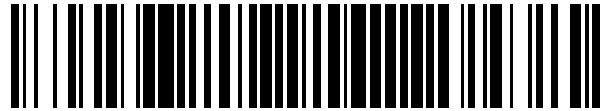


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 623 147**

51 Int. Cl.:

C12N 15/82 (2006.01)

C07K 14/415 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.05.2011 PCT/EP2011/058864**

87 Fecha y número de publicación internacional: **01.12.2011 WO11147995**

96 Fecha de presentación y número de la solicitud europea: **30.05.2011 E 11723923 (6)**

97 Fecha y número de publicación de la concesión europea: **25.01.2017 EP 2576797**

54 Título: **Secuencias polipeptídicas inductoras de cuerpos proteicos**

30 Prioridad:

13.08.2010 EP 10382231
28.05.2010 US 349655 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.07.2017

73 Titular/es:

ZERA INTEIN PROTEIN SOLUTIONS, S.L.
(100.0%)
Calle Cavallers, 50
08034 Barcelona, ES

72 Inventor/es:

MARZÁBAL LUNA, PABLO;
LLOP, M^a INMACULADA;
LUDEVID MUGICA, MARÍA DOLORES;
LLOMPART ROYO, BLANCA;
MINU, JOSEPH;
TORRENT QUETGLAS, MARGARITA y
BASTIDA VIRGILI, MIRIAM

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 623 147 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Secuencias polipeptídicas inductoras de cuerpos proteicos

5 **Campo de la invención**

La presente invención contempla péptidos que son capaces de inducir la formación de un ensamblaje recombinante similar a cuerpos proteicos (RPBLA, por sus siglas en inglés) en una célula eucariota. Además, polipéptidos heterólogos que están fusionados a secuencias que median la inducción de la formación de ensamblajes recombinantes similares a cuerpos proteicos (RPBLA), se expresan de forma estable y se acumulan en sistemas huéspedes.

Antecedentes de la invención

15 Hace varias décadas, se describieron las prolaminas como proteínas de almacenamiento muy especializadas implicadas en la formación de cuerpos proteicos naturales (PB, por sus siglas en inglés) en el endospermo de cereales (Sherry et al., 1990, *Biochem. J.* 267:1-12). Sin embargo, hasta la fecha, se sabe poco sobre los mecanismos y requisitos para inducir este tipo de orgánulo.

20 El endospermo es un tejido vegetal especializado que parece tener una mayor tendencia a transportar proteínas a los PB que otros tejidos y tipos celulares. Esto es cierto incluso cuando las proteínas no están fusionadas a prolaminas. Por ejemplo, la proteína fitasa recombinante, que se secreta de las células de hoja de arroz, se retiene en los PB cuando se expresa en el endospermo. Drakakaki et al., 2006, *Plant Physiology*, 141, 578-586. De forma similar, la glucoproteína (gB) principal del citomegalovirus humano (Wright et al., 2001, *Transgenic Research* 10: 25 177-181) y lisozima (Yang et al., 2003, *Planta* 216: 597-603) también se acumulan en los PB en endospermos de plantas dicotiledóneas y monocotiledóneas incluso cuando no están fusionadas a prolaminas. De forma interesante, incluso cuando lisozima estaba fusionada a un péptido señal no relacionado con prolamina y se expresó bajo el control de un promotor no de prolamina (puroindolina b) se acumuló en los PB en endospermo de arroz. Hennegan et al., 2005, *Transgenic Research* 14:583-592. Estos datos sugieren que la acumulación de proteínas en PB podría 30 requerir el medio de almacenamiento especializado del endospermo.

La propensión aumentada a ser transportado en endospermo también se demuestra por experimentos usando proteínas recombinantes etiquetadas con KDEL (SEQ ID NO: 110). La etiqueta KDEL también se conoce como una "señal de retención en el RE" porque ayuda a mantener proteínas en el retículo endoplásmico (RE). Por tanto, cuando seroalbúmina se fusiona a la etiqueta KDEL, se localiza en la luz del RE en células de hojas. Sin embargo, cuando se expresa en el endospermo, la seroalbúmina humana etiquetada con KDEL se depositó en agregados de prolamina dentro de vacuolas. Además, un anticuerpo monoclonal etiquetado con KDEL, que se retiene eficazmente en el retículo endoplásmico en hojas, se secretó parcialmente y se transportó parcialmente a vacuolas de almacenamiento de proteínas en semillas. Petruccelli et al., 2006, *Plant Biotechnol J.* 4:511-27.

40 Por tanto, el transporte de proteínas en células de endospermo de cereal está afectado por el medio específico del endospermo, incluyendo la abundancia de cuerpos proteicos vacuolares y derivados del retículo endoplásmico (RE). Según esto, las proteínas que se transportan a PB en el endospermo pueden no ser transportadas allí en otras células o tejidos, y las secuencias y estructuras que son suficientes para inducir la formación de PB en el 45 endospermo pueden no ser necesariamente suficientes para la formación de PB en otras células y tejidos.

Además, las secuencias y estructuras específicas que son suficientes para inducir la formación de PB no se han identificado. De hecho, cuando se comparan todas las proteínas implicadas en la formación PB, no es evidente ninguna homología clara en términos de secuencias, estructura o características físicas y químicas.

50 La gamma zeína es un constituyente principal de los cuerpos proteicos en maíz (Ludevid et al., 1984, *Plant Mol. Biol.* 3, 227-234). El dominio N-terminal de gamma zeína contiene una región Pro-X (P-X) y una secuencia muy repetitiva (PPPVHL)₈ (SEQ ID NO: 27) (RD) necesarias para el transporte de zeína en el RE (Geli, et al. *Plant Cell* 6:1911 (1994)) y para la formación de cuerpos proteicos. Véase también, la solicitud publicada de EE UU No. 2007/0243198. Un estudio de dicroísmo circular de una serie de péptidos sintéticos de la secuencia (VHLPPP)_n (SEQ ID NO: 28) (n=3, 5, 8) en agua a pH 5, mostró que estos péptidos adoptan una hélice de poliprolina II (PII) (Rabanal, *Biopolymers* 33: 1019-28 (1993)). Además, se ha mostrado que la secuencia PPPVAL (SEQ ID NO: 16) adopta conformación PPII en solución acuosa, aunque no se ha divulgado la formación de cuerpos proteicos (Dalcol, et al., 1996 *Journal of Organic Chemistry*, vol. 61, no. 20, 1996, páginas 6775-6782). La gamma zeína también 60 contiene varias cisteínas que se mostró que se requerían para la formación de PB estables (Pompa, *Plant Cell* 18: 2608-2621 (2006)). Los documentos WO2007096192, WO2004/003207 y WO2006/056484 divulgan una secuencia que induce cuerpos proteicos que contiene la secuencia (PPPVHL)₆ (SEQ ID NO: 29), una secuencia anfipática, flanqueada por dos cisteínas en el extremo N o C y el documento WO2004/003207 también divulga una secuencia que induce cuerpos proteicos que comprende el P4 que contiene una repetición PPPVHL (SEQ ID NO: 4), una 65 secuencia rica en prolina anfipática flanqueada por dos cisteínas en el extremo N y el C.

La hélice PPII del RD de gamma zeína tiene un marcado carácter anfipático. Estudios previos han sugerido que la naturaleza anfipática de la hélice PPII era importante para la formación de PB estables, y se han demostrado las propiedades tensoactivas de la hélice PPII anfipática (VHLPPP)₈ (SEQ ID NO: 27) mediante varios planteamientos (Kogan et al., 2001, *J. Mol. Biol.* 312, 907-913003; Kogan et al., 2002, *Biophysical Journal*. Volumen 83. 1194–1204). Por ejemplo, se ha mostrado que el péptido octamérico sintético (VHLPPP)₈ era capaz de disminuir la tensión de superficie del agua, debido en gran parte a la adsorción del anfifilo a la interfaz aire-agua con la fracción hidrofóbica orientada lejos de la fase acuosa (Ludevid et al., 1984, *Plant Mol. Biol.* 3, 227-234). También se demostró que este péptido anfipático interacciona con liposomas de fosfatidilcolina de soja y se ensambla para formar dominios extendidos sobre la membrana, lo que aumenta su estabilidad y permeabilidad (Kogan et al., 2004, *Biopolymers*, Vol. 73, 258–268). El ensamblaje anfipático espontáneo de (VHLPPP)₈ (SEQ ID NO: 27) en la membrana sugiere un mecanismo de deposición de gamma zeína dentro de los cuerpos proteicos de maíz. Basado en las características anfipáticas del RD de gamma zeína, se ha propuesto que esta proteína interacciona con la cara interna de la membrana del RE induciendo un recubrimiento interno que podría ser un elemento clave en el mecanismo de inducción de PB (Ludevid et al., 1984, *Plant Mol. Biol.* 3, 227-234). Este recubrimiento se podría estabilizar después covalentemente a través de entrecruzamiento disulfuro intramolecular que implica residuos de cisteína que flanquean la secuencia repetitiva de gamma zeína.

Mientras que se han caracterizado algunas de las características de la proteína gamma zeína, no se entendía previamente cuál de estas características o combinación de características era relevante para la formación de cuerpos proteicos. Además, otras secuencias inductoras de cuerpos proteicos contienen poca o ninguna similitud estructural o de secuencia con gamma zeína. Como se describe en más detalle posteriormente, se ha identificado una secuencia mínima capaz de inducir cuerpos proteicos. Además, se han identificado secuencias recombinantes inductoras de cuerpos proteicos con propiedades mejoradas, tales como una capacidad mejorada para formar ensamblajes recombinantes similares a cuerpos proteicos (RPBLA) y capacidad de formar RPLBA con características mejoradas.

Breve compendio de la invención

En un aspecto, la invención se refiere a una secuencia polipeptídica recombinante inductora de cuerpos proteicos (PBIS, por sus siglas en inglés) que comprende

- (i) una estructura de poliprolina II (PPII) que tiene al menos 36 aminoácidos de longitud;
 - en donde la estructura PPII se localiza entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C;
 - en donde no más del 10% de los aminoácidos en la estructura PPII son lisina o arginina; y
 - en donde la estructura PPII es una secuencia no anfipática y comprende al menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO:17), o
- (ii) una secuencia rica en prolina que tiene al menos 36 aminoácidos de longitud;
 - en donde la secuencia rica en prolina se localiza entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C;
 - en donde no más del 10% de los aminoácidos en la secuencia rica en prolina son lisina o arginina; y
 - en donde la secuencia rica en prolina es una secuencia no anfipática y comprende al menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO:17).

En otro aspecto, la invención se refiere a una proteína de fusión que comprende la PBIS recombinante según la invención y una proteína heteróloga.

En aspectos adicionales, la invención se refiere a una molécula de ácido nucleico que comprende una secuencia que codifica la PBIS recombinante según la invención o la proteína de fusión según la invención.

En otro aspecto, la invención se refiere a un RPBLA que comprende la PBIS recombinante según la invención o la proteína de fusión según la invención. En otro aspecto, la invención se refiere a una célula huésped que comprende la PBIS recombinante de la invención, la proteína de fusión según la invención, el ácido nucleico según la invención o el RPBLA según la invención.

En aspectos adicionales, la invención se refiere a métodos para producir el RPBLA según la invención y a un método para purificar una proteína de fusión según la invención.

En aspectos adicionales, la invención se refiere a una composición farmacéutica que comprende una cantidad terapéuticamente eficaz del RPBLA de la invención, a una vacuna que comprende una cantidad inmunológicamente eficaz del RPBLA de la invención y a un producto alimenticio que comprende el RPBLA según la invención.

Breve descripción de los dibujos/figuras

Figura 1. (A) Representación esquemática del dominio de repetición (RD) de diferentes variantes de péptidos RX3 fusionados a una proteína fluorescente indicadora, y transformadas mediante agroinfiltración en hojas de tabaco. La posición relativa de los aminoácidos diferentes de prolina en la hélice PPII se representa mediante un triángulo. Se muestra una línea de puntos que separa el lado hidrofóbico (abajo) del lado hidrofílico (arriba) en variantes con una hélice antipática. Se muestra el dominio de repetición de RX3 salvaje en el centro (caja gris). Las variantes completamente hidrofóbicas donde los residuos de histidina se han cambiado por alanina o leucina (RX3(A) y RX3(L), respectivamente) se muestran a la derecha. Las variantes anfipáticas cargadas positivamente donde los residuos de histidina se han cambiado por arginina o lisina (RX3(R) y RX3(K), respectivamente) se muestran a la izquierda. **(B)** Análisis por inmunotransferencia de cantidades equivalentes de homogeneizados de hojas de tabaco que expresan RX3-GFP, RX3(R)-GFP, RX3-ECFP y RX3(K)-ECFP, probados con anticuerpo anti-GFP. La punta de flecha blanca marca las proteínas de fusión y la punta de flecha negra marca una proteína fluorescente parcialmente degradada. **(C)** Imágenes de microscopía confocal de células de hoja de tabaco que expresan el péptido RX3 salvaje fusionado a GFP (RX3-GFP) o diferentes variantes del péptido RX3 en las que los residuos de histidina del dominio repetitivo se mutaron a alanina, leucina, arginina o lisina (RX3(A)-GFP, RX3(L)-GFP, RX3(R)-GFP y RX3(K)-ECFP, respectivamente). Las flechas marcan la fluorescencia correspondiente a proteína secretada. Las puntas de flecha blancas marcan los RPBLA derivados del RE y las puntas de flecha negras señalan cloroplastos. Las barras corresponden a 5 micrómetros.

Figura 2. (A) Fotografías de los resultados de gradiente de densidad por capas para varias proteínas de fusión (RX3-GFP, RX3(A)-GFP y RX3(L)-GFP) expresadas en hojas de tabaco y analizadas mediante inmunotransferencia usando un anticuerpo anti-GFP. **(B)** Tinción con azul de Coomassie que muestra el enriquecimiento en RX3(A)-GFP después de un gradiente de densidad en capas. Se indican las diferentes fracciones del gradiente de densidad ((H) homogeneizado, (S) sobrenadante, (f) interfaz encima del colchón de Optiprep™ correspondiente).

Figura 3. (A) Recuperación de RPBLA mediante centrifugación a baja velocidad. El panel izquierdo muestra tinción con plata de SDS-PAGE después del aislamiento de RPBLA de RX3-GFP y RX3(A)-GFP expresados en hojas de tabaco. El panel derecho muestra tinción con plata de SDS-PAGE de RPBLA de RX3-GFP y RX3(A)-GFP lavados en condiciones moderadas (borato 50 mM pH 10, bME 10 mM). **(B)** La tinción con azul de Coomassie (izquierda) y la inmunotransferencia (derecha; anticuerpo anti-GFP) muestran la recuperación mediante centrifugación a baja velocidad de los RPBLA inducidos en plantas de tabaco por la expresión de RX3(H)-GFP. (H0) homogeneizado preclarificado, (H1) homogeneizado clarificado mediante filtración, (SN) sobrenadante después de la recuperación de los RPBLA mediante centrifugación, (Ws) sobrenadante del paso de lavado, (wPB) RPBLA recuperados mediante centrifugación después del paso de lavado, (sPB) proteína de fusión solubilizada de wPB e (iPB) fracción insoluble después del paso de solubilización. Las puntas de flecha indican las proteínas de fusión monoméricas correspondientes. El asterisco indica las formas multiméricas de la proteína de fusión.

Figura 4. (A) Representación esquemática del dominio de repetición (RD) de RX3(A3) como se describe en la figura 1. **(B)** Imágenes de microscopía confocal de células de hoja de tabaco 3 y 6 días después de la infiltración (dpi) con el vector que expresa RX3(A3)-ECFP. Los RPBLA se indican con flechas. Las barras corresponden a 2 micrómetros.

Figura 5. (A) Representación esquemática del dominio de repetición de variantes anfipáticas cargadas negativamente del péptido RX3 (RX3(E) y RX3(D)) como se describe para la figura 1. **(B)** Inmunotransferencia con anti-GFP en cantidades equivalentes de homogeneizados de hojas de tabaco que expresan RX3-GFP, RX3(E)-GFP, RX3-ECFP y RX3(D)-ECFP. **(C)** Imágenes de microscopía confocal de células de hoja de tabaco que expresan RX3(E)-GFP (izquierda) y RX3(D)-ECFP (derecha). Los RPBLA se indican con flechas. Las barras corresponden a 2 micrómetros.

Figura 6. (A) Representación esquemática del dominio de repetición de variantes polares sin carga de péptidos RX3 (RX3(T), RX3(N) y RX3(Q)) como se describe para la figura 1. **(B)** Imágenes de RPBLA inducidos por la expresión de estos péptidos ensambladores fusionados a ECFP 3 y 6 días después de la agroinfiltración (dpi). Las barras mostradas en las imágenes de 3 y 6 dpi corresponden a 5 y 2 micrómetros, respectivamente.

Figura 7. (A) Alineamiento de secuencia de péptidos ensambladores RX3 maduro, PP y PA. La identidad entre los tres péptidos se indica en negrita, y los residuos de cisteína se indican mediante cajas grises. SP indica péptido señal. **(B)** Inmunotransferencia con anti-GFP de cantidades equivalentes de homogeneizados de hojas de tabaco que expresan PP-ECFP, PA-ECFP y RX3-ECFP. Las flechas marcan las proteínas de fusión. **(C)** Imágenes de microscopía confocal de células de hoja de tabaco que expresan RX3-ECFP, PP-ECFP y PA-ECFP. Los RPBLA se indican con flechas.

Figura 8. (A) Diagrama que muestra las posiciones de los residuos de cisteína en el péptido ensamblador RX3 fusionado a ECFP. **(B)** SDS-PAGE/tinción con azul de Coomassie (panel superior) e inmunotransferencia con anti-GFP (panel inferior) que muestran el análisis de proteínas totales de hojas de tabaco: tabaco sin transformar (carril

1); tabaco que expresa RX3-ECFP (carril 2); tabaco que expresa RX3 Cys7 (carril 3); tabaco que expresa RX3 Cys9 (carril 4); tabaco que expresa RX3 Cys64 (carril 5); tabaco que expresa RX3 Cys82 (carril 6); tabaco que expresa RX3 Cys84 (carril 7); tabaco que expresa RX3 Cys92 (carril 8); tabaco que expresa RX3 Cys7-Cys9 (carril 9); y tabaco que expresa RX3 Cys82- Cys84- Cys92- (carril 10). Las flechas indican las bandas electroforéticas de RX3-ECFP y mutantes Cys de RX3-ECFP y la punta de flecha indica una banda inmunorreactiva adicional. **(C-K)** Imágenes confocales que muestran el patrón de fluorescencia de células epidérmicas transformadas con los mutantes RX3-Cys fusionados a ECFP. Las barras en C-H corresponden a 10 μm . Las barras en I-K corresponden a 20 μm .

Figura 9. (A) Alineamiento de secuencia de PP maduro y PP2. La identidad entre los dos péptidos se indica en negrita y los residuos de cisteína se indican mediante una caja gris. SP indica péptido señal. **(B)** Representación esquemática de la hélice PPII de PP y PP2. Se indican los residuos de cisteína. **(C)** Imágenes de microscopía confocal de células de hoja de tabaco que expresan PP2-GFP. El panel izquierdo es un aumento del panel derecho.

Figura 10. (A) Representación esquemática del péptido RX3 fusionado a ECFP que presenta una reducción progresiva en el número de unidades en el RD. Las unidades se muestran en cajas grises numeradas y los residuos de cisteína se indican mediante asteriscos. **(B)** Gel de SDS-PAGE teñido con azul de Coomassie que muestra el patrón de proteínas de cantidades equivalentes de homogeneizados de hojas de tabaco que expresan proteínas de fusión RX3- R8(4C)-, R7(4C)-, R6(4C)- y R4(4C)-ECFP (flechas). **(C)** Imágenes de microscopía confocal de células de hoja de tabaco que expresan R8(4C)-, R6(4C)- y R4(4C)-ECFP. Las flechas en los recuadros marcan los RPBLA. Las barras corresponden a 20 micrómetros en la imagen completa y a 5 micrómetros en los recuadros.

Figura 11. (A, B y C) Imágenes de microscopía confocal de células de hoja de tabaco que expresan la proteína fluorescente mCherry fusionada a los péptidos RX3, RX3(A) y RX(E) (A, B y C, respectivamente). **(A', B' y C')** Mayor aumento de las imágenes en A, B y C para señalar los RPBLA (puntas de flecha). Las flechas muestran RX3-mCherry secretada. Las barras corresponden a 5 micrómetros.

Figura 12. (A) Inmunotransferencia que compara la recuperación mediante centrifugación a baja velocidad de RPBLA inducidos por la expresión de RX3-, RX3(E)- y RX3(A)-EGF. Se muestran el homogeneizado clarificado (carril 1), el sobrenadante después de la recuperación de RPBLA mediante centrifugación (carril 2), el sobrenadante del paso de lavado (carril 3) y los RPBLA recuperados mediante centrifugación después del paso de lavado (carril 4). **(B)** Inmunotransferencia que muestra RX3(E)-EGF (carril 1), RX3(A)-EGF (carril 2) y RX3-EGF (carril 3) solubilizadas después de incubar en condiciones moderadas. La fracción no solubilizada de la proteína de fusión correspondiente se recuperó mediante centrifugación a 16000 x g durante 10' y se muestra en los carriles 4, 5 y 6, respectivamente. **(C)** Análisis por inmunotransferencia de cantidades equivalentes de homogeneizados de hojas de tabaco que expresan PP-, PA-, RX3(E)- y RX3-EGF. **(D)** Inmunotransferencias que muestran las proteínas de fusión RX3(E)-EGF (carril 2), PP-EGF (carril 5) y PA-EGF (carril 8) solubilizadas de los correspondientes RPBLA aislados mediante centrifugación a baja velocidad (carriles 1, 4 y 7, respectivamente). La proteína de fusión restante no solubilizada se muestra en los carriles 3, 6 y 9.

Figura 13. (A) Tinción con plata de SDS-PAGE del proceso posterior de RX3(E)-EGF del aislamiento de RPBLA mediante centrifugación a baja velocidad al corte de la proteína de fusión por digestión con FXa: marcador molecular (carril 1); homogeneizado preclarificado (carril 2); homogeneizado clarificado por centrifugación (carril 3); sobrenadante después de recuperar los RPBLA mediante centrifugación (carril 4); sobrenadante del paso de lavado (carril 5, 6); RPBLA recuperados mediante centrifugación después del paso de lavado (carril 7); fracción insoluble después del paso de solubilización (carril 8); proteína de fusión solubilizada del wPB (carril 9); y proteína de fusión RX3(E)-EGF cortada (carril 10). **(B)** Cromatograma de la purificación de EGF por FPLC de fase reversa. La flecha indica el pico de EGF que corresponde al 30% de acetonitrilo. **(C)** Tinción con azul de Coomassie que muestra el aporte a RF-FPLC (carril 1) y las dos fracciones (carriles 2-3) que contienen el EGF puro y que corresponden al pico indicado por una flecha en (B).

Figura 14. (A) La tinción con azul de Coomassie de SDS-PAGE (izquierda) y la inmunofluorescencia (derecha) muestran RPBLA de RX3(A)-hGH aislados. **(B)** Inmunofluorescencia con anti-hGH que muestra la eficacia de solubilización de RX3-hGH y RX3(A)-hGH. (H0) homogeneizado preclarificado, (H1) homogeneizado clarificado mediante filtración, (SN) sobrenadante después de la recuperación de los RPBLA mediante centrifugación, (Ws) sobrenadante del paso de lavado, (wPB) RPBLA recuperados mediante centrifugación después del paso de lavado, (sPB) proteína de fusión solubilizada de wPB e (iPB) fracción insoluble después del paso de solubilización.

Figura 15. (A y B) Representación esquemática de las proteínas de fusión que indican la secuencia inductora de cuerpos proteicos (PBIS), el enlazador de 5 glicinas ((Gly)₅), el sitio de corte (CS), Intein MxeGyrA de New England Biolabs (I) y la proteína de interés (PDI). Los CS son el sitio de corte de la enteroquinasa (EK) y el sitio de corte de FXa (FXa). Las PDI son hormona de crecimiento humana (hGH), factor de crecimiento epidérmico (EGF), proteína fluorescente cian mejorada (ECFP), proteasa enteroquinasa (EKp), xilanasa (Xyl) y proteína fluorescente verde (GFP). Las proteínas de fusión Zera(Adh) son las proteínas de fusión basadas en las siguientes PBIS derivadas de un fragmento de adhesina que forma una hélice PPII: Z(Adh), Z(Adh2), Z(Adh)Px y Z(Adh2)Px. Las proteínas de

fusión Zera(Col) son las proteínas de fusión basadas en las siguientes PBIS derivadas de un fragmento de colágeno que forma una hélice PPII: Z(Col), Z(Col2), Z(Col)Px y Z(Col2)Px.

Figura 16. (A) En el panel superior, tinción con azul de Coomassie de 20 microgramos de proteína total de homogeneizados preclarificados de hojas de tabaco que expresan RX3ΔCys64,82,84,92-ECFP-KDEL (carril 1), RX3-ECFP (carril 2), y SP-ECFP-KDEL (carril 3). En la parte inferior, inmunotransferencia de los mismos homogeneizados mediante anticuerpo anti-GFP. La flecha indica las proteínas de fusión. **(B)** Imágenes de microscopía confocal de hojas de tabaco que expresan RX3ΔCys64,82,84,92-ECFP-KDEL, tres y siete días después de la infiltración (dpi). Las barras blancas corresponden a 5 micrómetros. **(C)** Determinación de la densidad de RX3-ECFP (izquierda) y RX3ΔCys64,82,84,92-ECFP-KDEL (derecha) por gradiente de densidad por capas. Se analizaron cantidades equivalentes de las fracciones recuperadas por tinción con azul de Coomassie (paneles superiores) e inmunotransferencia usando anticuerpo anti-GFP (paneles inferiores). Los carriles muestran: (1) muestra cargada en el gradiente de densidad Optiprep™, (2) sobrenadante recuperado de encima del colchón de Optiprep™ al 18%, (3) fracción entre el 18-30%, (4) fracción entre el 30-34%, (5) fracción entre el 34-38%, (6) fracción entre el 38-42%, (7) fracción entre el 42-46%; y (8) fracción recuperada en el precipitado.

Figura 17. (A) Alineamiento de secuencia de los péptidos ensambladores Z(Adh) (SEQ ID NO:98), Z(Adh)Px (SEQ ID NO:94), Z(Col) (SEQ ID NO:100), y Z(Col)Px (SEQ ID NO:96). El péptido señal (de gamma zeína de 27 KDa en todos los casos) no se muestra. Los aminoácidos subrayados (Adh – SEQ ID NO: 128; Col – SEQ ID NO: 129) corresponden a los fragmentos de proteína de los genes de adhesina o colágeno. La barra negra indica los fragmentos con una alta propensión a adoptar una estructura PPII. **(B)** Tinción con azul de Coomassie de homogeneizados preclarificados que provienen de cantidades equivalentes (20 microgramos de proteína total) de hojas de tabaco que expresan Z(Adh)Px-GFP, Z(Col)Px-GFP, Z(Adh)-GFP y Z(Col)-GFP (carriles 1 a 4, respectivamente). Las flechas marcan las proteínas de fusión. **(C)** De izquierda a derecha, imágenes de microscopía confocal de hojas de tabaco que expresan Z(Adh)-GFP, Z(Col)-GFP, Z(Adh)Px-GFP, y Z(Col)Px-GFP. Las barras blancas corresponden a 5 micrómetros. **(D)** Proceso posterior de la recuperación de los RPBLA por medio de centrifugación a baja velocidad a 1500xg analizado por tinción con azul de Coomassie. El homogeneizado preclarificado (carril 1) se centrifugó a 1500xg y el sobrenadante (carril 2) se desechó. Después de tres rondas de pasos de lavados (carriles 3-5), se obtuvo el precipitado correspondiente a la fracción de los RPBLA (carril 6). Se muestra el proceso se realizó en cantidades equivalentes de hojas de tabaco que expresan Z(Adh)-GFP (D1), Z(Col)-GFP (D2), Z(Adh)Px-GFP (D3), y Z(Col)Px-GFP (D4).

Figura 18. (A) Representación esquemática de los péptidos ensambladores maduros: (i) RX3 y (ii) RX3 invertido (iRX3). La orientación de los diferentes dominios: (N) fragmento N-terminal, (RD, en gris) dominio repetitivo y (PX) dominio Pro-X se indica mediante flechas. También se muestra la posición de los residuos de cisteína. **(B)** Imágenes de microscopía confocal de células de tabaco que expresan RX3-ECFP, ECFP-RX3, iRX3-ECFP y ECFP-iRX3. Los RPBLA se indican con flechas. Las barras corresponden a 5 micrómetros. **(C)** Determinación de la densidad de los RPBLA de plantas de tabaco que expresan RX3-ECFP, ECFP-RX3, iRX3-ECFP y ECFP-iRX3. El homogeneizado (carril 1) se cargó en la parte superior de un gradiente de densidad Optiprep multicapa, y se recogieron las siguientes fracciones a 80.000xg: (carril 2) sobrenadante, (carril 3) interfase por encima del colchón de 1,117 g/cm³, (carril 4) interfase por encima del colchón de 1,175 g/cm³, (carril 5) interfase por encima del colchón de 1,21 g/cm³, (carril 6) interfase por encima del colchón de 1,233 g/cm³, (carril 7) interfase por encima del colchón de 1,26 g/cm³ y (carril 8) precipitado en el fondo del tubo. Se analizaron cantidades volumétricas equivalentes de cada fracción por inmunotransferencia por medio de anticuerpo anti-RX3.

Figura 19. (A) Acumulación de hGH-iRX3 en RPBLA en células CHO. El panel izquierdo muestra la determinación de la densidad de los RPBLA inducidos por la expresión de hGH-iRX3 en células CHO. El homogeneizado (H) se cargó en un gradiente de densidad de sacarosa multicapa y se recogieron las siguientes fracciones después de centrifugación a 80.000xg: (S) sobrenadante, (F27) interfase por encima del colchón de sacarosa al 27%, (F35) interfase por encima del colchón de sacarosa al 35%, (F42) interfase por encima del colchón de sacarosa al 42%, (F56) interfase por encima del colchón de sacarosa al 56%, y (P) precipitado en el fondo del tubo. Los marcadores moleculares se indican a la izquierda en kDa y la posición esperada de la proteína de fusión hGH-iRX3 y hGH se indican a la derecha por puntas de flecha. El anticuerpo usado en la inmunotransferencia corresponde a anti-hGH. El panel derecho muestra inmunohistoquímica de la proteína de fusión hGH-iRX3 en células de mamífero CHO. Imágenes de microscopía confocal de células CHO que expresan hGH-iRX3 que se incubaron con un anticuerpo anti-hGH muestran la acumulación de la proteína de fusión en RPBLA intracelulares (flechas). **(B)** Inmunohistoquímica de proteínas de fusión EK-RX3 y DsRED-iRX3 en células de mamífero CHO. El panel izquierdo muestra imágenes de microscopía confocal de células CHO que expresan EK-RX3 incubadas con un anticuerpo anti-RX3 (αR8). El panel derecho muestra que los RPBLA se pueden observar directamente por la fluorescencia intrínseca de DsRED-iRX3. Los RPBLA intracelulares que contienen la proteína de fusión DsRED-iRX3 se muestran mediante flechas. N corresponde al núcleo celular. **(C)** Inducción de RPBLA por la expresión de hGH-iRX3 en células de insecto Sf9. El panel izquierdo muestra imágenes de microscopía confocal de células de insecto que expresan hGH-iRX3 e incubadas con anticuerpo anti-hGH. La imagen superior corresponde a células Sf9 no infectadas que muestran marcaje de fondo. La imagen inferior muestra una célula Sf9 que expresa la proteína de fusión hGH-iRX3 en RPBLA (flechas). El panel derecho muestra la recuperación de RPBLA por centrifugación a baja

velocidad. Un homogeneizado preclarificado de células de insecto Sf9 que expresan hGH-I-RX3 (carril 1) se centrifugó a 5000xg. El sobrenadante (carril 2) se desechó y el correspondiente precipitado que contiene los RPBLA (carril 3) se obtuvo después de varios pasos de lavado (carril 3). La punta de flecha indica la posición de la proteína de fusión hGH-I-RX3.

5

Descripción detallada de la invención

A continuación se proporciona una descripción de secuencias inductoras de cuerpos proteicos (PBIS) recombinantes que son útiles para la formación de ensamblajes recombinantes similares a cuerpos proteicos (RPBLA). Las PBIS recombinantes se pueden fusionar a proteínas de interés y los RPBLA formados por la expresión de las proteínas de fusión en las células se pueden usar para purificar de forma sencilla y eficaz grandes cantidades de la proteína de interés. Además, los RPBLA se pueden usar en terapéutica tal como vacunaciones.

10

Los encabezamientos de sección usados en el presente documento son sólo para fines de organización y no se deben interpretar en ninguna manera limitantes del objeto descrito.

15

I. Definiciones

A menos que expresamente se defina de otra manera, los términos usados en el presente documento se deben entender según su significado normal en la técnica. Los términos usados en singular o referidos como “un” o “una” también incluyen el plural y viceversa, a menos que se especifique o indique de otra manera por el contexto. Las técnicas y procedimientos estándar en general se realizan según métodos convencionales en la técnica y varias referencias generales (véase, en general, Sambrook *et al.* Molecular Cloning: A Laboratory Manual, 2ª ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), que se proporcionan en todo este documento.

20

25

Los términos “polipéptido”, “péptido” y “proteína” se usan en el presente documento de forma intercambiable para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y puede estar interrumpido por no aminoácidos. Los términos también abarcan un polímero de aminoácidos que se ha modificado de forma natural o mediante intervención; por ejemplo, formación de puentes disulfuro, glucosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación o modificación, tal como conjugación con un componente marcador. También se incluyen en la definición, por ejemplo, polipéptidos que contienen uno o más análogos de un aminoácido (incluyendo, por ejemplo, aminoácidos no naturales, etc.), así como otras modificaciones conocidas en la técnica. Se entiende que, puesto que los polipéptidos de esta divulgación se basan en anticuerpos, en ciertas divulgaciones, los polipéptidos se pueden producir como cadenas únicas o cadenas asociadas.

30

35

Un “polipéptido de fusión” o “proteína de fusión” es un polipéptido compuesto de al menos dos polipéptidos y opcionalmente una secuencia enlazadora para unir operativamente los dos polipéptidos en un polipéptido continuo. Los dos polipéptidos unidos en un polipéptido de fusión típicamente derivan de dos fuentes independientes y por tanto un polipéptido de fusión comprende dos polipéptidos unidos que normalmente no se encuentran unidos en la naturaleza. Los dos polipéptidos pueden estar operativamente unidos directamente por un enlace peptídico o pueden estar unidos indirectamente a través de un enlazador descrito aquí o de otra manera conocido en la técnica.

40

Un “ácido nucleico”, “polinucleótido” o “molécula de ácido nucleico” es un compuesto polimérico compuesto de subunidades covalentemente unidas denominadas nucleótidos. Ácido nucleico incluye ácido polirribonucleico (ARN) y ácido polidesoxirribonucleico (ADN), ambos pueden ser monocatenarios o bicatenarios. El ADN incluye ADNc, ADN genómico, ADN sintético y ADN semisintético.

45

Los términos “idéntico” o porcentaje de “identidad” en el contexto de dos o más ácidos nucleicos o polipéptidos, se refiere a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de nucleótidos o residuos de aminoácidos que son iguales, cuando se comparan y alinean (introduciendo huecos, si es necesario) para una correspondencia máxima, sin considerar ninguna sustitución conservadora de aminoácidos como parte de la identidad de secuencia. Se puede medir el porcentaje de identidad usando software o algoritmos de comparación de secuencias o mediante inspección visual. Se conocen varios algoritmos y software en la técnica que se pueden usar para obtener alineamientos de secuencias de aminoácidos o nucleótidos. Uno de tales ejemplos no limitantes de un algoritmo de alineamiento de secuencias es el algoritmo descrito en Karlin *et al.*, 1990, *Proc. Natl. Acad. Sci.*, 87:2264-2268, modificado en Karlin *et al.*, 1993, *Proc. Natl. Acad. Sci.*, 90:5873-5877, e incorporado en los programas NBLAST y XBLAST (Altschul *et al.*, 1991, *Nucleic Acids Res.*, 25:3389-3402). En ciertas formas de realización, se puede usar Gapped BLAST como describen Altschul *et al.*, 1997, *Nucleic Acids Res.* 25:3389-3402. BLAST-2, WU-BLAST-2 (Altschul *et al.*, 1996, *Methods in Enzymology*, 266:460-480), ALIGN, ALIGN-2 (Genentech, South San Francisco, California) o Megalign (DNASTAR) son programas informáticos adicionales públicamente disponibles que se pueden usar para alinear secuencias. En ciertas formas de realización, se determina el porcentaje de identidad entre dos secuencias de nucleótidos usando el programa GAP en el software CGC (por ejemplo, usando una matriz NWSgapdna.CMP y un peso de brecha de 40, 50, 60, 70 o 90 y un peso de longitud de 1, 2, 3, 4, 5 o 6). En ciertas formas de realización alternativas, se puede usar el programa GAP en el paquete de

50

55

60

65

software CGC, que incorpora el algoritmo de Needleman y Wunsch (*J. Mol. Biol.* 48:444-453 (1970)) para determinar el porcentaje de identidad entre dos secuencias de aminoácidos (por ejemplo, usando una matriz Blossum 62 o una matriz PAM250 y un peso de brecha de 16, 14, 12, 10, 8, 6 o 4 y un peso de longitud de 1, 2, 3, 4, 5). De forma alternativa, en ciertas formas de realización, se determina el porcentaje de identidad entre secuencias de nucleótidos o aminoácidos usando el algoritmo de Myers y Miller (CABIOS, 4:11-17 (1989)). Por ejemplo, se puede determinar el porcentaje de identidad usando el programa ALIGN (versión 2.0) y usando PAM120 con una tabla de residuos, una penalización por longitud de brecha de 12 y una penalización de brecha de 4. Los parámetros apropiados para el alineamiento máximo mediante un software de alineamiento particular los puede determinar el experto en la materia. En ciertas formas de realización, se usan los parámetros por defecto del software de alineamiento. En ciertas formas de realización, el porcentaje de identidad "X" de una primera secuencia de aminoácidos respecto a una segunda secuencia de aminoácidos se calcula como $100 \times (Y/Z)$, donde Y es el número de residuos de aminoácidos puntuados como emparejamientos idénticos en el alineamiento de la primera y segunda secuencias (alineadas mediante inspección visual o un programa de alineamiento de secuencias particular) y Z es el número total de residuos en la segunda secuencia. Si la longitud de una primera secuencia es mayor que la segunda secuencia, el porcentaje de identidad de la primera secuencia respecto a la segunda secuencia será más largo que el porcentaje de identidad de la segunda secuencia respecto a la primera secuencia.

Como ejemplo no limitante, si cualquier polinucleótido particular tiene un cierto porcentaje de identidad de secuencia (por ejemplo, es al menos el 80% idéntico, al menos el 85% idéntico, al menos el 90% idéntico y en algunas formas de realización al menos el 95%, el 96%, el 97%, el 98% o el 99% idéntico) a una secuencia de referencia se puede determinar, en ciertas formas de realización, usando el programa Bestfit (Wisconsin Sequence Analysis Package, Versión 8 para Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit usa el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2: 482-489 (1981), para determinar el mejor segmento de homología entre dos secuencias. Cuando se usa Bestfit o cualquier otro programa de alineamiento de secuencia para determinar si una secuencia particular es, por ejemplo, el 95% idéntica a una secuencia de referencia según la presente invención, los parámetros se ajustan de modo que el porcentaje de identidad se calcula en la longitud completa de la secuencia de nucleótidos de referencia y que se permitan brechas en homología de hasta el 5% del número total de nucleótidos en la secuencia de referencia.

En algunas formas de realización, dos ácidos nucleicos o polipéptidos de la invención son sustancialmente idénticos, lo que significa que tienen al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90% y en algunas formas de realización al menos el 95%, el 96%, el 97%, el 98%, el 99% de identidad de residuos de nucleótidos o aminoácidos, cuando se comparan y alinean para la máxima correspondencia, medido usando un algoritmo de comparación de secuencias o mediante inspección visual. La identidad puede existir en una región de las secuencias que tiene al menos aproximadamente 10, aproximadamente 20, aproximadamente 40-60 residuos de longitud o cualquier valor integral entre ellos, y puede ser en una región más larga de 60-80 residuos, por ejemplo al menos aproximadamente 90-100 residuos y en algunas formas de realización, las secuencias son sustancialmente idénticas en la longitud completa de las secuencias que se comparan, tal como la región codificante de una secuencia de nucleótidos, por ejemplo.

El término "vector" significa una construcción, que es capaz de administrar y opcionalmente expresar uno o más gen(es) o secuencia(s) de interés en una célula huésped. Ejemplos de vectores incluyen, pero no están limitados a, vectores víricos, vectores de expresión de ADN o ARN desnudo, vectores de plásmidos, cósmidos o fagos, vectores de expresión de ADN o ARN asociados con agentes de condensación catiónicos, vectores de expresión de ADN o ARN encapsulados en liposomas, y ciertas células eucariotas, tales como células productoras. Los vectores pueden ser estables y pueden ser autoreplicantes. Un "vector de expresión" es un vector que es capaz de dirigir la expresión de genes a los que está operativamente asociado.

"Promotor" se refiere a un fragmento de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, una región codificante se localiza en 3' respecto a un promotor. Los promotores pueden derivar en su totalidad de un gen nativo o estar compuestos de diferentes elementos derivados de diferentes promotores encontrados en la naturaleza o incluso comprender segmentos de ADN sintético. Los expertos en la materia entienden que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos celulares o en diferentes fases del desarrollo o en respuesta a diferentes condiciones medioambientales o fisiológicas. Los promotores que producen que un gen se exprese en la mayoría de los tipos celulares en la mayoría de los tiempos comúnmente se denominan "promotores constitutivos". Se reconoce además que, puesto que en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido por completo, fragmentos de ADN de diferente longitud pueden tener actividad promotora idéntica. Un promotor en general está unido en su extremo 3' al sitio de iniciación de la transcripción y se extiende corriente arriba (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables sobre el fondo. Dentro del promotor se encontrará un sitio de iniciación de la transcripción (convenientemente definido, por ejemplo, mapeando con nucleasa S1), así como dominios de unión de proteínas (secuencias consenso) responsables de la unión de ARN polimerasa.

El término “heterólogo” como se usa en el presente documento se refiere a un elemento de un vector, plásmido o célula huésped que deriva de una fuente diferente de la fuente endógena. De esta manera, por ejemplo, una secuencia heteróloga pudiera ser una secuencia que deriva de un gen o plásmido diferente del mismo huésped, de una cepa diferente de célula huésped o de un organismo de un grupo taxonómico diferente (por ejemplo, reino, filo, clase, orden, familia, género o especie diferente o cualquier subgrupo dentro de una de estas clasificaciones). El término “heterólogo” también se usa aquí como sinónimo del término “exógeno”.

Una “región codificante” de ADN o ARN es una molécula de ADN o ARN que se transcribe y/o traduce en un polipéptido en una célula in vitro o in vivo cuando se coloca bajo el control de secuencias reguladoras apropiadas. “Regiones reguladoras adecuadas” se refiere a regiones de ácido nucleico localizadas antes (secuencias 5' no codificantes), dentro o después (secuencias 3' no codificantes) de una región codificante y que influyen en la transcripción, procesamiento o estabilidad del ARN, o la traducción de la región codificante asociada. Las regiones codificantes pueden incluir promotores, secuencias líderes de traducción, sitio de procesamiento de ARN, sitio de unión de efectores y estructura tallo-bucle. Los límites de la región codificante se determinan por un codón de iniciación en el extremo 5' (amino) y un codón de parada en el extremo 3' (carboxilo). Una región codificante puede incluir, pero no está limitada a, regiones procariotas, ADNc de ARNm, moléculas de ADN genómico, moléculas de ADN sintético o moléculas de ARN. Si se pretende la región codificante para expresión en una célula eucariota, normalmente una señal de poliadenilación y secuencia de terminación de la transcripción estará localizada en 3' de la región codificante.

“Marco abierto de lectura” se abrevia ORF y significa una longitud de ácido nucleico, sea ADN, ADNc o ARN, que comprende una señal de inicio de la traducción o codón de iniciación, tal como un ATG o AUG, y un codón de terminación y se puede traducir potencialmente en una secuencia polipeptídica.

Una región codificante está “bajo el control” de elementos de control transcripcional y traduccional en una célula cuando la ARN polimerasa transcribe la región codificante a ARNm, que después se somete a ajuste en trans del ARN (si la región codificante contiene intrones) y se traduce a la proteína codificada por la región codificante.

“Regiones de control transcripcional y traduccional” son regiones reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una región codificante en una célula huésped. En células eucariotas, las señales de poliadenilación son regiones de control.

Los términos “operativamente asociado” y “operativamente unido” se refieren a la asociación de dos moléculas de modo que la función de una está afectada por la otra. Por ejemplo, un promotor está operativamente asociado con una región codificante cuando es capaz de afectar la expresión de esa región codificante (es decir, que la región codificante está bajo el control transcripcional del promotor). Las regiones codificantes pueden estar operativamente asociadas a regiones reguladoras en orientación sentido o antisentido. Dos moléculas están “operativamente unidas” si están unidas directa (por ejemplo, una proteína de fusión) o indirectamente (por ejemplo, a través de un enlazador).

