

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 372 041**

51 Int. Cl.:  
**C12N 15/82** (2006.01)  
**C07K 14/65** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07008978 .4**  
96 Fecha de presentación: **03.05.2007**  
97 Número de publicación de la solicitud: **1865064**  
97 Fecha de publicación de la solicitud: **12.12.2007**

54 Título: **MÉTODO DE EXPRESIÓN DE PÉPTIDOS PEQUEÑOS USANDO PROTEÍNAS DE CEREAL QUE NO SON PROTEÍNAS DE ALMACENAMIENTO COMO PORTADOR DE FUSIÓN EN EL ENDOSPERMO Y USO DE LOS MISMOS.**

30 Prioridad:  
**08.06.2006 CN 200610019285**

45 Fecha de publicación de la mención BOPI:  
**13.01.2012**

45 Fecha de la publicación del folleto de la patente:  
**13.01.2012**

73 Titular/es:  
**HEALTHGEN BIOTECHNOLOGY INC. IN WUHAN  
SPECIAL 1ST, GUIYUAN ROAD WUHAN  
UNIVERSITY, LUOJIA HILL  
WUHAN HUBEI, CN**

72 Inventor/es:  
**Yang, Daichang y  
Xie, Tingting**

74 Agente: **Miltenyi, Peter**

ES 2 372 041 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

Método de expresión de péptidos pequeños usando proteínas de cereal que no son proteínas de almacenamiento como portador de fusión en el endospermo y uso de los mismos

## Campo de la invención

5 La presente invención se refiere al campo de la ingeniería genética. Particularmente, la presente invención se refiere a un método de expresión de péptidos pequeños en el endospermo usando proteínas de cereal que no son proteínas de almacenamiento como portadores de fusión y el uso de los mismos, es decir, un método en el que se usaron células de endospermo de cereales tales como arroz y cebada como biorreactores y se aplicó la estrategia de proteína de fusión para introducir el casete de expresión específico del endospermo usando cualquier proteína  
10 que no es una proteína de almacenamiento (por ejemplo, Bip y PDI de cultivos de cereales) como las parejas de fusión en las células de arroz o cebada, lo que conduce a la acumulación masiva de los péptidos pequeños en células de endospermo de un arroz o cebada transgénico. La presente invención también se refiere a péptidos de origen vegetal producidos mediante el método.

## Antecedentes de la invención

15 Se denomina péptidos pequeños a aquéllos con menos de 100 aminoácidos de longitud. En los últimos años, se usaron ampliamente péptidos pequeños en los campos de la medicina, tratamiento de enfermedades, vacunas moleculares, etc., lo que incluye antígenos de superficie, diagnóstico de enfermedades y tratamiento de SIDA y cáncer. Con el rápido desarrollo de la biotecnología, se descubrieron cada vez más péptidos pequeños. Por consiguiente, se necesitan una gran cantidad de péptidos para satisfacer el requisito en una variedad de industrias  
20 tales como investigación funcional, experimentos clínicos y tratamiento de enfermedades. Generalmente, la síntesis química es la vía principal para producir péptidos con menos de 40 aminoácidos de longitud. Durante el proceso de síntesis química, debido a la existencia de ciertas modificaciones químicas y reacciones químicas incompletas, incluso péptidos con menos de 40 aminoácidos son difíciles de sintetizar (Dobeli, *et al.*, 1998, Protein Expression & Purification, Vol, 12: 404-414). Por tanto, resultó ventajoso y necesario utilizar un biosistema para producir péptidos.

25 En la década de 1950, se usaban bacterias como biorreactores para producir productos farmacéuticos. Sin embargo, debido a que las bacterias son procariontes, que no poseen el sistema de procesamiento del eucariota, su aplicación está limitada seriamente para algunas proteínas cuyas bioactividades se basan en la modificación de proteínas. Como biorreactor de segunda generación, se ha usado la levadura en la producción de productos de medicina desde la década de 1970. Sin embargo, los problemas de bajo rendimiento y procesamiento/modificación incompleta limitaron seriamente el uso extensivo de la levadura. El biorreactor de tercera generación utilizó células animales y de plantas superiores. En la actualidad, los biorreactores eucariotas se clasifican en biorreactores animales y biorreactores vegetales, y el biorreactor animal incluye además cultivos celulares y animales transgénicos. Actualmente, los anticuerpos principales para uso farmacéutico se producen mediante el cultivo de líneas celulares de CHO (o de ratón). Los estudios de animales transgénicos principalmente se centran en la expresión de proteína recombinante en células mamarias de vaca transgénica o albúmina de huevo. Sin embargo, los problemas de contaminación por patógenos animales, alto coste y alta necesidad de inversión limitan gravemente su uso. Se estima que la capacidad de producción máxima de anticuerpo monoclonal es de aproximadamente 1.000 kg por año a nivel mundial. Para obtener otros 1.000 kg, se necesitan adicionalmente 40 mil millones de dólares estadounidenses de inversión y más de diez años de tiempo. Todos estos datos indican que el actual sistema productivo y la capacidad de producción de la proteína recombinante están muy lejos de satisfacer la necesidad del mercado. Por consiguiente, se necesita un sistema de expresión altamente eficaz y seguro para satisfacer la gran demanda del mercado de producción de péptidos pequeños.

45 En la mayoría de los casos, se necesitan péptidos con al menos 80 aminoácidos para la expresión de proteína recombinante. Incluso en tal caso, el nivel de expresión es bastante bajo. Por tanto, una manera principal para mejorar el nivel de expresión del péptido es usar una estrategia de proteína de fusión. Hasta ahora, los estudios en sistemas de expresión de proteína de fusión son adecuados principalmente para sistemas de levadura y *Escherichia coli* (*E. coli*). Por ejemplo, se usaron proteína de unión a maltosa (MBP), FLAG (Einhauer *et al.*, 2001, J. Biochem. Biophys. Methods, Vol 49:455-465) y glutatión (GST) (Papaioannou *et al.*, 2002, Protein Expression & Purification, Vol.13: 462-466) como parejas de fusión en *E. coli* y levadura. Aunque algunos sistemas de expresión de proteína de fusión se han comercializado, se usaron meramente en investigaciones básicas en laboratorios. Los estudios sobre la expresión de péptidos en un sistema de expresión vegetal son relativamente nuevos y no han logrado mucho progreso hasta ahora. Recientemente, se ha notificado que se usó el portador de fusión de la dismutasa unida a disulfuro (PDI) y proteína fluorescente verde (GFP) para expresar un péptido. Desafortunadamente, el nivel de expresión fue bastante bajo. Por otro lado, muchos péptidos se han expresado satisfactoriamente en sistemas de *Escherichia coli* y levadura, existiendo un riesgo obvio de contaminación por patógenos de los huéspedes. Además, los problemas de nivel de expresión bajo y formación de cuerpos de inclusión insolubles en células de *E. coli*, y el peso molecular superior de las parejas de fusión usadas en procariontes provocan problemas al procesamiento posterior y por tanto no son adecuados para usarse en eucariotas, especialmente en células de plantas superiores. Aunque se han usado células vegetales para expresar péptidos, el

nivel de expresión bajo siempre ha sido un problema que ha obstaculizado las investigaciones.

Debido a los defectos y las limitaciones mencionados anteriormente, desarrollar vectores de expresión de proteína de fusión en plantas superiores se vuelve cada vez más importante. Usar una planta superior como biorreactor tiene las ventajas del bajo coste, expresión de nivel alto, fácil de escalar, libre de contaminación por patógenos, etc., convirtiéndose en un candidato prometedor para la futura producción de péptidos. Hasta ahora, debido a los problemas de peso molecular relativamente alto, ausencia de señales de transporte a orgánulos celulares, etc., los sistemas de expresión de fusión de procariotas no son adecuados para expresiones en células de plantas superiores. Por tanto, es importante explorar y desarrollar un sistema de expresión de péptidos pequeños que sea adecuado para plantas superiores. Usando la proteína de almacenamiento de arroz como portador de fusión, Ventría Bioscience Inc. en Estados Unidos ha expresado satisfactoriamente péptidos pequeños. No obstante, aunque se logró una expresión de nivel superior por la compañía con el uso de globulina como portador de proteína de fusión, su aplicación era limitada en muchos aspectos puesto que el uso de globulina para expresar péptidos provocó problemas de solubilidad. Aparte de las proteínas de almacenamiento, otras dos proteínas expresadas masivamente en el endospermo del arroz son la proteína de unión a retículo endoplásmico (BiP) y la proteína dismutasa unida a disulfuro (PDI), que están almacenadas ambas en el cuerpo de proteína I. El extremo C-terminal de la proteína Bip tiene actividad chaperona molecular, facilitando el plegado correcto de la proteína en una conformación de proteína funcional. El uso del extremo C-terminal de la proteína Bip como portador de fusión puede no sólo acumular la proteína de fusión dentro de los cuerpos de proteína (similar a la acumulación de la proteína de almacenamiento en cuerpos de proteína), sino también aumentar la solubilidad de la proteína de fusión, superando de ese modo el problema de insolubilidad que resulta del uso de la proteína de almacenamiento como portador de fusión en métodos convencionales. Otra proteína que no es una proteína de almacenamiento que se expresa altamente en el endospermo de arroz o trigo es proteína disulfuro isomerasa (PDI). La PDI tiene dos funciones. Una es la actividad disulfuro dismutasa en el extremo N-terminal, y la otra es que el extremo C-terminal tiene actividad chaperona molecular. Por consiguiente, el uso de su extremo C-terminal como portador de proteína de fusión puede lograr también los propósitos de mejorar tanto la expresión como la solubilidad de la proteína. Usando el extremo C-terminal de proteínas que no son proteínas de almacenamiento PDI y Bip como portador de fusión que se expresan específicamente en el endospermo, la presente invención puede no solamente mejorar el nivel de expresión de proteína sino también superar el problema de solubilidad que existía en otras patentes internacionales en las que se usaron proteínas de almacenamiento como proteínas de fusión, confiriendo de ese modo un carácter innovador y patentabilidad a la presente invención.

El factor de crecimiento similar a la insulina (IGF) es uno de los factores de crecimiento más importantes involucrados en diversos tipos de proliferaciones y metabolismos. No sólo asume un papel importante en el crecimiento del esqueleto humano, sino que también facilita la maduración de células relevantes y se asocia con la cicatrización. El IGF-1 es un péptido monocatenario con 70 aminoácidos, con 3 enlaces disulfuro y sin sitio de glicosilación (De Bree, *et al.*, 1998, Protein Expression & Purification, Vol:13, 319-325). Basándose en el análisis de la posición del enlace disulfuro reconocible, se cree que la estructura secundaria de IGF-1 podría ser similar a la de la insulina, que tienen ambos las mismas glicinas conservativas en las mismas posiciones y tienen un núcleo de residuo de aminoácido no polar similar. El IGF-1 se usa ampliamente en la práctica clínica. El IGF-1 humano recombinante (rhIGF-1) y su complejo se han usado eficazmente en el tratamiento del síndrome de insensibilidad a la hormona del crecimiento (GHIS), que incluye deficiencia del receptor de GH, deficiencia del gen IGF y el bloqueo de la ruta de transducción de señales de la hormona de crecimiento. Además, el IGF-1 se ha usado para tratar pacientes que padecían de diabetes tipo I o tipo II o pacientes con síntomas de resistencia a insulina graves. Con la administración de rhIGF-1, los síntomas se aliviaron en gran medida. El rhIGF-1 o su complejo rhIGF-1/IGFBP-3 pueden usarse adicionalmente para tratar inflamación crónica, trastorno de nutrición y otros estados tales como enfermedad de Crohn (también llamada enteritis segmentaria), artritis crónica juvenil, fibrosis de vejiga/vesícula biliar, etc. Los estudios relevantes sobre la farmacodinámica de los IGF son muy limitados, sin embargo, se cree que la escasez de suministro de IGF es uno de los problemas principales (Savage, *et al.*, 2005, Edocr Development, Vol. 9: 100-106).

