

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 621 320**

51 Int. Cl.:

**C12N 15/70** (2006.01)

**C07K 16/24** (2006.01)

**C07K 16/28** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.08.2013 PCT/US2013/053562**

87 Fecha y número de publicación internacional: **13.02.2014 WO14025663**

96 Fecha de presentación y número de la solicitud europea: **05.08.2013 E 13756743 (4)**

97 Fecha y número de publicación de la concesión europea: **11.01.2017 EP 2880168**

54 Título: **Sistema de coexpresión inducible**

30 Prioridad:

**05.08.2012 US 201261679751 P**  
**29.12.2012 US 201261747246 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**03.07.2017**

73 Titular/es:

**ABSCI, LLC (100.0%)**  
**101 E. 6th Street, Suite 300**  
**Vancouver, WA 98660, US**

72 Inventor/es:

**MCCLAIN, SEAN y**  
**VALASEK, MARK**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 621 320 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

Sistema de coexpresión inducible

Campo de la invención

5 La presente invención pertenece a los campos técnicos generales de la biología molecular y la fabricación biotecnológica. Más particularmente, la presente invención pertenece al campo técnico de la expresión de proteínas recombinantes.

Antecedentes de la invención

10 La producción de sustancias biotecnológicas es un proceso complejo, más aún cuando el producto deseado es una combinación de moléculas codificadas por genes diferentes, tal como una proteína multímera formada a partir de dos o más polipéptidos diferentes. Una coexpresión con éxito de múltiples productos génicos requiere superar una serie de retos, que se ven agravados por la expresión simultánea de más de un producto génico. Los problemas que deben superarse incluyen la creación de vectores de expresión compatibles cuando se emplea más de un tipo de vector; la obtención de la relación estequiométrica correcta de los productos; la producción de productos génicos que se pliegan correctamente y con la conformación apropiada con respecto a las parejas de unión; la purificación de los productos deseados, alejados de células y proteínas no deseadas, tales como proteínas que se pliegan de forma incorrecta y/o que tienen una conformación incorrecta; y la reducción al mínimo de la formación de cuerpos de inclusión, como un aspecto para maximizar el rendimiento del o de los productos deseados. Se han adoptado muchos enfoques diferentes para hacer frente a estos retos, pero todavía existe una necesidad de métodos de coexpresión más adecuados.

20 Se han ideado diversos sistemas de expresión inducible de proteínas bacterianas, que incluyen plásmidos que contienen los promotores *lac* y *ara*, para expresar proteínas individuales. Estos sistemas tienen una utilidad limitada en la coexpresión de proteínas difíciles de expresar, ya que no pueden inducir a las proteínas de forma homogénea en toda la población del cultivo en crecimiento en *E. coli* de tipo silvestre (Khlebnikov y Keasling, "Effect of *lacY* expression on homogeneity of induction from the  $P_{lac}$  and  $P_{trc}$  promoters by natural and synthetic inducers", Biotechnol Prog 2002 May-Jun; 18(3): 672- 674). Cuando la expresión de las proteínas de transporte para inductores es dependiente de la presencia de un inductor, como es el caso para los sistemas *lac* y *ara* de *E. coli* de tipo silvestre, la concentración celular del inductor debe alcanzar un nivel umbral para iniciar la producción de las proteínas de transporte, pero una vez que se ha alcanzado ese umbral, se puede producir un bucle de retroalimentación positiva descontrolado, en donde el resultado es un alto nivel de inductor en la célula y niveles elevados correspondientes de expresión desde los promotores inducibles: el fenómeno de "todo o nada". El aumento de la concentración del inductor en el medio de crecimiento, aumenta la proporción de células en la población que está en modo de expresión elevada. Aunque este tipo de sistema da como resultado una inducción dependiente de la concentración de la expresión de proteínas a escala de población, es subóptimo para la expresión y la producción de proteínas que requieren un control estricto de la expresión, incluyendo las que son tóxicas, tienen una solubilidad inadecuada o requieren concentraciones específicas por otras razones.

35 Se han realizado algunos esfuerzos para abordar el fenómeno de inducción "todo o nada" en sistemas de expresión de un solo promotor, eliminando el transporte del inductor dependiente de inductor. Un ejemplo es tener una mutación nula en el gen de la lactosa permeasa (*lacYam*) y emplear un inductor alternativo del promotor *lac*, tal como IPTG (isopropil-tio- $\beta$ -D-galactósido), que puede pasar a través de la membrana celular en cierto grado en ausencia de un transportador (Jensen *et al.*, "The use of *lac*-type promoters in control analysis", Eur J Biochem 1993 Jan 15; 211(1-2): 181-191). Otro enfoque es el uso de un promotor inducible con arabinosa en una cepa que carece de los genes transportadores de arabinosa, pero que tiene una mutación en el gen de la lactosa permeasa, *lacY(A117C)*, que le permite el transporte de arabinosa dentro de la célula (Morgan-Kiss *et al.*, "Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity", Proc Natl Acad Sci U S A, 2002 May 28; 99(11): 7373-7377).

40 Los componentes de sistemas de expresión de proteínas individuales son frecuentemente incompatibles, lo que impide su uso en sistemas de coexpresión, ya que pueden verse afectados negativamente por los efectos de "comunicación cruzada" entre diferentes sistemas de inductor-promotor, o requieren modificaciones genómicas que se excluyen mutuamente, o están sujetos a una regulación del metabolismo en general. Un intento de abordar el problema de "comunicación cruzada" entre los sistemas de promotores inducibles *lac* y *ara* incluía un desarrollo dirigido del activador de la transcripción AraC para mejorar su capacidad para inducir el promotor *araBAD* en presencia de IPTG, un inductor del promotor *lac* (Lee *et al.*, "Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters", Appl Environ Microbiol 2007 Sep; 73(18): 5711-5715; Epub 20 de julio 2007). Sin embargo, la compatibilidad entre vectores de expresión basados en los promotores inducibles *lac* y *ara* es todavía limitada debido a la necesidad de modificaciones genómicas mutuamente excluyentes: una mutación puntual de *lacY* (*lacY(A117C)*) para la inducción homogénea del promotor *araBAD* mediante arabinosa, y un gen *lacY* nulo para la inducción homogénea del promotor *lac* mediante IPTG. La regulación del metabolismo en general - por ejemplo, la represión por catabolito de carbono (CCR) - también puede afectar a la compatibilidad de los promotores inducibles. La CCR se caracteriza por la represión de genes necesarios para la utilización de un compuesto que

contiene carbono cuando está presente un compuesto más preferido, como se observa en la utilización preferente de la glucosa antes que otros azúcares. En el caso de los sistemas de promotores inducibles *ara* y *prp*, la presencia de arabinosa reduce la capacidad del propionato para inducir la expresión a partir del promotor *prpBCDE*, un efecto que se cree que implica CCR (Park *et al.*, "The mechanism of sugar-mediated catabolite repression of the propionate catabolic genes in *Escherichia coli*", Gene 2012 Aug 1; 504(1): 116-121, Epub 3 de mayo 2012). Existe claramente una necesidad de un sistema de coexpresión inducible que supere estos problemas.

El concepto de emplear dos promotores inducibles diferentes en dos estructuras artificiales en *E. coli*, está descrito por VAN DIEN *et al.*, "Manipulation of independent synthesis and degradation of polyphosphate in *Escherichia coli* for investigation of phosphate secretion from the cell" APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 63, n° 5, 1997 May 1 (1-05-1997), páginas 1689-1695, en donde un sistema de promotor de este tipo se emplea para la expresión secuencial de la polifosfato cinasa (PPK) y exopolifosfatasa (PPX) de *E. coli*. IPTG y arabinosa se emplearon como inductores diferentes.

En el documento EP 2386641, se describe un método que implica la coexpresión de una porina en la membrana externa del microorganismo a partir de un plásmido y una proteína de interés a partir de otro plásmido. Los plásmidos tienen diferentes promotores inducibles (propionato y arabinosa a través de pBAD).

En HAYHURST A *et al.*: "*Escherichia coli* Skp Chaperone Coexpression Improves Solubility and Phage Display of Single-Chain Antibody Fragments", PROTEIN EXPRESSION AND PURIFICATION, ACADEMIC PRESS, SAN DIEGO, CA, vol. 15, n° 3, 1 de abril 1999 (1-04-1999), páginas 336-343, se emplean dos inductores diferentes para expresar un fragmento de anticuerpo de cadena sencilla y la chaperona Skp de *E. coli*. Dicha Skp estaba mutada y presente en un plásmido bajo el control de un promotor que requería un inductor diferente (arabinosa) que el del fragmento de anticuerpo (IPTG).

#### Compendio de la invención

La presente invención proporciona sistemas de coexpresión inducible, capaces de una inducción controlada de cada componente del producto génico.

En un primer aspecto, la invención proporciona una célula hospedadora que comprende dos o más tipos de estructuras artificiales de expresión, en donde la estructura artificial de expresión de cada tipo comprende un promotor inducible y una secuencia de polinucleótidos que codifica un producto génico, que se va a transcribir a partir del promotor inducible; y al menos uno de dichos promotores inducibles es sensible a un inductor que es diferente del inductor de otro de dichos promotores inducibles; y en donde cada inductor inducible no es un promotor inducible con lactosa; y en donde la célula hospedadora tiene un nivel reducido de función génica de al menos un gen que codifica una proteína que metaboliza un inductor de al menos uno de dichos promotores inducibles; y al menos uno de dichos productos génicos se selecciona a partir del grupo que consiste en un producto génico que forma un multímero con otro de dichos productos génicos, y un polipéptido que forma al menos dos enlaces disulfuro.

En una realización de la invención, al menos uno de dichos productos génicos se selecciona a partir del grupo que consiste en: (a) un polipéptido que carece de un péptido señal y que forma al menos tres enlaces disulfuro; (b) un polipéptido seleccionado a partir del grupo que consiste en enzimas que utilizan arabinosa y xilosa; y (c) un polipéptido seleccionado a partir del grupo que consiste en peroxidases que degradan lignina. En otra realización de la invención, al menos uno de dichos productos génicos es un polipéptido que forma al menos dos y menos de diecisiete enlaces disulfuro, o forma al menos dos y menos de diez enlaces disulfuro, o forma un número de enlaces disulfuro seleccionado a partir del grupo que consiste en dos, tres, cuatro, cinco, seis, siete, ocho y nueve. En algunas realizaciones de la invención, esta célula hospedadora es una célula procarionta, y en algunos casos, es una célula de *E. coli*. En otras realizaciones de la invención, la célula hospedadora es una célula eucariota, y en algunos casos se trata de una célula de levadura, y en algunos casos adicionales, es una célula de *Saccharomyces cerevisiae*. En realizaciones adicionales, las estructuras artificiales de expresión comprendidas en una célula hospedadora, comprenden cada una al menos un promotor inducible, en donde el promotor inducible es un promotor inducible con L-arabinosa o un promotor inducible con propionato, o se selecciona a partir del grupo que consiste en: el promotor *araBAD*, el promotor *prpBCDE*, el promotor *rhaSR* y el promotor *xylA*. En realizaciones adicionales, al menos una estructura artificial de expresión comprendida en una célula hospedadora comprende además una secuencia de polinucleótidos que codifica un regulador transcripcional que se une a un promotor inducible; en algunas realizaciones, la secuencia de polinucleótidos que codifica un regulador transcripcional y el promotor inducible al que se une dicho regulador transcripcional, están en la misma estructura artificial de expresión; y en otros casos, el regulador de transcripcional se selecciona a partir del grupo que consiste en: AraC, PrpR, RhaR y XylR; o en particular es AraC o PrpR. En ciertas realizaciones, al menos una estructura artificial de expresión comprendida en una célula hospedadora, se produce mediante un método que comprende una etapa de inserción de una secuencia de polinucleótidos en un plásmido seleccionado a partir del grupo que consiste en: pBAD18, pBAD18-Cm, pBAD18-Kan, pBAD24, pBAD28, pBAD30, pBAD33, pPRO18, pPRO18-Cm, pPRO18-Kan, pPRO24, pPRO30 y pPRO33; o particularmente en pBAD24 o pPRO33. También se describe en esta memoria una célula hospedadora que comprende dos o más tipos de estructuras artificiales de expresión, en donde la estructura artificial de expresión de cada tipo comprende un promotor inducible y una secuencia de polinucleótidos que codifica un

producto génico, en donde al menos un producto génico es un polipéptido, o se selecciona a partir del grupo que consiste en: (a) una cadena pesada de inmunoglobulina; (b) una cadena ligera de inmunoglobulina; y (c) un fragmento de cualquiera de (a) - (b), o es una cadena ligera de inmunoglobulina, o es una cadena pesada de inmunoglobulina; o es una cadena pesada o ligera de infliximab o un fragmento del mismo, o tiene una identidad de secuencia de aminoácidos de al menos 80% o 90% con SEQ ID NO: 30 o SEQ ID NO: 31 a lo largo de al menos 50% o 80% de la longitud de SEQ ID NO: 30 o SEQ ID NO: 31, respectivamente, o tiene la secuencia de aminoácidos de SEQ ID NO: 30 o SEQ ID NO: 31.

En realizaciones adicionales de la invención, al menos un producto génico es un polipéptido que carece de un péptido señal y que forma al menos tres enlaces disulfuro, o al menos tres y menos de diecisiete enlaces disulfuro, o al menos dieciocho y menos de un centenar de enlaces disulfuro, o al menos tres y menos de diez enlaces disulfuro, o al menos tres y menos de ocho enlaces disulfuro; o forma una cantidad de enlaces disulfuro seleccionada a partir del grupo que consiste en tres, cuatro, cinco, seis, siete, ocho y nueve; o es un polipéptido seleccionado a partir del grupo que consiste en: (a) una cadena pesada de inmunoglobulina; (b) una cadena ligera de inmunoglobulina; (c) manganeso peroxidasa; y (d) un fragmento de cualquiera de (a) - (c); o es una cadena pesada o ligera de infliximab o un fragmento del mismo, o tiene al menos 80% o 90% de identidad de secuencia de aminoácidos con SEQ ID NO: 30 o SEQ ID NO: 31 a lo largo de al menos 50% o 80% de la longitud de SEQ ID NO: 30 o SEQ ID NO: 31, respectivamente, o tiene la secuencia de aminoácidos de SEQ ID NO: 30 o SEQ ID NO: 31; o tiene al menos 80% o 90% de identidad de secuencia de aminoácidos con SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23 a lo largo de al menos 50% u 80% de la longitud de SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23, respectivamente, o tiene la secuencia de aminoácidos de SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23; o es un polipéptido seleccionado a partir del grupo que consiste en enzimas que utilizan arabinosa y xilosa - tal como xilosa isomerasa; o es un polipéptido seleccionado a partir del grupo que consiste en peroxidases que degradan lignina, tales como manganeso peroxidasa o peroxidasa versátil.

En realizaciones adicionales de la invención, se proporciona una célula hospedadora que comprende dos tipos de estructuras artificiales de expresión, y en ciertos casos, un tipo de estructura artificial de expresión se produce mediante un método que comprende una etapa de inserción de una secuencia de polinucleótidos en una secuencia de polinucleótidos de pBAD24, y el otro tipo de estructura artificial de expresión se produce mediante un método que comprende una etapa de inserción de una secuencia de polinucleótidos en una secuencia de polinucleótidos de pPRO33.

En algunas realizaciones, la célula hospedadora tiene una alteración de la función génica de al menos un gen que codifica una proteína transportadora para un inductor de al menos uno de dichos promotores inducibles, y como otro ejemplo, en donde el gen que codifica la proteína transportadora se selecciona a partir del grupo que consiste en *araE*, *araF*, *araG*, *araH*, *rhaT*, *xylF*, *xylG* y *xylH*, o particularmente es *araE*. En algunas realizaciones, el gen que codifica una proteína que metaboliza un inductor de al menos uno de dichos promotores inducibles, se selecciona a partir del grupo que consiste en *araA*, *araB*, *araD*, *prpB*, *prpD*, *rhaA*, *rhaB*, *rhaD*, *xylA* y *xylB*. En algunas realizaciones, la célula hospedadora tiene un nivel reducido de función génica de al menos un gen que codifica una proteína implicada en la biosíntesis de un inductor de al menos uno de dichos promotores inducibles, que en realizaciones adicionales se selecciona a partir del grupo que consiste en *scpA/sbm*, *argK/ygfD*, *scpB/ygfG*, *scpC/ygfH*, *rmlA*, *rmlB*, *rmlC* y *rmlD*.

En algunas realizaciones, la célula hospedadora tiene una función génica alterada de un gen que afecta al entorno de reducción/oxidación del citoplasma de la célula hospedadora, el cual en algunos ejemplos se selecciona a partir del grupo que consiste en *gor* y *gshB*; o en donde la célula hospedadora tiene un nivel reducido de función génica de un gen que codifica una reductasa, que en algunas realizaciones es *trxB*; o en donde la célula hospedadora comprende al menos una estructura artificial de expresión que codifica al menos una proteína isomerasa de enlace disulfuro, que en algunas realizaciones es DsbC; o en donde la célula hospedadora comprende al menos un polinucleótido que codifica una forma de DsbC que carece de un péptido señal; o en donde la célula hospedadora comprende al menos un polinucleótido que codifica Erv1p.

En otras realizaciones, la estructura artificial de expresión de cada tipo comprende un promotor inducible, que no es un promotor inducible de la estructura artificial de expresión de cada otro tipo, o en donde la estructura artificial de expresión de cada tipo comprende un origen de replicación que es diferente del origen de replicación de la estructura artificial de expresión de cada otro tipo.

También se describe en esta memoria una célula hospedadora de *E. coli*, que comprende dos tipos de estructuras artificiales de expresión, en donde un tipo de estructura artificial de expresión se produce por un método que comprende una etapa de inserción de una secuencia de polinucleótidos en una secuencia de polinucleótidos de pBAD24, y el otro tipo de estructura artificial de expresión se produce por un método que comprende una etapa de inserción de una secuencia de polinucleótidos en una secuencia de polinucleótidos de pPRO33; y en donde la célula hospedadora comprende además dos o más de los siguientes: (a) una delección de los genes *araBAD*; (b) una función génica alterada de los genes *araE* y *araFGH*; (c) un gen *lacY(A117C)*; (d) una función génica reducida de los genes *prpB* y *prpD*; (e) una función génica reducida de los genes *sbm/scpA-ygfD/argK-ygfGH/scpBC*, sin alterar la expresión del gen *ygfI*; (f) una función génica reducida de los genes *gor* y *trxB*; (g) una función génica reducida del gen *AscG*; (h) un polinucleótido que codifica una forma de DsbC que carece de péptido señal; e (i) un polinucleótido

que codifica Erv1p, ChuA o una chaperona; y en ciertos ejemplos, la célula hospedadora comprende además al menos una estructura artificial de expresión que comprende una secuencia de polinucleótidos que codifica un producto génico, dicha secuencia de polinucleótidos que se transcribe a partir de un promotor inducible, y en algunos casos, el producto génico se selecciona a partir del grupo que consiste en: (a) una cadena pesada de inmunoglobulina; (b) una cadena ligera de inmunoglobulina; (c) manganeso peroxidasa; y (d) un fragmento de cualquiera de (a) - (c); o es una cadena pesada o ligera de infliximab o un fragmento del mismo, o tiene al menos 80% o 90% de identidad de secuencia de aminoácidos con SEQ ID NO: 30 o SEQ ID NO: 31 a lo largo de al menos 50% u 80% de la longitud de SEQ ID NO: 30 o SEQ ID NO: 31, respectivamente, o tiene la secuencia de aminoácidos de SEQ ID NO: 30 o SEQ ID NO: 31; o tiene al menos 80% o 90% de identidad de secuencia de aminoácidos con SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23 a lo largo de al menos 50% o 80% de la longitud de SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23, respectivamente, o tiene la secuencia de aminoácidos de SEQ ID NO: 13, SEQ ID NO: 15 o SEQ ID NO: 23; o es un polipéptido seleccionado a partir del grupo que consiste en enzimas que utilizan arabinosa y xilosa tales como la xilosa isomerasa; o es un polipéptido seleccionado a partir del grupo que consiste en peroxidasa que degradan lignina, tales como manganeso peroxidasa o peroxidasa versátil.