Como se usa en el presente documento, el término “expresión” se refiere a la transcripción de ARN (por ejemplo, ARNm) a partir de un molde de ácido nucleico y/o la traducción de ARNm a un polipéptido. El término “expresión aumentada” se pretende que incluya una alteración en la expresión génica a nivel de producción aumentada de ARNm y/o a nivel de expresión del polipéptido, que en general produce una cantidad aumentada de un producto génico o proteína. En algunos casos, “expresión aumentada” se usa de forma intercambiable con el término “sobrexpresión” o “sobrexpresado”.

Un “marcador de selección” es un gen, cuya expresión crea un fenotipo detectable y que facilita la detección de las células huésped que contienen un plásmido que tiene el marcador de selección. Ejemplos no limitantes de marcadores de selección incluyen genes de resistencia a fármacos y marcadores nutricionales. Por ejemplo, el marcador de selección puede ser un gen que confiere resistencia a un antibiótico seleccionado del grupo que consiste en: ampicilina, kanamicina, eritromicina, cloranfenicol, gentamicina, kasugamicina, rifampicina, espectinomicina, D-cicloserina, ácido nalidíxico, estreptomycinina o tetraciclina. Otros ejemplos no limitantes de marcadores de selección incluyen adenosina desaminasa, aminoglucósido fosfotransferasa, dihidrofolato reductasa, higromicina-B-fosfotransferasa, timidina quinasa y xantina-guanina fosforribosiltransferasa. Un único plásmido puede comprender uno o más marcadores de selección.

Como se usan en el presente documento, los términos “tratar” o “tratamiento” se refieren tanto al tratamiento terapéutico como a medidas profilácticas o preventivas, en donde el objeto es prevenir o ralentizar (reducir) un cambio fisiológico o trastorno no deseado, tal como el desarrollo o extensión del cáncer. Los resultados clínicos beneficiosos o deseados incluyen, pero no están limitados a, mejora de síntomas, disminución de la extensión de la enfermedad, enfermedad estabilizada (es decir, sin empeoramiento), retraso o disminución de la evolución de la enfermedad, mejora o paliación de la enfermedad, y remisión (parcial o total), ya sean detectables o indetectables. “Tratamiento” también puede significar prolongar la supervivencia comparada con la supervivencia esperada si no se

recibe tratamiento. Esos en necesidad de tratamiento incluyen los que ya tienen la afección o trastorno, así como los propensos a tener la afección o trastorno o esos en los que se debe prevenir la afección o trastorno.

5 Mediante "sujeto" o "individuo" o "animal" o "paciente" o "mamífero" se quiere decir cualquier sujeto, en particular un sujeto mamífero, para el que se desea diagnóstico, pronóstico o terapia. Los sujetos mamíferos incluyen seres humanos, animales domésticos, animales de granja y animales de zoo, deporte o mascotas tales como perros, gatos, cobayas, conejos, ratas, ratones, caballos, ganado, vacas, etcétera.

10 II. Secuencias inductoras de cuerpos proteicos (PBIS) recombinantes

10 Las secuencias inductoras de cuerpos proteicos (PBIS) son secuencias polipeptídicas que son capaces de mediar la formación de cuerpos proteicos o RPBLA, que se describen en más detalle posteriormente, en una célula diferente de tejido de endospermo. La capacidad de una PBIS candidata para mediar la formación de RPBLA se puede determinar usando técnicas disponibles para el experto en la materia tales como: (i) detección de agregados fluorescentes mediante expresión de una proteína de fusión de la PBIS candidata y una proteína fluorescente en 15 células vegetales como se describe en el ejemplo 1 de la presente invención; (ii) detección de agregados que tienen una densidad de aproximadamente 1,175 hasta aproximadamente 1,194 mediante fraccionamiento de extractos vegetales que expresan la PBIS candidata en un gradiente de densidad como se describe en el ejemplo 2 de la presente invención; (iii) identificación de la PBIS candidata mediante inmunotransferencia en el sedimento después de centrifugación a baja velocidad (típicamente a 10.000 x g durante 30 minutos a 4°C) como se describe en el 20 ejemplo 3 de la presente invención.

25 Se ha identificado una PBIS natural en la proteína de maíz gamma zeína y se describe en más detalle en la patente de EE UU No. 7.575.898, solicitud publicada de EE UU No. 2006/0121573, solicitud publicada de EE UU No. 2006/0123509 y solicitud publicada de EE UU No. 2007/0243198.

30 Sorprendentemente, se han identificado PBIS que median la formación de ensamblajes recombinantes similares a cuerpos proteicos (RPBLA) con propiedades mejoradas y se describen en el presente documento. En algunas formas de realización, la PBIS recombinante comprende una estructura poliprolina II (PPII). En algunas formas de realización, la PBIS recombinante comprende una secuencia rica en prolina.

En un primer aspecto, la invención se refiere a una secuencia polipeptídica inductora de cuerpos proteicos (PBIS) recombinante que comprende

35 (i) una estructura poliprolina II (PPII) que tiene al menos 36 aminoácidos de longitud; en donde la estructura PPII se localiza entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C; en donde no más del 10% de los aminoácidos en la estructura PPII son lisina o arginina; y 40 en donde la estructura PPII es una secuencia no anfipática y comprende al menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO:17), o

45 (ii) una secuencia rica en prolina que tiene al menos 36 aminoácidos de longitud; en donde la secuencia rica en prolina se localiza entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C; en donde no más del 10% de los aminoácidos en la secuencia rica en prolina son lisina o arginina; y 50 en donde la secuencia rica en prolina es una secuencia no anfipática y comprende al menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO:17).

En una divulgación preferida, la estructura PPII no consiste en la secuencia P4 como se describe en el documento WO2004003207 y que tiene la secuencia QPPPPVHLPPPPCHYPTQPPRPQHPQPHP (SEQ ID NO:130).

55 En una divulgación preferida, la estructura PPII no contiene la secuencia (PPPVHL)₆PPPVHVPPPVHL (SEQ ID NO: 2) o no contiene la secuencia (PPPVHL)₆PPPVHVPPPVHLPPPPCH (SEQ ID NO: 3).

En una forma de realización preferida, la estructura PPII tiene al menos 37, al menos 38, al menos 39 o al menos 40 aminoácidos. En otra forma de realización, la estructura PPII no contiene residuos de cisteína. En aún otra forma de 60 realización, la estructura PPII no contiene más del 10% de residuos de histidina.

En una divulgación preferida, la secuencia rica en prolina no consiste en la PBIS de la secuencia P4 como se describe en el documento WO2004003207 y que tiene la secuencia QPPPPVHLPPPPCHYPTQPPRPQHPQPHP (SEQ ID NO:130).

En una divulgación preferida, la secuencia rica en prolina no contiene la secuencia (PPPVHL)₆PPPVHVPPPVL (SEQ ID NO: 2) o no contiene la secuencia (PPPVHL)₆PPPVHVPPPVLPPPCH (SEQ ID NO: 3).

5 En una forma de realización preferida, región rica en prolina tiene al menos 37, al menos 38, al menos 39 o al menos 40 aminoácidos. En otra forma de realización, la región rica en prolina no contiene residuos de cisteína. En aún otra forma de realización, la región rica en prolina no contiene más del 10% de residuos de histidina o no consiste en la PBIS de la secuencia P4 como se describe en el documento WO2004003207 y que tiene la secuencia QPPPVHLPPPCHYPTQPPRPQHPQPHP (SEQ ID NO:130).

10 En algunas formas de realización, la PBIS recombinante comprende una estructura PPII entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C y comprende además una cisteína adicional y una secuencia rica en prolina entre la estructura PPII y las dos cisteínas C-terminales.

15 En algunas formas de realización, la PBIS recombinante comprende una primera secuencia rica en prolina entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C y comprende además una cisteína adicional y una segunda secuencia rica en prolina entre la primera secuencia rica en prolina y las dos cisteínas C-terminales.

20 No más de aproximadamente el 10% de los aminoácidos en la estructura PPII o en la secuencia rica en prolina de la PBIS de la invención son lisina o arginina. En algunas formas de realización, no más de aproximadamente el 9%, aproximadamente el 8%, aproximadamente el 7%, aproximadamente el 6%, aproximadamente el 5%, aproximadamente el 4%, aproximadamente el 3%, aproximadamente el 2% o aproximadamente el 1% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son lisina o arginina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene lisina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene arginina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene lisina ni arginina.

30 En algunas divulgaciones, no más de aproximadamente el 15% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son histidina. En algunas divulgaciones, no más de aproximadamente el 14%, aproximadamente el 13%, aproximadamente el 12%, aproximadamente el 11%, aproximadamente el 10%, aproximadamente el 9%, aproximadamente el 8%, aproximadamente el 7%, aproximadamente el 6%, aproximadamente el 5%, aproximadamente el 4%, aproximadamente el 3%, aproximadamente el 2% o aproximadamente el 1% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son histidina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene histidina.

35 En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene la secuencia (PPPVHL)₆PPPVHVPPPVL (SEQ ID NO: 2). En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene la secuencia (PPPVHL)₆PPPVHVPPPVLPLCH (SEQ ID NO: 3). En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina no contiene la secuencia PPPVHL (SEQ ID NO: 4).

40 En algunas formas de realización, la PBIS recombinante es capaz de mediar la formación de un RPBLA cuando se expresa a una concentración mínima particular. De esta manera, en algunas formas de realización, la PBIS recombinante es capaz de mediar la formación de un RPBLA cuando se expresa a aproximadamente 0,5 gramos/kilogramo de peso fresco de hojas de tabaco.

45 La capacidad de una secuencia recombinante de funcionar como una PBIS se puede probar según métodos descritos en el presente documento u otros métodos conocidos en la técnica.

50 III. Estructuras de poliprolina II y secuencias ricas en prolina

Como se ha descrito anteriormente, la PBIS recombinante usada en el presente documento puede comprender una estructura de poliprolina de tipo II (PPII).

55 El término "poliprolina II" o "PPII" como se usa en el presente documento, se refiere a un tipo de estructura proteica secundaria helicoidal. Las características particulares de estructura PPII ejemplares se han descrito, por ejemplo, por Eisenberg *et al.*, *J. Mol. Biol.* 179: 125-142 (1984), Bicudo *et al* (2008, *Biopolymers*, 89:175-178) Fernández-Carneado, *J. Mol. Biol.* 372: 708-22 (2004); Bochicchio and Tamburro, *Chirality* 14:782-92 (2002); Knighton *et al. Science* 253: 414-420 (1991); y Caldwell *et al. Biopolymers* 10:1891-1904 (1984).

60 Una hélice PPII es una estructura proteica secundaria. Es una estructura helicoidal a la izquierda con una forma global que recuerda un prisma triangular. Puede estar bastante extendida con un paso helicoidal de 9,3 Å/giro, 3 residuos por giro y los ángulos Φ y Ψ centrados aproximadamente a -75° y 145° , respectivamente.

65 La estructura PPII se ha descrito en la bibliografía (véase, por ejemplo, Eisenberg D, *et al. J. Mol. Biol.* 179:125-142 (1984). Los residuos de prolina están muy favorecidos en las hélices PPII. Glicina y tirosina en general están

desfavorecidas, pero se encuentran en algunas estructuras PPII tal como colágeno. El análisis conformacional de péptidos sintéticos de secuencias repetitivas donde la histidina de gamma zeína se sustituyó por Ala, Glu y Lys indica que todos estos péptidos adoptaban una estructura de tipo PPII. Por tanto, la conformación PPII se adopta independientemente tanto del signo de la carga en el aminoácido cargado (Lys o Glu) como de que el residuo esté cargado o sin cargar (Ala). Dalcol, *J. Org. Chem* 61:6775-6782 (1996). La longitud de la secuencia repetitiva, pH ácido y altas concentraciones de péptido aumentan el contenido en PPII. A pH 3,0 hay más contenido en PPII que a pH 7,0, excepto cuando la histidina se sustituyó por ácido glutámico. En este caso, la desestabilización de PPII probablemente es debida a la protonación de los grupos carboxilo del ácido glutámico a pH 3 y las posteriores interacciones cadena lateral-cadena lateral por puentes de hidrógeno (Dalcol, 1996).

La estructura PPII es un rasgo dinámico de una proteína. Bicudo y col (2008, *Biopolimers*, 89:175-178) analizaron la estructura secundaria de gamma zeína purificada de PB de maíz mediante dicroísmo circular cuando se solubilizaba en agua, SDS y 2-propanol, con y sin agente reductor. La conformación PPII era sólo el 1% en SDS, el 4% en propanol y aproximadamente el 7% en agua. Considerando que el RD representa el 22% de la secuencia total de gamma zeína, estos resultados indican que al menos el 30% de este dominio adoptará una estructura PPII en agua. Fernández-Carneado, *J. Mol. Biol.* 372: 708-22 (2004); Bicudo, 2007. Además, el nivel de PPII puede aumentar mucho mediante las interacciones zeína-zeína que producen el ensamblaje de proteínas y la formación de PB. Por tanto, se considera que los polipéptidos tienen estructura PPII si tienen propensión a formar la estructura PPII.

Los métodos para determinar si una secuencia forma una estructura PPII también son conocidos. Por ejemplo, se pueden usar espectroscopías basadas en actividad óptica, tales como dicroísmo circular (DC), dicroísmo circular vibratorio (DCV) y actividad óptica de Raman (AOR). Véase, por ejemplo, Bochicchio y Tamburro, *Chirality* 14:782-92 (2002). Además, la estructura PPII en proteínas cristalizadas se puede determinar mediante difracción de rayos X Knighton et al. *Science*; 253:414-420 (1991); Caldwell et al. *Biopolymers*; 10:1891-1904 (1984). El DC es muy sensible a la estructura secundaria. Por tanto, en algunas divulgaciones, el método de determinar la estructura PPII es DC. Los espectros de DC de un péptido de aproximadamente 30 a 100 aminoácidos de longitud se puede determinar a aproximadamente pH 7 y a aproximadamente 5°C. La presencia de estructura PPII se puede caracterizar por un patrón de DC con un mínimo a aproximadamente $\lambda=202$ nm y un máximo a aproximadamente $\lambda=228$ nm. El porcentaje de estructura PPII en un péptido muestra se determinará por la proporción de las magnitudes $[\theta_{max}]_M$ (la elipticidad molar en el máximo aproximadamente $\lambda=228$ nm) del péptido muestra comparado con un péptido de referencia H-(Pro) $_n$ -OH. Se puede considerar que el péptido de referencia, que tiene una longitud similar al péptido muestra, tiene el 100% de estructura PPII. Se encuentra una descripción más detallada del método basado en el espectro de DC en Dalcol et al. *Org. Chem.* 61: 6675-6782 (1996).

Por ejemplo, un péptido poliprolina (H-(Pro) $_n$ -OH) se puede considerar como la referencia, y se puede considerar que forma el 100% de hélice PPII. En algunas formas de realización, la $[\theta_{max}]_M$ de la estructura PPII es al menos aproximadamente el 25%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 55%, al menos aproximadamente el 60%, al menos aproximadamente el 65%, al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, o el 100% de la $[\theta_{max}]_M$ del polipéptido de referencia (H-(Pro) $_n$ -OH).

Como se ha descrito anteriormente, la PBIS recombinante puede comprender una secuencia rica en prolina. En algunas divulgaciones, la estructura PPII comprende una secuencia rica en prolina. En algunas divulgaciones, la PBIS recombinante comprende una estructura PPII y una secuencia rica en prolina que no tiene propensión a formar una hélice PPII. En algunas divulgaciones, una secuencia rica en prolina es capaz de formar una estructura PPII.

El término "región rica en prolina", como se usa en el presente documento, se refiere a una secuencia polipeptídica que contiene una proporción relativamente alta de prolina. En algunas divulgaciones, la región rica en prolina comprende al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 100% de prolina.

En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina tiene al menos 30 aminoácidos de longitud. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina tiene al menos 38, al menos 40, al menos 42, al menos 44, al menos 46 o al menos 48 aminoácidos de longitud.

En algunas formas de realización, la estructura PPII o la secuencia rica en prolina tiene no más de 100, no más de 98, no más de 96, no más de 94, no más de 92, no más de 90, no más de 88, no más de 86, no más de 84, no más de 82, no más de 80, no más de 78, no más de 76, no más de 74, no más de 72, no más de 70, no más de 68, no más de 66, no más de 64, no más de 62, no más de 60, no más de 58, no más de 56, no más de 54 o no más de 52 aminoácidos de longitud.

En algunas formas de realización, la estructura PPII o la secuencia rica en prolina tiene desde aproximadamente 36 hasta aproximadamente 100 aminoácidos, desde aproximadamente 36 hasta aproximadamente 90, desde

homoméricas. En algunas divulgaciones, al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14 o al menos 15 de las repeticiones de 6 aminoácidos son repeticiones heteroméricas.

5 En algunas divulgaciones, la estructura PPII o la región rica en prolina son moléculas anfipáticas.

Una molécula anfipática es una molécula que tiene una región con carácter hidrofóbico significativamente mayor comparado con el resto de la molécula, que es hidrofóbica en comparación con esa región. El carácter hidrofóbico o hidrofílico de una molécula o región de la misma tiene un efecto importante en su comportamiento con respecto al medio circundante. Las superficies hidrofóbicas tienden a evitar interacción con medios acuosos, mientras que las superficies hidrofílicas están estabilizadas por los mismos. Puesto que una proteína es una estructura tridimensional, solo algunos de sus residuos de aminoácidos están expuestos al medio. Como consecuencia, solo los aminoácidos localizados en la superficie, por tanto, expuestos al medio, se deben considerar al determinar si una proteína o región proteica es de carácter hidrofóbica, hidrofílica o anfipática.

15 Como se ha mencionado en el presente documento, la estructura PPII es una hélice a la izquierda extendida con tres aminoácidos por giro. Cada giro individual se puede representar como un triángulo (por ejemplo, figura 1A) con cada uno de los 3 aminoácidos en un giro colocados en uno de los vértices del triángulo en un orden en sentido de las agujas del reloj. Como resultado, el apilado de triángulos consecutivos (giros) de la estructura PPII generará un prisma triangular que representa la estructura entera de la proteína o región proteica. En una hélice PPII, los aminoácidos se están colocados en los bordes del prisma triangular están expuestos al medio y todos se deben considerar en determinaciones hidrofóbico/hidrofílico. Una estructura PPII se puede considerar anfipática cuando uno de los bordes del prisma triangular tiene un carácter hidrofóbico significativamente diferente con respecto a los otros dos bordes considerados independientemente.

25 El carácter hidrofóbico/hidrofílico de los aminoácidos se determina por la naturaleza de su cadena lateral. La escala de hidrofobicidad de Kyte-Doolittle (Kyte J., Doolittle R.F., J. Mol. Biol. 157:105-132(1982)) se deriva de las propiedades fisicoquímicas de las cadenas laterales de los aminoácidos, y es comúnmente usada y bien aceptada. Los aminoácidos se clasificarán en tres grupos basados en valores de Kyte-Doolittle: (i) aminoácidos que varían de 4,5 a 1,8 que se considerarán como no polares o hidrofóbicos (I, V, L, F, C, M y A), (ii) aminoácidos que varían de -0,4 a -1,6 que se considerarán parcialmente polares o parcialmente hidrofílicos (G, S, T, W, Y y P) y aminoácidos que varían de -3,2 a -4,5 que se considerarán como polares o hidrofílicos (R, H, D, E, N, Q y K). Entre los aminoácidos del grupo polar, se pueden encontrar aminoácidos que están positivamente cargados (R, H y K), negativamente cargados (D y E) y no cargados (Q y N). Para simplificar el cálculo de la hidrofobicidad de una hélice PPII, se ha asignado un valor de polaridad consenso considerando esta clasificación: (i) 0 para aminoácidos no polares, (ii) 0,5 para aminoácidos parcialmente polares, y (iii) 1 para aminoácidos polares.

40 El carácter anfipático de una estructura PPII se puede calcular considerando la clasificación de aminoácidos descrita en el presente documento en el contexto de la distribución espacial de aminoácidos en el prisma triangular. El porcentaje de aminoácidos hidrofílicos se puede calcular para cada borde (borde 1: i, i+4, i+7 ...; borde 2: i+1, i+5, i+8 ...; borde 3: i+2, i+6, i+9 ...), y la estructura PPII se puede considerar anfipática cuando la diferencia del porcentaje de uno de los bordes es al menos aproximadamente 35 con respecto a los otros dos bordes considerados independientemente.

45 Por tanto, puesto que cada tercer residuo en la estructura PPII está alineado a lo largo de un lado de la hélice, si la hidrofobicidad de los residuos a lo largo de un lado de la hélice es diferente de la hidrofobicidad de los residuos a lo largo de otro lado de la hélice, la hélice PPII será anfipática. La hélice PPII del RD de gamma zeína tiene un carácter anfipático marcado. Con 3,0 residuos por giro, los residuos de valina y leucina de gamma zeína se alinean en el mismo lado de la hélice (borde 1 y 2, respectivamente) mientras que los residuos polares de histidina, que están cargados se alinean en el lado opuesto (borde 3). Esta anfipaticidad está claramente señalada por el cálculo mencionado justo antes, donde el porcentaje de aminoácidos hidrofílicos es 29,4, 29,4 y 78,1 en el borde 1, 2, y 3, respectivamente (véase la tabla 1).

50

Tabla 1: Valores de anfipaticidad de las PBIS según la invención o divulgadas en el presente documento

grupos PPII	PBIS	residuos*	% de prolina	% de identidad**	Polaridad de los bordes***	Diferencia en polaridad en los bordes****	carga neta
Anfipáticos positivamente cargados	RX3	11-63	54,7	100	(29,4/29,4/78,1)	(0/48,7/48,7)	+8
	RX3(K)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	+8
	RX3(R)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	+8
Anfipáticos negativamente cargados	RX3(D)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	-8
	RX3(E)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	-8
	RX3(N)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	0
Anfipáticos no cargados	RX3(Q)	11-63	54,7	84,9	(29,4/29,4/78,1)	(0/48,7/48,7)	0
	RX3(A)	11-63	54,7	86,8	(29,4/29,4/34,4)	(0/5/5)	+1
No anfipáticos	RX3(A)2	11-63	54,7	84,9	(29,4/29,4/28,1)	(0/1,3/1,3)	0
	RX3(A3)	11-63	54,7	54,7	(29,4/29,4/28,1)	(0/1,3/1,3)	0
	RX3(L)	11-63	54,7	84,9	(29,4/29,4/28,1)	(0/1,3/1,3)	0
	RX3(V)	11-63	54,7	84,9	(29,4/29,4/28,1)	(0/1,3/1,3)	0
	RX3(T)	11-63	54,7	84,9	(29,4/29,4/53,1)	(0/23,7/23,7)	0
	PP	11-63	96,2	58,5	(50/52,9/56,3)	(0/1,3/1,3)	+1
	PP3	11-63	100	54,7	(50/50/50)	(0/0/0)	0
	PA	11-63	66	58,5	(29,4/29,4/56,7)	(0/27,3/27,3)	+1
	PA2	11-63	67,9	54,7	(29,4/29,4/53,1)	(0/23,7/23,7)	0
	Z(Col)	11-63	39,6	38,6	(52,9/52,9/56,3)	(0/3,4/3,4)	+1
	Z(Col2)	11-56	36,9	27,8	(50/50/50)	(0/0/0)	0
	Z(Adh)	11-63	39,6	35,8	(64,7/61,8/50)	(2,9/14,7/11,8)	-5
Z(Adh2)	11-56	39,6	27,8	(68,8/65,6/46,7)	(4,4/22,1/18,9)	-6	

* Residuos usados en el análisis (posición relativa con respecto al extremo N de la proteína madura)

** % de aminoácidos idénticos tomando RX3 como referencia

*** Porcentaje de aminoácidos polares en el borde 1, 2 y 3 de la representación de prisma triangular de la estructura PPII

**** Diferencias en el porcentaje de aminoácidos polares entre los bordes del prisma triangular de la representación de prisma triangular de la estructura PPII (borde 1 frente a borde 2, borde 1 frente a borde 3 y borde 2 frente a borde 3)

En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina es anfipática y está cargada negativamente. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina es anfipática y no cargada. La estructura PPII o secuencia rica en prolina es no anfipática.

5 En algunas formas de realización, el porcentaje de prolinas a lo largo de la longitud de la estructura PPII o la secuencia rica en prolina es consistente. Por tanto, por ejemplo, en algunas formas de realización, ventanas de 10 aminoácidos a través de la estructura PPII o secuencia rica en prolina se diferencian en el porcentaje en prolina en no más de aproximadamente el 50%, aproximadamente el 45%, aproximadamente el 35%, aproximadamente el 30%, aproximadamente el 25%, aproximadamente el 20%, aproximadamente el 15%, aproximadamente el 10%, o aproximadamente el 5%.

15 En algunas formas de realización, los aminoácidos prolina presentes en la estructura PPII o la secuencia rica en prolina pueden estar hidroxilados dando hidroxiprolina (por ejemplo, (2S,4R)-4-hidroxiprolina o L-hidroxiprolina ($C_5H_9O_3N$)). Esta es una modificación postraduccional común de prolina que se diferencia solo por la presencia de un grupo hidroxilo (OH) unido al átomo de carbono gamma de prolina. Hidroxiprolina está presente en secuencias ricas en prolina tal como colágeno. Por ejemplo, en la triada canónica GXY (donde X e Y son independientemente cualquier aminoácido), una prolina que ocupa la posición Yaa puede estar hidroxilada. Las hidroxiprolinas no interfieren con la formación de hélice PPII.

20 En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina comprende una secuencia relacionada con colágeno. Por tanto, en algunas divulgaciones, la estructura PPII o la secuencia rica en prolina comprende la secuencia $(GXY)_n$, en donde n es al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 12, al menos 14, al menos 16, al menos 18 o al menos 20. En algunas divulgaciones, la secuencia que comprende $(GXY)_n$ comprende al menos aproximadamente el 30% de prolinas.

25 La estructura PPII o la secuencia rica en prolina tiene al menos 36 aminoácidos de longitud. La PBIS recombinante comprende al menos dos cisteínas en el extremo amino (extremo N) de la estructura PPII o la secuencia rica en prolina (es decir, dos cisteínas antes de la estructura PPII o la región rica en prolina). La PBIS recombinante comprende al menos dos cisteínas en el extremo carboxi (extremo C) de la estructura PPII o la secuencia rica en prolina (es decir, dos cisteínas después de la estructura PPII o la región rica en prolina).

30 En algunas formas de realización, al menos aproximadamente el 95%, aproximadamente el 90%, aproximadamente el 85%, aproximadamente el 80%, aproximadamente el 75%, aproximadamente el 70%, aproximadamente el 65%, aproximadamente el 60%, aproximadamente el 55% o aproximadamente el 50% de PBIS muestra estructura de poliprolina II.

35 Como se ha descrito anteriormente, la PBIS recombinante puede comprender una secuencia rica en prolina. En algunas formas de realización, la estructura PPII comprende una secuencia rica en prolina. En algunas formas de realización, la PBIS recombinante comprende una estructura PPII y una secuencia rica en prolina que no es propensa a formar una hélice PPII.

40 Puesto que cada tercer residuo en la estructura PPII se alinea en un lado de la hélice, si la hidrofobicidad de los residuos en un lado de la hélice es diferente de la hidrofobicidad de los residuos en el otro lado de la hélice, la hélice PPII será anfipática. La hélice PPII del RD de gamma zeína tiene un marcado carácter anfipático. Con 3 residuos por giro, los residuos de valina y leucina de la gamma zeína se alinean en el mismo lado de la hélice mientras que los residuos polares de histidina, que están cargados dependiendo del pH, se alinean en el lado opuesto.

45 En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina comprende una repetición rica en prolina. Como se ha descrito, una repetición rica en prolina es una secuencia que comprende al menos dos copias de una secuencia que comprende prolina. La repetición puede tener al menos dos aminoácidos de longitud (por ejemplo, PX), al menos tres aminoácidos de longitud (por ejemplo, PPX, PXP, XPP o PXX), al menos cuatro aminoácidos de longitud (por ejemplo, PPPX (SEQ ID NO:5), PPXX (SEQ ID NO:6), PXXX (SEQ ID NO:7), PPXP (SEQ ID NO:8), PXPP (SEQ ID NO:65)), al menos cinco aminoácidos de longitud (por ejemplo, PPPXX (SEQ ID NO:9)), al menos seis aminoácidos de longitud (por ejemplo, PPPXXX (SEQ ID NO:10) o PPPXPX (SEQ ID NO:11)), al menos siete aminoácidos de longitud, al menos ocho aminoácidos de longitud, al menos nueve aminoácidos de longitud o al menos diez aminoácidos de longitud. Las repeticiones enumeradas aquí se proporcionan sólo a modo de ejemplo.

50 La repetición rica en prolina puede comprender al menos dos copias, al menos tres copias, al menos cuatro copias, al menos cinco copias, al menos seis copias, al menos siete copias, al menos ocho copias, al menos nueve copias o al menos diez copias de una secuencia que contiene prolina. La repetición rica en prolina puede contener todas las copias de la misma repetición (es decir, una repetición rica en prolina homomérica) o puede contener una combinación de repeticiones ricas en prolina (es decir, una repetición rica en prolina heteromérica). A modo de ejemplo, la secuencia PPPAAAPPPAAAPPPAAA (SEQ ID NO:12) es una repetición rica en prolina homomérica que contiene tres copias de la misma repetición PPPAAA (SEQ ID NO:13) y la secuencia PPPAAAPPPAAAPPAPPPPPAPPP (SEQ ID NO: 14) es una repetición rica en prolina heteromérica que contiene

dos copias de una secuencia, PPPAAA (SEQ ID NO:13) y dos copias de una secuencia diferente, PPAPPP (SEQ ID NO:15).

5 Algunas estructuras PPII o secuencias ricas en prolina de la invención consisten esencialmente en aminoácidos seleccionados del grupo que consiste en (i) prolina, (ii) valina, (iii) leucina y (iv) alanina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina consisten esencialmente en aminoácidos seleccionados del grupo que consiste en (i) prolina, (ii) valina, (iii) leucina, (iv) ácido aspártico y (v) glutamina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina consisten esencialmente en aminoácidos seleccionados del grupo que consiste en (i) prolina, (ii) valina, (iii) leucina, (iv) treonina, (v) asparragina y (vi) glutamina. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina consiste esencialmente en aminoácidos seleccionados del grupo que consiste en (i) prolina, (ii) aminoácidos cargados negativamente (es decir, D y E), (iii) aminoácidos con cadenas laterales polares sin carga (es decir, N y Q), (iv) aminoácidos con cadenas laterales no cargadas parcialmente polares (es decir, S, T, W, Y y G) o con aminoácidos con cadenas laterales no cargadas parcialmente polares (es decir, S, T, y G), (v) aminoácidos con cadenas laterales hidrofóbicas (es decir, A, I, L, M, F, y V) o con aminoácidos con cadenas laterales hidrofóbicas (es decir, A, I, L, M y V). En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina consiste esencialmente en aminoácidos seleccionados del grupo que consiste en (i) prolina y (ii) alanina.

20 En algunas formas de realización, al menos aproximadamente el 40% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son prolina. En algunas formas de realización, al menos aproximadamente el 45%, al menos aproximadamente el 50%, al menos aproximadamente el 55%, al menos aproximadamente el 60%, al menos aproximadamente el 65%, al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o el 100% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son prolina.

30 En algunas divulgaciones, no más de aproximadamente el 95%, aproximadamente el 90%, aproximadamente el 85%, aproximadamente el 80%, aproximadamente el 75%, aproximadamente el 70%, aproximadamente el 65%, aproximadamente el 60%, aproximadamente el 55% o aproximadamente el 50% de los aminoácidos en la estructura PPII o la secuencia rica en prolina son prolina.

35 La estructura PPII o la secuencia rica en prolina es no anfipática. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina es anfipática y negativamente cargada. En algunas divulgaciones, la estructura PPII o la secuencia rica en prolina es anfipática y no cargada.

40 En una divulgación, la estructura PPII recombinante o la región rica en prolina comprende una secuencia anfipática negativamente cargada. El término "secuencia anfipática negativamente cargada" se ha explicado anteriormente y se usa para referirse a estructuras PPII o regiones ricas en prolina que muestran una carga neta negativa calculada usando el método definido anteriormente. En una divulgación aún más preferida, la secuencia anfipática negativamente cargada comprende al menos una repetición que se selecciona independientemente del grupo que consiste en PPPVDL (SEQ ID NO: 19) y PPPVEL (SEQ ID NO: 18). En una divulgación preferida, la secuencia anfipática comprende una secuencia seleccionada de $(PPPVEL)_n$, (SEQ ID NO:18) y $(PPPVDL)_n$, (SEQ ID NO:19) en donde n es 4, 5, 6, 7 o 8.

45 En aún otra divulgación, la secuencia anfipática cargada negativamente consiste esencialmente de la secuencia definida a continuación:

50 PPPVELPPPP VELPPPVELP PPVELPPPVE LPPPVELPPP VEVPPPVELPPPP (SEQ ID NO:30),
 PPPVELPPP VELPPPVELP PPVELPPPVE LPPPVELPPP VELPPPVELPPPP (SEQ ID NO:123),
 PPPVDLPPP VDLPPPVDLP PPVDLPPPVD LPPPVDLPPP VDVPPPVDLPPPP (SEQ ID NO:124), y
 PPPVDLPPP VDLPPPVDLP PPVDLPPPVD LPPPVDLPPP VDLPPPVDLPPPP (SEQ ID NO:31).

55 En otra divulgación, la estructura PPII o región rica en prolina que forma parte de la PBIS según la invención es una secuencia anfipática no cargada. El término "secuencia anfipática no cargada" se refiere a una secuencia que tiene una carga neta de 0 y en donde los residuos en el borde 2 de la representación en prisma triangular de la estructura PPII se seleccionan del grupo que consiste en Asn y Gln. En una divulgación aún más preferida, la secuencia anfipática no cargada comprende la menos una repetición que se selecciona independientemente del grupo PPPVNL (SEQ ID NO:21) y PPPVQL (SEQ ID NO:22). En una divulgación aún más preferida, la secuencia anfipática no cargada comprende seis repeticiones en donde cada repetición se selecciona independientemente del grupo de PPPVNL (SEQ ID NO:21) y PPPVQL (SEQ ID NO:22). En una divulgación preferida, la PBIS comprende una secuencia seleccionada de $(PPPVNL)_n$, (SEQ ID NO:21) y $(PPPVQL)_n$, (SEQ ID NO:22) en donde n es 4, 5, 6, 7 o 8.

En una divulgación preferida, la secuencia anfipática no cargada que forma parte de la estructura PPII o la región rica en prolina que forma parte de la PBIS se selecciona independientemente de una secuencia como se define a continuación:

- 5 PPPVNLPPP VNLPPPVNLP PPVNLPPPVN LPPPVNLP VNVPPPVNLP (SEQ ID NO:32),
 PPPVNLPPP VNLPPPVNLP PPVNLPPPVN LPPPVNLP VNLPPPVNLP (SEQ ID NO:33),
 PPPVQLPPP VQLPPPVQLP PPVQLPPP VQVPPPVQLPPP (SEQ ID NO:34), y
 PPPVQLPPP VQLPPPVQLP PPVQLPPP LPPPVQLPPP VQLPPPVQLPPP (SEQ ID NO:35).
- 10 La estructura PPII o la región rica en prolina que forma parte de la PBIS según la invención es una secuencia no anfipática. El término "secuencia no anfipática" se refiere a una secuencia que tiene una carga neta de 0 y en donde los residuos en el borde 2 de la representación en prisma triangular de la estructura PPII se seleccionan independientemente del grupo que consiste en Ala, Leu, Val, Thr, y Ser. La secuencia no anfipática comprende al menos una repetición que se selecciona del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO:17). En una divulgación preferida, la PBIS comprende una secuencia seleccionada de (PPPVAL)_n, (SEQ ID NO:16), (PPPVTL)_n, (SEQ ID NO:20), (PPPVSL)_n, (SEQ ID NO:36), (PPPVVL)_n, (SEQ ID NO:37) y (PPPVLL)_n, (SEQ ID NO:17) en donde n es 4, 5, 6, 7 o 8.

- 20 En una forma de realización aún más preferida, la secuencia no anfipática se selecciona del grupo que consiste en:
- PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VAVPPPVALPPPP (SEQ ID NO:38),
 PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VALPPPVALPPPP (SEQ ID NO:39),
 PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VAVPPPVLPPPP (SEQ ID NO:40),
 25 PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VALPPPVLPPPP (SEQ ID NO:41),
 PPPVLLPPP VLLPPPVLPP PPVLLPPPVL LPPPVLPPPP VLVPPPVLPPPP (SEQ ID NO:42),
 PPPVLLPPP VLLPPPVLPP PPVLLPPPVL LPPPVLPPPP VLLPPPVLPPPP (SEQ ID NO:43),
 PPPVLLPPP VLVPPPVLPP PPVLLPPPVL LPPPVLPPPP VVPPPVLPPPP (SEQ ID NO:44),
 PPPVLLPPP VLVPPPVLPP PPVLLPPPVL LPPPVLPPPP VLVPPPVLPPPP (SEQ ID NO:45),
 30 PPPVTLPPP VTLPPPVTLP PPVTLPPPVT LPPPVTLP VLVPPPVTLP (SEQ ID NO:46),
 PPPVTLPPP VTLPPPVTLP PPVTLPPPVT LPPPVTLP VTLPPPVTLP (SEQ ID NO:47),
 PPPVSLPPP VSLPPPVSLP PPVSLPPPVS LPPPVSLPPP VSVPPPVSLPPPP (SEQ ID NO:48) y
 PPPVSLPPP VSLPPPVSLP PPVSLPPPVS LPPPVSLPPP VSLPPPVSLPPPP (SEQ ID NO:49).

- 35 En otra divulgación, la secuencia no anfipática recombinante consiste esencialmente en los aminoácidos seleccionados del grupo que consiste en prolina y alanina. En una divulgación preferida, la secuencia que consiste esencialmente en prolina y alanina comprende al menos una repetición seleccionada del grupo que consiste en PPPAAA (SEQ ID NO: 13), PPPAPA (SEQ ID NO:23) y PPPPPP (SEQ ID NO:24). En una divulgación preferida, la secuencia anfipática comprende una secuencia seleccionada de (PPPAAA)_n, (SEQ ID NO:13), (PPPAPA)_n (SEQ ID NO:23) y (PPPPPP)_n (SEQ ID NO:24) en donde n es 4, 5, 6, 7 o 8.

En aún otra divulgación más preferida, la secuencia que consiste esencialmente en aminoácidos seleccionados del grupo que consiste en prolina y alanina se selecciona del grupo:

- 45 PPPAAAPPP AAAPPPAAAP PPAAPPPAA APPPAAAPPP AAAPPPAAAPPP (SEQ ID NO:50),
 PPPAAAPPP AAAPPPAAAP PPAAPPPAA APPPAAAPPP AAAPPPVHLPPPP (SEQ ID NO:51),
 PPPPPPPPP PPPPPPPPP PPPPPPPPP PPPPPPPPP PPPPPPPPPPPPPPP (SEQ ID NO:52),
 PPPPPPP PPPPPPPPP PPPPPPPPP PPPPPPPPP PPPPPPPVHLPPPP (SEQ ID NO:53),
 PPPAPAPPP APAPPPAPAP PPAPAPPPAP APPPAPAPPP APAPPPVHLPPPP (SEQ ID NO:54), y
 50 PPPAPAPPP APAPPPAPAP PPAPAPPPAP APPPAPAPPP APAPPPAPAPPP (SEQ ID NO:55).

En otra divulgación, la secuencia no anfipática consiste esencialmente en la secuencia (GXY)_n, en donde X e Y son independientemente cualquier aminoácido y n es un número entero de 1 a 20 y en donde la secuencia que comprende (GXY)_n comprende al menos el 30% de prolina. En aún otra divulgación, la secuencia no anfipática consiste esencialmente en aminoácidos seleccionados del grupo que consiste en aminoácidos que muestran al menos el 70% de identidad respecto a los aminoácidos 135 a 179 del número de registro de GenBank CAA34683. En aún otra divulgación, la secuencia no anfipática consiste esencialmente en las secuencias:

- 60 GPMGPRGPPG PAGAPGPQGF QGNPGEPEGEP GVS GPMGPRG PPGPPVHL (SEQ ID NO:56) y
 GPMGPRGPPG PAGAPGPQGF QGNPGEPEGEP GVS GPMGPRG PPGPP (SEQ ID NO:68)

En otra divulgación, la secuencia anfipática consiste esencialmente en aminoácidos que muestran al menos el 70% de identidad respecto a los aminoácidos 884 a 927 de la adhesina de superficie Agl/II (*Streptococcus mutans* cepa NG8; GenBank: CQ456171). En una divulgación preferida, la secuencia anfipática consiste esencialmente en los aminoácidos seleccionados del grupo que consiste en:

- 65

PPTPPTYETE KPLEPAPVEP SYEAEPTPPT RAPDQAEPNK PTPPTPVHLP PPP (SEQ ID NO:69) y
PPTPPTYETE KPLEPAPVEP SYEAEPTPPT RAPDQAEPNK PTPPTP (SEQ ID NO:70).

- 5 En algunas divulgaciones, la PPII/secuencia rica en Pro contiene cuatro residuos de prolina en el extremo C. Esta secuencia PPPP se encuentra antes de las cisteínas C-terminales, es decir, entre el extremo C de la PPII/secuencia rica en Pro y las cisteínas C-terminales.

IV. Residuos de cisteína

- 10 Si no se describe específicamente de otra manera, la posición de los residuos de cisteína son relativas a los extremos N o C de una estructura PPII, una secuencia rica en poliprolina o tanto una estructura PPII como una secuencia rica en prolina.

- 15 Como se ha descrito anteriormente, la PBIS recombinante puede comprender una estructura PPII, una secuencia rica en poliprolina o tanto una estructura PPII como una secuencia rica en prolina entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C. En algunas formas de realización, la PBIS recombinante comprende al menos tres cisteínas o al menos cuatro cisteínas en el extremo C.

- 20 En algunas formas de realización, la PBIS recombinante comprende múltiples estructuras PPII y/o secuencias ricas en prolina separadas por una cisteína adicional. A modo de ejemplo sólo, la PBIS recombinante puede comprender desde el extremo N al extremo C: dos residuos de cisteínas, una estructura PPII, un residuo de cisteína, una secuencia rica en prolina y dos residuos de cisteína. En otro ejemplo, la PBIS recombinante puede comprender, desde el extremo N al extremo C: dos residuos de cisteína, una primera secuencia rica en prolina, un residuo de
25 cisteína, una segunda secuencia rica en prolina y dos residuos de cisteína.

- En algunas formas de realización, el número de aminoácidos entre las dos cisteínas en el extremo C es un número entero de 0 a 20 y/o el número de aminoácidos entre las dos cisteínas en el extremo N es un número entero desde 0 a 20. Preferiblemente, las dos cisteínas en el extremo N pueden estar separadas por al menos 20, 19, 18, 17, 16,
30 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, o 1 aminoácido o pueden estar adyacentes. Preferiblemente, las dos cisteínas en el extremo C pueden estar separadas por al menos 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, o 1 aminoácido o pueden estar adyacentes. En una forma de realización preferida, las al menos dos cisteínas en el extremo N están separadas por un aminoácido (es decir, CXC), por dos aminoácidos (es decir, CXXC (SEQ ID NO: 57)).

- 35 En algunas formas de realización, las dos cisteínas en el extremo N de la secuencia PPII o la región rica en prolina y la secuencia PPII o rica en prolina están separadas por aproximadamente de 0 a 40 aminoácidos (medido como el número de aminoácidos entre la más C-terminal de las dos cisteínas y el aminoácido N-terminal de la poliprolina II o la región rica en prolina). En algunas formas de realización, las dos cisteínas en extremo N de la secuencia PPII o la
40 región rica en prolina y la secuencia PPII o rica en prolina están separadas por al menos 40, al menos, 35, al menos 30, al menos 25, al menos 20, al menos 15, al menos 10, al menos 5 aminoácidos.

- 45 En algunas formas de realización, las dos cisteínas en el extremo C de la secuencia PPII o la región rica en prolina y la secuencia PPII o rica en prolina están separadas por aproximadamente de 0 a 40 aminoácidos (medido como el número de aminoácidos entre la más N-terminal de las dos cisteínas y el aminoácido C-terminal de la poliprolina II o la región rica en prolina). En algunas formas de realización, las dos cisteínas en extremo C de la secuencia PPII o la
50 región rica en prolina y la secuencia PPII o rica en prolina están separadas por al menos 40, al menos, 35, al menos 30, al menos 25, al menos 20, al menos 15, al menos 10, al menos 5 aminoácidos.

- 55 En algunas formas de realización, las al menos dos cisteínas en el extremo N no están en un dominio globular. En algunas formas de realización, las al menos dos cisteínas en el extremo C no están en un dominio globular. En algunas formas de realización, la PBIS recombinante no contiene ninguna cisteína en un dominio globular. En una forma de realización preferida, cuando la PBIS se proporciona como una proteína de fusión a una proteína heteróloga, las cisteínas no forman parte de la proteína heteróloga.