Yasuda Hiroshi *et al.* (2005) Transgenic Research, 14:677-684 dan a conocer la expresión de una proteína de fusión entre GFP y el péptido similar al glucagón 1 (GLP-1) bajo el control del promotor de glutelina de proteína de almacenamiento del arroz (GluB-1) en las semillas de plantas de arroz transgénicas.

El documento WO 01/72959 da a conocer la expresión de una proteína heteróloga tal como insulina en los cloroplastos de plantas de tabaco transgénicas.

Hosoyama Hiroshi *et al.* (1995) Plant Cell Reports, 15:174-177 dan a conocer la generación de una planta de arroz transgénico que expresa una proteína de fusión entre orizacistatina (OC) y betaglucuronidasa (GUS).

El documento WO 2004/022697 da a conocer una proteína novedosa homo-oligomérica, resistente a la proteasa, estable a la desnaturalización, de una planta y su uso en la expresión de proteínas.

El documento WO 2005/056578 da a conocer el uso de proteínas de almacenamiento de semilla de planta monocotiledónea como portadores de proteína de fusión para la expresión de péptidos o polipéptidos heterólogos en

plantas transgénicas.

Liu Yang *et al.* (2005) *Protein Expression and Purification*, 44:155-161 dan a conocer que la fusión entre un factor transcripcional específico de planta DREBIII-1 y el extremo N- o C-terminal de la proteína disulfuro isomerasa (PDI) humana se expresó como una proteína soluble en *E. coli*.

- 5 El documento CN-A-1 884 517 / EP-A1-1 621 556 da a conocer la construcción de una proteína de fusión compuesta por una chaperona molecular, una proteína interviniente que tiene una actividad de escindir un enlace peptídico, y una proteína diana para expresar una proteína que apenas puede expresarse mediante técnicas de ADN recombinante habituales.

### Descripción de la invención

- 10 Por tanto, es un objeto solucionar los defectos de expresión de nivel bajo, poca solubilidad, bioactividades pobres, falta de seguridad, etc. que existían en los biorreactores convencionales en los que se usaban organismos procariontes y eucariotes como huéspedes. Puesto que el extremo C-terminal de Bip o PDI (proteínas de cereales que no son proteínas de almacenamiento) tiene función de chaperona molecular, se usaron como portadores de fusión para fusionarse con péptidos diana pequeños. Bajo el control del promotor específico del endospermo del arroz y el péptido señal, la proteína de fusión expresada entró al sistema de endomembrana de las células de endospermo del arroz y se almacenó en el cuerpo de proteína del endospermo del arroz. Como resultado, una gran cantidad de proteínas de fusión se acumularon a un nivel alto en las semillas de arroz. La presente invención puede no solamente solucionar los problemas de expresión de nivel bajo, poca solubilidad, bioactividades pobres, falta de seguridad, etc. que existían en los sistemas de expresión convencionales, sino también evitar el problema de contaminación por patógenos de células animales.

- 20 Un objeto es proporcionar un método, en el que se usen el promotor y el péptido señal del gen de glutelina de arroz Gt13a para preparar el casete de expresión de la proteína de fusión expresado específicamente y almacenado en el cuerpo de proteína de la célula de endospermo del arroz. El método puede evitar que la proteína de fusión se vea atacada por proteasas en el citoplasma y permitir la acumulación de la proteína en el endospermo del arroz, produciendo de ese modo la proteína con alto rendimiento.

- 25 Utilizando el dominio funcional de chaperona molecular del extremo C-terminal de Bip y PDI de la proteína de cereal que no es una proteína de almacenamiento como portador de fusión, otro objeto es mejorar tanto el nivel de expresión como la solubilidad de la proteína de fusión en células de endospermo. Un objeto adicional es establecer una nueva plataforma técnica para la expresión de manera eficaz de péptidos pequeños en células de endospermo de cereales tales como arroz o cebada. Con esta plataforma, es fácil obtener un sistema de expresión de este tipo, que es más seguro, está libre de patógenos y es fácil de escalar en comparación con aquellos de pollo y vaca transgénicos. Además, también es rentable y puede producir proteína de fusión con alto rendimiento.

- 30 Es otro objeto proporcionar el uso de la proteína de cereal que no es una proteína de almacenamiento como portador de fusión para expresar IGF-1 en endospermos de arroz y cebada. Se optimizaron el portador de fusión y genes diana para dar codones genéticos de preferencia de arroz, de modo que el nivel de expresión podía mejorarse en células de endospermo de arroz.

#### A. Obtención del promotor específico del endospermo y péptido señal a partir del genoma del arroz

- 40 A través de análisis de información biológica, se encontró un promotor fuerte del gen Gt13a, un miembro de la familia génica de la glutelina de arroz. Con el fin de obtener un promotor Gt13a y su secuencia de péptido señal, se sintetizaron un par de cebadores (SEQ ID NO: 1 y SEQ ID NO:2 de la solicitud de patente china número 200510019084.4), según la información proporcionada por el Genbank Gt13a (número de acceso: AP003256) para amplificación por PCR. Para la conveniencia en el clonaje, se añadió un sitio de restricción para extremo cohesivo al extremo 5' del cebador directo, y se añadió otro sitio de restricción para extremo como al extremo 3' del cebador inverso. Se usó ADN genómico extraído de las hojas de la variedad de arroz Taipei 309 como molde para la PCR. Usando los cebadores, se obtuvo un fragmento de ADN con 1284 pares de bases de longitud siguiendo un protocolo de PCR convencional. El análisis de la secuenciación de ADN indica que el fragmento es idéntico a la secuencia de Genbank y tiene una estructura de promotor típica. Por tanto, se produjeron el promotor Gt13a y la secuencia de péptido señal que pueden usarse para controlar la expresión de la proteína recombinante en célula de endospermo de cereal (SEQ ID NO: 1).

- 50 B. Construcción del casete de expresión específico del endospermo de arroz

- Tras la obtención del promotor Gt13a y la secuencia de péptido señal tal como se describe anteriormente, se digirió el producto de PCR con endonucleasas de restricción de extremo cohesivo y de extremo como, seguido por una reacción de ligación con pBI221 (Clontech) digerido por las mismas endonucleasas de restricción. Se sometió a electroporación la mezcla de ligación en la cepa DH10B de *Escherichia coli* (Invitrogen). El plásmido resultante denominado pOsPMP2 (véase la figura 1) contiene el promotor Gt13a, el péptido señal Gt13a y el terminador Nos, con la secuencia de nucleótidos mostrada en SEQ ID NO.2.

C. Optimización del codón genético del portador de fusión y gen diana y síntesis de los mismos

5 Se obtuvieron las secuencias de nucleótidos C-terminales del gen Bip de arroz (número de acceso de Genebank: AAB63469), gen PDI de trigo (número de acceso de Genebank: AJ277377) y gen del factor de crecimiento similar a la insulina 1 (número de acceso de Genebank: CAA01955) del Genebank del National Center of Biotechnology Information (NCBI). Se usó el software analítico MacVector (producto de Accelrys) para convertir los genes en codones genéticos de preferencia de arroz. Entonces, los genes optimizados se sintetizaron por Blue Heron Biotechnology Incorporation en Estados Unidos. El extremo C-terminal del gen Bip de arroz de preferencia de arroz así obtenido tiene la secuencia mostrada en SEQ ID NO: 3. El extremo C-terminal del gen PDI de trigo de preferencia de arroz así obtenido tiene la secuencia mostrada en SEQ ID NO: 4. Las secuencias del gen de IGF-1 humano de preferencia de arroz así obtenido tiene la secuencia mostrada en SEQ ID NO: 5. En comparación con sus secuencias de nucleótidos originales, la razón de secuencias cambiadas de los genes optimizados oscila desde el 11,2 hasta el 21,4%, y la de los codones genéticos oscila desde el 30,5-54,3%. Por el contrario, las secuencias de aminoácidos se mantienen sin cambios (véase la tabla 1).

Tabla 1. Comparación entre los portadores de fusión optimizados y el gen diana

Elementos	Bip-C	PDI-C	IGF-1
Nos de códigos genéticos en total	256	133	70
Nos de códigos genéticos cambiados	78	64	38
Razón de códigos genéticos cambiados (%)	30,5	48,1	54,3
Nos de desoxinucleótidos en total	768	399	210
Nos de desoxinucleótidos cambiados	86	67	45
Razón de desoxinucleótidos cambiados (%)	11,2	16,8	21,4
Razón de aminoácidos cambiados (%)	0	0	0

15 Para la conveniencia en el clonaje de genes, se añadieron sitios de restricción de un extremo romo y de un extremo cohesivo a los extremos 5' y 3' respectivamente durante la síntesis de genes.

D. Construcción de diversos vectores para la expresión de proteínas de fusión

20 **1). Construcción de pOsPMP25 (Gt13a-PDIC-IGF-1-Nos):** En primer lugar, se digirió el ADN de pOsPMP2 por *MscI* y *XhoI*. Entonces, se clonó el gen optimizado de IGF-1 humano amplificado mediante amplificación por PCR en pOsPMP2. Se usó el resultante para transformar la cepa DH10B de *E. coli* para producir el pOsPMP3 resultante, un plásmido intermedio que contiene el gen de IGF-1. Se digirió adicionalmente el ADN de pOsPMP3 por *NaeI* y *NcoI*, tras lo cual se clonó el fragmento de ADN de *PDIC* amplificado por PCR en pOsPMP3, el plásmido resultante se denominó pOsPMP25 (Gt13a-BipC-IGF-1-Nos).

25 **2). Construcción de pOsPMP26 (Gt13a-BipC-IGF-1-Nos):** En primer lugar, se digirió el ADN de pOsPMP3 por *NaeI* y *NcoI*. Entonces, se clonó el ADN de BipC en pOsPMP3 tras amplificación por PCR para producir el plásmido resultante pOsPMP26 (Gt13a-BipC-IGF-1-Nos).

