



(19)



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 289 185**

(51) Int. Cl.:

**C12N 15/62** (2006.01)

**C12N 15/18** (2006.01)

**C12N 15/82** (2006.01)

**C07K 14/585** (2006.01)

**C07K 14/425** (2006.01)

**A01H 5/00** (2006.01)

**A01H 5/10** (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **02807566 .1**

(86) Fecha de presentación : **05.08.2002**

(87) Número de publicación de la solicitud: **1523558**

(87) Fecha de publicación de la solicitud: **20.04.2005**

(54)

**Título: Producción de péptidos y proteínas por acumulación en cuerpos proteicos derivados del retículo endoplásmico en plantas.**

(30)

Prioridad: **28.06.2002 ES 200201508**

(73)

Titular/es: **ERA Biotech, S.A.**  
**Parc. Científic de Barcelona**  
**Josep Samitier, 1-5**  
**08028 Barcelona, ES**

(45)

Fecha de publicación de la mención BOPI:  
**01.02.2008**

(72)

Inventor/es: **Ludevid Múgica, María Dolores;**  
**Torrent Quetglas, Margarita y**  
**Lasserre-Ramassamy, Sabine**

(45)

Fecha de la publicación del folleto de la patente:  
**01.02.2008**

(74)

Agente: **Carpintero López, Francisco**

**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

Producción de péptidos y proteínas por acumulación en cuerpos proteicos derivados del retículo endoplásmico en plantas.

### Campo de la invención

La invención se refiere a la producción de péptidos y proteínas de interés en sistemas vegetales hospedadores por acumulación de los mismos en cuerpos proteicos derivados del retículo endoplásmico, a ácidos nucleicos que codifican tales productos y al uso de dichos ácidos nucleicos en la fabricación de construcciones y vectores para transformar sistemas vegetales hospedadores. Se describe específicamente un procedimiento para expresar y aislar productos heterólogos de interés, tales como calcitonina (CT), en plantas.

### Antecedentes de la invención

Ya que se espera que la demanda de productos biofarmacéuticos aumente considerablemente debido a los notables avances en el conocimiento del genoma y en investigaciones biomédicas relacionadas, existe un considerable interés en elaborar sistemas de producción recombinante de bajo coste.

La ingeniería genética de plantas para producir productos biofarmacéuticos es relativamente reciente, ya que para la bioproducción se han adoptado ampliamente y durante mucho tiempo otros sistemas transgénicos que incluyen bacterias, hongos y células de mamífero cultivadas. Sin embargo, algunas proteínas terapéuticas recombinantes que usan sistemas de expresión vegetales ya están en el mercado o en diversas etapas de ensayos clínicos humanos, como la hirudina, una proteína anticoagulante para tratar la trombosis (Parmenter y col., 1995), una vacuna quimérica de IgG-IgA contra la caries dental (Ma y col., 1998), un vacunógeno bacteriano contra una cepa enterotoxigénica de *E. coli* (Haq y col., 1995), y una lipasa gástrica recombinante de perro para tratar la fibrosis quística (Bénicourt y col., 1993).

Los sistemas de expresión vegetales son atractivos porque el nivel de expresión de las proteínas recombinantes se puede potenciar explotando los mecanismos innatos de aislamiento y fijación como diana que usan las plantas para fijar como diana las proteínas hospedadoras a los orgánulos. Además, los productos biofarmacéuticos derivados de plantas se pueden aumentar a escala fácilmente para la producción en masa y tienen la ventaja de que minimizan los riesgos para la salud que surgen de la contaminación con patógenos o toxinas.

Parece que las plantas son, cada vez más, un sistema de expresión atractivo, debido a su potencial para proporcionar cantidades ilimitadas de material biológicamente activo a bajo coste de producción y con riesgos reducidos para la salud. La capacidad de las plantas para acumular niveles elevados de proteínas recombinantes y para llevar a cabo la mayor parte de las modificaciones post-traduccionales hace que se las considere biorreactores para el cultivo molecular de productos terapéuticos recombinantes (para revisión véase Fischer y Emans, 2000). Sin embargo, las decisiones importantes en lo que se refiere a las estrategias de selección de especies de cultivo, elección de tejidos, expresión y recuperación y procesamiento post-traduccionales son determinantes para la viabilidad de la producción basada en plantas dirigida a la comercialización (Cramer y col., 1999).

La fijación como diana subcelular de proteínas recombinantes es una consideración importante para la acumulación de niveles elevados y el ensamblaje y plegamiento correctos de tales proteínas en las plantas. La compartimentalización de las proteínas hospedadoras en orgánulos intracelulares de almacenamiento se consigue generalmente usando péptidos señal apropiados o fusiones de proteínas enteras. Se ha dirigido una diversidad de proteínas terapéuticas recombinantes a los siguientes compartimientos de las plantas, espacio apoplástico (McCormick y col., 1999), cloroplastos (Staub y col., 2000), retículo endoplásmico (ER) (Stoger y col., 2000). Las inmunoglobulinas dirigidas al compartimiento del ER en plantas transgénicas han mostrado proporcionar rendimientos 10-100 veces superiores que cuando se dirigen a otros compartimientos tales como el apoplasto o el citosol (Conrad y Fiedler, 1998).

La fijación como diana de proteínas complejas tales como anticuerpos en el compartimiento del ER es particularmente interesante, ya que la mayor parte de las modificaciones post-traduccionales necesarias para obtener un producto funcional tienen lugar dentro del ER (Düring y col., 1990; Ma y Hein, 1995; Conrad y Fiedler, 1998). De hecho, dentro del ER, el péptido señal se escinde y las proteínas de estrés tales como la proteína de unión a IgG (BiP) y las enzimas tales como la proteína disulfuro isomerasa (PDI), funcionan como chaperones, se unen a la proteína no ensamblada y dirigen el plegamiento y ensamblaje subsiguientes. Además de estas características particulares, la evidencia disponible indica que el ER de las plantas es muy flexible, lo cual hace que sea un depósito ideal para las proteínas farmacéuticas heterólogas. El ER, aunque parezca que es la puerta de entrada a la ruta secretora, también es capaz de almacenar proteínas durante periodos de tiempo cortos o largos. Las plantas almacenan aminoácidos durante periodos largos en forma de proteínas de almacenamiento específicas. Un mecanismo para proteger estas proteínas de almacenamiento frente a la degradación prematura no controlada es depositarlas en orgánulos de almacenamiento derivados del ER, denominados cuerpos proteicos (PB) (para revisión, Müntz, 1998). El ensamblaje de tales orgánulos como una simple acumulación de proteínas recombinantes en el lumen del ER requiere como primera etapa la retención de la proteína hospedadora. Las proteínas secretoras, una vez plegadas y ensambladas correctamente en el ER, tienen una diversidad de destinos celulares en su mayor parte por medio de la progresión a través del aparato de Golgi. Sin embargo, la retención en el ER de proteínas solubles competentes para el transporte se puede inducir

por medio de la señal de retención/recuperación carboxi-terminal KDEL (o HDEL) (Munro y col., 1987; Wandelt y col., 1992; Vitale y col., 1993). Este motivo C-terminal conservado, reconocido en el aparato de Golgi a través de los receptores de la membrana, permite el reciclaje de las proteínas residentes del ER escapadas de vuelta al ER (Vitale y Denecke, 1999; Yamamoto y col., 2001). Muchos fragmentos de anticuerpos recombinantes se han extendido con la señal KDEL con el fin de acumularse de forma estable en el ER de las plantas (Verch y col., 1998; Torres y col., 1999). Un camino alternativo para generar la retención y acumulación de proteínas recombinantes en el compartimiento del ER es crear una fusión apropiada con un residente del ER natural tal como una proteína de almacenamiento de semillas.

El documento WO 01/75312 describe un procedimiento para producir una citocina en un sistema vegetal hospedador en el que dicho sistema vegetal hospedador se ha transformado con una secuencia de ácido nucleico quimérica que codifica dicha citocina, comprendiendo dicha secuencia de ácido nucleico quimérica una primera secuencia de ácido nucleico capaz de regular la transcripción en dicho sistema vegetal hospedador de una segunda secuencia de ácido nucleico, en la que dicha segunda secuencia de ácido nucleico codifica una secuencia señal que está unida en fase de lectura a una tercera secuencia de ácido nucleico que codifica una citocina y una cuarta secuencia de ácido nucleico unida en fase de lectura al extremo 3' de dicha tercera secuencia de ácido nucleico que codifica una secuencia de aminoácidos "KDEL".

Las zeínas son un grupo de proteínas que se sintetizan durante el desarrollo de endospermo del maíz y se pueden dividir en cuatro grupos  $\alpha$ ,  $\beta$ ,  $\gamma$  y  $\delta$ , en base a su solubilidad. Las zeínas pueden agregarse en los PB directamente en el ER. Se han descrito (documento WO 00/40738) plantas o tejidos de plantas que comprenden cuerpos proteicos estables ante rumina, expresados como proteínas de fusión que comprenden una proteína zeína de longitud completa y un material proteico unido operativamente.

La  $\gamma$ -zeína, una proteína de almacenamiento del maíz, es una de las cuatro prolaminas del maíz y representa el 10%-15% de la proteína total en el endospermo del maíz. Al igual que otras prolaminas de cereales, las  $\alpha$  y  $\gamma$ -zeínas se biosintetizan en polisomas unidos a la membrana en la parte citoplásmica del ER rugoso, se ensamblan dentro del lumen y después se complejan formando PB derivados del ER (Herman y Larkins, 1999; Ludevid y col., 1984; Torrent y col., 1986). La  $\gamma$ -zeína está compuesta por cuatro dominios característicos i) un péptido señal de 19 aminoácidos, ii) el dominio repetido que contiene ocho unidades del hexapéptido PPPVHL (53 aa), iii) el dominio proX en el que restos de prolina se alternan con otros aminoácidos (29 aa) y iv) el dominio C-terminal hidrófobo rico en cisteína (111aa). La capacidad de la  $\gamma$ -zeína de ensamblarse para formar PB derivados del ER no se limita a las semillas. De hecho, cuando el gen de la  $\gamma$ -zeína se expresó de forma constitutiva en plantas transgénicas de Arabidopsis, la proteína de almacenamiento se acumuló dentro de los PB derivados del ER en las células de mesófilo de hoja (Geli y col., 1994). Al buscar una señal responsable para la deposición de la  $\gamma$ -zeína en los PB derivados del ER (las prolaminas no tienen señal KDEL), se ha demostrado que el dominio N-terminal rico en prolina, incluyendo el dominio de repetición en tándem, era necesario para la retención en el ER y que el dominio C-terminal estaba implicado en la formación de los PB. Sin embargo, todavía se desconocen los mecanismos por los que estos dominios fomentan el ensamblaje de PB.

La calcitonina (CT), un péptido hormonal de 32 aminoácidos, es esencial para el metabolismo correcto del calcio y ha encontrado un uso clínico ampliamente extendido en el tratamiento de osteoporosis, choque hipercalcémico y enfermedad de Paget (Reginster y col., 1993; Azria y col., 1995; Silverman y col., 1997). La CT humana se sintetiza en forma de una preproteína con un péptido señal de 25 aminoácidos y dos propéptidos en los extremos N- y C-terminal (57 aa y 21 aa respectivamente). El péptido activo resultante tiene una longitud de 32 aminoácidos con un solo enlace disulfuro (Cys1-Cys7) y está amidado en el extremo carboxi. *In vitro*, la CT humana se agrega, lo cual limita su utilidad como producto terapéutico. Por consiguiente, comúnmente se usa en su lugar la CT del salmón que es menos propensa a agregarse (Cudd y col., 1995). La producción de CT se consigue actualmente por síntesis química, pero el coste de esta producción animó a algunos grupos de investigación a explorar enfoques alternativos. La CT humana y del salmón se han producido en *E. coli* (Ray y col., 1993; Hong y col., 2000), en células de pituitaria de ratón (Merli y col., 1996), en las líneas celulares no endocrinas Cos-7 y CHO (Takahashi y col., 1997) y más recientemente en la leche de conejos transgénicos (McKee y col., 1998). La producción de calcitonina bioactiva por medio de procedimientos biotecnológicos requiere al menos dos etapas de procesamiento: i) generación de una calcitonina con una extensión de glicina (Bradbury y col., 1988) y ii) formación de una prolinamida carboxi terminal a través de la acción de la enzima de amidación, peptidil glicina monooxigenasa de  $\alpha$ -amidación (PAM) (Eipper y col., 1992). Ya que no se sabe actualmente si la carboxil-amidación se produce en células vegetales, la amidación *in vitro* de la calcitonina de plantas con extensión de glicina con la enzima PAM proporcionaría la amida C-terminal (Ray y col., 1993).

## Sumario de la invención

El problema que tiene que resolver la presente invención es proporcionar un sistema alternativo para producir péptidos y proteínas de interés en un sistema vegetal hospedador.

La solución presentada en la presente memoria descriptiva se basa en la capacidad de los dominios ricos en prolina de  $\gamma$ -zeína para auto-ensamblarse y para proporcionar estabilidad a las proteínas de fusión en el ER de un sistema vegetal hospedador. El uso de un sistema basado en proteínas de fusión de  $\gamma$ -zeína para acumular un producto de interés en un sistema vegetal hospedador constituye un enfoque de éxito para acumular dicho producto de interés dentro de los PB derivados del ER de plantas.

La invención se ilustra en el Ejemplo, en el que se describe un sistema basado en proteínas de fusión para acumular CT recombinante en PB derivados del ER en plantas de tabaco. Diversos dominios ricos en prolina se modificaron por ingeniería genética a partir de  $\gamma$ -zeína para servir de parejas de fusión a través de un sitio escindible por proteasa. La región que codifica la calcitonina madura se condensó en la parte C-terminal de los dominios de  $\gamma$ -zeína y se expresó en plantas transgénicas de tabaco. Las proteínas de fusión se acumularon en los PB derivados del ER en hojas de tabaco. Después de la purificación, las proteínas de fusión se sometieron a escisión con enteroquinasa permitiendo la liberación de calcitonina.

Por consiguiente, un aspecto de esta invención se refiere a una secuencia de ácido nucleico que comprende (i) una secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el retículo endoplásmico; (ii) una secuencia de nucleótidos que codifica una secuencia de aminoácidos que es escindible específicamente por medios químicos o enzimáticos; y (iii) una secuencia de nucleótidos que codifica un producto de interés; en la que dichas secuencias nucleótidas se unen de forma operativa entre sí.

En otro aspecto, esta invención se refiere a una construcción de ácido nucleico que comprende dicha secuencia de ácido nucleico.

En otro aspecto adicional, la invención se refiere a un vector que contiene dicha secuencia o construcción y a una célula transformada con dicho vector.

En otro aspecto adicional, la invención se refiere a un sistema vegetal hospedador transformado que tiene dicha secuencia, construcción o vector de ácido nucleico.

En otro aspecto adicional, la invención se refiere a un sistema vegetal hospedador transgénico que comprende, integrado en su genoma, dicha secuencia de ácido nucleico.

En otro aspecto adicional, la invención se refiere a un procedimiento para producir un producto de interés en un sistema vegetal hospedador.

En otro aspecto adicional, la invención se refiere a un procedimiento para producir calcitonina en un sistema vegetal hospedador.

En otro aspecto adicional, la invención se refiere a una proteína de fusión, teniendo dicha proteína de fusión una secuencia de aminoácidos que corresponde a la secuencia de ácido nucleico arriba mencionada.

### Breve descripción de los dibujos

La Figura 1 muestra las secuencias de nucleótidos y traducciones de  $\gamma$ -zeína (Figura 1.A) y de derivados RX3 (Figura 1.B, superior), R3 (Figura 1.B, inferior), P4 (Figura 1.C, superior) y X10 (Figura 1.C, inferior) de  $\gamma$ -zeína.

La Figura 2 muestra la secuencia de nucleótidos (banda 2) y la traducción (banda 1) de calcitonina (CT) sintética. El gen de calcitonina sintética se construyó usando codones de plantas preferentes. Las modificaciones de codones destacan en comparación con el gen de CT del salmón de tipo salvaje (banda 3). El gen sintético contiene en el extremo 5' una secuencia de unión que corresponde al sitio de escisión de la enteroquinasa (EK) y se extiende al 3' para producir una sola glicina C-terminal.