En un segundo aspecto, la invención proporciona un método para producir un producto génico que comprende hacer crecer un cultivo de una célula hospedadora de la invención como se ha descrito anteriormente; y añadir al cultivo un inductor de al menos un promotor inducible. Un producto génico o un producto múltimero producido por este método también se describe en esta memoria, y en algunos casos es un anticuerpo, o en casos más particulares, es un anticuerpo aglicosilado, un anticuerpo quimérico o un anticuerpo humano.

En un tercer aspecto, la invención proporciona un kit que comprende una célula hospedadora de la invención. También se describen en esta memoria, kits que comprenden un producto génico o un producto múltimero producido mediante el cultivo de una célula hospedadora de la invención y la adición de al menos un inductor al cultivo, en donde en algunos casos, el producto múltimero es un anticuerpo, o en casos más particulares, es un anticuerpo aglicosilado, un anticuerpo quimérico o un anticuerpo humano.

Breve descripción de los dibujos

La Fig. 1 es una ilustración esquemática del sistema de coexpresión inducible, que incluye una célula hospedadora (1) que comprende dos vectores de expresión inducible diferentes (3) y (4), que expresan diferentes productos génicos después de la aplicación de inductores (5), formando un producto múltimero (6).

La Fig. 2 es una ilustración esquemática de un uso particular del sistema de coexpresión inducible, en el que el genoma de la célula hospedadora *E. coli* (2) codifica una forma citoplásmica de la disulfuro isomerasa DsbC que carece de un péptido señal; el vector de expresión pBAD24 (3) proporciona una expresión inducible con L-arabinosa de una cadena pesada de inmunoglobulina, y el vector de expresión pPRO33 (4) proporciona una expresión inducible con propionato de una cadena ligera de inmunoglobulina; formando después de la inducción (5) el producto de anticuerpo múltimero (6).

La Fig. 3 muestra el resultado de la coexpresión de las cadenas pesadas y ligeras de inmunoglobulina en células bacterianas. Las células SHuffle<sup>®</sup> Express y BL21 que contienen tanto los vectores de expresión inducible pBAD24-HC como pPRO33-LC, fueron inducidos mediante crecimiento en L-arabinosa y propionato. Extractos de proteínas solubles procedentes de células inducidas y controles no inducidos se separaron por electroforesis en gel de SDS en condiciones reductoras, sobre un gel de Bis-Tris al 4-12%. Carril 1: SHuffle<sup>®</sup> Express inducidas. Carril 2: SHuffle<sup>®</sup> Express no inducidas. Carril 3: BL21 inducidas. Carril 4: BL21 no inducidas. Las flechas indican una banda de proteína (la cadena pesada de IgG1) en 51 kDa y otra banda de proteína (cadena ligera de IgG1) en 26 kDa; estas bandas están presentes en las células inducidas, pero no en las células no inducidas.

La Fig. 4 muestra el resultado de la coexpresión de cadenas pesadas y ligeras de inmunoglobulina en células bacterianas. Los mismos extractos de proteínas solubles procedentes de células SHuffle<sup>®</sup> Express y BL21 inducidas y no inducidas que contenían tanto los vectores de expresión inducible pBAD24-HC como pPRO33-LC, tal como se ha descrito para la Figura 3, se separaron por electroforesis en gel en condiciones naturales (no reductoras) sobre un gel de Tris-glicina al 10-20%. Carril 1: SHuffle<sup>®</sup> Express inducidas. Carril 2: SHuffle<sup>®</sup> Express no inducidas. Carril 3: BL21 inducidas. Carril 4: BL21 no inducidas. La flecha indica una banda de proteína (anticuerpo IgG1 que comprende cadenas pesadas y ligeras) en 154 kDa; esta banda está presente en las células SHuffle<sup>®</sup> Express inducidas, pero se reduce significativamente o está ausente en las células BL21 inducidas y en las células no inducidas.

La Fig. 5 muestra el resultado de la coexpresión, en células bacterianas, de manganeso peroxidasa (MnP) y proteína disulfuro isomerasa (PDI) en presencia de grupos hemo. Las células SHuffle<sup>®</sup> Express que contenían tanto el vector de expresión inducible pPRO33-MnP-ChuA como pBAD24-PDI, fueron inducidas mediante el crecimiento en L-arabinosa y propionato. Extractos de proteínas solubles procedentes de células inducidas y no inducidas fueron separados por electroforesis en gel en condiciones reductoras en un gel de Bis-glicina al 10%.

Marcadores: Patrón de Bio-Rad Precision Plus Protein<sup>®</sup> (teñido previamente)

Carril 1: No se induce (sin hemo, sin propionato, sin arabinosa)

	Carril 2:	propionato 50 mM	0,002% de arabinosa
	Carril 3:	propionato 25 mM	0,002% de arabinosa
	Carril 4:	propionato 12,5 mM	0,002% de arabinosa
	Carril 5:	propionato 50 mM	0,01% de arabinosa
5	Carril 6:	propionato 25 mM	0,01% de arabinosa
	Carril 7:	propionato 12,5 mM	0,01% de arabinosa
	Carril 8:	propionato 50 mM	0,05% de arabinosa
	Carril 9:	propionato 25 mM	0,05% de arabinosa
	Carril 10:	propionato 12,5 mM	0,05% de arabinosa

10 Las flechas indican las bandas de proteínas, MnP en 39 kDa y PDI en 53 kDa; estas bandas están presentes en las células SHuffle<sup>®</sup> Express más fuertemente bajo algunas de las condiciones de inducción, pero se reducen de manera significativa en las células no inducidas.

15 La Fig. 6 muestra el resultado de la coexpresión, en células bacterianas, de una forma madura alternativa de manganeso peroxidasa (MnP\_FT) y proteína disulfuro isomerasa (PDI) en presencia de grupos hemo. Las células SHuffle<sup>®</sup> Express que contenían los vectores de expresión inducible pBAD24-MnP\_FT-ChuA y pPRO33-PDI, fueron inducidos mediante el crecimiento en 0,1% de L-arabinosa y propionato 50 mM. Extractos de proteínas solubles procedentes de células no inducidas e inducidas fueron separados por electroforesis en gel en condiciones reductoras sobre un gel de Bis-glicina al 10%.

	Carril 1:	Marcadores del peso molecular
20		Patrón de Bio-Rad Precision Plus Protein <sup>®</sup> (teñido previamente)
	Carril 2:	Coexpresión inducida (0,1% de L-arabinosa, propionato 50 mM)
	Carril 3:	Sin inducir (sin hemina, sin L-arabinosa, sin propionato)
	Carril 4:	Control inducido (no hay insertos que codifican proteínas)
	Carril 5:	Control no inducido (no hay insertos que codifican proteínas)

25 Descripción detallada de la invención

El problema de componentes de sistemas de coexpresión incompatibles se aborda con el desarrollo de sistemas de coexpresión bacteriana coordinados que utilizan sistemas de promotores compatibles, inducibles de forma homogénea, situados en estructuras artificiales de expresión distintas y, en algunas realizaciones, activados por diferentes inductores. Las ventajas de la presente invención incluyen, sin limitación: 1) mejora de la compatibilidad de los componentes en el sistema de coexpresión inducible; 2) expresión inducible de productos génicos que juntos forman multímeros, u otras combinaciones de productos génicos (coexpresión de dos o más productos génicos); 3) mejora del control de la coexpresión de productos génicos mediante una inducción que se puede valorar independientemente; 4) mejora de la expresión de complejos de productos génicos y otros productos que son difíciles de expresar, tales como productos multímeros y productos que forman enlaces disulfuro; 5) optimización simplificada de la coexpresión de productos génicos.

35 Productos génicos coexpresados. Los sistemas de coexpresión inducible de la invención están diseñados para coexpresar dos o más productos génicos diferentes que contribuyen a un producto deseado. El producto deseado puede ser un multímero, formado a partir de productos génicos coexpresados, o la coexpresión se puede utilizar para producir una combinación del producto deseado además de un producto o productos adicionales que ayudan en la expresión del producto deseado.

40 Un "producto multímero" se refiere a un conjunto de productos génicos que se ensamblan entre sí para llevar a cabo la función del producto multímero, y no se refiere a asociaciones transitorias entre productos génicos y otras moléculas, tales como enzimas modificadoras (cinasas, peptidasas y similares), chaperonas, transportadores, etc. En ciertas realizaciones de la invención, los productos multímeros son heteromultímeros. En muchas realizaciones, 45 los productos génicos coexpresados serán polipéptidos que son subunidades de proteínas multímeras. Sin embargo, también es posible utilizar los sistemas de coexpresión inducible de la invención para coexpresar múltiples moléculas de ARN no codificante diferentes, o una combinación de polipéptido y productos génicos de ARN no codificante. Las moléculas de ARN no codificante, también denominadas ARN que no codifica proteínas (npcARN), ARN no mensajero (nmARN) y ARN funcional (fARN), incluyen muchos tipos diferentes de moléculas de ARN tales

como microARNs que no son ARNs mensajeros y que, por lo tanto, no son moldes para la formación de polipéptidos a través de la traducción.

Muchos productos biológicamente importantes se forman a partir de combinaciones de diferentes cadenas de polipéptidos. Además de los anticuerpos y los fragmentos de anticuerpos, otros productos multímeros que se pueden producir por los métodos de coexpresión inducible de la invención, incluyen receptores acoplados a proteínas G y canales iónicos regulados por ligandos, tales como los receptores nicotínicos de acetilcolina, los receptores de GABA<sub>A</sub>, los receptores de glicina, los receptores de serotonina, los receptores de glutamato y los receptores regulados por ATP, tales como los receptores de P2X. La neurotoxina botulínica (a menudo referida como BoTN, BTX o como una de sus formas disponibles en el mercado, BOTOX<sup>®</sup> (toxina botulínica tipo A)) se forma a partir de una cadena pesada y una cadena ligera, unidas por un enlace disulfuro (Simpson *et al.*, "The role of the interchain disulfide bond in governing the pharmacological actions of botulinum toxin", *J Pharmacol Exp Ther* 2004 Mar; 308(3): 857-864, Epub 14 Nov 2003). Otro ejemplo de un producto formado a partir de diferentes cadenas de polipéptidos es la insulina, la cual en eucariotas se traduce primero como una sola cadena polipeptídica, se pliega y luego se escinde en última instancia en dos cadenas polipeptídicas unidas por enlaces disulfuro. Una producción eficaz de la neurotoxina botulínica o de insulina madura en una única célula hospedadora, son ejemplos de usos de los métodos de coexpresión inducible de la invención.

Los métodos de la invención están diseñados para producir productos génicos que han sido plegados y/o ensamblados correctamente en forma de productos funcionales, y que tienen un número deseado de enlaces disulfuro en los lugares deseados dentro de tales productos funcionales (que se pueden determinar por métodos tales como el del Ejemplo 11). El número de enlaces disulfuro para un producto génico tal como un polipéptido, es el número total de enlaces intramoleculares e intermoleculares formados por ese producto génico, cuando está presente en un producto funcional deseado. Por ejemplo, una cadena ligera de un anticuerpo de IgG humana, normalmente tiene tres enlaces disulfuro (dos enlaces intramoleculares y un enlace intermolecular), y una cadena pesada de un anticuerpo de IgG humana, normalmente tiene siete enlaces disulfuro (cuatro enlaces intramoleculares y tres enlaces intermoleculares). En algunas realizaciones, los productos génicos deseados se coexpresan con otros productos génicos, tales como chaperonas, que son beneficiosas para la producción del producto génico deseado. Las chaperonas son proteínas que ayudan al plegamiento no covalente o el desplegamiento, y/o el ensamblaje o desensamblaje de otros productos génicos, pero no se producen en las estructuras resultantes de productos génicos monómeros o multímeros, cuando las estructuras están realizando sus funciones biológicas normales (habiendo completado los procesos de plegamiento y/o ensamblaje). Las chaperonas se pueden expresar a partir de un promotor inducible o un promotor constitutivo dentro de una estructura artificial de expresión, o se pueden expresar a partir del cromosoma de la célula hospedadora; preferiblemente, la expresión de proteína(s) chaperona(s) en la célula hospedadora tiene un nivel suficientemente alto para producir productos génicos coexpresados que están correctamente plegados y/o ensamblados en el producto deseado. Ejemplos de chaperonas presentes en células hospedadoras de *E. coli* son los factores de plegamiento DnaK/DnaJ/GrpE, DsbC/DsbG, GroEL/GroES, IbpA/IbpB, Skp, Tig (factor de activación) y FkpA, que han sido utilizadas para evitar la agregación proteica de proteínas citoplasmáticas o periplásmicas. DnaK/DnaJ/GrpE, GroEL/GroES y ClpB pueden actuar sinérgicamente para ayudar en el plegamiento de una proteína y, por lo tanto, la expresión de estas chaperonas en combinaciones ha mostrado ser beneficiosa para la expresión de proteínas (Makino *et al.*, "Strain engineering for improved expression of recombinant proteins in bacteria", *Microb Cell Fact* 2011 May 14; 10: 32). Cuando se expresan proteínas eucariotas en células hospedadoras procariontas, una proteína chaperona eucariota, tal como una proteína disulfuro isomerasa (PDI) de la misma especie o de una especie eucariota relacionado, se coexpresa o se coexpresa de forma inducible con el producto génico deseado en ciertas realizaciones de la invención.

Promotores inducibles. Lo siguiente es una descripción de promotores inducibles que se pueden emplear en estructuras artificiales de expresión para la coexpresión de productos génicos, junto con algunas de las modificaciones genéticas que se pueden realizar en las células hospedadoras que contienen tales estructuras artificiales de expresión. Ejemplos de estos promotores inducibles y genes relacionados proceden, a menos que se especifique lo contrario, de la cepa MG1655 (depósito en la American Type Culture Collection ATCC 700926) de *Escherichia coli* (*E. coli*), que es una subcepa de *E. coli* K-12 (depósito en la American Type Culture Collection ATCC 10798). La Tabla 1 enumera las ubicaciones genómicas, en *E. coli* MG1655, de las secuencias de nucleótidos para estos ejemplos de promotores inducibles y genes relacionados. Los nucleótidos y otras secuencias genéticas, a las que se hace referencia a través de una localización genómica como en la Tabla 1, se incorporan expresamente como referencia en el presente documento. Una información adicional sobre los promotores, los genes y las cepas de *E. coli* descritas en este documento, se puede encontrar en muchas fuentes públicas, incluyendo el medio EcoliWiki en línea, ubicado en [ecoliwiki.net](http://ecoliwiki.net).

Promotor de arabinosa. (Tal como se emplea en este documento, 'arabinosa' significa L-arabinosa). Diversos operones de *E. coli* implicados en la utilización de arabinosa son inducibles con arabinosa - *araBAD*, *araC*, *araE* y *araFGH* - pero las expresiones "promotor de arabinosa" y "promotor ara" se utilizan normalmente para designar el promotor *araBAD*. Se han empleado diversos términos adicionales para indicar el promotor *araBAD* de *E. coli*, tales como P<sub>ara</sub>, P<sub>araB</sub>, P<sub>araBAD</sub> y P<sub>BAD</sub>. El uso en esta memoria de "promotor ara" o cualquiera de los términos alternativos proporcionados anteriormente, significa el promotor *araBAD* de *E. coli*. Como se puede observar a partir de la utilización de otro término, "promotor *araC-araBAD*", el promotor *araBAD* se considera que forma parte de un promotor bidireccional, en donde el promotor *araBAD* controla la expresión del operón *araBAD* en una dirección, y el

promotor *araC*, en estrecha proximidad y en la cadena opuesta al promotor *araBAD*, que controla la expresión de la secuencia que codifica *araC* en la otra dirección. La proteína AraC es un regulador transcripcional tanto positivo como negativo del promotor *araBAD*. En ausencia de arabinosa, la proteína AraC inhibe la transcripción desde  $P_{BAD}$ , pero en presencia de arabinosa, la proteína AraC, que altera su conformación después de la unión a arabinosa, se convierte en un elemento regulador positivo que permite la transcripción desde  $P_{BAD}$ . El operón *araBAD* codifica proteínas que metabolizan la L-arabinosa convirtiéndola, a través de los compuestos intermedios L-ribulosa y L-ribulosa-fosfato, en D-xilulosa-5-fosfato. Con el fin de maximizar la inducción de la expresión desde un promotor inducible con arabinosa, es útil eliminar o reducir la función de AraA, que cataliza la conversión de L-arabinosa a L-ribulosa y, opcionalmente, eliminar o reducir también la función de por lo menos una de AraB y AraD. Eliminar o reducir la capacidad de las células hospedadoras para disminuir la concentración eficaz de arabinosa en la célula, mediante la eliminación o la reducción de la capacidad de la célula para convertir la arabinosa en otros azúcares, permite que esté disponible más arabinosa para la inducción del promotor inducible con arabinosa. Los genes que codifican los transportadores que llevan la arabinosa dentro de la célula hospedadora, son *araE*, que codifica el simportador de protones y L-arabinosa de baja afinidad, y el operón *araFGH*, que codifica las subunidades de un transportador de L-arabinosa de la superfamilia ABC de alta afinidad. Otras proteínas que pueden transportar L-arabinosa dentro de la célula son ciertos mutantes de la lactosa permeasa LacY: las proteínas LacY(A177C) y LacY(A177V), que tienen un aminoácido cisteína o valina en lugar de alanina en la posición 177, respectivamente (Morgan-Kiss *et al.*, "Long-term and homogeneous regulation of the *Escherichia coli* *araBAD* promoter by use of a lactose transporter of relaxed specificity", Proc Natl Acad Sci U S A 2002 May 28; 99(11): 7373-7377). Con el fin de conseguir una inducción homogénea de un promotor inducible con arabinosa, es útil realizar el transporte de arabinosa en la célula, independientemente de la regulación con arabinosa. Esto se puede lograr mediante la eliminación o la reducción de la actividad de las proteínas transportadoras AraFGH y la alteración de la expresión de *araE*, de modo que solo se transcribe a partir de un promotor constitutivo. La expresión constitutiva de *araE* se puede lograr mediante la eliminación o la reducción de la función del gen *araE* natural, y la introducción en la célula de una estructura artificial de expresión que incluye una secuencia que codifica la proteína AraE, expresada a partir de un promotor constitutivo. Alternativamente, en una célula que carece de la función AraFGH, el promotor que controla la expresión del gen *araE* cromosómico de la célula hospedadora, se puede cambiar desde un promotor inducible con arabinosa a un promotor constitutivo. De forma similar, como alternativas adicionales para la inducción homogénea de un promotor inducible con arabinosa, una célula hospedadora que carece de la función AraE puede tener cualquier secuencia funcional que codifica AraFGH presente en la célula expresada a partir de un promotor constitutivo. Como otra alternativa, es posible expresar tanto el gen *araE* como el operón *araFGH* a partir de promotores constitutivos, mediante la sustitución de los promotores *araE* y *araFGH* naturales por promotores constitutivos en el cromosoma del hospedador. También es posible eliminar o reducir la actividad de los transportadores de arabinosa AraE y AraFGH, y en esa situación utilizar una mutación en la lactosa permeasa LacY que permite que esta proteína transporte arabinosa. Dado que la expresión del gen *lacY* no está regulada normalmente por arabinosa, el uso de un mutante de LacY, tal como LacY(A117C) o LacY(A177V), no dará lugar al fenómeno de inducción de 'todo o nada' cuando el promotor inducible con arabinosa se induce con la presencia de arabinosa. Debido a que la proteína LacY(A117C) parece ser más eficaz para el transporte de arabinosa dentro de la célula, el uso de polinucleótidos que codifican la proteína LacY(A117C) se prefiere el uso de polinucleótidos que codifican la proteína LacY(A177V).