30 **3). Construcción de vector con genes marcadores seleccionables:** Se usó el promotor de cisteína proteinasa β (CP) para mediar en la expresión específica de callo de un gen marcador seleccionable, que codifica para higromicina fosfotransferasa durante el cultivo de tejido tras la transformación. Se sintetizaron un par de cebadores (SEQ ID NO.6 y SEQ ID NO.7), con sitios de restricción *HindIII* y *SmaI* añadidos a los extremos de cada cebador. Se usó el genoma de la variedad de arroz Taipei 309 como molde de ADN basándose en un protocolo de PCR convencional, y se obtuvo un fragmento de 1.103 pb que contenía el promotor mediante amplificación por PCR. Se digirió el fragmento con *HindIII* y *SmaI*, y entonces se ligó con pBI221 (Clontech) que se digirió por las mismas enzimas de restricción. Se usó el resultante que contenía la secuencia del promotor de CP para transformar la cepa DH10B de *E. coli* para producir el plásmido intermedio denominado pOsPMP4. Se usó higromicina β fosfotransferasa (Hpt) como marcador seleccionable, que se obtiene de pCAMBIA 1301 (CAMBIA company de Australia). Para obtener el fragmento génico de Hpt, se usó pCAMBIA 1301 como molde siguiendo el protocolo de PCR convencional, con un cebador directo (SEQ ID NO.8) y un cebador inverso (SEQ ID NO.9), en el que se añadió un sitio de restricción *SmaI* al cebador directo y se añadió un sitio de restricción *XhoI* al cebador inverso. Se digirió el fragmento génico resultante de Hpt con *SmaI* y *XhoI*, mientras que el ADN de pOsPMP4 se digirió con *NaeI* y *XhoI*. Entonces, se ligó el fragmento génico de Hpt con el fragmento de ADN de pOsPMP4 (que contiene el promotor de CP) digerido por *NaeI* y *XhoI*. Finalmente, se usó el resultante para transformar la cepa DH10B de *E. coli* para producir pOsPMP5, un vector de expresión que guía un marcador seleccionable expresado específicamente en el

callo del arroz.

Transformación genética del gen del arroz:

5 Se retiraron las cascarillas de semilla de arroz y se esterilizó la semilla en hipoclorito de sodio al 20% durante 20 minutos, seguido por aclarado tres veces con agua esterilizada. Entonces, se cultivó la semilla en medio de cultivo de inducción de callo durante 20-25 días para producir callos. Entonces se transfirieron los callos inducidos a un medio PR para su cultivo durante otros 9-10 días antes de usarse para una futura transformación. Para el recubrimiento de ADN, se mezclaron 0,5 µg de pOsPMP25 o pOsPMP26 y 0,5 µg de pOsPMP5, que tienen marcadores seleccionables, con 50 µl de partícula de oro, 250 µl de CaCl<sub>2</sub> 1 M y 50 µl de espermidina 0,1 M rotando durante 30 minutos. Tras lavar tres veces con etanol al 100%, se recubrieron las partículas de oro con el ADN. 10 Siguiendo el protocolo Genegun proporcionado por DuPont Inc. Estados Unidos, se cotransformaron los dos plásmidos en los callos de Taipei 309. Tras 45 días de selección en el medio selectivo que contenía Higromicina B, se transfirieron adicionalmente callos positivos resistentes a higromicina B a un medio de regeneración en condición de luz durante aproximadamente 20 días. Cuando el callo se diferenció para dar una plántula, se transfirieron a un medio de enraizado durante otros 15-20 días. Las plantas transgénicas obtenidas se sometieron a prueba mediante PCR usando cebadores específicos para el gen diana. Tras confirmarse mediante PCR, las plántulas se cultivaron en un invernadero para producir semilla, que se denominó generación T0. 15

Selección de líneas transgénicas de expresión alta

20 Tras aproximadamente cuatro meses de crecimiento, la planta de arroz transgénica empezó a florecer y producir semillas T1. Un mes después, la generación T0 de la planta de arroz transgénica empezó la maduración, produciendo semillas el 50%-60%. Después de recoger las semillas T1, se seleccionó la línea transgénica de expresión alta mediante inmunotransferencia de tipo Western a partir de extractos brutos de endospermos del arroz. El nivel de expresión se analizó con ELISA de cuantificación de proteína (US R&D system Inc.). Las líneas transgénicas de expresión alta se seleccionaron de manera continua durante 1-2 generaciones para obtener líneas transgénicas genéticamente estables. Entonces, la línea se usó para la producción a gran escala de la proteína de fusión. 25

30 Se usaron los endospermos de los cultivos de cereales tales como arroz y cebada para producir una proteína de fusión bioactiva soluble. El nivel de expresión de la proteína de fusión que se obtiene usando este sistema de expresión es de al menos el 0,3% en peso seco de arroz pardo, es decir, más de tres gramos de proteína de fusión por kg de semillas. El nivel de expresión es aproximadamente 20 veces el de un sistema de expresión de cloroplasto, y aproximadamente 500 veces el de un sistema de expresión de tubérculo de patata.

Usando los endospermos de cultivos de cereales tales como arroz y cebada como biorreactores para producir péptidos pequeños, la presente invención supera los defectos de expresión de nivel bajo, coste alto, poca solubilidad; seguridad pobre, etc. que existían en otros sistemas de expresión de plantas, de microorganismos y de animales convencionales. 35

Usando proteína de cereal que no es una proteína de almacenamiento como portador de fusión, la presente invención supera eficazmente la insolubilidad y dificultad de los problemas de procesamiento posterior de proteína de fusión expresada descritos en la patente estadounidense titulada "High-Level Expression of Fusion Polypeptides in Plant Seeds Utilizing Seed-Storage Proteins as Fusion Carriers" (US60/527-753). 40

El sistema de proteína de fusión de la proteína que no es una proteína de almacenamiento ha logrado tanto un nivel de expresión alto (20-500 veces más que aquéllos en métodos convencionales) como una alta solubilidad de la proteína. Soluciona eficazmente el problema de solubilidad provocado por la expresión de nivel alto. La proteína de fusión expresada tiene bioactividades incluso sin eliminar el portador de fusión, lo que puede reducir el coste hasta el 40-50%. Las características incluyen: 45

a. Usar proteínas de semilla de cereal que no son proteínas de almacenamiento como portador de fusión para expresar diversos péptidos para el cuidado de la salud y medicinales en semillas de cereales, con un nivel de expresión de más del 0,3% en peso seco de semilla. 50

b. Usar los promotores y péptidos señal del gen de la proteína de almacenamiento de cultivos de monocotiledóneas para expresar específicamente diversos péptidos para el cuidado de la salud y medicinales en semilla; como ejemplo, el péptido pequeño de IGF-1 expresado en el endospermo del arroz alcanza el 0,75% en peso seco de la semilla. 55

c. Dichos diversos péptidos medicinales expresados en semillas se refieren a péptidos con 20-100 aminoácidos, que incluyen péptidos con una variedad de usos para el cuidado de la salud y de manera terapéutica, diversos tipos de péptidos antitumorales, diversos tipos de péptidos antibacterianos, así como otros péptidos que tienen funciones para el cuidado de la salud y de tratamiento para los organismos humanos.

d. Para expresar el factor de crecimiento similar a la insulina 1 humano recombinante en semillas de arroz/cebada

Abreviaturas de los vectores de expresión:

pOsPMP2----un vector de expresión génica específico del endospermo del arroz

pOsBipC-----un plásmido que porta el gen de proteína de fusión Bip-IGF-1 recombinante

pOsPDIC-----un plásmido que porta el gen de proteína de fusión PDI-IGF-1 recombinante

5 pOsPMP3-----un plásmido que porta códigos de preferencia de arroz del gen de IGF-1

pOsPMP25-----vectores de expresión específicos del endospermo del arroz que portan el gen de proteína de fusión recombinante (Bip-C-IGF)

pOsPMP26----- vectores de expresión específicos del endospermo del arroz que portan el gen de proteína de fusión recombinante (PDI-C-IGF)

10 pOsPMP5----- vectores de expresión específicos del callo de arroz que portan un marcador seleccionable

### Breve descripción de los dibujos

La figura 1 es el mapa de restricción del vector de expresión génica específico del endospermo del arroz pOsPMP2 (Gt13a Sp-Stuff-Nos).

La figura 2 es el mapa de restricción del vector pOsPMP25 (Gt13a-BipC-IGF-1).

15 La figura 3 es el mapa de restricción del vector pOsPMP26 (Gt13a-PDIC-IGF-1).

La figura 4 es el mapa de restricción de vector de marcador selectivo pOsPMP05 (CP-Hpt-Nos).

La figura 5 es un ejemplo que muestra la tinción con azul de Coomassie del gel de poliacrilamida de extractos de semilla de arroz, en el que la flecha indica que la proteína de fusión BipC-IGF-1 puede verse claramente mientras que no había una banda correspondiente en la muestra de Taipei 309 o la cepa aislada genéticamente única (control).

20 La figura 6 muestra la inmunotransferencia de tipo Western usando anticuerpo específico de IGF-1, en la que la flecha indica que la proteína de fusión BipC-IGF-1 puede verse claramente en el endospermo transgénico. Por el contrario, no había banda correspondiente en la muestra de Taipei 309 o la cepa aislada genéticamente única (control).

25 La figura 7 muestra el resultado de la inmunotransferencia de tipo Western cuando se usa el anticuerpo específico de Bip, en el que la flecha indica la proteína de fusión BipC-IGF-1 y la proteína Bip no de fusión endógena. La proteína de fusión BipC-IGF-1 y la proteína Bip endógena existen en el endospermo transgénico. Por el contrario, sólo había proteína Bip endógena, pero no proteína de fusión recombinante Bip-IGF-1, en la muestra de Taipei 309 o la cepa aislada genéticamente única (control).

30 La figura 8 es un ejemplo que muestra la tinción con azul de Coomassie del gel de poliacrilamida, en el que la flecha indica que la proteína de fusión PDIC-IGF-1 puede verse claramente. Por el contrario, no había banda correspondiente en la muestra de Taipei 309 o la cepa aislada genéticamente única (control).

35 La figura 9 muestra el resultado de la inmunotransferencia de tipo Western cuando se usa el anticuerpo específico de IGF-1, en el que la flecha indica que la proteína de fusión PDIC-IGF-1 puede verse claramente en el endospermo transgénico. Por el contrario, no había banda correspondiente en la muestra de Taipei 309 o la cepa aislada genéticamente única (control).

40 Las figuras 5 y 6 ilustran los resultados de la tinción del gel de poliacrilamida e inmunotransferencia de tipo Western de la proteína de fusión recombinante expresada en el endospermo del arroz. La proteína de fusión recombinante se extrajo del endospermo del arroz transgénico. Se obtuvieron once generaciones T1 de la cepa transgénica n.º 26-13, y se usó 1 ml de tampón de extracción de proteína para extraer la proteína total de una única semilla. Se cargó una muestra de diez µl en el gel de poliacrilamida al 12%. Tras la electroforesis, se usaron la tinción con azul de Coomassie y la inmunotransferencia de tipo Western para mostrar los resultados.

45 Las figuras 8 y 9 ilustran los resultados de la tinción del gel de poliacrilamida y la inmunotransferencia de tipo Western de la proteína de fusión recombinante expresada en el endospermo del arroz. Se extrajo la proteína de fusión recombinante del endospermo del arroz transgénico. Se obtuvieron trece semillas T1 de la cepa transgénica n.º 25-12, y se usó 1 ml de tampón de extracción de proteína para extraer la proteína total de una única semilla. Se cargó una muestra de diez µl en gel de poliacrilamida al 12%. Tras la electroforesis, se usaron la tinción con azul de Coomassie y la inmunotransferencia de tipo Western para mostrar los resultados.

SEQ ID No.1 es la secuencia promotora del gen de glutelina de arroz Gt13a;

SEQ ID No.2 es el vector expresado específicamente en el endospermo del arroz.

SEQ ID No. 3 es la secuencia de nucleótidos del extremo C-terminal de codones de preferencia de arroz sintetizados químicamente de Bip.