La Figura 3 muestra un perfil esquemático para la construcción del plásmido pCRX3CT. El procedimiento representado fue el mismo para la obtención de los siguientes plásmidos pCZeinCT, pCR3CT, pCP4CT y pCX10CT, siendo la diferencia entre ellos las secuencias de  $\gamma$ -zeína o derivadas de  $\gamma$ -zeína introducidas. Los diferentes plásmidos no se describen en proporción.

La Figura 4 muestra una representación esquemática de los plásmidos pBZeinCT, pBRX3CT, pBR3CT, pBP4CT y pBX10CT. Los diferentes plásmidos no se describen en proporción.

La Figura 5 muestra una representación esquemática de las diferentes proteínas de fusión. Los dominios de  $\gamma$ -zeína y los derivados de  $\gamma$ -zeína (RX3, R3, P4 y X10) se condensaron a calcitonina (CT) a través del sitio de escisión de la enteroquinasa (EK). SP, péptido señal; REPEAT, ocho unidades del dominio de repetición (PPPVHL); R1, una unidad de repetición; Pro-X, prolina-Xaa; PX, fragmento del dominio Pro-X; C-term, dominio C-terminal rico en cisteína; N, secuencia N-terminal de la proteína madura. El número de aminoácidos para cada proteína de fusión se indica a la derecha.

La Figura 6 muestra los resultados de un análisis por inmunotransferencia de las proteínas de fusión en plantas transgénicas de tabaco usando antisuero de  $\gamma$ -zeína. Las proteínas solubles se extrajeron de hojas de tabaco transgénicas (To) y de tipo salvaje (WT), se separaron sobre geles de SDS-poliacrilamida al 15% (20  $\mu$ g por banda) y se transfirieron a nitrocelulosa. Los números representan las líneas transgénicas independientes obtenidas para los diferentes genes quiméricos,  $\gamma$ -zein-CT, RX3-CT, R3-CT, P4-CT.

La Figura 7 muestra: A. Análisis por transferencia de western comparativo de las diferentes proteínas de fusión recombinantes usando antisuero de CT. Los extractos de proteínas solubles se prepararon a partir de plantas de tipo salvaje (WT) y líneas de tabaco transgénicas (T1) que tienen la expresión máxima de la proteína de fusión del gen quimérico relacionado. Se cargaron 8  $\mu$ g de proteínas solubles sobre gel de SDS-poliacrilamida al 15% y se transfirieron a nitrocelulosa. B. Análisis por transferencia de northern comparativo de las diferentes transcripciones de genes quiméricos. Los ARN totales se aislaron de las líneas transgénicas analizadas por inmunotransferencia (Figura 7 A), se fraccionaron sobre electroforesis en gel de formamida desnaturizante (30  $\mu$ g por banda) y se transfirieron por capilaridad sobre membranas de nylon. Las transferencias se hibridaron con una sonda cebada aleatoria (129 bases) obtenida a partir de ADNc de calcitonina.

La Figura 8 muestra la localización subcelular de las proteínas RX3-CT y P4-CT en plantas de tabaco transgénicas: (A) Inmunolocalización de la proteína RX3-CT en líneas RX3-CT transgénicas usando antisuero de CT (dilución 1:100). (B) Inmunolocalización de la proteína P4-CT en líneas P4-CT transgénicas usando antisuero CT (dilución 1:100). (C) Inmunolocalización de la proteína RX3-CT en líneas RX3-CT transgénicas usando antisuero de zeína  $\gamma$  (dilución 1:1,500). (D) Inmunolocalización de la proteína BiP en líneas RX3-CT transgénicas usando antisuero de BiP (dilución 1:250). (E) Inmunolocalización en plantas de tipo salvaje usando antisuero de  $\gamma$ -zeína (dilución 1:1,500). (F) Inmunolocalización en plantas RX3-CT transgénicas sin anticuerpos primarios (dilución 1:1,500). La inmunocitoquímica sobre secciones de hojas de tabaco se realizó usando los anticuerpos primarios indicados y la proteína A-oro coloidal (15 nm). cw: pared celular; ch: cloroplasto; pb: cuerpo proteico; v: vacuola.

La Figura 9 muestra los resultados del análisis de inmunotransferencia de las proteínas de fusión RX3-CT y P4-CT escindidas por EK. Se incubaron 12  $\mu$ g de cada proteína de fusión parcialmente purificada con 0,2 U de EK durante 24 horas a 20°C. Las proteínas de fusión digeridas se fraccionaron por electroforesis en gel de Tris-Tricina poliácridamida al 18% y se transfirieron a nitrocelulosa. Banda 1, proteínas de fusión no digeridas (1  $\mu$ g); banda 2, productos de digestión; banda 3, patrón CT de salmón sintética.

La Figura 10 muestra los resultados del fraccionamiento por RP-HPLC de la proteína de fusión RX3-CT digerida por EK. La pCT liberada de la proteína de fusión RX3-CT se detectó en la fracción 3 (Tr = 13 min) por TOF-MALDI usando CT de salmón sintética como patrón.

La Figura 11 muestra los resultados de caracterización de espectrometría de masas TOF-MALDI de (A) CT de salmón sintética (PM = 3433,24) y (B) CT de planta (PM = 3491,93) eluidas a Tr = 13 min en la fraccionación por RP-HPLC.

### Descripción detallada de la invención

El primer aspecto de la invención proporciona una secuencia de ácido nucleico, denominada en lo sucesivo la secuencia de ácido nucleico de la invención, que comprende:

una primera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el retículo endoplásmico (ER);

una segunda secuencia de ácido nucleico que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos; y

una tercera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica un producto de interés;

donde el extremo 3' de dicha primera secuencia de ácido nucleico está unido al extremo 5' de dicha segunda secuencia de ácido nucleico y el extremo 3' de dicha segunda secuencia de ácido nucleico está unido al extremo 5' de dicha tercera secuencia de ácido nucleico.

La primera secuencia de ácido nucleico contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER.

El término " $\gamma$ -zeína", como se usa en la presente memoria descriptiva, se refiere a una proteína de almacenamiento del maíz que se compone de los cuatro dominios característicos mencionados previamente en los Antecedentes de la Invención. Dicho término incluye proteínas  $\gamma$ -zeína nativas, así como variantes de la misma y proteínas  $\gamma$ -zeína recombinantes que son capaces de dirigir y retener una proteína en el ER.

Prácticamente se puede usar cualquier secuencia de nucleótidos que codifique una proteína  $\gamma$ -zeína, o un fragmento de la misma que contenga una secuencia de nucleótidos que codifique una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER.

Por consiguiente, en una realización preferida, la primera secuencia de ácido nucleico contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína de longitud completa. En una realización particular, la secuencia de

nucleótidos que codifica una proteína  $\gamma$ -zeína de longitud completa se muestra en la Figura 1A y se identifica en la SEC ID NO: 1.

En otra realización preferida, la primera secuencia de ácido nucleico contiene una secuencia de nucleótidos que codifica un fragmento de la proteína  $\gamma$ -zeína, conteniendo dicho fragmento una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER. En este caso, la primera secuencia de ácido nucleico puede contener:

- una o más secuencias de nucleótidos que codifican todo o parte del dominio de repetición de la proteína  $\gamma$ -zeína;
- una o más secuencias de nucleótidos que codifican todo o parte del dominio ProX de la proteína  $\gamma$ -zeína; o
- una o más secuencias de nucleótidos que codifican todo o parte del dominio de repetición de la proteína  $\gamma$ -zeína, y una o más secuencias de nucleótidos que codifican todo o parte del dominio ProX de la proteína  $\gamma$ -zeína.

En una realización particular, dicha primera secuencia de ácido nucleico contiene una secuencia de nucleótidos que codifica un fragmento de la proteína  $\gamma$ -zeína, conteniendo dicho fragmento una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER, seleccionada entre el grupo compuesto por:

- la secuencia de nucleótidos mostrada en la SEC ID NO: 2 [secuencia de nucleótidos identificada como RX3 (Figura 1B)],
- la secuencia de nucleótidos mostrada en la SEC ID NO: 3 [secuencia de nucleótidos identificada como R3 (Figura 1B)],
- la secuencia de nucleótidos mostrada en la SEC ID NO: 4 [secuencia de nucleótidos identificada como P4 (Figura 1C)], y
- la secuencia de nucleótidos mostrada en la SEC ID NO: 5 [secuencia de nucleótidos identificada como X10 (Figura 1C)]

La segunda secuencia de ácido nucleico contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos. En una realización particular, dicha segunda secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica un sitio de escisión por proteasas, por ejemplo, un sitio de aminoácidos que se puede escindir por una proteasa tal como una enteroquinasa, endoproteasa Arg-C, endoproteasa Glu-C, endoproteasa Lys-C, factor Xa y similares.

Como alternativa, la segunda secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica un aminoácido que se puede escindir específicamente por un reactivo químico, tal como, por ejemplo, bromuro de cianógeno, que escinde restos de metionina, o cualquier otro reactivo químico adecuado.

La segunda secuencia de ácido nucleico se puede generar como resultado de la unión entre dicha primera secuencia de ácido nucleico y dicha tercera secuencia de ácido nucleico. En ese caso, cada secuencia contiene un número de nucleótidos de tal forma que cuando dichas primera y tercera secuencias de ácido nucleico se llegan a unir, se forma una secuencia funcional de nucleótidos que codifica una secuencia de aminoácidos que se escinde específicamente por medios enzimáticos o químicos, es decir, la segunda secuencia de ácido nucleico. En una realización alternativa, la segunda secuencia de ácido nucleico es una secuencia foránea insertada operativamente entre dichas primera y tercera secuencias de ácido nucleico.

La tercera secuencia de ácido nucleico contiene la secuencia de nucleótidos que codifica un producto de interés. En principio, cualquier producto de interés se puede expresar por el sistema proporcionado por la presente invención. En una realización preferida, el producto de interés es un fármaco proteico (es decir, una proteína o péptido), por ejemplo, una hormona peptídica, tal como calcitonina, eritropoyetina, trombopoyetina, hormona de crecimiento y similares, un interferón, es decir, una proteína producida en respuesta a infecciones virales y como una citocina durante una respuesta inmune, etc. Preferiblemente, dichos productos terapéuticos de interés son eficaces para el tratamiento de un cuerpo humano o animal.

En una realización particular, la tercera secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica la calcitonina (CT), por ejemplo, calcitonina humana (hCT) o calcitonina de salmón (sCT). En general, en este caso, dicha tercera secuencia de ácido nucleico preferiblemente incluye un codón para glicina en el extremo 3' de dicha secuencia de ácido nucleico que codifica la calcitonina, obteniéndose así una calcitonina con extensión de glicina.

De acuerdo con la invención, el extremo 3' de dicha primera secuencia de ácido nucleico está unido al extremo 5' de dicha segunda secuencia de ácido nucleico y el extremo 3' de dicha segunda secuencia de ácido nucleico está unido al extremo 5' de dicha tercera secuencia de ácido nucleico, es decir, dichas primera, segunda y tercera secuencias de ácidos nucleicos están en fase de lectura.

La secuencia de ácido nucleico de la invención se puede obtener usando técnicas convencionales conocidas por los especialistas en la técnica. En general, dichas técnicas implican la unión de diferentes fragmentos de la secuencia de ácido nucleico de la invención en un vector adecuado. Por ejemplo, en "Molecular cloning, a Laboratory Manual", 2ª ed., por Sambrook y col., Cold Spring Harbor Laboratory Press, 1989, se puede encontrar una revisión de dichas técnicas convencionales. La construcción de algunos vectores que contienen un ácido nucleico de la invención se describe en el Ejemplo y se ilustra en las Figuras 3 y 4. Como se muestra en este documento, se modificaron por ingeniería genética diversos dominios ricos en prolina a partir de  $\gamma$ -zeína para servir como parejas de fusión a través de un sitio de escisión de proteasas. Se condensó la región que codifica la calcitonina madura (32 aa) en el extremo C-terminal de los dominios de  $\gamma$ -zeína y se expresó en plantas de tabaco transgénicas. Las proteínas de fusión se acumularon en cuerpos proteicos derivados del ER en hojas de tabaco. Después de la purificación, las proteínas de fusión se sometieron a escisión con enteroquinasa permitiendo la liberación de calcitonina, que puede purificarse adicionalmente a partir de la mezcla de digestión por medio de una cromatografía de fase inversa.

En otro aspecto, la invención proporciona una proteína de fusión, denominada en lo sucesivo la proteína de fusión de la invención, que comprende (i) la secuencia de aminoácidos de la proteína  $\gamma$ -zeína, o un fragmento de la misma capaz de dirigir y retener una proteína en el ER, (ii) una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos, y (iii) un producto de interés; siendo dicha proteína de fusión el producto de expresión de la secuencia de ácido nucleico de la invención en un sistema vegetal hospedador.

La proteína de fusión de la invención se acumula en PB derivados del ER estables, en un sistema vegetal hospedador. El sitio que se puede escindir enzimática o químicamente, que está presente en el extremo C de los dominios de  $\gamma$ -zeína, permite recuperar posteriormente el producto de interés. Después, el producto de interés se puede aislar y purificar por medios convencionales. Por lo tanto, la proteína de fusión de la invención constituye un enfoque nuevo y satisfactorio para acumular un producto de interés.

En una realización, la proteína de fusión de la invención comprende una proteína de  $\gamma$ -zeína de longitud completa. En la Figura 1A se muestra una secuencia de aminoácidos de  $\gamma$ -zeína de longitud completa y se identifica en la SEC ID NO: 6.

En otra realización, la proteína de fusión de la invención comprende un fragmento de una proteína  $\gamma$ -zeína, conteniendo dicho fragmento una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER. En una realización particular, la proteína de fusión de la invención comprende un fragmento de una proteína  $\gamma$ -zeína, seleccionada entre el grupo compuesto por:

- La secuencia de aminoácidos mostrada en la SEC ID NO: 7 [secuencia de aminoácidos correspondiente a RX3 (Figura 1B)],

- la secuencia de aminoácidos mostrada en la SEC ID NO: 8 [secuencia de aminoácidos correspondiente a R3 (Figura 1B)],

- la secuencia de aminoácidos mostrada en la SEC ID NO: 9 [secuencia de aminoácidos correspondiente a P4 (Figura 1C)], y

- la secuencia de aminoácidos mostrada en la SEC ID NO: 10 [secuencia de aminoácidos correspondiente a X10 (Figura 1C)].

La proteína de fusión de la invención contiene una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos. En una realización particular, dicho sitio escindible comprende un sitio de escisión de proteasas, por ejemplo, un sitio de aminoácidos que se puede escindir por una proteasa tal como la enteroquinasa, la endoproteasa Arg-C, la endoproteasa Glu-C, la endoproteasa Lys-C, el factor Xa y similares, o un sitio de aminoácidos que se puede escindir por un reactivo químico, tal como, por ejemplo, bromuro de cianógeno, que escinde restos de metionina, o cualquier otro reactivo químico adecuado.

La proteína de fusión de la invención también contiene un producto de interés, por ejemplo, un fármaco proteico (es decir, una proteína o péptido), tal como una hormona peptídica, un interferón y similares. Preferiblemente, dicho producto de interés es eficaz para el tratamiento del cuerpo humano o animal. En una realización particular, la proteína de fusión de la invención comprende una calcitonina (CT), por ejemplo, una calcitonina humana (hCT) o calcitonina de salmón (sCT) opcionalmente con una extensión de glicina.

En un aspecto adicional, la invención proporciona una construcción de ácido nucleico que comprende (i) la secuencia de ácido nucleico de la invención y (ii) una secuencia de nucleótidos reguladora que regula la transcripción del ácido nucleico de la invención (i), siendo dicha secuencia reguladora (ii) funcional en plantas. Dichas secuencias de ácido nucleico están unidas operativamente.

Prácticamente se puede usar cualquier secuencia reguladora funcional en plantas. En una realización, preferiblemente dicha secuencia reguladora (ii) es específica de tejidos, es decir, puede regular la transcripción del ácido nucleico de la invención en un tejido específico, tal como semillas, hojas, tubérculos, etc.

La secuencia reguladora (ii) puede comprender un promotor funcional en plantas. Prácticamente se puede usar cualquier promotor funcional en plantas. En una realización particular, dicha secuencia reguladora (ii) comprende el promotor de 35SCaMV. En otra realización particular, dicha secuencia reguladora (ii) comprende el promotor de “patatina”, un promotor de una proteína de almacenamiento, el promotor del gen de ubiquitina, las secuencias reguladoras del gen de la  $\gamma$ -zeína, o similares.