V. Dominio ProX

- 60 En otra forma de realización preferida, la PBIS comprende una secuencia como se ha definido en el párrafo previo en donde las cisteínas que flanquean la estructura PPII o el dominio rico en prolina se encuentran en secuencias adicionales conocidas como N1 y N2. Estas secuencias N1 y N2 flanquean PPII o la región rica en prolina. Las secuencias N1 y N2 típicas que se pueden usar en la presente invención incluyen sin limitaciones:

- Las secuencias N1 adecuadas incluyen THTSGGCGCQ (SEQ ID NO:25), THPPPCPCP (SEQ ID NO:26) o
65 QCPSPHPQQCPCPHPQPHPQPRPPQTPYHC (SEQ ID NO:126). En una forma de realización, la secuencia N1 aparece antes de PPII o la región rica en prolina.

La secuencia N2 puede ser una región rica en prolina. El término “región rica en prolina” se ha definido anteriormente como una secuencia que está predominantemente formada por prolinas. En una forma de realización preferida, la región rica en prolina está predominantemente formada por prolinas y tiene prolinas uniformemente distribuidas todo a lo largo de la secuencia. Por tanto, por ejemplo, en algunas formas de realización, cuando se dibujan ventanas de 10 aminoácidos a través de la longitud de la estructura PPII o la secuencia rica en prolina, entonces cada ventana se diferencia en el porcentaje en prolina en no más de aproximadamente el 50%, aproximadamente el 45%, aproximadamente el 35%, aproximadamente el 30%, aproximadamente el 25%, aproximadamente el 20%, aproximadamente el 15%, aproximadamente el 10%, o aproximadamente el 5% con respecto a las ventanas restantes. En aún una forma de realización más preferida, la región rica en prolina es una secuencia que está predominantemente formada por prolinas, en donde las prolinas están uniformemente distribuidas todo a lo largo de la secuencia y en donde el porcentaje de aminoácidos que tienen una cadena lateral voluminosa (Trp, Tyr y/o Phe) no es mayor del 10% con respecto a la longitud total de la secuencia.

15 Las regiones ricas en prolina adecuadas que se pueden usar en la presente invención incluyen el dominio ProX encontrado en γ -zeína HYPYQPPRPQHPQHPHPCPC (SEQ ID NO:71) o una secuencia modificada seleccionada del grupo que consiste en las secuencias:

20 CHYPTQPPRPQHPQHPHPCPCQQPHPS (SEQ ID NO:72),
 CPPPPPPPPPPPPPPPPPCPCPPPPPPPCP (SEQ ID NO:73),
 CPPPPPPPPPPPPPPPPPPPCPPCPPPPPPCPP (SEQ ID NO:74),
 CPPPAPAPPAPAPPAPPCPCPAPAPP (SEQ ID NO:75),
 CHYPPAPAPPAPAPPAPACPCAAPAPSPCQ (SEQ ID NO:76) y
 CHYPTAPARPAPAPAPAPPCPCAAPAPSPCQ (SEQ ID NO:127)

25 La secuencia N2 se puede encontrar después o antes de la región PPII o de la región rica en prolina. En una forma de realización preferida, la secuencia N2 se encuentra después de la región PPII.

Por tanto, en una divulgación preferida, la PBIS comprende una secuencia seleccionada del grupo que consiste en:

30 THTSGGCGCQP(PPPVEL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:107),
 THTSGGCGCQP(PPPVDL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:108),
 THTSGGCGCQP(PPPVNL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:109),
 THTSGGCGCQP(PPPVQL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:111),
 35 THTSGGCGCQP(PPPVAL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:112),
 THTSGGCGCQP(PPPRTL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:113),
 THTSGGCGCQP(PPPVSL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:114),
 THTSGGCGCQP(PPPVLL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:115),
 THTSGGCGCQP(PPPVVL)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:116),
 40 THTSGGCGCQP(PPAAAA)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:117),
 THTSGGCGCQP(PPAAAA)_nPPAPAPPAPAPPAPPCPCPAPAPPCCP (SEQ ID NO:118),
 THTSGGCGCQP(PPPPPP)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:119),
 THTSGGCGCQP(PPPPPP)_nPPPCPPPPPPPPPPPPPPPPPCPCPPPPPPPCP (SEQ ID NO:120),
 THTSGGCGCQP(PPPPA)_nPPPPCHYPTQPPRPQHPQHPHPCPCQQPHPSPCQ (SEQ ID NO:121), y
 45 THTSGGCGCAP(PPAAAA)_nPPPPCHYPPAPAPPAPAPPAPACPCAAPAPSPCQ (SEQ ID NO:122)

en donde n es 4, 5, 6, 7, o 8. En una forma de realización preferida n es 8.

En otra divulgación, la PBIS consiste esencialmente de las secuencias mencionadas en la tabla 2.

SEQ ID NO:	PBIS	PBIS
77	THTSGGCGCQ(PPPVEL) ₆ PPPVEVPPVLPPLPPPCCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(E)
78	THTSGGCGCQ(PPPVDL) ₆ PPPVDVPPVLDLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(D)
79	THTSGGCGCQ(PPPVNL) ₆ PPPVNLPVPPVNLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(N)
80	THTSGGCGCQ(PPPVQL) ₆ PPPVQLVPPVQLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(Q)
81	THTSGGCGCQ(PPPVAL) ₆ PPPVALVPPVALPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(A)
82	THTSGGCGCQ(PPPVTL) ₆ PPPVTVPVTLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(A)2
85	THTSGGCGCQ(PPPVSL) ₆ PPPVSVPPVSLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(T)
86	THTSGGCGCQ(PPPVLL) ₆ PPPVLVPPVLLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(S)
83	THTSGGCGCQ(PPPVLL) ₆ PPPVLVPPVLLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(L)
84	THTSGGCGCQ(PPPVLL) ₆ PPPVLVPPVLLPPPPCHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(V)
87	THTSGGCGCQ(PPPAAA) ₈ PPPPACHYTPQPRPQPHPCQPCQCPHFSPCC	RX3(A3)
88	THTSGGCGCQ(PPPPPP) ₇ PPPHLPPPP ₇ PPPHLPPPPCPPPPPPPPPPPPPPPPPPPPPPCP	PP
125	THPPPCPCP(PPPPPP) ₇ PPPHLPPPP ₇ PPPHLPPPPCPPPPPPPPPPPPPPPPPPPPPPCP	PP2
89	THTSGGCGCQ(PPPPPP) ₈ PPPPCPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPCP	PP3
90	THTSGGCGCQ(PPPAPA) ₇ PPPAHLPPPPCPPPAPAPAPAPAPAPAPAPAPAPAPAP	PA
91	THTSGGCGCQ(PPPAPA) ₈ PPPPCPPPAPAPAPAPAPAPAPAPAPAPAPAPAPAP	PA2
92	THPPPCPCP(PPPPPP) ₇ PPPHLPPPPCPPPPPPPPPPPPPPPPPPPPPPPPPPPPCP	NPP
93	THPPPCPCP(PPPAPA) ₇ PPPAHLPPPPCPPPAPAPAPAPAPAPAPAPAPAPAPAP	NPA
94	THTSGGCGCQPPPTTYETE KLEPAPVEP SYEAETPPT RAPDQAEPNK PTPPTPVHLP PPPTPTTYETEKLEPAPVEPSYEAETPTTRAPDQAEPNKPTPPTPVHLP CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(Adh)Px
95	THTSGGCGCQPPPTTYETE KLEPAPVEP SYEAETPPT RAPDQAEPNK PTPPT CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(Adh2)Px
96	THTSGGCGCQPGMPRGPAGAPGQGFQGNPGEPEGVSGPMGPRGPPPVHLP CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(col)Px
97	THTSGGCGCQPGMPRGPAGAPGQGFQGNPGEPEGVSGPMGPRGPP CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(col2)Px
98	THTSGGCGCQPPPTTYETE KLEPAPVEPSYEAETPTTRAPDQAEPNKPTPPTTYETEKLEPAP VEPSYEAETPTTRAPDQAEPNKPTPPTPVHLP CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(Adh)
99	THTSGGCGCQPPPTTYETE KLEPAPVEPSYEAETPTTRAPDQAEPNKPTPPTTYETEKLEPAPVEPSYEA EPTPTTRAPDQAEPNKPTPPTPCPPC	Z(Adh)2
100	THTSGGCGCQPGMPRGPAGAPGQGFQGNPGEPEGVSGPMGPRGPPPVHLP PGQGFQGNPGEPEGVSGPMGPRGPPPVHLP CHYTPQPRPQPHPCQPCQCPHFSPCC	Z(col)
101	THTSGGCGCQ PGMPRGPAGAPGQGFQGNPGEPEGVSGPMGPRGPPPCPPCGMPRGPAGAPGQGFQGNPGE GEPVSGPMGPRGPPPCPPC	Z(col2)

Tabla 2: Resumen de PBIS ejemplares de la presente invención o divulgadas en el presente documento indicando la secuencia N1, P1I/rica en prolina y la segunda secuencia rica en prolina. Esas PBIS que no tienen segunda secuencia rica en prolina contienen las dos cisteínas C-terminales unidas a la región P1I/rica en prolina como un tetrapéptido CPPC.

También se divulga una PBIS que comprende una región N1 (que comprende las dos cisteínas N-terminales) y una región PPII/rica en prolina que comprende todo o parte de PPII de RX3, pero que carece de la segunda región rica en prolina. Para proporcionar las cisteínas C-terminales necesarias, esta PBIS comprende dos cisteínas que forman un tetrapéptido CPPC que está directamente unido al extremo C de la región PPII. Por tanto, también se divulga una PBIS que comprende una secuencia seleccionada del grupo que consiste en:

R8(4C) (SEQ ID NO:102):
THTSGGCGCQ PPPPVHLPPP VHLPPPVHLP PPVHLPPP VH LPPP VH LPPP VHVPPP VHLC PPC

R7(4C) (SEQ ID NO:103):
THTSGGCGCQ PPPPVHLPPP VHLPPPVHLP PPVHLPPP LH LPPP VH LPPP VHVCPPC

R6(4C) (SEQ ID NO:104):
THTSGGCGCQ PPPPVHLPPP VHLPPPVHLP PPVHLPPP LH LPPP VH LCPP C

R4(4C) (SEQ ID NO:105):
THTSGGCGCQ PPPPVHLPPP VHLPPPVHLP PPVHLPPP LH LCPPC

También se divulga una PBIS que comprende una región N1 (que comprende las dos cisteínas N-terminales), una región PPII/rica en Pro que comprende todo o parte de PPII de RX3 y una región N2, pero en donde la región N1 comprende los aminoácidos del dominio ProX encontrado en RX3 en la orientación inversa y en donde la región N2 comprende los aminoácidos encontrados en la región N1 de RX3 en la orientación inversa. Por tanto, también se divulga una PBIS que comprende seleccionado del grupo que consiste en:

iRX3 (SEQ ID NO:106)
QCPSHPQQC PCPHPQPHPQ PRPPQTPYHC PPPPVHLPPP VHLPPPVHLP PPVHLPPP LH LPPP VH LPPP VHLPPP VHLPPP VHLPPP PPPQCGCGGS THT

VI. Proteínas de fusión de PBIS

Como se describe en el presente documento, la PBIS recombinante se puede fusionar a cualquier producto de interés. Por tanto, en otro aspecto, la invención se refiere a una proteína de fusión que comprende la PBIS recombinante de la invención y una proteína heteróloga. El término "proteína heteróloga" se usa de forma intercambiable en el presente documento con "producto de interés" o "PDI" y puede ser, por ejemplo, una proteína o péptido. La PBIS recombinante se puede fusionar en el extremo N o C a una proteína o péptido de interés. Además, la PBIS recombinante se puede fusionar directamente a la proteína o péptido de interés o se puede fusionar indirectamente, por ejemplo, a través de un espaciador, a la proteína o péptido de interés.

Por tanto, la PBIS recombinante se puede fusionar, por ejemplo, a una enzima, una hormona, tales como calcitonina, eritropoyetina, trombopoyetina, hormona del crecimiento humana, factor de crecimiento epidérmico, y similares, un interferón o una citoquina. Otros ejemplos de proteínas o péptidos de interés incluyen cualquier proteína que tenga uso terapéutico, nutracéutico, agrícola o industrial. Por ejemplo, en algunas formas de realización, la PBIS recombinante se puede fusionar a péptidos enriquecidos en aminoácidos esenciales.

Actividades ilustrativas de tales otras proteínas incluyen (a) captura y emisión de luz como proporcionan la proteína fluorescente verde (GFP), proteína fluorescente cian mejorada (ECFP), proteína fluorescente roja (DsRed) y similares; (b) actividad enzimática que se puede asociar con señalización intracelular primaria y secundaria y rutas metabólicas, ejemplificadas por enteroquinasa, beta-glucoronidasa (GUS), fitasa, anhidrasa carbónica y enzimas industriales (hidrolasas, glucosidasas, celulasas, oxido-reductasas y similares); (c) interacción proteína-proteína, proteína-receptor y proteína-ligando tales como, por ejemplo, anticuerpos (Acm tales como IgG, IgM, IgA, etc.) y fragmentos de los mismos, hormonas (calcitonina, hormona del crecimiento humana (hGH), factor de crecimiento epidérmico (EGF) y similares), inhibidores de proteasas, antibióticos, antimicrobianos, inhibidores de la entrada de VIH (Ryser *et al.*, 2005 *Drug Discov Today*. 10:1085-1094), colágeno, lactoferrina humana y citoquinas; (d) antígenos proteicos y peptídicos para vacunas (virus de la inmunodeficiencia humana, VIH; antígenos de pre-superficie, superficie y núcleo de hepatitis B; gen P1 de poliproteína estructural del virus de la fiebre aftosa (FMDV) (Dus Santos *et al.*, 2005 *Vaccine*.23:1838-1843), péptidos estimuladores de células T de la patente de EE UU No. 4.882.145, coronavirus de la gastroenteritis, virus del papiloma humano y similares); (e) interacciones proteína-no proteína tales como fitohemaglutinina (PHA), la subunidad B de la toxina ricina (RTB) y otras lectinas.

Como se describe en el presente documento, el producto de interés puede mantener su actividad funcional cuando se expresa como una fusión con la PBIS recombinante o cuando se purifica de un RPBLA. Los ensayos para la bioactividad de tales polipéptidos expresados se conocen bien en la técnica y están disponibles en una o más publicaciones. Por ejemplo, se puede medir la actividad de ECFP cuantificando la fluorescencia emitida a una longitud de onda de 470-530 nm cuando la proteína se ha excitado a 458 nm. Véase Richards *et al.*, 2003 *Plant Cell Rep.* 22:117-121. Se puede medir la actividad de la enteroquinasa (EK), por ejemplo, con dos enfoques diferentes.

La actividad se puede determinar analizando el corte de una proteína de fusión que contenga el sitio de corte específico de enteroquinasa mediante inmunotransferencia, como se discute en el catálogo de Invitrogen Life Technologies (E180-01 y E180-2) y también cuantificando la actividad EK usando un sustrato peptídico fluorogénico para EK (Sigma G-5261, CAS RN 70023-02-8); se mide la actividad enzimática mediante un aumento de la fluorescencia (excitación a 337 nm, emisión a 420 nm) producida por la liberación de β -naftilamina del péptido durante el tiempo. Véase, LaVallie *et al.*, 1993 *J. Biol. Chem.* 268:23311-23317. La actividad de la enzima beta-glucuronidasa (GUS) se puede medir mediante la conversión del sustrato MUG (4-metil umbeliferil glucuronido) en el producto MU. Este producto se puede cuantificar midiendo la fluorescencia con excitación a 365 nm, emisión a 455 nm en un espectrofluorímetro. Véase, Pai-Hsiang *et al.*, 2001 *J. Plant Physiol.* 158:247-254 y Jefferson *et al.*, 1987 *EMBO J.* 6:3901-3907. Los ensayos de fitasa se llevan a cabo mediante la cuantificación de ortofosfatos inorgánicos liberados del reactivo AAM que consiste en acetona, ácido sulfúrico 5,0 N y molibdato de amonio 10 mM. Véase Ullah *et al.*, 1999 *Biochem. Biophys. Res. Commun.* 264:201-206.

Ensayos similares están disponibles para otras proteínas biológicas. Los ensayos de actividad de RTB se pueden realizar midiendo la unión de RTB a asialofetúina, lactosa y galactosa, como se describe en Reed *et al.*, 2005 *Plant Cell Rep.* 24:15-24. EGF es un factor de crecimiento implicado en proliferación de fibroblastos. La actividad de EGF se puede ensayar mediante la cuantificación de la inducción de la síntesis de ADN medida mediante la incorporación del análogo de pirimidina 5-bromo-2'-desoxiuridina (BrdU), en lugar de timidina, en el ADN de células en proliferación usando el kit de proliferación celular ELISA. Oliver, *et al.*, 2004 *Am. J. Physiol. Cell Physiol.* 286:1118-1129; No. de Catálogo. 1647229, Roche Diagnostics, Mannheim, Alemania.

Se advierte que la captura y emisión de luz constituye un tipo separado y especial de "actividad biológica" que es la actividad luminiscente. Estas proteínas son útiles, por ejemplo, como moléculas indicadoras en muchos tipos de ensayos o cribados usados en el análisis o descubrimiento de moléculas biológicamente importantes, y su actividad luminiscente requiere la presencia de la estructura proteica secundaria y terciaria correcta.

En algunas formas de realización, la proteína de fusión de PBIS recombinante comprende una secuencia de aminoácidos espaciadora. La secuencia espaciadora de aminoácidos puede ser una secuencia de aminoácidos cortable por medios enzimáticos o químicos, o no cortable. Mediante "no cortable" se quiere decir que no se produce el corte del espaciador sin destrucción de parte o todo el polipéptido biológicamente activo.

El espaciador se puede colocar entre la PBIS recombinante y el polipéptido biológicamente activo. Un espaciador ilustrativo es una secuencia de aminoácidos que es cortable por una proteasa tal como enteroquinasa, endoproteasa Arg-C, endoproteasa Glu-C, endoproteasa Lys-C, Factos Xa, proteasas SUMO (Tauseef *et al.*, 2005 *Protein Expr. Purif.* 43:1-9) y similares. El espaciador también puede corresponder a una secuencia autocortable tal como la secuencia 2A autoprocresable vírica de FMDV, intrones proteicos (inteínas) tales como la inteína Ssp ADNb y similares, comercialmente disponibles de New England Biolabs y otros. El uso de un enlazador inteína puede ser ventajoso porque se puede inducir a que tales secuencias produzcan selectivamente ajuste de proteínas y por tanto se eliminan a sí mismas de una proteína expresada, recuperada. Las inteínas son particularmente interesantes puesto que no requieren grandes enzimas proteicas para alcanzar su sitio diana para cortar la PBIS recombinante de la proteína de interés. Esta propiedad puede ser particularmente útil para el aislamiento directo de proteínas de interés a partir de RPBLA intactos. De forma alternativa, un espaciador puede ser una secuencia de aminoácidos que es específicamente cortable por un reactivo químico, tal como, por ejemplo, bromuro de cianógeno que corta en residuos de metionina.

En una forma de realización preferida, el espaciador está compuesto de un péptido pentaglicina. En una forma de realización aún más preferida, el espaciador comprende un péptido pentaglicina seguido por un sitio de corte de proteasa. En una forma de realización aún más preferida, el sitio de corte de la proteasa se selecciona del grupo del sitio de corte de la enteroquinasa (D-D-D-D-K, que corta después del residuo K) y el sitio de corte del factor Xa (I-D/E-G-R, que corta después del residuo R).

VII. Ácidos nucleicos que codifican PBIS recombinante y proteínas de fusión de PBIS recombinante

También se describen en el presente documento polinucleótidos que codifican una PBIS recombinante. De forma similar, también se describen polinucleótidos que codifican una proteína de fusión que comprende una PBIS recombinante. Los polinucleótidos de la invención pueden estar en forma de ARN o en forma de ADN. ADN incluye ADNc, ADN genómico y ADN sintético; y puede ser bicatenario o monocatenario y si es monocatenario puede ser la hebra codificante o la hebra no codificante (antisentido). En ciertas formas de realización, los polinucleótidos están aislados. En ciertas formas de realización, los polinucleótidos están sustancialmente puros. En ciertas formas de realización, los polinucleótidos tienen codones optimizados para el huésped en donde se van a expresar los RPBLA.

Los polinucleótidos según la presente invención pueden comprender además una secuencia que codifica una secuencia señal que dirige la PBIS o la proteína de fusión al retículo endoplásmico, en donde dicha secuencia señal está en el mismo marco abierto de lectura que la PBIS o la proteína de fusión.

Como se ha descrito anteriormente, la PBIS recombinante puede comprender una señal que dirige la PBIS recombinante a una localización particular en la célula. Por ejemplo, la PBIS recombinante se puede dirigir a la luz del retículo endoplásmico (RE) a través de un péptido señal.

5 La señal puede ser cualquier dominio que dirija la PBIS recombinante al RE. A modo de ejemplo solo, un dominio de señalización a RE puede ser un péptido de señalización a RE derivado de una proteína zeína tal como gamma-zeína o alfa-zeína, una proteína gliadina tal como alfa-gliadina o gamma-gliadina, o la proteína relacionada con patogénesis de PR10 clase 25. En una forma de realización preferida, la secuencia señal comprende la secuencia MRVLLVALALLALAASATS (SEQ ID NO:58). Alternativamente, la secuencia señal puede derivar del gen relacionado con patogénesis, en particular el gen PR-10 y, lo más preferiblemente, la proteína PR-2 de tabaco (Veroverd et al.,
10 Plant Physiol., 1995, 109:1199-1205). En una forma de realización preferida, la secuencia señal comprende la secuencia MFLKSPFYAFLCFGQYFVAVTHA (SEQ ID NO:59).

15 Las características de los péptidos señal responsables para dirigir la proteína al RE se han estudiado extensamente (von Heijne *et al.*, 2001 *Biochim. Biophys. Acta* 1541:114-119). Los péptidos señal no comparten homología en la estructura primaria, pero tienen una estructura tripartita común: una región hidrofóbica central y regiones flanqueantes N- y C-terminales hidrofílicas. Estas similitudes, y el hecho de que las proteínas se translocan a través de la membrana del RE usando rutas aparentemente comunes, permite el intercambio de los péptidos señal entre diferentes proteínas o incluso de diferentes organismos que pertenecen a diferentes filos. Véase, Martoglio *et al.*,
20 1998 *Trends Cell Biol.* 8:410-415.

El péptido señal se puede cortar una vez que la PBIS recombinante ha alcanzado la localización celular adecuada, por ejemplo, el RE. En la mayoría de los eucariotas, el péptido señal se corta co-traduccionalmente, por tanto, la gran mayoría de la proteína encontrada en el compartimento de endomembrana (por ejemplo, el RE, Golgi, vacuolas) es la proteína madura, esto es, la proteína sin el péptido señal. En consecuencia, la proteína madura es la responsable de inducir la formación de los RPBLA. En una divulgación preferida, la secuencia señal es el péptido señal de gamma zeína que tiene la secuencia MRVLLVALALLALAASATS (SEQ ID NO:58).

30 Tales polinucleótidos se pueden, por ejemplo, incorporar en un vector de expresión para producir una PBIS recombinante o una proteína de fusión que comprende una PBIS recombinante en una célula. Los vectores de expresión son construcciones de ADN replicables que tienen fragmentos de ADN sintético o derivados de ADNc que codifican una PBIS recombinante o una proteína de fusión que comprende una PBIS recombinante, operativamente unidos a elementos reguladores transcripcionales o traduccionales adecuados. Los elementos reguladores transcripcionales o traduccionales pueden derivar, por ejemplo, de mamífero, microbio, virus, insecto o genes. Una unidad transcripcional en general comprende un ensamblaje de (1) un elemento o elementos genético(s) que tiene(n) un papel regulador en la expresión génica, por ejemplo, promotores o potenciadores de transcripción, (2) una secuencia estructural o codificante que se transcribe a ARNm y se traduce a proteína, y (3) secuencias apropiadas de iniciación y terminación de transcripción y traducción, como se describe en detalle posteriormente. Tales elementos reguladores pueden incluir una secuencia operadora para controlar la transcripción. Se puede
40 incorporar adicionalmente la capacidad para replicarse en un huésped, normalmente conferida por un origen de replicación, y un gen de selección para facilitar el reconocimiento de transformantes. Las regiones de ADN están operativamente unidas cuando están funcionalmente relacionadas entre sí. Por ejemplo, el ADN para un péptido señal está operativamente unido al ADN para un polipéptido si se expresa como un precursor que participa en la secreción del polipéptido; un promotor está operativamente unido a una secuencia codificante si controla la transcripción de la secuencia; o un sitio de unión al ribosoma está operativamente unido a una secuencia codificante si está colocado de modo que permite la traducción.

La elección de la secuencia de control de la expresión y el vector de expresión dependerá de la elección del huésped. Se puede emplear una gran variedad de combinaciones de huésped/vector de expresión. Los vectores de expresión útiles para huéspedes eucariotas incluyen, por ejemplo, vectores que comprenden secuencias de control de expresión de SV40, virus del papiloma bovino, adenovirus y citomegalovirus. Los vectores de expresión útiles para huéspedes bacterianos incluyen plásmidos bacterianos conocidos, tales como plásmidos de *Escherichia coli*, incluyendo pCR1, pBR322, pMB9 y sus derivados, plásmidos de gama de huéspedes más amplia, tales como M13 y fagos filamentosos de ADN monocatenario.

55 En algunas formas de realización, el vector que comprende un polinucleótido que codifica una PBIS recombinante comprende además un sitio múltiple de clonación. Un sitio múltiple de clonación es una secuencia de polinucleótido que comprende uno o más sitios de restricción únicos. Ejemplos no limitantes de sitios de restricción incluyen EcoRI, SacI, KpnI, SmaI, XmaI, BamHI, XbaI, HincII, PstI, SphI, HindIII, Aval o cualquier combinación de los mismos.

60 Los sitios múltiples de clonación se pueden usar en vectores que comprenden un polinucleótido que codifica una PBIS recombinante para simplificar la inserción de un polinucleótido que codifica una proteína o péptido de interés en el vector de modo que el vector se pueda usar para expresar una proteína de fusión que comprende la PBIS recombinante y la proteína o péptido de interés. En algunas formas de realización, el polinucleótido que codifica la

PBIS recombinante está en 5' respecto al sitio múltiple de clonación. En algunas formas de realización, el polinucleótido que codifica la PBIS recombinante está en 3' respecto al sitio múltiple de clonación.

5 Los vectores pueden comprender al menos un promotor. El promotor puede ser cualquier secuencia que sea adecuada para dirigir la expresión de una PBIS recombinante o proteína de fusión que comprende una PBIS recombinante. En una divulgación particular, el promotor dirige la expresión en hojas de tabaco.

10 Con frecuencia diferentes huéspedes tienen preferencias para usar un codón particular para codificar un residuo particular de aminoácido. Tales preferencias de codones se conocen bien y se puede cambiar una secuencia de ADN que codifica una secuencia de proteína de fusión deseada, usando mutagénesis *in vitro*, por ejemplo, de modo que se utilicen los codones preferidos por el huésped para un huésped particular en el que se va a expresar la proteína de fusión.

15 También se contempla una molécula recombinante de ácido nucleico tal como una molécula de ADN, que comprende un vector génico o construcción que contiene una o más secuencias reguladoras (elementos de control) tal como un promotor adecuado para dirigir la expresión del gen en un organismo celular eucariota huésped operativamente unido a un segmento de ácido nucleico exógeno (por ejemplo, un segmento o secuencia de ADN) que define un gen que codifica una proteína de fusión contemplada, como se ha discutido anteriormente. Más en particular, también se contempla una molécula recombinante de ADN que comprende un vector génico que
20 comprende un promotor para dirigir la expresión de la proteína de fusión en células de un organismo huésped operativamente unido a un segmento de ADN que define un gen que codifica una secuencia inductora de cuerpos proteicos (PBIS) unida a un polipéptido de interés. Esa molécula de ADN recombinante, tras transfección y expresión adecuadas en una célula eucariota huésped, proporciona una proteína de fusión contemplada como RPLBA.

25 Como se sabe bien en la técnica, siempre que el ácido nucleico requerido, ilustrativamente secuencia de ADN, esté presente (incluyendo señales de iniciación y parada), normalmente pueden estar presentes pares de bases adicionales en cualquier extremo del segmento de ADN y ese segmento se puede utilizar todavía para expresar la proteína. Esto, por supuesto, presume la ausencia en el segmento de una secuencia de ADN operativamente unida que reprima la expresión, exprese un producto adicional que consuma la proteína de fusión deseada a ser expresada, exprese un producto que consuma un producto de reacción querido producido por esa proteína de fusión deseada o interfiera de otra manera con la expresión del gen del segmento de ADN.

30 Por tanto, siempre que el segmento de ADN esté libre de tales secuencias interferentes de ADN, un segmento de ADN puede tener de aproximadamente 500 hasta aproximadamente 1.500 pares de bases de longitud. El tamaño máximo de una molécula recombinante de ADN, particularmente un vector de expresión, está regido principalmente por conveniencia y el tamaño del vector que puede acomodar una célula huésped, una vez que todas las secuencias mínimas de ADN requeridas para replicación y expresión, cuando se desea, están presentes. Los tamaños mínimos de vectores se conocen bien.

40 Un segmento de ADN que codifica una proteína de fusión anteriormente descrita se puede sintetizar mediante técnicas químicas, por ejemplo, el método del fosfotriéster de Matteucci *et al.*, 1981 J. Am. Chem. Soc., 103:3185. Por supuesto, mediante la síntesis química de la secuencia codificante, se puede hacer cualquier modificación deseada simplemente sustituyendo las bases apropiadas por las que codifican la secuencia nativa de residuos de aminoácidos.

Los segmentos de ADN que contienen un gen que codifica la proteína de fusión también se pueden obtener de moléculas recombinantes de ADN (vectores plásmidos) que contienen ese gen.

50 Un vector que dirige la expresión de un gen de una proteína de fusión se denomina en el presente documento un "vector de expresión". Un vector de expresión contiene elementos de control de la expresión incluyendo el promotor. El gen codificante de la proteína de fusión está operativamente unido al vector de expresión para permitir que la secuencia promotora dirija la unión de la ARN polimerasa y la expresión del gen codificante de la proteína de fusión. Útiles en la expresión de genes que codifican polipéptidos son promotores que son inducibles, víricos, sintéticos, constitutivos como describen Paszkowski *et al.*, 1989 EMBO J., 3:2719 y Odell *et al.*, 1985 Nature, 313:810, así como temporalmente regulados, espacialmente regulados y regulados espaciotemporalmente como se da en Chua *et al.*, 1989 Science, 244:174-181.

60 En el presente documento se contemplan vectores de expresión compatibles con células eucariotas, tales como los que son compatibles con células de mamífero, algas o insectos y similares. Tales vectores de expresión también se pueden usar para formar las moléculas recombinantes de ADN de la presente invención. Los vectores de expresión de células eucariotas se conocen bien en la técnica y están disponibles de varias fuentes comerciales. Normalmente, tales vectores contienen uno o más sitios de restricción convenientes para la inserción del segmento de ADN y secuencias promotoras deseadas. Opcionalmente, tales vectores contienen un marcador de selección específico para su uso en células eucariotas.

65

La elección de qué vector de expresión y por último a qué promotor se une operativamente un gen que codifica una proteína de fusión depende directamente de las propiedades funcionales deseadas, por ejemplo, la localización y el momento de la expresión proteica y la célula huésped que se va a transformar. Estas son limitaciones bien conocidas inherentes en la técnica de construcción de moléculas recombinantes de ADN. Sin embargo, un vector útil en practicar la presente invención puede dirigir la replicación y preferiblemente también la expresión (para un vector de expresión) del gen de la proteína de fusión incluido en el segmento de ADN al que se fusiona operativamente.

Vectores típicos útiles para la expresión de genes en células de plantas y mamíferos superiores son bien conocidos en la técnica e incluyen vectores vegetales derivados del plásmido inductor de tumores (Ti) de *Agrobacterium tumefaciens* descrito por Rogers *et al.* (1987) *Meth. in Enzymol.*, 153:253-277 y vectores de expresión en mamíferos pKSV-10, anteriormente, y pCl-neo (Promega Corp., #E1841, Madison, Wis.). Sin embargo, se sabe que algunos otros sistemas de vectores de expresión funcionan en plantas incluyendo el vector de control de transferencia pCaMVCN descrito por Fromm *et al.* (1985) *Proc. Natl. Acad. Sci. USA*, 82:58-24. El plásmido pCaMVCN (disponible de Pharmacia, Piscataway, N.J.) incluye el promotor del virus del mosaico de la coliflor CaMV 35S.

VIII. Ensamblajes recombinantes similares a cuerpos proteicos (RPBLA), células huésped y métodos de hacer ensamblajes recombinantes similares a cuerpos proteicos (RPBLA)

Los cuerpos proteicos naturales se han descrito en el endospermo de cereales. Se inducen por la expresión de prolaminas y glutelinas. Los cuerpos proteicos son orgánulos que están muy enriquecidos en proteínas ensambladas (por ejemplo, prolaminas y glutelinas) y rodeados por una membrana, que puede derivar del RE o la vacuola. La mayoría de los cuerpos proteicos son estructuras de forma redonda (generalmente esféricas), con diámetros desde aproximadamente 0,5 hasta aproximadamente 3,0 micrómetros.

Los ensamblajes recombinantes similares a cuerpos proteicos (RPBLA) se pueden formar por la expresión de una PBIS recombinante en una célula. De forma similar a los cuerpos proteicos naturales, que son orgánulos que están muy enriquecidos en PBIS, los RPBLA son orgánulos recombinantes que están muy enriquecidos en PBIS recombinantes. La PBIS ensamblada o PBIS recombinante en los orgánulos están rodeados por una membrana. En células, los RPBLA típicamente se encuentran en el citoplasma de la célula y, por tanto, están rodeados de una membrana adicional (membrana plasmática). Estas membranas (membrana plasmática y la membrana del orgánulo) se pueden eliminar durante el proceso de recuperación de RPBLA, y el orgánulo aún se considera un RPBLA.

Cuando se expresan en células animales, los RPBLA en general tienen forma esférica, tienen diámetros desde aproximadamente 0,5 hasta aproximadamente 3,0 micrómetros y tienen una membrana que los rodea. Los RPBLA expresados en células vegetales normalmente también son en general esféricos, tienen diámetros desde aproximadamente 0,5 hasta aproximadamente 2,0 micrómetros y están rodeados por una membrana. Sin embargo, los RPBLA pueden algunas veces tener forma amorfa y tamaño no uniforme.

En algunas formas de realización, los RPBLA tienen al menos aproximadamente 0,3, al menos aproximadamente 0,4 o al menos aproximadamente 0,5 micrómetros. En algunas formas de realización, los RPBLA tienen aproximadamente 3 micrómetros o menos, aproximadamente 2,5 micrómetros o menos, o aproximadamente 2 micrómetros o menos. En algunas formas de realización, los RPBLA tienen desde aproximadamente 0,3 hasta aproximadamente 3,0 micrómetros, desde aproximadamente 0,3 hasta aproximadamente 2,5 micrómetros o desde aproximadamente 0,3 hasta aproximadamente 2 micrómetros. En algunas formas de realización, los RPBLA tienen desde aproximadamente 0,5 hasta aproximadamente 3,0 micrómetros, desde aproximadamente 0,5 hasta aproximadamente 2,5 micrómetros o desde aproximadamente 0,5 hasta aproximadamente 2 micrómetros.

En algunas formas de realización, los RPBLA tienen una densidad predeterminada que puede ser diferente entre diferentes proteínas de fusión, pero es predecible en huéspedes para una proteína de fusión particular que se prepara. Esa densidad predeterminada de los RPBLA es típicamente mayor que esa de sustancialmente todas las proteínas endógenas de la célula huésped presente en el homogeneizado y es típicamente desde aproximadamente 1 hasta aproximadamente 1,4 g/ml. La alta densidad de los RPBLA se debe a la capacidad general de las proteínas de fusión recombinantes de autoensamblarse y acumularse en agregados ordenados asociados con membranas.

En algunas formas de realización, un RPBLA tiene una densidad de al menos aproximadamente 1,0 g/ml. En algunas formas de realización, el RPBLA tiene una densidad desde aproximadamente 1,1 hasta aproximadamente 1,4 g/ml, desde aproximadamente 1,15 hasta aproximadamente 1,35 g/ml o desde aproximadamente 1,0 hasta aproximadamente 1,3 g/ml. En algunas formas de realización, el RPBLA tiene una densidad que no es más de 1,4 g/ml.

Los RPBLA contemplados se pueden caracterizar por su densidad como se apuntó anteriormente y su tamaño y forma. El gradiente de densidad iodixanol (Optiprep®) en colchón por capas descrito en los ejemplos y materiales y métodos es un método útil para determinar la densidad de RPBLA determinados. Otros métodos complementarios o

adecuados que se pueden usar para determinar la densidad de los RPBLA incluyen gradientes de densidad en colchón por capas que se basan en otros solutos que proporcionan densidad tal como sacarosa, glicerol y Percoll.

También se divulga un método para producir un RPBLA que comprende:

- 5
- (i) transformar un sistema huésped vegetal con el ácido nucleico o vector según la divulgación;
 - (ii) generar plantas a partir de dicho sistema vegetal transformado; y
 - (iii) hacer crecer dichas plantas en condiciones adecuadas para la formación de RPBLA.
- 10 En algunas formas de realización, los RPBLA se producen en una célula eucariota. Por ejemplo, se pueden producir RPBLA en plantas, animales, insectos u hongos. Las células huésped adecuadas para la producción de RPBLA incluyen, a modo de ejemplo, plantas superiores (por ejemplo, tomate, tabaco, Arabidopsis, alfalfa), células de mamífero (por ejemplo, células CHO, COS y 293T), hongos filamentosos (por ejemplo, Trichoderma resei y Aspergillum sp.), células de insecto. Véase también la patente de EE UU No. 7.575.898 y la solicitud publicada de EE UU No. 2006/0121573, que se incorporan en el presente documento mediante referencia y describen otras
- 15 células huésped de ejemplo que se pueden usar para producir un RPBLA como se describe en el presente documento. En algunas formas de realización, se expresa un RPBLA en una célula de planta de tabaco.

20 En aún otras formas de realización, la célula huésped es una célula eucariota superior. Las células eucariotas superiores incluyen líneas celulares establecidas de mamíferos. Se emplean de forma ventajosa varios sistemas de cultivo de células de mamífero para expresar RPBLA porque las proteínas en general se pliegan correctamente, se modifican apropiadamente y son completamente funcionales. Los ejemplos de líneas celulares huésped de mamíferos incluyen las líneas COS-7 de células de riñón de mono, descritas por Gluzman (*Cell* 23:175, 1981) y otras líneas celulares capaces de expresar un vector apropiado incluyendo, por ejemplo, células L, líneas celulares C127, 3T3, de ovario de hámster chino (CHO), HeLa y BHK. Se revisan los sistemas de baculovirus para la

25 producción de proteínas heterólogas en células de insecto en Luckow y Summers, *Bio/Technology* 6:47 (1988).

También se pueden producir RPBLA en diferentes tejidos tales como hojas, granos, raíces y cotiledones. En algunas formas de realización, se expresa un RPBLA en una célula de hoja de tabaco.

30 Como se describe en más detalle posteriormente, se puede producir un RPBLA cultivando una célula huésped que comprende un polinucleótido que codifica una PBIS recombinante o una proteína de fusión que comprende una PBIS recombinante en condiciones adecuadas para la formación de RPBLA. Por ejemplo, se puede producir un RPBLA en una célula huésped vegetal transformando una célula huésped vegetal con un polinucleótido que

35 comprende una secuencia que codifica una PBIS recombinante o una proteína de fusión que comprende una PBIS recombinante, generando plantas transformadas a partir de la célula huésped y haciendo crecer las plantas en condiciones que son adecuadas para la formación de RPBLA.

40 Se describen ciertas condiciones que son adecuadas para la formación de RPBLA en los ejemplos posteriores. Los expertos en la materia pueden determinar otras condiciones que sean adecuadas para la formación de RPBLA expresando la PBIS recombinante como se describe en el presente documento y evaluando la capacidad de tal PBIS recombinante conocida para formar un RPBLA en las condiciones ensayadas.

45 **IX. Purificación de RPBLA, de las proteínas de fusión que comprenden PBIS recombinante y de las proteínas de interés**

Los RPBLA según la invención se pueden purificar usando métodos conocidos. En particular, puesto que los RPBLA son densos relativos a otras proteínas presentes en las células huésped, los RPBLA son particularmente susceptibles de ser recogidos por centrifugación de un homogeneizado celular.

50 También se divulga un método para purificar un RPBLA que comprende (i) cultivar una célula huésped según la invención en condiciones adecuadas para la formación de RPBLA; y (ii) purificar el cuerpo proteico recombinante. En aún otra divulgación, el método para producir un RPBLA además comprende

- 55
- (i) transformar un sistema huésped vegetal con el ácido nucleico o vector según la divulgación;
 - (ii) generar plantas a partir de dicho sistema vegetal transformado;
 - (iii) hacer crecer dichas plantas en condiciones adecuadas para la formación de RPBLA y
 - (iv) purificar el RPBLA.

60 Por tanto, en una divulgación preferida, la purificación del RPBLA comprende

- (i) preparar un homogeneizado celular acuoso;
 - (ii) formar regiones de diferente densidad en el homogeneizado celular acuoso para proporcionar una región que contenga una concentración aumentada de los RPBLA y una región que contenga una concentración
- 65 disminuida de los RPBLA; y

- (iii) separar la región con RPBLA disminuidos de la región de concentración relativamente aumentada de RPBLA, purificando mediante ello el RPBLA.

5 Por tanto, se forman regiones de diferente densidad en el homogeneizado para proporcionar una región que contiene una concentración relativamente aumentada de los RPBLA y una región que contiene una concentración relativamente disminuida de los RPBLA. La región con RPBLA disminuidos se separa de la región de concentración relativamente aumentada de RPBLA, purificándose de esta manera dichos RPBLA.

10 Una vez se purifican los RPBLA, los RPBLA recogidos se usan como están, sin necesidad de aislar la proteína de fusión, como donde los RPBLA se usan como una vacuna oral. Sin embargo, la región de concentración relativamente aumentada se puede recoger después de ello o se puede tratar con uno o más reactivos o someter a uno o más procedimientos antes del aislamiento de los RPBLA o la proteína de fusión en los mismos. Por tanto, en otro aspecto, la invención se refiere a un método de purificar una proteína de fusión que comprende

- 15 (i) proporcionar RPBLA que comprenden una proteína de fusión encerrada por una membrana, en donde la proteína de fusión es una proteína de fusión según la invención;
 (ii) poner en contacto los RPBLA con un tampón acuoso que contiene una cantidad de un tensioactivo que desensambla la membrana;
 20 (iii) mantener el contacto durante un periodo de tiempo suficiente para desensamblar la membrana y a una temperatura que no desnaturalice la proteína de fusión para separar la membrana de la proteína de fusión;
 y
 (iv) recoger la proteína de fusión separada.

25 La proteína de fusión que contiene el polipéptido biológicamente activo se puede obtener de los RPBLA recogidos por disolución de la membrana que la rodea en tampón acuoso que contiene un detergente y/o un agente reductor. Los agentes reductores ilustrativos incluyen 2-mercaptoetanol, ácido tioglicólico y sales de tioglicolato, ditiotreitól (DTT), iones sulfito o bisulfito, seguido por métodos de aislamiento de proteínas habituales. Dodecilsulfato de sodio (SDS) es el detergente preferido, aunque se pueden usar otros tensioactivos iónicos (desoxicolato, N-laurilsarcosina, y similares), no iónicos (Tween 20, Nonidet P-40, glucósido de octilo y similares), y dipolares (CHAPS, Zwittergent.TM, serie 3-X y similares). Se utiliza una cantidad mínima de tensioactivo que disuelve o dispersa la proteína de fusión.

30 La PBIS recombinante o proteínas de fusión que comprenden la PBIS recombinante se pueden purificar según cualquier otro método adecuado. Los métodos estándar incluyen cromatografía (por ejemplo, cromatografía de intercambio iónico, de afinidad y en columna por tamaño), centrifugación, solubilidad diferencial o mediante cualquier otra técnica estándar para purificación de proteínas. Se pueden unir a la proteína etiquetas de afinidad tales como hexahistidina, dominio de unión a maltosa, secuencia de recubrimiento de la gripe y glutatión S-transferasa para permitir la purificación fácil mediante paso por una columna de afinidad adecuada.

40 Las proteínas aisladas se pueden caracterizar físicamente usando técnicas conocidas tales como proteólisis, resonancia magnética nuclear y cristalografía de rayos X.

También se divulga un método de purificar una proteína que comprende

- 45 (i) proporcionar RPBLA que comprenden una proteína de fusión encerrada por membrana, en donde la proteína de fusión es la proteína de fusión según la invención;
 (ii) poner en contacto los RPBLA con un tampón acuoso que contiene una cantidad de un tensioactivo que desensambla la membrana;
 (iii) mantener el contacto durante un periodo de tiempo suficiente para desensamblar la membrana y a una
 50 temperatura que no desnaturaliza la proteína de fusión para separar la membrana de su proteína de fusión;
 (iv) recoger la proteína de fusión separada; y
 (v) cortar el sitio de corte entre la PBIS recombinante y la proteína heteróloga.