5 SEQ ID No.4 es la secuencia de nucleótidos del extremo C-terminal de codones de preferencia de arroz sintetizados químicamente de PDI.

SEQ ID No.5 es la secuencia de nucleótidos del extremo C-terminal de codones de preferencia de arroz sintetizados químicamente de IGF-1.

SEQ ID No.6 es el cebador directo de promotor de CP de arroz usado en PCR.

SEQ ID No.7 es el cebador inverso de promotor de CP de arroz usado en PCR.

10 SEQ ID No.8 es el cebador directo del gen de higromicina fosfotransferasa usado en PCR.

SEQ ID No.9 es el cebador inverso de gen de higromicina fosfotransferasa usado en PCR.

### Ejemplos

#### Ejemplo 1. Clonaje del promotor Gt13a y el péptido señal

15 Para clonar el promotor y la secuencia de péptido señal del gen Gt13a (que codifica para glutelina), se usaron los cebadores de SEQ ID NO: 1 para amplificar el ADN genómico de Taipei 309 usando un protocolo de PCR convencional. Se obtuvo un fragmento de ADN con 1284 pb. Se digirió el fragmento por *NaeI* y *XhoI*, y entonces se clonó en pBI221 para obtener pOsPMP2, un vector expresado específicamente en células de endospermo de arroz (véase la figura 1). El resultado del análisis de ADN indicó que el fragmento de ADN aparentemente tiene el promotor y la secuencia de péptido señal (SEQ ID NO:1).

#### 20 Ejemplo 2. Síntesis química del portador de fusión y gen diana con codones genéticos de preferencia de arroz

25 Se obtuvieron las secuencias de nucleótidos del extremo C-terminal del gen Bip de arroz (número de acceso de Genebank: AAB63469), extremo C-terminal del gen PDI de trigo (número de acceso de Genebank: AJ277377) y gen del factor de crecimiento similar a la insulina (número de acceso de Genebank: CAA01955) de la base de datos del NCBI. Se usó el software analítico MacVector para convertir los genes en códigos genéticos de preferencia de arroz. Los cambios de codones genéticos y secuencia de ADN optimizados se muestran en la tabla 1. Puede observarse a partir de la tabla 1 que todas las secuencias de aminoácidos se mantienen sin cambios. Se sintetizaron los genes optimizados por US Blue Heron Biotechnology Incorporation. Durante el proceso de síntesis, se añadieron sitios de restricción *MylI* y *XhoI* a los dos extremos de cada gen. Entonces, se clonaron los genes en pUC119 (Blue Heron Biotechnology) para producir pOsBipC, pOsPDIC y pOsPMP3, que contiene los genes BipC, PDIC e IGF-1 con codones genéticos de preferencia de arroz, respectivamente (SEQ ID NO.3, SEQ ID NO.4 y SEQ ID NO.5).

#### 30 Ejemplo 3. Construcción del vector de expresión específico del endospermo de la proteína de fusión.

35 En primer lugar, se amplificó el IGF-1 humano optimizado mediante PCR, y entonces se clonó en pOsPMP2 digerido con *MscI* y *XhoI*. Se usó el resultante para transformar la cepa DH10B de *E. coli* para producir el plásmido intermedio pOsPMP3. Se digirió el plásmido por *NaeI* y *NcoI*, mientras que los plásmidos de pOsPMP2 (véase SEQ ID NO.2), pOsBipC y pOsPDIC se digirieron con enzimas de restricción apropiadas. Se ligó el gen del portador de fusión con pOsPMP2, y se usó el resultante para transformar la cepa DH10B de *E. coli* para producir plásmidos de expresión de pOsPMP25 y pOsPMP26. Los mapas de restricción de los plásmidos se muestran en las figuras 2 y 3.

#### Ejemplo 4. Clonaje del promotor de cisteína proteinasa $\beta$ de arroz

40 Se usó la PCR para producir el promotor de cisteína proteinasa  $\beta$  de arroz a partir del genoma del arroz. Se diseñaron dos cebadores (SEQ ID NO.6 y SEQ ID NO.7) según la secuencia de nucleótidos en Genebank. Siguiendo un protocolo de PCR convencional, se seleccionó un clon positivo 42M2 (número de clon BAC) de una biblioteca de cromosomas artificiales bacterianos de arroz (BAC) de IR64. Se obtuvo un fragmento de 5 kb tras digerir el clon BAC por *XhoI*. El resultado de la transferencia de tipo Southern confirmó que el fragmento contiene la secuencia de nucleótidos completa del gen de la cisteína proteinasa  $\beta$ . Se usó el clon BAC como molde en la reacción de PCR para producir un fragmento de ADN con 1113 pb, que se clonó adicionalmente en pBI221 para producir el vector intermedio pOsPMP04.

#### Ejemplo 5. Construcción de vectores con el gen marcador seleccionable

50 Se usó ADN del plásmido pCAMBIA 1301 como molde en una PCR convencional con un cebador directo (SEQ ID NO.8) y un cebador inverso (SEQ ID NO.9). En primer lugar, se digirió el fragmento de PCR resultante con *SmaI* y



*XhoI*, y entonces se clonó en pOsPMP4 digerido con *NaeI* y *XhoI* para obtener pOsPMP5, un vector de expresión específico de callo del arroz que tiene un marcador seleccionable (figura 4).

#### Ejemplo 6. Transformación mediada por la pistola génica

5 Se retiraron las cascarillas de la semilla de la variedad de arroz Taipei 309 y se esterilizó la semilla sin cascarilla en hipoclorito de sodio al 20% durante 20 minutos, seguido por lavado tres veces con agua esterilizada, 10 minutos cada vez. Entonces, se cultivó la semilla en un medio de inducción de callo durante 20-30 días para producir callos. Se mezclaron 0,5 µg de pOsPMP25 o pOsPMP26 y 0,5 µg de pOsPMP5 que tienen un marcador seleccionable con 50 µl de partícula de oro, 250 µl de CaCl<sub>2</sub> 1 M y 50 µl de espermidina 0,1 M durante 30 minutos a temperatura ambiente (20-25°C). Tras lavar tres veces con etanol, se recubrieron las partículas de oro con el ADN. Siguiendo el protocolo de pistola génica proporcionado por US DuPont Inc., se co-transformaron los dos plásmidos en el callo de Taipei 309. Tras 45 días de selección en un medio selectivo que contenía 50 µg/ml de higromicina B, se indujeron adicionalmente los callos positivos con resistencia a higromicina B (crecimiento continuo en el medio) en un medio de regeneración en condición de luz durante aproximadamente 20 días. Cuando los callos se diferenciaron para dar plántulas verdes pequeñas, se transfirieron a un medio de enraizado para su cultivo durante otros 15-20 días. Se transfirieron las plantas transgénicas obtenidas a un invernadero para su crecimiento adicional hasta la obtención de semillas maduras.

#### Ejemplo 7. Selección de la línea transgénica con expresión de nivel alto de proteína de fusión

20 Se cultivaron las plantas transgénicas en invernaderos hasta que se inició la floración, y se inició la etapa de semillas. Las semillas se denominaron T1. Tras su recogida, se homogenizaron 10 semillas de cada planta transgénica en 10 ml de disolución de extracción (Tris 50 mM, pH 8,0, NaCl 50 mM, EDTA 10 mM). Se centrifugó la disolución durante 10 minutos en una centrifuga a 14.000 g. Se sometió a prueba el sobrenadante con un kit de ELISA de cuantificación. El resultado de la prueba indica que el nivel de expresión de proteína de fusión Bip-IGF-1 es de aproximadamente 150 µg por semilla, equivalente al 0,75% en peso seco de semilla parda. Los individuos transgénicos con expresión de nivel alto se seleccionaron repetidas veces entre las siguientes generaciones hasta que se obtuvieron las líneas transgénicas que expresan de manera estable las proteínas de fusión. Entonces, se usaron las líneas para la producción a gran escala de la proteína de fusión.

30 Se seleccionaron las líneas transgénicas con el nivel de expresión más alto de la proteína de fusión a partir de las semillas T1. Para someter a prueba el nivel de expresión, se cargaron los extractos brutos de diferentes líneas de generación T1 en SDS-PAGE para electroforesis, que se sometió adicionalmente a tinción con azul de Coomassie o inmunotransferencia de tipo Western. Las figuras 5-7 muestran los resultados de 11 muestras de las líneas transgénicas n.º 26-13 y muestran segregación genética. Los carriles 1-11 son muestras de la línea transgénica 26-13, en la que los carriles 2, 4 y 10 son muestras de individuos negativos segregados genéticamente; el carril 12 es una muestra de la variedad Taipei 309 no transgénica (control negativo). La expresión de la proteína de fusión puede observarse claramente en gel de poliacrilamida (tal como se muestra mediante la flecha en la figura 5). Cuando se usó el anticuerpo IGF-1 en la prueba, puede observarse una proteína de fusión, con el mismo peso molecular que el previsto, en los extractos de endospermo transgénico. Por el contrario, no había banda de proteína correspondiente en los extractos de Taipei 309 no transgénica e individuos negativos aislados genéticamente, lo que concuerda con la ley de Mendel clásica (véase la figura 6). Cuando se usó el anticuerpo de Bip de arroz en la prueba, sólo había proteína Bip endógena en semillas negativas segregadas o muestra de Taipei 309. Por el contrario, había dos bandas de proteína, es decir, Bip endógena y proteína de fusión recombinante Bip-IGF-1, en semillas transgénicas positivas.

45 Las figuras 8 y 9 muestran los resultados de 13 muestras de una línea transgénica n.º 25-12 y muestran la segregación genética. Los carriles 1-13 son muestras de la línea transgénica n.º 25-12, en la que los carriles 3, 5, 8, 10 y 11 son muestras de semillas negativas segregadas genéticamente; el carril 14 es una muestra de la variedad Taipei 309 no transgénica (control negativo). La expresión de proteína de fusión puede observarse claramente en el gel de poliacrilamida (tal como se muestra mediante la flecha en la figura 8). Cuando se usó el anticuerpo de IGF-1 en la prueba, puede observarse una proteína de fusión, con el mismo peso molecular que el previsto, en los extractos del endospermo transgénico. Por el contrario, no había banda de proteína correspondiente en Taipei 309 no transgénica e individuos negativos segregados genéticamente, lo que concuerda con la ley de Mendel clásica (véase la figura 9).