Promotor de propionato. El "promotor de propionato" o "promotor *prp*" es el promotor para el operón *prpBCDE* de *E. coli*, y también se denomina  $P_{prpB}$ . Del mismo modo que el promotor *ara*, el promotor *prp* forma parte de un promotor bidireccional, que controla la expresión del operón *prpBCDE* en una dirección, y con el promotor *prpR* controla la expresión de la secuencia que codifica *prpR* en la otra dirección. La proteína PrpR es el regulador transcripcional del promotor *prp*, y activa la transcripción desde el promotor *prp* cuando la proteína PrpR se une a 2-metilcitrato ("2-MC"). El propionato (también denominado propanoato) es el ion,  $CH_3CH_2COO^-$ , de ácido propiónico (o 'ácido propanoico'), y es el más pequeño de los ácidos 'grasos' que tienen la fórmula general  $H(CH_2)_nCOOH$ , que comparte ciertas propiedades de esta clase de moléculas: producir una capa aceitosa cuando se produce una precipitación salina con agua y tener una sal de potasio jabonosa. El propionato disponible comercialmente se vende generalmente como una sal de catión monovalente de ácido propiónico, tal como propionato de sodio ( $CH_3CH_2COONa$ ), o como una sal de catión divalente, tal como propionato de calcio ( $Ca(CH_3CH_2COO)_2$ ). El propionato es permeable a membranas y se metaboliza a 2-MC por la conversión de propionato a propionil-CoA a través de PrpE (propionil-CoA sintetasa), y luego la conversión de propionil-CoA a 2-MC a través de PrpC (2-metilcitrato sintasa). Las otras proteínas codificadas por el operón *prpBCDE*, PrpD (2-metilcitrato deshidratasa) y PrpB (2-metilcitrato liasa), están involucradas en un catabolismo adicional de 2-MC en productos más pequeños, como piruvato y succinato. Con el fin de maximizar la inducción de un promotor inducible con propionato mediante propionato añadido al medio de crecimiento celular, es deseable, por lo tanto, disponer de una célula hospedadora con actividad PrpC y PrpE, para convertir el propionato en 2-MC, pero también que no tenga actividad PrpD o que esté reducida y, opcionalmente, también una actividad PrpB eliminada o reducida, para evitar que 2-MC sea metabolizado. Otro operón que codifica proteínas que participan en la biosíntesis de 2-MC es el operón *scpA-argK-scpBC* también denominado el operón *sbm-ygfDGH*. Estos genes codifican proteínas requeridas para la conversión de succinato en propionil-CoA, que luego se puede convertir en 2-MC a través de PrpC. La eliminación o la reducción de la función de estas proteínas podrían eliminar una ruta paralela para la producción del inductor 2-MC y, por lo tanto, podría reducir los niveles de fondo de expresión de un promotor inducible con propionato, y aumentar la sensibilidad del promotor inducible con propionato para suministrar propionato de forma exógena. Se ha encontrado

que una delección de *sbm-ygfD-ygfG-ygfH-ygfI*, introducida en *E. coli* BL21(DE3) para crear la cepa JSB ("Lee y Keasling, "A propionate-inducible expression system for enteric bacteria", Appl Environ Microbiol 2005 Nov; 71(11): 6856-6862), era útil en la reducción de la expresión de fondo en ausencia de inductor suministrado de forma exógena, pero esta delección también reducía la expresión general del promotor *prp* en la cepa JSB. Cabe señalar, sin embargo, que la delección de *sbm-ygfD-ygfG-ygfH-ygfI* también afecta aparentemente a *ygfI*, que codifica un regulador transcripcional putativo de la familia LysR, con función desconocida. Los genes *sbm-ygfDGH* se transcriben como un operón, e *ygfI* se transcribe desde la hebra opuesta. Los extremos 3' de las secuencias que codifican *ygfH* y *ygfI* se solapan en unos pocos pares de bases, por lo que una delección que elimina todo el operón *sbm-ygfDGH*, aparentemente elimina también la función que codifica *ygfI*. La eliminación o la reducción de la función de un subconjunto de los productos génicos de *sbm-ygfDGH*, tal como YgfG (también denominado ScpB, metilmalonil-CoA descarboxilasa), o la delección de la mayoría del operón *sbm-ygfDGH* (o *scpA-argK-scpBC*), aunque deja suficiente del extremo 3' del gen *ygfH* (o *scpC*) de modo que la expresión de *ygfI* no se ve afectada, podría ser suficiente para reducir la expresión de fondo desde un promotor inducible con propionato, sin reducir el nivel máximo de expresión inducida.

Promotor de ramnosa. (Tal como se usa en este documento, 'ramnosa' significa L-ramnosa). El "promotor de ramnosa" o "promotor *rha*" o  $P_{rhaSR}$ , es el promotor del operón *rhaSR* de *E. coli*. Al igual que los promotores *ara* y *prp*, el promotor *rha* forma parte de un promotor bidireccional, que controla la expresión del operón *rhaSR* en una dirección, y con el promotor *rhaBAD* controla la expresión del operón *rhaBAD* en la otra dirección. El promotor *rha*, sin embargo, tiene dos reguladores transcripcionales implicados en la modulación de la expresión: RhaR y RhaS. La proteína RhaR activa la expresión del operón *rhaSR* en presencia de ramnosa, mientras que la proteína RhaS activa la expresión de los operones catabólicos y de transporte de L-ramnosa, *rhaBAD* y *rhaT*, respectivamente (Wickstrum *et al.*, "The AraC/XylS family activator RhaS negatively autoregulates *rhaSR* expression by preventing cyclic AMP receptor protein activation", J Bacteriol Jan 2010; 192(1): 225-232). Aunque la proteína RhaS también puede activar la expresión del operón *rhaSR*, de hecho RhaS autorregula negativamente esta expresión al interferir con la capacidad de la proteína receptora de AMP cíclico (CRP) para coactivar la expresión con RhaR a un nivel mucho mayor. El operón *rhaBAD* codifica las proteínas catabólicas de ramnosa RhaA (L-ramnosa isomerasa), que convierte L-ramnosa en L-ramnulosa; RhaB (ramnulocinasa), que fosforila la L-ramnulosa para formar L-ramnulosa-1-P; y RhaD (ramnulosa-1-fosfato aldolasa), que convierte L-ramnulosa-1-P a L-lactaldehído y DHAP (fosfato de dihidroxiacetona). Para maximizar la cantidad de ramnosa en la célula disponible para una inducción de la expresión desde un promotor inducible con ramnosa, es deseable reducir la cantidad de ramnosa que se descompone por catálisis, eliminando o reduciendo la función de RhaA, u opcionalmente de RhaA y al menos una de RhaB y RhaD. Las células de *E. coli* también pueden sintetizar L-ramnosa a partir de alfa-D-glucosa-1-P a través de las actividades de las proteínas Rm1A, Rm1B, Rm1C y Rm1D (también denominadas RfbA, RfbB, RfbC y RfbD, respectivamente) codificadas por el operón *rmBDACX* (o *rfbBDACX*). Para reducir la expresión de fondo de un promotor inducible con ramnosa, y para mejorar la sensibilidad de la inducción del promotor inducible con ramnosa a través de ramnosa suministrada de forma exógena, puede ser útil eliminar o reducir la función de una o varias de las proteínas Rm1A, Rm1B, Rm1C y Rm1D. L-ramnosa se transporta dentro de la célula a través de RhaT, la permeasa de ramnosa o el simportador de L-ramnosa:protón. Como se ha señalado anteriormente, la expresión de RhaT se activa a través del regulador transcripcional RhaS. Para hacer que la expresión de RhaT sea independiente de la inducción con ramnosa (que induce la expresión de RhaS), la célula hospedadora puede ser alterada de modo que todas las secuencias que codifican RhaT funcional en la célula, se expresen a partir de promotores constitutivos. Además, las secuencias que codifican RhaS se pueden delecionar o inactivar, de modo que no se produce RhaS funcional. Mediante la eliminación o la reducción de la función de RhaS en la célula, el nivel de expresión a partir del promotor *rhaSR* se incrementa debido a la ausencia de una autorregulación negativa a través de RhaS, y el nivel de expresión del operón catalítico de ramnosa *rhaBAD* se reduce, aumentando aún más la capacidad de la ramnosa para inducir la expresión a partir del promotor *rha*.

Promotor de xilosa. (Tal como se usa en este documento, 'xilosa' significa D-xilosa). El promotor de xilosa, o "promotor *xyl*" o  $P_{xylA}$ , significa el promotor para el operón *xylAB* de *E. coli*. La región promotora de xilosa es similar en la organización a otros promotores inducibles, en donde el operón *xylAB* y el operón *xylFGHR* se expresan ambos a partir de promotores adyacentes inducibles con xilosa, en direcciones opuestas sobre el cromosoma de *E. coli* (Song y Park, "Organization and regulation of the D-xylose operons in *Escherichia coli* K-12: XylR acts as a transcriptional activator", J Bacteriol. Nov 1997; 179(22): 7025-7032). El regulador transcripcional tanto del promotor  $P_{xylA}$  como  $P_{xylF}$  es XylR, que activa la expresión de estos promotores en presencia de xilosa. El gen *xylR* se expresa ya sea como parte del operón *xylFGHR* o a partir de su propio promotor débil, que no es inducible con xilosa, situado entre las secuencias que codifican proteínas *xylH* y *xylR*. La D-xilosa se cataboliza con XylA (D-xilosa isomerasa), que convierte D-xilosa en D-xilulosa, que luego se fosforila con XylB (xilulocinasa) para formar D-xilulosa-5-P. Para maximizar la cantidad de xilosa en la célula, disponible para una inducción de la expresión a partir de un promotor inducible con xilosa, es deseable reducir la cantidad de xilosa que se descompone por catálisis, mediante la eliminación o la reducción de la función de al menos XylA, u opcionalmente tanto de XylA como de XylB. El operón *xylFGHR* codifica XylF, XylG y XylH, las subunidades de un transportador de alta afinidad de D-xilosa de la superfamilia ABC. El gen *xylE*, que codifica el simportador de xilosa-protón de baja afinidad de *E. coli*, representa un operón distinto, cuya expresión también es inducible con xilosa. Para hacer que la expresión de un transportador de xilosa sea independiente de la inducción con xilosa, la célula hospedadora puede ser alterada de modo que todos los transportadores de xilosa se expresen a partir de promotores constitutivos. Por ejemplo, el operón

5 *xyIFGHR* se podría modificar de manera que se delecionen las secuencias que codifican *xyIFGH*, dejando a XylR como la única proteína activa expresada a partir del promotor inducible con xilosa P<sub>xyIF</sub>, y en donde la secuencia que codifica *xyIE* se expresa a partir de un promotor constitutivo, en lugar de su promotor natural. Como otro ejemplo, la secuencia que codifica *xyIR* se expresa a partir del promotor P<sub>xyIA</sub> o el promotor P<sub>xyIF</sub> en una estructura artificial de expresión, mientras que el operón *xyIFGHR* se deleciona y *xyIE* se expresa constitutivamente o, alternativamente, un operón *xyIFGH* (que carece de la secuencia que codifica *xyIR* ya que está presente en una estructura artificial de expresión) se expresa a partir de un promotor constitutivo y la secuencia que codifica *xyIE* se deleciona o se altera de manera que no produce una proteína activa.

10 Promotor de lactosa. La expresión "promotor de lactosa" se refiere al promotor inducible con lactosa para el operón *lacZYA*, un promotor que también se denomina lacZp1; este promotor de lactosa se encuentra en aprox. 365603 - 365568 (cadena negativa, con el sitio de unión a la polimerasa de ARN (" -35 ") sitio a aprox. 365603 - 365598, la caja de Pribnow ('-10') a 365579 - 365573, y un sitio de iniciación de la transcripción a 365567) en el secuencia genómica de la subcepa MG1655 de *E. coli* K-12 (Secuencia de referencia en NCBI NC\_000913.2, 11I-ENE-2012).

Tabla 1. Ubicaciones genómicas de promotores y genes relacionados inducibles de *E. coli* [1]

Promotor o Gen	Ubicación genómica:	Comentarios:
promotor <i>araBAD</i>	[2] (aprox. 70165) - 70074 (cadena negativa)	Smith y Schleif [3]: unión a ARN pol [4] ('-5') 70110-70104, caja de Pribnow ('-10') 70092-70085
operón <i>araBAD</i>	70075 - 65855 (cadena negativa)	Smith y Schleif [3]: inicio transcrito 70075, <i>araB</i> ATG 70048; NCBI: <i>araB</i> final de TAA 68348; <i>araA</i> ATG 68337, final de TAA 66835; <i>araD</i> ATG 66550, final de TAA 65855
promotor <i>araC</i>	[2] (aprox. 70166) - 70241 (cadena positiva)	Smith y Schleif [3]: unión a ARN pol ('-35') 70210-7021, caja de Pribnow ('-10') 70230-70236
gen <i>araC</i>	70242 - 71265 (cadena positiva)	Miyada [5]: inicio transcrito 70242, <i>araC</i> ATG 70387; NCBI: final de TAA 71265
promotor <i>araE</i>	[2] (aprox. 2980349) - 2980231 (cadena negativa)	Stoner y Schleif [6]: unión a CRP 2980349-2980312, unión a ARN pol ('-35') 2980269-2980264, caja de Pribnow ('-10') 2980244-2980239
gen <i>araE</i>	2980230-2978786 (cadena negativa)	Stoner y Schleif [6]: inicio transcrito 2980230, ATG 2980204; NCBI: final de TGA 2978786
promotor <i>araFGH</i>	[2] (aprox. 1984423) - 1984264 (cadena negativa)	Hendrickson [7]: unión a AraC aprox. 1984423-aprox. 1984414 y 1984326-1984317, unión a CRP 1984315-1984297, unión a ARN pol ('-35') 1984294-1984289, caja de Pribnow ('-10') 1984275-1984270

ES 2 621 320 T3

Promotor o Gen	Ubicación genómica:	Comentarios:
operón <i>araFGH</i>	1984263 - 1980578 (cadena negativa)	Hendrickson [7]: inicio transcrito 1984263; NCBI: <i>araF</i> ATG 1984152, final de TAA 1983163; <i>araG</i> ATG 1983093, final de TGA 1981579; <i>araH</i> ATG 1981564, final de TGA 1980578
gen <i>lacY</i>	362403 - 361150 (cadena negativa)	Expresado como parte del operón <i>lacZYA</i> . NCBI: ATG 362403, final de TAA 361150
promotor <i>prpBCDE</i>	[2] aprox. 347790 - aprox. 347870 (cadena positiva)	Keasling [8]: unión a ARN pol ('-24') 347844-347848, caja de Pribnow ('-12') 347855-347859
operón <i>prpBCDE</i>	(aprox. 347871) - 353816 (cadena positiva)	Keasling [8]: inicio transcrito deducido aprox. 347871, <i>prpB</i> ATG 347906; NCBI: <i>prpB</i> final de TAA 348796; <i>prpC</i> ATG 349236, final de TAA 350405; <i>prpD</i> ATG 350439, final de TAA 351890; <i>prpE</i> ATG 351930, final de TAG 353816
promotor <i>prpR</i>	[2] aprox. 347789 - aprox. 347693 (cadena negativa)	Keasling [8]: unión a CRP 347775-347753, unión a ARN pol ('-35') 347728-347723, caja de Pribnow ('-10') 347707-347702
gen <i>prpR</i>	(aprox. 347692) - 346081 (cadena negativa)	Keasling [8]: inicio transcrito deducido aprox. 347692, <i>prpR</i> ATG 347667; NCBI: final de TGA 346081
operón <i>scpA-argK-scpBC</i> (o <i>sbm-ygfDGH</i> )	3058872 - 3064302 (cadena positiva)	NCBI: <i>scpA</i> ATG 3058872, final de TAA 3061016; <i>argK</i> ATG 3061009, final de TAA 3062004; <i>scpB</i> ATG 3062015, final de TAA 3062800; <i>scpC</i> ATG 3062824, final de TAA 3064302
promotor <i>rhaBAD</i>	[2] (aprox. 4095605) - 4095496 (cadena negativa)	Wickstrum [9]: unión a CRP 4095595-4095580, unión a ARN pol ('-35') 4095530-4095525, caja de Pribnow ('-10') 4095506-4095501
operón <i>rhaBAD</i>	4095495 - 4091471 (cadena negativa)	Wickstrum [9]: inicio transcrito 4095495, <i>rhaB</i> ATG 4095471; NCBI: <i>rhaB</i> final de TGA 4094002; <i>rhaA</i> ATG 4094005, final de TAA 4092746; <i>rhaD</i> ATG 4092295, final de TAA 4091471
promotor <i>rhaSR</i>	[2] (aprox. 4095606) - 4095733 (cadena positiva)	Wickstrum [9]: unión a CRP 4095615-4095630, unión a ARN pol ('-35') 4095699-4095704, caja de Pribnow ('-10') 4095722-4095727

ES 2 621 320 T3

Promotor o Gen	Ubicación genómica:	Comentarios:
operón <i>rhaSR</i>	4095734 - 4097517 (cadena positiva)	Wickstrum [9]: inicio transcrito 4095734, <i>rhaS</i> ATG 4095759; NCBI: <i>rhaS</i> final de TAA 4096595; <i>rhaR</i> ATG 4096669, final de TAA 4097517
operón <i>rfbBDACX</i> (o <i>rmlBDACX</i> )	2111085 - 2106361 (cadena negativa)	NCBI: <i>rfbB</i> GTG 2111085, final de TAA 2110000; <i>rfbD</i> ATG 2110000, final de TAA 2109101; <i>rfbA</i> ATG 2109043, final de TAA 2108162; <i>rfbC</i> ATG 2108162, final de TGA 2107605; <i>rfbX</i> ATG 2107608, final de TGA 2106361
promotor <i>rhaT</i>	[2] (aprox. 4098690) - 4098590 (cadena negativa)	Vía [10]: unión a CRP 4098690-4098675, unión a ARN pol ('-35') 4098621-4098616, caja de Pribnow ('-10') 4098601-4098596
gen <i>rhaT</i>	4098589 - 4097514 (cadena negativa)	Vía [10]: inicio transcrito 4098589, <i>rhaT</i> ATG 4098548; NCBI: <i>rhaT</i> final de TAA 4097514
promotor <i>xyIAB</i>	[2] (aprox. 3728960) - 3728831 (cadena negativa)	Song y Park [11]: unión a CRP 3728919-3728901, unión a ARN pol ('-35') 3728865-3728860, caja de Pribnow ('-10') 3728841-3728836
operón <i>xyIAB</i>	3728830 - 3725940 (cadena negativa)	Song y Park [11]: inicio transcrito 3728830, <i>xyIA</i> ATG 3728788; NCBI: <i>xyIA</i> final de TAA 3727466; <i>xyIB</i> ATG 3727394, final de TAA 3725940
promotor <i>xyIFGHR</i>	[2] (aprox. 3728961) - 3729091 (cadena positiva)	Song y Park [11]: unión a ARN pol ('-35') 3729058-3729063, caja de Pribnow ('-10') 3729080-3729085
operón <i>xyIFGHR</i>	3729092 - 3734180 (cadena positiva)	Song y Park [11]: inicio transcrito 3729092, <i>xyIF</i> ATG 3729154; NCBI: <i>xyIF</i> final de TAA 3730146, <i>xyIG</i> ATG 3730224, final de TGA 3731765; <i>xyIH</i> ATG 3731743, final de TGA 3732924; <i>xyIR</i> ATG 3733002, final de TAG 3734180
promotor <i>xyIE</i>	[2] aprox. 4240482 - aprox. 4240320 (cadena negativa)	Davis y Henderson [12]: posible caja de Pribnow ('-10') 4240354-4240349, posible caja de Pribnow ('-10') 4240334-4240329
gen <i>xyIE</i>	(aprox. 4240319) - 4238802 (cadena negativa)	Davis y Henderson [12]: inicio transcrito deducido aprox. 4240319, <i>xyIE</i> ATG 4240277, final de TAA 4238802

Notas para la Tabla 1:

[1] Todas las ubicaciones de secuencias genómicas se refieren a la secuencia genómica de la subcepa MG1655 de *E. coli* K-12, proporcionada por el Centro Nacional de Información Biotecnológica (NCBI) como Referencia de Secuencia NCBI NC\_000913.2, 11-ENE-2012.