55 Se pueden preparar proteínas de fusión según un método que comprende transformar un sistema de células huésped eucariotas tal como un animal, cultivo de células animales, planta o cultivo de célula vegetales, cultivo de hongos, cultivo de células de insecto o cultivo de algas con una secuencia de ácido nucleico (ADN o ARN) que comprende (i) un primer ácido nucleico que codifica una PBIS recombinante que está operativamente unida a (ii) una segunda secuencia de ácido nucleico que codifica un producto polipeptídico de interés que es biológicamente activo. También se divulga el uso de medios indirectos para introducir el ADN, tal como mediante transducción o infección vírica, y se debe usar de forma intercambiable con métodos de administración directa de ADN tal como transfección.

60 La célula o entidad huésped transformada se mantiene durante un periodo de tiempo y en condiciones de cultivo adecuadas para la expresión de la proteína de fusión y ensamblaje de la proteína de fusión expresada en RPBLA. Tras la expresión, la proteína de fusión resultante se acumula en el sistema huésped transformado como RPBLA. La proteína de fusión se puede recuperar después de las células huésped o se pueden usar las células huésped que

contienen la proteína fusión como se desee, como para un pienso animal que contiene un nutriente o suplemento añadido. La proteína de fusión se puede aislar como parte de los RPBLA o libre de los RPBLA.

Las condiciones de cultivo adecuadas para la expresión de la proteína de fusión son típicamente diferentes para cada tipo de entidad huésped o célula huésped. Sin embargo, esas condiciones las conoce el experto en la materia y se determinan fácilmente. De forma similar, la duración del mantenimiento puede ser diferente con las células huésped y con la cantidad de proteína de fusión deseada a preparar. De nuevo, esas condiciones son bien conocidas y se pueden determinar fácilmente en situaciones específicas. Además, se pueden obtener condiciones específicas de cultivo de las citas y ejemplos en el presente documento.

En otra divulgación particular, se prepara una proteína de fusión según un método que comprende transformar el sistema de célula huésped tal como un animal, cultivo de células animales, planta, cultivo de células vegetales, hongos o algas con una secuencia de ácido nucleico que comprende además de las secuencias de ácidos nucleicos (i) y (ii) anteriormente mencionadas, una secuencia de ácido nucleico en el mismo marco de lectura (iii) que codifica una secuencia espaciadora de aminoácidos. La secuencia espaciadora de aminoácidos puede ser una secuencia de aminoácidos cortable por medios enzimáticos o químicos, o no cortable, como se apuntó anteriormente. En una divulgación particular, la secuencia de ácido nucleico (iii) se coloca entre dichas secuencias de ácido nucleico (i) y (ii), por ejemplo, el extremo 3' de la tercera secuencia de ácido nucleico (iii) está unido al extremo 5' de la segunda secuencia de ácido nucleico (ii). En otra forma de realización, el extremo 5' de la tercera secuencia de ácido nucleico (iii) está unido al extremo 3' de la segunda secuencia de ácido nucleico (ii).

También se puede usar un sistema de células de insecto para expresar una proteína de fusión contemplada. Por ejemplo, en uno de tales sistemas se usa el virus de la poliedrosis nuclear de *Autographa californica* (AcNPV) o baculovirus como vector para expresar genes exógenos en células de *Spodoptera frugiperda* o en larvas de *Trichoplusia*. Las secuencias que codifican una proteína de fusión se pueden clonar en una región no esencial del virus, tal como el gen de la poliedrina, y colocar bajo el control del promotor de la poliedrina. La inserción exitosa de una secuencia de proteína de fusión inactiva el gen de la poliedrina y produce virus recombinantes que carecen de proteína de recubrimiento. Los virus recombinantes se pueden usar después para infectar, por ejemplo, células de *S. frugiperda* o larvas de *Trichoplusia* en las que se puede expresar la proteína de fusión, por ejemplo, como se describe en Engelhard *et al.* (1994) *Proc. Natl. Acad. Sci.*, USA, 91:3224-3227 y V. Luckow, "Insect Cell Expression Technology", páginas 183-218, en *Protein Engineering: Principles and Practice*, J. L. Cleland *et al.* eds., Wiley-Liss, Inc, 1996). Los genes heterólogos colocados bajo el control del promotor de la poliedrina del virus de la poliedrosis nuclear de *Autographa californica* (AcNPV) con frecuencia se expresan a niveles altos durante las últimas fases de la infección.

Los baculovirus recombinantes que contienen el gen de la proteína de fusión se construyen usando el sistema de vector lanzadera de baculovirus (Luckow *et al.*, 1993 *J. Virol.*, 67:4566-4579], vendido comercialmente como el sistema de expresión de baculovirus Bac-To-Bac.TM. (Life Technologies). Se preparan soluciones madre de virus recombinantes y la expresión de la proteína recombinante se controla mediante protocolos estándar (O'Reilly *et al.*, *Baculovirus Expression Vectors: A Laboratory Manual*, W.H. Freeman and Company, Nueva York, 1992 y King *et al.*, *The Baculovirus Expression System: A Laboratory Guide*, Chapman & Hall, Londres, 1992). El uso de baculovirus u otros vectores de administración en células de mamífero, tal como el sistema 'BacMam' descrito por T. Kost y colaboradores (véase, por ejemplo, Merrihew *et al.*, 2004 *Methods Mol. Biol.* 246:355-365) u otros de tales sistemas los conoce el experto en la materia y también se contemplan en la presente divulgación.

Los sistemas de expresión vegetales típicamente proporcionan expresión sistémica o constitutiva de un transgén insertado. La expresión sistémica puede ser útil cuando la mayoría o toda la planta se usa como fuente de los RPBLA y sus proteínas de fusión. Sin embargo, puede ser más eficaz expresar los RPBLA y su contenido en proteína de fusión en un órgano de almacenamiento de la planta tal como una raíz, semilla o fruto del que las partículas se pueden aislar o ingerir más fácilmente.

Una manera de alcanzar expresión en un órgano de almacenamiento es usar un promotor que exprese su gen controlado en uno o más órganos vegetales no fotosintéticos preseleccionados o predeterminados. La expresión en uno o más órganos de almacenamiento preseleccionados con poca o ninguna expresión en otros órganos tales como raíces, semilla o fruto frente a hojas o tallos se denomina en el presente documento expresión mejorada o preferente. La expresión en sustancialmente sólo un órgano de almacenamiento y sustancialmente sin expresión en otros órganos de almacenamiento se denomina expresión específica de órgano; es decir, una relación de productos de expresión en un órgano de almacenamiento relativo a otro de aproximadamente 100:1 o mayor indica especificidad de órgano. Los promotores específicos de órganos de almacenamiento son por tanto miembros de la clase de promotores mejorados de órganos de almacenamiento.

Los órganos de almacenamiento vegetales de ejemplo incluyen las raíces de zanahorias, taro o yuca, tubérculos de patata y la carne de frutos tales como guayaba roja, fruta de la pasión, mango, papaya, tomate, aguacate, cereza, clementina, mandarina, palma, melones tal como cantalupo y sandías y otros frutos carnosos tales como

calabacines, pepinos, mangos, albaricoques, melocotones, así como las semillas de maíz, soja, arroz, aceite de semillas de colza y similares.

5 La transfección de células vegetales usando *Agrobacterium tumefaciens* típicamente se lleva a cabo mejor en plantas dicotiledóneas. Las monocotiledóneas normalmente se transforman más fácilmente por la denominada transferencia génica directa de protoplastos. La transferencia génica directa normalmente se lleva a cabo mediante electroporación, mediante transferencia mediada por polietilenglicol o bombardeo de células por microproyectiles que llevan el ADN necesitado. Estos métodos de transfección se conocen bien en la técnica y no necesitan discusión adicional en el presente documento. Los métodos para regenerar plantas enteras a partir de células
10 transfectadas y protoplastos también son bien conocidos, como lo son las técnicas para obtener una proteína deseada de tejidos vegetales. Véase también, las patentes de EE UU No, 5.618.988 y 5.679.880 y las citas en las mismas.

15 Una planta transgénica formada usando transformación con *Agrobacterium*, electroporación u otros métodos típicamente contiene un único gen en un cromosoma. Una planta transgénica también puede ser homocigota para el gen estructural añadido, es decir, una planta transgénica que contiene dos genes añadidos, un gen en el mismo locus en cada cromosoma de un par de cromosomas. Se puede obtener una planta transgénica homocigota apareando sexualmente (autofecundación) una planta transgénica segregante independiente que contiene un único gen añadido, germinando algunas de las semillas producidas y analizando las plantas resultantes producidas por la
20 acumulación mejorada de partículas quimeras relativo a una planta control (nativa, no transgénica) o transgénica segregante independiente. Una planta transgénica homocigota muestra acumulación mejorada de partículas quimeras comparada tanto con una planta nativa no transgénica como con una planta transgénica segregante independiente.

25 Un experto en la materia apreciará que la elección de la proteína o péptido que se puede producir usando la invención como se ha descrito es grande y variada. Pueden ser, por ejemplo, enzimas industriales, antígenos, citoquinas, receptores, agonistas, proteínas nutracéuticas, un producto con valor añadido, proteína farmacéuticamente activa y similares. Esto incluye, sin limitación, 16ESH, CTB, Gb, Les, TB, PAP, Cap, E2, NP, Her, glucosa oxidasa, glucosa isomerasa, peroxidasa, oxidasa alternativa, GOOX, beta-galactosidasa, glucosa amilasa,
30 lipasa, lipasa versátil, cloroperoxidasa, xilosa isomerasa, Mn peroxidasa, catalasa, formiato deshidrogenasa, alcohol deshidrogenasa, alfa-1-antitripsina, defensina, hormona del crecimiento humana, GM-CSF, EGF, factor de crecimiento de hepatocitos. Se apreciará que esta lista no es en ninguna manera exhaustiva.

35 X. Usos de los RPBLA como fármacos

Como se describe en el presente documento, la formación de RPBLA permite purificar proteínas de fusión que comprenden una PBIS recombinante usando técnicas muy sencillas. Por tanto, las proteínas de fusión que comprenden una PBIS recombinante y una proteína de interés se pueden expresar y purificar fácilmente. En algunas divulgaciones, la proteína de interés es una proteína terapéutica. Como tal, se puede formular un fármaco
40 purificando una proteína de fusión que comprende una PBIS recombinante y una proteína de interés y opcionalmente eliminando la PBIS recombinante usando métodos descritos en el presente documento o conocidos en la técnica. La proteína de interés aislada se puede formular después para uso farmacéutico según técnicas conocidas.

45 Por tanto, en algunas divulgaciones, la proteína de fusión o proteína de interés obtenida mediante los métodos descritos en el presente documento se formula en una forma "farmacéuticamente aceptable". "Farmacéuticamente aceptable" se refiere a un bioproducto que es, en el ámbito del juicio médico razonable, adecuado para contacto con los tejidos de seres humanos y animales sin toxicidad excesiva u otras complicaciones equivalentes a una relación beneficio/riesgo razonable.

50 La proteína de fusión o proteína de interés obtenida por los métodos descritos en el presente documento se puede formular en composiciones farmacéuticas para administración a mamíferos, incluyendo seres humanos. Las composiciones farmacéuticas usadas en los métodos de esta invención comprenden soportes farmacéuticamente aceptables, incluyendo, por ejemplo, intercambiadores de iones, alúmina, estearato de aluminio, lecitina, proteínas de suero, tal como seroalbúmina humana, sustancias reguladoras tal como fosfatos, glicina, ácido sórbico, sorbato de potasio, mezclas de glicéridos parciales de ácidos grasos vegetales saturados, agua, sales o electrolitos tales como sulfato de protamina, hidrogenofosfato disódico, hidrogenofosfato de potasio, cloruro de sodio, sales de zinc, sílice coloidal, trisilicato de magnesio, polivinilpirrolidona, sustancias basadas en celulosa, polietilenglicol, carboximetilcelulosa sódica, poliacrilatos, ceras, polímeros en bloque de polietileno-polioxipropileno, polietilenglicol y
60 lanolina.

Las composiciones usadas en los métodos de la presente invención se pueden administrar por cualquier método adecuado, por ejemplo, por vía parenteral, intraventricular, oral, por aerosol de inhalación, por vía tópica, rectal, nasal, bucal, vaginal o mediante un depósito implantado. El término "parenteral" como se usa aquí incluye técnicas

de inyección o infusión subcutánea, intravenosa, intramuscular, intraarticular, intrasinovial, intraesternal, intratecal, intrahepática, intralesión e intracraneal.

5 Una dosis y pauta de tratamiento específicas para cualquier paciente particular dependerá de una variedad de factores, incluyendo el anticuerpo particular usado, la edad, peso corporal, salud general, sexo y dieta del paciente, y el tiempo de administración, velocidad de excreción, combinación de fármacos y la gravedad de la enfermedad particular que se trata. El juicio de tales factores entre los profesionales médicos está dentro de de las habilidades de la técnica. La cantidad también dependerá del paciente individual que se va a tratar, la vía de administración, el tipo de formulación, las características del compuesto usado, la gravedad de la enfermedad y el efecto deseado. La cantidad usada se puede determinar mediante principios farmacológicos y farmacocinéticos bien conocidos en la técnica.

15 Además, los RPBLA se pueden usar terapéuticamente ellos mismos. Por ejemplo, el uso de RPBLA en vacunas e inóculos se ha descrito previamente en la solicitud publicada de EE UU No. 2007/0243198.

Los RPBLA se pueden usar como el inmunógeno de un inóculo o vacuna en un paciente humano u otro huésped animal adecuado tal como un chimpancé, ratón, rata, caballo, oveja, bovino, perro, gato o similar. Un inóculo puede inducir una respuesta de células B o células T (estimulación) tal como la producción de anticuerpos que inmunorreaccionan con el epítipo inmunogénico o determinante antigénico, o la activación de células T a tal epítipo, mientras que una vacuna proporciona protección contra la entidad de la que deriva el inmunógeno mediante una o ambas de una respuesta de células B o células T.

25 Los RPBLA de una vacuna contemplada pueden actuar en células presentadoras de antígeno (CPA) tal como células dendríticas y monocitos/macrófagos que engullen los RPBLA y procesan su contenido. Al actuar sobre estos tipos celulares, los RPBLA pueden mejorar la administración de antígeno a células presentadoras de antígeno. Esos RPBLA también pueden mejorar el procesamiento y la presentación del antígeno a células presentadoras de antígeno.

30 Por tanto, se puede producir una vacuna o inóculo disolviendo o dispersando una cantidad inmunogénica eficaz de ensamblajes recombinantes similares a cuerpos proteicos (RPBLA) en un diluyente farmacéuticamente aceptable. Los RPBLA pueden contener una proteína de fusión recombinante que ella misma comprende dos secuencias unidas: una secuencia es una PBIS recombinante y la otra es un polipéptido biológicamente activo contra el que se va a inducir una respuesta inmunológica mediante dicha vacuna o inóculo.

35 La activación de células T se puede medir mediante varias técnicas. En la práctica normal, un animal huésped se inocula con una vacuna de RPBLA contemplada y después de ello se recogen las células mononucleares de sangre periféricas (CMSP). Esas CMSP se cultivan después in vitro en presencia del polipéptido biológicamente activo (inmunógeno de células T) durante un periodo de tiempo de aproximadamente tres a cinco días. Se ensayan después la proliferación o secreción de una citoquina tales como IL-2, GM-CSF o IFN-gamma de las CMSP cultivadas. Los ensayos para la activación de células T son bien conocidos en la técnica. Véase, por ejemplo, la patente de EE UU No. 5.478.726 y la técnica citada en ella.

45 Un inóculo o vacuna divulgados comprende una cantidad inmunológicamente eficaz de RPBLA que se disuelven o dispersan en una composición diluyente farmacéuticamente aceptable que típicamente también contiene agua. Cuando se administra a un animal huésped en el que se va a inducir una respuesta inmunológica contra el péptido biológicamente activo por la vacuna o inóculo, tal como un animal huésped en necesidad de inmunización o en el que se desea que se induzcan anticuerpos, tal como un mamífero (por ejemplo, un ratón, pero, cabra, oveja, caballo, bovino, mono, simio o ser humano) o ave (por ejemplo, un pollo, pavo, pato o ganso), un inóculo induce anticuerpos que inmunorreaccionan con uno o más determinantes antigénicos del polipéptido diana biológicamente activo.

50 La cantidad de inmunógeno de RPBLA utilizado en cada inmunización se denomina una cantidad inmunogénicamente eficaz y puede variar ampliamente, dependiendo entre otros, del inmunógeno de RPBLA, paciente inmunizado y la presencia de un adyuvante en la vacuna. Cantidades inmunogénicamente eficaces para una (i) vacuna y un (ii) inóculo proporcionan la (i) protección o (ii) anticuerpo o actividad de células T, respectivamente, discutidas aquí anteriormente.

60 Las vacunas o inóculos típicamente contienen una concentración de inmunógeno de RPBLA desde aproximadamente 1 microgramo hasta aproximadamente 1 miligramo por inoculación (dosis unitaria) y preferiblemente desde aproximadamente 10 microgramos hasta aproximadamente 50 microgramos por dosis unitaria. El término "dosis unitaria" perteneciente a una vacuna o inóculo de la presente invención se refiere a unidades físicamente discretas adecuadas como dosis unitarias para animales, cada unidad contiene una cantidad predeterminada de material activo calculada para producir de forma individual o colectiva el efecto inmunogénico deseado en asociación con el diluyente requerido, es decir, soporte o vehículo.

Las vacunas o inóculos típicamente se preparan a partir de un inmunógeno de RPBLA recuperado dispersando el inmunógeno, en forma particulada, en un vehículo diluyente fisiológicamente tolerable (aceptable) tales como agua, solución salina, solución salina tamponada en fosfato (PBS), solución salina tamponada en acetato (ABS), solución de Ringer o similares para formar una composición acuosa. El vehículo diluyente también puede incluir materiales oleaginosos tales como aceite de cacahuete, escualano o escualeno como se discute en el presente documento posteriormente.

La preparación de inóculos y vacunas que contienen materiales proteicos como principios activos también está bien entendida en la técnica. Típicamente, tales inóculos o vacunas se preparan como parenterales, sea como soluciones o suspensiones líquidas; también se pueden preparar formas sólidas adecuadas para solución en, o suspensión en, líquido antes de la inyección. La preparación también se puede emulsionar.

Los RPBLA inmunogénicamente activos con frecuencia se mezclan con excipientes que son farmacéuticamente aceptables y compatibles con el principio activo. Los excipientes adecuados son, por ejemplo, agua, solución salina, dextrosa, glicerol, etanol o similares y combinaciones de los mismos. Además, si se desea, un inóculo o vacuna puede contener cantidades minoritarias de sustancias auxiliares tales como agentes humectantes o emulsionantes, agentes reguladores del pH que aumentan la eficacia inmunogénica de la composición.

XI. Uso de los RPBLA como productos alimenticios

Los RPBLA también se pueden usar para aumentar el valor nutricional de productos alimenticios. Por ejemplo, proteínas de fusión que comprenden una PBIS recombinante y una proteína o péptido con valor nutricional (por ejemplo, péptidos enriquecidos en aminoácidos esenciales) y se pueden expresar en un producto alimenticio. Tales proteínas de fusión pueden, en algunas divulgaciones, incluir una PBIS recombinante que sea menos alergénica que una PBIS de tipo salvaje. Los autores de la presente invención han observado que la PBIS según la presente invención muestra una alergenicidad predicha reducida cuando se compara con la PBIS de zeína no modificada (RX3). Además, basado en esta observación, los autores de la presente invención también han desarrollado PBIS nuevas que muestran alergenicidad reducida, denominados de aquí en adelante RX3-LA1 y RX3-LA2.

Por tanto, la divulgación también se refiere a una versión no alergénica de la PBIS, en donde dicha PBIS se selecciona del grupo de RX3-LA1 (SEQ ID NO: 60) y RX3-LA2 (SEQ ID NO: 61).

RX3-LA1 (SEQ ID NO:60)

THTSGGCGCA PPPPAAAPPP AAAPPPAAAP PPAAPPPAA APPPAAAPPP AAAPPPAAAP PPPCHYPTAP
ARPAPAPAPA PCPCAPAPS PCQ

RX3-LA2 (SEQ ID NO:61)

THTSGGCGCA PPPPAAAPPP AAAPPPAAAP PPAAPPPAA APPPAAAPPP AAAPPPAAAP PPPCHYPPAP
APPAPAPPAP ACPCAPAPS PCQ

En algunas divulgaciones, la PBIS recombinante es menos alergénica que una PBIS de tipo salvaje, por ejemplo, la PBIS de la gamma-zeína del maíz.

La alergenicidad se puede determinar según métodos descritos en el presente documento u otros métodos conocidos en la técnica. Por ejemplo, se ha usado la homología de la secuencia de aminoácidos para evaluar el potencial alergénico. Se pueden usar comparaciones de homología de secuencia para determinar el grado al que una proteína recién expresada o identificada es similar en estructura a un alérgeno conocido. Esta información puede predecir si esa proteína tiene potencial alergénico. Estas comparaciones se pueden realizar usando varios algoritmos tal como FASTA o BLASTP para predecir similitudes estructurales globales. La reactividad cruzada de IgE entre la proteína recién expresada o identificada y un alérgeno conocido se puede considerar una posibilidad cuando hay más del 35 por ciento de identidad en un segmento de 80 o más aminoácidos (Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) y la Organización Mundial de la Salud (OMS), 2001). También se pueden usar otros criterios científicamente justificados para predecir la reactividad cruzada de IgE.

En algunas divulgaciones la alergenicidad de una PBIS recombinante se determina usando la Base de Datos Allergen Online (versión 10.0, Enero 2010; <http://www.allergenonline.com>) desarrollada por Food Allergy Research and Resource Program. Por tanto, en algunas divulgaciones, la PBIS recombinante contiene menos resultados con más del 35% de identidad a péptidos alergénicos que la PBIS de la gamma-zeína de maíz en la Base de Datos Allergen Online.

En algunas divulgaciones, la PBIS recombinante contiene no más de 9 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 8 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 7 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 6 resultados con al menos el 35% de identidad

a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 5 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 4 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 3 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 2 resultados con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 1 resultado con al menos el 35% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante no contiene ningún resultado con al menos el 35% de identidad a un péptido alergénico conocido.

En algunas divulgaciones, la PBIS recombinante contiene no más de 9 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 8 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 7 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 6 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 5 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 4 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 3 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 2 resultados con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante contiene no más de 1 resultado con al menos el 30% de identidad a un péptido alergénico conocido. En algunas divulgaciones, la PBIS recombinante no contiene ningún resultado con al menos el 30% de identidad a un péptido alergénico conocido.

Se han desarrollado varias técnicas para determinar la respuesta alergénica desarrollada después de la administración de un alergen. Por ejemplo, la presencia de IgE, una inmunoglobulina asociada con alergenidad, se puede determinar por ensayos de ELISA y por pruebas competitivas de ELISA. Kim et al., *Yonsei Medical Journal* 47: 505-12 (2006); Fritsché, *Toxicology letters* 140-141:303-309 (2003). En algunas divulgaciones, la PBIS recombinante genera una respuesta de anticuerpos IgE disminuida in vivo comparada con la PBIS de la gamma-zeína de maíz. En algunas divulgaciones, la PBIS recombinante genera una respuesta disminuida en una prueba de punción comparada con la PBIS de la gamma-zeína de maíz.

Además, se ha observado resistencia a digestión con pepsina en varios alergenios alimenticios. Por tanto, existe una correlación entre resistencia a digestión con pepsina y potencial alergénico. El método esbozado en la Farmacopea de los Estados Unidos (1995) se usó en el establecimiento de la correlación. Astwood et al. *Nat Biotechnol* 14:1269-73 (1996). Por tanto, la resistencia de una proteína a degradación en presencia de pepsina en condiciones apropiadas indica que una proteína puede tener potencial alergénico. En algunas divulgaciones, la PBIS recombinante es menos resistente a degradación proteica en presencia de pepsina que la PBIS de la gamma-zeína de maíz.

Se pueden realizar análisis adicionales para determinar la probabilidad del potencial alergénico de la proteína recién expresada o identificada. También se han descrito procedimientos *ex vivo* como prueba para alergenidad usando cultivo de células o tejido. Evaluation of Allergenicity of Genetically Modified Foods: Report of Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology; Organización de las Naciones Unidas para Alimentación y Agricultura (FAO); Roma Italia (2001). Una de estas técnicas se basa en el hecho de que se puede medir la capacidad de provocar células cebadas mediada por IgE de epítomos alergénicos usando un ensayo in vitro funcional. Basado en células cebadas peritoneales de ratas sensibilizadas pasivamente con IgE de rata específica y marcadas con 3H-serotonina, las células se provocan para liberación de mediadores con diluciones estándar del alergen. Fritsché, *Toxicology letters* 140-141:303-309 (2003). Por tanto, en algunas divulgaciones, la PBIS recombinante genera liberación de mediadores disminuida comparado con la PBIS de la gamma-zeína de maíz.

La reacción alérgica dependiente de IgE está compuesta de dos fases. La primera fase es un paso de inducción, donde el sistema inmunitario del huésped se sensibiliza por el alergen. Como resultado, se producen anticuerpos IgE anti-alergen específicos, y los anticuerpos se fijan después por las células cebadas en órganos diana. La segunda fase es una fase de desencadenamiento mediada por la unión del alergen a estos anticuerpos IgE y estimulando la liberación de mediador (histamina) de las células cebadas. Para evaluar la alergenidad de antígenos, se pueden examinar una o ambas fases mediante pruebas in vivo apropiadas, esencialmente como se describe en Fritsché, *Toxicology letters* 140-141:303-309 (2003). Por tanto, por ejemplo, en algunas divulgaciones, la PBIS recombinante genera una producción de anticuerpos IgE disminuida comparada con la PBIS de la gamma-zeína de maíz. En algunas divulgaciones, la PBIS recombinante genera una liberación de mediador disminuida de células cebadas comparada con la PBIS de la gamma-zeína de maíz.

La divulgación también proporciona alimentos, suplementos alimenticios, preparaciones cosméticas, formulaciones nutricionales que contienen los RPBLA según la invención o materiales derivados de células u organismos

huéspedes que comprenden dichos RPBLA. Los RPBLA pueden estar formados por la PBIS como tales o por proteínas de fusión que comprenden los RPBLA. En una divulgación preferida, el RPBLA que forma parte del producto alimenticio se forma mediante el ensamblaje de una proteína de fusión que comprende una PBIS seleccionada del grupo de PA, RX3(A3), R8(4C), RX3-LA1, RX3-LA2 y una combinación de las mismas.

5 El término "producto alimenticio" como se usa en el presente documento, se refiere a una sustancia o composición que es adecuada para consumo humano y/o animal. El término "producto alimenticio" como se usa en el presente documento puede significar un producto alimenticio en una forma que está lista para consumir. De forma alternativa o adicional, sin embargo, el término producto alimenticio como se usa en el presente documento puede significar uno o más materiales alimenticios que se usan en la preparación de un producto alimenticio. A modo de ejemplo sólo, el término producto alimenticio abarca tanto productos horneados producidos de masa como la masa usada en la preparación de tales productos horneados.

15 El producto alimenticio según la presente invención puede derivar de todo o una parte de los organismos recombinantes que expresan los RPBLA según la invención. Por ejemplo, el producto alimenticio puede derivar del endospermo de una planta que expresa el RPBLA. Los productos alimenticios y suplementos alimenticios que contienen el material propagante de la invención, o partes del mismo, pueden incluir alimentos basados en cereales, por ejemplo, cereales de desayuno, harinas y alimentos que contienen estas harinas, por ejemplo, panes, pan rallado, masa líquida, tartas, pasteles, galletas, productos de panadería y pasta. Además, también se proporcionan alimentos y suplementos alimenticios que contienen el material propagante de vegetales, o partes de los mismos, por ejemplo, tubérculos y ñames. El producto alimenticio se puede, por ejemplo, seleccionar de (a) un alimento infantil o leche maternizada, o (b) un producto de panadería (por ejemplo, un pan, productos con levadura o una tarta); (c) un producto de suministro de panadería (por ejemplo, crema o un relleno o recubrimiento de pastelería); (d) una masa líquida; o (e) un empanado; o (f) un cereal; o (g) un dulce, o (h) un saborizante o bebida en emulsión; o (i) un relleno de fruta; o (j) una salsa concentrada, sopa, salsa o espesante alimenticio; o (k) una comida o componente de comida; o (l) un producto cárnico; o (m) un pienso para mascotas; o (n) un fármaco o nutracéutico, (o) un producto de patata o ñame; o (p) un producto lácteo (por ejemplo un postre o yogur); o (q) un aliño de ensalada; o (r) un aperitivo o galleta; o (s) una pasta untada; o (t) un producto de pasta (por ejemplo fideos).

30 Las plantas transgénicas de esta divulgación como se han descrito anteriormente son aplicables a seres humanos, así como a animales domésticos, aves de corral, mascotas y otros animales.

La invención se describe en detalle a continuación por medio de los siguientes ejemplos que se deben considerar como meramente ilustrativos y no limitantes del ámbito de la invención.

35

Ejemplos

Materiales y métodos

40

Construcción de plásmidos para transformación de plantas

Todas las construcciones de plásmidos se hicieron clonando un fragmento de ADN que codifica una proteína de fusión mostrado en las figuras 15a y 15b mediante digestión con Sall/BamHI en el vector pC2300 (AF234315) abierto con las mismas enzimas de restricción.

45

Material vegetal

50 Se hicieron crecer plantas de tabaco (*Nicotiana benthamiana*) en una cámara de crecimiento in vitro a 24-26°C con un fotoperiodo de 16 horas. Las plantas adultas se hicieron crecer en un invernadero entre 18-28°C. La humedad se mantuvo entre el 55 y el 65% con un fotoperiodo medio de 16 horas.

Las plántulas para agroinfiltración (Vaquero *et al.*, 1999 Proc. Natl. Acad. Sci., USA 96(20):11128-11133; Kapila *et al.*, 1997 Plant Sci. 122:101-108) se hicieron crecer de semillas durante 4-6 semanas en las condiciones in vitro descritas anteriormente.

55

Agroinfiltración de tabaco por vacío

60 Se hizo crecer la cepa EHA 105 de *A. tumefaciens* que contenía una construcción deseada en medio LB (Tripton 10 g/l, extracto de levadura 5 g/l, NaCl 10 g/l) suplementado con kanamicina (50 mg/ml) y rifampicina (100 mg/l) a 28°C con agitación (250 rpm) durante la noche (aproximadamente 18 horas). Después se inocularon las agrobacterias en 30 ml de LB también suplementado con kanamicina (50 mg/ml) y rifampicina (100 mg/l). Después de cultivar durante la noche (aproximadamente 18 horas) a 28°C, las células agrobacterianas se recogieron por centrifugación durante 10 minutos a 3.000 x g y se resuspendieron en 10 ml de medio MS líquido con MES (Sigma Chemical) 4,9 g/l y sacarosa 30 g/l a pH 5,8. El cultivo bacteriano se ajustó a una DO600 final de 0,1 para la

65

agroinfiltración. Después, el cultivo celular se suplementó con acetosiringona a una concentración final de 0,2 mM y se incubó durante 90 minutos a 28°C (Torrent, M., Llop-Tous, I. y Ludevid, M.D. (2009) En Recombinant Proteins from plants. Methods in Molecular Biology. Vol. 483. Ed por Gomord V y Faye L, Springer Verlag, Humana Press, Heidelberg; pp. 193-208; Voinnet, O., Rivas, S., Mestre, P. y Baulcombe, D. (2003) Plant J. 33, 949-956). Se mezclaron cultivos individuales de *Agrobacterium* que llevaban las construcciones de RX3 y las construcciones supresoras de silenciador HC-Pro (Goytia et al., 2006). Las plantas se cubrieron totalmente con la suspensión y se aplicó vacío (100 KPa) durante 5-6 segundos. La suspensión se retiró y las plantas se mantuvieron en el invernadero. El material vegetal se recuperó y la extracción de proteína total se analizó mediante inmunotransferencia usando un anticuerpo apropiado.

Agroinfiltración de tabaco con jeringuilla

Se hizo crecer la cepa EHA 105 de *Agrobacterium tumefaciens* a 28°C en caldo L suplementado con kanamicina 50 µg ml⁻¹ y rifampicina 50 µg ml⁻¹ hasta fase estacionaria. Las bacterias se sedimentaron por centrifugación a 5000 g durante 15 minutos a temperatura ambiente y se resuspendieron en tampón MES 10 mM pH 5,6, MgCl₂ 10 mM y acetosiringona 200 µM hasta una DO600 final de 0,2. Las células se dejaron en su medio durante 3 horas a temperatura ambiente. Se mezclaron cultivos individuales de *Agrobacterium* que llevaban las construcciones de RX3 y las construcciones supresoras de silenciador HC-Pro (Goytia et al., 2006) y se infiltraron en la cara abaxial de hojas de plantas de *Nicotiana benthamiana* de 2-4 semanas de edad.

Extracción de proteína y proteína total o análisis de inmunotransferencia

Se extrajeron las proteínas solubles totales (PST) de hojas transformadas en tampón Tris-HCl 100 mM pH 7,5 que contenía NaCl 100 mM, SDS al 0,5% y DTT 200 mM durante 1 hora a temperatura ambiente. Los extractos resultantes se centrifugaron a 10.000 x g durante 30 minutos a 4°C y las PST se separaron en geles de SDS-poliacrilamida. Las proteínas se detectaron por tinción con azul de Coomassie o tinción con plata o por inmunotransferencia usando el antisuero anti-R8 (Torrent, M., Llompert, B., Lasserre-Ramassamy, S., Llop-Tous, I., Bastida, M., Marzábal, P., Westerholm-Pavinen, A., Saloheimo, M., Heifetz, P. B. y Ludevid, M.D. (2009) BMC Biology 7, 5), un anti-GFP o un anti-hGH producidos en conejos inyectados con una proteína recombinante GFP o hGH expresada y purificada de células de *E. coli* y un anti-EGF comercial (Abcam).

Fraccionamiento subcelular y determinación de la densidad de RPBLA

Se molieron tejidos de hojas de tabaco agroinfiltrados en un mortero a 0°C en el tampón de homogenización HB (Tris 10 mM pH 8,0, sacarosa 0,25 M e inhibidores de proteasas). El homogeneizado se filtró a través de dos capas de Miracloth (22-24 micrómetros, Calbiochem) para eliminar restos de tejido antes de centrifugar a 50 x g durante 5 minutos a 4°C. Los homogeneizados clarificados resultantes de los diferentes tejidos se cargaron en gradientes de densidad multicapa de iodixanol (Optiprep, Sigma) (capas preferidas para el cálculo de densidad: 1,11, 1,17, 1,19, 1,21, 1,23 y 1,25 g/cm³) tamponados con el tampón HB. Los gradientes se centrifugaron a 4°C durante 2 horas a 80.000 x g en un rotor SW40 Ti de Beckman. Se analizaron alícuotas equivalentes del sobrenadante, fracciones de interfase y precipitado mediante SDS-PAGE e inmunotransferencia usando anticuerpos específicos.

Como se indica anteriormente, esta técnica es una técnica adecuada para determinar la densidad de los RPBLA. La siguiente tabla proporciona una relación detallada entre el % de iodixanol (peso/volumen) en el colchón determinado y la densidad correspondiente:

Porcentaje de iodixanol (peso/volumen)	Densidad (g/ml)	Porcentaje de iodixanol (peso/volumen)	Densidad (g/ml)
8	1,069	30	1,175
10	1,079	32	1,185
12	1,088	34	1,194
14	1,098	36	1,204
16	1,107	38	1,214
18	1,117	40	1,223
20	1,127	42	1,233
22	1,136	44	1,243
24	1,146	46	1,252
26	1,156	48	1,262
28	1,165	50	1,272

Tabla 3: Relación entre el % de iodixanol (peso/volumen) en el colchón determinado y la densidad correspondiente.

Inmunocitoquímica e imágenes: microscopía confocal

Se montaron secciones de tejido de hoja de tabaco transformado con secuencias fluorescentes en agua para observación confocal directa. Se obtuvieron micrografías usando el microscopio confocal de barrido laser Leica TCS SP (Heidelberg, Alemania). Se recogieron imágenes fluorescentes verdes después de excitación a 515 nm usando una ventana de emisión de 530-630 nm. Se recogieron imágenes fluorescentes cian a una excitación de 458 nm y una ventana de emisión de 470-530 nm. Se recogieron imágenes de proteína fluorescente verde a excitación de 488 nm con el láser iónico de argón usando una ventana de emisión ajustada a 495-535 nm. Se recogieron imágenes fluorescentes rojas después de excitación a 543 nm con un laser de HeNe y una ventana de emisión de 550-600 nm. Las secciones ópticas tenían 0,5 μm de espesor. Se registraron imágenes y proyecciones digitales usando el software del microscopio confocal. Las imágenes presentadas en las figuras de microscopía son representativas de al menos cinco experimentos independientes.

Determinación del número y tamaño de los RPBLA por microscopía confocal

Se determinó la distribución por tamaño (en porcentaje relativo) de los PB durante el tiempo midiendo los diámetros aparentes de aproximadamente 500 PB por tiempo (2, 4, 7 y 10 dpi). Se analizaron tres plantas transformadas independientes por tiempo y se usó el software Olympus fluoview v. 1.6a para medir los PB fluorescentes observados en el microscopio confocal FV1000 de Olympus. Para la determinación del número de PB, se usaron proyecciones confocales correspondientes a $10^5 \mu\text{m}^3$ (70 x 70 x 20) de tejido transformado. Se analizaron cuarenta imágenes confocales de 8 plantas transformadas independientes por tiempo. Los resultados se analizaron estadísticamente mediante análisis de varianza (ANOVA) unidireccional y la prueba de comparaciones múltiples de Bonferroni ($p < 0,05$ se consideró significativamente diferente).

Aislamiento de RPBLA a centrifugación de baja densidad

Se molió aproximadamente 1 gramo de tejidos de hoja de tabaco agroinfiltrados en un mortero a 0°C en 5 ml de tampón de homogenización PBP3 (Tris 100 mM pH 8,0, KCl 50 mM, MgCl₂ 6 mM, EDTA 10 mM y NaCl 0,5 M). El homogeneizado (H0) se filtró a través de una capa de Miracloth (22-24 micrómetros, Calbiochem). Los homogeneizados clarificados resultantes de los homogeneizados preclarificados (H1) se centrifugaron a baja velocidad (1.500 x g) durante 10 minutos y se analizaron el precipitado y sobrenadante (SN) resultantes. El precipitado de la centrifugación a baja velocidad se resuspendió mediante sonicación moderada (ciclo del 50%, control de salida 3, 10 segundos, Branson sonifier 250) en 3-5 ml de tampón de lavado (Triton® X-100 al 0,5%) y por último se incubaron durante 15-60 minutos a temperatura ambiente. Después de una segunda centrifugación a baja velocidad, se analizaron cantidades equivalentes del precipitado (wPB) y sobrenadante (Ws) resultantes.

Purificación de EGF

Se solubilizó una fracción wPB de 80 gramos (peso fresco) de hojas de planta de tabaco agroinfiltradas con Zera(E)-EGF como se indica anteriormente y se cortó con 30 microlitros de FXa (Qiagen) en presencia de CaCl₂ 2 mM a 37°C. La reacción se paró después de 3 horas añadiendo EDTA 50 mM. La muestra se diluyó luego 5 veces en tampón A. El EGF se purificó por medio de una columna de 3 ml de fase inversa Resource de Amersham. Se realizó el gradiente de acetonitrilo con 20 volúmenes de columna con tampón A (acetato 10 mM pH 4, bME 2 mM y acetonitrilo al 5%) y tampón B (acetato 10 mM pH 4, bME 2 mM y acetonitrilo al 75%). Se recuperó EGF puro en dos fracciones a aproximadamente el 40% de acetonitrilo y se dializó contra Tris 50 mM pH 8, bME 2 mM y NaCl 100 mM durante la noche a 4°C en una membrana de diálisis de 3,5 kDa.

Ensayo de actividad de EGF

Se sembraron células de la línea celular de carcinoma epitelial humano (A431), que sobreexpresa el receptor de EGF, en placas (P-35) a $0,5 \times 10^5$ células/placa. Las células se incubaron 48 horas en medio de cultivo (MEMx1, glutamina 2 mM, aminoácidos no esenciales al 1%) y SBF (suero bovino fetal) al 10%. Después de ello se ayunaron durante la noche en medio de cultivo sin SBF. A continuación, se añadieron el EGF estándar (de 0 a 10 ng EGF/ml) de Promega y las muestras correspondientes (RX3(E)-EGF y EGF solubilizados) en el mismo intervalo de concentraciones y las células se incubaron durante 9 minutos. Después, las células se lavaron dos veces con PBS frío y se congelaron en N₂ líquido para detener el metabolismo celular. Para analizar la fosforilación del receptor de EGF, las células se levantaron y se analizaron cantidades equivalentes de proteína total mediante el kit de ELISA sándwich PathScan® Phospho-EGF Receptor (Tyr1068) esencialmente como describe el fabricante.

Ejemplo 1: Acumulación de RX3-GFP, RX3(R)-GFP, RX3(K)-ECFP, RX3(A)-GFP y RX3(L)-GFP en RPBLA de plantas de tabaco transformadas

Se ha descrito que la anfipaticidad del dominio de repetición de RX3 es esencial en el autoensamblaje y la formación de RPBLA (Ludevid *et al.*, Plant Mol. Biol. (1984) 3:277-234; Kogan *et al.*, J. Mol. Biol. (2001) 312:907-913). Para caracterizar la importancia de la anfipaticidad en la capacidad de inducción de RPBLA, se analizaron un conjunto de etiquetas de RX3 fusionadas a una proteína fluorescente indicadora (GFP o ECFP) (Fig. 1A): RX3 nativo con histidinas (PPPVHL)₈ (SEQ ID NO:62) [RX3], dos variantes de RX3 con un carácter anfipático aumentado obtenidas

por la sustitución de las histidinas por arginina (PPPVRL)₈ (SEQ ID NO:63) [RX3(R)] y lisina (PPPVKL)₈ (SEQ ID NO:64) [RX3(K)] y dos variantes de RX3 con un RD completamente hidrofóbico donde las histidinas se cambiaron por alanina (PPPVAL)₈ (SEQ ID NO:66) [RX3(A)] y leucina y (PPPVLL)₈ (SEQ ID NO:67) [RX3(L)].

5 Como se esperaba, las plantas de tabaco transformadas con RX3-GFP acumularon un gran número de RPBLA fluorescentes de forma redonda de aproximadamente 1 micrómetro de diámetro en la célula. Sorprendentemente, a pesar de la ausencia de un RD anfipático, las proteínas de fusión RX3(A)-GFP y RX3(L)-GFP también se retuvieron dentro del retículo endoplásmico (RE) de la célula y formaron RPBLA (Fig. 1C).

10 La proteína de fusión RX3(A)-GFP se acumuló en grandes RPBLA, que eran ligeramente mayores que los RPBLA obtenidos de la expresión de RX3-GFP. Los RPBLA de RX3(L)-GFP medían aproximadamente 0,5 micrómetros. No se observó fluorescencia en la superficie de la célula, lo que indica que la proteína se acumuló eficazmente en RPBLA y no se secretaba. Estos estudios también indican que las etiquetas de RX3 con un RD completamente hidrofóbico son capaces de ensamblarse e inducir la formación de RPBLA en plantas.

15 Basado en los sorprendentes resultados obtenidos con las etiquetas de RX3 que contenían RD hidrofóbico, se decidió caracterizar más la importancia de la anfipaticidad de RD de las etiquetas de RX3 analizando la acumulación de RX3 anfipático en la que todas las histidinas se cambiaron por argininas [RX3(R)] o lisinas [RX3(K)]. Arginina y lisina tiene un pKa alto (12 y 10,5, respectivamente) comparadas con histidina. De esta manera, estas sustituciones produjeron un RD con mayor carga neta positiva al valor de pH del retículo endoplásmico, una carga neta positiva aumentada y una anfipaticidad aumentada del RD. Sorprendentemente, el análisis de microscopía confocal mostró fluorescencia muy baja al usar estas construcciones. Cuando se aplicaron las mismas condiciones usadas para las otras variantes de RX3 a RX3(R)-GFP y RX3(K)-ECFP, no se observó señal significativa. De forma similar, se observó baja acumulación de proteínas de fusión RX3(R)-GFP y RX3(K)-ECFP mediante inmunotransferencia usando un anticuerpo anti-GFP. Como se muestra en la figura 1B, RX3(R)-GFP se acumuló poco comparada con RX3-GFP, y no se observó proteína de fusión RX3(K)-ECFP. De forma interesante, en los homogeneizados que expresan RX3(K)-ECFP, sólo se observó una banda de mayor movilidad electroforética que era inmunorreactiva con el anticuerpo anti-GFP. Esta banda probablemente corresponde a ECFP parcialmente degradada (Fig. 1B, carril 4, punta de flecha negra). Las imágenes muy expuestas de hojas de tabaco agroinfiltradas con pRX3(R)-GFP o pRX3(K)-ECFP, mostraron que la proteína no se retiene eficazmente en los RPBLA derivados del RE. La expresión de RX3(R)-GFP produjo la secreción de la mayoría de la proteína de fusión y solo se acumularon en el interior de la célula unos pocos RPBLA muy pequeños (Fig. 1C). Respecto a RX3(K)-ECFP los resultados fueron incluso más sorprendentes, porque no se observaron RPBLA y la fluorescencia se asociaba a cloroplastos y secreción (Fig. 1C, recuadro). Basado en este experimento y en el hecho de que parece que la mayoría de la proteína de fusión se degrada, es posible que la fluorescencia asociada con cloroplastos se deba a ECFP parcialmente degradada. Es poco probable que RX3(K)-ECFP se transporte a este orgánulo.