Realizaciones a modo de ejemplo de la invención son:

1. Un método de uso de una proteína de cereal que no es una proteína de almacenamiento como portador de fusión para expresar altamente un péptido pequeño en células de endospermo del huésped, que comprende las etapas de:
  - 55 (i) proporcionar el promotor específico del endospermo y la secuencia de péptido señal de SEQ ID NO:1;
  - (ii) proporcionar el gen de la proteína de cereal que no es una proteína de almacenamiento como portador de fusión y un gen diana que codifica para el péptido pequeño;

- (iii) construir un vector de expresión que contiene la secuencia líder de ADN y el promotor, el gen del portador de fusión y el gen diana; y
- (iv) expresar el vector de expresión en una célula de endospermo del huésped,
- 5 en el que el péptido pequeño tiene menos de 100 aminoácidos, y en el que la proteína de cereal que no es una proteína de almacenamiento se selecciona del grupo que consiste en el extremo C-terminal de la proteína de unión a retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4.
- 10 2. El método de la realización 1, que comprende además la etapa de optimizar al menos uno del promotor y la secuencia de péptido señal, el gen del portador de fusión y el gen diana en codones de preferencia del huésped, antes de la etapa de construir el vector de expresión;
3. El método de la realización 1 ó 2, que comprende además la etapa de incorporar un marcador seleccionable en el vector.
- 15 4. El método de la realización 3, en el que el huésped es arroz, trigo o cebada; y el gen diana es un péptido pequeño terapéutico.
5. El método de la realización 4, en el que el huésped es arroz, y el péptido pequeño terapéutico es IGF-1 humano.
6. El método de la realización 4, en el que el huésped es cebada, y el péptido pequeño terapéutico es IGF-1 humano.
- 20 7. Un vector de expresión para la expresión de alto nivel y específica de un péptido pequeño heterólogo en células de endospermo de cereal, que comprende un gen quimérico que comprende:
- (a) una secuencia líder de ADN que codifica para el promotor específico del endospermo y la secuencia del péptido señal de SEQ ID NO:1;
- 25 (b) un gen seleccionado del grupo que consiste en el extremo C-terminal de la proteína de unión al retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4 como portador de fusión; y
- (c) un gen diana que codifica para el péptido pequeño,
- 30 todos operativamente unidos para permitir la expresión de la proteína de fusión del portador de fusión y el gen diana en una célula de endospermo del huésped,
- en el que el péptido pequeño tiene menos de 100 aminoácidos.
8. El vector de expresión de la realización 7, en el que al menos uno del promotor y la secuencia de péptido señal, el gen del portador de fusión y el gen diana tiene codones optimizados para su expresión en la célula huésped.
- 35 9. El vector de expresión de la realización 8, en el que el gen diana es un péptido pequeño terapéutico.
10. El vector de expresión de la realización 9, en el que el vector comprende además al menos uno seleccionado del grupo que consiste en: un origen de replicación, un marcador seleccionable, un terminador de traducción y un terminador de transcripción.
11. El vector de expresión de la realización 10, en el que el gen diana es IGF-1 humano.
- 40 12. Uso de una proteína de cereal que no es una proteína de almacenamiento como portador de fusión para expresar un péptido pequeño en una célula de endospermo vegetal, en el que el péptido pequeño tiene menos de 100 aminoácidos y en el que la proteína de cereal que no es una proteína de almacenamiento se selecciona del grupo que consiste en el extremo C-terminal de la proteína de unión a retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4.
- 45 13. El uso de la realización 12, en el que el péptido pequeño es un péptido terapéutico.
14. El uso de la realización 13, en el que el péptido pequeño es IGF-1.

LISTADO DE SECUENCIAS

- <110> Wuhan University
- <120> Método de expresión de péptidos pequeños usando proteínas de cereal que no son proteínas de almacenamiento como portador de fusión en el endospermo y uso de los mismos
- 5 <130> Método de expresión de péptidos pequeños usando proteínas de cereal que no son proteínas de almacenamiento como portador de fusión en el endospermo y uso de los mismos
- <160> 9
- <170> PatentIn versión 3.3
- <210>1
- 10 <211> 1223
- <212> ADN
- <213> *Oryza sativa*
- <400> 1

```

gaagaacaac tgacgggcat aaggagaggg agcttttcga taggtgccgt gcagttcaaa      60
gagttagtta gcagtaggat gaagatTTTT gcacatggca atgagaagtt aattatggtg      120
taggcaacc aaatgaaaca ccaaaatatg cacaagacag tttgtgtat tctgtagtac      180
agaataaact aaagtaatga aagaagatgg tgttagaaaa tgaacaata ttatgagtaa      240
tgtgtgagca ttatgggacc acgaaataaa aaaagaacat ttttatgagc agtgtgttct      300
caatgagcct tgaatgttat cacccaggat aagaaaccct taagcaatga aacatgcaag      360
cgtttaatgt gcaaagtgg cattctccac gacataatgc aaaagaagat ataatctatg      420
acatagcaag tcatgcatca tttcatgcct ctgtcaacct attcatttct agtcatctag      480
gtaagtatct taagctaaag tgttagaact tccatacat aagtcataac tgatgacaat      540
tgggtgtaac acatgacaaa ccagagagtc aagcaagata aagcaaaagg atgtgtacat      600
aaaactacag agctatatgt catgttgcga aaagaggaga gcttataaga caagccatga      660
ctcaaaaaaa attcacatgc ctactgtggc ccatatatca tgcaacaatc caaaaactca      720
caggtctcgg tgttgatcgt gtcaacatgt gaccacccta aaaactcttc actaaatatt      780

```

ES 2 372 041 T3

aaagtattgc tagaacagag cttcaagata taagtcatga tcaccaacaa ccatgttcaa 840  
aaagaaatag aaagctatgg cacagcaaca aaaagcaaaa gcatgcatgg atataatctt 900  
taacatcatc catgtcatat tgcaaaagaa agaaagagag aacaatacaa atgatgtgtc 960  
aattacacat ccatcattat ccatccacct tccgtgtacc acacttcata tatcatgagt 1020  
cacttcatgt ctggacatta acaaactcta tcttaacatt caaatgcatg agactttatc 1080  
tcactataaa tgcacaatga tttagcattg tttctcaca aaccattcaa gttcattagt 1140  
actacaacaa catggcatcc ataaatgcc ccatagtttt cttcacagtt tgcttgttcc 1200  
tcttgtgcaa tggctctcta gcc 1223

<210> 2

<211> 6038

<212> ADN

5 <213> *Escherichia coli*

<400> 2

aagcttcaac ctgctgagaa gaacaactga cggtcataag gagagggagc ttttcgatag 60  
gtgccgtgca gttcaaagag ttagttagca gtaggatgaa gatttttgca catggcaatg 120  
agaagttaat tatgggtgtag gcaacccaaa tgaaacacca aaatatgcac aagacagttt 180  
gttgattct gtagtacaga ataaactaaa gtaatgaaag aagatgggtg tagaaaatga 240  
aacaatatta tgagtaatgt gtgagcatta tgggaccacg aaataaaaa agaacatttt 300  
tatgagcagt gtgttctcaa tgagccttga atgttatcac ccaggataag aaacccttaa 360  
gcaatgaaac atgcaagcgt ttaatgtgca aagttggcat tctccacgac ataatgcaaa 420  
agaagatata atctatgaca tagcaagtca tgcatcattt catgcctctg tcaacctatt 480  
catttctagt catctaggtg agtatcttaa gctaaagtgt tagaacttcc catacataag 540  
tcataactga tgacaattgg gtgtaacaca tgacaaacca gagagtcaag caagataaag 600  
caaaaggatg tgtacataaa actacagagc tatatgtcat gttgcgaaaa gaggagagct 660  
tataagacaa gccatgactc aaaaaaatt cacatgccta ctgtggccca tataatcatgc 720  
aacaatccaa aaactcacag gtctcgggtg tgatcgtgtc aacatgtgac caccctaaaa 780  
actcttcaat aatatataaa gtattgctag aacagagctt caagatataa gtcatgatca 840  
ccaacaacca tgttcaaaaa gaaatagaaa gctatggcac agcaacaaaa agcaaaagca 900

tgcattgata taatcttta catcatccat gtcatttgc aaaagaaaga aagagagaac 960  
 aatacaaatg atgtgtcaat tacacatcca tcattatcca tccaccttc gtgtaccaca 1020  
 cttcatatat catgagtcac ttcattgtctg gacattaaca aactctatct taacattcaa 1080  
 atgcatgaga ctttatctca ctataaatgc acaatgattt agcattgttt ctcacaaaac 1140  
 cattcaagtt cattagtact acaacaacat ggcattccata aatcgcccca tagttttctt 1200  
 cacagtttgc ttgttctct tgtgcaatgg ctctctagcc ggccccggcg ggggtgtcag 1260  
 tcccttatgt tacgtcctgt agaaacccca acccgtgaaa tcaaaaaact cgacggcctg 1320  
 tgggcattca gcttgatcg cgaaaactgt ggaattgatc agcgttggg gaaaagcgcg 1380  
 ttacaagaaa gccgggcaat tgctgtgcca ggcagtitta acgatcagtt cgccgatgca 1440  
 gatattcgta attatgcggg caacgtctgg taccagcgcg aagtctttat accgaaaggt 1500  
 tgggcaggcc agcgtatcgt gctgcgttcc gatgcggtca ctcattacgg caaagtgtgg 1560  
 gtcaataatc aggaagtgat ggagcatcag ggcggctata cgccatttga agccgatgct 1620  
 acgccgatg ttattgccgg gaaaagtgta cgtatcacccg tttgtgtgaa caacgaactg 1680  
 aactggcaga ctatccgcc gggaaatggg attaccgacg aaaacggcaa gaaaaagcag 1740  
 tcttacttcc atgatttctt taactatgcc ggaatccatc gcagcgtaat gctctacacc 1800  
 acgccgaaca cctgggtgga cgatatcacc gtggtgacgc atgtcgcgca agactgtaac 1860  
 cacgcgtctg ttgactggca ggtggtggcc aatggtgatg tcagcgttga actgctgat 1920  
 gcggatcaac aggtggttgc aactggacaa ggcactagcg ggactttgca agtgggtgaat 1980  
 ccgcacctct ggcaaccggg tgaaggttat ctctatgaac tgtgcgtcac agccaaaagc 2040  
 cagacagagt gtgatatcta cccgcttcgc gtcggcatcc ggtcagtggc agtgaagggc 2100  
 gaacagttcc tgattaacca caaacgctc tactttactg gctttggtcg tcatgaagat 2160  
 gcggacttgc gtggcaaagg attcgataac gtgctgatgg tgcacgacca cgcattaatg 2220  
 gactggattg gggccaactc ctaccgtacc tcgcattacc cttacgctga agagatgctc 2280  
 gactgggcag atgaacatgg catcgtggtg attgatgaaa ctgctgctgt cggctttaac 2340  
 ctctcttag gcattggttt cgaagcgggc aacaagccga aagaactgta cagcgaagag 2400  
 gcagtcaacg gggaaactca gcaagcgcac ttacaggcga ttaaagagct gatagcgcgt 2460  
 gacaaaaacc acccaagcgt ggtgatgtgg agtattgcca acgaaccgga taccctccg 2520  
 caaggtgcac gggaaatatt cgcgccactg gcggaagcaa cgcgtaaact cgaccggacg 2580  
 cgtccgatca cctgcgtcaa tgtaatgttc tgcgacgctc acaccgatac catcagcgat 2640  
 ctctttgatg tgctgtgcct gaaccgttat tacggatggt atgtccaaag cggcgatttg 2700