La secuencia reguladora (ii) también puede contener una secuencia de terminación de la transcripción. Prácticamente, se puede usar cualquier secuencia de terminación de la transcripción funcional en plantas. En una realización particular, dicha secuencia de terminación de la transcripción comprende el terminador de 35SCaMV, el terminador del gen de la octopina sintasa (ocs), el terminador del gen de la nopalina sintasa (nos), el terminador del gen de la  $\gamma$ -zeína y similares.

La secuencia reguladora (ii) también puede contener un potenciador de la traducción funcional en plantas. Prácticamente se puede usar cualquier potenciador de la traducción funcional en plantas, por ejemplo, la secuencia promotora de la transcripción del virus de las manchas del tomate, y similares.

La secuencia de ácido nucleico de la invención, o la construcción proporcionada por esta invención, se puede insertar en un vector apropiado. Por lo tanto, en un aspecto adicional, la invención proporciona un vector que comprende la secuencia de ácido nucleico de la invención o una construcción de ácido nucleico proporcionada por la presente invención. Los vectores adecuados incluyen plásmidos, cósmidos y vectores virales. En una realización, dicho vector es adecuado para transformar plantas. La elección del vector puede depender de la célula hospedadora en la que posteriormente se va a introducir. A modo de ejemplo, el vector en el que se introduce la secuencia de ácido nucleico de la invención puede ser un plásmido, un cósmido o un vector viral que, cuando se introduce en una célula hospedadora, se integra en el genoma de dicha célula hospedadora y se replica junto con el cromosoma (o cromosomas) en el que se ha integrado. Para obtener dicho vector pueden usarse procedimientos convencionales (Sambrook y col., 1989).

En un aspecto adicional, la invención proporciona un sistema vegetal hospedador, habiéndose transformado dicho sistema vegetal hospedador con el ácido nucleico de la invención, o con una construcción o un vector proporcionado por la presente invención.

Como se usa en la presente memoria descriptiva, la expresión “sistema vegetal hospedador” incluye plantas, incluyendo, pero sin limitación, monocotiledóneas, dicotiledóneas y, específicamente, cereales (por ejemplo, maíz, arroz, avena, etc.), legumbres (por ejemplo, soja, etc.), crucíferas (por ejemplo, *Arabidopsis thaliana*, colza, etc.) o solanáceas (por ejemplo, patata, tomate, tabaco, etc.). El sistema vegetal hospedador también abarca células vegetales. Las células vegetales incluyen cultivos de suspensión, embriones, regiones meristemáticas, tejidos de callo, hojas, raíces, brotes, gametofitos, esporofitos, polen, semillas y microsporas. El sistema vegetal hospedador puede estar en diversas fases de madurez y puede crecer en cultivos líquidos o sólidos, o en suelo o medio adecuado en tiestos, invernaderos o campos. La expresión en sistemas vegetales hospedadores puede ser transitoria o permanente. El sistema vegetal hospedador también se refiere a cualquier clon de tales plantas, semillas, progenie consanguínea o híbrida, propágulos generados sexual o asexualmente, y descendientes de cualquiera de éstos, tales como esquejes o semillas.

La transformación de los sistemas vegetales hospedadores puede realizarse usando procedimientos convencionales. Se puede ver una revisión de la transferencia genética a plantas en el libro de texto titulado “Ingeniería genética y transferencia génica”, de Marta Izquierdo, Ed. Pirámide (1999), en particular, el Capítulo 9, “Transferencia génica a plantas”, páginas 283-316.

En un aspecto adicional, la invención proporciona un sistema vegetal hospedador transgénico, modificado por ingeniería genética para contener un nuevo transgen diseñado en laboratorio, comprendiendo dicho sistema vegetal hospedador transgénico, integrado en su genoma, el ácido nucleico de la invención. Dicho sistema vegetal hospedador transgénico se puede obtener por medio de técnicas convencionales, por ejemplo, por medio del uso de técnicas convencionales de ARNm sin sentido y/o sobreexpresión (en sentido silencioso) u otras, como por ejemplo, usando vectores binarios u otros vectores disponibles para las diferentes técnicas de transformación de plantas actualmente en uso. Los ejemplos de sistemas vegetales hospedadores transgénicos proporcionados por la presente invención incluyen tanto plantas monocotiledóneas como plantas dicotiledóneas y, específicamente, cereales, legumbres, crucíferas, solanáceas, etc.

La secuencia de ácido nucleico de la invención es útil para producir un producto de interés en un sistema vegetal hospedador. Por lo tanto, en un aspecto adicional, la invención proporciona un procedimiento para producir un producto de interés en un sistema vegetal hospedador, que comprende desarrollar un sistema vegetal hospedador transformado o transgénico proporcionado por la presente invención, en condiciones que permitan la producción y expresión de dicho producto de interés en forma de una proteína de fusión. Como se ha mencionado anteriormente, dicha proteína de fusión se acumula en los PB derivados de ER estables, en dicho sistema vegetal hospedador. El sitio de escisión enzimática o química, que está presente en el extremo C-terminal de los dominios de la  $\gamma$ -zeína, permite recuperar el producto de interés posteriormente. Después el producto de interés se puede aislar y purificar por medios convencionales. Por consiguiente, el procedimiento proporcionado por la presente invención además comprende, si se desea, el aislamiento y purificación de dicha proteína de fusión y, opcionalmente, la liberación de dicho producto de interés de dicha proteína de fusión. La proteína de fusión se escinde en el sitio de escisión por un reactivo enzimático o químico adecuado, según convenga.

En un aspecto adicional, la invención proporciona un procedimiento para producir calcitonina en un sistema vegetal hospedador, que comprende:

5 a) transformar un sistema vegetal hospedador con un vector de expresión o con una construcción de ácido nucleico, que comprende una secuencia reguladora para la transcripción de una secuencia de ácido nucleico (secuencia de ácido nucleico de la invención) que está constituida por:

10 una primera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el retículo endoplásmico (ER);

una segunda secuencia de ácido nucleico que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos; y

15 una tercera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica calcitonina;

20 en la que el extremo 3' de dicha primera secuencia de ácido nucleico está unido al extremo 5' de dicha segunda secuencia de ácido nucleico y el extremo 3' de dicha segunda secuencia de ácido nucleico está unido al extremo 5' de dicha tercera secuencia de ácido nucleico;

b) generar plantas completas a partir de dichos sistemas vegetales hospedadores transformados con dicho vector de expresión o construcción de ácido nucleico;

25 c) desarrollar tales plantas transformadas en condiciones que permitan la producción y expresión de la calcitonina en forma de una proteína de fusión; y, si se desea

d) aislar, purificar dicha proteína de fusión y tratar dicha proteína de fusión con el fin de liberar la calcitonina.

30 Por lo tanto, la invención proporciona un sistema basado en proteínas de fusión para acumular productos recombinantes de interés en PB derivados del ER en sistemas vegetales hospedadores. La invención se ilustra adicionalmente con el siguiente ejemplo no limitante.

#### Ejemplo 1

#### 35 *Producción de calcitonina en plantas de tabaco*

A continuación se describe un ejemplo satisfactorio de la producción de CT en plantas de tabaco. Se modificaron por ingeniería genética diversos dominios ricos en prolina a partir de  $\gamma$ -zeína para servir como parejas de fusión a 40 través de un sitio escindible por proteasa. La región codificante de la CT madura (32 aa) se condensó en el extremo C-terminal de los dominios de la  $\gamma$ -zeína y se expresó en plantas de tabaco transgénicas. Se introdujo un sitio escindible por proteasa en el extremo C-terminal de los dominios de zeína  $\gamma$  para recuperar posteriormente la calcitonina pura. Este enfoque proporciona una alta acumulación de proteínas de fusión dentro del ER y la formación de PB derivados del ER en plantas de tabaco. Las proteínas de fusión se acumularon de forma abundante en PB derivados de ER en 45 hojas de tabaco. El nivel de expresión de dichas proteínas de fusión alcanzó en algunos casos hasta el 12,44% de las proteínas solubles totales. Después de sólo dos etapas de purificación, las proteínas de fusión se sometieron a escisión por enteroquinasa permitiendo la liberación de calcitonina. La calcitonina pura se obtuvo a partir de la mezcla de digestión por cromatografía de fase inversa. El producto de calcitonina acumulado en plantas de tabaco se validó por espectroscopía de masas. También se presentan la purificación de las proteínas de fusión, la digestión con proteasa y 50 la caracterización completa de la calcitonina de las plantas liberada (pCT).

#### I. Procedimiento experimental

#### 55 *Construcción de genes quiméricos y vectores*

60 El gen de la  $\gamma$ -zeína natural y cuatro secuencias derivadas de la  $\gamma$ -zeína denominadas RX3, R3, P4 y X10 que codifican diferentes dominios de la  $\gamma$ -zeína (Figura 1A, 1B y 1C) se condensaron con un gen CT sintético que contenía un sitio de digestión por enteroquinasa (Figura 2) y se introdujeron en vectores de transformación de plantas como se describe a continuación y en la Figura 3.

65 Las secuencias de ADNc de la  $\gamma$ -zeína, RX3 y R3 se generaron por PCR usando pKSG2 (Torrent y col., 1994) como molde. El ADNc de X10 se amplificó a partir de pDRT20, un plásmido producido a partir de pKSG2 después de la delección de la secuencia correspondiente al dominio de repetición. Los cebadores usados para las diferentes PCR fueron:

para la secuencia de ADNc de la  $\gamma$ -zeína:

T1: 5'TCATGAGGGTGTGCTCGTTGCCCTG3', y  
T2: 5'CCATGGCGTGGGGGACACCGCCGGC3'

para las secuencias de ADNc de RX3 y X10:

T1 y  
T2: 5'CCATGGTCTGGCACGGGCTTGGATGCGG3', y

para la secuencia de ADNc de R3:

T1 y  
T2: 5'CCATGGTCCGGGGCGGTTCACTAGGGTA3'

Los productos de PCR se subclonaron en un vector pUC 18 (SureClone Ligation Kit, Pharmacia) y los plásmidos resultantes se denominaron pUCZein, pUCRX3, pUCR3 y pUCX10. El vector pUCP4 que contiene la secuencia P4 derivada de la  $\gamma$ -zeína (Figura 1C) se obtuvo durante la selección de los clones derivados de pUCRX3. Los fragmentos de ADNc de la  $\gamma$ -zeína, RX3, R3, P4 y X10 que contenían extremos cohesivos de BspHI y NcoI, se insertaron en el vector pCKGFPS65C (Reichel y col., 1996) digerido previamente con NcoI. Este vector se seleccionó porque contiene las secuencias reguladoras para la expresión en plantas y la secuencia codificante de GFP que se usaría para estudios paralelos de fijación de diana de proteínas derivadas de  $\gamma$ -zeína en plantas transgénicas. Los vectores generados, pCZeinGFP, pCRX3GFP, pCR3GFP, pCP4GFP y pCX10GFP contenían las siguientes secuencias reguladoras para la expresión en sistemas de plantas: i) el promotor 35S potenciado derivado del virus de mosaico de la coliflor (CaMVp35S), ii) el potenciador traduccional del virus de las manchas del tomate (TL) e iii) la secuencia de terminación de la transcripción de CaMV35S (pA35S). Las construcciones quiméricas derivadas de  $\gamma$ -zeína/CT se generaron por sustitución de la secuencia codificante de GFP con el gen sintético de CT como se describe más adelante (véase la Figura 3).

El gen sintético que codifica los 32 aminoácidos de la CT de salmón activa (Figura 2) se generó a partir de dos oligonucleótidos complementarios de 122 bases. Los oligonucleótidos se diseñaron para usar codones de plantas preferentes con el fin de conseguir una elevada expresión en plantas. Los oligonucleótidos 5' fosforilados sintetizados que usan un sintetizador de ADN Applied Biosystems 394 tenían las siguientes secuencias:

Call:

5'CATGGACGACGACGACAAGTGCTCCAACCTCTCTACCTGCGTTCTTGGTAAGCT  
CTCTCAGGAGCTTCACAAGCTCCAGACTTACCCTAGAACCAACACTGGTTCCGGT  
ACCCCTGGTTGAT 3',

Calll:

5'CTAGATCAACCAGGGGTACCGGAACCAGTGTTGGTTCTAGGGTAAGTCTGGAG  
CTTGTGAAGCTCCTGAGAGAGCTTACCAAGAACGCAGGTAGAGAGGTTGGAGCA  
CTTGTCTGTCGTCGTC3'.

Después de la purificación en gel de poliácridamida al 12%, se usaron 60 pmol de cada oligonucleótido para formar la molécula de doble cadena. La mezcla de hibridación calentada a 95°C durante 5 minutos se mantuvo a 70°C durante 1 hora y se dejó enfriar a temperatura ambiente. El fragmento de ADNc sintético contenía extremos cohesivos NcoI y XbaI en los extremos 3' y 5' terminales, respectivamente. El ADNc de la CT sintética incluía una secuencia de unión 5' que correspondía al sitio de escisión específico por enteroquinasa ((Asp)<sub>4</sub>-Lys) y se extendió en el extremo 3' para producir una sola glicina para la amidación adicional del péptido CT. El ADNc de CT NcoI/XbaI se subclonó en un vector pUC 18 y después se insertó en los sitios de restricción NcoI y BamHI de los vectores pCZeinGFP, pCRX3GFP, pCR3GFP, pCP4GFP y pCX10GFP que contenían las secuencias codificantes de  $\gamma$ -zeína derivadas y seleccionadas de la secuencia codificante de GFP. Las construcciones resultantes se denominaron pCZeinCT, pCRX3CT, pCR3CT, pCP4CT y pCX10CT (Figura 3). Finalmente, se obtuvieron los vectores de transformación de plantas eficaces pBZeinCT, pBRX3CT, pBR3CT, pBP4CT y pBX10CT (Figura 4) insertando los diferentes módulos de expresión HindIII/HindIII en el vector binario pBin19 (Bevan, 1984).

*Transformación estable de plantas de tabaco*

Se transfirieron vectores binarios a cepas LBA 4404 de *Agrobacterium tumefaciens*. Se transformaron discos de hoja de tabaco (*Nicotiana tabacum*, W38) de acuerdo con el método de Draper y col. (1988). Las plantas regeneradas se seleccionaron en medio que contenía 200 mg/l de kanamicina y se transfirieron a un invernadero. Se cultivaron las plantas de tabaco transgénicas que tenían los niveles más elevados de producto transgénico para la obtención de la generación T1. Se recogieron las hojas en desarrollo (de aproximadamente 12 cm de longitud), se congelaron inmediatamente con nitrógeno líquido y se almacenaron a -80°C para experimentos adicionales.

*Extracción y análisis por transferencia de western de proteínas recombinantes*

Se trituraron hojas de tabaco en nitrógeno líquido y se homogeneizaron usando 4 ml de tampón de extracción (Tris-HCl 50 mM, pH 8, ditiotritol (DTT) 200 mM e inhibidores de proteasa (aprotinina 10 µM, pepstatina 1 µM, leupeptina 100 µM, fluoruro de fenilmetilsulfonilo 100 µM y E64 [(N-(N-(L-3-trans-carboxioxirano-2-carbonil)-Leucil)-agmantina] 100 µM por gramo de material de hoja fresca. Los homogeneizados se agitaron durante 30 minutos a 4°C y después se centrifugaron dos veces (15.000 rpm 30 minutos, 4°C) para retirar el material insoluble. Las proteínas solubles totales se cuantificaron usando el ensayo de proteínas de Bradford (Bio-Rad). Las proteínas se separaron en un gel de SDS- poliacrilamida al 15% y se transfirieron a membranas de nitrocelulosa (0,22 µM) usando un aparato semiseco. Las membranas se incubaron con antisuero de  $\gamma$ -zeína (dilución 1/7.000) (Ludevid y col., 1985) o un antisuero activado contra KLH-calcitonina (antisuero de CT) (dilución 1/1.000) y después se incubaron con anticuerpos conjugados con peroxidasa de rábano picante (dilución 1/10.000). Las bandas inmunorreactivas se detectaron por un aumento de quimioluminiscencia (sistema de transferencia de western ECL, Amersham). Se indujeron anticuerpos contra calcitonina en conejos por inoculación de calcitonina de salmón sintética acoplada a KLH. Después de cuatro inoculaciones del antígeno (de 200 µg cada una), los sueros se recogieron, se extrajeron alícuotas y se almacenaron a -80°C. La titulación de los sueros se realizó por inmunotransferencia de puntos usando calcitonina sintética y por ensayos ELISA usando BSA-calcitonina como antígeno.