En conjunto, estos resultados indican que, al contrario de lo que se sugiere en la bibliografía, la anfipaticidad de RX3 no es un elemento clave necesario para el ensamblaje de proteínas e inducción de RPBLA. Además, el aumento de la anfipaticidad de RD por la sustitución de histidina con arginina o lisina reduce significativamente la capacidad de estos péptidos de inducir la formación de RPBLA.

Ejemplo 2: Caracterización de la densidad de RPBLA inducidos por RX3-GFP, RX3(A)-GFP y RX3(L)-GFP en plantas de tabaco

45 Una característica de los RPBLA es su alta densidad, que se puede determinar mediante gradientes de densidad en capas (Torrent, BMC Biology 2009, 7:5). En el presente estudio, se llevaron a cabo gradientes de Optiprep™ cargando homogeneizados de plantas filtrados en los siguientes colchones de Optiprep™:

Optiprep™ %	Densidad (g/ml)
18	1,117
30	1,175
34	1,194
38	1,214
42	1,233
46	1,252

50 Tabla 4: Tabla de correspondencia de concentración de Optiprep (%) y densidad (g/ml).

Los microsomas (RE, Golgi, etc.) sedimentan en la fracción f18, la interfase por encima del colchón de baja densidad con el 18% de Optiprep™. Típicamente, los RPBLA son más densos que los microsomas y se recuperan de fracciones más densas del 26-28% de Optiprep™.

55 Como se muestra en la figura 2, cuando el homogeneizado de una planta de tabaco que expresa la proteína de fusión RX3-GFP se ultracentrifugó en el gradiente densidad en capa determinado, como se esperaba, la mayoría de

la proteína se recuperó en las fracciones densas f34 y f38. Este resultado demuestra que la proteína de fusión se ensambla estrechamente dentro del RE y forma RPBLA densos. También se observa algo de RX3-GFP en la fracción de baja densidad f18 que corresponde a microsomas, pero esto probablemente representa proteína de fusión recién sintetizada que aún no está totalmente ensamblada en estructuras densas.

5 Como se ha mencionado anteriormente, dos de las construcciones analizadas tenían un RD completamente hidrofóbico (RX3(A)-GFP y RX3(L)-GFP) y ambos inducían la formación de RPBLA dentro de la célula. Cuando se analizó RX3(A)-GFP mediante gradiente de densidad, la mayoría de la proteína de fusión se acumuló en fracciones densas (f30 y f34), confirmando que RX3(A)-GFP se acumula en RPBLA densos. Se obtuvieron resultados similares cuando se analizaron homogeneizados clarificados de hojas de tabaco transformadas con RX3(L)-GFP. En ese caso, incluso aunque los RPBLA eran menos densos que los inducidos por la expresión de RX3(A)-GFP, se recuperó una fracción significativa de los mismos en fracciones densas (Fig. 2B; f30 y f34).

15 Estos datos muestran que, a pesar de la ausencia de un RD anfipático, RX3(A)-GFP y en algún grado RX3(L)-GFP, son capaces de autoensamblarse eficazmente en RPBLA que son lo suficientemente densos para ser aislados de la mayoría de los contaminantes celulares. El análisis por tinción con azul de Coomassie de volúmenes equivalentes de las fracciones recuperadas de un gradiente de densidad cargado con homogeneizado de RX3(A)-GFP se muestra en la figura 2B. Se puede ver que mientras que la mayoría de los contaminantes celulares se recuperan en las fracciones S y f18, f30 y f34 contenían una fracción muy pura de RX3(A)-GFP.

20 **Ejemplo 3: Procedimiento posterior para el aislamiento de RPBLA de plantas de tabaco mediante centrifugación a baja velocidad**

25 La centrifugación en gradiente de densidad es un procedimiento analítico apropiado para la determinación de la densidad de los RPBLA, pero no es adecuado para purificar estos orgánulos a gran escala. Se ha desarrollado un proceso posterior basado en una simple centrifugación a baja velocidad y algunos pasos de lavado del sedimento recuperado (véase los procedimientos experimentales) que permite la recuperación de una fracción limpia de RPBLA.

30 Este sencillo proceso posterior se aplicó a plantas de tabaco transformadas con RX3-GFP y RX3(A)-GFP, y se analizaron cantidades equivalentes de los diferentes pasos del proceso mediante inmunotransferencia y SDS-PAGE teñido con azul de Coomassie. Como se muestra en la figura 3A, casi toda la proteína de fusión presente en los homogeneizados correspondientes (H0) se recuperó en la fracción de RPBLA (wPB). El rendimiento en ambos casos fue similar y muy alto ya que no se pierde proteína de fusión en el sobrenadante (SN) y paso de lavado (Ws). Esta observación se confirmó por inmunotransferencia. El hecho de que se recupere más proteína de fusión RX3(A)-GFP que RX3-GFP (Fig. 3A, compárense las fracciones wPB) de la misma cantidad de biomasa procesada no refleja un rendimiento de recuperación mejor, sino una mayor velocidad de acumulación de RX3(A)-GFP en las hojas de tabaco. La eficaz recuperación es un resultado bastante sorprendente dado que los RPBLA de RX3(A)-GFP son menos densos que RX3-GFP.

40 El enriquecimiento alcanzado en el proceso también es muy alto. La mayoría de los contaminantes presentes en el homogeneizado (H0) se eliminan durante el proceso y la fracción correspondiente a los RPBLA lavados (wPB) contenía casi exclusivamente la proteína de fusión. Se observan varias bandas en la fracción wPB, pero corresponden a varios estados de oligomerización de la misma proteína de fusión y no a impurezas proteicas.

45 La recuperación posterior de los RPBLA de RX3(L)-GFP se adaptó centrifugando el homogeneizado correspondiente a una velocidad mayor (4.000-5.000 x g) para recuperar más eficazmente los orgánulos pequeños y menos densos que contenían esta proteína de fusión (Fig. 2). Sorprendentemente, incluso para RX3(L)-GFP, el rendimiento fue muy eficaz y la mayoría de la proteína de fusión se recuperó en la fracción de los RPBLA (wPB), como se puede observar en la figura 3B (panel derecho). El hecho de que se necesite una velocidad mayor para recuperar los RPBLA que contienen RX3(L)-GFP no afecta significativamente el enriquecimiento del proceso como se demostró mediante geles de SDS-PAGE teñidos con azul de Coomassie (Fig. 3B, panel izquierdo). Estos resultados confirman que etiquetas de RX3 con un RD completamente hidrofóbico son adecuadas para la producción y purificación de proteínas de proteínas mediante aislamiento de RPBLA.

55 **Ejemplo 4: Solubilización de proteínas de fusión de RX3**

60 La solubilización de proteínas de fusión basadas en RX3 de la fracción de los RPBLA es uno de los pasos clave en el procesamiento posterior de los RPBLA. Se usaron las fracciones de RPBLA recuperadas mediante centrifugación a baja velocidad esencialmente como se muestra en el ejemplo 3 para comparar la solubilidad de RX3-GFP y RX3(A)-GFP. Se esperaba que la sustitución de histidinas por alaninas en el RD disminuyera la solubilidad de las proteínas de fusión en solución acuosa porque se reduce la polaridad total de la proteína, pero sorprendentemente, RX3(A)-GFP era mucho más soluble en cada una de las condiciones ensayadas.

Como ejemplo, en la figura 3A se comparan fracciones solubilizadas (sPB) con no solubilizadas (uPB) después de incubar la fracción limpia de RPBLA (wPB) durante 4 horas en tampón moderado (Tris 50 mM pH 8, TCEP 5 mM y 2bME 10 mM). Justo después del paso de solubilización, la muestra se centrifugó a 16.000g durante 10 minutos y se recuperó sPB como sobrenadante. Se recuperó uPB como precipitado.

5 Como se esperaba, RX3-GFP sólo se solubilizó parcialmente. La mayoría de la proteína de fusión en forma monomérica se solubiliza en las condiciones dadas (Fig. 3A, compárense sPB frente a uPB, punta de flecha), pero una gran cantidad de las formas multiméricas estrechamente agregadas de la proteína de fusión permanece en la fracción no solubilizada (Fig. 3A, asterisco). Por el contrario, en las mismas condiciones de solubilización
10 prácticamente toda la proteína RX3(A)-GFP se solubilizó, independientemente de la forma de agregación. Este sorprendente resultado, que permite un rendimiento muy alto al recuperar la proteína de fusión de los RPBLA lavados se confirmó con un amplio espectro de péptidos ensambladores completamente hidrofóbicos fusionados a varias proteínas como se muestra posteriormente.

15 **Ejemplo 5: Acumulación de RX3(A3)-ECFP en RPBLA de plantas de tabaco transformadas**

En experimentos previos todas las nuevas variantes de RX3 consistían en mutaciones únicas de las histidinas presentes en el RD, de modo que el papel de los residuos de valina y lisina antes y después de las histidinas no se había analizado aún. Para hacer eso, todos los residuos de valina, histidina y leucina presentes en el RD se
20 cambiaron por alaninas en una nueva variante no anfipática de RX3 (RX3(A3)) (Fig. 4A). Se transformaron hojas de tabaco con un vector que expresaba el péptido RX3(A3) fusionado a ECFP y se analizó la formación de RPBLA mediante microscopía confocal 3 y 6 días después de la infiltración (dpi). Incluso en los primeros días después de la infiltración (3 dpi), se observó la presencia de un gran número de RPBLA pequeños, lo que sugiere que la variante RX3(A3) era muy eficaz. Esta observación se confirmó a 6 dpi por la presencia de un inmenso número de RPBLA
25 grandes organizados en grupos (Fig. 4B) lo que demuestra que este nuevo péptido ensamblador es inesperadamente más eficaz que RX3 salvaje.

El análisis de la expresión de RX3(A3)-ECFP en plantas indica que los residuos de valina y leucina presentes en el RD no son necesarios para el ensamblaje de proteínas y formación de RPBLA. Una consecuencia clara de esta
30 conclusión es que las valinas, histidinas y leucinas presentes en el RD se pueden sustituir por cualquier otro aminoácido hidrofóbico, siempre que la estructura del RD no esté desplegada. Los siguientes experimentos demuestran que estos residuos también se pueden sustituir por aminoácidos polares sin carga y cargados negativamente.

35 **Ejemplo comparativo 6: Acumulación de RX3(E)-GFP y RX3(D)-ECFP en RPBLA de plantas de tabaco transformadas**

Como se menciona anteriormente, el cambio de los residuos de histidina del RD por lisina o arginina aumenta la carga neta y por consiguiente la anfipaticidad de este dominio. Sorprendentemente, también reduce drásticamente la
40 eficacia del péptido ensamblador en términos de formación de RPBLA y en acumulación total de proteína. Basado en estos resultados, también se esperaba que la sustitución de histidinas por ácido aspártico y glutámico produjera una reducción en la acumulación en RPBLA, porque la adición de estos residuos también aumenta la carga neta del RD, por tanto, su anfipaticidad. Sin embargo, como se muestra en la figura 5B, la inmunotransferencia de extractos de proteína total de plantas de tabaco transformadas con RX3(E)-GFP y RX3(D)-ECFP y los controles
45 correspondientes (RX3-GFP y RX3-ECFP, respectivamente) indicó que las proteínas de fusión que tenían aminoácidos ácidos se acumulaban ligeramente mejor que los controles. La menor movilidad de RX3(E)-GFP y RX3(D)-ECFP se puede explicar por su mayor contenido en aminoácidos ácidos, como se ha descrito previamente para otras proteínas.

Además, las plantas de tabaco transformadas con RX3(E)-GFP y RX3(D)-ECFP indujeron la formación de un gran número de RPBLA fluorescentes grandes de forma redonda organizados en grupos (Fig. 5C). De hecho, con
50 frecuencia están presentes grandes RPBLA de 1 a 2 micrómetros en la mayoría de las células transformadas. De forma interesante, el tamaño medio de estos RPBLA era significativamente mayor que los inducidos por la expresión de GFP o ECFP fusionadas a RX3.

55 **Ejemplo 7: Acumulación de RX3(T)-ECFP, RX3(N)-ECFP y RX3(Q)-ECFP en RPBLA de plantas de tabaco**

Los resultados anteriores demuestran que la sustitución de histidina por aminoácidos hidrofóbicos o ácidos aumentaba la capacidad del péptido ensamblador de inducir los RPBLA. Además, se observó el efecto negativo de
60 aumentar la carga positiva del RD sustituyendo las histidinas por aminoácidos básicos. Sin embargo, el efecto de aminoácidos polares sin carga sobre la funcionalidad del péptido ensamblador era aún desconocido. Por tanto, se crearon tres nuevas construcciones en las que todas las histidinas del RD se cambiaron por residuos de treonina, asparragina y glutamina y se fusionaron a ECFP (RX3(T)-ECFP, RX3(N)-ECFP y RX3(Q)-ECFP, respectivamente). La adición de estos aminoácidos produce un RD que es anfipático, pero sin carga neta (Fig. 6A).

65

Se analizaron hojas de tabaco que expresaban todas estas construcciones mediante microscopia confocal 3 y 6 días después de la infiltración (dpi). A 3 dpi, se observó un gran número de RPBLA pequeños, lo que sugiere que todas estas variantes de RX3 son muy eficaces. Esta observación se confirmó a 6 dpi por la presencia de muchos RPBLA grandes organizados en grupos (Fig. 4B).

Estos resultados demuestran que la expresión en plantas de variantes de RX3 que tienen aminoácidos polares sin carga produce la formación de RPBLA.

Ejemplo 8: Acumulación de PP-ECFP y PA-ECFP en RPBLA de plantas de tabaco transformadas

Los dos dominios principales de RX3 (RD y PX) son ambas secuencias ricas en prolina, que típicamente adoptan una estructura PPII. El estudio de un RD sintético mediante dicroísmo circular confirmó la presencia de la hélice PPII en este dominio en solución acuosa (Dacol, J. Org. Chem., 1996, 61 (20), pp. 6775–6782). Las proteínas que tienen dominios PPII son abundantes en plantas y en otros organismos. Sin embargo, en general no tienen la capacidad de inducir la formación de RPBLA. Por ejemplo, colágeno (Caldwell JW, Applequist J., Biopolymers 1984; 10:1891–1904), proteínas alfa-caseína de leche (Syme CD. et al., Eur J Biochem 2002; 269:148–156.) y PKA (Knighton D.R. et al., Science 1991; 253:414–420) son ejemplos de proteínas que tienen hélices PPII pero que no tienen la capacidad de inducir RPBLA. Como resultado, se esperaba que otros elementos en RX3, además de la secuencia con estructura PPII, fueran necesarias para inducir la formación de los RPBLA.

Para identificar tales elementos en la secuencia de RX3, se sustituyeron los dominios RD y PX por una secuencia basada en prolina sin homología a RX3. La estructura PPII se mantuvo y el número y posición relativa de los residuos de cisteína no se modificó (PP; Fig. 7A). Además, se sintetizó un péptido similar, pero más hidrofóbico en el que se cambiaron los dominios RD y PX por el hexapéptido PPPAPA repetido a lo largo de la secuencia (PA; Fig. 7A). Estos dos péptidos tienen esencialmente la misma estructura PPII que RX3, pero no comparten homología primaria más que el alto contenido en prolina y el número y posición de las cisteínas.

El análisis de hojas de tabaco que sobreexpresaban RX3-ECFP, PP-ECFP o PA-ECFP por inmunotransferencia mostró que sorprendentemente los dos péptidos sintéticos fusionados a ECFP se acumulaban con mayor eficacia que RX3 natural (Fig. 7B; compárense los carriles 1 y 2 frente a 3). PP-ECFP, y en menor grado PA-ECFP, presentaban menor movilidad en el gel de SDS-PAGE que RX3-ECFP, incluso aunque las tres proteínas de fusión tienen el mismo número de aminoácidos. Esto sugiere que PP y PA presentan una hélice PPII extendida que es más estable que RX3, incluso en las condiciones de inmunotransferencia. Como consecuencia, el ensamblaje de estos péptidos sintéticos estaría favorecido y la proteína mostraría acumulación aumentada, como se determinó por inmunotransferencia. Esta hipótesis se confirmó por la observación de hojas de tabaco en el microscopio confocal seis días después de la transformación con vectores que expresaban proteínas de fusión RX3-, PP- y PA-ECFP. La figura 7C muestra imágenes confocales de hojas transformadas con PP-ECFP y PA-ECFP, que presentaban un número mayor de RPBLA comparadas con hojas que expresan RX3-ECFP. Además, el tamaño medio de los RPBLA inducidos por la expresión de PP- y PA-ECFP era dos veces mayor que el tamaño medio de los RPBLA de RX3-ECFP.

La comparación de secuencia de estas 3 secuencias inductoras de cuerpos proteicos (Fig. 7A), revela que la única región de homología entre ellas es los primeros diez aminoácidos después del péptido señal (THTSGGCGCQ) (SEQ ID NO:25). Para confirmar que esta secuencia no desempeña un papel específico en el ensamblaje de proteínas, la secuencia se cambió por una secuencia sintética basada en prolina (THPPPPCPCP) (SEQ ID NO:26). La secuencia basada en prolina mantenía las posiciones de las cisteínas en el contexto de las construcciones PA-ECFP. Esta construcción (NPA-ECFP) eran tan eficaz en la formación de RPBLA como PA-ECFP.

Además, también se cambió el péptido señal de gamma zeína de RX3 por el péptido señal de PR10, también conocido como PR-S (Veroverd et al., Plant Physiol., 1995, 109:1199-1205). La construcción resultante se fusionó a GFP (RX3(PR10)-GFP). Es improbable que el péptido señal influya en la formación de RPBLA ya que se corta cotraduccionalmente a la entrada de la proteína en el RE y como se esperaba la velocidad de acumulación y la inducción de RPBLA era completamente independiente del péptido señal usado.

El hecho de que los péptidos PP, PA, NPP y NPA no compartan homología de secuencia primaria con RX3 indica que inesperadamente, la capacidad de ensamblarse e inducir RPBLA no depende de la secuencia de RX3, sino de la estructura tridimensional de la misma. La hélice PPII extendida presente en los péptidos RX3, PP y PA, y sus variantes, es un rasgo clave responsable de la inducción de los RPBLA.

Debido a razones de clonación, los péptidos PP y PA mantienen parte de la última repetición del dominio de repetición de RX3. Para descartar que la última unidad de repetición de RX3 del RD se requiera para formar los RPBLA, se realizaron un nuevo conjunto de construcciones donde esta repetición está ausente: PP3 y PA2. Estas dos construcciones mantienen la capacidad para formar RPBLA.

Ejemplo 9: Determinación de los residuos de cisteína de RX3 implicados en la inducción o estabilización de RPBLA en plantas de tabaco transformadas

El trabajo presentado aquí proporciona revelaciones clave en rasgos específicos de RX3 que permiten su multimerización y la formación de RPBLA. Un rasgo relevante, pero no único, es la presencia de residuos de cisteína que podrían participar en los enlaces inter-disulfuro entre moléculas de RX3.

Se usaron mutaciones puntuales de cada una de las seis cisteínas como un modo directo de determinar el papel de los puentes disulfuro en la polimerización de RX3. La sobreexpresión de RX3-ECFP que carecía de C7 o C9 (RX3C7G-ECFP y RX3C9G-ECFP) claramente perturbó el proceso de multimerización de proteínas de fusión: ambos mutantes se secretan mucho (Fig. 8C-D). De hecho, en análisis de imágenes célula a célula de células epidérmicas transformadas con los mutantes individuales C7 o C9, las proteínas de fusión principalmente se secretaban y sólo unas pocas células individuales muy fluorescentes mostraban manchas fluorescentes similares a RPBLA. Esta observación sugiere que grandes tasas de expresión de la proteína recombinante en el RE producen un aumento de la eficacia de agregación, lo que indica que se necesita una concentración crítica de proteína mayor para la agregación proteica comparada con RX3 de tipo salvaje (Fig. 8C-D, recuadros). Como se esperaba, cuando se mutaron ambos residuos de cisteína al mismo tiempo (Fig. 8I; RX3C7,9G-ECFP) no se observó acumulación de RPBLA. Estos resultados demuestran que la presencia de dos residuos de cisteína cerca del extremo N-terminal del dominio RX3 es necesaria para retener la proteína en el RE para inducir la formación de RPBLA. Sin embargo, el mutante múltiple RX3C64,82,84,92G-ECFP, que mantiene exclusivamente las dos primeras cisteínas, demostró que C7 y C9 no son suficientes para retener la proteína de fusión en el RE y permitirle formar RPBLA (Fig. 8K). De esta manera, algunas de las otras cisteínas presentes en el péptido RX3 desempeñan un papel en su ensamblaje de proteínas apropiado.

La mutación del residuo de cisteína localizado en medio del dominio RX3 (C64) no tuvo efecto significativo en el proceso de oligomerización y las proteínas acumuladas en los RPBLA (Fig. 8E; RX3C64G-ECFP). Además, las mutaciones individuales de los tres residuos de cisteína localizados en el extremo C-terminal de RX3 (C82, C84 o C92) no tuvieron efecto significativo en la capacidad de acumularse en RPBLA (Fig. 8F-H; RX3C82G-ECFP, RX3C84G-ECFP o RX3C92G-ECFP, respectivamente). Las capacidades normales de agregación de estos mutantes individuales apoyan la hipótesis de que ninguno de estos residuos se requiere por sí mismo para inducir y estabilizar RPBLA. Sin embargo, puesto que el mutante múltiple de Cys C-terminales (Fig. 8J; RX3C82,84,92G-ECFP) fue incapaz de progresar con la biogénesis de RPBLA, es evidente que RX3 también requiere al menos dos residuos de cisteína cerca del extremo C-terminal para formar RPBLA.

Además, se sabe que la concentración de proteínas es un parámetro clave que controla la agregación de péptidos autoensamblantes (Wetzel, R. (1999) En *Methods in Enzymology*, vol. 309, Academic Press, San Diego, California). Sin embargo, la expresión de RX3-ECFP y mutantes de Cys de RX3 se visualizó claramente en geles de SDS-PAGE teñidos con azul de Coomassie y no se observaron diferencias significativas en los niveles de proteínas entre las proteínas que formaban RPBLA y las que se secretaban y no agregaban (Fig. 8B). Por tanto, diferencias en la multimerización de proteínas de los mutantes de RX3 probados están relacionadas con diferencias en sus propiedades intrínsecas que afectan las capacidades de agregación. Las modificaciones de la secuencia de RX3 podrían deteriorar la capacidad de agregación o aumentar el umbral de concentración de proteína necesario para la oligomerización.

La determinación del número mínimo de cisteínas requeridas para que un péptido ensamblador permita la inducción eficaz de RPBLA es crítica. En primer lugar, los residuos de cisteína en un péptido ensamblador pueden afectar negativamente al plegamiento apropiado de la proteína de interés, por tanto, su actividad, principalmente cuando esta proteína también contiene residuos de cisteína. Sin embargo, la presencia de cisteínas en el péptido ensamblador asegura el entrecruzamiento de la proteína de fusión mediante formación de puentes disulfuro. Este entrecruzamiento produce la estabilización y solidez de los RPBLA. Estas características son deseables porque permiten el aislamiento fácil de este orgánulo. Además, cualquier aplicación industrial en el que se usa el orgánulo RPBLA dependerá del entrecruzamiento mediante la formación de puentes disulfuro. Como se describe en la introducción, PBIS diferentes de prolaminas son capaces de inducir la formación de RPBLA mediante la adición de una secuencia KDEL en el extremo C de la proteína. Incluso aunque este planteamiento permite la formación de RPBLA, el aislamiento de estos orgánulos será difícil debido a su estabilidad reducida.

Ejemplo 10: La orientación de los residuos de cisteína no afecta la capacidad del péptido PP para acumularse en RPBLA en plantas de tabaco transformadas

La importancia de los residuos de cisteína en el ensamblaje de proteínas e inducción de RPBLA se demostró mediante mutaciones individuales y múltiples de estos residuos en el ejemplo 9. Sin embargo, quedaba por explorar la influencia de la orientación de estos residuos. Para hacer eso, se sintetizó un nuevo péptido ensamblador (PP2). PP2 es homólogo a PP excepto por la posición de algunos de los residuos de cisteína (Fig. 9A; C9→10, C84→85 y C92→91).

Los péptidos muy ricos en prolina forman una hélice PPII que se caracteriza por tener tres aminoácidos por giro (Brochicchio, 2002, Chrality, 14: 782-92), como se muestra en la figura 9B. En RX3 natural, los residuos de cisteína están orientados hacia cada uno de los tres lados de la hélice. Se hipotetizó que esta orientación podría ser importante para fomentar el ensamblaje y estabilización en las tres dimensiones que permite que los naturales y RPBLA formen los orgánulos habituales grandes (hasta 3 micrómetros) de forma redonda llenos de proteínas ensambladas y estabilizadas por puentes disulfuro. De esta manera, se diseñó PP2 de modo que todos los residuos de cisteína estén orientados hacia el mismo lado de la hélice. Inesperadamente, PP2 fusionado o GFP fue capaz de acumularse tanto como PP-GFP e incluso más sorprendentemente, de inducir la formación de grandes RPBLA (Fig. 9C).

Ejemplo 11: Acumulación de R8(4C)-ECFP, R7(4C)-ECFP, R6(4C)-ECFP y R4(4C)-ECFP en RPBLA de plantas de tabaco transformadas

También se determinó la longitud mínima de la PPII requerida para la biogénesis de RPBLA. Para hacer eso, se creó una proteína mínima derivada de RX3-ECFP, R8(C4)-ECFP, mediante delección de la secuencia PX de RX3 y adición de una secuencia mínima adicional. De esta manera, R8(C4)-ECFP carece de C82, C84 y C92, pero contiene un nuevo residuo adicional de cisteína unido a C64 por dos prolinas para reforzar los puentes disulfuro entre cadenas (Fig. 10A). El patrón de distribución de la proteína R8(C4)-ECFP cuando se expresaba en hojas de *N. benthamiana* se acumulaba en manchas fluorescentes esféricas evidentes que, a gran aumento, parecían ser pequeños RPBLA (Fig. 10C). Mientras que los RPBLA inducidos por RX3-ECFP alcanzaban un diámetro medio de 1,4 μm a 4 dpi y aumentaban progresivamente para alcanzar hasta 2 μm a 7 dpi, las manchas fluorescentes inducidas por R8(C4)-ECFP raramente alcanzaban diámetros de 1 μm a 7 dpi (véase la tabla posterior). Por tanto, como se ha sugerido anteriormente, el par de cisteínas en cada lado de la hélice PPII (RD en el contexto de R8(C4)) son suficientes para nuclear agregados de proteína que evolucionarán a RPBLA. La observación de que RX3-ECFP tiene una tendencia más fuerte a formar RPLBA grandes que R8(C4)-ECFP puede ser debida al hecho de que los oligómeros RX3-ECFP tienen más probabilidades de crecimiento al aumentar el entrecruzamiento entre cadenas en virtud de sus seis residuos de cisteína.

Por tanto, también se examinó el grado al que la polimerización se relaciona con la longitud de la PPII. Estos experimentos se realizaron acortando de forma progresiva el RD. Se generaron tres construcciones más a partir de R8(C4)-ECFP. El dominio de repetición se acortó a 7, 6 y 4 unidades de PPPVHL (Fig. 10A). Como se muestra en la figura 10C, las proteínas sobreexpresadas que contenían seis o cuatro unidades de repetición (R6(C4)-ECFP y R4(C4)-ECFP, respectivamente) aún eran capaces de formar pequeños agregados, pero la secreción aumenta con el acortamiento del dominio de repetición (Fig. 10B recuadro). R7(C4)-EGF produjo resultados similares a R8(C4)-EGF.

El análisis por tinción de azul de Coomassie de las proteínas totales extraídas de hojas que sobreexpresan las proteínas RX3 truncadas (Fig. 10B) mostró que la concentración de proteínas recombinantes era similar en todos los casos. Esto indica que cada una de las proteínas de fusión es estable y que un buen nivel de expresión no es una prueba de que los RPBLA se formarán eficazmente. En otras palabras, la menor eficacia de agregación en mutantes no estaba relacionada con los niveles de expresión de proteína. Estos resultados sugieren que hay un tamaño crítico que favorece las interacciones óptimas del péptido RX3 y determina la eficacia de la formación de PB.

Proteína de fusión	Intervalo de tamaño (μm)	
	4 dpi	7 dpi
RX3-ECFP	1,1-1,85	1,1-1,85
R8(4C)-ECFP	0,4-0,7	0,9-1,3
R7(4C)-ECFP	0,4-0,75	0,78-1,3
R6(4C)-ECFP	0,3-0,6	1-1,4
R4(4C)-ECFP	0,2-0,4	0,75-1

Tabla 5: Tamaño de los RPBLA después de 4 y 7 días de la transformación de las plantas.

Basado en el aumento significativo en la secreción de proteína de fusión de R4(C4)-ECFP comparado con las otras construcciones, R4(C4) marca la longitud mínima de un péptido ensamblador eficaz. Estos resultados indican que péptidos más largos de 24 aminoácidos (longitud del dominio de repetición de R4(C4)) que adoptan una hélice PPII y están flanqueados por un par de residuos de cisteína en cada extremo son suficientes para que se retengan en el RE y para inducir la formación de RPBLA.

Ejemplo 12: Acumulación de RX(A)-mCherry, RX(E)-mCherry en RPBLA de plantas de tabaco transformadas

La proteína fluorescente mCherry es muy soluble y reticente a acumularse en RPBLA cuando se fusiona a RX3. El análisis microscópico confocal de hojas de tabaco transformadas con un vector que expresaba RX3-mCherry mostró un patrón de secreción claro, como se muestra en la figura 11A donde la mayoría de la fluorescencia roja observada se localizaba en la periferia de la célula. De forma interesante, cuando se transformaron hojas de tabaco con

RX3(A)-mCherry, la mayoría de la fluorescencia se observó dentro de la célula, lo que indica que la proteína de fusión se retenía eficazmente en la célula.

Además, se observó un patrón particulado claro (Fig. 11B y 11B'), lo que sugiere que la proteína de fusión se acumulaba en los RPBLA. El tamaño de estos cuerpos proteicos recombinantes (aproximadamente 1 micrómetro de diámetro) se ajusta a los RPBLA caracterizados previamente en plantas transformadas con varias proteínas de fusión (véase, la patente de EE UU No. 7.575.898 y las publicaciones de patente de EE UU No. 20060123509, No. 20060121573 y No. 20070243198).

Este resultado fue inesperado al menos por dos razones. Primero, la eficacia de la formación de RPBLA aumentó cuando mCherry se fusionó a RX3(A) comparada con RX3 no modificado derivado de prolamina que contiene el dominio de repetición natural. Además, se describió que la anfipaticidad del dominio de repetición de RX3 es importante para la formación de cuerpos proteicos (Kogan *et al.*, 2001, J. Mol. Biol. 312:907-913) y RX3(A) se mutó para tener un dominio de repetición completamente hidrofóbico (véase el ejemplo 1).

Como se indica anteriormente, la alta solubilidad de mCherry podría ser uno de los principales factores que reducen la capacidad de la fusión de mCherry con RX3 natural de formar RPBLA. Para determinar si el mayor rendimiento de RX3(A) se debe al simple aumento en el contenido hidrofóbico del RD (sustitución de histidinas (aminoácidos cargados positivamente) por alaninas hidrofóbicas), se expresó una fusión de mCherry con RX3(E). Como se muestra en la figura 11C, RX3(E)-mCherry se retuvo eficazmente en la célula, y además, se acumuló en RPBLA que son incluso mayores que los inducidos por la expresión de RX3(A)-mCherry.

El hecho de que RX3(E), que tiene un RD anfipático cargado, también sea capaz de inducir la formación de RPBLA cuando se fusiona a mCherry, sugiere que el aumento en el rendimiento de RX3(A) comparada con RX3 no es debido a un aumento inespecífico del carácter hidrofóbico del péptido, sino que se debe al aumento de la capacidad de ensamblaje y, por tanto, a la capacidad de inducción de cuerpos proteicos de estos dos péptidos sintéticos derivados de RX3.

Ejemplo 13: Acumulación de EGF fusionado a los péptidos ensambladores RX3, RX3(A), RX3(E), PP y PA en plantas de tabaco transformadas

La producción y purificación del factor de crecimiento epidérmico (EGF) fusionado a RX3 en plantas de tabaco agroinfiltradas se ha descrito previamente (documento WO2006056484). El proceso de aislamiento de RPBLA mediante centrifugación de baja velocidad, la solubilización y el corte de la proteína de fusión RX3-EGF y la posterior purificación de EGF mediante FPLC de fase inversa (RF) se han desarrollado en el documento WO2007096192. Para probar la eficacia de los nuevos péptidos ensambladores, se fusionó EGF a RX3(A) y RX3(E) y se realizó un proceso posterior equivalente. Los RPBLA inducidos por la expresión de RX3(A)-EGF y RX3(E)-EGF se recuperaron tan eficazmente mediante centrifugación de baja velocidad como los RPBLA control inducidos por la expresión de RX3-EGF (Fig. 12A). En los tres casos, casi toda la proteína de fusión se recuperó en la fracción wPB (Fig. 12A, carril 4) con sólo pérdidas insignificantes en el sobrenadante (Fig. 12A, carril 2) y el paso de lavado (Fig. 12A, carril 3). A partir de estos resultados, es aparente que casi toda la proteína de fusión está estrechamente ensamblada en los orgánulos RPBLA se recupera fácilmente en la fracción wPB.

Un punto importante posterior es la solubilización de la proteína de fusión de los RPBLA. Después de la solubilización de los RPBLA de RX3-EGF en condiciones moderadas (tampón borato 50 mM a pH 10 y BME 10 mM a 25°C durante 4 horas), sólo se solubilizó un porcentaje bajo de la proteína de fusión. De hecho, sólo se solubilizó el 50% del monómero RX3-EGF (Fig. 12B, compárese el carril 3 (sPB) con el carril 6 (iPB)). Sorprendentemente, la solubilización de los RPBLA que contenían RX3(A)-EGF o RX3(E)-EGF fue mucho más eficaz. Casi toda la forma monomérica de RX3(A)-EGF se solubilizó y casi nada de RX3(E)-EGF parece permanecer insoluble en las condiciones usadas (Fig. 12B, compárese el carril 1 (sPB) con el carril 4 (iPB)).

También se examinaron los efectos de PP-EGF y PA-EGF en el procesamiento posterior. La figura 12C muestra que estas dos proteínas de fusión se acumulan al mismo nivel que RX3-EGF y RX3(E)-EGF en hojas de tabaco agroinfiltradas. Además, después de recuperar los RPBLA mediante centrifugación a baja velocidad, PP-EGF y PA-EGF se solubilizaron muy eficazmente, similar a RX3(E)-EGF, en las condiciones moderadas descritas anteriormente (Fig. 12D). Como se ha mencionado anteriormente, las diferencias en la movilidad electroforética entre todas estas proteínas de fusión no están relacionadas con una diferencia en el tamaño de la proteína, sino que se deben a ligeras diferencias en la conformación proteica incluso en condiciones de SDS-PAGE.

Inesperadamente, los nuevos péptidos ensambladores RX3(A), RX3(E), PP y PA aumentan todos mucho la solubilidad de las proteínas de fusión sin afectar a su capacidad de ensamblarse e inducir la formación de RPBLA. Este resultado chocante es de principal importancia en el procesamiento posterior de los RPBLA porque aumenta drásticamente el rendimiento total del proceso (al menos dos veces para RX3(A) y más de diez veces para RX3(E), PP y PA).

Para evaluar la conformación del EGF producido por medio de estos nuevos péptidos ensambladores, se probó la actividad de RX3(E)-EGF solubilizada, así como el EGF cortado por Fxa de esta proteína de fusión y purificado mediante RF-FPLC. La figura 13A muestra un gel teñido con plata con todos los pasos posteriores desde el homogeneizado no clarificado hasta el paso de corte. Después de cortar con la proteasa específica de sitio Fxa, el EGF liberado se purificó como se describe en materiales y métodos mediante RF-FPLC. La proteína EGF se recuperó al 30% de acetonitrilo en sólo dos fracciones correspondientes al pico agudo indicado por una flecha en la figura 13B. Se estimó la pureza de una mezcla de las dos fracciones recuperadas (figura 13C) en más del 92% por HPLC. Se midió la actividad del EGF in vitro mediante el análisis de la fosforilación del receptor de EGF de células A431 incubadas con cantidades crecientes de RX3(E)-EGF y EGF. Como control negativo estas células se incubaron en paralelo con RX3. Tres experimentos independientes demostraron que RX3(E)-EGF y EGF eran activos al 50% y el 100%, respectivamente comparados con un EGF comercial de Promega (GS02A).

Ejemplo 14: Acumulación de hGH fusionada a los péptidos ensambladores RX3 y RX3(A) en plantas de tabaco transformadas

Se eligió la hormona del crecimiento humana como un ejemplo adicional para comprobar la capacidad de las variantes de RX3 para aumentar el proceso posterior. Se homogenizaron plantas de tabaco que expresaban la proteína de fusión RX3(A)-hGH y los RPBLA inducidos se aislaron mediante centrifugación a baja velocidad. Como muestra la figura 14A (panel izquierdo), se obtuvo una fracción muy pura de RPBLA (wPB) del homogeneizado (H0). La proteína de fusión RX3(A)-hGH en sus diferentes formas oligoméricas es la única proteína observada en esta fracción cuando se analizó por tinción con Coomassie. De forma interesante, cuando el rendimiento del proceso se analizó por inmunotransferencia, la mayoría de la proteína de fusión se acumulaba en RPBLA estables estrechamente ensamblados, que se puede recuperar por centrifugación y solo una pequeña cantidad de la misma se pierde en el sobrenadante correspondiente (fig. 14A, panel izquierdo).

Una vez obtenidos los RPBLA que contenían RX3(A)-hGH mediante este sencillo método de centrifugación, se incubaron durante 3 horas en tampón de solubilización moderado (borato 50 mM pH 10, beta-mercaptoetanol (bME) 10 mM a temperatura ambiente). Como referencia, se incubaron cantidades equivalentes de RPBLA que contenían RX3-hGH en las mismas condiciones y se analizaron en paralelo mediante inmunotransferencia usando anticuerpos anti-hGH. Sorprendentemente, la proteína de fusión RX3(A)-hGH era mucho más soluble que RX3-hGH, incluso aunque el alto rendimiento en recuperación de RPBLA sugiere que los orgánulos inducidos por la expresión de cualquiera de estas dos proteínas de fusión son tanto estables como ensamblados estrechamente. El aumento en la solubilidad en agua es sorprendente dado el hecho de que cambiar los residuos de histidina de RX3 con alaninas en RX3(A) aumenta la hidrofobicidad del péptido ensamblador (Eisenberg, 1984, *J. Mol. Biol.* 179: 125-142).

Ejemplo 15: Dependencia de residuos de Cys en biogénesis de RPBLA densos en plantas de tabaco transformadas

Como se ha mencionado anteriormente, uno de los elementos esenciales de PBIS en la formación de RPBLA densos es la presencia de al menos 2 residuos de cisteína en cada extremo de un dominio poliprolina de tipo II. No obstante, hay otros péptidos ensambladores (por ejemplo, hidrofobina, ELP) que se retienen en el RE y se acumulan en estructuras de tipo vesícula cuando se fusionan a una secuencia KDEL, incluso en ausencia de formación de puentes disulfuro. La presencia de KDEL en el extremo C-terminal de ciertas proteínas ensambladoras es suficiente para retener la molécula en el RE, y la capacidad ensambladora de la proteína produce la formación de estructuras de vesícula.

Para determinar si la adición de una secuencia de retención de RE al final del péptido ensamblador RX3 podría sustituir la necesidad de los residuos de cisteína, la variante de RX3 $RX3\Delta Cys_{64,82,84,92}$ -ECFP, que no era capaz de inducir RPBLA y se secretaba como se muestra en el ejemplo 9, se fusionó a una ECFP que tenía un KDEL ($RX3\Delta Cys_{64,82,84,92}$ -ECFP-KDEL).

El análisis de hojas de tabaco que sobreexpresan RX3-ECFP y $RX3\Delta Cys_{64,82,84,92}$ -ECFP-KDEL por tinción de Coomassie e inmunotransferencia (figura 16A) mostró que las dos proteínas de fusión basadas en RX3 se acumulaban a niveles similares. Para determinar si $RX3\Delta Cys_{64,82,84,92}$ -ECFP-KDEL se acumula en RPBLA, se analizaron hojas que expresan esta proteína de fusión por microscopía confocal 3 y 7 días después de la infiltración (dpi). Como se observó anteriormente en hojas de tabaco que expresan RX3-ECFP, se observaron vesículas de aproximadamente 1 micrómetro de forma redonda inducidas por la expresión de $RX3\Delta Cys_{64,82,84,92}$ -ECFP-KDEL incluso solo 3 dpi (Figura 16B). Esta observación demuestra que la secuencia KDEL unida al extremo C-terminal de ECFP permite una retención eficaz de la proteína de fusión dentro de la célula. Además, también muestra que la capacidad ensambladora de péptido basado en RX3 permite el autoensamblaje e inducción de estructuras de vesícula. Estas estructuras de vesícula siguen creciendo después de 3 dpi, y, de forma interesante, a 7 dpi se observaron estructuras de vesícula grandes irregulares, algunas de más de 5 micrómetros (Figura 16B). No se observaron RPBLA de este tamaño en hojas de tabaco que sobreexpresan la proteína de fusión control RX3-ECFP, lo que sugiere que el mecanismo de autoensamblaje de RX3-ECFP en RPBLA no es el mismo que los mecanismos de ensamblaje de vesículas de $RX3\Delta Cys_{64,82,84,92}$ -ECFP-KDEL.

Un enfoque técnico para caracterizar indirectamente lo apretado del ensamblaje de la proteína de fusión dentro de estructuras de vesícula es la determinación de la densidad de las estructuras de vesículas por gradientes de colchón por capas Optiprep®. Por tanto, estructuras de vesículas inducidas por la expresión de RX3ΔCys_{64,82,84,92}-ECFP-KDEL (o RX3-ECFP) en plantas de tabaco se llevaron a cabo cargando homogeneizados de plantas filtrados sobre los siguientes colchones por capas de Optiprep®. La densidad de cada colchón por capas es como se ha definido anteriormente en la tabla 4.

Como se ejemplifica por el gradiente de RX3-ECFP, los RPBLA densos se recuperan de fracciones que tienen densidades de aproximadamente 1,2 g/ml (Fig. 16C, izquierda, carriles 5 y 6), que están libres de la mayoría de las proteínas endógenas. De hecho, la densidad y dureza de los RPBLA es de gran interés en el aislamiento de RPBLA como ya se ha descrito en la solicitud publicada en EE UU No. 2006/0123509. Cuando se analizaron estructuras de vesículas de RX3ΔCys_{64,82,84,92}-ECFP-KDEL por el mismo medio, sorprendentemente casi toda la proteína de fusión se recuperó en la fracción del sobrenadante (Fig. 16C, derecha). Este resultado claramente demuestra que las proteínas de fusión RX3ΔCys_{64,82,84,92}-ECFP-KDEL se acumulan en estructuras similares a vesículas que no pueden considerar verdaderos RPBLA. Los orgánulos intracelulares derivados del RE o aparato de Golgi, por ejemplo, típicamente sedimentan entre el 18% y 30% de colchones de Optiprep, pero durante el proceso de homogenización de tejido una gran proporción de las proteínas solubles presentes en la luz de los microsomas se libera y se recuperará en el sobrenadante. Por tanto, se puede concluir que la proteína de fusión RX3ΔCys_{64,82,84,92}-ECFP-KDEL se retiene en el RE, pero no se ensambla o se ensambla solo débilmente, y no se forman RPBLA densos.

Ejemplo 16: Acumulación de Z(Adh)-GFP, Z(Adh)Px-Gfp, Z(Col)-Gfp, Z(Col)Px-Gfp en RPBLA de plantas de tabaco transformadas

Se ha demostrado que PBIS con baja homología de secuencia a RX3 son capaces de generar RPBLA. PP y PA comparten menos del 60% de identidad con la secuencia RX3 entre los residuos C9 y C64. Por otra parte, es importante señalar que las secuencias PP y PA tienen un alto porcentaje de prolina (96,2 y 67,9 por ciento, respectivamente); significativamente mayor que RX3 de tipo salvaje (54,7%). Se evaluó la capacidad de inducción de RPBLA de otras secuencias de proteínas que adoptan una estructura PPII. Las secuencias cumplían los siguientes criterios: (i) menos del 40% a RX3 y (ii) menos del 40% de contenido en prolina. Se seleccionaron un fragmento de colágeno humano COL2A1 que comprende los aminoácidos 135 a 179 (AccN: CAA34683) y el fragmento 884 a 927 de la adhesina de superficie Agl/II (*Streptococcus mutans* cepa NG8; GeneBank: GQ456171. AccN: ACV69919). Como se muestra en la figura 17A, tales secuencias se usaron para sustituir RD en R8(C4) (Z(Adh) y Z(Col) o en los péptidos ensambladores de RX3 (Z(Adh)Px y Z(Col)Px).