gaaacggcag agaaggtact ggaaaaagaa ctctcggcct ggcaggagaa actgcatcag 2760  
 ccgattatca tcaccgaata cggcgtggat acgtagccg ggctgcactc aatgtacacc 2820  
 gacatgtgga gtgaagagta tcagtgtgca tggctggata tgtatcaccg cgtctttgat 2880  
 cgcgtcagcg ccgtcgtcgg tgaacaggta tggaaattcg ccgattttgc gacctcgcaa 2940  
 ggcatattgc gcgttggcgg taacaagaaa gggatcttca ctccgcaccg caaacccaag 3000  
 tcggcggcctt ttctgctgca aaaacgctgg actggcatga acttcgggta aaaaccgag 3060  
 cagggaggca aacaatgaat caacaactct cctggcgcac catcgtcggc tacagcctcg 3120  
 ggaattgctc tcgagctcga atttccccga tcgttcaaac atttggcaat aaagtttctt 3180  
 aagattgaat cctgttgcgg gtcttgcgat gattatcata taatttctgt tgaattacgt 3240  
 taagcatgta ataattaaca tgtaatgcat gacgttattt atgagatggg tttttatgat 3300  
 tagagtcccg caattataca ttaatacgc gatagaaaac aaaatatagc gcgcaacta 3360  
 ggataaatta tcgcgcggg tgcatctat gttactagat cgggaattca ctggccgtcg 3420  
 ttttacaacg tcgtgactgg gaaaaccctg gcgttaccga acttaatcgc cttgcagcac 3480  
 atccccctt cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ctttccaac 3540  
 agttgcgcag cctgaatggc gaatggcgcc tgatgcggta ttttctctt acgcatctgt 3600  
 gcggtatttc acaccgcata tgggtgcactc tcagtacaat ctgctctgat gccgatagt 3660  
 taagccagcc ccgacaccg ccaacaccg ctgacgcgcc ctgacgggct tgtctgtctc 3720  
 cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt 3780  
 caccgtcacc accgaaacgc gcgagacgaa agggcctcgt gatacgccta tttttatagg 3840  
 ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaatgtgc 3900  
 gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg ctcatgagac 3960  
 aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt 4020  
 tccgtgtcgc cttattccc ttttttgcgg cattttgcct tcctgttttt gctcaccag 4080  
 aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg ggttacatcg 4140  
 aactggatct caacagcggc aagatccttg agagttttcg cccgaagaa cgttttccaa 4200  
 tgatgagcac ttttaaagtt ctgctatgtg gcgcggatt atcccgtatt gacgccggc 4260  
 aagagcaact cggtcgccg atacactatt ctcaaatga cttggttgag tactcaccag 4320  
 tcacagaaaa gcatcttac gatggcatga cagtaagaga attatgcagt gctgccataa 4380  
 ccatgagtga taacactgcg gccaaacttac ttctgacaac gatcggagga ccgaaggagc 4440  
 taaccgcttt tttgcacaac atgggggatc atgtaactcg cttgatcgt tgggaaccgg 4500

agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta gcaatggcaa 4560  
 caacgttgcg caaactatta actggcgaac tacttactct agcttcccgg caacaattaa 4620  
 tagactggat ggaggcggat aaagttgcag gaccacttct gcgctcggcc cttccggctg 4680  
 gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcgcggt atcattgcag 4740  
 cactggggcc agatggtaag cccctccgta tcgtagttat ctacacgacg gggagtcagg 4800  
 caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg attaagcatt 4860  
 ggtaactgtc agaccaagtt tactcatata tactttagat tgatttaaaa cttcattttt 4920  
 aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaa atcccttaac 4980  
 gtgagtttcc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag 5040  
 atcctttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg 5100  
 tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact ggcttcagca 5160  
 gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga 5220  
 actctgtagc accgcctaca tacctcgtc tgctaactct gttaccagtg gctgctgcca 5280  
 gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc 5340  
 agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca 5400  
 ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc gaaggagaaa 5460  
 aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc 5520  
 cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc 5580  
 gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg 5640  
 cctttttacg gttcctggcc ttttctggc cttttgctca catgttcttt cctgcgttat 5700  
 cccctgattc tgtggataac cgtattaccg cttttgagtg agctgatacc gctcgcgca 5760  
 gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc ccaatagca 5820  
 aaccgctct ccccgcgctg tggccgattc attaatgcag ctggcacgac aggtttcccg 5880  
 actggaaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttagctcact cattaggcac 5940  
 cccaggcttt acactttatg cttccggctc gtatgttggtg tggaattgtg agcggataac 6000  
 aatttcacac aggaaacagc tatgaccatg attacgcc 6038

ES 2 372 041 T3

<211> 768

<212> ADN

<213> secuencia artificial

<400> 3

```

atcctctccg gcgagggcgg cgacgagacc aaggacatcc tcctcctcga cgtggccccg      60
ctcaccctcg gcatcgagac cgtgggcggc gtgatgacca agctcatccc gcgcaacacc      120
gtgatcccgga ccaagaagtc ccaggtgttc accacctacc aggaccagca gaccaccgtg      180
tccatccagg tgttcgaggg cgagcgtcc atgaccaagg actgccgctt cctcggcaag      240
ttcgacctct ccggcatccc ggccgccccg cgcggcacc cgcagatcga ggtgaccttc      300
gaggtggacg ccaacggcat cctcaacgtg aaggccgagg acaagggcac cggcaagtcc      360
gagaagatca ccatcaccaa cgagaagggc gcctctccc aggaggagat cgaccgatg      420
gtgcgcgagg ccgaggagtt cgccgaggag gacaagaagg tgaaggagcg catcgacgcc      480
cgcaaccagc tggagaccta cgtgtacaac atgaagaaca cctgggcgga caaggacaag      540
ctcgccgaca agctggagtc cgaggagaag gagaaggtgg aggaggcctt caaggaggcc      600
ctggagtggc tcgacgagaa ccagaccgcc gagaaggagg agtacgagga gaagctcaag      660
gaggtggagg ccgtgtgcaa cccgatcatc tccgccgtgt accagcgcac cggcggcgcg      720
ccgggcbggc gccgccggcg ccgcctcgac gacgagcacg acgagctc      768

```

5

<210> 4

<211> 399

<212> ADN

<213> secuencia artificial

10

<400> 4

```

ccggtgaagg tggtggtggc cgacaacatc cagcagtggt tgttcaagtc cggcaagaac      60
gtgctcatcg agttctacgc cccgtgggtg gccactgca agaagctcgc cccgatcctc      120
gacgaggccg ccgccaccct ccagtccgag gaggacgtgg tgatcgccaa gatcgacgcc      180
accgccaacg acgtgccggg cgagttcgac gtgcagggct acccgaccct ctacttcgtg      240
accccgctcc gcaagaaggt gtcctacgag ggcggccgca ccgccgacga gatcgtggac      300
tacatcaaga agaacaagga gaccgccggc caggccggcg ccgccgccac cgagaaggcc      360

```



**gccgagccgg ccgccaccga gccgctcaag gacgagctc**

**399**

<210> 5

<211> 399

<212> ADN

5 <213> secuencia artificial

<400> 5

<b>ccggtgaagg tgggtggtggc cgacaacatc cacgacgtgg tgttcaagtc cggcaagaac</b>	<b>60</b>
<b>gtgctcatcg agttctacgc cccgtggtgc ggccactgca agaagctcgc cccgatcctc</b>	<b>120</b>
<b>gacgaggccg ccgccaccct ccagtccgag gaggacgtgg tgatcgccaa gatcgacgcc</b>	<b>180</b>
<b>accgccaacg acgtgccggg cgagtccgac gtgcagggct acccgaccct ctacttcgtg</b>	<b>240</b>
<b>accccgtccg gcaagaaggt gtcctacgag ggccggccgca ccgccgacga gatcgtggac</b>	<b>300</b>
<b>tacatcaaga agaacaagga gaccgccggc caggccgccg ccgccgccac cgagaaggcc</b>	<b>360</b>
<b>gccgagccgg ccgccaccga gccgctcaag gacgagctc</b>	<b>399</b>

<210> 6

<211> 34

10 <212> ADN

<213> secuencia artificial

<400> 6

**cccaagcttc aacctgctga gaagaacaac tgac**

**34**

<210> 7

15 <211> 31

<212> ADN

<213> secuencia artificial

<400> 7

**cggtgccggc tagagagcca ttgcacaaga g**

**31**

20 <210> 8

<211> 33

<212> ADN

<213> secuencia artificial

<400> 8

**aaagtactat gaaaaagcct gaactcaccg cga**

**33**

<210> 9

<211> 35

5 <212> ADN

<213> secuencia artificial

<400> 9

**ggagaaactc gagcttgctg atcgacagat ccggt**

**35**

**REIVINDICACIONES**

1. Método de uso de una proteína de cereal que no es una proteína de almacenamiento como portador de fusión para expresar altamente un péptido pequeño en células de endospermo del huésped, que comprende las etapas de:
  - 5 (i) proporcionar el promotor específico del endospermo y la secuencia de péptido señal de SEQ ID NO:1;
  - (ii) proporcionar el gen de la proteína de cereal que no es una proteína de almacenamiento como portador de fusión y un gen diana que codifica para el péptido pequeño;
  - (iii) construir un vector de expresión que contiene la secuencia líder de ADN y el promotor, el gen del portador de fusión y el gen diana; y
  - 10 (iv) expresar el vector de expresión en una célula de endospermo del huésped,

en el que el péptido pequeño tiene menos de 100 aminoácidos, y en el que la proteína de cereal que no es una proteína de almacenamiento se selecciona del grupo que consiste en el extremo C-terminal de la proteína de unión a retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4.
- 15 2. Método de la reivindicación 1, que comprende adicionalmente la etapa de optimizar al menos uno del promotor y la secuencia de péptido señal, el gen del portador de fusión y el gen diana en codones de preferencia del huésped, antes de la etapa de construir el vector de expresión.
- 20 3. Método de la reivindicación 1 ó 2, que comprende además la etapa de incorporar un marcador seleccionable en el vector.
4. Método de la reivindicación 3, en el que el huésped es arroz, trigo o cebada; y el gen diana es un péptido pequeño terapéutico.
5. Método de la reivindicación 4, en el que el huésped es arroz, y el péptido pequeño terapéutico es IGF-1 humano.
- 25 6. Método de la reivindicación 4, en el que el huésped es cebada, y el péptido pequeño terapéutico es IGF-1 humano.
7. Vector de expresión para la expresión de alto nivel y específica de un péptido pequeño heterólogo en células de endospermo de cereal, que comprende un gen quimérico que comprende:
  - 30 (a) una secuencia líder de ADN que codifica para el promotor específico del endospermo y la secuencia del péptido señal de SEQ ID NO:1;
  - (b) un gen seleccionado del grupo que consiste en el extremo C-terminal de la proteína de unión al retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4 como portador de fusión; y
  - 35 (c) un gen diana que codifica para el péptido pequeño,

todos operativamente unidos para permitir la expresión de la proteína de fusión del portador de fusión y el gen diana en una célula de endospermo del huésped,

en el que el péptido pequeño tiene menos de 100 aminoácidos.
- 40 8. Vector de expresión de la reivindicación 7, en el que al menos uno del promotor y la secuencia de péptido señal, el gen del portador de fusión y el gen diana tiene codones optimizados para su expresión en la célula huésped.
9. Vector de expresión de la reivindicación 8, en el que el gen diana es un péptido pequeño terapéutico.
10. Vector de expresión de la reivindicación 9, en el que el vector comprende además al menos uno seleccionado del grupo que consiste en: un origen de replicación, un marcador seleccionable, un terminador de traducción y un terminador de transcripción.
- 45 11. Vector de expresión de la reivindicación 10, en que el gen diana es IGF-1 humano.
12. Uso de una proteína de cereal que no es una proteína de almacenamiento como portador de fusión para

- 5 expresar un péptido pequeño en una célula de endospermo vegetal, en el que el péptido pequeño tiene menos de 100 aminoácidos y en el que la proteína de cereal que no es una proteína de almacenamiento se selecciona del grupo que consiste en el extremo C-terminal de la proteína de unión a retículo endoplásmico (Bip) codificada por la secuencia de nucleótidos de SEQ ID NO:3, y el extremo C-terminal de la proteína disulfuro isomerasa (PDI) codificada por la secuencia de nucleótidos de SEQ ID NO:4.
13. Uso de la reivindicación 12, en el que el péptido pequeño es un péptido terapéutico.
14. Uso de la reivindicación 13, en el que el péptido pequeño es IGF-1.