*Análisis por transferencia de northern*

Se aisló el ARN total a partir de hojas de tabaco (T1) transgénicas y de tipo salvaje de acuerdo con Logemann y col., 1987. El ARN se fraccionó por electroforesis en gel de agarosa-formamida desnaturalizante (30 µg por banda) y se transfirió por capilaridad sobre una membrana de nylon (Hybond N, Amersham Pharmacia Biotech). Las transferencias de ARN se hibridaron con una sonda de ADN de 129 bases obtenida a partir del ADNc de CT y se marcaron con ( $\alpha$ -<sup>32</sup>P) dCTP usando un kit de marcaje de ADN cebado aleatoriamente (Roche). La hibridación se realizó durante una noche a 42°C y los filtros se lavaron tres veces durante 15 minutos en 3X SSC y SDS al 0,5% (P/V) a 65°C. Las transferencias se detectaron con un explorador phosphorImager (Fluor-S<sup>TM</sup> MultiImager, BIO-RAD).

*Ensayos ELISA*

Se realizaron ensayos ELISA para cuantificar la calcitonina de las plantas (pCT) en extractos de proteína de hoja soluble y proteínas de fusión  $\gamma$ -zein-CT parcialmente purificadas. Se cargaron placas de microtitulación (MaxiSorp, Nalgene Nunc International) con proteínas solubles (100 µl) diluidas en solución salina tamponada con fosfato (PBS) pH 7,5, y se incubaron durante una noche a 4°C. Después de lavar los pocillos 3 veces, los sitios de unión no específica se bloquearon con albúmina de suero bovino (BSA) al 3% en PBS-T (PBS que contenía Tween 20 al 0,1%), durante una hora a temperatura ambiente. Las placas se incubaron con antisuero de CT (dilución 1/1.000) durante dos horas y, después de cuatro lavados con PBS-T, se incubaron con anticuerpos secundarios conjugados con peroxidasa (dilución 1/8.000) (Sigma) durante 2 horas. Los anticuerpos primarios y secundarios se diluyeron en PBS-T que contenía BSA al 1%. Después de un lavado extensivo con PBS-T, la reacción enzimática se realizó a 37°C con 100 µl de tampón de sustrato (acetato sódico 100 mM, pH 6, 0,01 mg/ml de TMB (3,3',5,5'-tetrametilbenzidina) y peróxido de hidrógeno al 0,01%). La reacción se detuvo después de 10 minutos con ácido sulfúrico 2N y la densidad óptica se midió a 450 nm usando un espectrofotómetro Multiskan EX (Labsystems). La concentración de antígeno en los extractos de plantas se extrapoló desde una curva patrón obtenida usando calcitonina-BSA y antisuero de CT (dilución 1/1.000).

*Microscopía electrónica*

Se fijaron hojas de plantas naturales y transgénicas por infiltración al vacío con glutaraldehído al 1% y paraformaldehído al 2,5% en tampón fosfato 20 mM, pH 7,4 durante una hora a temperatura ambiente. Después del lavado con tampón fosfato 20 mM y cloruro amónico 200 mM sucesivamente, las muestras se deshidrataron a través de una serie de etanol y se incluyeron en resina Lowicryl K4M. La inmunquímica se realizó esencialmente como se describe por Moore y col., 1991. Las secciones ultrafinas se incubaron con antisueros contra KLH-calcitonina (1/500),  $\alpha$ BiP (1/500) y  $\gamma$ -zeína (1/1.500). Para la detección de los anticuerpos se usó proteína A-oro coloidal (partículas de oro 15 nm). Como control, se realizaron incubaciones paralelas en muestras de plantas no transgénicas usando diluciones idénticas de anticuerpos primarios y en plantas transgénicas sin anticuerpo primario. Las secciones se tiñeron con acetato de uranilo y citrato de plomo y se examinaron con un microscopio electrónico modelo 301 (Phillips, Eindhoven, Países Bajos).

*Purificación y escisión con enteroquinasa de proteínas de fusión RX3-CT y P4-CT*

Se obtuvieron extractos solubles de RX3-CT y P4-CT a partir de hojas de plantas de tabaco (T1) transgénicas en tampón de extracción como se ha descrito anteriormente. Se añadió progresivamente  $(\text{NH}_4)_2\text{SO}_4$  a 0°C a extractos solubles de RX3-CT y P4-CT a una saturación del 45% y 60% respectivamente. Las muestras se agitaron durante 30 minutos a 0°C y después se centrifugaron a 15.000 rpm durante 45 minutos a 4°C. Las proteínas precipitadas se volvieron a suspender en Tris-HCl 20 mM, pH 8,6 y se desalaron en una columna PD 10 (Sephadex G-25 M, Amersham Pharmacia). Los extractos de la proteína desalada se fraccionaron por Cromatografía Líquida de Resolución Rápida (FPLC) usando una columna de intercambio aniónico (HiTrap Q sepharose, Amersham Pharmacia) equilibrada con Tris HCl 20 mM, pH 8,6, DTT 100 mM. La elución de las proteínas se realizó con un gradiente lineal de sales de NaCl de 0 a 200 mM en Tris-HCl 20 mM, pH 8,6, DTT 100 mM. La presencia de RX3-CT y P4-CT en las fracciones eluidas se evaluó por electroforesis en gel de SDS-poliacrilamida al 15% y detección de inmunotransferencia usando antisuero de CT. Las fracciones positivas se desalaron y se concentraron con filtros de centrifuga 5 K NMWL (BIOMAX, Millipore). La cuantificación de las proteínas de fusión RX3-CT y P4-CT se realizó por medio de un ELISA.

Para la digestión por EK, se incubaron 15  $\mu\text{g}$  de proteínas de fusión parcialmente purificadas con 0,2 U EK (EK Max, Invitrogen) en 30  $\mu\text{l}$  de tampón de digestión (Tris-HCl 50 mM, pH8, NaCl 1 mM, Tween-20 al 0,1%) durante 24 horas a 20°C. El tampón de digestión EK se suplementó con DTT 100 mM. La presencia del agente reductor permite optimizar la escisión de la enteroquinasa. Los productos de digestión se analizaron en electroforesis de gel de Tris-Tricina poliacrilamida al 18% y la pCT liberada se detectó por inmunotransferencia. Como control positivo se usó CT sintética de salmón.

*Purificación y análisis de la pCT liberada*

La calcitonina de plantas (pCT) liberada de las proteínas de fusión por digestión con EK se purificó por RP-HPLC. La mezcla de digestión se aplicó a una columna RP-C18 analítica (250 x 4 mm, tamaño de partículas 10  $\mu\text{m}$ , tamaño de poros 120 Å) y la columna se eluyó usando un gradiente que variaba entre el 25% y el 60% de acetonitrilo con TFA al 0,036% en 20 minutos a una velocidad de caudal de 1 ml/min. Las fracciones recogidas se concentraron por liofilización y se almacenaron a -20°C para la caracterización de pCT. En un experimento distinto, se eluyó CT convencional de salmón en las mismas condiciones cromatográficas. Se usó espectrometría de masas TOF-MALDI para la caracterización de la pCT. Se mezclaron alícuotas de la fracción de RP-HPLC con un volumen igual de una solución de matriz (10 mg/ml de ácido  $\alpha$ -ciano-4-hidroxycinámico y TFA al 0,1%) y 1  $\mu\text{l}$  de la mezcla se depositó en el soporte y se analizó con un espectrómetro de masas Voyager-DE-RP (Applied Biosystems). En la espectrometría de masas TOF-MALDI, siempre se usó CT de salmón convencional como control. El análisis C-terminal de la pCT se realizó incubando el péptido purificado (20 pmoles/ $\mu\text{l}$ ) durante 60 minutos a 37°C con carboxipeptidasa Y (0,1 U/ $\mu\text{l}$ ) y el análisis de los productos de digestión por espectrometría de masas TOF-MALDI.

**II. Resultados***Construcción de varios genes quiméricos de derivados  $\gamma$ -zeína-CT*

La expresión y ensamblaje satisfactorio de dominios ricos en prolina de  $\gamma$ -zeína en cuerpos proteicos derivados del ER en hojas de plantas (Geli y col., 1994) proporcionan una herramienta valiosa para acumular proteínas terapéuticas en el ER de tejidos de plantas. El gen de la  $\gamma$ -zeína se delecionó para crear diversas proteínas truncadas ricas en prolina usadas como pareja de fusión para producir CT en plantas de tabaco. Los genes quiméricos comprendían los dominios de  $\gamma$ -zeína y un gen sintético de CT unido por un enlace que correspondía a un sitio escindible por proteasas. El gen sintético que codificaba la calcitonina de salmón activa de 32 aminoácidos se generó a partir de dos oligonucleótidos complementarios (122 bases) diseñados para usar codones de plantas preferentes para conseguir una alta expresión del péptido recombinante en plantas. El ADNc de la CT sintética (Figura 2) incluía en el extremo 5' una secuencia de unión correspondiente al sitio de escisión de enteroquinasa ((Asp)<sub>4</sub>-Lys) y en el extremo 3' un codón adicional para producir una glicina. Esta glicina es un sustrato necesario para que la enzima de amidación (PAM) genere la prolinamida C-terminal esencial para la actividad biológica de la CT. El ADNc de la calcitonina se condensó a las secuencias que codificaban los dominios de  $\gamma$ -zeína en una fusión C-terminal. Para la expresión óptima de los genes quiméricos de derivados de  $\gamma$ -zeína-CT en sistemas de plantas, los vectores de transformación de plantas contenían las siguientes secuencias reguladoras i) el promotor de 35S potenciado constitutivo y el terminador de 35S del virus del mosaico de la coliflor, y ii) el promotor de la traducción del virus de las manchas del tomate (TL). Las diferentes proteínas de fusión generadas se representan en la Figura 5. La proteína de fusión  $\gamma$ -zeína-CT contiene la  $\gamma$ -zeína entera condensada a la CT. Las proteínas de fusión RX3-CT, R3-CT, P4-CT y X10-CT contienen los dominios derivados de la  $\gamma$ -zeína unidos a la CT de la misma forma que la  $\gamma$ -zeína entera. Estas proteínas de fusión difieren esencialmente en la presencia o ausencia de la repetición y dominios proX.

*Producción de proteínas de fusión en plantas de tabaco*

Todos los genes de fusión se usaron para la transformación estable de plantas de tabaco por medio de *Agrobacterium tumefaciens*. Se regeneraron al menos veinte plantas resistentes a kanamicina independientes (To) para cada gen de fusión. La selección de las plantas transgénicas se realizó por medio del análisis por transferencia de western de extractos de proteínas solubles usando un antisuero policlonal de  $\gamma$ -zeína. En la Figura 6 se muestran los patrones de inmunotransferencia de líneas transgénicas representantes de cada gen de fusión. Según lo observado, se obtuvieron

proteínas de fusión recombinantes en todas las líneas transgénicas con la excepción del gen de fusión X10-CT, en el que no se detectaron trazas de proteínas de fusión. Esta pequeña proteína de fusión (80 aminoácidos) probablemente es inestable en plantas de tabaco. Se detectaron dos bandas inmunomarcadas en las líneas transgénicas R3-CT, una con una masa molecular aparente elevada atípica. Probablemente, esta proteína de fusión se sometió a modificaciones post-traduccionales tales como glicosilación. De hecho, se ha demostrado que el dominio de repetición rico en prolina de la  $\gamma$ -zeína es capaz de glicosilarse cuando se expresa en plantas de Arabidopsis (Alvarez y col., 1998). El nivel de expresión de proteínas fue bastante variable entre las diferentes líneas de un mismo gen de fusión, con la excepción del gen de fusión RX3-CT, que mostró un alto nivel de expresión de proteína recombinante en todas las líneas transgénicas. Se realizó una selección de inmunotransferencia adicional usando un antisuero inducido específicamente contra el péptido sCT (Figura 7 A). Como se observó, las proteínas RX3-CT y P4-CT se reconocían fuertemente por el antisuero sCT, indicando que estas fusiones proporcionan una mejor acumulación del péptido CT en plantas de tabaco. Podría indicarse que los patrones de inmunotransferencia de RX3-CT y P4-CT presentaron varias bandas marcadas, correspondiendo la banda principal a la masa molecular aparente correcta de la proteína recombinante relacionada. Una hipótesis podría ser que las bandas marcadas de alto peso molecular eran el resultado de un proceso de oligomerización en dominios de la  $\gamma$ -zeína que se formaban durante la acumulación de las proteínas de fusión en tejidos de plantas. Con el fin de comprobar los niveles de expresión de los genes de fusión en relación con los niveles de proteína, se realizó un análisis de transferencia de northern comparativo (Figura 7 B) usando las líneas transgénicas analizadas por inmunotransferencia en la Figura 7 A. Como se muestra, los transcritos de RX3-CT y P4-CT fueron los más abundantes, lo que demuestra una acumulación estable de estos transcritos. Sorprendentemente, los transcritos de R3-CT fueron relativamente abundantes en comparación con el bajo nivel de proteínas de fusión R3-CT detectado por inmunotransferencia. Probablemente, la modificación post-traducciona evita el autoensamblaje correcto de la proteína de fusión y posteriormente su estabilidad en el ER.

El máximo nivel de expresión de las proteínas RX3-CT y P4-CT, medido por ELISA en extractos de proteína de hoja de plantas T1, fueron respectivamente del 12,44% y del 10,65% de las proteínas totales solubles, mientras que los niveles de expresión de  $\gamma$ -zein-CT y R3-CT permanecieron tan bajos como el 0,01% de las proteínas solubles totales. Con respecto a estos resultados, se eligieron las líneas transgénicas RX3-CT y P4-CT para experimentos adicionales dirigidos a la producción de calcitonina de plantas (pCT).

#### *Localización subcelular de proteínas de fusión RX3-CT y P4-CT*

La expresión de la  $\gamma$ -zeína y dos mutantes de delección de  $\gamma$ -zeína en plantas Arabidopsis demostró que estas proteínas se localizaban dentro del ER de células de mesófilo formando PB derivados del ER (Geli y col., 1994). Sin embargo, no hubo indicios de que la calcitonina condensada a derivados de  $\gamma$ -zeína se clasificara en orgánulos similares, los PB del ER. Para examinar la localización subcelular en hojas de tabaco de las proteínas de fusión de  $\gamma$ -zeína que contenían calcitonina, los inventores usaron microscopía inmunoelectrónica (Figura 8). Se incubaron secciones ultrafinas de hojas de tabaco transgénicas que expresaban proteínas RX3-CT y P4-CT, con anticuerpos contra CT y proteína A-oro. Se observaron unos orgánulos grandes de tipo PB fuertemente marcados en las células de mesófilo de tabaco que expresaban RX3-CT y P4-CT (Figura 8 A y B, respectivamente). Se detectaron pocas vesículas por célula y su tamaño era bastante heterogéneo. Como las proteínas de fusión contenían la proteína calcitonina y fragmentos de  $\gamma$ -zeína, las secciones ultrafinas también se incubaron con anticuerpo contra  $\gamma$ -zeína. Como era de esperar, los PB se marcaron con el anticuerpo contra  $\gamma$ -zeína confirmando que dentro de estos orgánulos se acumulaban las proteínas de fusión (Figura 8 C). Para demostrar que los PB se formaban a partir del ER, las secciones se incubaron con un anticuerpo contra la proteína residente en el ER, BiP (Figura 8 D). La existencia concomitante de las proteínas de fusión de CT y BiP en estos orgánulos indicó que RX3-CT y P4-CT se acumulaban dentro del lumen del ER para formar otras vesículas derivadas del ER independientes. Como los inventores no pudieron detectar orgánulos de tipo PB en secciones ultrafinas de plantas no transgénicas (Figura 8 E), los experimentos de control se realizaron sin anticuerpo primario en plantas transgénicas (Figura 8 F). Como era de esperar, en los experimentos de control no se detectó ningún marcador específico.