Estos péptidos ensambladores fusionados a GFP se acumularon a altos niveles en hojas de tabaco y se representaron como una banda principal en homogeneizados preclarificados de estas hojas teñidos por Coomassie (Fig. 17B). Z(Adh)-Gfp y Z(Adh)Px-Gfp, con un MW predicho de 37,7 kDa y 34,6 kDa, respectivamente, tienen menor movilidad en gel de SDS-PAGE. Este cambio en migración, también observado en la mayoría de los péptidos ensambladores que adoptan una estructura PPII descritos anteriormente (tal como PP, RX3(E), etc.), se puede atribuir a la fuerte estabilidad de esta estructura secundaria.

El análisis por microscopía confocal de plantas de tabaco que expresan Z(Adh)-GFP, Z(Adh)Px-GFP, Z(Col)-GFP y Z(Col)Px-GFP mostró que estas proteínas de fusión se acumulaban dentro de la célula en RPBLA fluorescentes de forma redonda de aproximadamente 1 micrómetro de diámetro. A pesar del bajo porcentaje de homología respecto a RX3 y el bajo contenido en prolina, las proteínas de fusión también se retuvieron dentro del retículo endoplásmico (RE), y su ensamble indujo eficazmente la formación de RPBLA (Fig. 17C).

A diferencia de los cuerpos producidos por ELP, hidrofobina o RX3ΔCys_{64,82,84,92}-ECFP-KDEL, una característica importante de los RPBLA es el empaquetamiento apretado de la proteína dentro de los orgánulos densos. Esta característica permite el aislamiento de RPBLA por centrifugación. Por tanto, la observación de que Z(Adh)-GFP, Z(Adh)Px-GFP, Z(Col)-GFP y Z(Col)Px-GFP se recuperan en el precipitado (fracción de RPBLA) después de centrifugar a 1500xg (Fig. 17D), demuestra que estas proteínas de fusión inducen la formación de RPBLA densos. La presencia del dominio Px, de ahí el aumento del número de residuos de Cys (de 4 a 6), y también el agrandamiento del péptido ensamblador mediante una secuencia rica en prolina, produjo un aumento significativo en lo apretado del ensamblaje y recuperación de mayor eficacia de RPBLA (compárese Z(Adh)-GFP frente a Z(Adh)Px-GFP y Z(Col)-GFP frente a Z(Col)Px-GFP en la figura 17). Experimentos de gradiente de densidad y FRAP apoyan estos datos.

Como se puede observar de la comparación de secuencia de Z(Adh)-GFP, Z(Adh)Px-GFP, Z(Col)-GFP y Z(Col)Px-GFP (Fig. 17A), la última repetición del dominio de repetición de RX3 se mantuvo en todos los casos debido a razones de clonación. Para descartar que las PPII de Adh y Col necesitan la última unidad de repetición de RX3 del RD para formar RPBLA, se realizó un nuevo conjunto de construcciones donde esta repetición estaba ausente: Z(Adh2)-GFP, Z(Adh2)Px-GFP, Z(Col2)-GFP y Z(Col2)Px-GFP. Todas estas construcciones nuevas mantuvieron la capacidad para inducir RPBLA.

Ejemplo 17: Determinación de la actividad xilanasa en RPBLA inducidos por la expresión de RX3-Xyl, RX3(L)-Xyl, RX3(E)-Xyl, RX3(A)-Xyl, PP-Xyl y PA-Xyl en plantas de tabaco transformadas

5 Como se ha descrito previamente (documento WO2007/096192A1), los RPBLA inducidos por la expresión de proteínas de fusión derivadas de RX3 de tipo salvaje tienen la capacidad de mantener la actividad de la proteína de interés (PDI) fusionada a tal PBIS.

10 Para determinar cómo los péptidos ensambladores descritos en el presente documento afectan la actividad de las proteínas fusionadas a ellos, se estudiaron péptidos ensambladores no anfipáticos (PP, PA, RX3(A) y RX3(L)) y anfipáticos negativamente cargados (RX3(E)). Se eligió la enzima xilanasa (DQ465452) como la PDI indicadora y se fusionó a través de un enlazador que comprendía 5 glicinas a todos estos péptidos ensambladores, así como a RX3 de tipo salvaje como referencia.

15 Se recogieron hojas de tabaco que sobreexpresaban PP-Xyl, PA-Xyl, RX3(E)-Xyl, RX3(L)-Xyl, RX3(A)-Xyl y RX3-Xyl 6 días después de agroinfiltración, evitando tejido necrótico para disminuir la inconsistencia y variabilidad. Se llevó a cabo procesamiento posterior por centrifugación a baja velocidad para cada construcción, y se analizó la fracción enriquecida en RPBLA en SDS-PAGE por tinción con azul de Coomassie. En todos los casos, la fracción de RPBLA estaba muy enriquecida en la proteína de fusión, que era la proteína más abundante en la fracción. El contenido en
20 proteína se determinó mediante el kit de cuantificación de proteína EZQ (Invitrogen, Molecular Probes), y la actividad xilanasa asociada con cada muestra se cuantificó con un sustrato sintético (DiMUX2), midiendo el correspondiente producto fluorescente cada 2 minutos (longitud de onda: excitación 360/40 nm, emisión 460/40 nm).

25 Inesperadamente, entre todas las muestra analizadas, la fracción de RPBLA que tiene la menor actividad específica (11,7 nmol/(minuto*microgramo de proteína)) correspondía a la proteína de fusión RX3-Xyl. Tomando esta proteína de fusión como referencia, RX3(A)-Xyl mostró casi el doble de actividad específica; RX3(E)-Xyl y RX3(L)-Xyl mostraron aproximadamente un aumento de 3 veces; y PP-Xyl y PA-Xyl tenían actividades específicas por encima de 7 veces mayores que la referencia (véase la tabla 4). La actividad se midió usando la misma cantidad de proteína de fusión, de modo que la actividad específica aumentada de estos ensambladores comparados con RX3 podría
30 estar relacionada con las diferentes propiedades de los RPBLA. Esta observación es de gran importancia considerando que una alta actividad específica es un atributo útil en aplicaciones de enzimas

	nmol/ min/μg prot	Desv Est	VECES	Desv Est Rel
RX3-Xyl	11,7	0,63	1,0	0,05
PA-Xyl	84,7	0,57	7,3	0,05
PP-Xyl	84,4	4,51	7,2	0,50
RX3(A)-Xyl	21	0,5	1,8	0,05
RX3(E)-Xyl	33,5	2,89	2,9	0,25
RX3(L)-Xyl	41,2	4,00	3,5	0,34

Tabla 6: Actividad específica de los RPBLA inducidos por la expresión de distintos péptidos ensambladores.

35 Ejemplo 18: Independencia de la orientación del dominio de repetición de RX3 con respecto a los residuos de cisteína en su capacidad para inducir RPBLA en plantas de tabaco y con respecto al polipéptido de interés

40 El dominio repetitivo (RD) de RX3 de tipo salvaje está flanqueado por 2 y 4 residuos de cisteína localizados en el N-terminal y C-terminal, respectivamente. Esta distribución asimétrica podría ser de alguna importancia respecto a la capacidad de ensamblaje y/o la eficacia en la inducción de RPBLA. Para probar esto, se generó una nueva construcción (iRX3) de modo que las regiones flanqueantes del RD se intercambiaron y clonaron en la orientación invertida (Fig. 18A). Cuando hojas de tabaco se agroinfiltraron con una construcción que codifica el péptido ensamblador RX3 invertido fusionado a ECFP (iRX3-ECFP), se observaron RPBLA grandes de forma redonda.
45 Sorprendentemente, el tamaño de los RPBLA inducidos era significativamente mayor que los RPBLA correspondientes obtenidos por la expresión de la proteína de fusión RX3-ECFP usada como referencia (Fig. 18B). El tamaño medio de RX3-ECFP e iRX3-ECFP era aproximadamente 1 y 2,5 micrómetros, respectivamente.

50 La alta densidad y apretura en RPBLA permite el aislamiento eficaz por centrifugación del resto de los orgánulos celulares y proteínas solubles (solicitud publicada en EE UU No. 2006/0123509), de modo que la densidad de los RPBLA de iRX3-ECFP se determinó por un gradiente de densidad multicapa de Optiprep. Una comparación de la distribución de los RPBLA a lo largo del gradiente de densidad inducido por la expresión de iRX3-ECFP y RX3-ECFP no mostró diferencias significativas. En ambos casos, la mayoría de los RPBLA se formaron 6 días después de agroinfiltración y se recuperaron de la fracción de alta densidad que varía de 1,175 a 1,26 g/cm³ (Fig. 18C;
55 carriles 4 a 7). Las fracciones también estaban libres de la mayoría de las proteínas celulares endógenas.

La importancia de la posición relativa del péptido ensamblador (RX3 e iRX3) con respecto a la proteína de interés (ECFP) también se analizó. Se generaron dos construcciones adicionales clonando el péptido ensamblador RX3 o iRX3 en el extremo C-terminal de ECFP (Fig. 18; ECFP-RX3 y ECFP-iRX3). Hojas de tabaco que expresaban estas proteínas de fusión fueron capaces de inducir RPBLA, aunque menos eficazmente que las proteínas de fusión N-terminales correspondientes (RX3-ECFP e iRX3-ECFP). En la figura 18B, las imágenes confocales muestran que ECFP-RX3 y ECFP-iRX3 inducen un gran número de pequeños RPBLA, la mayoría de ellos de aproximadamente 0,5 micrómetros. De forma interesante, a pesar del tamaño reducido, algunos RPBLA se recuperaron en las fracciones densas 1,175 y 1,21 g/cm³ (Fig. 18C, carriles 4 y 5), que estaban libres de la mayoría de los contaminantes celulares. La presencia de algo de proteína de fusión en las fracciones con menor densidad (Fig. 18C, carriles 2 y 3) puede representar la proteína de fusión que no se ha ensamblado en RPBLA completos, probablemente debido a una cinética de ensamblaje más lenta. También se puede concluir que los RPBLA inducidos por la expresión de proteínas de fusión que tienen el péptido ensamblador unido al extremo C-terminal de la proteína de interés se pueden aislar por centrifugación.

Ejemplo comparativo 19: Acumulación en RPBLA de hGH-I-RX3, DsRED-I-RX3 y EK-RX3 en células CHO

Para demostrar que la PBIS fusionada en el extremo C-terminal de la PDI puede inducir la formación de RPBLA en células CHO, se generaron tres construcciones que expresaban las siguientes proteínas de fusión: (i) hGH-I-RX3, (ii) EK-RX3, y (iii) DsRED-I-RX3 (véase la figura 19). En los tres casos, después de transfección de células CHO, la observación del patrón característico de manchas intracelulares demostró que las correspondientes proteínas de fusión se retenían dentro de las células en RPBLA (figuras 19A y B). La heterogeneidad de los tamaños de los RPBLA que se puede observar claramente en la figura 19A se puede asociar con diferentes fases de formación de RPBLA o a diferencias en la eficacia de transfección y las diferencias resultantes en los niveles de expresión de las proteínas de fusión.

Un enfoque técnico para caracterizar indirectamente lo apretado del ensamblaje de la proteína de fusión dentro del cuerpo proteico es la determinación de la densidad de los RPBLA por gradientes en colchón por capas de sacarosa. Por tanto, se cargaron RPBLA inducidos por la expresión de hGH-I-RX3 en células CHO encima del siguiente gradiente en colchón por capas de sacarosa:

% de sacarosa	Densidad (g/ml)
20	1,08
27	1,12
35	1,16
42	1,2
56	1,28

Tabla 7: Concentración de sacarosa (%) y densidad (g/ml)

El análisis de volúmenes equivalentes de las diferentes fracciones recuperadas del gradiente de densidad por inmunotransferencia mostró que la mayoría de la proteína de fusión cargada (Fig. 19A, carril H) se recuperó en fracciones densas (Fig. 19A, carril F42 y F56). Una pequeña cantidad de proteína de fusión no ensamblada también se observó en las fracciones S y F27, probablemente debido a cinética de ensamblaje más lenta. Como los RPBLA muy densos típicamente se recuperan de fracciones que tienen densidades de aproximadamente 1,2 g/ml, se puede concluir que se pueden inducir RPBLA por proteínas de fusión que tienen el dominio RX3 en el extremo C-terminal de la proteína.

Ejemplo comparativo 20: Acumulación de RPBLA de hGH-I-RX3 en células de insecto SF9

Una demostración de que proteínas de fusión que tienen el péptido ensamblador RX3 en el extremo C-terminal de la proteína son capaces de inducir RPBLA en células de insecto se realizó con hGH-I-RX3. Células de insecto infectadas con un virus recombinante pBacPAK8 que expresa hGH-RX3 mostraron el patrón de inmunofluorescencia característico de acumulación de RPBLA (Figura 19C). Las manchas distribuidas uniformemente a lo largo de la célula corresponden a RPBLA y demostraban que la proteína de fusión se retenía eficazmente en el RE. Cuando células SF9 que expresan hGH-I-RX3 se homogenizaron y centrifugaron a baja velocidad (3000xg), una gran proporción de la proteína de fusión se recuperó en la fracción de los RPBLA (figura 19C, panel derecho, carril 3). Como se observa en proteínas de fusión con RX3 N-terminal (Publicación en EE UU No. 2006/0123509), el ensamblaje apretado en orgánulos densos (RPBLA) inducido por interacciones RX3:RX3 permite una recuperación eficaz de una fracción muy enriquecida en RPBLA.

Ejemplo 21: Acumulación de RPBLA en células de mamífero

Se fusionaron secuencias que codifican las siguientes proteínas a secuencias que codifican xilanasas y se clonaron en el vector pcDNA 3.1 (Invitrogen) para expresión en células de mamífero: RX3, RX3(A), RX3(L), RX3(A3), RX3(E), RX3(D), RX3(T), RX3(N), RX3(Q), PP, PA, RX3C64G, RX3C82G, RX3C84G, RX3C92G, PP2, R8(C4), R7(C4),

R6(C4), R4(C4), Z(Adh), Z(Adh)Px, Z(Col), Z(Col)Px, e iRX3. Los vectores resultantes se introdujeron en células 293T, Cos1 y CHO usando el método de transfección basado en lipofectamina.

5 Las inmunotransferencias de células transfectadas muestran acumulación de todas las proteínas de fusión. Además, la localización de las proteínas de fusión por inmunocitoquímica indica que las proteínas de fusión se acumulan en RPBLA esféricos que tienen diámetros desde aproximadamente 0,5 hasta aproximadamente 3 micrómetros. La densidad de los RPBLA se determina cargando en colchones por capas de sacarosa y es de aproximadamente 1,1 a aproximadamente 1,4 g/ml. Los RPBLA se purifican usando centrifugación a baja velocidad (menos de aproximadamente 5000xg), y los RPBLA recuperados son al menos aproximadamente el 95% puros. Los RPBLA se solubilizan incubando en tampón moderado (Tris 50 mM pH 8, TCEP 5 mM y 2bME 10 mM) durante aproximadamente 4 horas y después centrifugación a aproximadamente 16.000 g durante aproximadamente 10 minutos. Se recuperan altos rendimientos de proteína en la porción solubilizada. La actividad xilanasa se mide usando un sustrato sintético (DiMux2) y se observan altos niveles de actividad.

15 **Ejemplo 22: Acumulación de RPBLA en células de insecto**

Se fusionaron secuencias que codifican las siguientes proteínas a secuencias que codifican xilanasa y se clonaron en el sistema de vector de expresión de baculovirus pFastBacK (Invitrogen): RX3, RX3(A), RX3(L), RX3(A3), RX3(E), RX3(D), RX3(T), RX3(N), RX3(Q), PP, PA, RX3C64G, RX3C82G, RX3C84G, RX3C92G, PP2, R8(C4), R7(C4), R6(C4), R4(C4), Z(Adh), Z(Adh)Px, Z(Col), Z(Col)Px, e iRX3. El virus recombinante se produce usando el kit de transfección BaculoGold (PharMingen, San Diego, California, EE UU). Las células Sf9 se dejan adherir al fondo de placas de cultivo y después de 15 minutos a 1 hora de incubación, la solución madre de virus se añade a los cultivos que se mantienen a 27°C en aire humidificado durante desde aproximadamente 30 hasta aproximadamente 36 horas.

25 Las inmunotransferencias de células infectadas muestran acumulación de todas las proteínas de fusión. Además, la localización de las proteínas de fusión por inmunocitoquímica indica que las proteínas de fusión se acumulan en RPBLA esféricos que tienen diámetros desde aproximadamente 0,5 hasta aproximadamente 3 micrómetros. La densidad de los RPBLA se determina cargando en colchones por capas de sacarosa y es de aproximadamente 1,1 a aproximadamente 1,4 g/ml. Los RPBLA se purifican usando centrifugación a baja velocidad (menos de aproximadamente 5000xg), y los RPBLA recuperados son al menos aproximadamente el 95% puros. Los RPBLA se solubilizan incubando en tampón moderado (Tris 50 mM pH 8, TCEP 5 mM y 2bME 10 mM) durante aproximadamente 4 horas y después centrifugación a aproximadamente 16.000 g durante aproximadamente 10 minutos. Se recuperan altos rendimientos de proteína en la porción solubilizada. La actividad xilanasa se mide usando un sustrato sintético (DiMux2) y se observan altos niveles de actividad

35 **Ejemplo 23: Acumulación de RPBLA en células de hongos filamentosos**

40 Se fusionan secuencias que codifican las siguientes proteínas a secuencias que codifican xilanasa y se clonaron en vectores de expresión de *Trichoderma reesei* se introducen en la cepa RutC-30 de *T. reesei* [Montenecourt BS, Eveleigh DE: Selective screening methods for the isolation of high yielding cellulase mutants of *Trichoderma reesei*. *Adv Chem Ser* 1979, 181:289-301] esencialmente como se ha descrito [Penttilä M, Nevalainen H, Rättö M, Salminen E, Knowles J: A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. *Gene* 1987, 61:155-164]: RX3, RX3(A), RX3(L), RX3(A3), RX3(E), RX3(D), RX3(T), RX3(N), RX3(Q), PP, PA, RX3C64G, RX3C82G, RX3C84G, RX3C92G, PP2, R8(C4), R7(C4), R6(C4), R4(C4), Z(Adh), Z(Adh)Px, Z(Col), Z(Col)Px, e iRX3. Se seleccionan transformantes en placas que contienen 125 µg/ml de higromicina B. Las transformantes se extienden en medio selectivo que contiene lactosa para expresión inducida y se criban por microscopía de fluorescencia. Se recogen micelios de los transformantes que producen las mayores cantidades de las proteínas de fusión por filtración.

50 Las inmunotransferencias de células de micelio transformadas muestran acumulación de todas las proteínas de fusión. Además, la localización de las proteínas de fusión por inmunocitoquímica indica que las proteínas de fusión se acumulan en RPBLA esféricos que tienen diámetros desde aproximadamente 0,5 hasta aproximadamente 3 micrómetros. La densidad de los RPBLA se determina cargando en colchones por capas de Optiprep y es de aproximadamente 1,1 a aproximadamente 1,4 g/ml. Los RPBLA se purifican usando centrifugación a baja velocidad (menos de aproximadamente 5000xg), y los RPBLA recuperados son al menos aproximadamente el 95% puros. Los RPBLA se solubilizan incubando en tampón moderado (Tris 50 mM pH 8, TCEP 5 mM y 2bME 10 mM) durante aproximadamente 4 horas y después centrifugación a aproximadamente 16.000 g durante aproximadamente 10 minutos. Se recuperan altos rendimientos de proteína en la porción solubilizada. La actividad xilanasa se mide usando un sustrato sintético (DiMux2) y se observan altos niveles de actividad.

60 **Ejemplo 24: Resumen de las propiedades de los RPBLA obtenidos usando las PBIS según la invención y divulgadas en el presente documento**

Se determinaron el diámetro y la densidad de los RPBLA preparados de plantas de tabaco agroinfiltradas con las correspondientes construcciones que expresan proteína de fusión de las PBIS de RX3, RX3(E), RX3(D), RX3(Q), RX3(N), RX3(T), RX3(A), RX3(A3), RX3(L), PA, PP e iRX3 y ECFP (cuando se usa la PBIS iRX3, las proteínas de fusión se construyeron con ambas orientaciones de la PBIS con respecto a la ECFP) como se ha explicado anteriormente. Los resultados se proporcionan en la tabla 9.

5

Proteínas de fusión	Diámetro * en µm.	RPBLA* Densidad (g/ml)
RX3-ECFP	1<d<2	1,21-1,26
RX3(E)-ECFP	1<d<5	1,166-1,185
RX3(D)-ECFP	1<d<2	nd
RX3(Q)-ECFP	0,5<d<2	aproximadamente 1,17
RX3(N)-ECFP	0,5<d<1	nd
RX3(T)-ECFP	0,5<d<1	nd
RX3(A)-ECFP	1<d<2	1,194-1,204
RX3(A3)-ECFP	1<d<2	aproximadamente 1,17
RX3(L)-ECFP	aproximadamente 0,5	1,160-1,204
PA-ECFP	0,5<d<1	1,175-1,204
PP-ECFP	0,5<d<1	1,175-1,204
ECFP-iRX3	0,5<d<1	1,11-1,17
iRX3-ECFP	1<d<3	1,21-1,23

Tabla 8: Diámetro y densidad de RPBLA obtenidos usando diferentes construcciones

Ejemplo comparativo 25: Construcción de PBIS no alergénicas

10

Las proteínas prolaminas y péptidos derivados de prolaminas pueden ser alergénicos. De forma interesante, cuando se analizó la putativa alergenicidad del péptido RX3 en la base de datos AllergenOnline (versión 10.0, enero, 2010; <http://www.allergenonline.com/>), desarrollada por el Programa de Investigación y Recursos de Alergias Alimenticias (FARRP, por sus siglas en inglés), se encontraron 10 resultados con más del 35% de identidad a péptidos alergénicos. Este resultado sugiere que RX3 de tipo salvaje tiene potencial alergénico. Versiones poco alergénicas o no alergénicas de este péptido son útiles en varias aplicaciones (es decir, nutrición). Por tanto, el mismo análisis se realizó con diferentes variantes de péptidos RX3 descritos en esta solicitud, y los resultados demostraron que PA y RX3(A), son significativamente menos alergénicos que RX3 de tipo salvaje. El análisis de secuencia de PA y RX3(A) en la base de datos AllergenOnline mostró sólo tres y dos resultados con más del 35% de identidad a péptidos alergénicos, respectivamente. Basado en estas observaciones, se sintetizaron varias versiones nuevas no alergénicas del péptido ensamblador RX3: RX3-(LA1) (SEQ ID NO: 110) y RX3-(LA2) (SEQ ID NO: 111). Estos dos péptidos no tienen resultados con más del 35% de identidad a péptidos alergénicos, lo que indica que no son alergénicos. Las proteínas de fusión de RX3(LA1) y RX3(LA2) con GFP y ECFP se expresaron en hojas de tabaco, y su capacidad de inducción de RPBLA no está afectada.

15

20

25

RX3-LA1 (SEQ ID NO:60)
THTSGGCGCA PPPPAAAPPP AAAPPPAAAP PPAAAPPPAA APPPAAAPPP AAAPPPAAAP PPPCHYPTAP
ARPAPAPAPA PCPCAAPAPS PCQ

30

RX3-LA2 (SEQ ID NO:61)
THTSGGCGCA PPPPAAAPPP AAAPPPAAAP PPAAAPPPAA APPPAAAPPP AAAPPPAAAP PPPCHYPPAP
APPAPAPPAP ACPCAAPAPS PCQ

35

Notablemente, R8(4C) no tiene resultados con más del 35% de identidad a péptidos alergénicos, lo que indica que no es alergénico en absoluto. Este resultado indica que la alergenicidad de RX3 se debe principalmente a la secuencia de aminoácidos en el dominio Pro-X. Para evitar putativos efectos alergénicos en los péptidos ensambladores, se han sintetizado varias variantes no alergénicas de R8(4C) y se han probado con éxito para la inducción de RPBLA en plantas de tabaco.

40

* * *

Se debe apreciar que se pretende que se use la sección Descripción Detallada, y no las secciones Compendio y Resumen, para interpretar las reivindicaciones. Las secciones Compendio y Resumen pueden explicar una o más

pero no todas las formas de realización ejemplares de la presente invención como contemplan el/los inventor(es), y por tanto, no se pretende que limiten la presente invención y las reivindicaciones adjuntas en modo alguno.

5 La presente invención se ha descrito anteriormente con la ayuda de bloques constructores funcionales que ilustran la realización de funciones especificadas y las relaciones de las mismas.

Lista de secuencias

10 <110> ERA BIOTECH S.A.
 <120> SECUENCIAS POLIPEPTÍDICAS INDUCTORAS DE CUERPOS PROTEICOS
 <130> P5854PC00
 15 <150> EP10382231.8
 <151> 13-08-2010
 <150> US 61/349655
 <151> 28-05-2010
 20 <160> 130
 <170> PatentIn versión 3.5
 25 <210> 1
 <211> 36
 <212> PRT
 <213> Secuencia artificial
 30 <220>
 <223> PPVHLx6
 <400> 1
 Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15
 His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30
 Pro Val His Leu
 35 35
 <210> 2
 <211> 48
 <212> PRT
 <213> Secuencia artificial
 40 <220>
 <223> (PPPVHL)x6 - PPPVHVPPPVHL
 <400> 2
 Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15
 His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30
 Pro Val His Leu Pro Pro Pro Val His Val Pro Pro Pro Val His Leu
 45 35 40 45

<210> 3
 <211> 54
 <212> PRT
 <213> Secuencia artificial
 5
 <220>
 <223> (PPPVHL)x6 - PPPVHVPPPVLPPPPCH
 <400> 3
 Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15
 His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30
 Pro Val His Leu Pro Pro Pro Val His Val Pro Pro Pro Val His Leu
 35 40 45
 Pro Pro Pro Pro Cys His
 10 50
 <210> 4
 <211> 6
 <212> PRT
 15 <213> Secuencia artificial
 <220>
 <223> PPPVHL
 20 <400> 4
 Pro Pro Pro Val His Leu
 1 5
 <210> 5
 <211> 4
 25 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> PPPX
 30 <220>
 <221> característica mezcla
 <222> (4)..(4)
 <223> Xaa puede ser cualquier aminoácido natural
 35 <400> 5
 Pro Pro Pro Xaa
 1
 <210> 6
 <211> 4
 <212> PRT
 <213> Secuencia artificial
 40 <220>
 <223> PPXX
 45 <220>
 <221> característica mezcla

<222> (3)..(4)
 <223> Xaa puede ser cualquier aminoácido natural

 <400> 6
Pro Pro Xaa Xaa
 5 **1**

 <210> 7
 <211> 4
 <212> PRT
 10 <213> Secuencia artificial

 <220>
 <223> PXXX

 15
 <220>
 <221> característica mezcla
 <222> (2)..(4)
 <223> Xaa puede ser cualquier aminoácido natural
 20 <400> 7
Pro Xaa Xaa Xaa
1

 <210> 8
 <211> 4
 25 <212> PRT
 <213> Secuencia artificial

 <220>
 <223> PPXP
 30

 <220>
 <221> característica mezcla
 <222> (3)..(3)
 35 <223> Xaa puede ser cualquier aminoácido natural

 <400> 8
Pro Pro Xaa Pro
1

 40 <210> 9
 <211> 5
 <212> PRT
 <213> Secuencia artificial

 45 <220>
 <223> PPPXX

 <220>
 <221> característica mezcla
 <222> (4)..(5)
 <223> Xaa puede ser cualquier aminoácido natural

 <400> 9
Pro Pro Pro Xaa Xaa
 55 **1** **5**

 <210> 10
 <211> 6
 <212> PRT
 60 <213> Secuencia artificial

<220>
 <223> PPPXXX

5
 <220>
 <221> característica mezcla
 <222> (4)..(6)
 <223> Xaa puede ser cualquier aminoácido natural

10
 <400> 10
 Pro Pro Pro Xaa Xaa Xaa
 1 5

15
 <210> 11
 <211> 6
 <212> PRT
 <213> Secuencia artificial

20
 <220>
 <223> PPPXPX

25
 <220>
 <221> característica mezcla
 <222> (4)..(4)
 <223> Xaa puede ser cualquier aminoácido natural

30
 <220>
 <221> característica mezcla
 <222> (6)..(6)
 <223> Xaa puede ser cualquier aminoácido natural

35
 <400> 11
 Pro Pro Pro Xaa Pro Xaa
 1 5

40
 <210> 12
 <211> 18
 <212> PRT
 <213> Secuencia artificial

45
 <220>
 <223> PPPAAAPPPAAAPPPAAA

50
 <400> 12
 Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala
 1 5 10 15
 Ala Ala

55
 <210> 13
 <211> 6
 <212> PRT
 <213> Secuencia artificial

60
 <220>
 <223> PPPAAA

65
 <400> 13
 Pro Pro Pro Ala Ala Ala
 1 5

70
 <210> 14

ES 2 623 147 T3

<211> 24
<212> PRT
<213> Secuencia artificial

5 <220>
<223> PPPAAAPPPAAAPPAPPPPPAPPP

<400> 14
Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Ala Pro
1 5 10 15

Pro Pro Pro Pro Ala Pro Pro Pro
20

10 <210> 15
<211> 6
<212> PRT
<213> Secuencia artificial

15 <220>
<223> PPAPPP

<400> 15
Pro Pro Ala Pro Pro Pro
1 5

20 <210> 16
<211> 6
<212> PRT
25 <213> Secuencia artificial

<220>
<223> PPPVAL

30 <400> 16
Pro Pro Pro Val Ala Leu
1 5

<210> 17
<211> 6
35 <212> PRT
<213> Secuencia artificial

<220>
<223> PPPVLL

40 <400> 17
Pro Pro Pro Val Leu Leu
1 5

<210> 18
<211> 6
45 <212> PRT
<213> Secuencia artificial

<220>
<223> PPPVEL

50 <400> 18
Pro Pro Pro Val Glu Leu
1 5

55 <210> 19

<211> 6
 <212> PRT
 <213> Secuencia artificial
 5 <220>
 <223> PPPVDL
 <400> 19
Pro Pro Pro Val Asp Leu
 1 5
 10 <210> 20
 <211> 6
 <212> PRT
 <213> Secuencia artificial
 15 <220>
 <223> PPPVTL
 <400> 20
Pro Pro Pro Val Thr Leu
 1 5
 20 <210> 21
 <211> 6
 <212> PRT
 25 <213> Secuencia artificial
 <220>
 <223> PPPVNL
 30 <400> 21
Pro Pro Pro Val Asn Leu
 1 5
 <210> 22
 <211> 6
 35 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> PPPVQL
 40 <400> 22
Pro Pro Pro Val Gln Leu
 1 5
 <210> 23
 <211> 6
 45 <212> PRT
 <213> Secuencia artificial
 <220>
 50 <223> PPPAPA
 <400> 23
Pro Pro Pro Ala Pro Ala
 1 5
 55 <210> 24
 <211> 6
 <212> PRT
 <213> Secuencia artificial
 60 <220>

<223> P P P P P P

<400> 24
Pro Pro Pro Pro Pro Pro
 1 5

5 <210> 25
 <211> 10
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> THTSGGCGCQ

<400> 25
Thr His Thr Ser Gly Gly Cys Gly Cys Gln
 1 5 10

15 <210> 26
 <211> 10
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> THPPPPPCPCP

25 <400> 26
Thr His Pro Pro Pro Pro Cys Pro Cys Pro
 1 5 10

30 <210> 27
 <211> 48
 <212> PRT
 <213> Secuencia artificial

35 <220>
 <223> (PPPVHL) x 8

<400> 27
Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15

His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30

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

40 <210> 28
 <211> 6
 <212> PRT
 <213> Secuencia artificial

45 <220>
 <223> VHLPPP

<400> 28
Val His Leu Pro Pro Pro
 1 5

50 <210> 29

ES 2 623 147 T3

<211> 36
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> (PPPVHL) x 6

<400> 29
 Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15

His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30

Pro Val His Leu

35

10

<210> 30
 <211> 52
 <212> PRT
 <213> Secuencia artificial

15

<220>
 <223> PPPVELPPP VELPPPVELP PPVELPPPVE LPPPVELPPP VEVPPPVELPPPP

20

<400> 30
 Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val
 1 5 10 15

Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro
 20 25 30

Pro Val Glu Leu Pro Pro Pro Val Glu Val Pro Pro Pro Val Glu Leu
 35 40 45

Pro Pro Pro Pro
 50

25

<210> 31
 <211> 48
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> PPPVDLPPP VDLPPPVDLP PPVDLPPPVD LPPPVDLPPP VDLPPPVDL

30

<400> 31
 Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val
 1 5 10 15

Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro
 20 25 30

Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu
 35 40 45

<210> 32
 <211> 48
 <212> PRT
 5 <213> Secuencia artificial

 <220>
 <223> PPPVNLPPP VNLPPPVNLP PPVNLPPPVN LPPPVNLPPP VNVPPPVN

 10 <400> 32
 Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val
 1 5 10 15

 Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro
 20 25 30

 Pro Val Asn Leu Pro Pro Pro Val Asn Val Pro Pro Pro Val Asn Leu
 35 40 45

 <210> 33
 15 <211> 48
 <212> PRT
 <213> Secuencia artificial

 <220>
 20 <223> PPPVNLPPPV NLPPPVNLP PVNLPPPVN PPPVNLPPPV NLPPPVN

 <400> 33
 Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val
 1 5 10 15

 Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro
 20 25 30

 Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu
 35 40 45

 25 <210> 34
 <211> 48
 <212> PRT
 <213> Secuencia artificial

 30 <220>
 <223> PPPVQLPPP VQLPPPVQLP PPVQLPPPVQ LPPPVQLPPP VQVPPPVQL

 <400> 34
 Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val
 1 5 10 15

 Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro
 20 25 30

 Pro Val Gln Leu Pro Pro Pro Val Gln Val Pro Pro Pro Val Gln Leu
 35 40 45

 35 <210> 35
 <211> 48
 <212> PRT
 <213> Secuencia artificial

ES 2 623 147 T3

<220>

<223> PPPVQLPPP VQLPPPVQLP PPVQLPPPVQ LPPPVLPPP VQLPPPVQL

5 <400> 35

Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val
1 5 10 15

Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro
20 25 30

Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu
35 40 45

<210> 36

10 <211> 6

<212> PRT

<213> Secuencia artificial

<220>

15 <223> PPPVSL

<400> 36

Pro Pro Pro Val Ser Leu
1 5

20 <210> 37

<211> 6

<212> PRT

<213> Secuencia artificial

25 <220>

<223> PPPVVL

<400> 37

Pro Pro Pro Val Val Leu
1 5

30

<210> 38

<211> 48

<212> PRT

<213> Secuencia artificial

35

<220>

<223> PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VAVPPPVAL

<400> 38

Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val
1 5 10 15

Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro
20 25 30

Pro Val Ala Leu Pro Pro Pro Val Ala Val Pro Pro Pro Val Ala Leu
35 40 45

40

<210> 39

<211> 48

<212> PRT

45 <213> Secuencia artificial

ES 2 623 147 T3

<220>

<223> PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VALPPPVAL

<400> 39

Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val
1 5 10 15

Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro
20 25 30

Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu
35 40 45

5

<210> 40

<211> 48

<212> PRT

10 <213> Secuencia artificial

<220>

<223> PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VAVPPPVHL

15 <400> 40

Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val
1 5 10 15

Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro
20 25 30

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

<210> 41

<211> 48

20 <212> PRT

<213> Secuencia artificial

<220>

<223> PPPVALPPP VALPPPVALP PPVALPPPVA LPPPVALPPP VALPPPVHL

25

<400> 41

Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val
1 5 10 15

Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro
20 25 30

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

<210> 42

30 <211> 48

<212> PRT

<213> Secuencia artificial

<220>

35 <223> PPPVLLPPP VLLPPPVLLP PPVLLPPPVL LPPPVLLPPP VLVPPPVLL

<400> 42

ES 2 623 147 T3

Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val
1 5 10 15

Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro
20 25 30

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

<210> 43

<211> 48

5 <212> PRT

<213> Secuencia artificial

<220>

<223> PPPVLLPPP VLLPPPVLLP PPVLLPPPVL LPPPVLLPPP VLLPPPVLL

10

<400> 43

Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val
1 5 10 15

Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro
20 25 30

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

<210> 44

<211> 48

15 <212> PRT

<213> Secuencia artificial

<220>

<223> PPPVLPPP VVLLPPPVLPPV PPVVLPPPVL LPPPVLPPP VVPPPVL

20

<400> 44

Pro Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro Val
1 5 10 15

Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro
20 25 30

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

<210> 45

<211> 48

<212> PRT

<213> Secuencia artificial

<220>

<223> PPPVLPPP VVLLPPPVLPPV PPVVLPPPVL LPPPVLPPP VVLLLPPPVL

30

<400> 45

ES 2 623 147 T3

Pro Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro Val
 1 5 10 15

Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro
 20 25 30

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

<210> 46
 <211> 48
 5 <212> PRT
 <213> Secuencia artificial

<220>
 <223> PPPVTLPPP VTLPPPVTLP PPVTLPPPVT LPPPVTLPPP VTVPPPVTL
 10

<400> 46
 Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val
 1 5 10 15

Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro
 20 25 30

Pro Val Thr Leu Pro Pro Pro Val Thr Val Pro Pro Pro Val Thr Leu
 35 40 45

<210> 47
 <211> 48
 15 <212> PRT
 <213> Secuencia artificial

<220>
 <223> PPPVTLPPPV TLPPPVTLP PVTLPVTL PPPVTLPPPV TLPPPVTL
 20

<400> 47
 Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val
 1 5 10 15

Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro
 20 25 30

Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu
 35 40 45

<210> 48
 <211> 48
 25 <212> PRT
 <213> Secuencia artificial

<220>
 <223> PPPVSLPPP VSLPPPVS LPVSLPPPVS LPPPVS LPPP VSVPPPVS
 30

<400> 48

ES 2 623 147 T3

Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val
 1 5 10 15

Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro
 20 25 30

Pro Val Ser Leu Pro Pro Pro Val Ser Val Pro Pro Pro Val Ser Leu
 35 40 45

5 <210> 49
 <211> 48
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> PPPVSLPPP VSLPPPVS LPVSLPPPVS LPPPVS LPPPVS LPPPVS

<400> 49
 Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val
 1 5 10 15

Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro
 20 25 30

Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu
 35 40 45

15 <210> 50
 <211> 48
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> PPPAAAPPP AAAPPPAAAP PPAAAPPPAA APPPAAAPPP AAAPPPAAA

<400> 50
 Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala
 1 5 10 15

Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro
 20 25 30

Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala
 35 40 45

25 <210> 51
 <211> 48
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> PPPAAAPPP AAAPPPAAAP PPAAAPPPAA APPPAAAPPP AAAPPPVHL

<400> 51

ES 2 623 147 T3

<223> PPPAPAPPP APAPPPAPAP PPAPAPPPAP APPPAPAPPP APAPPPVHL

<400> 54

Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala
1 5 10 15

Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro
20 25 30

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

5

<210> 55

<211> 48

<212> PRT

<213> Secuencia artificial

10

<220>

<223> PPPAPAPPP APAPPPAPAP PPAPAPPPAP APPPAPAPPP APAPPPAPA

<400> 55

Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala
1 5 10 15

Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro
20 25 30

Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala
35 40 45

15

<210> 56

<211> 49

<212> PRT

20

<213> Secuencia artificial

<220>

<223> GPMGPRGPPG PAGAPGPQGF QGNPGEPGEP GVSGPMGPRG PPGPPPVHL

25

<400> 56

Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly
1 5 10 15

Pro Gln Gly Phe Gln Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val
20 25 30

Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Pro Val His
35 40 45

Leu

30

<210> 57

<211> 4

<212> PRT

<213> Secuencia artificial

<220>

35

<223> CXXC

<220>
 <221> característica mezcla
 <222> (2)..(3)
 <223> Xaa puede ser cualquier aminoácido natural
 5 <400> 57
Cys Xaa Xaa Cys
1
 <210> 58
 10 <211> 19
 <212> PRT
 <213> Secuencia artificial
 <220>
 15 <223> Secuencia señal
 <400> 58
Met Arg Val Leu Leu Val Ala Leu Ala Leu Leu Ala Leu Ala Ala Ser
1 5 10 15

Ala Thr Ser
 20 <210> 59
 <211> 23
 <212> PRT
 <213> Secuencia artificial
 25 <220>
 <223> Secuencia señal
 <400> 59
Met Phe Leu Lys Ser Pro Phe Tyr Ala Phe Leu Cys Phe Gly Gln Tyr
1 5 10 15

Phe Val Ala Val Thr His Ala
 30 **20**
 <210> 60
 <211> 93
 <212> PRT
 35 <213> Secuencia artificial
 <220>
 <223> RX3-LA1
 40 <400> 60

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Ala Pro Pro Pro Pro Ala Ala
1 5 10 15

Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro
20 25 30

Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro
35 40 45

Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Ala Pro Ala Arg Pro Ala Pro Ala Pro Ala Pro Ala
65 70 75 80

Pro Cys Pro Cys Ala Ala Pro Ala Pro Ser Pro Cys Gln
85 90

5 <210> 61
<211> 93
<212> PRT
<213> Secuencia artificial

10 <220>
<223> RX3-LA2

<400> 61
Thr His Thr Ser Gly Gly Cys Gly Cys Ala Pro Pro Pro Pro Ala Ala
1 5 10 15

Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro
20 25 30

Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro
35 40 45

Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Pro Ala Pro Ala Pro Pro Ala Pro Ala Pro Pro Ala Pro
65 70 75 80

Ala Cys Pro Cys Ala Ala Pro Ala Pro Ser Pro Cys Gln
85 90

15 <210> 62
<211> 48
<212> PRT
<213> Secuencia artificial

20 <220>
<223> RX3

<400> 62

ES 2 623 147 T3

Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val
 1 5 10 15

His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro
 20 25 30

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

5 <210> 63
 <211> 48
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> RX3(R)

<400> 63
 Pro Pro Pro Val Arg Leu Pro Pro Pro Val Arg Leu Pro Pro Pro Val
 1 5 10 15

Arg Leu Pro Pro Pro Val Arg Leu Pro Pro Pro Val Arg Leu Pro Pro
 20 25 30

Pro Val Arg Leu Pro Pro Pro Val Arg Leu Pro Pro Pro Val Arg Leu
 35 40 45

15 <210> 64
 <211> 48
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> RX3(K)

<400> 64
 Pro Pro Pro Val Lys Leu Pro Pro Pro Val Lys Leu Pro Pro Pro Val
 1 5 10 15

Lys Leu Pro Pro Pro Val Lys Leu Pro Pro Pro Val Lys Leu Pro Pro
 20 25 30

Pro Val Lys Leu Pro Pro Pro Val Lys Leu Pro Pro Pro Val Lys Leu
 35 40 45

25 <210> 65
 <211> 4
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> PXPP

35 <220>
 <221> característica mezcla
 <222> (2)..(2)
 <223> Xaa puede ser cualquier aminoácido natural

ES 2 623 147 T3

<400> 65
 Pro Xaa Pro Pro
 1

5 <210> 66
 <211> 48
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> RX3(A)

<400> 66
 Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val
 1 5 10 15

Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro
 20 25 30

Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu
 35 40 45

15 <210> 67
 <211> 48
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> RX3(L)

<400> 67
 Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val
 1 5 10 15

25 Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro
 20 25 30

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

30 <210> 68
 <211> 45
 <212> PRT
 <213> Secuencia artificial

35 <220>
 <223> GPMGPRGPPG PAGAPGPQGF QGNPGEPEGEP GVSGPMGPRG PPGPP

<400> 68
 Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly
 1 5 10 15

Pro Gln Gly Phe Gln Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val
 20 25 30

Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro
 35 40 45

ES 2 623 147 T3

<210> 69
 <211> 53
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> PPTPPTYETE KPLEPAPVEP SYEAEPTPPT RAPDQAEPNK PTPPTPVHLP PPP

<400> 69
 Pro Pro Thr Pro Pro Thr Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala
 1 5 10 15

Pro Val Glu Pro Ser Tyr Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala
 20 25 30

Pro Asp Gln Ala Glu Pro Asn Lys Pro Thr Pro Pro Thr Pro Val His
 35 40 45

10 Leu Pro Pro Pro Pro
 50

<210> 70
 <211> 46
 <212> PRT
 15 <213> Secuencia artificial

<220>
 <223> PPTPPTYETE KPLEPAPVEP SYEAEPTPPT RAPDQAEPNK PTPPTP

20 <400> 70
 Pro Pro Thr Pro Pro Thr Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala
 1 5 10 15

Pro Val Glu Pro Ser Tyr Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala
 20 25 30

Pro Asp Gln Ala Glu Pro Asn Lys Pro Thr Pro Pro Thr Pro
 35 40 45

25 <210> 71
 <211> 20
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> dominio ProX de gamma-zeína

<400> 71
 His Tyr Pro Tyr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
 1 5 10 15

Pro Cys Pro Cys
 20

35 <210> 72
 <211> 27
 <212> PRT

ES 2 623 147 T3

<213> Secuencia artificial

<220>

<223> CHYPTQPPRPQPHQPHPCPCQQPHPS

5

<400> 72

Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro
1 5 10 15

His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
20 25

<210> 73

10

<211> 30

<212> PRT

<213> Secuencia artificial

<220>

15

<223> CPPPPPPPPPPPPPPPPPCPCPPPPPPPCP

<400> 73

Cys Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro Pro Cys Pro Pro
20 25 30