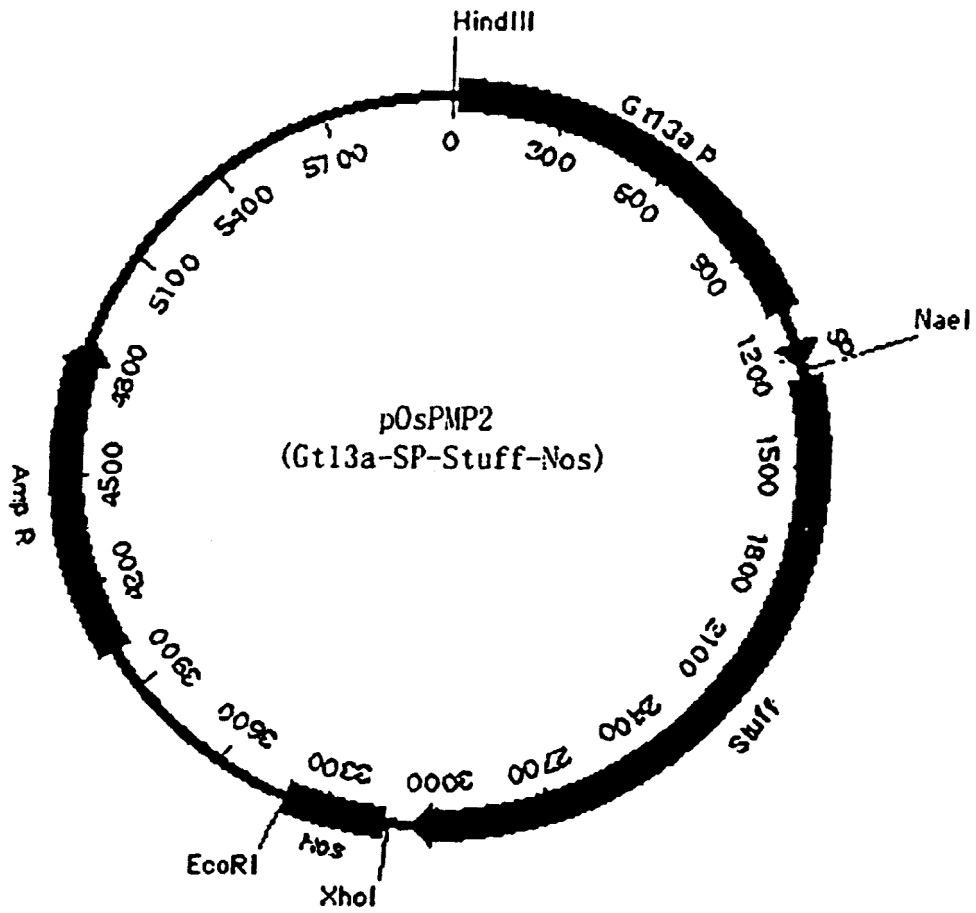


Fig. 1

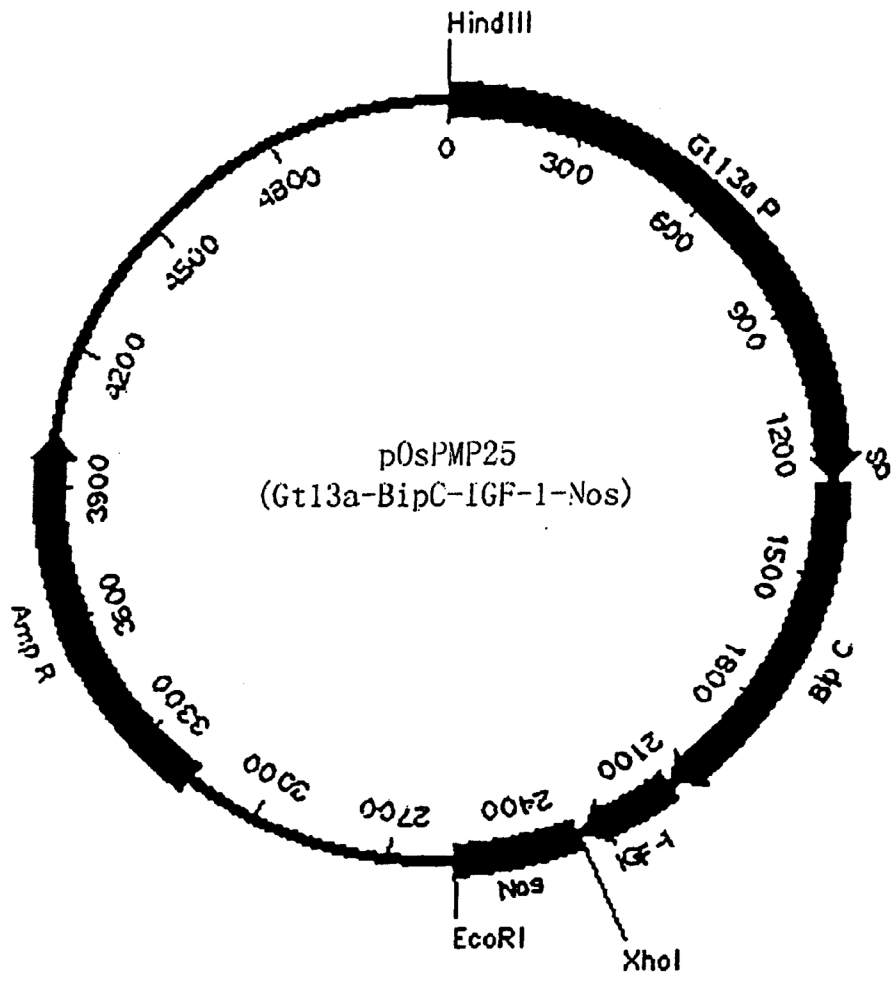


Fig. 2

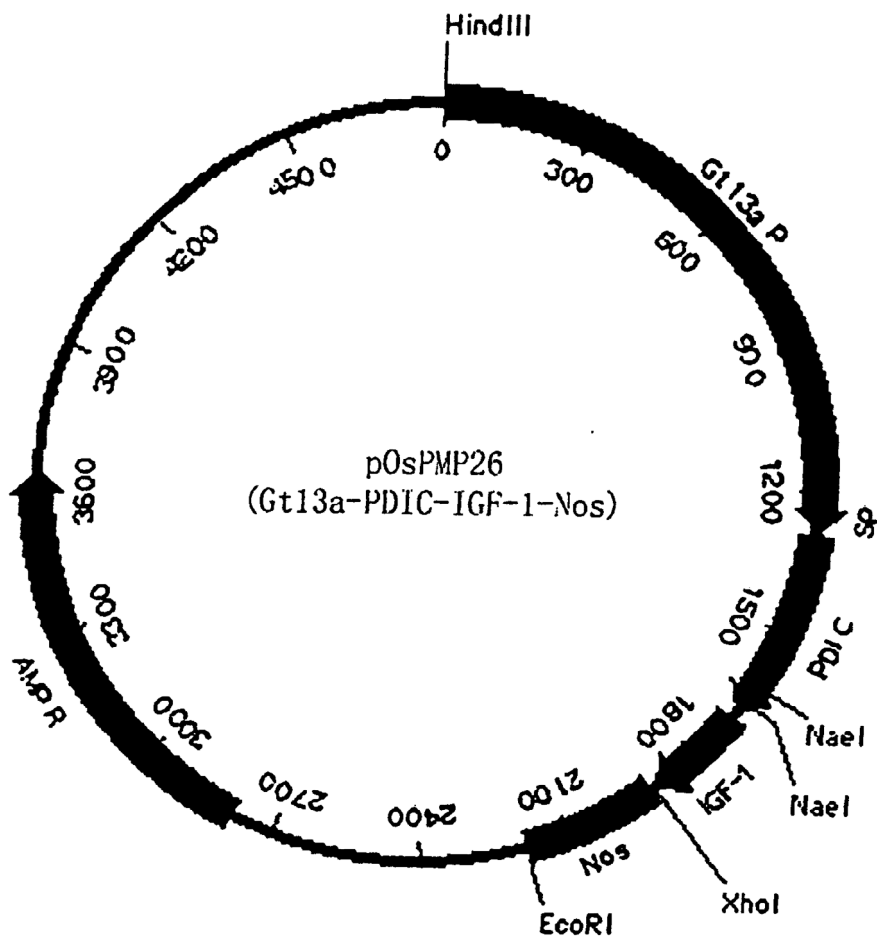


Fig. 3

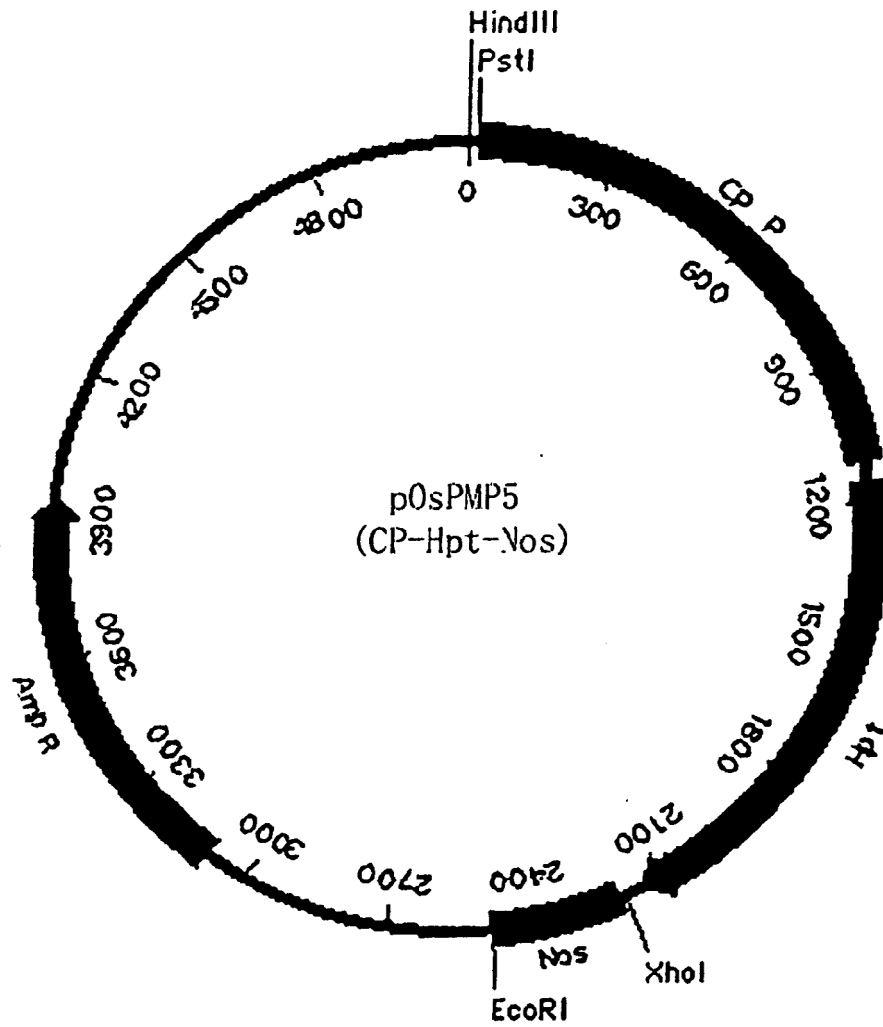


Fig. 4



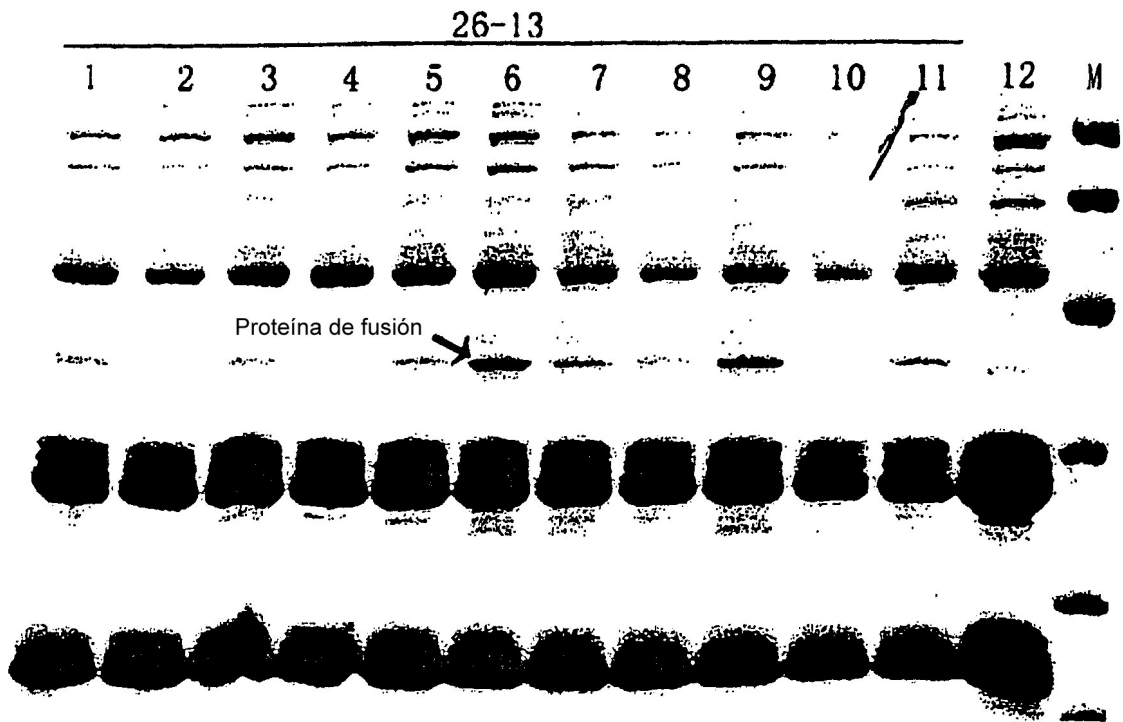