5 [2] La ubicación del extremo 5' (o "aguas arriba") de la región promotora es aproximada; para los promotores "bidireccionales", una ubicación de la secuencia de nucleótidos que es aproximadamente equidistante entre los sitios de inicio de la transcripción se selecciona como el "extremo" 5' designado para ambos promotores individuales. En la práctica, la porción de promotor de una estructura artificial de expresión puede tener algo menos de secuencia en su extremo 5' que las secuencias promotoras, tal como se indica en la tabla, o puede tener una  
10 secuencia de nucleótidos que incluye una secuencia adicional desde la región 5' (o "aguas arriba") de las secuencias promotoras tal como se indica en la tabla, a la vez que conserva la capacidad de favorecer la transcripción de una secuencia codificante aguas abajo, de una manera inducible.

[3] Smith y Schleif, "Nucleotide sequence of the L-arabinose regulatory region of *Escherichia coli* K12", J Biol Chem 1978 Oct. 10; 253(19): 6931-6933.

15 [4] 'ARN pol' indica ARN polimerasa a lo largo de la tabla.

[5] Miyada, *et al.*, "DNA sequence of the *araC* regulatory gene from *Escherichia coli* B/r", Nucleic Acids Res 1980 Nov. 25; 8(22): 5267-5274.

[6] Stoner y Schleif, "*E. coli araE* regulatory region *araE* codes for the low affinity L-arabinose uptake protein", Número de registro de la base de datos de GenBank X00272.1, fecha de revisión 6 JUL. 1989.

20 [7] Hendrickson *et al.*, "Sequence elements in the *Escherichia coli araFGH* promoter", J Bacteriol 1992 Nov; 174(21): 6862-6871.

[8] Documento de patente de EE.UU. nº 8178338 B2; 15 de Mayo, 2012; Keasling, Jay; Figura 9.

[9] Wickstrum *et al.*, "The AraC/XylS family activator RhaS negatively autoregulates *rhaSR* expression by preventing cyclic AMP receptor protein activation", J Bacteriol 2010 Jan; 192(1): 225-232.

25 [10] Vía *et al.*, "Transcriptional regulation of the *Escherichia coli rhaT* gene", Microbiology 1996 Jul; 142(Pt 7): 1833-1840.

[11] Song y Park, "Organization and regulation of the D-xylose operons in *Escherichia coli* K-12: XylR acts as a transcriptional activator", J Bacteriol. 1997 Nov; 179(22): 7025-7032.

30 [12] Davis y Henderson, "The cloning and DNA sequence of the gene *xylE* for xylose-proton symport in *Escherichia coli* K12", J Biol Chem 1987 Oct 15; 262(29): 13928-13932.

Estructuras artificiales de expresión. Las estructuras artificiales de expresión son polinucleótidos diseñados para la expresión de uno o varios productos génicos de interés y, por lo tanto, no son moléculas presentes en la naturaleza. Las estructuras artificiales de expresión se pueden integrar en un cromosoma de la célula hospedadora, o mantenerse dentro de la célula hospedadora como moléculas de polinucleótidos que se replican  
35 independientemente del cromosoma de la célula hospedadora, tales como plásmidos o cromosomas artificiales. Un ejemplo de una estructura artificial de expresión es un polinucleótido resultante de la inserción de una o varias secuencias de polinucleótidos en un cromosoma de la célula hospedadora, en donde las secuencias de polinucleótidos insertadas alteran la expresión de las secuencias que codifican cromosomas. Un vector de expresión es una estructura artificial de expresión plasmídica, utilizada específicamente para la expresión de uno o varios  
40 productos génicos. Una o varias estructuras artificiales de expresión se pueden integrar en un cromosoma de la célula hospedadora o se pueden mantener en un polinucleótido extracromosómico, tal como un plásmido o un cromosoma artificial. Las siguientes son descripciones de los tipos particulares de secuencias de polinucleótidos que se pueden utilizar en estructuras artificiales de expresión para la coexpresión de productos génicos.

Orígenes de replicación. Las estructuras artificiales de expresión deben comprender un origen de replicación, también denominado replicón, con el fin de mantenerse dentro de la célula hospedadora como polinucleótidos que se replican independientemente. Diferentes replicones que utilizan el mismo mecanismo para la replicación, no se pueden mantener juntos en una única célula hospedadora a lo largo de divisiones celulares repetidas. Como resultado de ello, los plásmidos se pueden clasificar en grupos de incompatibilidad en función del origen de replicación que contienen, como se muestra en la Tabla 2.

50

Tabla 2. Orígenes de replicación y plásmidos representativos para uso en estructuras artificiales de expresión [1]

Grupo de Incompatibilidad:	Origen de Replicación	Número de Copias:	Plásmidos Representativos (nº de depósito ATCC):
colE1, pMB1	colE1	15 - 20	colE1 (ATCC 27138)
	pMB1	15 - 20	pBR322 (ATCC 31344)
	pMB1 modificado	500 - 700	pUC9 (ATCC 37252)
IncFII, pT181	R1(ts)	15 - 120	pMOB45 (ATCC 37106)
F, P1, p15A, pSC101, R6K, RK2 [2]	p15A	18 - 22	pACYC177 (ATCC 37031); pACYC184 (ATCC 37033); pPRO33 (Addgene 17810) [3]
	pSC101	~5	pSC101 (ATCC 37032); pGBM1 (ATCC 87497)
	RK2	4 - 7 [2]	RK2 (ATCC 37125)
CloDF13 [4]	CloDF13	20 - 40 [4]	pCDFDuet™-1 (EMD nº de catálogo de Millipore 71340-3)
ColA [4]	ColA	20 - 40 [4]	pCOLADuet™-1 (EMD nº de catálogo de Millipore 71406-3)
RSF1030 [4]	RSF1030 (también denominado NTP1)	>100 [4]	pRSFDuet™-1 (EMD nº de catálogo de Millipore 71341-3)

Notas para la Tabla 2:

[1] Adaptado de [www.bio.davidson.edu/courses/Molbio/Protocols/ORIs.html](http://www.bio.davidson.edu/courses/Molbio/Protocols/ORIs.html), y Sambrook y Russell, "Molecular Cloning: A Laboratory Manual", 3ª Ed., Cold Spring Harbor Laboratory Press, 2001.

[2] Kües y Stahl, "Replication of plasmids in gram-negative bacteria", *Microbiol Rev* 1989 Dec; 53(4): 491-516.

[3] El plásmido pPRO33 (documento de Patente de Estados Unidos nº 8178338 B2, 15 mayo de 2012; Keasling, Jay) está disponible en Addgene ([www.addgene.org](http://www.addgene.org)) como plásmido 17810 de Addgene.

[4] [openwetware.org/wiki/CH391L/S12/Origins\\_of\\_Replication](http://openwetware.org/wiki/CH391L/S12/Origins_of_Replication); visitada 03 Ago de 2013.

10 Los orígenes de replicación se pueden seleccionar para uso en estructuras artificiales de expresión basándose en el grupo de incompatibilidad, el número de copias y/o el tipo de hospedador, entre otros criterios. Como se ha descrito anteriormente, si dos o más estructuras artificiales de expresión diferentes se van a utilizar en la misma célula hospedadora para la coexpresión de múltiples productos génicos, lo mejor es que las diferentes estructuras artificiales de expresión contengan orígenes de replicación procedentes de diferentes grupos de incompatibilidad: un replicón pMB1 en una estructura artificial de expresión y un replicón p15A en otra, por ejemplo. El número promedio de copias de una estructura artificial de expresión en la célula, en relación con el número de moléculas de cromosomas del hospedador, se determina por el origen de replicación contenido en esa estructura artificial de expresión. El número de copias puede variar desde unas pocas copias por célula a varios cientos (Tabla 2). En una realización de la invención, se emplean diferentes estructuras artificiales de expresión que comprenden promotores inducibles que se activan con el mismo inductor, pero que tienen diferentes orígenes de replicación. Mediante la selección de orígenes de replicación que conservan cada estructura artificial de expresión diferente en un cierto número de copias aproximado en la célula, es posible ajustar los niveles de producción total de un producto génico expresado a partir de una estructura artificial de expresión, en relación con otro producto génico expresado a partir de una diferente estructura artificial de expresión. A modo de ejemplo, para coexpresar subunidades A y B de una proteína multímera, se crea una estructura artificial de expresión que comprende el replicón colE1, el promotor *ara* y una secuencia codificadora de la subunidad A expresada a partir del promotor *ara*: 'colE1-P<sub>ara</sub>-A'. Otra estructura artificial de expresión se crea de modo que comprenda el replicón p15A, el promotor *ara* y una secuencia codificadora de la subunidad B: 'p15A-P<sub>ara</sub>-B'. Estas dos estructuras artificiales de expresión se pueden mantener

juntas en las mismas células hospedadoras, y la expresión de ambas subunidades A y B se induce mediante la adición de un inductor, arabinosa, al medio de crecimiento. Si el nivel de expresión de la subunidad A se tiene que aumentar significativamente en relación con el nivel de expresión de la subunidad B, con el fin de que la relación estequiométrica de las cantidades expresadas de las dos subunidades se acerque a una proporción deseada, por ejemplo, se podría crear una nueva estructura artificial de expresión para la subunidad A, que tenga un replicón pMB1 modificado como se encuentra en el origen de replicación del plásmido pUC9 ('pUC9ori'): pUC9ori-P<sub>ara</sub>-A. Expresar la subunidad A a partir de una estructura artificial de expresión con un número elevado de copias, tal como pUC9ori-P<sub>ara</sub>-A, debe aumentar la cantidad de subunidad A producida, en relación con la expresión de la subunidad B desde p15A-P<sub>ara</sub>-B. De manera similar, el uso de un origen de replicación que conserva estructuras artificiales de expresión con un número de copias inferior, tal como pSC101, podría reducir el nivel general de un producto génico expresado a partir de esa estructura artificial. La selección de un origen de replicación también puede determinar qué células hospedadoras pueden mantener una estructura artificial de expresión que comprenda ese replicón. Por ejemplo, estructuras artificiales de expresión que comprenden el origen de replicación colE1 tienen un margen relativamente reducido de hospedadores disponibles, las especies dentro de la familia Enterobacteriaceae, cuyas estructuras artificiales de expresión que comprenden el replicón RK2, se pueden mantener en *E. coli*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, *Azotobacter vinelandii* y *Alcaligenes eutrophus*, y si una estructura artificial de expresión comprende el replicón RK2 y algunos genes reguladores procedentes del plásmido RK2, se puede mantener en células hospedadoras tan diversas como *Sinorhizobium meliloti*, *Agrobacterium tumefaciens*, *Caulobacter crescentus*, *Acinetobacter calcoaceticus* y *Rhodobacter sphaeroides* (Kues y Stahl, "Replication of plasmids in gram-negative bacteria", Microbiol Rev 1989 Dec; 53(4): 491-516).

Consideraciones similares se pueden emplear para crear estructuras artificiales de expresión para la coexpresión inducible en células eucariotas. Por ejemplo, el plásmido circular de 2 micras de *Saccharomyces cerevisiae* es compatible con plásmidos de otras cepas de levadura, tales como pSR1 (depósito de la ATCC nº 48233 y 66069; Araki *et al.*, "Molecular and functional organization of yeast plasmid pSR1", J Mol Biol 20 Mar 1985; 182(2): 191-203) y pKD1 (depósito de la ATCC nº 37519; Chen *et al.*, "Sequence organization of the circular plasmid pKD1 from the yeast *Kluyveromyces drosophilum*", Nucleic Acids Res 11 Jun 1986; 14(11): 4471-4481).

**Marcadores seleccionables.** Las estructuras artificiales de expresión generalmente comprenden un gen de selección, también denominado un marcador seleccionable, que codifica una proteína necesaria para la supervivencia o el crecimiento de células hospedadoras en un medio de cultivo selectivo. Las células hospedadoras que no contienen la estructura artificial de expresión que comprende el gen de selección, no sobrevivirán en el medio de cultivo. Los genes de selección típicos codifican proteínas que confieren resistencia a antibióticos u otras toxinas, o que complementan deficiencias auxotróficas de la célula hospedadora. Un ejemplo de un esquema de selección utiliza un fármaco, tal como un antibiótico, para detener el crecimiento de una célula hospedadora. Aquellas células que contienen una estructura artificial de expresión que comprende el marcador seleccionable, producen una proteína que confiere resistencia al fármaco y sobreviven al régimen de selección. Algunos ejemplos de antibióticos que se utilizan comúnmente para la selección de marcadores seleccionables (y abreviaturas que indican genes que proporcionan fenotipos de resistencia a antibióticos) son: ampicilina (Amp<sup>R</sup>), cloranfenicol (Cml<sup>R</sup> o Cm<sup>R</sup>), kanamicina (Kan<sup>R</sup>), espectinomina (Spc<sup>R</sup>), estreptomina (Str<sup>R</sup>) y tetraciclina (Tet<sup>R</sup>). Muchos de los plásmidos representativos de la Tabla 2 comprenden marcadores seleccionables, tales como pBR322 (Amp<sup>R</sup>, Tet<sup>R</sup>); pMOB45 (Cm<sup>R</sup>, Tet<sup>R</sup>); pACYC177 (Amp<sup>R</sup>, Kan<sup>R</sup>); y pGBM1 (Spc<sup>R</sup>, Str<sup>R</sup>). La región del promotor natural para un gen de selección, generalmente se incluye junto con la secuencia que codifica su producto génico, como parte de una porción de marcador seleccionable de una estructura artificial de expresión. Alternativamente, la secuencia que codifica el gen de selección, se puede expresar a partir de un promotor constitutivo.

**Promotor inducible.** Como se describe en el presente documento, existen varios promotores inducibles diferentes que se pueden incluir en estructuras artificiales de expresión como parte de los sistemas de coexpresión inducible de la invención. Los promotores inducibles preferidos comparten al menos 80% de identidad de secuencia de polinucleótidos (más preferiblemente, al menos 90%) de identidad y lo más preferiblemente, al menos 95% de identidad) con al menos 30 (más preferiblemente, al menos 40 y lo más preferiblemente, al menos 50) bases contiguas de una secuencia de polinucleótidos de promotor, tal como se definen en la Tabla 1, en función de la secuencia genómica de la subcepa MG1655 de *E. coli* K-12, en donde el porcentaje de identidad de secuencia de los polinucleótidos se determina usando los métodos del Ejemplo 13. Por condiciones de inducción "convencionales" (véase el Ejemplo 5), los promotores inducibles preferidos tienen al menos 75% (más preferiblemente, al menos 100%, y lo más preferiblemente, al menos 110%) de la fuerza del promotor inducible de "tipo silvestre" correspondiente de la subcepa MG1655 de *E. coli* K-12, como se determina utilizando el método de PCR cuantitativa de De Mey *et al.* (Ejemplo 8). Dentro de la estructura artificial de expresión, un promotor inducible se coloca en 5' (o "aguas arriba de") la secuencia que codifica el producto génico que se va a expresar de manera inducible, por lo que la presencia del promotor inducible dirigirá la transcripción de la secuencia que codifica el producto génico en dirección 5' a 3' con respecto a la cadena codificante del polinucleótido que codifica el producto génico.

**Sitio de unión al ribosoma.** Para los productos génicos de polipéptidos, la secuencia de nucleótidos de la región entre el sitio de iniciación de la transcripción y el codón de iniciación de la secuencia codificante del producto génico que se va a expresar de manera inducible, se corresponde con la región 5' no traducida ('UTR') del ARNm para el producto génico del polipéptido. Preferiblemente, la región de la estructura artificial de expresión que se corresponde con la UTR 5', comprende una secuencia de polinucleótidos similar al sitio de unión a ribosoma de consenso (RBS,

también llamado la secuencia de Shine-Dalgarno) que se encuentra en las especies de la célula hospedadora. En procariontes (arqueas y bacterias), la secuencia de consenso de RBS es GGAGG o GGAGGU, y en bacterias tales como *E. coli*, la secuencia de consenso de RBS es AGGAGG o AGGAGGU. El RBS se separa normalmente del codón de iniciación por 5 a 10 nucleótidos intermedios. En estructuras artificiales de expresión, la secuencia de RBS es preferiblemente al menos 55% idéntica a la secuencia de consenso AGGAGGU, más preferiblemente al menos 70% idéntica, y lo más preferiblemente al menos 85% idéntica, y se separa del codón de iniciación por 5 a 10 nucleótidos intermedios, más preferentemente por 6 a 9 nucleótidos intermedios, y lo más preferiblemente por 6 o 7 nucleótidos intermedios. La capacidad de un determinado RBS para producir una tasa de iniciación de la traducción deseable se puede calcular en el sitio web [salis.psu.edu/software/RBSLibraryCalculatorSearchMode](http://salis.psu.edu/software/RBSLibraryCalculatorSearchMode), usando la calculadora RBS; la misma herramienta se puede utilizar para optimizar un RBS sintético para una tasa de traducción a lo largo de un margen de 100.000 veces (Salis, "The ribosome binding site calculator", *Methods Enzymol* 2011; 498: 19-42).

Sitio de clonación múltiple. Un sitio de clonación múltiple (MCS), también denominado un polienlazador, es un polinucleótido que contiene múltiples sitios de restricción próximos o solapados entre sí. Los sitios de restricción en el MCS se presentan normalmente una vez dentro de la secuencia de MCS, y preferiblemente no aparecen dentro del resto del plásmido u otra estructura artificial de polinucleótidos, permitiendo que las enzimas de restricción corten el plásmido u otra estructura artificial de polinucleótidos solo dentro del MCS. Los ejemplos de secuencias de MCS son los de la serie pBAD de vectores de expresión, que incluyen pBAD18, pBAD18-Cm, pBAD18-Kan, pBAD24, pBAD28, pBAD30 y pBAD33 (Guzman *et al.*, "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter", *J Bacteriol* 1995 Jul; 177(14): 4121-4130); o los de la serie pPRO de vectores de expresión obtenidos a partir de los vectores pBAD, tales como pPRO18, pPRO18-Cm, pPRO18-Kan, pPRO24, pPRO30 y pPRO33 (documento de Patente de EE.UU. n° 8178338 B2; 15 de mayo 2012; Keasling, Jay). Un sitio de clonación múltiple se puede utilizar en la creación de una estructura artificial de expresión: mediante la colocación de un sitio de clonación múltiple en 3' (o aguas abajo) de una secuencia de promotor, el MCS se puede utilizar para insertar la secuencia que codifica un producto génico que se va a coexpresar en la estructura artificial, en la ubicación correcta en relación con el promotor, de modo que se produzca la transcripción de la secuencia codificante. Dependiendo de qué enzima de restricción se utiliza para cortar dentro del MCS, puede haber alguna parte de la secuencia de MCS que permanezca dentro de la estructura artificial de expresión después de que se inserte la secuencia codificante u otra secuencia de polinucleótidos en la estructura artificial de expresión. Cualquier secuencia de MCS que permanezca, puede ser aguas arriba o aguas abajo, o en ambos lados de la secuencia insertada. Un sitio de unión al ribosoma se puede colocar aguas arriba del MCS, preferiblemente inmediatamente adyacente, o separado del MCS por solo unos pocos nucleótidos, en cuyo caso el RBS estará aguas arriba de cualquier secuencia codificante insertada en el MCS. Otra alternativa es incluir un sitio de unión al ribosoma dentro del MCS, en cuyo caso la elección de las enzimas de restricción usadas para cortar dentro del MCS, determinará si el RBS se conserva, y en qué relación frente a, las secuencias insertadas. Una alternativa adicional es incluir un RBS dentro de la secuencia de polinucleótidos que se va a insertar en la estructura artificial de expresión en el MCS, preferiblemente en una relación apropiada con cualquiera de las secuencias codificadoras, para estimular el inicio de la traducción desde el ARN mensajero transcrito.

