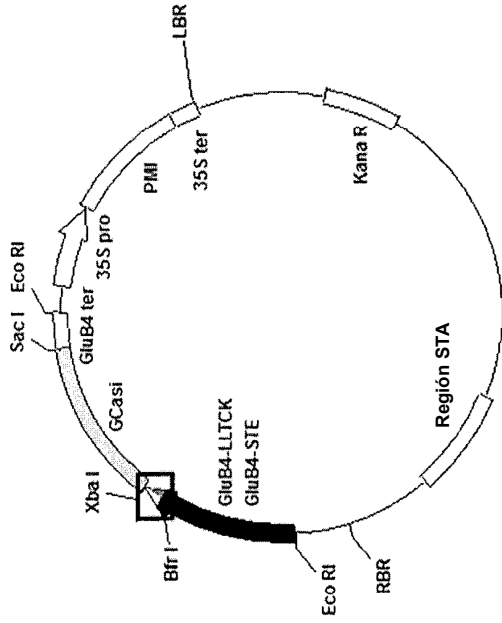


fig. 4A



fig. 4B

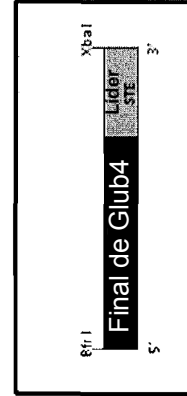


fig. 4C

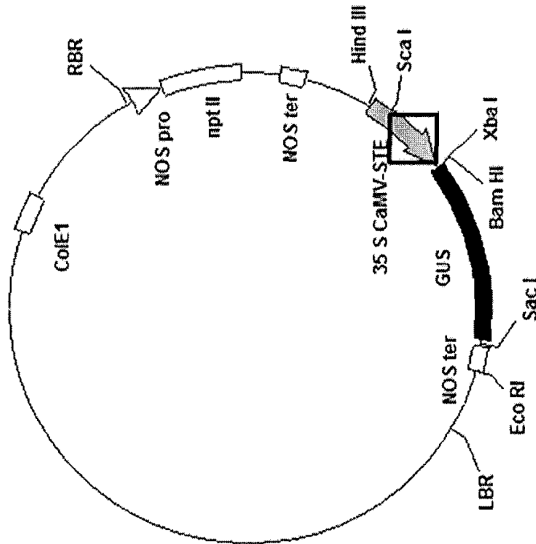


fig. 3A

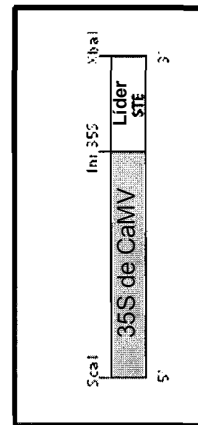


fig. 3B