#### *Purificación de proteínas de fusión y liberación de pCT*

Se extrajeron eficazmente las proteínas de fusión RX3-CT y P4-CT a partir de hojas de tabaco transgénicas (T1) usando un tampón de extracción que incluía un agente reductor tal como DTT (200 mM). Se recuperaron aproximadamente 85  $\mu$ g de RX3-CT y 73  $\mu$ g de P4-CT por gramo de material fresco. Las proteínas RX3-CT y P4-CT se concentraron respectivamente por precipitación con sulfato amónico al 45% y 60%. Los extractos de proteína desalados se fraccionaron por FPLC usando una cromatografía de intercambio aniónico y las proteínas de fusión recuperadas se cuantificaron por ELISA. La proteína RX3-CT representó aproximadamente el 80% de las proteínas purificadas totales, mientras que la P4-CT fue sólo aproximadamente el 50% de las proteínas purificadas totales. Tal diferencia podía explicarse por el hecho de que con sulfato amónico al 60% precipitan más proteínas que con sulfato amónico al 45% y, por consiguiente, las proteínas P4-CT precipitadas contenían muchas más proteínas contaminantes. Las proteínas de fusión RX3-CT y P4-CT parcialmente purificadas se digirieron por EK y la liberación de pCT se controló por una electroforesis en gel de Tris-Tricina poliacrilamida e inmunodetección. Como se muestra en la Figura 9, por la escisión tanto de RX3-CT como de P4-CT se generó una sola banda marcada correspondiente a la calcitonina. Pequeñas cantidades de las proteínas de fusión RX3-CT y P4-CT permanecieron sin digerir, probablemente debido a la no accesibilidad de la enzima a algunos sitios de escisión.

### Purificación y caracterización de pCT

La calcitonina de las plantas (pCT) se aisló por fraccionamiento de las mezclas de digestión por EK en una columna C18 RP-HPLC (Figura 10) y por análisis de las fracciones eluidas por espectrometría de masas TOF-MALDI usando sCT sintética como patrón (PM 3.433,24, Figura 11. A). La calcitonina pCT se eluyó a 13 minutos (sCT sintética Tr = 14 min) y proporcionó un solo espectro con un masa de 3.491,93 Da por espectrometría de masas TOF-MALDI, que es coherente con la masa molecular teórica de la calcitonina con una extensión C-terminal de glicina reducida (Figura 11 B). El análisis de espectrometría de masas de la pCT sometida a digestión con carboxipeptidasa Y confirmó la integridad de la glicina C-terminal, que es esencial para producir la prolina C-terminal.

### III. Discusión

Se presenta un sistema satisfactorio basado en proteínas de fusión para acumular calcitonina de salmón en plantas de tabaco. Se descubrió que dos proteínas de fusión, RX3-Cal y P4-Cal, se acumulaban fuertemente en PB derivados del ER de hojas de tabaco. Estas proteínas de fusión contienen el péptido CT y los dominios ricos en prolina de la  $\gamma$ -zeína que constan de i) el dominio de repetición compuesto de ocho unidades del hexapéptido PPPVHL (sólo una unidad en la proteína de fusión P4-Cal) y ii) el dominio proX donde los restos de prolina alternan con otros aminoácidos. Los dominios ricos en prolina de la  $\gamma$ -zeína son necesarios para la retención y el ensamblaje correctos de la  $\gamma$ -zeína dentro del ER de plantas del género *Arabidopsis* (Geli y coll., 1994). El plegamiento y la estabilización de las cadenas polipeptídicas de la  $\gamma$ -zeína en el ER se han atribuido a la capacidad de los dominios de repetición y proX de autoensamblarse y de promover la formación de oligómeros. La conformación particular adoptada por estos dominios muy hidrófobos se debería a las secuencias ricas en prolina que pueden formar una estructura secundaria anfipática. Como resultado de su conformación apropiada, los dominios ricos en prolina inducirían mecanismos de agregación que implican interacciones proteína-proteína y enlaces disulfuro cruzados que conducen a la retención en el ER y a la formación de PB derivados del ER. Este ejemplo demuestra que cuando se expresan en una forma de fusión N-terminal, los dominios ricos en prolina de la  $\gamma$ -zeína conservan la capacidad entera de autoensamblarse y de promover los acontecimientos complejos que conducen a la retención y a la acumulación en los PB derivados de ER. También se descubrió que la CT de salmón implicada en la proteína de fusión se acumulaba en gran medida en los PB. El alto nivel de expresión de CT en las plantas de tabaco transgénicas se puede atribuir a la capacidad de los dominios ricos en prolina de plegarse y estabilizar la proteína de fusión. La deposición de la proteína de fusión en los PB naturalmente contribuye al enriquecimiento de los tejidos de plantas en CT por medio de su eliminación del medio intracelular hidrolítico. Como los péptidos pequeños son inestables en sistemas biológicos, actualmente se usa el procedimiento de las proteínas de fusión para producir calcitonina en sistemas heterólogos, por ejemplo, en *E. coli* (Ray y col., 1993; Yabuta y col., 1995; Hong y col., 2000), en *Staphylococcus carnosus* (Dilsen y col., 2000) y en la leche de conejos transgénicos (Mckee y col., 1998). En este último caso, la fusión de CT con lactoalbúmina alfa humana también tiene el fin de enmascarar la actividad de la calcitonina para evitar una posible interferencia con el desarrollo normal del animal.

Los inventores tuvieron éxito en la rápida producción de sCT extendida con glicina a partir de plantas de tabaco:

- i) las proteínas de fusión RX3-Cal y P4-Cal se recuperaron eficazmente a partir de tejidos de tabaco debido a su alta solubilidad en presencia de agentes reductores,
- ii) la liberación de la calcitonina por la acción de una enteroquinasa a partir de las proteínas de fusión se realizó después de una etapa de purificación de la proteína de fusión por una cromatografía de intercambio aniónico, y
- iii) una cromatografía de fase inversa condujo a la CT purificada por su eliminación de la mezcla de la digestión de EK.

El análisis de espectrometría de masas de la CT liberada confirmó que se producía una CT extendida con glicina correcta por las plantas de tabaco.

### Bibliografía

Álvarez, I., Gelim M.I., Pimentel, E., Ludevid, D., Torrent, M. 1998. Lysine-rich  $\gamma$ -zeins are secreted in transgenic *Arabidopsis* plants. *Planta* 205:420-427.

Azria, M., Copp, D.H., Zanelli, J.M. 1995. 25 years of salmon calcitonin: from synthesis to therapeutic use. *Calcif. Tissue Int.* 57 (6): 405-408.

Bénicourt, C., Blanchard, C., Carrière, F., Verger, R. and Junien, J.L. 1993. Potential use of a recombinant dog gastric lipase as an enzymatic supplement to pancreatic extracts in cystic fibrosis, pp. 291-295 in *Clinical Ecology of Cystic Fibrosis*. Escobar H, Baquero CF, Suarez L (eds). Elsevier Sciences Publishers BV, Amsterdam. Bevan, M. 1984. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Research* 12: 8711-8721.

Bradbury, A.F. and Smyth, D. G. 1988. Biosynthesis of peptide neurotransmitters: studies on the formation of peptide amides. *Physiol. Bohemoslov.* 37:267-274.

**Conrad, U. and Fiedler, U.** 1998. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Mol. Biol.* 38: 101-109.

**Cramer, C.L., Boothe, J.G. and Oishi, K.K.** 1999. Transgenic plants for therapeutic proteins linking upstream and downstream strategies. *Curr. Top. Microbiol. Immunol.* 240: 95-118.

**Cudd, A., Arvinte, T., Das, R.E., Chinni, C. and MacIntyre, I.** 1995. Enhanced potency of human calcitonin when fibrillation is avoided. *J. Pharm. Sci.* 84: 717-719.

**Dilsen, S., Paul, W., Sandgathe, D., Tippe, R., Freudl, J., Kula, M.R., Takors, C., Wandrey, D. and Weuster-Botz, D.** 2000. Fed-batch production of recombinant human calcitonin precursor fusion protein using *Staphylococcus carnosus* as an expression-secretion system. *Appl. Microbiol. Biotechnol.* 54: 361-369.

**Draper, J., Scott, R. and Hamil, J.** 1988. In: *Plant Genetic Transformation and Gene Expression. A Laboratory Manual* (Eds. Draper, J., Scott, R., Armitage, P. and Walden, R.), Blackwell Scientific Publications Düring, K., Hippe, S., Kreuzaler, F. and Schell, J. 1990. Synthesis and self-assembly of a functional monoclonal antibody in transgenic *Nicotiana tabacum*. *Plant Mol. Biol.* 15: 281-293.

**Eipper, B.A., Stoffers D.A. and Mains R.E.** 1992. The biosynthesis of neuropeptides: peptide  $\alpha$ -amidation. *Annu. Rev. Neurosci.* 15: 57-85.

**Fischer, R. and Emans, N.** 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research* 9: 279-299.

**Geli, M.I., Torrent, M. and Ludevid, M.D.** 1994. Two structural domains mediate two sequential events in  $\gamma$ -zein targeting: protein endoplasmic reticulum retention and protein body formation. *Plant Cell* 6: 1911-1922.

**Haq, T.A., Mason, H.S., Clements, J.D. and Arntzen, C.J.** 1995. Oral immunization with a recombinant bacterial antigen produced in transgenic plants. *Science* 268: 714-715.

**Herman, E.M. and Larkins, B.A.** 1999. Protein storage bodies. *Plant Cell* 11: 601-613.

**Hong, D., Mingqiang, Z., Min, L., Changqing, C. and Jifang, M.** 2000. production of recombinant salmon calcitonin by amidation of precursor peptide using enzymatic transacylation and photolysis *in vitro*. *Biochem. Biophys. Res. Com.* 267: 362-367.

**Kogan, M.J., Dalcol, I., Gorostiza, P., López-Iglesias, C., Pons, M., Sanz, F., Ludevid, M.D., and Giral, E.** 2001. Self-assembly of the amphipathic helix (VHLPPP)<sub>8</sub>. A mechanism for zein protein body formation. *J. Mol. Biol.* 300: 1-7.

**Logemann, J., Schell, J., Willmitzer, L.** 1987. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* 163: 16-20.

**Ludevid, M.D., Martínez-Izquierdo, J.A., Armengol, M., Torrent, M., Puigdomènech, P. and Palau, P., J.** 1985. Immunological relations between glutelin-2 and low molecular weight zein-2 proteins from maize (*Zea mays*) endosperm. *Plant sci.* 41: 41-48.

**Ludevid, M.D., Torrent, M., Martínez-Izquierdo, J.A., Puigdomènech, P. and Palau, P.** 1984. Subcellular localization of glutelin-2 in maize. *Plant Mol. Biol.* 3: 277-284.

**Ma, J.K., Hikmat, B.Y., Wycoff, K., Vine, N.D., Chargelegue, D., Yum L., Hein, M.B. and Lehner, T.** 1998. Characterization of a recombinant plant monoclonal secretory antibody and preventive immunotherapy in humans. *Nat. Med.* 4: 601-606.

**Ma, J.K. and Hein, M.B.** 1995. Plant antibodies for immunotherapy. *Plant Physiol.* 109: 341-346.

**McCormick, A.A., Kumagai, M.H., Hanley, K., Turpen, T.H., Hakim, L., Grill, L.K., Tusé, D., Levy, S. and Levy, R.** 1999. Rapid production of specific vaccines for lymphoma by expression of the tumor-derived single-chain Fv epitopes in tobacco plants. *Proc. Natl. Aca. Sci. USA* 96 (2) 703-708.

**McKee, C., Gibson, A., Dalrymple, M., Emslie, L., Garner, I. and Cottingham, I.** 1998. Production of biologically active salmon calcitonin in the milk of transgenic rabbits. *Natura Biotechnology* 16: 647-651.

**Merli, S., De Falco, S., Verdoliva, A., Tortora, M., Villain, M., Silvi, P., Cassani, G. and Fassina, G.** 1996. An expression system for the single-step production of recombinant human amidated calcitonin. *Protein Express. Purif.* 7: 347-354.

**Moore, P.J., Swords, K.M.M., Lynch, M.A. and Staehelin, L.A. 1991.** Spatial organization of the assembly pathways of glycoproteins and complex polysaccharides in the Golgi apparatus of plants. *J. Cell Biol.* 112: 589-602.

**Munro, S. and Pelham, H.R.B. 1987.** A C-terminal signal prevents secretion of luminal ER proteins. *Cell* 48: 899-907.

**Müntz, K. 1998.** Deposition of storage proteins. *Plant Mol. Biol.* 38: 77-99.

**Parmenter, D.L., Boothe, J.G., van Rooijen, G.J., Yeung, E.C. and Moloney, M.M. 1995.** Production of biologically active hirudin in plant seeds using oleosin partitioning. *Plant Mol. Biol.* 29: 1167-1180.

**Ray, M.V., van Duyne, P., Bertelsen, A.H., Jackson, Matthews, D.E., Sturmer, A.M., Merkler, D.J., Consalvo, A.P., Young, S.D., Gilligan, J.P. and Shields, P.P. 1993.** Production of recombinant salmon calcitonin by *in vitro* amidation of an *Escherichia coli* produced precursor peptide. *BioTechnology* 11: 64-70.

**Reginster, J.Y. 1993.** Calcitonin for prevention and treatment of osteoporosis. *Am. J. Med.* 95 (5A):44S-47S.

**Reichel, C., Mathur, J., Eckes, P., Langenkemper, K., Koncz, C., Schell, J., Reiss, B., and Maas, C. 1996.** Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono- and dicotyledonous plant cells. *Proc. Natl. Acad. Sci. USA* 93: 5888-5893.

**Sambrook, J., Fritsch, E., Maniatis, T., 1989.** MolecularCloning: A laboratory manual. Cols Spring HarborLaboratory, Cold Spring Harbor, New York. Silverman, S.L. 1997. Calcitonin. *Am. J. Med. Sci.* 313 (1): 13-16.

**Staub, J.M., Garcia, B., Graves, J., Hajdukiewicz, P.T., Hunter, P., Nehra, N., Paradkar, V., Schlittler, M., Carroll, J.A., Spatola, L., Ward, D., Ye, G. and Russell, D.A. 2000.** High-yield production of a human therapeutic protein in tobacco chloroplasts. *Natura Biotechnology* 18: 333-338.

**Stöger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P. and Fischer, R. 2000.** Cereal crops as viable production and storage systems for pharmaceutical scfv antibodies. *Plant Mol. Biol.* 42: 583-590.

**Takahashi, K.I., Liu, Y.C., Hayashi, N., Goto, F., Kato, M., Kawashima, H. and Takeuchi, T. 1997.** Production of bioactive salmon calcitonin from the nonendocrine cell lines COS-7 and CHO. *Peptides* 18: 439-444.

**Torrent, M., Poca, E., Campos, N., Ludevid, M.D. and Palau, J. 1986.** In maize, glutelin-2 and low molecular weight zeins are synthesized by membrane-bound polyribosomes and translocated into microsomal membranes. *Plant Mol. Biol.* 7: 393-403.

**Torrent, M., Geli, M.I., Ruir-Avila, L., Canals, JM., Puigdomènech, P. and Ludevid, D. 1994.** Role of structural domains for maize g-zein retention in *Xenopus oocytes*. *Planta* 192: 512-518.

**Torres, E., Vaquero, C., Nicholson, L., Sack, M., Stöger, E., Drossard, J., Christou, P., Fischer, R. and Perrin, Y. 1999.** Rice cell cultura as an alternativa production system for functional diagnostic and therapeutic antibodies. *Transgenic Research* 8: 441-449.

**Verch, T., Yusibov, V. and Koprowski, H. 1998** Expression and assembly of a full-length monoclonal antibody in plants using a plant virus vector. *J. Immunol. Methods* 220: 69-75.

**Vitale, A., Ceriotti, A. and Denecke, J. 1993.** The role of the endoplasmic reticulum in protein synthesis, modification and intracellular transport. *J. Exp. Bot.* 44: 1417-1444.

**Vitale, A. and Denecke, J. 1999.** The endoplasmic reticulum-gateway of the secretory pathway. *Plant Cell* 11: 615-628.

**Wandelt, C.I., Khan, M.R.I., Craig, S., Schroeder, H.E., Spencer, D. and Higgins, T.J.V. 1992.** Vicilin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants. *Plant J.* 2: 181-192.