Expresión desde promotores constitutivos. Las estructuras artificiales de expresión de la invención también pueden comprender secuencias codificadoras que se expresan a partir de promotores constitutivos. A diferencia de los promotores inducibles, los promotores constitutivos inician una producción continua de producto génico bajo la mayoría de las condiciones de crecimiento. Un ejemplo de un promotor constitutivo es el del gen *Tn3 bla*, que codifica beta-lactamasa y es responsable del fenotipo de resistencia a la ampicilina (Amp<sup>R</sup>) conferido a la célula hospedadora a través de muchos plásmidos, incluyendo pBR322 (ATCC 31344), pACYC177 (ATCC 37031) y pBAD24 (ATCC 87399). Otro promotor constitutivo que se puede utilizar en estructuras artificiales de expresión, es el promotor para el gen de la lipoproteína de *E. coli*, *lpp*, que se encuentra en las posiciones 1755731-1755406 (hebra positiva) en la subcepa MG1655 de *E. coli* K-12 (Inouye e Inouye, "Up-promoter mutations in the *lpp* gene of *Escherichia coli*", *Nucleic Acids Res* 1985 May 10; 13(9): 3101-3110). Un ejemplo adicional de un promotor constitutivo que se ha utilizado para la expresión génica heteróloga en *E. coli* es el promotor *trpLEDCBA*, que se encuentra en las posiciones 1321169-1321133 (hebra negativa) de la subcepa MG1655 de *E. coli* K-12 (Windass *et al.*, "The construction of a synthetic *Escherichia coli* *trp* promoter and its use in the expression of a synthetic interferon gene", *Nucleic Acids Res* 1982 Nov 11; 10(21): 6639-6657). Los promotores constitutivos se pueden utilizar en estructuras artificiales de expresión para la expresión de marcadores de selección, como se describe en el presente documento, y también para la expresión constitutiva de otros productos génicos, útiles para la coexpresión del producto deseado. Por ejemplo, reguladores de la transcripción de los promotores inducibles, tales como AraC, PrpR, RhaR y XylR, si no se expresan a partir de un promotor inducible bidireccional, se pueden expresar, alternativamente, a partir de un promotor constitutivo, ya sea en la misma estructura artificial de expresión que el promotor inducible que regulan, o una estructura artificial de expresión diferente. Del mismo modo, los productos génicos útiles para la producción o el transporte del inductor, tal como PrpEC, AraE o Rha, o proteínas que modifican el entorno de reducción-oxidación de la célula, como algunos ejemplos, se pueden expresar a partir de un promotor constitutivo dentro de una estructura artificial de expresión. Los productos génicos útiles para la producción de productos génicos coexpresados, y el producto resultante deseado, también incluyen proteínas chaperonas, transportadores de cofactores, etc.

Péptidos señal. Los productos génicos de polipéptidos coexpresados por los métodos de la invención pueden contener péptidos señal o carecer de ellos, dependiendo de si es deseable que tales productos génicos se exporten desde el citoplasma de la célula hospedadora al periplasma, o que queden retenidos en el citoplasma, respectivamente. Los péptidos señal (también denominados secuencias señal, secuencias líder o péptidos líder) se caracterizan estructuralmente por un segmento de aminoácidos hidrófobos, con aproximadamente cinco a veinte aminoácidos de longitud y frecuentemente aproximadamente diez a quince aminoácidos de longitud, que tiene una tendencia a formar una sola hélice alfa. Este segmento hidrófobo, frecuentemente está precedido inmediatamente por un segmento más corto enriquecido con aminoácidos cargados positivamente (particularmente lisina). Los péptidos señal que se van a escindir del polipéptido maduro, por lo general terminan en un segmento de aminoácidos que es reconocido y escindido por la peptidasa señal. Los péptidos señal se pueden caracterizar funcionalmente por la capacidad para dirigir el transporte de un polipéptido, ya sea de modo cotraduccional o postraduccional, a través de la membrana plasmática de procariontes (o la membrana interna de bacterias Gram-negativas tales como *E. coli*), o dentro del retículo endoplásmico de las células eucariotas. El grado en que un péptido señal permite que un polipéptido sea transportado al espacio periplásmico de una célula hospedadora tal como *E. coli*, por ejemplo, se puede determinar mediante una separación de las proteínas periplásmicas procedentes de proteínas retenidas en el citoplasma, utilizando un método tal como el que se proporciona en el Ejemplo 12.

Células hospedadoras. Los sistemas de coexpresión inducible de la invención están diseñados para expresar múltiples productos génicos. En la invención, los productos génicos se coexpresan en una célula hospedadora. Se proporcionan ejemplos de células hospedadoras que permiten una coexpresión inducible eficaz y rentable de componentes de productos multímeros. Las células hospedadoras pueden incluir, además de las células aisladas en cultivo, células que forman parte de un organismo multicelular, o células que crecen dentro de un organismo o un sistema de organismos diferente. Además, las estructuras artificiales de expresión de los sistemas de coexpresión inducible de la invención se pueden utilizar en sistemas exentos de células, tales como los basados en extractos de germen de trigo o en extractos de células bacterianas, tales como un sistema de síntesis proteica exento de células con intercambio continuo (CECF), utilizando extractos de *E. coli* y un aparato de incubación, como el RTS ProteoMaster (Roche Diagnostics GmbH, Mannheim, Alemania). (Jun *et al.*, "Continuous-exchange cell-free protein synthesis using PCR-generated DNA and an RNase E-deficient extract", *Biotechniques* 2008 Mar; 44(3): 387-391).

Células hospedadoras procariontes. Las estructuras artificiales de expresión diseñadas para la coexpresión de productos génicos, se proporcionan en células hospedadoras, preferiblemente células hospedadoras procariontes. Las células hospedadoras procariontes pueden incluir arqueas (tales como *Haloferax volcanii*, *Sulfolobus solfataricus*), bacterias Gram-positivas (tales como *Bacillus subtilis*, *Bacillus licheniformis*, *Brevibacillus choshinensis*, *Lactobacillus brevis*, *Lactobacillus buchneri*, *Lactococcus lactis* y *Streptomyces lividans*), o bacterias Gram-negativas, que incluyen las alfa-proteobacterias (*Agrobacterium tumefaciens*, *Caulobacter crescentus*, *Rhodobacter sphaeroides* y *Sinorhizobium meliloti*), betaproteobacterias (*Alcaligenes eutrophus*), y gammaproteobacterias (*Acinetobacter calcoaceticus*, *Azotobacter vinelandii*, *Escherichia coli*, *Pseudomonas aeruginosa* y *Pseudomonas putida*). Las células hospedadoras preferidas incluyen gammaproteobacterias de la familia Enterobacteriaceae, tales como *Enterobacter*, *Erwinia*, *Escherichia* (incluyendo *E. coli*), *Klebsiella*, *Proteus*, *Salmonella* (incluyendo *Salmonella typhimurium*), *Serratia* (incluyendo *Serratia marcescens*) y *Shigella*.

Células hospedadoras eucariotas. Muchos otros tipos de células hospedadoras se pueden utilizar para los sistemas de coexpresión inducible de la invención, incluyendo células eucariotas tales como levadura (*Candida shehatae*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, otras especies de *Kluyveromyces*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Saccharomyces pastorianus* también conocido como *Saccharomyces carlsbergensis*, *Schizosaccharomyces pombe*, especies de *Dekkera/Brettanomyces* y *Yarrowia lipolytica*); otros hongos (*Aspergillus nidulans*, *Aspergillus niger*, *Neurospora crassa*, *Penicillium*, *Tolyposcladium*, *Trichoderma reesia*); líneas celulares de insecto (células Schneider 2 de *Drosophila melanogaster* y células Sf9 de *Spodoptera frugiperda*); y líneas celulares de mamífero incluyendo líneas celulares inmortalizadas (células de ovario de hámster chino (CHO), células HeLa, células de riñón de hámster neonato (BHK), células de riñón de mono (COS), células de riñón embrionario humano (HEK, 293 o HEK-293) y células de carcinoma hepatocelular humano (Hep G2)). Las células hospedadoras anteriores están disponibles de la American Type Culture Collection.

Alteraciones de las funciones génicas de células hospedadoras. Se pueden realizar ciertas alteraciones de las funciones génicas de las células hospedadoras que comprenden estructuras artificiales de expresión inducible, para favorecer una inducción eficaz y homogénea de la población de células hospedadoras a través de un inductor. Preferiblemente, la combinación de estructuras artificiales de expresión, el genotipo de la célula hospedadora y las condiciones de la inducción, da como resultado que al menos 75% (más preferiblemente al menos 85%, y lo más preferiblemente, al menos 95%) de las células en el cultivo expresan el producto génico de cada promotor inducido, tal como se mide por el método de Khlebnikov *et al.*, descrito en el Ejemplo 8. Para células hospedadoras distintas de *E. coli*, estas alteraciones pueden implicar la función de genes que son estructuralmente similares a un gen de *E. coli*, o a genes que llevan a cabo una función dentro de la célula hospedadora similar a la del gen de *E. coli*. Las alteraciones de las funciones de genes de células hospedadoras incluyen eliminar o reducir la función génica mediante una delección de la secuencia génica que codifica la proteína en su totalidad, o la delección de una gran porción suficiente del gen, la inserción de una secuencia en el gen, o alterando de otra forma la secuencia del gen, de modo que se produce un nivel reducido de producto génico funcional desde ese gen. Las alteraciones de las

funciones génicas de células hospedadoras también incluyen el aumento de la función génica mediante, por ejemplo, una alteración del promotor natural para crear un promotor más fuerte que dirige un mayor nivel de transcripción del gen, o la introducción de una mutación de sentido erróneo en la secuencia que codifica la proteína que da lugar a un producto génico mucho más activo. Las alteraciones de las funciones génicas de células hospedadoras incluyen la alteración de la función génica de cualquier modo, incluyendo, por ejemplo, la alteración de un promotor inducible natural para crear un promotor que está activado de forma constitutiva. Además de las alteraciones en las funciones de genes para el transporte y el metabolismo de los inductores, como se describen en este documento con relación a los promotores inducibles, y una expresión alterada de las proteínas chaperonas, también es posible alterar el sistema regulador mediante represión por catabolito de carbono (CCR) y/o el entorno de oxidación-reducción de la célula hospedadora.

Represión por catabolito de carbono (CCR). La presencia de un sistema regulador por CCR activo dentro de un hospedador puede afectar a la capacidad de un inductor para activar la transcripción desde un promotor inducible. Por ejemplo, cuando una célula hospedadora, tal como *E. coli*, se cultiva en un medio que contiene glucosa, los genes necesarios para la utilización de otras fuentes de carbono, tales como los operones *araBAD* y *prpBCDE*, se expresan como mucho a un nivel bajo, incluso si el inductor arabinosa o propionato también está presente en el medio de crecimiento. También existe una jerarquía en la utilización de fuentes de carbono diferentes de la glucosa: como en el caso de los sistemas de promotores inducibles *ara* y *prp*, en donde la presencia de arabinosa reduce la capacidad del propionato para inducir la expresión a partir del promotor *prpBCDE* (Park *et al.*, "The mechanism of sugar-mediated catabolite repression of the propionate catabolic genes in *Escherichia coli*", *Gene* 2012 Aug 1; 504(1): 116-121; Epub 2012 May 3). El mecanismo de CCR de la célula, por tanto, hace que sea más difícil utilizar dos o más inductores de fuentes de carbono en un sistema de coexpresión inducible, ya que la presencia del inductor que es la fuente de carbono preferida, inhibirá la inducción con fuentes de carbono menos preferidas. Los autores Park *et al.* intentaron atenuar la represión del promotor *prp* mediante arabinosa, ya fuera mediante el uso de un gen *crp* mutante que produce una proteína alterada del receptor de AMPc que puede actuar de manera independiente del AMPc, o una delección de genes de PTS (sistema de fosfotransferasa) implicados en la regulación de CCR; ambos enfoques no tuvieron mucho éxito. Sin embargo, la cepa inactivada para PTS, utilizada por los autores Park *et al.*, se basa en la cepa TP2811 que es una delección del operón *ptsHI-crr* de *E. coli* (Hernandez-Montalvo *et al.*, "Characterization of sugar mixtures utilization by an *Escherichia coli* mutant devoid of the phosphotransferase system", *Appl Microbiol Biotechnol* 2001 Oct; 57(1-2): 186-191). Se ha encontrado que la delección de todo el operón *ptsHI-crr* afecta a la síntesis total de AMPc de forma más significativa que una delección de solo el gen *crr* (Lévy *et al.*, "Cyclic AMP synthesis in *Escherichia coli* strains bearing known deletions in the *pts* phosphotransferase operon", *Gene* 1990 Jan 31; 86(1): 27-33). Un enfoque diferente es eliminar o reducir la función del gen *ptsG* en la célula hospedadora, que codifica EII A específico de glucosa (EII A<sup>glc</sup>), un elemento clave para la CCR en *E. coli* (Kim *et al.*, "Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass", *Appl Microbiol Biotechnol* 2010 Nov; 88(5): 1077-1085, Epub 2010 Sep 14). Otra alteración en el genoma de una célula hospedadora tal como *E. coli*, que conduce a un aumento de la transcripción del promotor *prp*, es eliminar o reducir la función génica del gen *ascG*, que codifica AscG. AscG es el represor del operón *ascFB* de utilización de beta-D-glucósido en condiciones de crecimiento normales, y también inhibe la transcripción del promotor *prp*; se ha observado que una interrupción de la secuencia que codifica AscG aumenta la transcripción desde el promotor *prp* (Ishida *et al.*, "Participation of regulator AscG of the beta-glucoside utilization operon in regulation of the propionate catabolism operon", *J Bacteriol* 2009 Oct; 191(19): 6136-6144; Epub 2009 Jul 24). Una alternativa adicional es aumentar la expresión del regulador transcripcional de promotores, inducible a través del inductor de fuente de carbono menos preferida, colocándolo ya sea bajo el control de un promotor constitutivo fuerte, o bajo el control del inductor de fuente de carbono más preferida. Por ejemplo, para aumentar la inducción de genes necesarios para la utilización de la xilosa que es una fuente de carbono menos preferida, en presencia de la arabinosa más preferida, la secuencia que codifica XylR se coloca en el operón *araBAD* de *E. coli* (Groff *et al.*, "Supplementation of intracellular XylR leads to cointilization of hemicellulose sugars", *Appl Environ Microbiol* 2012 Apr; 78(7): 2221-2229, Epub 2012 Jan 27). Por tanto, las células hospedadoras que comprenden estructuras artificiales de coexpresión inducible, incluyen preferiblemente un mayor nivel de función génica para reguladores de la transcripción de promotores, inducible con el o los inductores de fuentes de carbono menos preferidas, y una función génica eliminada o reducida para los genes implicados en el sistema CCR, tales como *crr* y/o *ptsG* y/o *ascG*.

Transporte celular del grupo hemo y otros cofactores. Cuando se utilizan los sistemas de coexpresión inducible de la invención para producir enzimas que requieren cofactores para actuar, es útil usar una célula hospedadora capaz de sintetizar el cofactor a partir de precursores disponibles, o de tomarlos desde el entorno. Los cofactores comunes incluyen ATP, coenzima A, dinucleótido de flavina adenina (FAD), NAD<sup>+</sup>/NADH y grupos hemo. Los grupos hemo comprenden un ion de hierro en el centro de un anillo heterocíclico orgánico grande llamado porfirina. El tipo más común entre los grupos hemo es hemo B; otros tipos principales son hemo A, hemo C y hemo O, que varían en las cadenas laterales de la porfirina. La hemina es cloruro de hemo B y se puede añadir al medio de crecimiento bacteriano como fuente de hemo. Otras fuentes potenciales de hemo incluyen citocromos, hemoglobina y sustancias que contienen hemoglobina, tales como la sangre. Las cepas de laboratorio de *E. coli* obtenidas a partir de *E. coli* K12, normalmente carecen de un receptor de hemo en la membrana externa, y por lo tanto no transportan hemo dentro de la célula, no pudiendo crecer en medios en los que hemo es la única fuente de hierro. Cepas patógenas de *E. coli* tales como O157:H7 y CFT073 contienen un segmento genómico de aproximadamente 9 kb que no está

presente en *E. coli* K12, y que contiene dos operones transcritos de forma divergente que codifican proteínas implicadas en la captación y utilización de hemo: el operón de *chuAS* y el operón *chuTWXYUhmuv*. Este segmento genómico se encuentra en *E. coli* CFT073 e incluye la secuencia de referencia de NCBI NC\_004431.1 (20-ENE-2012) desde la posición 4.084.974 a 4.093.975. La transformación con el gen *chuA* (por ejemplo, n° de id. del gen de NCBI 1037196), que codifica un receptor específico de hemina en la membrana externa, era suficiente para conferir a una cepa de *E. coli* obtenida a partir de K12, la capacidad para crecer sobre hemina como fuente de hierro (Torres y Payne, "Haem iron-transport system in enterohaemorrhagic *Escherichia coli* 0157:H7", *Mol Microbiol* 1997 Feb; 23(4): 825-833). Además de *ChuA*, algunos otros receptores de hemo heterólogos pueden permitir que cepas de *E. coli* obtenidas a partir de K12 capten el grupo hemo: *Yersinia enterocolitica* HemR, *Serratia marcescens* HasR y *Shigella dysenteriae* Shua; y a partir de bacterias Gram-negativas: *Bordetella pertussis* y *B. bronchiseptica* y Bhur, *Pseudomonas aeruginosa* Phur y *P. fluorescens* PfhR. La proteína *ChuS* también está involucrada en la utilización de hemo: es una hemo oxigenasa degradante de hemo. En una cepa de *E. coli aroB*, que es deficiente en la síntesis de la molécula enterobactina quelante de hierro, la transformación con *chuS* era útil para reducir la toxicidad celular causada por el crecimiento sobre hemina en ausencia de enterobactina. La transcripción de los operones *chuAS* y *chuTWXYUhmuv*, y varios otros operones que están involucrados en el metabolismo del hierro y presentes en las cepas de *E. coli* K12, está inhibida por el regulador transcripcional de *E. coli* Fur, cuando se asocia con Fe<sup>2+</sup>; la transcripción de estos genes se activa por tanto cuando hay una caída en la concentración intracelular de iones de hierro.