<210> 74

20

<211> 30

<212> PRT

<213> Secuencia artificial

<220>

25

<223> CPPPPPPPPPPPPPPPPPCPPCPPPPPPCPP

<400> 74

Cys Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Cys Pro Pro Cys Pro Pro Pro Pro Pro Pro Pro Cys Pro Pro
20 25 30

<210> 75

30

<211> 27

<212> PRT

<213> Secuencia artificial

<220>

35

<223> CPPPAPAPPPAPAPPPAPCPCPAPAPP

<400> 75

Cys Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro
1 5 10 15

Ala Pro Cys Pro Cys Pro Ala Pro Ala Pro Pro
20 25

40

<210> 76

<211> 30

<212> PRT

ES 2 623 147 T3

<213> Secuencia artificial

<220>

<223> CHYPPAPAPPAPAPPAPACPCAAPAPSPCQ

5

<400> 76

Cys His Tyr Pro Pro Ala Pro Ala Pro Pro Ala Pro Ala Pro Pro Ala
1 5 10 15

Pro Ala Cys Pro Cys Ala Ala Pro Ala Pro Ser Pro Cys Gln
20 25 30

<210> 77

10

<211> 93

<212> PRT

<213> Secuencia artificial

<220>

15

<223> RX3(E)

<400> 77

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Glu
1 5 10 15

Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro
20 25 30

Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro
35 40 45

Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

20

<210> 78

<211> 93

<212> PRT

<213> Secuencia artificial

25

<220>

<223> RX3(D)

<400> 78

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Asp
1 5 10 15

Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro
20 25 30

Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro
35 40 45

Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

<210> 79
<211> 93
5 <212> PRT
<213> Secuencia artificial

<220>
<223> RX3(N)

10

<400> 79
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Asn
1 5 10 15

Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro
20 25 30

Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro
35 40 45

Pro Pro Val Asn Leu Pro Pro Pro Val Asn Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

15 <210> 80
<211> 93
<212> PRT
<213> Secuencia artificial

20 <220>
<223> RX3(Q)

<400> 80

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Gln
1 5 10 15

Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro
20 25 30

Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro
35 40 45

Pro Pro Val Gln Leu Pro Pro Pro Val Gln Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

<210> 81
<211> 93
5 <212> PRT
<213> Secuencia artificial

<220>
<223> RX3(A)

10 <400> 81
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Ala
1 5 10 15

Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro
20 25 30

Val Ala Leu Pro Pro Pro Val Ala Leu Pro Pro Pro Val Ala Leu Pro
35 40 45

Pro Pro Val Ala Leu Pro Pro Pro Val His Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

15 <210> 82
<211> 98
<212> PRT
<213> Secuencia artificial

20 <220>
<223> RX3(A)2

<400> 82

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Ala
1 5 10 15

Leu Pro Pro Pro Pro Val Ala Leu Pro Pro Pro Pro Val Ala Leu Pro
20 25 30

Pro Pro Pro Val Ala Leu Pro Pro Pro Pro Val Ala Leu Pro Pro Pro
35 40 45

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

Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro
65 70 75 80

His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro
85 90 95

Cys Gln

<210> 83

<211> 86

5 <212> PRT

<213> Secuencia artificial

<220>

<223> RX3(L)

10

<400> 83

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Leu
1 5 10 15

Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro
20 25 30

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

Pro Pro Val Leu Leu Pro Pro Pro Val Leu Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln
85

<210> 84

<211> 93

15 <212> PRT

<213> Secuencia artificial

<220>

ES 2 623 147 T3

<223> RX3(V)

<400> 84

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Val
1 5 10 15

Leu Pro Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro
20 25 30

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

Pro Pro Val Val Leu Pro Pro Pro Val Val Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

5

<210> 85

<211> 93

<212> PRT

10 <213> Secuencia artificial

<220>

<223> RX3(T)

15 <400> 85

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Thr
1 5 10 15

Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro
20 25 30

Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro
35 40 45

Pro Pro Val Thr Leu Pro Pro Pro Val Thr Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

<210> 86

<211> 93

20 <212> PRT

<213> Secuencia artificial

<220>

<223> RX3(S)

25

ES 2 623 147 T3

<400> 86

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Ser
1 5 10 15

Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro
20 25 30

Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro
35 40 45

Pro Pro Val Ser Leu Pro Pro Pro Val Ser Leu Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

- 5 <210> 87
- <211> 93
- <212> PRT
- <213> Secuencia artificial

- 10 <220>
- <223> RX3(A3)

<400> 87

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Ala
1 5 10 15

Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro
20 25 30

Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro
35 40 45

Pro Pro Ala Ala Ala Pro Pro Pro Ala Ala Ala Pro Pro Pro Pro Cys
50 55 60

His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His
65 70 75 80

Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
85 90

- 15 <210> 88
- <211> 93
- <212> PRT
- <213> Secuencia artificial

- 20 <220>
- <223> PP

<400> 88

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
20 25 30

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
35 40 45

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro His Leu Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
65 70 75 80

Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro Pro Cys Pro
85 90

5 <210> 89
<211> 93
<212> PRT
<213> Secuencia artificial

10 <220>
<223> PP3

<400> 89
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
20 25 30

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
35 40 45

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
65 70 75 80

Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro Pro Cys Pro
85 90

15 <210> 90
<211> 93
<212> PRT
<213> Secuencia artificial

20 <220>
<223> PA

<400> 90

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Pro
1 5 10 15

Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro
20 25 30

Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro
35 40 45

Pro Pro Ala Pro Ala Pro Pro Pro Ala His Leu Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala
65 70 75 80

Pro Cys Pro Cys Pro Ala Pro Ala Pro Pro Pro Cys Pro
85 90

5 <210> 91
<211> 93
<212> PRT
<213> Secuencia artificial

10 <220>
<223> PA2

<400> 91
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Pro
1 5 10 15

Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro
20 25 30

Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro
35 40 45

Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala
65 70 75 80

Pro Cys Pro Cys Pro Ala Pro Ala Pro Pro Pro Cys Pro
85 90

15 <210> 92
<211> 93
<212> PRT
<213> Secuencia artificial

20 <220>
<223> NPP

<400> 92

ES 2 623 147 T3

Thr His Pro Pro Pro Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
20 25 30

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
35 40 45

Pro Pro Pro Pro Pro Pro Pro Pro Pro His Leu Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
65 70 75 80

Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro Cys Pro
85 90

5 <210> 93
<211> 93
<212> PRT
<213> Secuencia artificial

10 <220>
<223> NPA

<400> 93
Thr His Pro Pro Pro Pro Cys Pro Cys Pro Pro Pro Pro Pro Ala Pro
1 5 10 15

Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro
20 25 30

Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro
35 40 45

Pro Pro Ala Pro Ala Pro Pro Pro Ala His Leu Pro Pro Pro Pro Cys
50 55 60

Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala
65 70 75 80

Pro Cys Pro Cys Pro Ala Pro Ala Pro Pro Pro Cys Pro
85 90

15 <210> 94
<211> 93
<212> PRT
<213> Secuencia artificial

20 <220>
<223> Z(Adh)Px

<400> 94

ES 2 623 147 T3

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

5 <210> 95
 <211> 86
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Z(Adh2)Px

<400> 95
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Thr Pro Pro Thr
 1 5 10 15
 Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala Pro Val Glu Pro Ser Tyr
 20 25 30
 Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala Pro Asp Gln Ala Glu Pro
 35 40 45
 Asn Lys Pro Thr Pro Pro Thr Pro Cys His Tyr Pro Thr Gln Pro Pro
 50 55 60
 Arg Pro Gln Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro
 65 70 75 80
 His Pro Ser Pro Cys Gln
 85

15 <210> 96
 <211> 94
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> Z(col)Px

<400> 96
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Gly Pro Met Gly Pro

ES 2 623 147 T3

```

1           5           10           15

Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln
          20           25           30

Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val Ser Gly Pro Met Gly
          35           40           45

Pro Arg Gly Pro Pro Gly Pro Pro Pro Val His Leu Pro Pro Pro Pro
          50           55           60

Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro
          65           70           75           80

His Pro Cys Pro Cys Gln Gln Pro His Pro Ser Pro Cys Gln
          85           90
    
```

<210> 97
 <211> 86
 5 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Z(col2)Px

10
 <400> 97
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Gly Pro Met Gly Pro
 1 5 10 15

```

Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln
          20           25           30
    
```

```

Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val Ser Gly Pro Met Gly
          35           40           45
    
```

```

Pro Arg Gly Pro Pro Gly Pro Pro Cys His Tyr Pro Thr Gln Pro Pro
          50           55           60
    
```

```

Arg Pro Gln Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro
          65           70           75           80
    
```

```

His Pro Ser Pro Cys Gln
          85
    
```

15 <210> 98
 <211> 67
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> Z(Adh)

<400> 98

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Thr Pro Pro Thr
1 5 10 15

Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala Pro Val Glu Pro Ser Tyr
20 25 30

Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala Pro Asp Gln Ala Glu Pro
35 40 45

Asn Lys Pro Thr Pro Pro Thr Pro Val His Leu Pro Pro Pro Pro Cys
50 55 60

Pro Pro Cys
65

<210> 99
<211> 60
5 <212> PRT
<213> Secuencia artificial

<220>
<223> Z(Adh2)

<400> 99

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Thr Pro Pro Thr
1 5 10 15

Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala Pro Val Glu Pro Ser Tyr
20 25 30

Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala Pro Asp Gln Ala Glu Pro
35 40 45

Asn Lys Pro Thr Pro Pro Thr Pro Cys Pro Pro Cys
50 55 60

<210> 100
15 <211> 64
<212> PRT
<213> Secuencia artificial

<220>
20 <223> Z(col)

<400> 100

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Gly Pro Met Gly Pro
1 5 10 15

Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln
20 25 30

ES 2 623 147 T3

Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val Ser Gly Pro Met Gly
 35 40 45

Pro Arg Gly Pro Pro Gly Pro Pro Pro Val His Leu Cys Pro Pro Cys
 50 55 60

5 <210> 101
 <211> 60
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Z(col2)

<400> 101
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Gly Pro Met Gly Pro
 1 5 10 15

Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln
 20 25 30

Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly Val Ser Gly Pro Met Gly
 35 40 45

Pro Arg Gly Pro Pro Gly Pro Pro Cys Pro Pro Cys
 50 55 60

15 <210> 102
 <211> 63
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> R8(4C)

<400> 102
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val His
 1 5 10 15

Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro
 20 25 30

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

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

25 <210> 103
 <211> 57
 <212> PRT
 <213> Secuencia artificial

30 <220>
 <223> R7(4C)

<400> 103

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val His
1 5 10 15

Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro
20 25 30

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

Pro Pro Val His Val Cys Pro Pro Cys
50 55

<210> 104

<211> 51

5 <212> PRT

<213> Secuencia artificial

<220>

<223> R6(4C)

10

<400> 104

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val His
1 5 10 15

Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro
20 25 30

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

Pro Pro Cys
50

<210> 105

15 <211> 45

<212> PRT

<213> Secuencia artificial

<220>

20 <223> R4(4C)

<400> 105

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val His
1 5 10 15

Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro
20 25 30

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

25

<210> 106

<211> 93

<212> PRT

<213> Secuencia artificial

<220>
<223> iRX3

5 <400> 106
 Gln Cys Pro Ser Pro His Pro Gln Gln Cys Pro Cys Pro His Pro Gln
 1 5 10 15
 Pro His Pro Gln Pro Arg Pro Pro Gln Thr Pro Tyr His Cys Pro Pro
 20 25 30
 Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His
 35 40 45
 Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro
 50 55 60
 Val His Leu Pro Pro Pro Val His Leu Pro Pro Pro Val His Leu Pro
 65 70 75 80
 Pro Pro Pro Gln Cys Gly Cys Gly Gly Ser Thr His Thr
 85 90

<210> 107
<211> 51
 10 <212> PRT
<213> Secuencia artificial

<220>
<223> PBIS

15

<220>
<221> REPETICIÓN
<222> (12)..(17)
 20 <223> La secuencia se puede repetir n veces

<220>
<221> REPETICIÓN
<222> (12)..(17)
 25 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 107
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Glu
 1 5 10 15
 Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
 20 25 30
 Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
 35 40 45
 Pro Cys Gln
 50

30

<210> 108

<211> 51
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> PBIS

<220>
 10 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir n veces

<220>
 15 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 108
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Asp
 1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
 20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
 35 40 45

20 Pro Cys Gln
 50

<210> 109
 <211> 51
 <212> PRT
 25 <213> Secuencia artificial

<220>
 <223> PBIS

30 <220>
 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir n veces

35 <220>
 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

40 <400> 109

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Asn
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

5 <210> 110
<211> 4
<212> PRT
<213> Secuencia artificial

10 <220>
<223> KDEL
<400> 110
Lys Asp Glu Leu
1

15 <210> 111
<211> 51
<212> PRT
<213> Secuencia artificial

20 <220>
<223> PBIS

25 <220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir n veces

30 <220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 111
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Gln
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

35 Pro Cys Gln
50

40 <210> 112
<211> 51
<212> PRT
<213> Secuencia artificial

<220>
<223> PBIS

5
<220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir n veces

10
<220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

15
<400> 112
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Ala
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

20
<210> 113
<211> 51
<212> PRT
<213> Secuencia artificial

25
<220>
<223> PBIS

30
<220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir n veces

35
<220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 113
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Thr
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

40
Pro Cys Gln
50

<210> 114
 <211> 50
 <212> PRT
 5 <213> Secuencia artificial

 <220>
 <223> PBIS

 10
 <220>
 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir n veces

 15
 <220>
 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

 20
 <400> 114
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Ser
 1 5 10 15

 Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
 20 25 30

 Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
 35 40 45

 Pro Cys
 50

 25 <210> 115
 <211> 51
 <212> PRT
 <213> Secuencia artificial

 <220>
 30 <223> PBIS

 <220>
 <221> REPETICIÓN
 35 <222> (12)..(17)
 <223> La secuencia se puede repetir n veces

 <220>
 <221> REPETICIÓN
 40 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

 <400> 115

ES 2 623 147 T3

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Leu
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

<210> 116
<211> 51
5 <212> PRT
<213> Secuencia artificial

<220>
10 <223> PBIS

<220>
<221> REPETICIÓN
<222> (12)..(17)
15 <223> La secuencia se puede repetir n veces

<220>
<221> REPETICIÓN
<222> (12)..(17)
20 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 116
Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Val Val
1 5 10 15

Leu Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

25 <210> 117
<211> 51
<212> PRT
<213> Secuencia artificial

30 <220>
<223> PBIS

<220>
35 <221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir n veces

<220>

<221> REPETICIÓN

<222> (12)..(17)

<223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

5 <400> 117

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Ala
1 5 10 15

Ala Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

<210> 118

<211> 46

10 <212> PRT

<213> Secuencia artificial

<220>

<223> PBIS

15

<220>

<221> REPETICIÓN

<222> (12)..(17)

20 <223> La secuencia se puede repetir n veces

<220>

<221> REPETICIÓN

<222> (12)..(17)

25 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 118

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Ala
1 5 10 15

Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro Ala Pro Ala Pro Pro Pro
20 25 30

Ala Pro Cys Pro Cys Pro Ala Pro Ala Pro Pro Pro Cys Pro
35 40 45

30 <210> 119

<211> 51

<212> PRT

<213> Secuencia artificial

35 <220>

<223> PBIS

<220>

40 <221> REPETICIÓN

<222> (12)..(17)

<223> La secuencia se puede repetir n veces

<220>

<221> REPETICIÓN

<222> (12)..(17)

5 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 119

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
20 25 30

Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
35 40 45

Pro Cys Gln
50

10 <210> 120
<211> 50
<212> PRT
<213> Secuencia artificial

15 <220>
<223> PBIS

20 <220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir n veces

25 <220>
<221> REPETICIÓN
<222> (12)..(17)
<223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 120

Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Pro Pro
1 5 10 15

Pro Pro Pro Pro Cys Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
20 25 30

30 Pro Pro Pro Pro Pro Pro Cys Pro Cys Pro Pro Pro Pro Pro Pro Pro
35 40 45

Cys Pro
50

35 <210> 121
<211> 51
<212> PRT
<213> Secuencia artificial

<220>
<223> PBIS

<220>
 <221> REPETICIÓN
 5 <222> (12)..(17)
 <223> La secuencia se puede repetir n veces

<220>
 <221> REPETICIÓN
 10 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

<400> 121
 Thr His Thr Ser Gly Gly Cys Gly Cys Gln Pro Pro Pro Pro Ala Pro
 1 5 10 15
 Ala Pro Pro Pro Pro Cys His Tyr Pro Thr Gln Pro Pro Arg Pro Gln
 20 25 30
 Pro His Pro Gln Pro His Pro Cys Pro Cys Gln Gln Pro His Pro Ser
 35 40 45
 Pro Cys Gln
 50

15 <210> 122
 <211> 51
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> PBIS

25 <220>
 <221> REPETICIÓN
 <222> (12)..(17)
 <223> La secuencia se puede repetir 4, 5, 6, 7 o 8 veces

30 <400> 122
 Thr His Thr Ser Gly Gly Cys Gly Cys Ala Pro Pro Pro Pro Ala Ala
 1 5 10 15
 Ala Pro Pro Pro Pro Cys His Tyr Pro Pro Ala Pro Ala Pro Pro Ala
 20 25 30
 Pro Ala Pro Pro Ala Pro Ala Cys Pro Cys Ala Ala Pro Ala Pro Ser
 35 40 45
 Pro Cys Gln
 50

35 <210> 123
 <211> 52
 <212> PRT
 <213> Secuencia artificial

40 <220>
 <223> PPPVELPPPV ELPPPVELPP PVELPPPVEL PPPVELPPPV ELPPPVELPP PP

ES 2 623 147 T3

<400> 123

Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val
1 5 10 15

Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro
20 25 30

Pro Val Glu Leu Pro Pro Pro Val Glu Leu Pro Pro Pro Val Glu Leu
35 40 45

Pro Pro Pro Pro
50

<210> 124

5 <211> 52

<212> PRT

<213> Secuencia artificial

<220>

10 <223> PPPVDLPPPV DLPPPVDLPP PVDLPPPVDL PPPVDLPPPV DVPPPVDLPP PP

<400> 124

Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val
1 5 10 15

Asp Leu Pro Pro Pro Val Asp Leu Pro Pro Pro Val Asp Leu Pro Pro
20 25 30

Pro Val Asp Leu Pro Pro Pro Val Asp Val Pro Pro Pro Val Asp Leu
35 40 45

Pro Pro Pro Pro
50

15 <210> 125

<211> 93

<212> PRT

<213> Secuencia artificial

20 <220>

<223> PP2

<400> 125

ES 2 623 147 T3

Thr His Pro Pro Pro Pro Cys Pro Pro Cys Pro Pro Pro Pro Pro Pro
 1 5 10 15

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
 20 25 30

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
 35 40 45

Pro Pro Pro Pro Pro Pro Pro Pro Pro His Leu Pro Pro Pro Pro Cys
 50 55 60

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
 65 70 75 80

Pro Cys Pro Pro Cys Pro Pro Pro Pro Pro Cys Pro Pro
 85 90

5 <210> 126
 <211> 30
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> QCPSPHPQQCPCPHPQPHPQPRPPQTPYHC

<400> 126
 Gln Cys Pro Ser Pro His Pro Gln Gln Cys Pro Cys Pro His Pro Gln
 1 5 10 15

Pro His Pro Gln Pro Arg Pro Pro Gln Thr Pro Tyr His Cys
 20 25 30

15 <210> 127
 <211> 30
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> CHYPTAPARPAPAPAPAPPCCAAPAPSPCQ

<400> 127
 Cys His Tyr Pro Thr Ala Pro Ala Arg Pro Ala Pro Ala Pro Ala Pro
 1 5 10 15

25 Ala Pro Cys Pro Cys Ala Ala Pro Ala Pro Ser Pro Cys Gln
 20 25 30

30 <210> 128
 <211> 46
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Adh

<400> 128

ES 2 623 147 T3

Pro Pro Thr Pro Pro Thr Tyr Glu Thr Glu Lys Pro Leu Glu Pro Ala
 1 5 10 15

Pro Val Glu Pro Ser Tyr Glu Ala Glu Pro Thr Pro Pro Thr Arg Ala
 20 25 30

Pro Asp Gln Ala Glu Pro Asn Lys Pro Thr Pro Pro Thr Pro
 35 40 45

<210> 129
 <211> 47
 5 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Col

10 <400> 129

Pro Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Ala Gly Ala Pro
 1 5 10 15

Gly Pro Gln Gly Phe Gln Gly Asn Pro Gly Glu Pro Gly Glu Pro Gly
 20 25 30

Val Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro Pro Pro
 35 40 45

15 <210> 130
 <211> 30
 <212> PRT
 <213> Secuencia artificial

20 <220>
 <223> QPPPPVHLPPPPCHYPTQPPRPQPHPQPHP

<400> 130
 Gln Pro Pro Pro Pro Val His Leu Pro Pro Pro Pro Cys His Tyr Pro
 1 5 10 15

Thr Gln Pro Pro Arg Pro Gln Pro His Pro Gln Pro His Pro
 20 25 30

25

REIVINDICACIONES

1. Una secuencia polipeptídica inductora de cuerpos proteicos (PBIS) recombinante que comprende
 - 5 (i) una estructura de poliprolina II (PPII) que tiene al menos 36 aminoácidos de longitud; en donde la estructura PPII está localizada entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C; en donde no más del 10% de los aminoácidos en la estructura PPII son lisina o arginina;
 - 10 y en donde la estructura PPII es una secuencia no anfipática y comprende la menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO: 17), o
 - 15 (ii) una secuencia rica en prolina que tiene al menos 36 aminoácidos de longitud, en donde de la secuencia rica en prolina está localizada entre al menos dos cisteínas en el extremo N y al menos dos cisteínas en el extremo C; en donde no más del 10% de los aminoácidos en la secuencia rica en prolina son lisina o arginina;
 - 20 y en donde la secuencia rica en prolina es una secuencia no anfipática y comprende la menos una repetición seleccionada del grupo que consiste en PPPVAL (SEQ ID NO:16), PPPVTL (SEQ ID NO:20), PPPVSL (SEQ ID NO:36), PPPVVL (SEQ ID NO:37) y PPPVLL (SEQ ID NO: 17).
 - 25 2. La PBIS recombinante de la reivindicación 1 en donde la estructura PPII o la región rica en prolina comprende al menos cinco repeticiones ricas en prolina, en donde cada una de las repeticiones ricas en prolina se selecciona independientemente del grupo que consiste en PPPXXX (SEQ ID NO:10), PPXX (SEQ ID NO:6), PX, PPPXX (SEQ ID NO:9), PPPX (SEQ ID NO:5), PPX y PPPXPX (SEQ ID NO:11) en donde X es cualquier aminoácido y/o en donde la longitud de la estructura PPII o la región rica en prolina es de aproximadamente 40 a aproximadamente 60 aminoácidos.
 - 30 3. La PBIS recombinante de la reivindicación 1 en donde la secuencia no anfipática se selecciona independientemente del grupo que consiste en SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49.
 - 35 4. La PBIS recombinante de cualquiera de las reivindicaciones 1 a 3 en donde el número de aminoácidos entre las dos cisteínas en el extremo C es un número entero de 0 a 20 y/o en donde el número de aminoácidos entre las dos cisteínas en el extremo N es un número entero de 0 a 20 y/o en donde el número de aminoácidos entre la cisteína C-terminal en el extremo N y el extremo N de la estructura PPII o entre la cisteína N-terminal en el extremo C y el extremo C de la estructura PPII es un número entero que se selecciona independientemente desde aproximadamente 1 hasta 30.
 - 40 5. La PBIS recombinante de cualquiera de las reivindicaciones 1 a 4 en donde las dos cisteínas en el extremo N o las dos cisteínas en el extremo C se encuentran en una segunda región rica en prolina.
 - 45 6. La PBIS recombinante de la reivindicación 5 en donde la segunda secuencia rica en prolina se selecciona del grupo que consiste en SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO: 73, SEQ ID NO:74, SEQ ID NO:75, SEQ ID NO:76, SEQ ID NO:127.
 - 50 7. La PBIS recombinante de la reivindicación 6 en donde la PBIS se selecciona del grupo que consiste en SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85 y SEQ ID NO: 86.
 - 55 8. Una proteína de fusión que comprende la PBIS recombinante de cualquiera de las reivindicaciones 1 a 7 y una proteína heteróloga.
 9. Una molécula de ácido nucleico que comprende una secuencia que codifica la PBIS recombinante de cualquiera de las reivindicaciones 1 a 7 o la proteína de fusión de la reivindicación 8.
 - 60 10. Un ensamblaje similar a cuerpo proteico recombinante (RPBLA) que comprende la PBIS recombinante de cualquiera de las reivindicaciones 1 a 7 o la proteína de fusión de la reivindicación 8.
 - 65 11. Una célula huésped que comprende la PBIS recombinante de cualquiera de las reivindicaciones 1 a 7, la proteína de fusión de la reivindicación 8, el ácido nucleico de la reivindicación 9 o el RPBLA de la reivindicación 10.

12. Un método para producir un RPBLA que comprende cultivar la célula de la reivindicación 11 en condiciones adecuadas para la formación del RPBLA y, opcionalmente purificar el cuerpo proteico recombinante.
- 5 13. Un método para purificar una proteína de fusión que comprende
- (i) proporcionar RPBLA que comprenden una proteína de fusión encerrada por membrana, en donde la proteína de fusión es la proteína de fusión de la reivindicación 8;
 - (ii) poner en contacto los RPBLA con un tampón acuoso que contiene una cantidad de un tensioactivo que desensambla la membrana;
 - 10 (iii) mantener el contacto durante un periodo de tiempo suficiente para desensamblar la membrana y a una temperatura que no desnaturaliza la proteína de fusión para separar la membrana de la proteína de fusión; y
 - (iv) recoger la proteína de fusión separada.
- 15 14. Una composición farmacéutica que comprende una cantidad terapéuticamente eficaz del RPBLA de la reivindicación 10, una vacuna que comprende una cantidad inmunogénicamente eficaz del RPBLA de la reivindicación 10 o un producto alimenticio que comprende el RPBLA de la reivindicación 10.

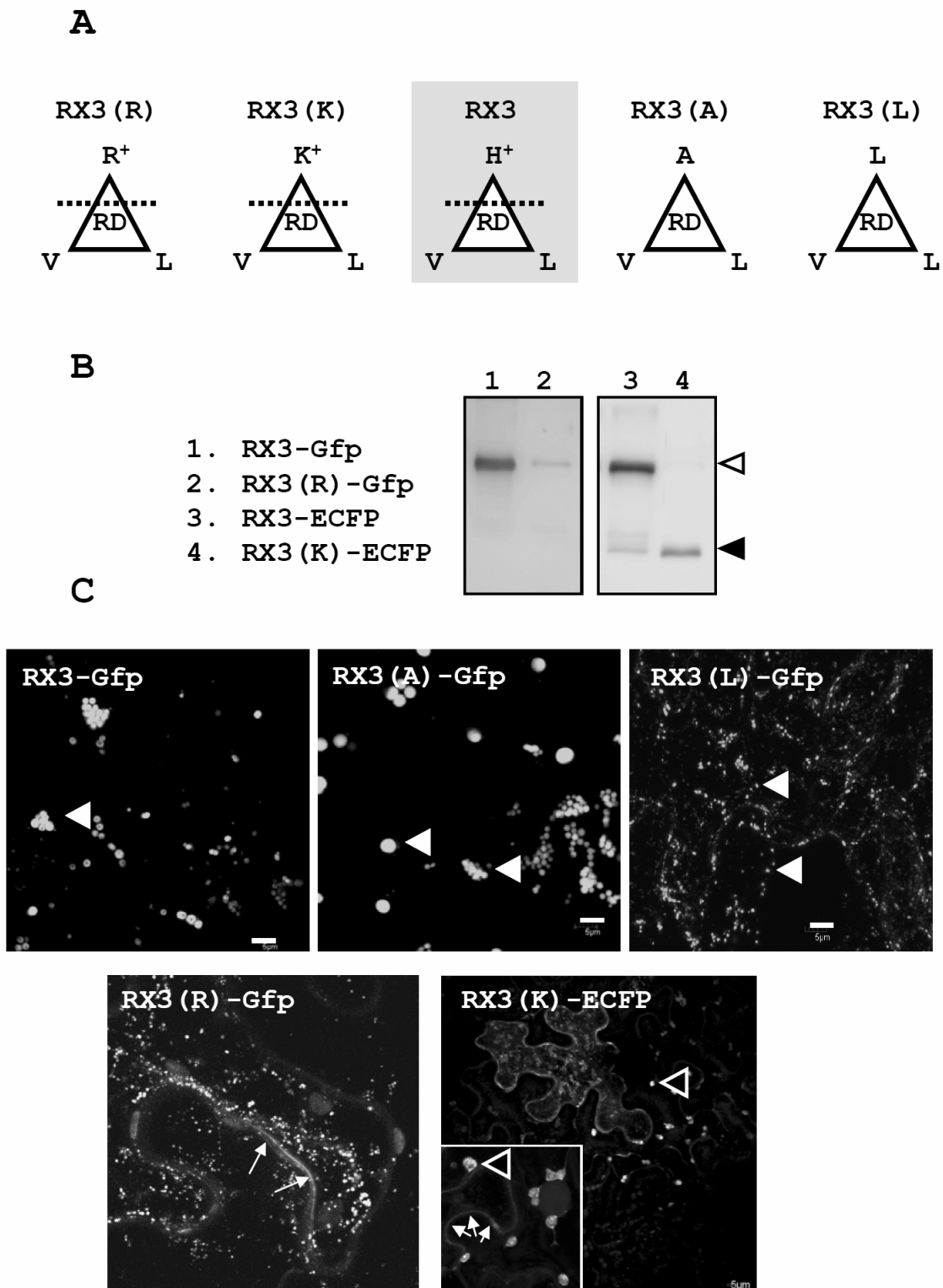


FIGURA 1

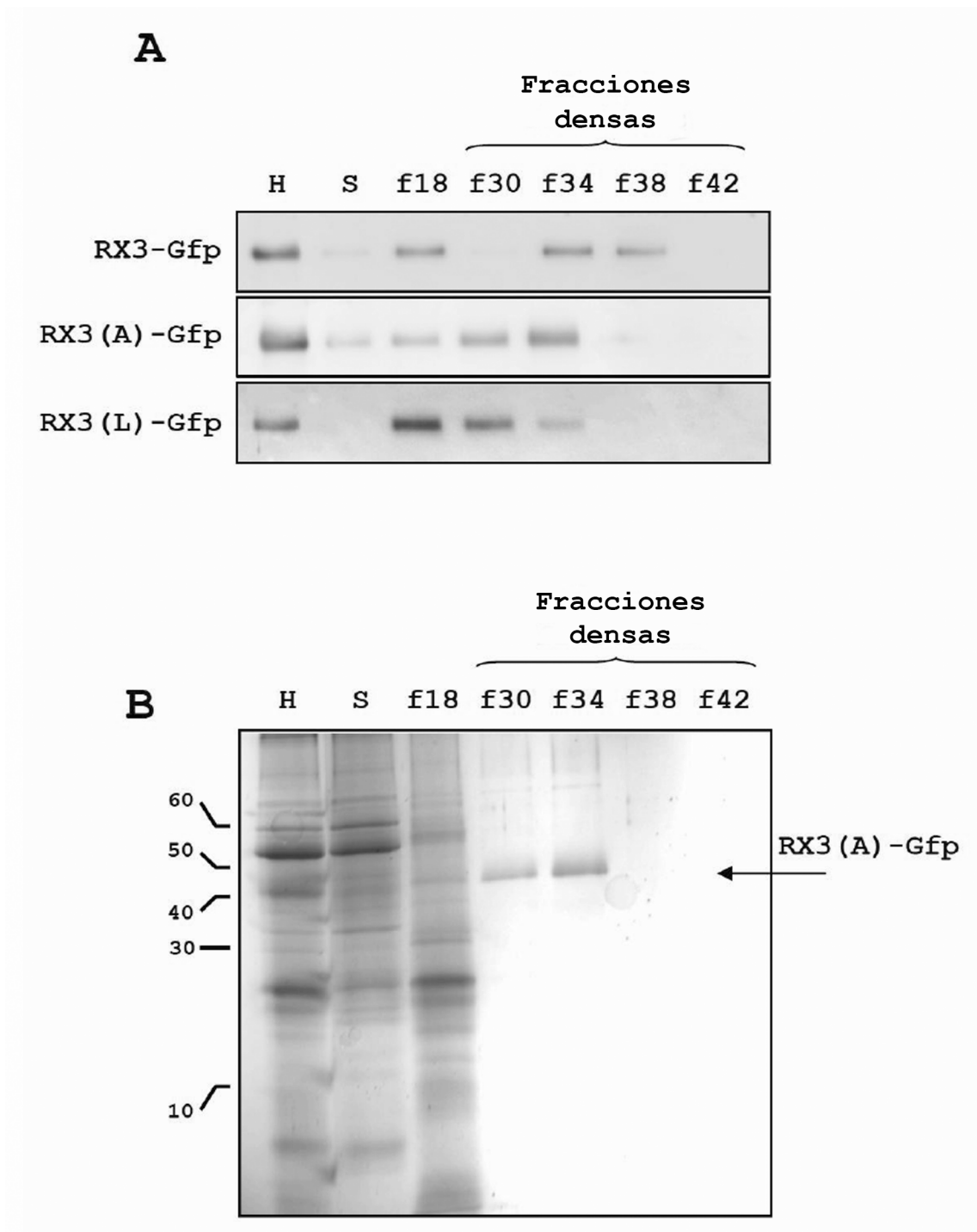
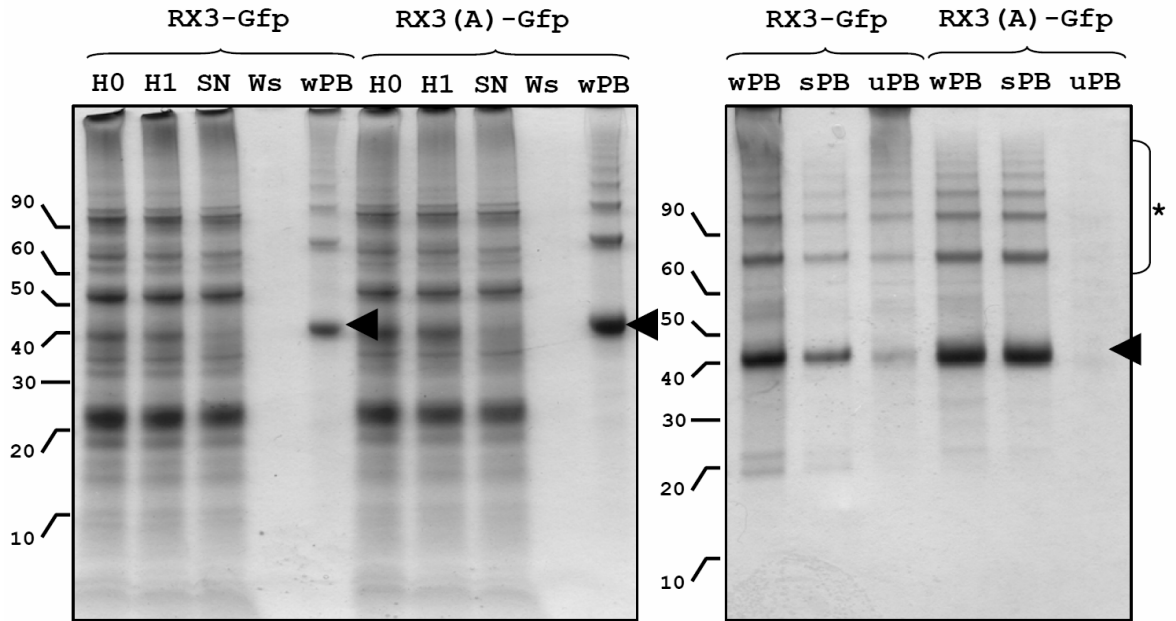


FIGURA 2

A



B

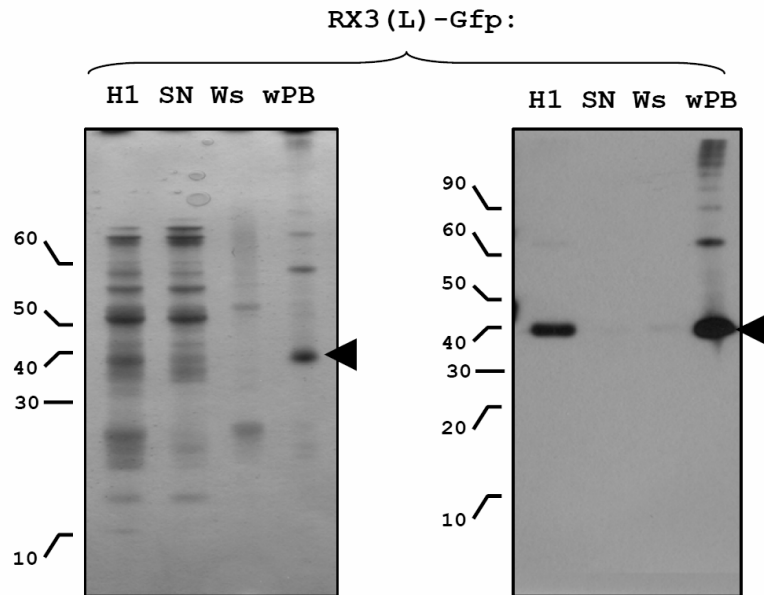
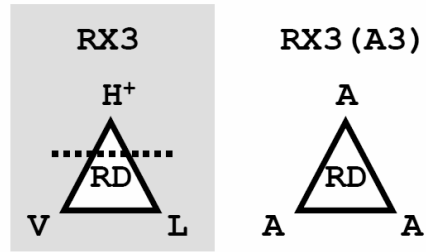


FIGURA 3

A



B

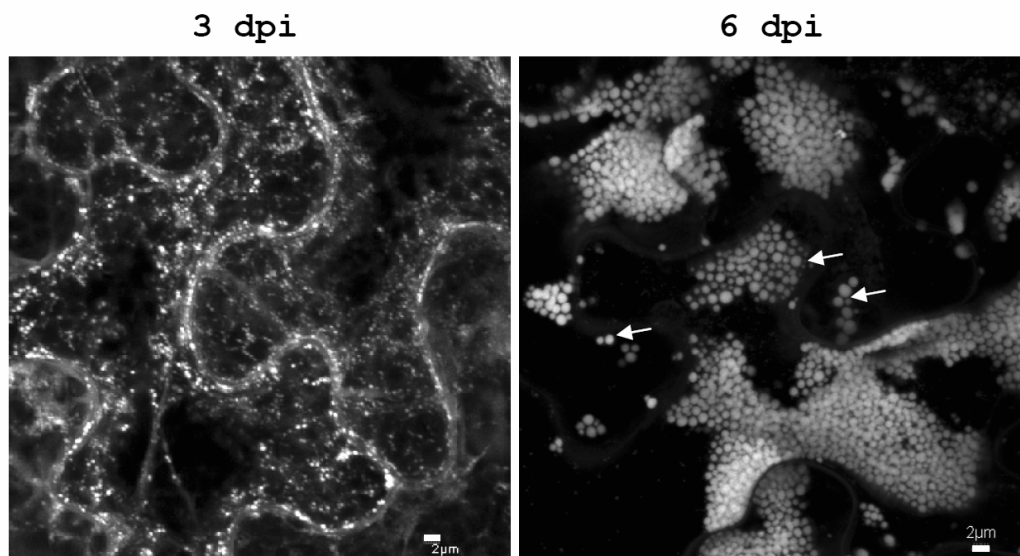
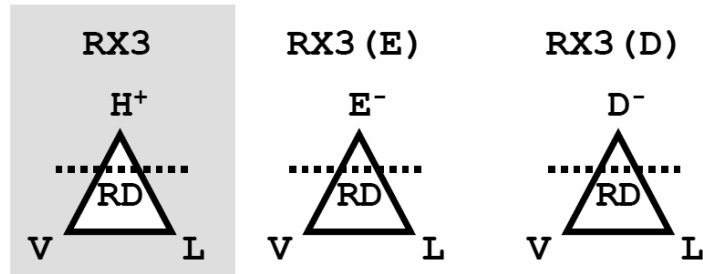
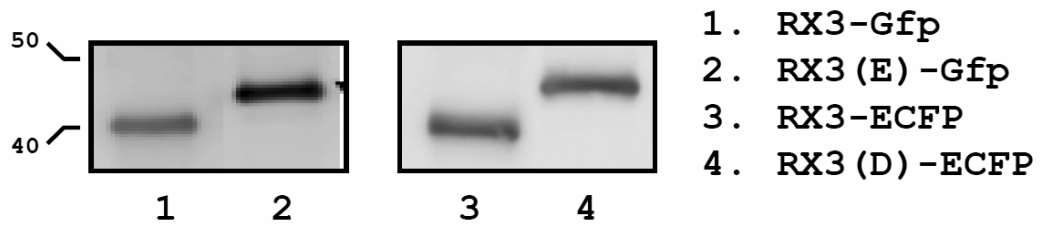


FIGURA 4

A



B



C

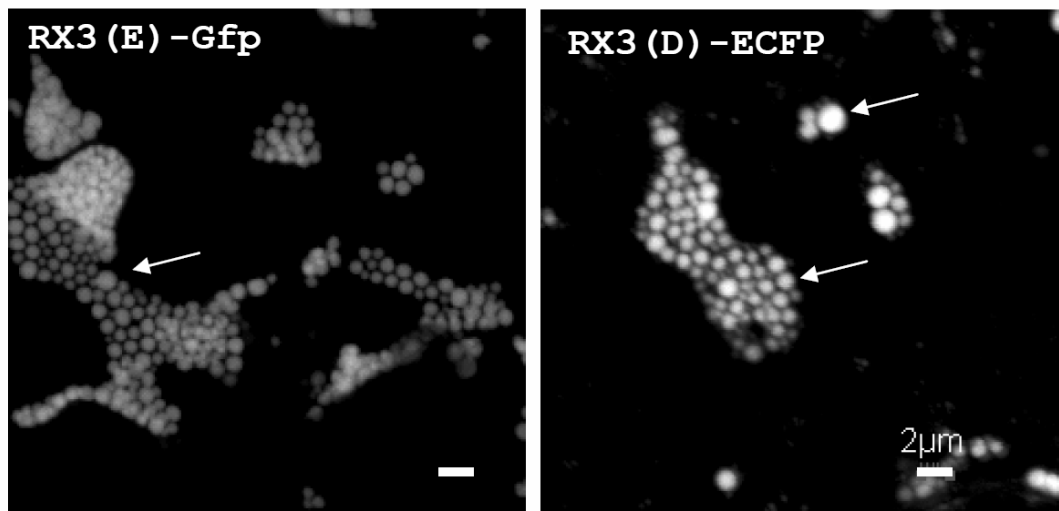


FIGURA 5

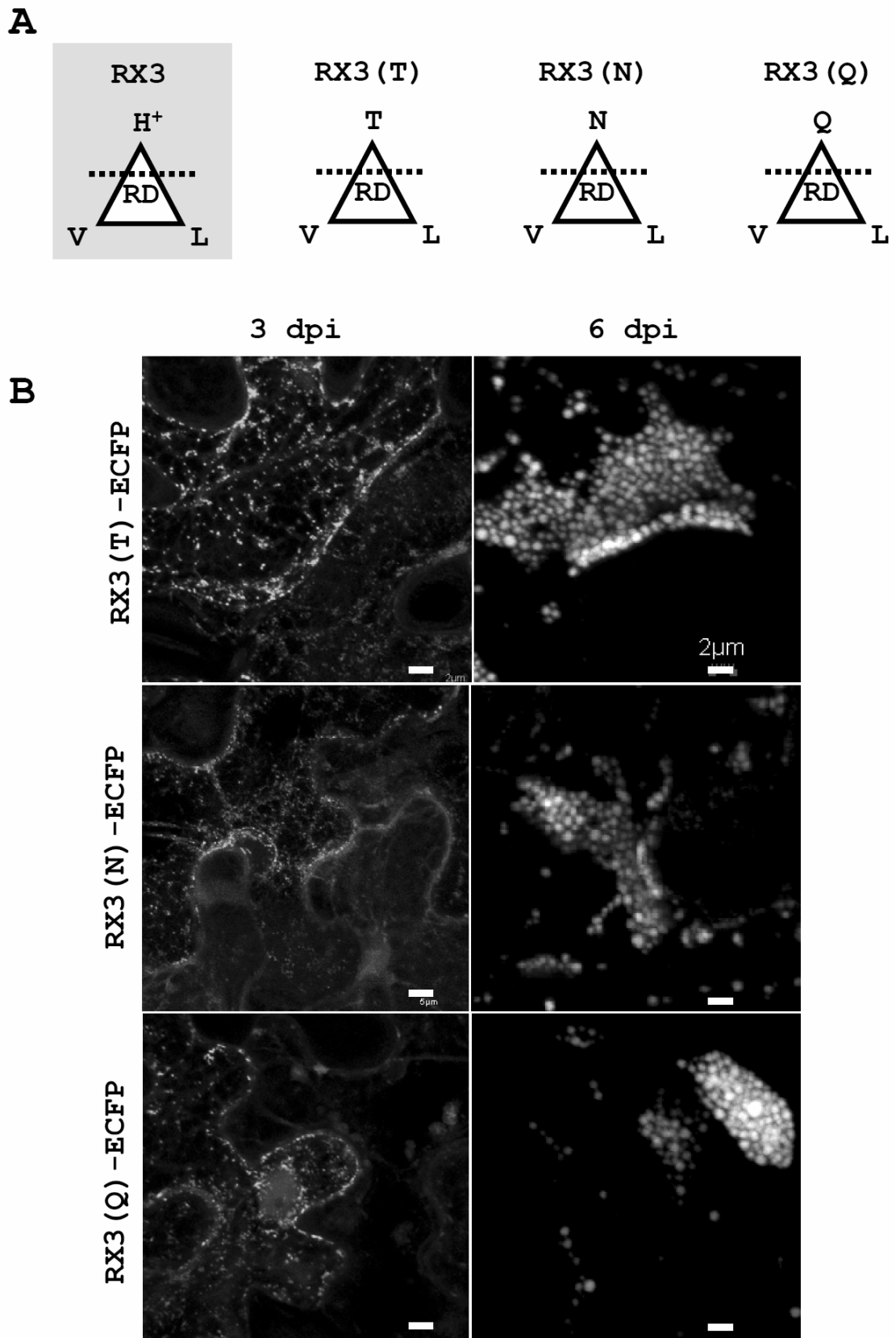


FIGURA 6

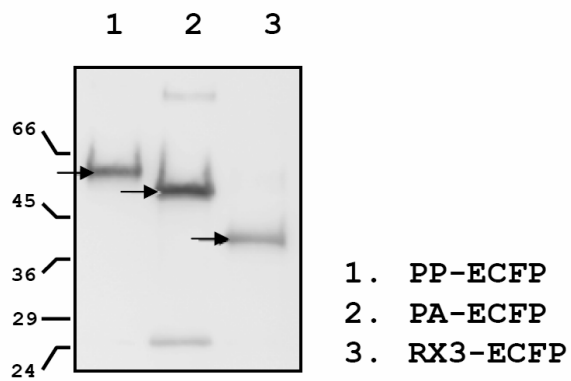
A

```

RX3: SP-THTSGGCGCQ P P P P V H L P P P V H L P P P V H L P P P V H L P P P V H L P P P V H L P P P
PP: SP-THTSGGCGCQ P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P
PA: SP-THTSGGCGCQ P P P P A P A P P P A P A P P P A P A P P P A P A P P P A P A P P P A P A P P P A P A P P P

V H V P P P V H L P P P C H Y P T Q P P R P Q P H P Q P H P C P C Q Q P H P S P C Q
P P P P P P H L P P P C P P P P P P P P P P P P P P P P P P P C P C P P P P P P P P C P
A P A P P P A H L P P P C P P P A P A P P P A P A P P P A P A P P P A P A P P P A P A P P P
    
```