**Yabuta, M., Suzuki, Y. and Ohsuye, K. 1995.** High expression of a recombinant human calcitonin precursor peptide in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 42 (5): 703-708.

**Yamamoto, K., Fujii, R., Toyofuku, Y., Saito, T., Koseki, H., Hsu V.W. and Aoe, T. 2001.** The KDEL receptor mediates a retrieval mechanism that contributes to quality control at the endoplasmic reticulum. *EMBO J.* 20 (12): 3082-3091.

## REIVINDICACIONES

1. Una secuencia de ácido nucleico que comprende:

una primera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el retículo endoplásmico (ER) de una célula vegetal;

una segunda secuencia de ácido nucleico que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos; y

una tercera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica un producto de interés;

donde el extremo 3' de dicha primera secuencia de ácido nucleico está unido al extremo 5' de dicha segunda secuencia de ácido nucleico y el extremo 3' de dicha segunda secuencia de ácido nucleico está unido al extremo 5' de dicha tercera secuencia de ácido nucleico.

2. Secuencia de ácido nucleico de acuerdo con la reivindicación 1, en la que dicha primera secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína de longitud completa.

3. Secuencia de ácido nucleico de acuerdo con la reivindicación 1, en la que dicha primera secuencia de ácido nucleico comprende:

- una o más secuencias de nucleótidos que codifican todo o parte del dominio de repetición de la proteína  $\gamma$ -zeína;

- una o más secuencias de nucleótidos que codifican todo o parte del dominio ProX de la proteína  $\gamma$ -zeína; o

- una o más secuencias de nucleótidos que codifican todo o parte del dominio de repetición de la proteína  $\gamma$ -zeína, y una o más secuencias de nucleótidos que codifican todo o parte del dominio ProX de la proteína  $\gamma$ -zeína.

4. Secuencia de ácido nucleico de acuerdo con la reivindicación 1, en la que dicha primera secuencia de ácido nucleico se elige entre el siguiente grupo:

- la secuencia de nucleótidos mostrada en la SEC ID NO: 1 [secuencia de nucleótidos que codifica la  $\gamma$ -zeína (Figura 1A)],

- la secuencia de nucleótidos mostrada en la SEC ID NO: 2 [secuencia de nucleótidos identificada como RX3 (Figura 1B)],

- la secuencia de nucleótidos mostrada en la SEC ID NO: 3 [secuencia de nucleótidos identificada como R3 (Figura 1B)],

- la secuencia de nucleótidos mostrada en la SEC ID NO: 4 [secuencia de nucleótidos identificada como P4 (Figura 1C)], y

- la secuencia de nucleótidos mostrada en la SEC ID NO: 5 [secuencia de nucleótidos identificada como X10 (Figura 1C)].

5. Secuencia de ácido nucleico de acuerdo con las reivindicaciones 1-4, en la que dicha segunda secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica una secuencia de aminoácidos que define un sitio de escisión por proteasas.

6. Secuencia de ácido nucleico de acuerdo con la reivindicación 5, en la que dicha proteasa es una enteroquinasa, una endoproteasa de Arg-C, una endoproteasa de Glu-C, una endoproteasa de Lys-C o el factor Xa.

7. Secuencia de ácido nucleico de acuerdo con las reivindicaciones 1-4, en la que dicha segunda secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica un aminoácido que se puede escindir específicamente por un reactivo químico.

8. Secuencia de ácido nucleico de acuerdo con la reivindicación 7, en la que dicho reactivo químico es bromuro de cianógeno.

9. Secuencia de ácido nucleico de acuerdo con las reivindicaciones 1-8, en la que dicho producto de interés es un fármaco proteico.

10. Secuencia de ácido nucleico de acuerdo con la reivindicación 9, en la que dicho fármaco proteico es una hormona peptídica o un interferón, siendo dicho fármaco eficaz en el tratamiento del cuerpo humano o animal.

11. Secuencia de ácido nucleico de acuerdo con la reivindicación 10, en la que dicha hormona peptídica se selecciona entre calcitonina, eritropoyetina, trombopoyetina y hormona de crecimiento.

12. Secuencia de ácido nucleico de acuerdo con la reivindicación 1, en la que dicha tercera secuencia de ácido nucleico comprende una secuencia de nucleótidos que codifica la calcitonina y un codón para glicina en el extremo 3' de dicha secuencia de ácido nucleico que codifica la calcitonina.

13. Una proteína de fusión que comprende:

(i) la secuencia de aminoácidos de la proteína  $\gamma$ -zeína, o un fragmento de la misma capaz de dirigir y retener una proteína en el ER de una célula vegetal,

(ii) una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos, y

(iii) un producto de interés;

siendo dicha proteína de fusión el producto de expresión de la secuencia de ácido nucleico de cualquiera de las reivindicaciones 1-12 en un sistema vegetal hospedador.

14. Proteína de fusión de acuerdo con la reivindicación 13, que comprende la proteína  $\gamma$ -zeína de longitud completa.

15. Proteína de fusión de acuerdo con la reivindicación 14, que comprende la secuencia de aminoácidos mostrada en la Figura 1A e identificada en la SEC ID NO: 6.

16. Proteína de fusión de acuerdo con la reivindicación 13, que comprende un fragmento de la proteína  $\gamma$ -zeína, conteniendo dicho fragmento una secuencia de aminoácidos capaz de dirigir y retener una proteína en el ER de una célula vegetal.

17. Proteína de fusión de acuerdo con la reivindicación 16, que comprende un fragmento de la proteína  $\gamma$ -zeína seleccionado entre el grupo que está constituido por:

- la secuencia de aminoácidos mostrada en la SEC ID NO: 7 [secuencia de aminoácidos correspondiente a RX3 (Figura 1B)],

- la secuencia de aminoácidos mostrada en la SEC ID NO: 8 [secuencia de aminoácidos correspondiente a R3 (Figura 1B)],

- la secuencia de aminoácidos mostrada en la SEC ID NO: 9 [secuencia de aminoácidos correspondiente a P4 (Figura 1C)], y

- la secuencia de aminoácidos mostrada en la SEC ID NO: 10 [secuencia de aminoácidos correspondiente a X10 (Figura 1C)].

18. Proteína de fusión de acuerdo con la reivindicación 13, en la que dicha secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos comprende un sitio de escisión por proteasas.

19. Proteína de fusión de acuerdo con la reivindicación 13, en la que dicha secuencia de aminoácidos que se puede escindir específicamente por medios químicos comprende un sitio de escisión que se puede escindir por un reactivo químico.

20. Proteína de fusión de acuerdo con la reivindicación 13, en la que dicho producto de interés es un fármaco proteico.

21. Una construcción de ácido nucleico que comprende (i) una secuencia de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 12, y (ii) una secuencia de nucleótidos reguladora que regula la transcripción del ácido nucleico de la invención (i), siendo dicha secuencia reguladora (ii) funcional en plantas.

22. Construcción de acuerdo con la reivindicación 21, en la que dicha secuencia reguladora (ii) es específica de tejidos.

23. Construcción de acuerdo con la reivindicación 21, en la que dicha secuencia reguladora (ii) comprende un promotor funcional en plantas.

24. Construcción de acuerdo con la reivindicación 22, en la que dicha secuencia reguladora (ii) comprende el promotor de 35SCaMV, el promotor de "patatina", un promotor de proteínas de almacenamiento, el promotor del gen de la ubiquitina o las secuencias reguladoras del gen de la  $\gamma$ -zeína.

5 25. Construcción de acuerdo con la reivindicación 21, en la que dicha secuencia reguladora (ii) comprende una secuencia de terminación de la transcripción funcional en plantas.

26. Construcción de acuerdo con la reivindicación 25, en la que dicha secuencia reguladora (ii) comprende el terminador de 35SCaMV, el terminador del gen de la octopina sintasa (ocs), el terminador del gen de la nopalina sintasa (nos) o el terminador del gen de la  $\gamma$ -zeína.

27. Construcción de acuerdo con la reivindicación 21, en la que dicha secuencia reguladora (ii) comprende además un potenciador de la traducción funcional en plantas.

15 28. Construcción de acuerdo con la reivindicación 27, en la que dicho potenciador de la traducción funcional en plantas comprende la secuencia promotora para la transcripción del virus de las manchas del tomate, y similares.

29. Un vector que comprende una secuencia de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 12, o una construcción de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 21 a 28.

20 30. Un sistema vegetal hospedador transformado, habiéndose transformado dicho sistema vegetal hospedador con una secuencia de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 12, o con una construcción de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 21 a 28, o con un vector de acuerdo con la reivindicación 29.

25 31. Un sistema vegetal hospedador transgénico, comprendiendo dicho sistema vegetal hospedador transgénico, integrado en su genoma, una secuencia de ácido nucleico de acuerdo con cualquiera de las reivindicaciones 1 a 13.

30 32. Sistema vegetal hospedador de acuerdo con la reivindicación 30 ó 31, en el que dicho sistema vegetal hospedador es una planta monocotiledónea o dicotiledónea.

33. Sistema vegetal hospedador, de acuerdo con la reivindicación 32, en el que dicho sistema vegetal hospedador es un cereal, una legumbre, una crucífera o una solanácea.

35 34. Sistema vegetal hospedador de acuerdo con la reivindicación 30 ó 31, que comprende una semilla.

35. Un procedimiento para producir un producto de interés en un sistema vegetal hospedador, que comprende desarrollar un sistema vegetal hospedador transformado o transgénico de acuerdo con las reivindicaciones 30-34, en condiciones que permiten la producción y expresión de dicho producto de interés en forma de una proteína de fusión.

40 36. Procedimiento de acuerdo con la reivindicación 35, que comprende además el aislamiento y purificación de dicha proteína de fusión.

45 37. Procedimiento de acuerdo con la reivindicación 35 ó 36, que comprende además la liberación de dicho producto de interés a partir de dicha proteína de fusión.

38. Un procedimiento para producir calcitonina en un sistema vegetal hospedador, que comprende:

50 a) transformar un sistema vegetal hospedador con un vector de expresión o con una construcción de ácido nucleico, que comprende una secuencia reguladora para la transcripción de una secuencia de ácido nucleico (secuencia de ácido nucleico de la invención) que consta de:

55 una primera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica la proteína  $\gamma$ -zeína, o un fragmento de la misma que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos capaz de dirigir y retener una proteína en el retículo endoplásmico (ER) de una célula de una planta;

una segunda secuencia de ácido nucleico que contiene una secuencia de nucleótidos que codifica una secuencia de aminoácidos que se puede escindir específicamente por medios enzimáticos o químicos; y

60 una tercera secuencia de ácido nucleico que contiene la secuencia de nucleótidos que codifica calcitonina;

donde el extremo 3' de dicha primera secuencia de ácido nucleico está unido al extremo 5' de dicha segunda secuencia de ácido nucleico y el extremo 3' de dicha segunda secuencia de ácido nucleico está unido al extremo 5' de dicha tercera secuencia de ácido nucleico;

65 b) generar plantas completas a partir de dichos sistemas vegetales hospedadores transformados con dicho vector de expresión o construcción de ácido nucleico;

## ES 2 289 185 T3

c) desarrollar tales plantas transformadas en condiciones que permitan la producción y expresión de la calcitonina en forma de una proteína de fusión; y, si se desea

d) aislar, purificar dicha proteína de fusión y tratar dicha proteína de fusión para liberar la calcitonina.

5

10

15

20

25

30

35

40

45

50

55

60

65

### Secuencia de la $\gamma$ -zeína

atgagggtgtgtgctcgtttgctcgttgcctcgcctcctcgtcctcgtcgcgcgcacccctccacg  
M R V L L V A L A L L A L A A S A T S T  
catacaugcggcggtgcggtgcccagccacgcgcgcgggttcattctaccgcgcgcgggtg  
H T S G G C G C Q P P P P V H L P P P V  
catctgccacctccggttcacctgccacctccgggtgcattctcccaccgcgcgggtccacctg  
H L P P P V H L P P P V H L P P P V H L  
ccgcgcgcgggtccacctgccaccgcgcgggtccattgtgcgcgcgcgggttcattctgcgcgcg  
P P P V H L P P P V H V P P P V H L P P  
ccacctgccactaccttactcaaccgcgcgcgcctcagcctcatccccagccacaccca  
P P C H Y P T O P P R P O P H P O P H P  
tgcccggtgccaacagccgcattccaagcccggtgccagctgcagggaacctgcggcgttggc  
C...C...C...Q...Q...P...H...P...S...P...C Q L Q G T C G V G  
agcaccgccgatcctgggcccagtgcgctcgagtttctgaggcatcagtgccagccgcagggc  
S T P I L G Q C V E F L R H Q C S P T A  
acgccttactgctcgcctcagtgccagtcggttgccggcagcagtggttgccagcagctcagg  
T P Y C S P Q C Q S L R Q Q C C Q Q L R  
cagggtggagccgcagcaccgggtaccaggcgatcttcggcttggtcctccagtcattcctg  
Q V E P Q H R Y Q A I F G L V L Q S I L  
cagcagcagccgcacaaagcggccaggtcgcggggctgttgccgggcgcagatagcgcagcaa  
Q Q Q P Q S G Q V A G L L A A Q I A Q Q  
ctgacggcgatgtgcggcctgcagcagccgactccatgcccttacgctgctgcccgcgggt  
L T A M C G L Q Q P T P C P Y A A A G G  
gtcccccacgcc  
V P H A

— Péptido señal

### — . Dominio de repetición

..... Dominio Pro-X

FIG. 1.A

# Secuencia de RX3

```

atgagggtgtgtgctcggtgcccctcgctctcctggctctcgctgcyagcgccacctccacg
M R V L L V A L A L L A L A A S A T S T
catacaagcggcggtgcggtgccagccacgcgcgcgggttcattaccgcgcgcgggtg
H T S G G C G C Q P P P P V H L P P P V
catctgccacctccgggttcacctgccacctccgggtgcattctccaccgcgcgggtccacctg
H L P P P V H L P P P V H L P P P V H L
ccgcgcgcgggtccacctgccacgcgcgggtccatgtgccgcgcgcgggttcattgcgcgcg
P P P V H L P P P V H V P P P V H L P P
ccaccatgccactacctactcaacgcgcgcgcgggtccgcctcattccccaqccacacca
P P C H Y P T Q P P R P Q P H P Q P H P
tgcccggtgccaacgcgcgcgcgtccagcccggtgccagacc
C P C Q Q P H P S P C Q T
.....

```

# Secuencia de R3

```

atgagggtgtgtgctcggtgcccctcgctctcctggctctcgctgcyagcgccacctccacg
M R V L L V A L A L L A L A A S A T S T
catacaagcggcggtgcggtgccagccacgcgcgcgggttcattaccgcgcgcgggtg
H T S G G C G C Q P P P P V H L P P P V
catctgccacctccgggttcacctgccacctccgggtgcattctccaccgcgcgggtccacctg
H L P P P V H L P P P V H L P P P V H L
ccgcgcgcgggtccacctgccacgcgcgggtccatgtgccgcgcgcgggttcattgcgcgcg
P P P V H L P P P V H V P P P V H L P P
ccaccatgccactacctactcaacgcgcgcgcgggtccgcctcattccccaqccacacca
P P C H Y P T Q P P R T
.....

```

— Péptido señal

— . Dominio de repetición

..... Dominio Pro-X

FIG. 1.B

# Secuencia de P4

```

atgagggtgttgctcgttgccctcgctctcctggctctcgtgagcgccacctccacg
M R V L L V A L A L L A L A S A T S T
catacaanccggcggtgcggtgcccagccacggcgcggttcctctgcccgcgcacca
H T S G G C G C Q P P P P V H L P P P P
tgccactaccctacacaacggccccggcctcagcctcatccccagccacccatgcccg
C H Y P T Q P P R P Q P H P Q P H P C P
.....
tgccaacagccgcatccaagcccgtgccagacc
C Q Q P H P S P C Q T
.....

```

# Secuencia de X10

```

atgagggtgttgctcgttgccctcgctctcctggctctcgtgagcgccacctccacg
M R V L L V A L A L L A L A S A T S T
catacaagcggcggtgcggtgccaatgccactaccctactcaaccgccccggcctcag
H T S G G C G C Q C H Y P T Q P P R P Q
cctcatccccagccacccatgcccggtgccaacagccgcatccaagcccgtgccagacc
P H P Q P H P C P C Q Q P H P S P C Q T
.....