Células hospedadoras tales como cepas de *E. coli* obtenidas a partir de K12 se pueden alterar para que puedan captar hemo mediante transformación del mismo con la totalidad o parte de la región de *chuAS* - *chuTWXYUhmuv* que contiene al menos *chuA*, o con el operón *chuAS*, que incluye opcionalmente genes adicionales procedentes del operón *chuTWXYUhmuv*. En realizaciones en las que el promotor que dirige la transcripción de *chuA* se puede inhibir con Fur, la represión con Fur se puede eliminar o reducir delecionando el gen de la célula hospedadora que codifica Fur, cultivando las células hospedadoras en ausencia de hierro libre, cultivando las células hospedadoras en presencia de un agente quelante de iones de hierro libres tal como EDTA (ácido etilendiaminotetraacético), o mediante la transformación de las células con estructuras artificiales de polinucleótidos (tales como plásmidos mantenidos con un número de copias elevado) que comprenden múltiples copias del sitio de unión a Fur, para reducir la cantidad de complejo Fur-Fe<sup>2+</sup> disponible inhibir operones del metabolismo del hierro. En una realización preferida, una estructura artificial de expresión se introduce en la célula hospedadora, en donde la estructura artificial de expresión comprende un polinucleótido que codifica *ChuA* y, opcionalmente, también codifica *ChuS*, bajo el control transcripcional de un promotor constitutivo.

Entorno de reducción-oxidación de una célula hospedadora. Muchos productos génicos multímeros, tales como anticuerpos, contienen enlaces disulfuro. El citoplasma de *E. coli* y de muchas otras células se mantiene normalmente en un estado reducido a través de los sistemas enzimáticos de tiorredoxina y glutarredoxina/glutación. Esto impide la formación de enlaces disulfuro en el citoplasma, y las proteínas que necesitan enlaces disulfuro se exportan al periplasma en donde la formación de enlaces disulfuro y la isomerización están catalizadas por el sistema de Dsb, que comprende DsbABCD y DsbG. El aumento de la expresión de la cisteína oxidasa DsbA, la disulfuro isomerasa DsbC o combinaciones de las proteínas Dsb, las cuales se transportan todas normalmente al periplasma, se ha utilizado en la expresión de proteínas heterólogas que requieren enlaces disulfuro (Makino *et al.*, "Strain engineering for improved expression of recombinant proteins in bacteria", *Microb Cell Fact* 2011 May 14; 10:32). También es posible expresar formas citoplasmáticas de estas proteínas Dsb, tales como una versión citoplásmica de DsbC ('cDsbC'), que carece de un péptido señal y por lo tanto no se transporta al periplasma. Proteínas Dsb citoplasmáticas tales como cDsbC son útiles para hacer el citoplasma de la célula hospedadora más oxidante y, por lo tanto, más propicio para la formación de enlaces disulfuro en proteínas heterólogas producidas en el citoplasma. El citoplasma de la célula hospedadora también se puede hacer más oxidante alterando directamente los sistemas enzimáticos de tiorredoxina y glutarredoxina/glutación: cepas mutantes defectuosas para la glutatión reductasa (*gor*) o glutatión sintetasa (*gshB*), junto con la tiorredoxina reductasa (*trxB*), vuelven el citoplasma oxidante. Estas cepas son incapaces de reducir ribonucleótidos y, por lo tanto, no pueden crecer en ausencia de un agente reductor exógeno, tal como ditiotreitól (DTT). Mutaciones supresoras (*ahpC*<sup>\*</sup>) en el gen *ahpC*, que codifica la peroxirredoxina AhpC, lo convierten en una disulfuro reductasa que genera glutatión reducido, lo que permite la canalización de electrones a la enzima ribonucleótido reductasa y permite que las células defectuosas para *gor* y *trxB*, o defectuosas para *gshB* y *trxB*, crezcan en ausencia de DTT. Una clase diferente de formas mutadas de AhpC puede permitir que las cepas defectuosas en la actividad gamma-glutamylcisteína sintetasa (*gshA*) y defectuosas en *trxB*, crezcan en ausencia de DTT; estas incluyen AhpC V164G, AhpC S71F, AhpC E173/S71F, AhpC E171Ter y AhpC dup162-169 (Faulkner *et al.*, "Functional plasticity of a peroxidase allows evolution of diverse disulfide-reducing pathways", *Proc Natl Acad Sci U S A* 2008 May 6; 105(18): 6735-6740, Epub 2008 May 2). En tales cepas con citoplasma oxidante, las cisteínas proteicas expuestas se oxidan fácilmente en un proceso que está catalizado por tiorredoxinas, en una inversión de su función fisiológica, lo que da lugar a la formación de enlaces disulfuro.

Otra alteración que se puede realizar en las células hospedadoras es expresar la sulfhidrilo oxidasa Erv1p desde el espacio de la membrana interna de las mitocondrias de levadura en el citoplasma de la célula hospedadora, lo que se ha observado que aumenta la producción de una variedad de complejos, proteínas enlazadas a disulfuro de origen eucariota en el citoplasma de *E. coli*, incluso en ausencia de mutaciones en *gor* o *trxB* (Nguyen *et al.*, "Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins

expressed in the cytoplasm of *E. coli*" Microb Cell Fact 2011 Jan 7; 10:1). Las células hospedadoras que comprenden estructuras artificiales con coexpresión inducible, preferiblemente también expresan cDsbC y/o Erv1p, que son deficientes en la función del gen *trx*B, también son deficientes en la función génica de cualquiera entre *gor*, *gsh*B o *gsh*A, y expresan una forma mutante adecuada de AhpC, de modo que las células hospedadoras se pueden cultivar en ausencia de DTT.

Glicosilación de productos génicos de polipéptidos. Las células hospedadoras pueden tener alteraciones en su capacidad para glicosilar polipéptidos. Por ejemplo, las células hospedadoras eucarióticas pueden haber eliminado o reducido la función génica en los genes de glicosiltransferasa y/o oligosacariltransferasa, impidiendo una glicosilación eucariótica normal de los polipéptidos para formar glicoproteínas. Las células hospedadoras procariontas tales como *E. coli*, que normalmente no glicosilan polipéptidos, se pueden alterar para expresar un conjunto de genes eucariotas y procariontas que proporcionan una función de glicosilación (DeLisa *et al.*, "Glycosylated protein expression in prokaryotes", documento WO2009089154A2, 2009 Jul 16).

Cepas de células hospedadoras disponibles con funciones génicas alteradas. Para crear cepas preferidas de células hospedadoras que se van a utilizar en los sistemas de coexpresión inducible y los métodos de la invención, es útil comenzar con una cepa que ya comprende las alteraciones genéticas deseadas (Tabla 3).

Tabla 3. Cepas de células hospedadoras

Cepa:	Genotipo:	Fuente:
<i>E. coli</i> TOP10	F- <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74</i> <i>recA1</i> <i>araD139</i> $\Delta$ ( <i>ara-leu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str <sup>R</sup> ) <i>endA1</i> <i>nupG</i> $\lambda$ -	Invitrogen Life Technologies nº de catálogo C4040-10, C4040-03, C4040-06, C4040-50 y C4040-52
<i>E. coli</i> Origami™ 2	$\Delta$ ( <i>ara-leu</i> )7697 $\Delta$ <i>lacX74</i> $\Delta$ <i>phoA</i> <i>PvuII</i> <i>phoR</i> <i>araD139</i> <i>ahpC</i> <i>galE</i> <i>galK</i> <i>rpsL</i> F' [ <i>lac</i> <sup>+</sup> <i>lacI</i> <sup>f</sup> <i>pro</i> ] <i>gor522::Tn10</i> <i>trx</i> B (Str <sup>R</sup> , Tet <sup>R</sup> )	Merck (EMD Millipore Chemicals) nº de catálogo 71344
<i>E. coli</i> SHuffle® Express	<i>fhuA2</i> [ <i>lon</i> ] <i>ompT</i> <i>ahpC</i> <i>gal</i> <i>latt::pNEB3-rl-</i> <i>cDsbC</i> (Spec, <i>lacI</i> ) $\Delta$ <i>trx</i> B <i>sulA11</i> <i>R(mcr-</i> <i>73::miniTn10--Tet<sup>S</sup>)</i> <sub>2</sub> [ <i>dcm</i> ] <i>R(zgb-</i> <i>210::Tn10--Tet<sup>S</sup>)</i> <i>endA1</i> $\Delta$ <i>gor</i> $\Delta$ ( <i>mcrC-</i> <i>mrr</i> )114::IS10	New England Biolabs nº de catálogo C3028H

Métodos para alterar las funciones génicas de células hospedadoras. Existen muchos métodos conocidos en la técnica para la realización de alteraciones en genes de células hospedadoras con el fin de eliminar, reducir o cambiar la función génica. Los métodos para realizar alteraciones específicas de genes en células hospedadoras tales como *E. coli* y otras procariontas, han sido descritos (Muyrers *et al.*, "Rapid modification of bacterial artificial chromosomes by ET- recombination", Nucleic Acids Res 1999 Mar 15; 27(6): 1555-1557; Datsenko y Wanner, "One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products", Proc Natl Acad Sci U S A 2000 Jun 6; 97(12): 6640- 6645), y los kits para el uso de métodos similares de recombinación Red/ET están comercialmente disponibles (por ejemplo, el kit de delección de genes en *E. coli* Quick & Easy de Gene Bridges GmbH, Heidelberg, Alemania). En una realización de la invención, la función de uno o varios genes de las células hospedadoras se elimina o se reduce mediante la identificación de una secuencia de nucleótidos dentro de la secuencia que codifica el gen que se va a alterar, tal como una de las secuencias codificadoras de la subcepa MG1655 de *E. coli* K-12 incorporada en este documento como referencia, para la ubicación genómica de la secuencia, y más específicamente mediante la selección de dos segmentos adyacentes de 50 nucleótidos cada uno, dentro de esa secuencia codificante. El kit de delección de genes en *E. coli* Quick & Easy se utiliza a continuación de acuerdo con las instrucciones del fabricante para insertar una estructura artificial de polinucleótido que contiene un marcador seleccionable entre los segmentos adyacentes seleccionados de la secuencia codificante, eliminando o reduciendo la función normal del gen. Los métodos de recombinación Red/ET también se pueden utilizar para reemplazar una secuencia de promotor con la de un promotor diferente, tal como un promotor constitutivo, o un promotor artificial que se prevé que favorece un cierto nivel de la transcripción (De Mey *et al.*, "Promoter knock-in: a novel rational method for the fine tuning of genes", BMC Biotechnol 2010 Mar 24; 10: 26). La función de los genes de la célula hospedadora también se puede eliminar o reducir por métodos de silenciamiento de ARN (Man *et al.*, "Artificial transencoded small non-coding RNAs specifically silence the selected gene expression in bacteria", Nucleic Acids Res 2011 Apr; 39(8): e50, Epub 2011 Feb 3). Además, mutaciones conocidas que alteran la función génica de la célula hospedadora se pueden introducir en las células hospedadoras a través de métodos genéticos tradicionales.

Sistemas de coexpresión inducible de la invención

Los sistemas de coexpresión inducible de la invención implican células hospedadoras que comprenden dos o más estructuras artificiales de expresión, en donde las estructuras artificiales de expresión comprenden promotores inducibles que dirigen la expresión de productos génicos, y las células hospedadoras tienen funciones alteradas de genes que permiten una expresión inducible homogénea de los productos génicos. La Fig. 1 muestra una representación esquemática de un sistema de coexpresión inducible de la invención, con los siguientes componentes: (1) célula hospedadora, (2) genoma del hospedador (incluyendo las alteraciones genéticas), (3) un vector de expresión 'X' que comprende un promotor inducible que dirige la expresión de un producto génico, (4) un vector de expresión diferente 'Y' que comprende un promotor inducible que dirige la expresión de otro producto génico, (5) inductores químicos de la expresión y (6) producto de coexpresión múltimera.

La Fig. 2 muestra una representación esquemática de un ejemplo particular de un sistema de coexpresión inducible de la invención, que emplea el promotor *araBAD* sobre un vector de expresión pBAD24 en combinación con un promotor inducible con propionato (promotor *prpBCDE*) sobre un vector de expresión pPRO33 (documento de Patente de EE.UU. n° 8178338 B2, 15 mayo de 2012; Keasling, Jay), en una célula hospedadora *E. coli* que alberga las alteraciones genómicas apropiadas que permiten una expresión inducible de forma homogénea. De esta manera, se puede lograr un estrecho control y una optimización de la expresión de cada componente de un producto múltimero, para uso en una variedad de aplicaciones de la coexpresión. En esta realización, la célula hospedadora (1) es la bacteria Gram-negativa *Escherichia coli*, que se utiliza comúnmente en la técnica para la expresión de proteínas. El genoma del hospedador (2) es el genoma del organismo de la célula hospedadora con mutaciones u otras alteraciones que facilitan la coexpresión inducible de proteínas de forma homogénea, incluyendo la expresión de una forma citoplásmica de la disulfuro isomerasa DsbC que carece de un péptido señal. En una realización, las alteraciones genómicas incluyen tanto una mutación que inactiva el operón *araBAD* como la expresión de *araE* y *araFGH* desde promotores constitutivos, o una mutación puntual en el gen *lacY* (A117C) en un fondo que carece de *araEFGH*, para facilitar una inducción homogénea de promotores *ara* basados en plásmidos con L-arabinosa aplicada de forma exógena, y también un gen del metabolismo del propionato inactivado, *prpD*, para facilitar una inducción homogénea de promotores de propionato basados en plásmidos, con propionato aplicado de forma exógena, que se convierte en 2-metilcitrato *in vivo*. Otras alteraciones genómicas que son útiles para el sistema de coexpresión inducible y que se pueden introducir en la célula hospedadora, incluyen, sin limitación: una inactivación dirigida del operón *scpA-argK-scpBC*, para reducir la expresión de fondo desde el promotor *prpBCDE*; una expresión del regulador transcripcional (*prpR*) para la fuente de carbono menos preferida (propionato) desde un promotor inducible con L-arabinosa tal como el promotor *araBAD*, y/o una función génica eliminada o reducida para genes implicados en el sistema CCR, como *crr* y/o *ptsG*, para evitar la supresión mediante el sistema CCR de una inducción con propionato en presencia de L-arabinosa; reducciones del nivel de la función génica para la glutatión reductasa (*gor*) o la glutatión sintetasa (*gshB*), junto con la tiorredoxina reductasa (*trxB*) y/o la expresión de la sulfhidrilo oxidasa Erv1p mitocondrial de levadura en el citoplasma de la célula hospedadora, para proporcionar un entorno reductor menos fuerte en el citoplasma de la célula hospedadora y favorecer la formación de enlaces disulfuro; aumento de los niveles de expresión, por ejemplo, desde un promotor constitutivo fuerte, de proteínas chaperonas tales como DnaK/DnaJ/GrpE, DsbC/DsbG, GroEL/GroES, IbpA/IbpB, Skp, Tig (factor desencadenante) y/o FkpA; y otras mutaciones para reducir la actividad de proteasas endógenas (tal como la de las proteasas Lon y OmpT) y actividades de recombinasa.

Como se muestra en la Fig. 2, dos vectores de expresión compatible (3, 4) se mantienen en la célula hospedadora para permitir una expresión simultánea (coexpresión) de dos productos génicos diferentes. En esta realización, un vector de expresión ("vector de expresión inducido con L-arabinosa") contiene un promotor inducido con L-arabinosa, y es similar o idéntico a pBAD o plásmidos relacionados en los que un promotor *araBAD* dirige la expresión de una secuencia de expresión insertada, clonada en el sitio de clonación múltiple (MCS). El vector de expresión inducido con L-arabinosa también contiene una secuencia que codifica un gen de resistencia a antibióticos (tal como el gen *Tn3 bla*, que codifica la beta-lactamasa y confiere resistencia a la ampicilina) para facilitar la selección de células hospedadoras (colonias bacterianas) que contienen un vector de expresión intacto. Se requiere un origen de replicación (ORI) para la propagación del plásmido dentro de las células hospedadoras bacterianas. El plásmido de expresión inducida con L-arabinosa también contiene una secuencia de polinucleótidos que codifica *araC*, un regulador transcripcional que permite la inducción con L-arabinosa del promotor *araBAD* y a través de represión transcripcional reduce la expresión de fondo "con fugas" en el estado no inducido. El otro vector de expresión ("vector de expresión inducido con propionato") es similar o idéntico a pPRO o a plásmidos relacionados, en los que un promotor de expresión inducido con propionato dirige la expresión de una secuencia de expresión insertada, clonada en el sitio de clonación múltiple (MCS). El plásmido también contiene una secuencia que codifica un gen de resistencia a antibióticos (tal como el gen *cat*, que codifica la cloranfenicol acetiltransferasa, que confiere resistencia al cloranfenicol) para facilitar la selección de las células hospedadoras que contienen un vector de expresión intacto. Se requiere un origen de replicación (ORI) para la propagación del plásmido dentro de las células hospedadoras bacterianas. Además, el vector de expresión inducido con propionato contiene una secuencia de polinucleótidos que codifica *prpR*, un regulador transcripcional que permite la inducción con propionato (2-metilcitrato) del promotor *prpBCDE* y reduce la expresión de fondo "con fugas" en el estado no inducido. Para facilitar una valoración por separado de la inducción, una compatibilidad de los plásmidos y la copropagación de los

vectores de expresión, es útil que los vectores de expresión contengan promotores que respondan a diferentes inductores, orígenes de replicación compatibles y diferentes marcadores de resistencia a antibióticos. En una realización de la invención, un pBAD24 o un vector de expresión relacionado (pMB1 o 'pBR322' ORI, Amp<sup>R</sup>) que contiene un promotor *araBAD* inducible con L-arabinosa, se combina en una célula hospedadora con un pPRO33 o un vector de expresión relacionado (p15A ORI, Cm<sup>R</sup>) que contiene un promotor *prpBCDE* inducible con propionato. Los vectores de expresión se propagan conjuntamente y se mantienen utilizando medio de crecimiento complementado con ampicilina y cloranfenicol. En una realización, un vector de expresión comprende una secuencia de polinucleótidos que codifica la cadena pesada de un anticuerpo de longitud completa, y el otro vector de expresión comprende una secuencia de polinucleótidos que codifica la cadena ligera de un anticuerpo de longitud completa, estando clonada cada secuencia codificadora en marco en el MCS del vector de expresión respectivo. Para la producción de ciertos productos génicos tales como anticuerpos, la optimización de la secuencia codificadora para el organismo hospedador (incluyendo el ajuste para el sesgo de codones y contenido en GC, entre otras consideraciones) determinará las secuencias codificadoras que se van a insertar en las estructuras artificiales de expresión del sistema de coexpresión.

Haciendo referencia de nuevo a la Fig. 2, la coexpresión de productos génicos es inducida por metabolitos químicos no costosos, aplicados exógenamente, L-arabinosa y propionato (5). El nivel de inducción de la expresión de cada producto génico se valora de forma independiente con su propio inductor químico, facilitando de este modo la optimización de la coexpresión de proteínas. Esto es útil para la expresión de complejos proteicos y proteínas que requieren una pareja de unión para la estabilización, y puede facilitar la expresión de proteínas que son difíciles de expresar de otro modo, tales como las que tienen una mala solubilidad o toxicidad celular. En este ejemplo, después de la inducción, las cadenas pesadas y ligeras de anticuerpos se expresan cada una por separado, a continuación, las proteínas se unen y forman puentes disulfuro intercatenarios (dentro del citoplasma de la bacteria hospedadora), lo que permite la formación y estabilización de un anticuerpo de longitud completa, compuesto por las cadenas pesadas y ligeras. Las proteínas se pueden dirigir a distintos compartimentos del organismo hospedador. Por ejemplo, en *E. coli* la proteína se puede expresar en el citoplasma, la membrana celular, el periplasma o se puede secretar en el medio. Después de un tiempo de incubación apropiado, se recogen las células y los medios, y la proteína total extraída, que incluye los productos génicos coexpresados (6). Después de la extracción, el producto deseado se puede purificar utilizando una serie de métodos bien conocidos en la técnica, dependiendo de la naturaleza de los productos génicos que se producen en el sistema de coexpresión (por ejemplo cromatografía líquida). En el ejemplo mostrado en la Fig. 2, el producto multímero (anticuerpo de longitud completa) se extrae y se purifica utilizando métodos cromatográficos. El anticuerpo intacto purificado se visualiza sobre un gel no desnaturante usando técnicas convencionales, incluyendo colorantes que se unen a proteínas o inmunohistoquímica. El producto de anticuerpo de longitud completa se puede utilizar entonces para una serie de aplicaciones en investigación, diagnóstico u otras aplicaciones.