Fig. 5

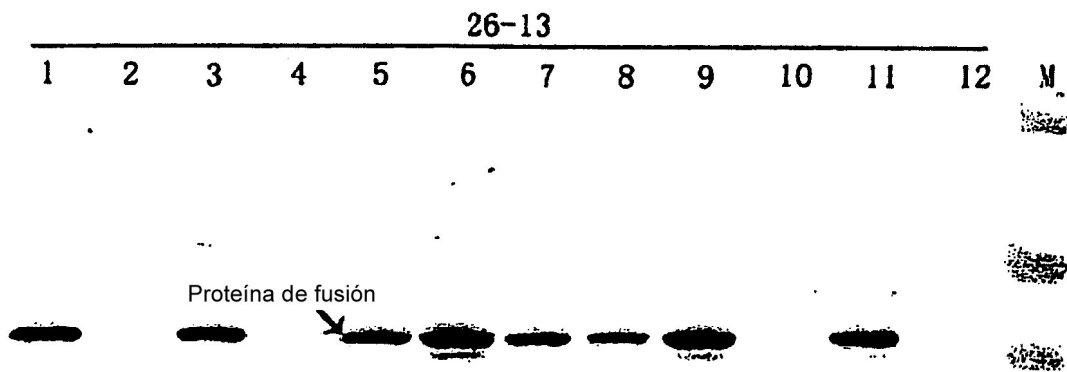


Fig. 6

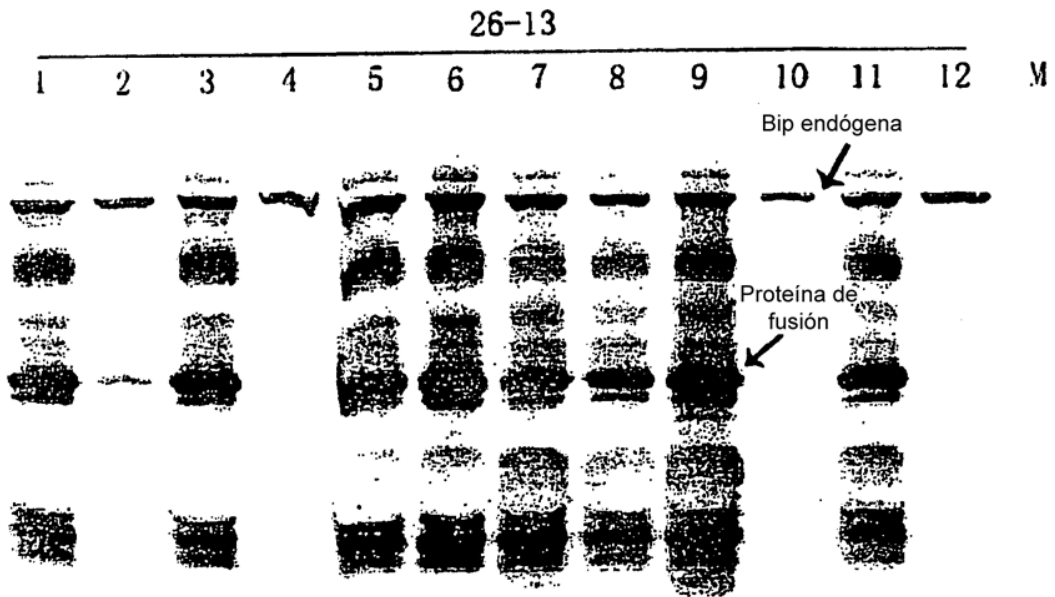
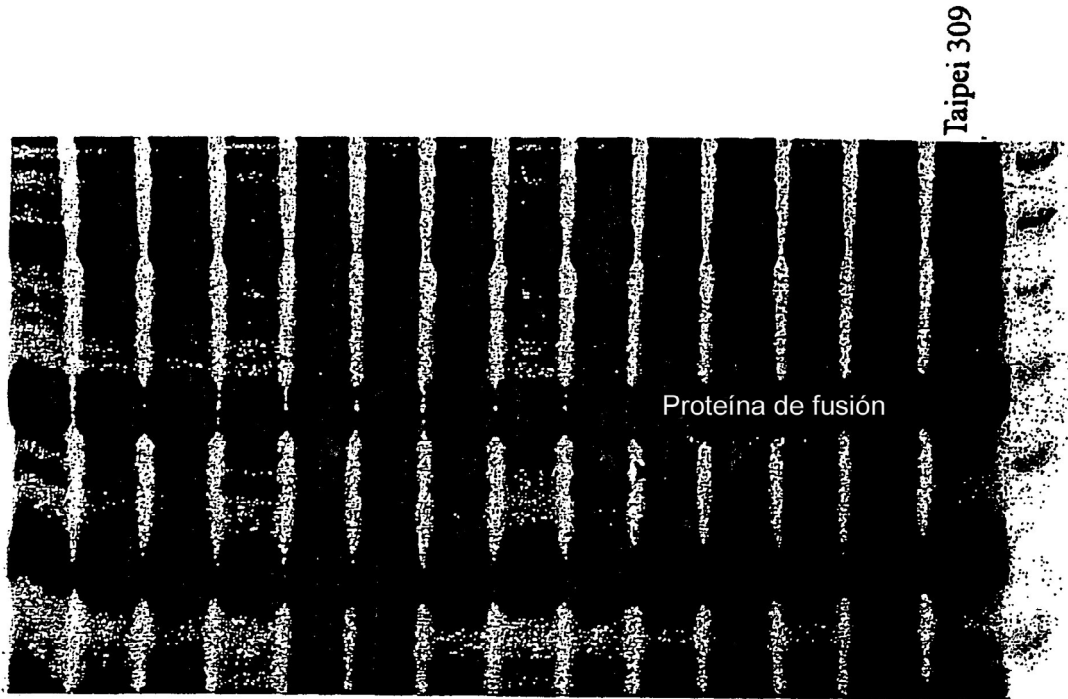
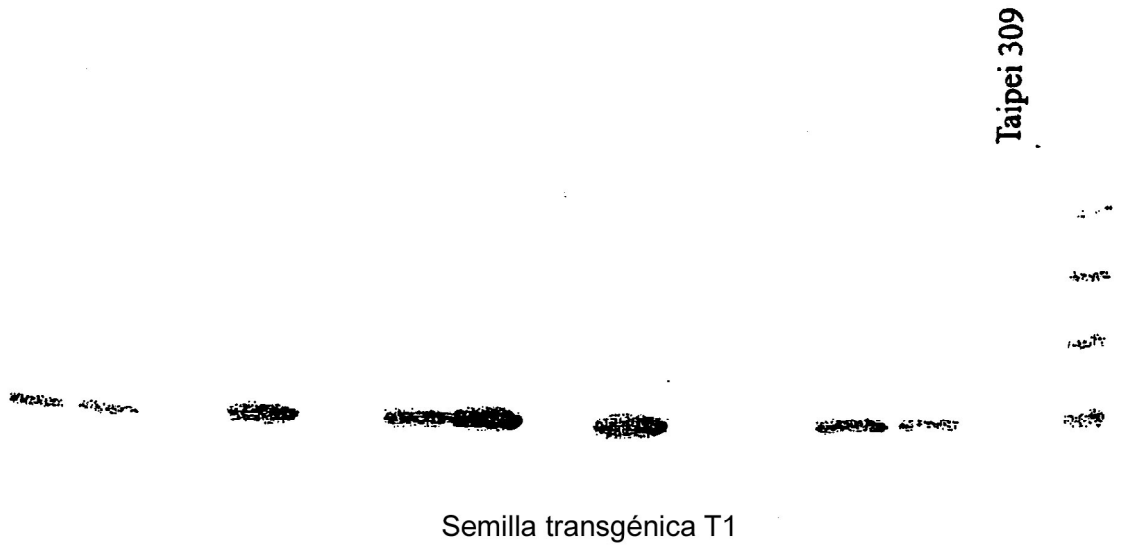


Fig. 7



Semilla transgénica T1

**Fig.8**



**Fig.9**