```

— Péptido señal  
 — . Dominio de repetición  
 ..... Dominio Pro-X

FIG. 1.C

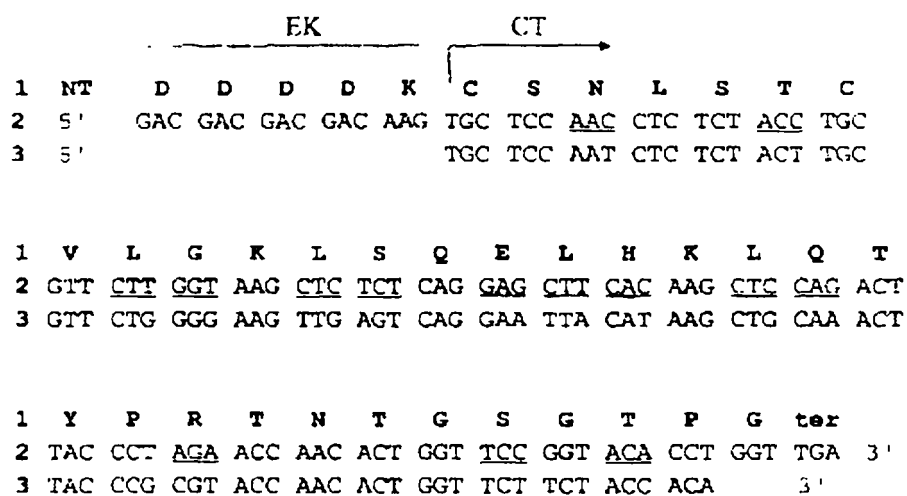


FIG. 2

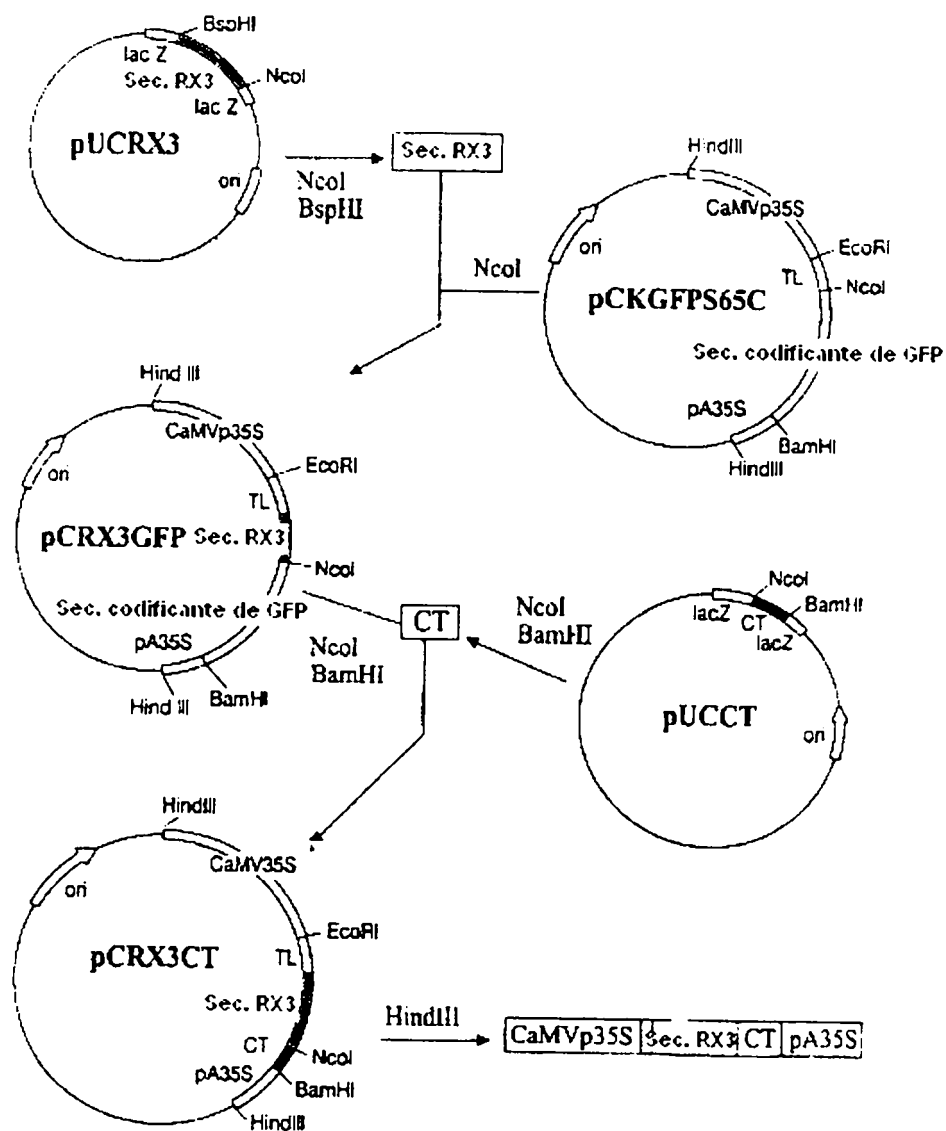


FIG. 3

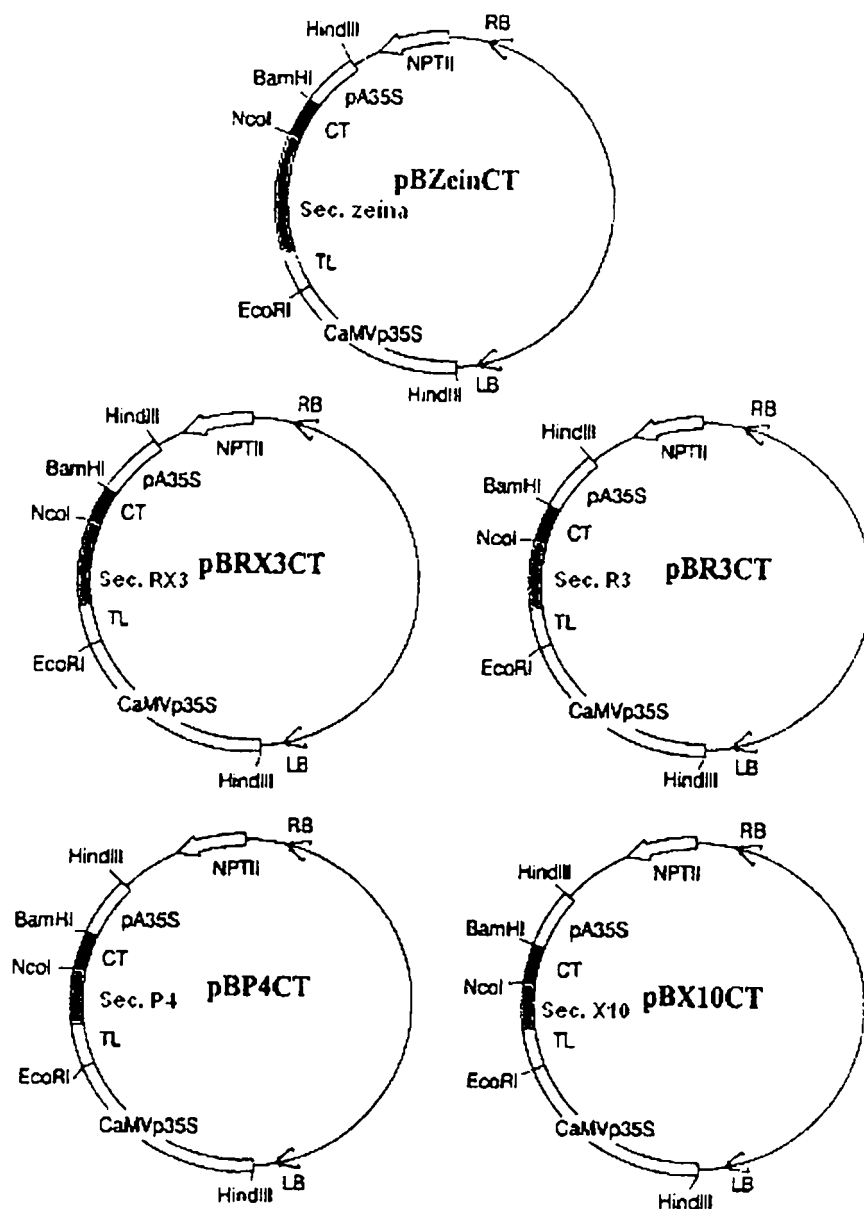


FIG. 4

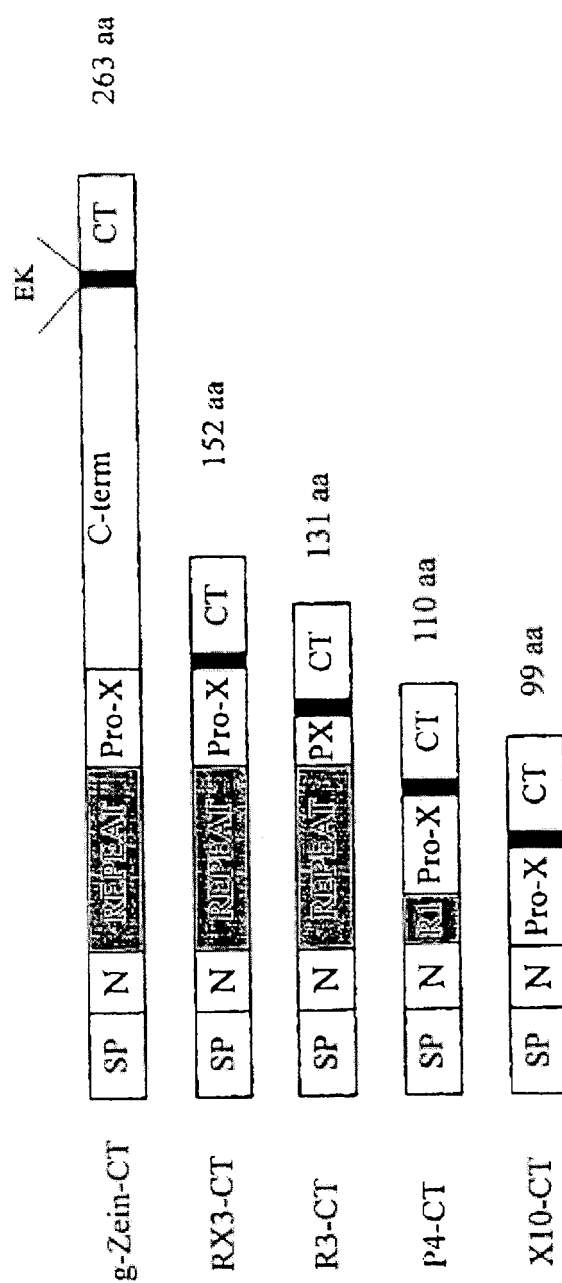


FIG. 5

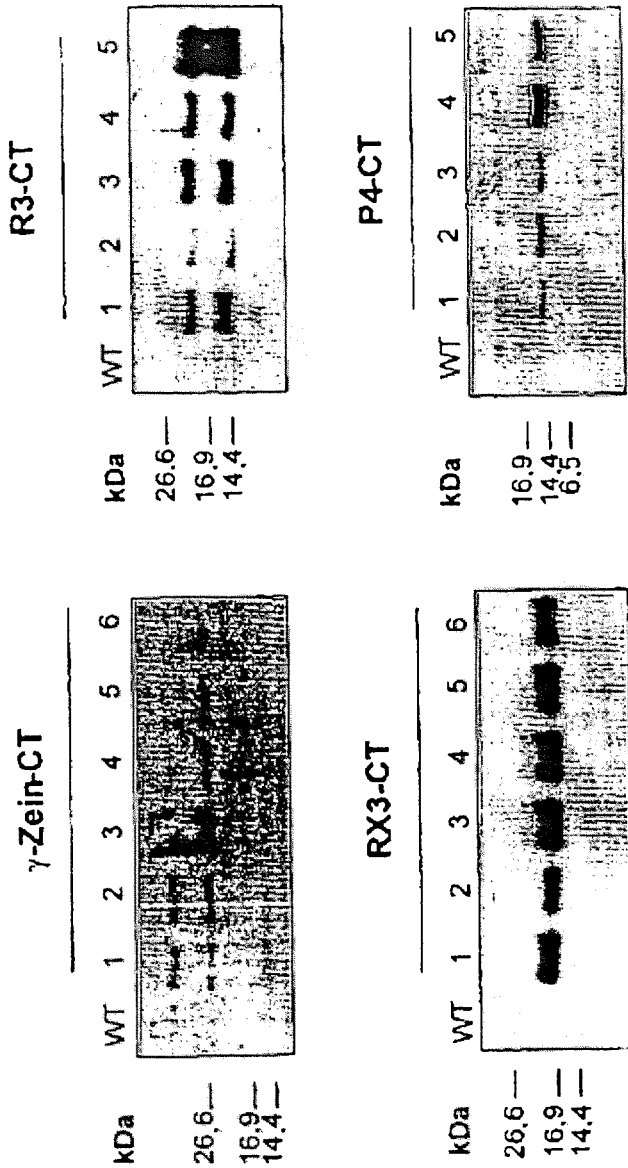


FIG. 6

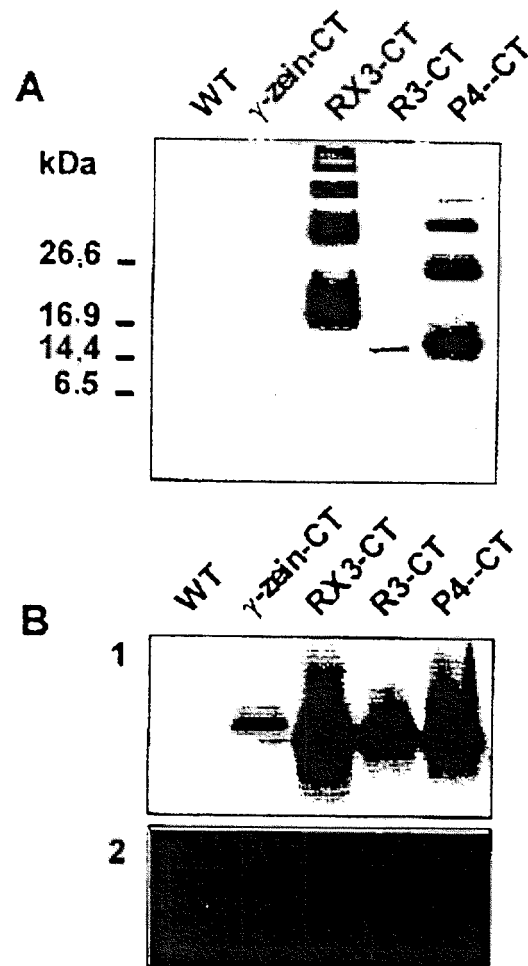


FIG. 7

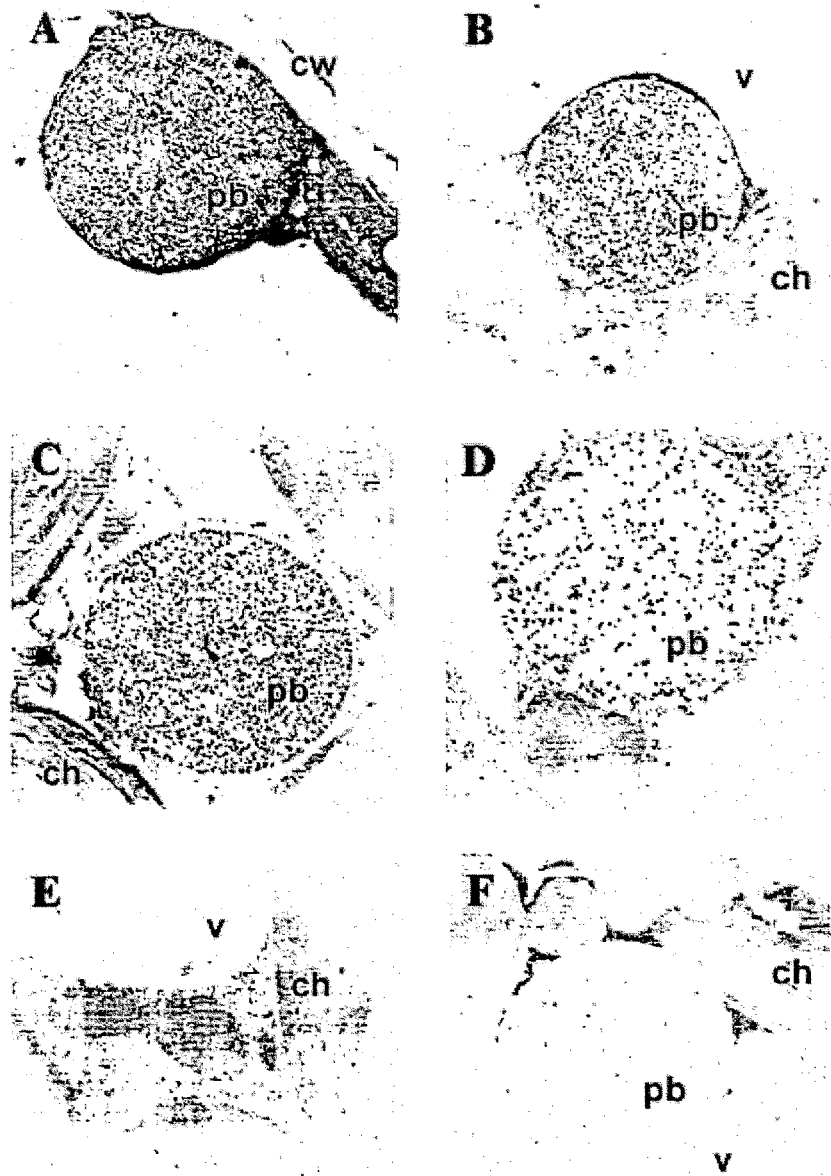


FIG. 8

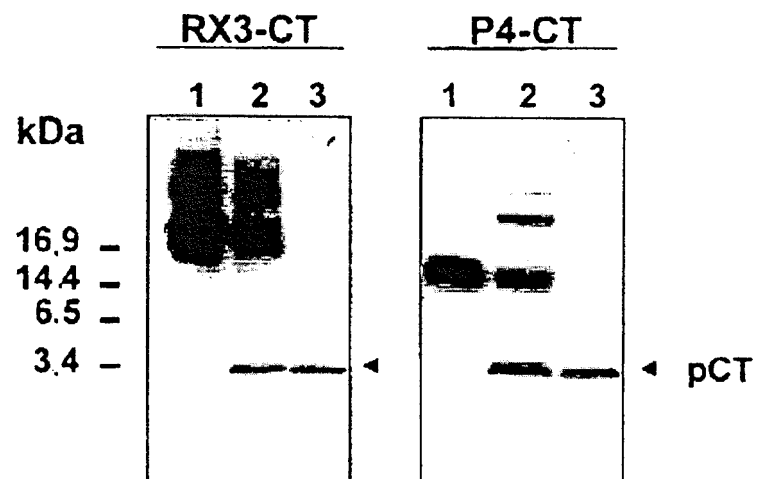


FIG. 9

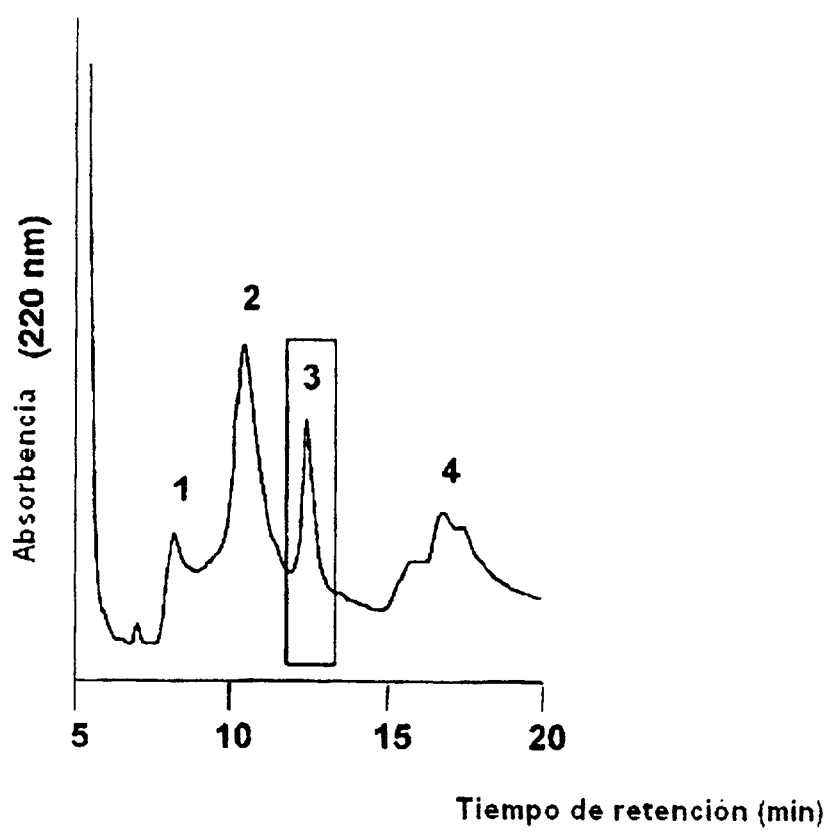


FIG. 10

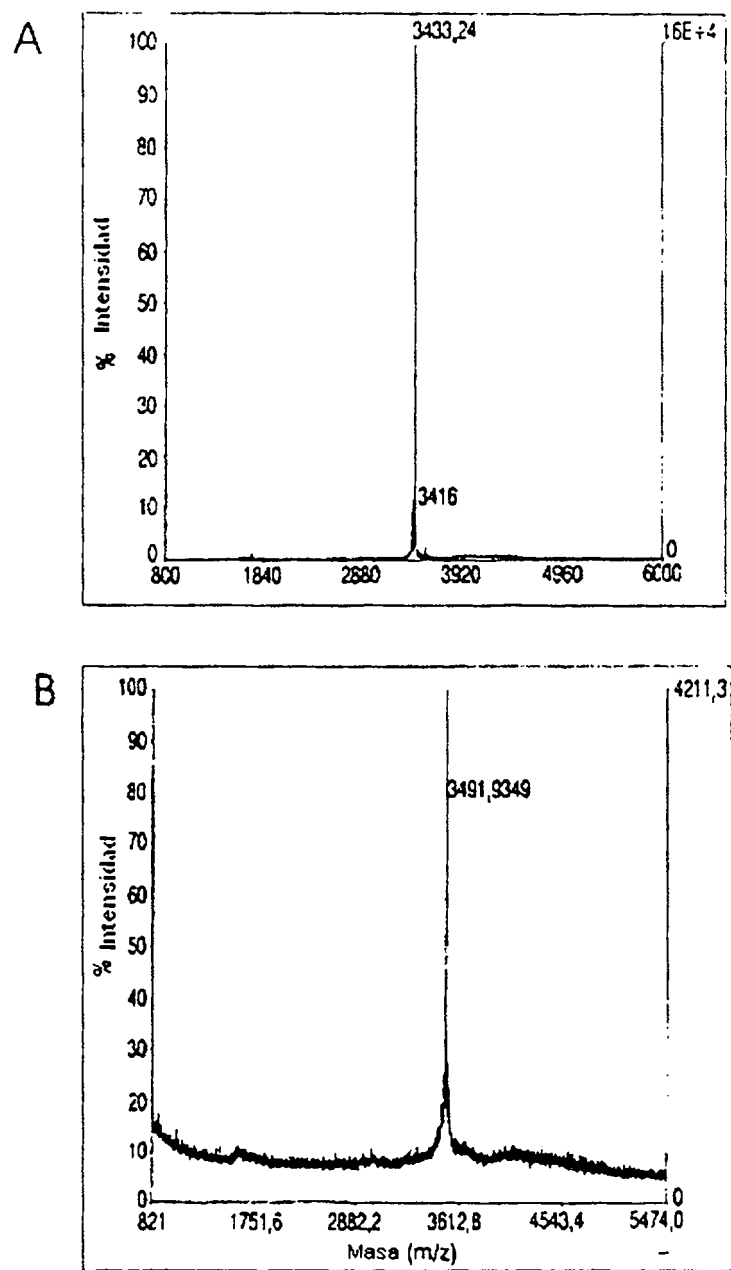


FIG. 11

# ES 2 289 185 T3

## LISTA DE SECUENCIAS

<110> Advanced *in Vitro* Cell Technologies, S.L.

5 <120> Producción de productos de interés por acumulación en cuerpos proteicos derivados del retículo endoplásmico.

<160> 10

10 <210> 1

<211> 672

<212> ADN

15 <213> *Zea mays*

<400> 1

|    |             |             |             |             |            |             |     |
|----|-------------|-------------|-------------|-------------|------------|-------------|-----|
| 20 | atgaggggtgt | tgctcgtttgc | cctcgtctctc | ctggctctctg | ctgcgagcgc | cacctccacg  | 60  |
|    | catacaagcg  | gcggtctgcg  | ctgccagcca  | ccgccgcggg  | ttcatctacc | gccgcgggtg  | 120 |
|    | catctgccac  | ctccggttca  | cctgccacct  | ccggtgcac   | tcccaccgcc | ggtccacctg  | 180 |
|    | ccgccgcggg  | tccacctgcc  | accgccggtc  | catgtgccgc  | cgccggttca | tctgccgcgg  | 240 |
|    | ccaccatgcc  | actaccctac  | tcaaccgccc  | cggcctcagc  | ctcatcccca | gccacaccca  | 300 |
|    | tgcccggtgcc | aacagccgca  | tccaagcccg  | tgccagctgc  | agggaacctg | cggcggttggc | 360 |
| 25 | agcaccgccga | tccctgggcca | gtgcgtcgag  | tttctgaggc  | atcagtgcag | cccgcgcggc  | 420 |
|    | acgcctact   | gtctgcctca  | gtgccagtcg  | ttgcggcagc  | agtgttgcca | gcagctcagg  | 480 |
|    | caggtggagc  | cgcagcaccg  | gtaccaggcg  | atcttcggct  | tggtcctcca | gtccatcctg  | 540 |
|    | cagcagcagc  | cgcacaagcg  | ccaggctcgc  | gggctgttgg  | cggcgagat  | agcgcagcaa  | 600 |
| 30 | ctgacggcga  | tgtgcggcct  | gcagcagccg  | actccatgcc  | cctacgctgc | tgccgcgggt  | 660 |
|    | gtcccccaag  | cc          |             |             |            |             | 672 |

<210> 2

35 <211> 339

<212> ADN

<213>

40 <400> 2

|    |             |             |             |             |            |             |     |
|----|-------------|-------------|-------------|-------------|------------|-------------|-----|
| 45 | atgaggggtgt | tgctcgtttgc | cctcgtctctc | ctggctctctg | ctgcgagcgc | caacctccacg | 60  |
|    | catacaagcg  | gcggtctgcg  | ctgccagcca  | ccgccgcggg  | ttcatctacc | gccgcgggtg  | 120 |
|    | catctgccac  | ctccggttca  | cctgccacct  | ccggtgcac   | tcccaccgcc | ggtccacctg  | 180 |
|    | ccgccgcggg  | tccacctgcc  | accgccggtc  | catgtgccgc  | cgccggttca | tctgccgcgg  | 240 |
|    | ccaccatgcc  | actaccctac  | tcaaccgccc  | cggcctcagc  | ctcatcccca | gccacaccca  | 300 |
|    | tgcccggtgcc | aacagccgca  | tccaagcccg  | tgccagacc   |            |             | 339 |

50 <210> 3

<211> 276

<212> ADN

<213>

55

<400> 3

|    |             |             |             |             |            |             |     |
|----|-------------|-------------|-------------|-------------|------------|-------------|-----|
| 60 | atgaggggtgt | tgctcgtttgc | cctcgtctctc | ctggctctctg | ctgcgagcgc | caacctccacg | 60  |
|    | catacaagcg  | gcggtctgcg  | ctgccagcca  | ccgccgcggg  | ttcatctacc | gccgcgggtg  | 120 |
|    | catctgccac  | ctccggttca  | cctgccacct  | ccggtgcac   | tcccaccgcc | ggtccacctg  | 180 |
|    | ccgccgcggg  | tccacctgcc  | accgccggtc  | catgtgccgc  | cgccggttca | tctgccgcgg  | 240 |
|    | ccaccatgcc  | actaccctac  | tcaaccgccc  | cggacc      |            |             | 276 |

65 <210> 4

<211> 213

# ES 2 289 185 T3

<212> ADN

<213>

5 <400> 4

```

          atgagggtgt tgcctgttgc cctcgtcttc ctggtctctg ctgcgagcgc caccctccacg    60
          catacaagcg gcggctgcgg ctgccagcca ccgccgcggg ttcattctgcc gccgccacca    120
          tgccactacc ctacacaacc gcccgcggct cagcctcctc cccagccaca cccatgcccg    180
10         tgccaacagc cgcattccaag cccgtgccag acc                                     213

```

<210> 5

<211> 180