#### Productos preparados por los métodos de la invención

Existe una amplia versatilidad en la utilización de los sistemas de coexpresión inducible de la presente invención en numerosas aplicaciones de la coexpresión, y en las propiedades de los productos.

Glicosilación. Los productos génicos coexpresados por los métodos de la invención pueden estar glicosilados o no. En una realización de la invención, los productos génicos coexpresados son polipéptidos. Los polipéptidos glicosilados son polipéptidos que comprenden un grupo glicosilo fijado covalentemente, e incluyen polipéptidos que comprenden todos los grupos glicosilo que están fijados normalmente a determinados residuos de ese polipéptido (polipéptidos totalmente glicosilados), polipéptidos parcialmente glicosilados, polipéptidos con glicosilación en uno o varios residuos, en donde la glicosilación no se produce normalmente (glicosilación alterada), y polipéptidos glicosilados con al menos un grupo glicosilo que difiere en la estructura del grupo glicosilo que está fijado normalmente a uno o varios residuos especificados (glicosilación modificada). Un ejemplo de glicosilación modificada es la producción de polipéptidos "desfucosilados" o "carentes de fucosa", polipéptidos que carecen de restos de fucosilo en los grupos glicosilo fijados a ellos, mediante la expresión de polipéptidos en células hospedadoras que carecen de la capacidad de fucosilar polipéptidos. Los polipéptidos no glicosilados son polipéptidos que no comprenden un grupo glicosilo unido covalentemente. Un polipéptido no glicosilado puede ser el resultado de la desglicosilación de un polipéptido, o de la producción de un polipéptido aglicosilado. Los polipéptidos desglicosilados se pueden obtener mediante polipéptidos glicosilados que están desglicosilados enzimáticamente, mientras que los polipéptidos aglicosilados se pueden producir mediante la expresión de polipéptidos en células hospedadoras que no tienen la capacidad de glicosilar polipéptidos, tales como las células procariontas o células en las que la función de al menos una enzima de glicosilación se ha eliminado o reducido. En una realización particular, los polipéptidos coexpresados están aglicosilados, y en una realización más específica, los polipéptidos aglicosilados se coexpresan en células procariontas tales como *E. coli*.

Otras modificaciones de productos génicos. Los productos génicos coexpresados por los métodos de la invención se pueden unir covalentemente a otros tipos de moléculas. Ejemplos de moléculas que pueden unirse covalentemente a los productos génicos coexpresados, sin limitar el alcance de la invención, incluyen polipéptidos (tales como receptores, ligandos, citocinas, factores de crecimiento, hormonas polipeptídicas, dominios de unión a ADN, dominios de interacción con proteínas, tales como dominios PDZ, dominios de cinasa, anticuerpos y fragmentos de cualquiera de tales polipéptidos); polímeros solubles en agua (tales como polietilenglicol (PEG),

carboximetilcelulosa, dextrano, poli(alcohol vinílico), polioles polioxietilados (por ejemplo, glicerol), propionaldehído de polietilenglicol, y compuestos similares, derivados o mezclas de los mismos; y agentes citotóxicos (tales como agentes quimioterapéuticos, agentes inhibidores del crecimiento, toxinas (tales como toxinas enzimáticamente activas de origen bacteriano, fúngico, vegetal o animal, o fragmentos de las mismas), e isótopos radiactivos).

5 Además, los productos génicos que se coexpresan por los métodos de la invención pueden estar diseñados para incluir restos moleculares que ayudan en la purificación y/o la detección de los productos génicos. Muchos de estos restos son conocidos en la técnica; a modo de ejemplo, un producto génico de polipéptido se puede diseñar para incluir una secuencia "marcadora" de polihistidina - una serie de seis o más histidinas, preferiblemente de seis a diez  
10 residuos de histidina, y lo más preferiblemente seis histidinas - en su extremo N-terminal o C-terminal. La presencia de una secuencia de polihistidina en el extremo de un polipéptido permite que se una mediante medios de afinidad a base de cobalto o níquel, y se separe de otros polipéptidos. La secuencia marcadora de polihistidina se puede eliminar mediante exopeptidasas. Como otro ejemplo, las secuencias de proteínas fluorescentes se pueden expresar como parte de un producto génico del polipéptido, añadiendo la secuencia de aminoácidos de la proteína fluorescente preferiblemente en el extremo N-terminal o C-terminal de la secuencia de aminoácidos del producto  
15 génico del polipéptido. La proteína de fusión resultante emite fluorescencia cuando se expone a la luz de ciertas longitudes de onda, lo que permite que la presencia de la proteína de fusión se detecte visualmente. Una proteína fluorescente bien conocida es la proteína verde fluorescente de *Aequorea victoria*, y muchas otras proteínas fluorescentes están disponibles comercialmente, junto con las secuencias de nucleótidos que las codifican.

Anticuerpos. En una realización de la invención, los productos génicos coexpresados son anticuerpos. El término "anticuerpo" se utiliza en el sentido más amplio e incluye específicamente anticuerpos "naturales", anticuerpos completamente humanos, anticuerpos humanizados, anticuerpos quiméricos, anticuerpos multiespecíficos (tales como anticuerpos biespecíficos), anticuerpos monoclonales, anticuerpos policlonales, fragmentos de anticuerpos y otros polipéptidos obtenidos a partir de anticuerpos que son capaces de unirse a un antígeno. A menos que se indique lo contrario en el presente documento, la numeración de los residuos en una cadena pesada de inmunoglobulina ("numeración EU") es la del índice EU (la numeración EU de residuos del anticuerpo humano IgG1) como en Kabat *et al.*, *Sequences of Proteins of Immunological Interest*, Quinta edición, 1991, National Institute of Health, Bethesda, Maryland.  
20  
25

Los anticuerpos "naturales" son normalmente glicoproteínas heterotetrámeras de aproximadamente 150.000 daltons, compuestas por dos cadenas ligeras (L) idénticas y dos cadenas pesadas (H) idénticas. Cada cadena ligera está unida a una cadena pesada por un enlace disulfuro covalente, aunque el número de enlaces disulfuro intercatenarios varía entre las cadenas pesadas de diferentes isotipos de inmunoglobulinas. Cada cadena pesada y ligera también tiene puentes disulfuro intracatenarios, espaciados regularmente. Cada cadena pesada tiene en su extremo N-terminal un dominio variable ( $V_H$ ) seguido por un número de dominios constantes. Cada cadena ligera tiene un dominio variable en su extremo N-terminal ( $V_L$ ) y un dominio constante en su extremo C-terminal; el dominio constante de la cadena ligera está alineado con el primer dominio constante de la cadena pesada, y el dominio variable de la cadena ligera está alineado con el dominio variable de la cadena pesada. El término "variable" se refiere al hecho de que ciertas porciones de los dominios variables difieren ampliamente en la secuencia entre los anticuerpos y se utilizan en la unión y la especificidad de cada anticuerpo particular con un antígeno. Sin embargo, la variabilidad no se distribuye uniformemente a través de los dominios variables de los anticuerpos. Se concentra en tres segmentos llamados regiones hipervariables (HVRs), tanto en los dominios variables de la cadena ligera como de la cadena pesada. Las porciones más conservadas de los dominios variables se denominan las regiones marco (FR). Los dominios variables de las cadenas pesadas y ligeras naturales comprenden cada uno cuatro regiones FR, conectadas por tres HVRs, y con las HVRs de la otra cadena, contribuyen a la formación del sitio de unión a antígeno de los anticuerpos.  
30  
35  
40

La expresión "región Fc" se refiere a una región C-terminal de una cadena pesada de inmunoglobulina, e incluye regiones Fc naturales y regiones Fc variantes. Aunque los límites de la región Fc de una cadena pesada de inmunoglobulina pueden variar, la región Fc de la cadena pesada de una IgG humana se puede definir por extenderse desde un residuo de aminoácido en la posición Cys226, o desde Pro230, hasta el carboxilo-terminal de la misma. Alternativamente, la región Fc se puede definir por extenderse desde el residuo N-terminal (Ala231) del dominio conservado de la inmunoglobulina  $C_H2$  hasta el extremo C-terminal, y puede incluir múltiples dominios conservados, tales como  $C_H2$ ,  $C_H3$  y  $C_H4$ . La lisina C-terminal (residuo 447 según el sistema de numeración EU) de la región natural Fc se puede eliminar, por ejemplo, durante la producción o la purificación del anticuerpo, o por ingeniería recombinante del ácido nucleico que codifica una cadena pesada del anticuerpo. En consecuencia, una composición de anticuerpos intactos puede comprender poblaciones de anticuerpos con todos los residuos K447 eliminados, poblaciones de anticuerpo sin residuos K447 eliminados, y poblaciones de anticuerpo que tienen una mezcla de anticuerpos con y sin el residuo K447. La región Fc de un anticuerpo es crucial para el reclutamiento de células inmunológicas y la citotoxicidad dependiente de anticuerpos (ADCC). En particular, la naturaleza de la respuesta de ADCC provocada por anticuerpos, depende de la interacción de la región Fc con los receptores (FcRs) que se encuentran en la superficie de muchos tipos de células. Los seres humanos contienen al menos cinco clases diferentes de receptores de Fc. La unión de un anticuerpo a FcRs determina su capacidad para reclutar otras células inmunológicas y el tipo de célula reclutada. Por lo tanto, la capacidad para diseñar anticuerpos con regiones Fc alteradas que pueden reclutar solo ciertos tipos de células, puede ser de importancia decisiva para la terapia (documento de Solicitud de Patente de EE.UU. 20090136936 A1, 05-28-2009, Georgiou, George). Los anticuerpos  
45  
50  
55  
60

naturales producidos por células de mamífero comprenden normalmente un oligosacárido ramificado, biantenarico que está fijado generalmente por un enlace N con Asn297 del dominio CH2 de la región Fc. En ciertas realizaciones, los anticuerpos producidos por los métodos de la invención no están glicosilados o están aglicosilados, por ejemplo, debido a una sustitución en el residuo 297 de la región Fc, o a la expresión en una célula hospedadora que no tiene la capacidad de glicosilar polipéptidos. Debido a las respuestas de ADCC alteradas, los anticuerpos no glicosilados pueden estimular un nivel más bajo de respuestas inflamatorias, tales como la neuroinflamación. Además, dado que un anticuerpo que tiene una región Fc aglicosilada tiene una afinidad de unión muy baja hacia los receptores Fc, tales anticuerpos no se unirán a una gran cantidad de células inmunes que son portadoras de estos receptores. Esto es una ventaja significativa, ya que se reduce la unión no específica, y también se aumenta la semivida del anticuerpo *in vivo*, por lo que este atributo es muy beneficioso en la terapéutica.

Las expresiones "anticuerpo de longitud completa", "anticuerpo intacto" y "anticuerpo completo" se usan indistintamente para referirse a un anticuerpo en su forma sustancialmente intacta "natural", no a fragmentos de anticuerpo como se definen a continuación. Las expresiones se refieren especialmente a un anticuerpo con cadenas pesadas que comprenden cada una un dominio variable y una región Fc. Los "fragmentos de anticuerpo" comprenden una porción de un anticuerpo intacto, que comprende preferiblemente la región de unión a antígeno del mismo. Ejemplos de fragmentos de anticuerpos incluyen fragmentos Fab, Fab', F(ab')<sub>2</sub>, Fc, Fd y Fv; diacuerpos; anticuerpos lineales; moléculas de anticuerpo de cadena sencilla, tales como scFv; y anticuerpos multiespecíficos formados a partir de fragmentos de anticuerpos.

Un "anticuerpo humano" es uno que posee una secuencia de aminoácidos correspondiente a la de un anticuerpo producido por un ser humano. Un anticuerpo "quimérico" es uno en el que una porción de la cadena pesada y/o ligera es idéntica a, o comparte un cierto grado de identidad de secuencia de aminoácidos con secuencias correspondientes en anticuerpos obtenidos a partir de una especie particular o que pertenecen a una clase o subclase de anticuerpo particular, mientras que el resto de la o las cadenas es idéntico a, o comparte un cierto grado de identidad de secuencia de aminoácidos con, secuencias correspondientes en anticuerpos obtenidos a partir de otras especies o pertenecientes a otra clase o subclase de anticuerpo, así como fragmentos de tales anticuerpos. Un anticuerpo "humanizado" es un anticuerpo quimérico que contiene residuos de aminoácidos mínimos obtenidos a partir de moléculas de inmunoglobulina no humana. En una realización, un anticuerpo humanizado es una inmunoglobulina humana (anticuerpo receptor) en la que los residuos de HVR del anticuerpo receptor se sustituyen por residuos de una HVR de inmunoglobulina de una especie no humana (anticuerpo donante), tal como ratón, rata, conejo o primate no humano. En algunos casos, los residuos de FR del anticuerpo receptor humano se sustituyen por residuos no humanos correspondientes. Además, los anticuerpos humanizados pueden comprender residuos que no se encuentran en el anticuerpo receptor o en el anticuerpo donante. La expresión "anticuerpo monoclonal" se refiere a un anticuerpo obtenido a partir de una población de anticuerpos sustancialmente homogéneos, en donde los anticuerpos individuales que comprenden la población son idénticos excepto por posibles mutaciones, tales como mutaciones de origen natural, que pueden estar presentes en cantidades menores. De este modo, el modificador "monoclonal" indica el carácter del anticuerpo como que no es una mezcla de anticuerpos discretos. En contraste con preparaciones de anticuerpos policlonales, que normalmente incluyen diferentes anticuerpos dirigidos contra diferentes determinantes (epítomos), cada anticuerpo monoclonal de una preparación de anticuerpos monoclonales está dirigido contra el mismo determinante particular en un antígeno. Además de su especificidad, las preparaciones de anticuerpos monoclonales son ventajosas porque normalmente no están contaminadas con otras inmunoglobulinas.

La "afinidad de la unión" de una molécula, tal como un anticuerpo, se refiere en general a la fuerza de la suma total de interacciones no covalentes entre un único sitio de unión de una molécula y su pareja de unión (tal como un anticuerpo y el antígeno al que se une). A menos que se indique lo contrario, "afinidad de la unión" se refiere a la afinidad de la unión intrínseca que refleja una interacción 1:1 entre los miembros de una pareja de unión (tal como anticuerpo y antígeno). La afinidad de una molécula X hacia su pareja Y, se puede representar en general por la constante de disociación (Kd). Los anticuerpos con baja afinidad (Kd más alta), en general se unen lentamente al antígeno y tienden a disociarse con facilidad, mientras que los anticuerpos de alta afinidad (Kd más baja) se unen al antígeno generalmente más rápido y tienden a permanecer unidos durante más tiempo. Se conoce en la técnica una variedad de formas de medir la afinidad de la unión, pudiéndose utilizar cualquiera de ellas para los fines de la presente invención. Métodos ilustrativos específicos para medir la afinidad de la unión se describen en el Ejemplo 10. Los anticuerpos y los fragmentos de anticuerpos producidos y/o utilizados en los métodos de la invención, preferiblemente tienen afinidades de unión de menos de 100 nM, más preferiblemente tienen afinidades de unión de menos de 10 nM y lo más preferiblemente tienen afinidades de unión de menos de 2 nM, según se mide mediante un ensayo de resonancia de plasmón superficial, como se describe en el Ejemplo 10.

Anticuerpos (secundarios) que reconocen anticuerpos aglicosilados. La producción de anticuerpos en sistemas de expresión procariotas basados en *E. coli* u otros sin enzimas de glicosilación, generalmente produce anticuerpos aglicosilados que se pueden utilizar como anticuerpos primarios. Además de utilizar los sistemas de coexpresión inducible de la invención para producir anticuerpos primarios aglicosilados, los sistemas de coexpresión inducible de la invención también se pueden utilizar para producir de forma eficaz anticuerpos secundarios que reconocen específicamente anticuerpos primarios aglicosilados. Un aspecto de la presente invención es un sistema de anticuerpo secundario capaz de detectar un anticuerpo primario no glicosilado o aglicosilado para fines de investigación, analítica, diagnóstico o terapéuticos. A modo de ejemplo, se proporciona un sistema de anticuerpo

secundario con los siguientes componentes: epítopo, anticuerpo primario, anticuerpo secundario y sistema de detección. El epítopo es una porción de un antígeno (por lo general una proteína) que es el determinante antigénico que produce una respuesta inmunológica cuando se introduce en un animal vivo o puede ser reconocido de otra manera por un anticuerpo. En la práctica, el epítopo de interés puede estar presente dentro de una mezcla o un tejido. En una realización, el epítopo es una proteína expresada en células de carcinoma en el tejido humano. El anticuerpo primario es un fragmento de anticuerpo, un anticuerpo aislado de longitud completa (monoclonal), o una mezcla de diferentes anticuerpos de longitud completa (policlonales), que reconoce y se une al epítopo, y preferiblemente se une específicamente al epítopo. Un anticuerpo de longitud completa en este ejemplo comprende dos cadenas polipeptídicas pesadas y dos cadenas polipeptídicas ligeras, unidas por puentes disulfuro. Cada una de las cadenas comprende una región constante (Fc) y una región variable (Fv). Hay dos sitios de unión a antígeno en el anticuerpo de longitud completa. En una realización de la presente invención, el anticuerpo primario es un anticuerpo aglicosilado de longitud completa (tal como el producido en un sistema de expresión basado en *E. coli*) que reconoce y se une a un epítopo de interés. El anticuerpo secundario es un fragmento de anticuerpo, un anticuerpo aislado de longitud completa (monoclonal), o una mezcla de diferentes anticuerpos de longitud completa (policlonales), que reconoce y se une al anticuerpo primario aglicosilado, y preferiblemente se une específicamente al anticuerpo primario aglicosilado. En una realización de la presente invención, el anticuerpo secundario es un anticuerpo de longitud completa que reconoce y se une a la porción Fc aglicosilada de un anticuerpo primario de longitud completa. En este caso, los sitios de unión al anticuerpo se seleccionan y/o se modifican genéticamente para reconocer específicamente la porción Fc del anticuerpo primario aglicosilado, con o sin el residuo de lisina C-terminal. En otras realizaciones, el anticuerpo secundario se podría modificar genéticamente para reconocer regiones adicionales (epítopos) del anticuerpo primario aglicosilado, o epítopos adicionales modificados genéticamente, incluyendo pero no limitados a secuencias de polipéptidos fijadas covalentemente al anticuerpo primario. El anticuerpo secundario se puede dirigir a sitios únicos o múltiples (epítopos) presentes en moléculas aglicosiladas de anticuerpos de longitud completa (incluyendo diversas clases de inmunoglobulinas, tales como IgG, IgA, etc.) o fragmentos de anticuerpos, tales como Fab o Fc. Por lo tanto, algunos anticuerpos secundarios generados de esta forma, tendrían una amplia especificidad para cualquier anticuerpo aglicosilado de longitud completa. Los anticuerpos primarios y secundarios de la presente invención también pueden incluir los producidos por métodos tradicionales (producción de anticuerpos policlonales usando conejos inmunizados o producción de anticuerpos monoclonales usando hibridomas de ratón) y tecnología de ADN recombinante, tal como los métodos de presentación en fagos para identificar polipéptidos que se unen a antígeno.