B



C

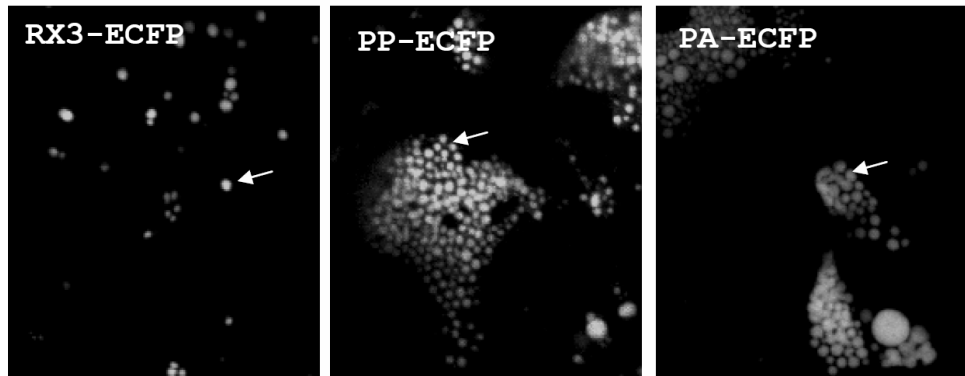


FIGURA 7

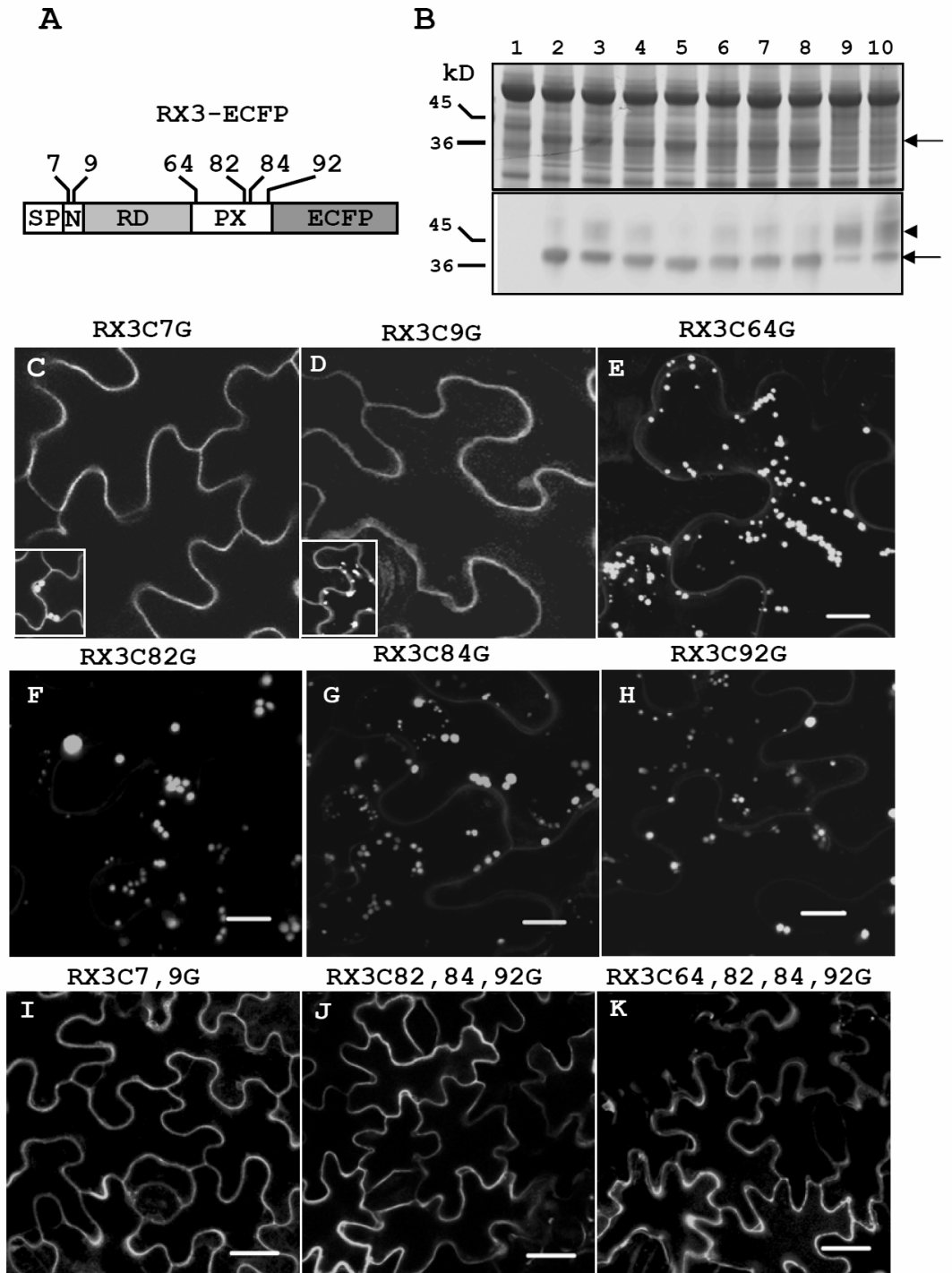


FIGURA 8

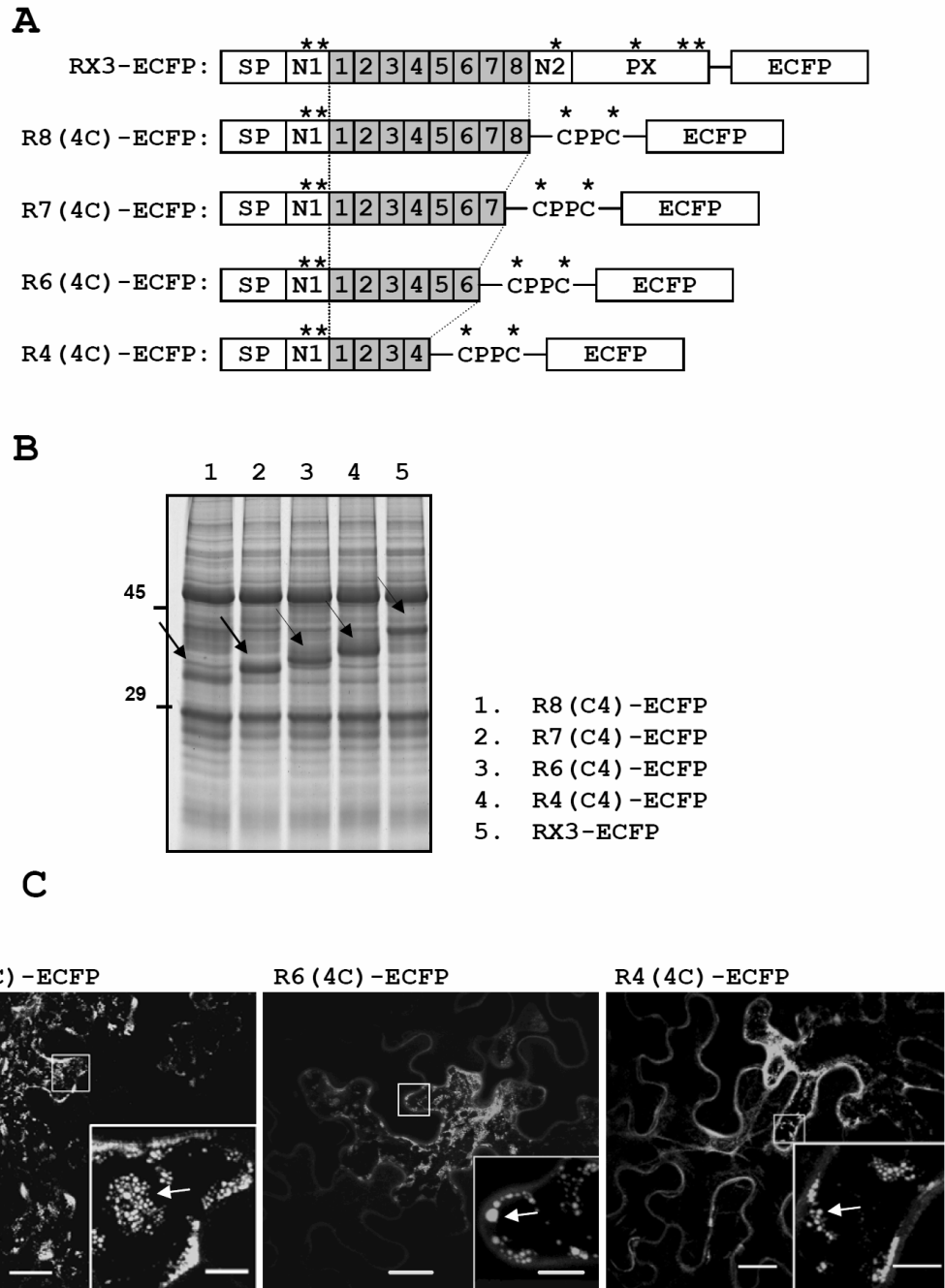


FIGURA 10

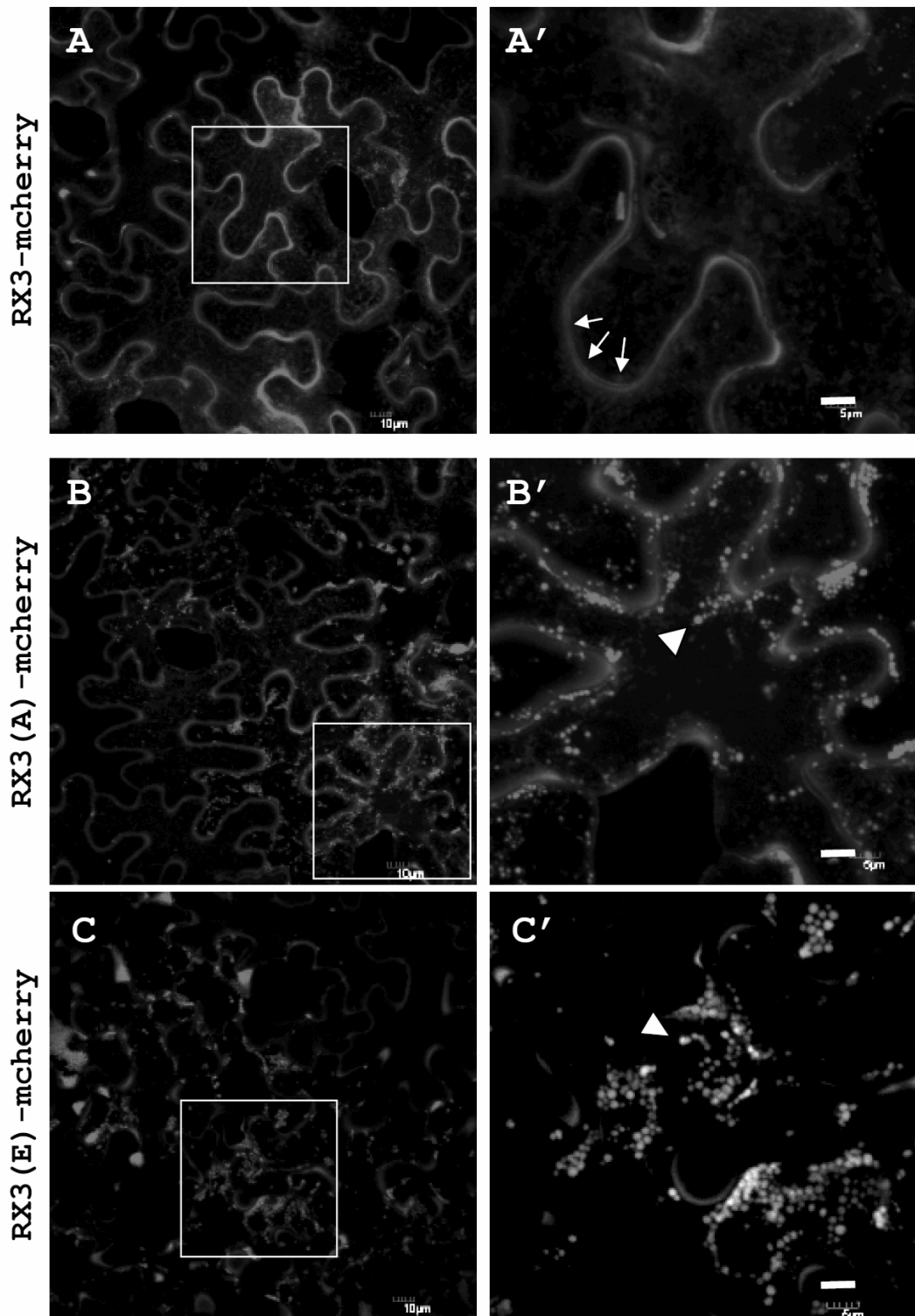


FIGURA 11

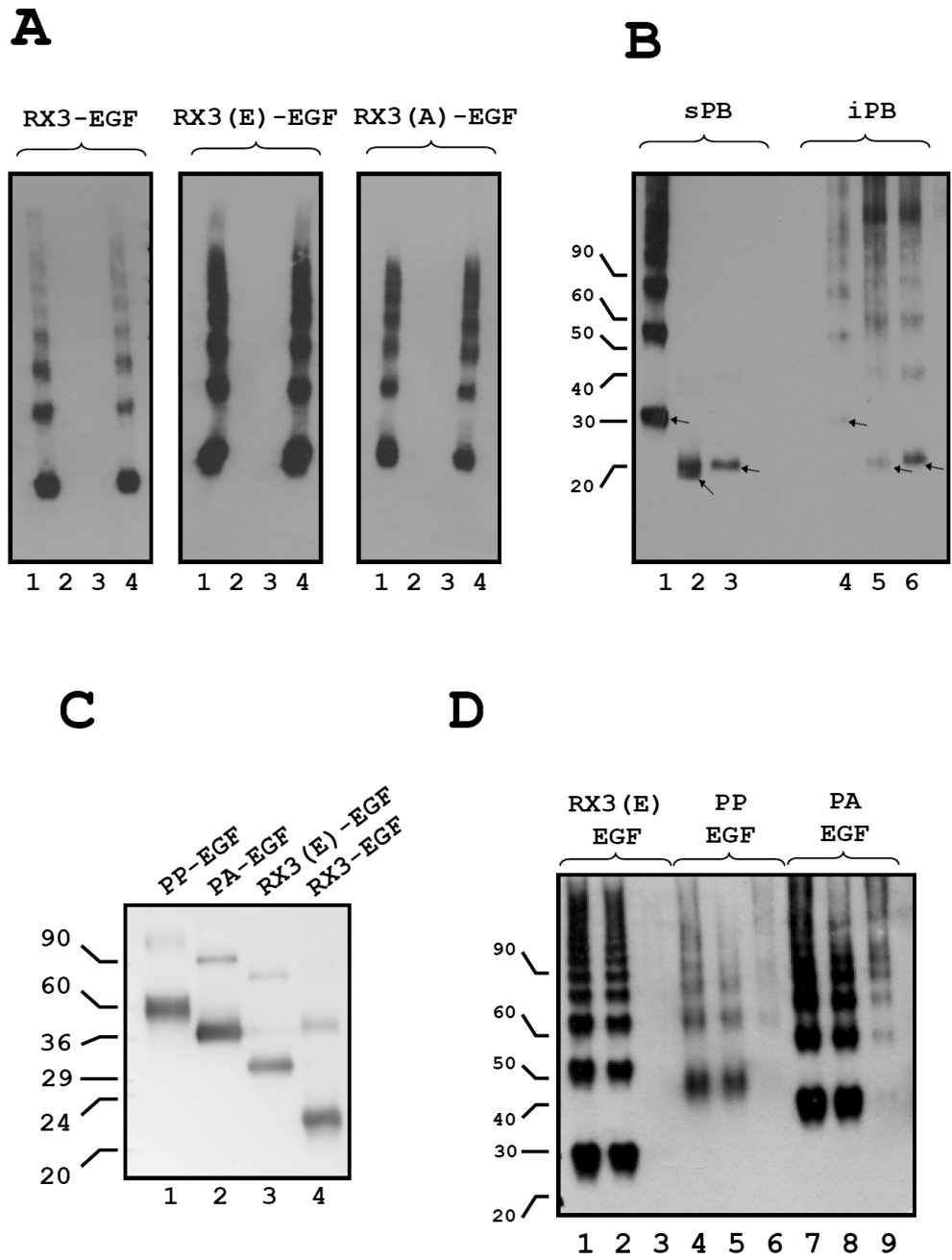


FIGURA 12

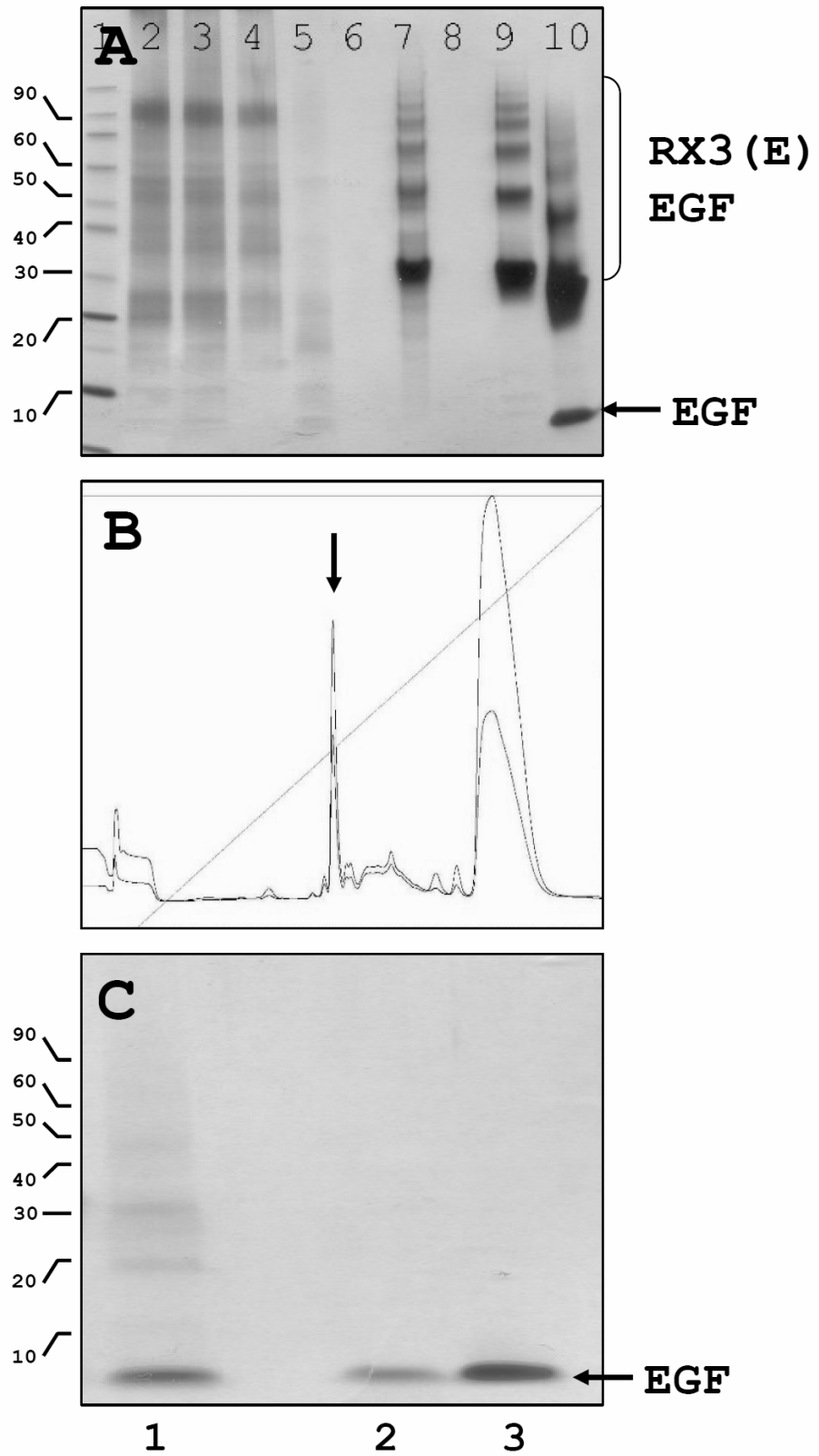


FIGURA 13

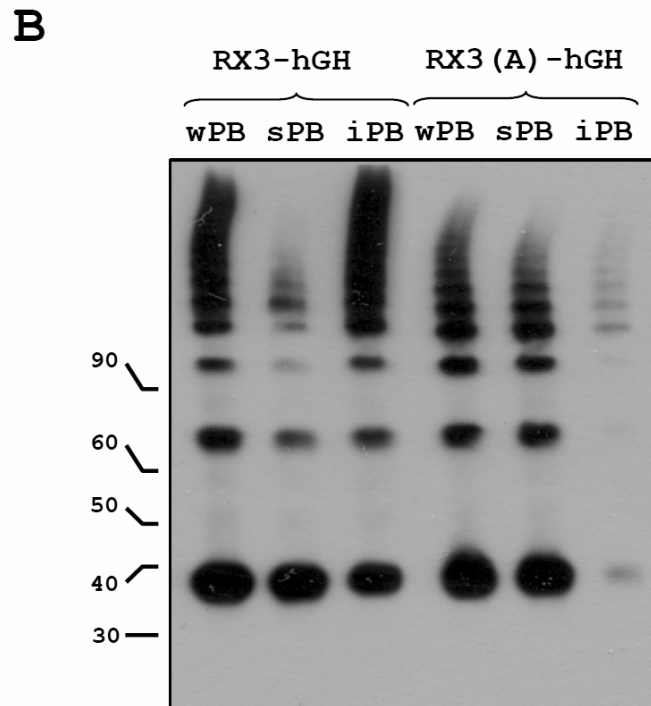
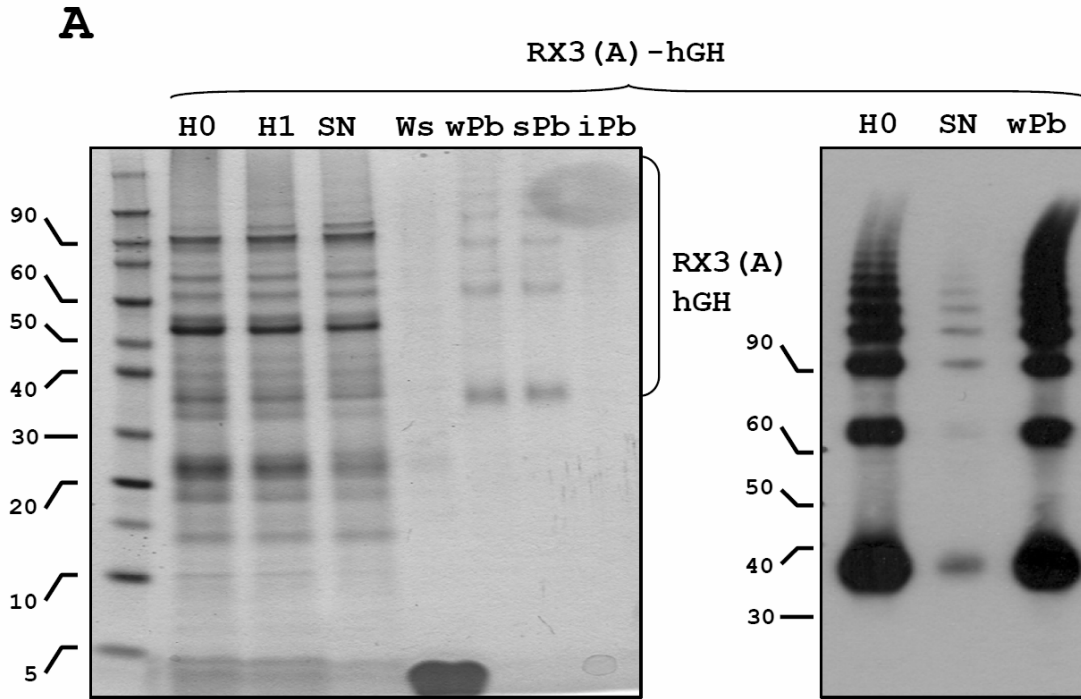


FIGURA 14

A

	PBIS	Enlazador	CS	PDI
RX3-mcherry:	RX3	(Gly) x5		mcherry
RX3 (A) -mcherry:	RX3 (A)	(Gly) x5		mcherry
RX3 (E) -mcherry:	RX3 (E)	(Gly) x5		mcherry
RX3-hGH:	RX3	(Gly) x5	EK	hGH
RX3 (A) -hGH:	RX3 (A)	(Gly) x5	EK	hGH
RX3-EGF:	RX3	(Gly) x5	FXa	EGF
RX3 (A) -EGF:	RX3 (A)	(Gly) x5	FXa	EGF
RX3 (E) -EGF:	RX3 (E)	(Gly) x5	FXa	EGF
PP-EGF:	PP	(Gly) x5	FXa	EGF
PA-EGF:	PP	(Gly) x5	FXa	EGF
RX3-Gfp:	RX3	(Gly) x5		Gfp
RX3 (A) -Gfp:	RX3 (A)	(Gly) x5		Gfp
RX3 (L) -Gfp:	RX3 (L)	(Gly) x5		Gfp
RX3 (R) -Gfp:	RX3 (R)	(Gly) x5		Gfp
RX3 (E) -Gfp:	RX3 (E)	(Gly) x5		Gfp
RX3dir-Gfp:	RX3dir			Gfp
RX3 (A) -Xyl:	RX3 (A)	(Gly) x5	EK	Xyl
RX3 (L) -Xyl:	RX3 (L)	(Gly) x5	EK	Xyl
RX3 (E) -Xyl:	RX3 (E)	(Gly) x5	EK	Xyl
RX3-Xyl:	RX3	(Gly) x5	EK	Xyl
PP-Xyl:	PP	(Gly) x5	EK	Xyl
PA-Xyl:	PA	(Gly) x5	EK	Xyl
Z (Adh) prot de fusión:	PBIS (Adh)	(Gly) x5		Gfp
Z (Col) prot de fusión:	PBIS (Col)	(Gly) x5		Gfp

	PDI	CS	Enlazador	PBIS
EK-RX3:	EKp		(Gly) x5	RX3
DsRED-I-RX3:	DsRED	I	(Gly) x5	RX3
hGH-I-RX3:	hGH	I	(Gly) x5	RX3
ECFP-RX3:	ECFP		(Gly) x5	RX3
ECFP-iRX3:	ECFP		(Gly) x5	iRX3

FIGURA 15

B

	PBIS	Enlazador	PDI
RX3-ECFP:	RX3	(Gly) x5	ECFP
RX3 (K) -ECFP:	RX3 (K)	(Gly) x5	ECFP
RX3 (A3) -ECFP:	RX3 (A3)	(Gly) x5	ECFP
RX3 (T) -ECFP:	RX3 (T)	(Gly) x5	ECFP
RX3 (N) -ECFP:	RX3 (N)	(Gly) x5	ECFP
RX3 (Q) -ECFP:	RX3 (Q)	(Gly) x5	ECFP
RX3 (D) -ECFP:	RX3 (D)	(Gly) x5	ECFP
PP-ECFP:	PP	(Gly) x5	ECFP
PA-ECFP:	PA	(Gly) x5	ECFP
NPP-ECFP:	NPP	(Gly) x5	ECFP
NPA-ECFP:	NPA	(Gly) x5	ECFP
R8 (C4) -ECFP:	R8 (C4)	(Gly) x5	ECFP
R7 (C4) -ECFP:	R7 (C4)	(Gly) x5	ECFP
R6 (C4) -ECFP:	R6 (C4)	(Gly) x5	ECFP
R4 (C4) -ECFP:	R4 (C4)	(Gly) x5	ECFP
RX3C7G-ECFP:	RX3C7G	(Gly) x5	ECFP
RX3C9G-ECFP:	RX3C9G	(Gly) x5	ECFP
RX3C64G-ECFP:	RX3C64G	(Gly) x5	ECFP
RX3C82G-ECFP:	RX3C82G	(Gly) x5	ECFP
RX3C84G-ECFP:	RX3C84G	(Gly) x5	ECFP
RX3C92G-ECFP:	RX3C92G	(Gly) x5	ECFP
RX3C7, 9G-ECFP:	RX3C7, 9G	(Gly) x5	ECFP
RX3C82, 84, 92G-ECFP:	RX3C82, 84, 92G	(Gly) x5	ECFP
RX3C64, 82, 84, 92G-ECFP:	RX3C64, 82, 84, 92G	(Gly) x5	ECFP
RX3 (PR10) -ECFP:	RX3 (PR10)	(Gly) x5	ECFP
PP2-ECFP:	PP2	(Gly) x5	ECFP
RX3C64, 82, 84, 92G-ECFP:	RX3C64, 82, 84, 92G	(Gly) x5	ECFP
iRX3-ECFP:	iRX3	(Gly) x5	ECFP

FIGURA 15 (cont.)

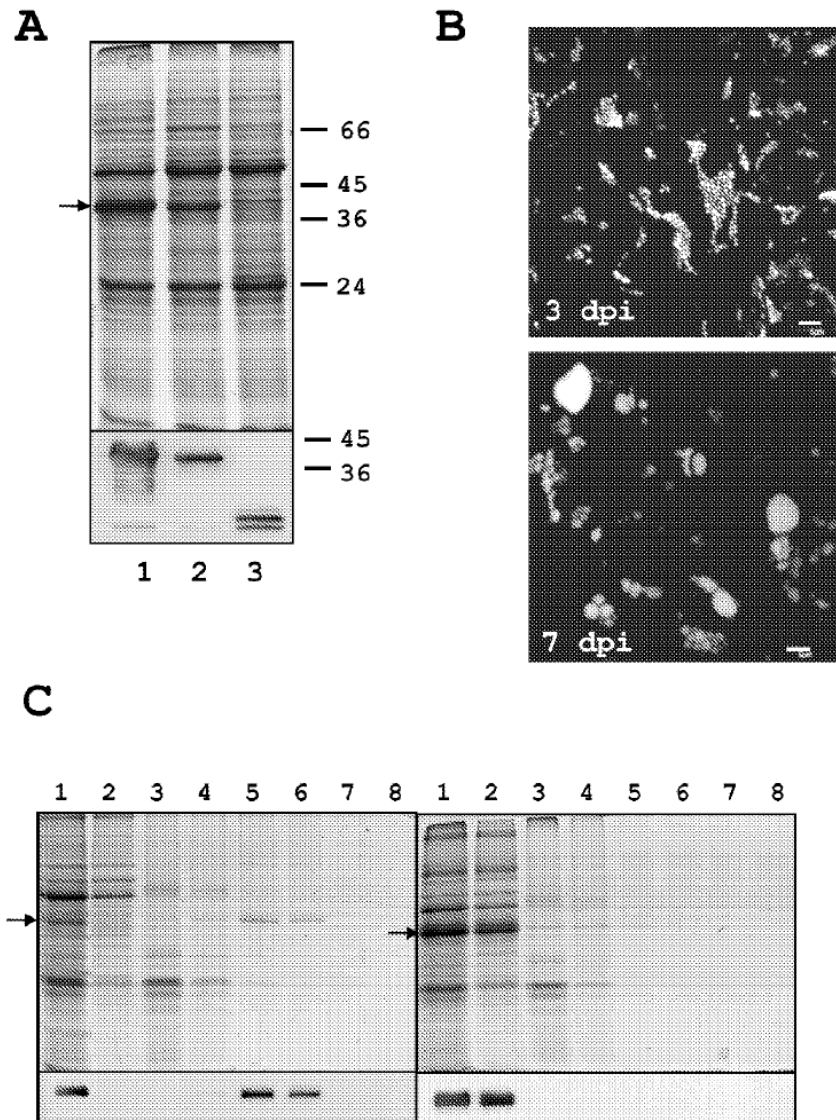


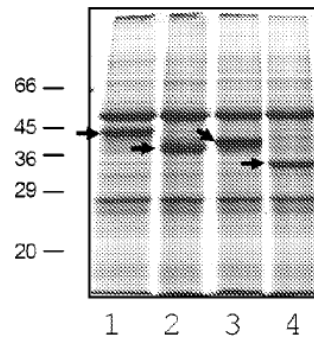
FIGURA 16

A

Z (Adh) : THTSGGCGCQ PPTPPTYETEKPLEPAPVEPSYEAETPPTRAPDQAEP
 Z (Adh) Ex: THTSGGCGCQ PPTPPTYETEKPLEPAPVEPSYEAETPPTRAPDQAEP
 Z (Col) : THTSGGCGCQ PGPMGPRGPPGAGAPGPQGFGNFGEPGEPGVSGPMG
 Z (Col) Ex: THTSGGCGCQ PGPMGPRGPPGAGAPGPQGFGNFGEPGEPGVSGPMG

Z (Adh) : NKPTPPTP- VHLPPPCPPC
 Z (Adh) Ex: NKPTPPTP- VHLPPPCHYPTQPPRPQPHQPHPCPCQQPHPSPCQ
 Z (Col) : PRGPPGPPP VHLPPPCPPC
 Z (Col) Ex: PRGPPGPPP VHLPPPCHYPTQPPRPQPHQPHPCPCQQPHPSPCQ

B



C

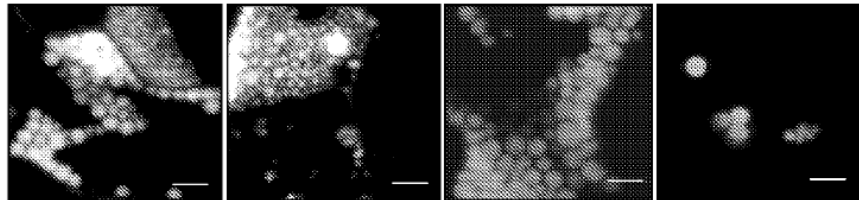


FIGURA 17

D

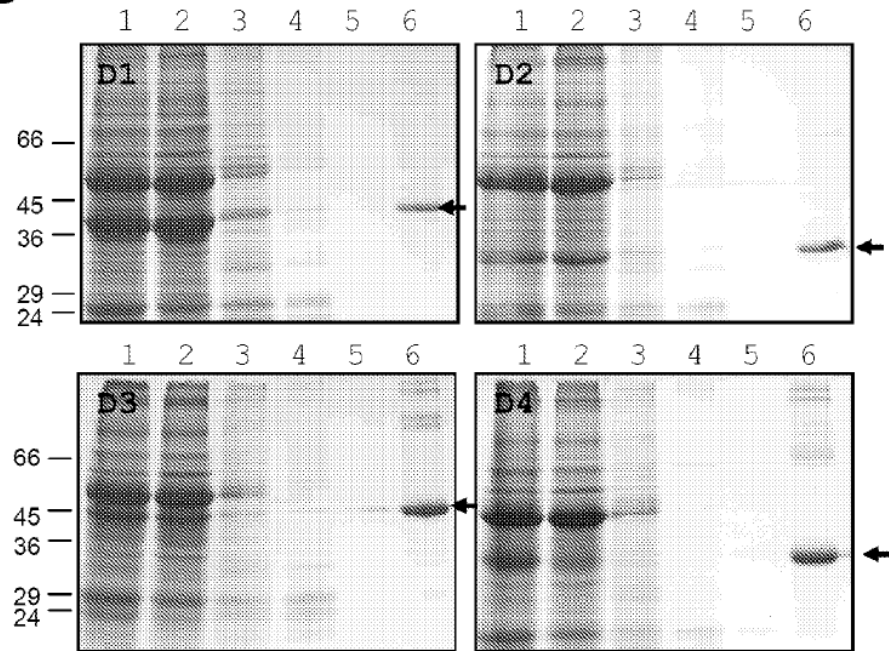


FIGURA 17 (cont.)

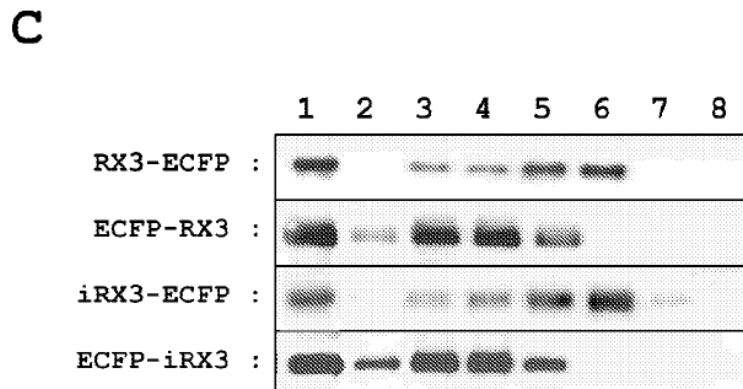
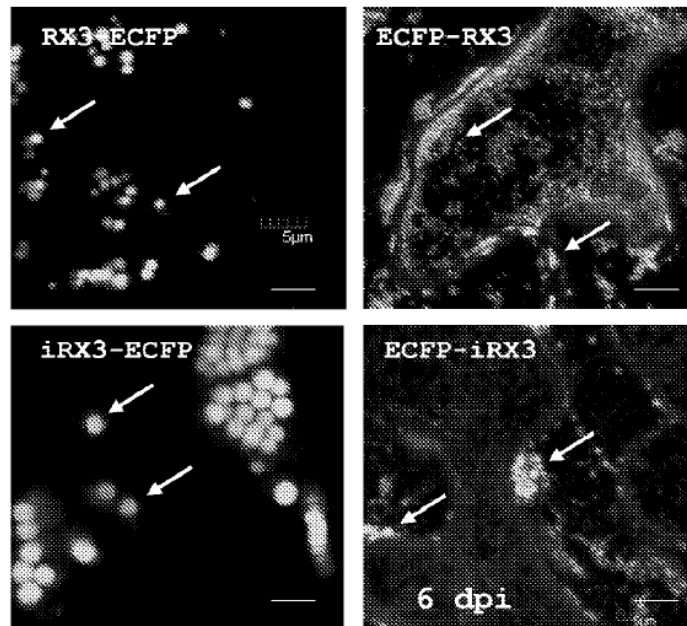
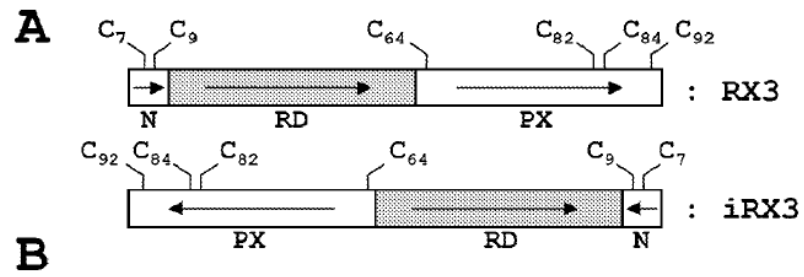


FIGURA 18

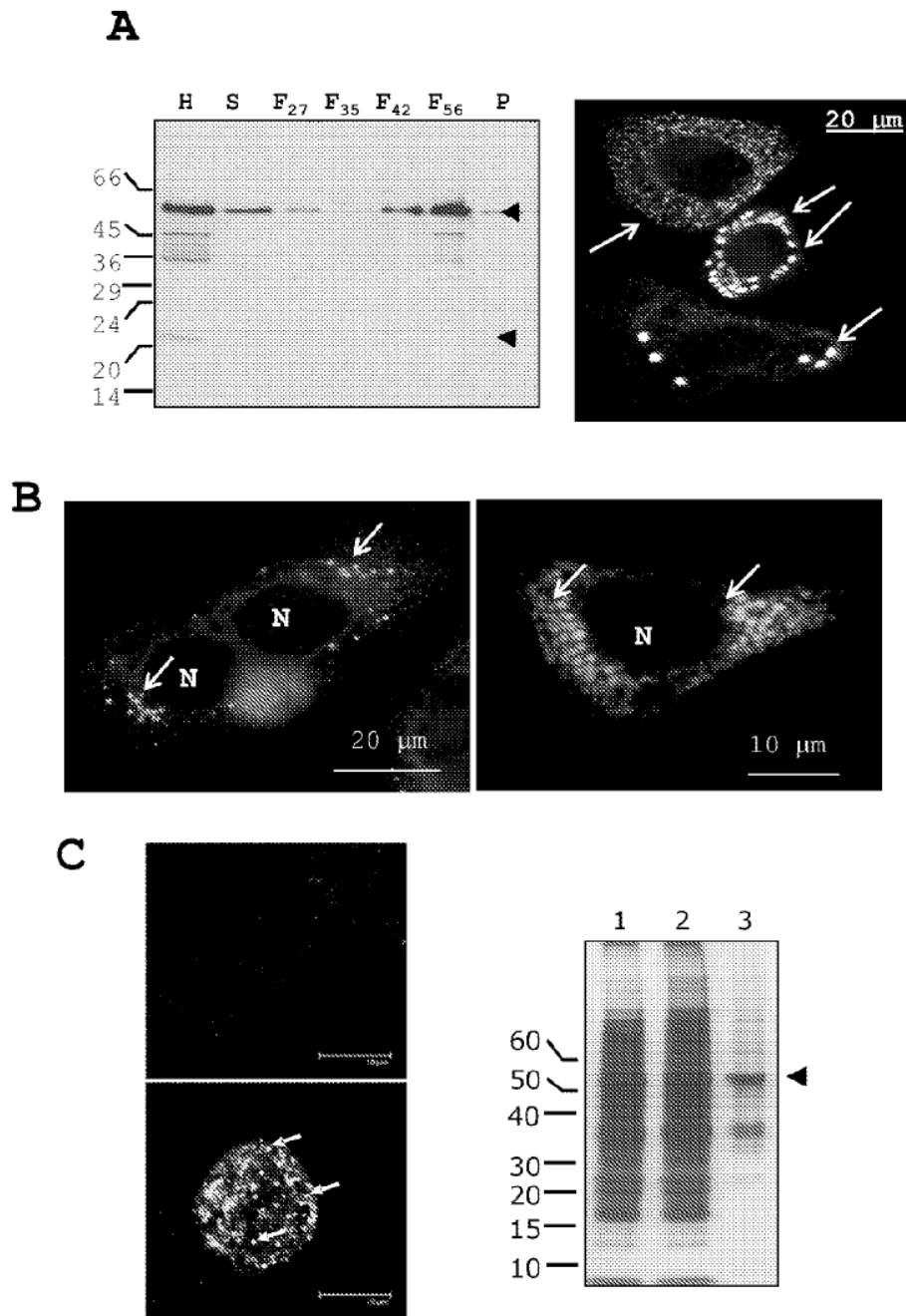


FIGURA 19