15 <212> ADN

<213>

<400> 5

```

20         atgagggtgt tgcctgttgc cctcgtcttc ctggtctctg ctgcgagcgc caccctccacg    60
          catacaagcg gcggctgcgg ctgccaatgc cactacccta ctcaaccgcc ccggcctcag    120
          cctcattccc agccacaccc atgcccgctgc caacagccgc atccaagccc gtgccagacc    180

```

25 <210> 6

<211> 224

<212> PRT

30 <213> *Zea mays*

<400> 6

```

35         Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
           1             5             10             15

          Ala Thr Ser Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro
           20             25             30

40         Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu
           35             40             45

          Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
           50             55             60

          His Leu Pro Pro Pro Val His Val Pro Pro Pro Val His Leu Pro Pro
           65             70             75             80

50         Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro
           85             90             95

          Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
           100            105            110

          Leu Gln Gly Thr Cys Gly Val Gly Ser Thr Pro Ile Leu Gly Gln Cys
           115            120            125

60         Val Glu Phe Leu Arg His Gln Cys Ser Pro Thr Ala Thr Pro Tyr Cys
           130            135            140

          Ser Pro Gln Cys Gln Ser Leu Arg Gln Gln Cys Cys Gln Gln Leu Arg
           145            150            155            160

```

# ES 2 289 185 T3

Gln Val Glu Pro Gln His Arg Tyr Gln Ala Ile Phe Gly Leu Val Leu  
 165 170 175  
 5 Gln Ser Ile Leu Gln Gln Gln Pro Gln Ser Gly Gln Val Ala Gly Leu  
 180 185 190  
 10 Leu Ala Ala Gln Ile Ala Gln Gln Leu Thr Ala Met Cys Gly Leu Gln  
 195 200 205  
 Gln Pro Thr Pro Cys Pro Tyr Ala Ala Ala Gly Gly Val Pro His Ala  
 210 215 220  
 15  
 <210> 7  
 <211> 113  
 20 <212> PRT  
 <213>  
 <400> 7  
 25 Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser  
 1 5 10 15  
 30 Ala Thr Ser Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro  
 20 25 30  
 Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu  
 35 35 40 45  
 Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val  
 50 55 60  
 40 His Leu Pro Pro Pro Val His Val Pro Pro Pro Val His Leu Pro Pro  
 65 70 75 80  
 Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro  
 45 85 90 95  
 Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln  
 100 105 110  
 50 Tyr  
 <210> 8  
 <211> 92  
 55 <212> PRT  
 60  
 65

# ES 2 289 185 T3

<400> 8

```

5      Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
      1          5          10          15
      Ala Thr Ser Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro
      20          25          30
10     Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu
      35          40          45
      Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
15     50          55          60
      His Leu Pro Pro Pro Val His Val Pro Pro Pro Val His Leu Pro Pro
      65          70          75          80
20     Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Tyr
      85          90

```

<210> 9

25 <211> 71

<212> PRT

<400> 9

```

30     Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
      1          5          10          15
35     Ala Thr Ser Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro
      20          25          30
      Pro Val His Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro
      35          40          45
40     Arg Pro Gln Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro
      50          55          60
45     His Pro Ser Pro Cys Gln Tyr
      65          70

```

<210> 10

<211> 60

50 <212> PRT

<400> 10

```

55     Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
      1          5          10          15
      Ala Thr Ser Thr His Thr Ser Gly Gly Cys Gly Cys Gln Cys His Tyr
      20          25          30
60     Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His Pro Cys
      35          40          45
      Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln Tyr
65     50          55          60

```