Los sistemas de detección comprenden generalmente un agente que está ligado a o que se une al anticuerpo secundario, lo que permite una detección, visualización y/o cuantificación del anticuerpo secundario. Se conocen bien en la técnica diversos sistemas de detección, incluyendo pero no limitados a colorantes fluorescentes, enzimas, isótopos radiactivos o metales pesados. Estos pueden implicar o no una conexión física directa de polipéptidos adicionales con el anticuerpo secundario. Las aplicaciones de este sistema de anticuerpos secundarios incluyen, pero no se limitan a la inmunohistoquímica, transferencia de tipo Western y ensayo de inmunoabsorción ligado a enzimas (ELISA). Por ejemplo, en una realización para uso en inmunohistoquímica, el epítopo de interés estará presente en una sección delgada de tejido, a continuación, un anticuerpo primario aglicosilado se aplica al tejido y se permite la unión con el epítopo. El anticuerpo primario no unido se elimina, y después un anticuerpo secundario capaz de unirse específicamente con el anticuerpo primario aglicosilado, se aplica al tejido y se permite que se una con el anticuerpo primario. El anticuerpo secundario no unido se retira, y después se aplican los reactivos del sistema de detección. Por ejemplo, si el anticuerpo secundario se uniera con una enzima, entonces sustratos enzimáticos con coloración se aplicarían al tejido y se dejaría que reaccionaran. Entonces se podría realizar una visualización microscópica o fluoroscópica directa de los sustratos enzimáticos reactivos. Otros métodos de detección son bien conocidos en la técnica. Las ventajas de un sistema que utiliza anticuerpos secundarios que reconocen anticuerpos aglicosilados incluyen, sin limitación, las siguientes: 1) aumento de la especificidad en la inmunohistoquímica debido a que el anticuerpo secundario está diseñado para unirse a la porción Fc aglicosilada del anticuerpo primario que no está presente de otro modo en los tejidos eucariotas; 2) disminución de la tinción de fondo debido a una mayor especificidad hacia el anticuerpo primario; 3) disminución de los costes de producción del sistema de anticuerpos secundarios debido a que los anticuerpos primarios y/o secundarios se pueden generar en procariontes tales como *E. coli*; y 4) se evita la utilización innecesaria de mamíferos, incluyendo ratones y conejos, ya que todo el proceso de desarrollo de los anticuerpos se puede realizar en procariontes tales como *E. coli*.

Enzimas utilizadas en aplicaciones industriales. Muchos procesos industriales emplean enzimas que se pueden producir por los métodos de la invención. Estos procesos incluyen el tratamiento de aguas residuales y otros procesos de biorremediación y/o descontaminación; blanqueo de materiales en la industria papelera y textil; y degradación de la biomasa en material que se puede fermentar de manera eficaz en biocombustibles. En muchos casos, sería deseable producir enzimas para estas aplicaciones en células hospedadoras microbianas o preferiblemente en células hospedadoras bacterianas, pero la enzima activa es difícil de expresar en grandes cantidades debido a problemas con el plegamiento de la enzima y/o el requisito de un cofactor. En las siguientes realizaciones de la invención, los métodos de coexpresión inducible de la invención se usan para producir enzimas con aplicaciones industriales.

Enzimas que emplean arabinosa y xilosa. La D-xilosa es la pentosa más abundante en la biomasa vegetal, que se encuentra en polisacáridos, hemicelulosa y pectina, siendo la L-arabinosa la segunda pentosa más abundante. Para

el desarrollo y la producción de biocarburantes y otros bioproductos, es útil convertir D-xilosa y L-arabinosa en hexosas, tales como glucosa y fructosa, ya que las hexosas fermentan de manera más eficaz en biocombustibles tales como el etanol. Como se ha descrito anteriormente, el operón *araBAD* de *E. coli* codifica proteínas que metabolizan la L-arabinosa del modo siguiente: L-arabinosa con L-arabinosa isomerasa (AraA, EC 5.3.1.4) a L-ribulosa; L-ribulosa con L-ribulocinasa (AraB, EC 2.7.1.16) a L-ribulosa-fosfato; L-ribulosa-fosfato con L-ribulosa-5-fosfato 4-epimerasa (AraD, EC 5.1.3.4) a D-xilulosa-5-fosfato (también denominada D-xilulosa-5-P), que forma parte de la vía de la pentosa fosfato para la formación de fructosa y glucosa. Otra vía enzimática (una "vía oxo-reductora") que convierte la L-arabinosa en xilitol, que luego se puede convertir en D-xilulosa-5-P, es del modo siguiente: L-arabinosa con L-arabinosa/D-xilosa reductasa (EC 1.1.1.21) a L-arabinitol; L-arabinitol con L-arabinitol deshidrogenasa (EC 1.1.1.12) a L-xilulosa; y L-xilulosa con L-xilulosa reductasa (EC 1.1.1.10) a xilitol. El operón *xylAB* de *E. coli* codifica una vía enzimática (la "vía isomerasa") para la utilización de D-xilosa, del modo siguiente: D-xilosa con D-xilosa isomerasa (XylA, EC 5.3.1.5) a D-xilulosa; D-xilulosa con xilulocinasa (XylB, EC 2.7.1.17) a D-xilulosa-5-P. Otra vía enzimática (una "vía oxo-reductora") para la conversión de D-xilosa en D-xilulosa-5-P es: D-xilosa con D-xilosa reductasa (EC 1.1.1.21) a xilitol; xilitol con xilitol deshidrogenasa (EC 1.1.1.9) a D-xilulosa; D-xilulosa con xilulocinasa (XylB, EC 2.7.1.17) a D-xilulosa-5-P. Debido a los diferentes cofactores necesarios en las vías oxo-reductoras, como NADPH, NAD<sup>+</sup> y ATP, y el grado en que estos cofactores están disponibles para el uso, un desequilibrio puede dar lugar a una sobreproducción de subproducto xilitol. D-xilulosa-5-P más eritrosa 4-fosfato se pueden convertir mediante una transcetolasa (EC 2.2.1.1) en gliceraldehído 3-fosfato más fructosa 6-fosfato.

Los métodos de coexpresión inducible de la invención se pueden usar para producir enzimas que utilizan arabinosa y xilosa, que se definen como las enzimas que se mencionan con el número EC en el párrafo anterior. El número EC (o "Comisión de Enzimas", EC del inglés "Enzyme Commission") para cada enzima se estableció por la Unión Internacional de Bioquímica y Biología Molecular (IUBBM). Más información sobre estas enzimas y ejemplos específicos de las mismas se puede obtener fácilmente a través de la base de datos de proteínas UniProt ([www.uniprot.org/uniprot](http://www.uniprot.org/uniprot)) y la base de datos BRENDA, ([www.brenda-enzymes.org/index](http://www.brenda-enzymes.org/index)). En algunas realizaciones una enzima que emplea arabinosa y xilosa se coexpresa con una proteína chaperona; en realizaciones adicionales de la invención, una enzima que emplea arabinosa o xilosa se coexpresa con un transportador para un cofactor. En ciertas realizaciones, la L-arabinosa isomerasa (AraA, EC 5.3.1.4) o la L-arabinosa reductasa (EC 1.1.1.21) es producida por los métodos de la invención; en estas realizaciones, un promotor inducible con arabinosa no se utiliza en ninguna estructura artificial de expresión debido a que el inductor, L-arabinosa, se convertiría en L-ribulosa o L-arabinitol, respectivamente. Para la producción de enzimas que emplean arabinosa, distintas de L-arabinosa isomerasa (AraA, EC 5.3.1.4) o L-arabinosa reductasa (EC 1.1.1.21), un promotor inducible con arabinosa se puede utilizar si la célula hospedadora carece de EC 5.3.1.4 y/o EC 1.1.1.21, tal como un mutante *araA*, y no puede catabolizar la L-arabinosa. Del mismo modo, en realizaciones en las que la D-xilosa isomerasa (XylA, EC 5.3.1.5) o D-xilosa reductasa (EC 1.1.1.21) se produce por los métodos de la invención, un promotor inducible con xilosa no se utiliza en ninguna estructura artificial de expresión porque el inductor, D-xilosa, se convierte en D-xilulosa o xilitol, respectivamente. Para la producción de enzimas que emplean xilosa, distintas de la D-xilosa isomerasa (XylA, EC 5.3.1.5) o D-xilosa reductasa (EC 1.1.1.21), un promotor inducible con xilosa se puede utilizar si la célula hospedadora carece de EC 5.3.1.5 y/o EC 1.1.1.21, tal como un mutante de *xylA*, y no puede catabolizar la D-xilosa.

La xilosa isomerasa (XylA, EC 5.3.1.5) es una enzima que se encuentra en microorganismos, hongos anaerobios y plantas, que cataliza la interconversión de un azúcar aldo (D-xilosa) a un azúcar ceto (D-xilulosa). También puede isomerizar D-ribosa a D-ribulosa y D-glucosa a D-fructosa. Esta enzima pertenece a la familia de las isomerasas, específicamente a las oxidorreductasas intramoleculares que interconvierten aldosas y cetosas. El nombre sistemático de esta clase de enzimas es D-xilosa aldosa-cetosa-isomerasa. Otros nombres de uso común incluyen D-xilosa isomerasa, D-xilosa cetoisomerasa, D-xilosa cetol-isomerasa y glucosa isomerasa. La enzima se utiliza industrialmente para convertir glucosa en fructosa en la fabricación de jarabe de maíz con alto contenido en fructosa y, como se ha descrito anteriormente, se puede utilizar en la conversión de pentosas a hexosas para la producción de biocombustible. La xilosa isomerasa es un homotetrámero y requiere dos cationes divalentes - Mg<sup>2+</sup>, Mn<sup>2+</sup> y/o Co<sup>2+</sup> - para una actividad máxima. La actividad xilosa isomerasa se puede medir usando un ensayo con arabitol deshidrogenasa ligado a NADH (Smith *et al.*, "D-Xylose (D-glucose) isomerase from *Arthrobacter* strain N.R.R.L. B3728. Purification and properties", *Biochem J* 1991 Jul 1; 277 (Pt 1): 255-261), en el que una unidad de actividad xilosa isomerasa es la cantidad de enzima que convierte 1 micromol de D-xilosa en D-xilulosa en un minuto. En al menos algunas especies, la xilosa isomerasa requiere magnesio (o manganeso en el caso de plantas) para su actividad, aunque el cobalto puede ser necesario para estabilizar la estructura tetrámera de la enzima. Cada subunidad de xilosa isomerasa contiene un barril alfa/beta plegado de forma similar a la de otras enzimas TIM divalentes con barril (triosafosfato isomerasa) que dependen de metal, y la parte C-terminal más pequeña forma un plegamiento helicoidal extendido, implicado en la multimerización. Los residuos conservados en todas las xilosa isomerasas conocidas son una histidina en la sección N-terminal de la enzima, que se ha mostrado que está involucrada en el mecanismo catalítico de la enzima, y dos residuos de glutamato, una histidina y cuatro residuos de aspartato que forman los dos sitios de unión a metal, cada uno de los cuales se une a un ion de magnesio, cobalto o manganeso (Katz *et al.*, "Locating active-site hydrogen atoms in D-xylose isomerase: time-of-flight neutron diffraction", *Proc Natl Acad Sci USA* 2006 May 30; 103(22): 8342-8347; Epub 2006 May 17).

En algunas realizaciones de la invención, la coexpresión inducible se utiliza para expresar una proteína xilosa isomerasa; en ciertas realizaciones, la xilosa isomerasa ("XI") se selecciona a partir del grupo que consiste en: XI de

5 *Arthrobacter* sp. de la cepa NRRL B3728 (UniProt P12070); XI de *Bacteroides stercoris* (UniProt B0NPH3); XI de *Bifidobacterium longum* (UniProt Q8G3Q1); XI de *Burkholderia cenocepacia* (UniProt Q1BG90); XI de *Ciona intestinalis* (UniProt F6WBF5); XI de *Clostridium phytofermentans* (UniProt A9KN98); XI ukk1 de *Orpinomyces* sp. (UniProt B7SLY1); XI E2 de *Piromyces* sp. (UniProt Q9P8C9); XI de *Streptomyces lividans* (UniProt Q9RFM4); XI TK24 de *Streptomyces lividans* (UniProt D6ESI7); XI JW 200 de *Thermoanaerobacter ethanolicus* (UniProt D2DK62); XI de *Thermoanaerobacter yonsei* (UniProt Q9KGU2); XI de *Thermotoga neapolitana* (UniProt P45687); XI de *Thermus thermophilus* (UniProt P26997); y XI de *Vibrio* sp. cepa XY-214 (UniProt C7G532). En realizaciones particulares, una xilosa isomerasa se coexpresa de manera inducible con un transportador de iones divalentes tal como CorA (UniProt P0ABI4); emplear un promotor inducible para controlar la coordinación y la duración del transporte de iones, puede ser útil para reducir la toxicidad en las células hospedadoras procedente de iones metálicos tales como  $\text{Co}^{2+}$ . En realizaciones adicionales, se introducen mutaciones en proteínas xilosa isomerasas que afectan a la interacción entre parejas de monómeros de XI; por ejemplo, la introducción de residuos de cisteína de modo que se pueden formar enlaces disulfuro entre una pareja de monómeros (véase, Varsani *et al.*, "Arthrobacter D-xylose isomerase: protein-engineered subunit interfaces", *Biochem J* 1993 Apr 15; 291 (Pt 2): 575-583). En este ejemplo, debido a que los residuos de cisteína se introducen en lugares recíprocos pero no idénticos dentro de los monómeros, se producen dos tipos diferentes de monómeros alterados, y los sistemas de coexpresión inducible de la invención se utilizan para valorar la expresión relativa de los dos tipos de monómeros para lograr la relación estequiométrica deseada.

20 Peroxidasas degradantes de lignina. Las peroxidases son un subgrupo de las oxidoreductasas y se usan para catalizar una variedad de procesos industriales. Las oxidoreductasas pueden descomponer la lignina o actuar como reductasas en la degradación de la celulosa y la hemicelulosa, permitiendo a las enzimas utilizadas en el procesamiento de la biomasa vegetal, un acceso más sencillo a los residuos de sacáridos durante la producción de biocombustibles, tales como el etanol. Las oxidoreductasas pueden ser oxidasas o deshidrogenasas. Las oxidasas utilizan oxígeno molecular como un aceptor de electrones, mientras que las deshidrogenasas oxidan un sustrato mediante la transferencia de un grupo  $\text{H}^-$  a un aceptor, tal como NAD/NADP<sup>+</sup> o una enzima dependiente de flavina. Las peroxidases catalizan la reducción de un peróxido, tal como peróxido de hidrógeno ( $\text{H}_2\text{O}_2$ ). Otros tipos de oxidoreductasas incluyen oxigenasas, hidroxilasas y reductasas. Además del oxígeno, el dinucleótido de flavina adenina (FAD) y los dinucleótidos de nicotinamida adenina NAD y NADP, cofactores potenciales de oxidoreductasas incluyen citocromos y grupos hemo, disulfuro y proteínas de hierro-azufre.

30 Las peroxidases degradantes de lignina, pueden oxidar una variedad de compuestos aromáticos, incluyendo compuestos de alto potencial redox como la lignina, colorantes industriales, plaguicidas, etc. Se han identificado y caracterizado cuatro tipos de enzimas que modifican la lignina: lignina peroxidasa (LiP, EC 1.11.1.14), manganeso peroxidasa (MnP, EC 1.11.1.13), peroxidasa versátil (VP, EC 1.11.1.16) y lacasa (EC 1.10.3.2). LiP, MnP y VP son hemoproteínas con cuatro o cinco enlaces disulfuro y dos sitios de unión para iones  $\text{Ca}^{2+}$  estructurales, y son peroxidases de alto potencial redox que pueden oxidar directamente sustratos de alto potencial redox y/o  $\text{Mn}^{2+}$ . La actividad peroxidasa de MnP se puede medir usando el ensayo de oxidación de 2,6-dimetoxifenol (2,6-DMP) descrito en el Ejemplo 4, Sección E, a continuación; la actividad de LiP se puede medir mediante la oxidación de alcohol veratrílico a aldehído veratrílico en presencia de  $\text{H}_2\text{O}_2$  (Orth *et al.*, "Overproduction of lignin-degrading enzymes by an isolate of *Phanerochaete chrysosporium*", *Appl Environ Microbiol* 1991 Sep; 57(9): 2591-2596). Las lacasas no están asociadas con grupos hemo, pero están asociadas con iones de cobre (generalmente 4 iones de cobre por proteína de lacasa), y son oxidoreductasas de bajo potencial redox, que solo pueden oxidar sustratos de alto potencial redox en presencia de mediadores redox. La actividad de las lacasas se puede medir usando el ensayo de oxidación ABTS (2,2'-azinobis-(ácido 3-etilbenzotiazolin-6-sulfónico)) como se describe en Zhao *et al.*, "Characterisation of a novel white laccase from the deuteromycete fungus *Myrothecium verrucaria* NF-05 and its decolourisation of dyes", *PLoS One* 2012; 7(6): e38817; Epub 2012 Jun 8; una unidad de actividad se define como la producción de 1 micromol de producto por minuto.

50 Los métodos de coexpresión inducible de la invención se pueden utilizar para producir peroxidases degradantes de lignina, que se definen como las enzimas que se mencionan por el número de EC en el párrafo anterior. En algunas realizaciones, una peroxidasa degradante de lignina se coexpresa con una proteína chaperona; en realizaciones adicionales de la invención, una peroxidasa degradante de lignina se coexpresa con un transportador para un cofactor, tal como un grupo hemo.

55 LiP se puede expresar utilizando los sistemas de coexpresión inducible de la invención; en ciertas realizaciones, la LiP se selecciona a partir del grupo que consiste en: isoenzimas de LiP de *Phanerochaete chrysosporium* (*Sporotrichum pruinosum*) "ligninasa A" (UniProt P31837), "ligninasa B" (UniProt P31838), "ligninasa H2" (UniProt P11542), "ligninasa H8" (UniProt P06181), "ligninasa LG2" (UniProt P49012), "ligninasa LG3" (UniProt P21764), "ligninasa LG5" (UniProt P11543) y "ligninasa LG6" (UniProt P50622); "ligninasa-3" de *Phlebia radiata* (UniProt P20010); isoenzimas de LiP de *Trametes versicolor* (*Coriolus versicolor*) "ligninasa C" (UniProt P20013), LP7 (UniProt Q99057) y LP12 (UniProt Q7LHY3); y LiP de *Trametesopsis cervina* (UniProt Q3C1R8) y (UniProt Q60FD2).

60 En algunas realizaciones de la invención, la coexpresión inducible se utiliza para expresar MnP; en ciertas realizaciones, la MnP se selecciona a partir del grupo que consiste en: MnP de *Agaricus bisporus* (UniProt Q5TJC2); MnP de *Agrocybe praecox* (UniProt G4WG41); MnP de *Ganoderma lucidum* (UniProt C0IMT8); MnP de *Lenzites gibbosa* (UniProt C3V8Q9); isoenzimas de MnP de *Phanerochaete chrysosporium* MnP1 (UniProt Q02567); H3

(UniProt P78733); y H4 (UniProt P19136); isoenzimas de MnP de *Phlebia radiata* MnP2 (UniProt Q70LM3) y MnP3 (UniProt Q96TS6); MnP b19 de *Phlebia* sp. (UniProt B2BF37); isoenzimas de MnP MG60 de *Phlebia* sp. MnP1, MnP2, MnP3 (UniProt B1B554, B1B555 y B1B556, respectivamente); MnP3 de *Pleurotus ostreatus* (UniProt B9VR21); MnP5 de *Pleurotus pulmonarius* (UniProt Q2VT17); MnP FERM P-18171 de *Spongipellis* sp. (UniProt Q2HWK0); e isoenzimas de MnP de *Trametes versicolor* (*Coriolus versicolor*) (UniProt Q99058, Q6B6M9, Q6B6N0, Q6B6N1 y Q6B6N2). La coexpresión de MnP con la proteína disulfuro isomerasa en presencia de hemo, se describe en el Ejemplo 4.

La lacasa se puede expresar utilizando los sistemas de coexpresión inducible de la invención; en ciertas realizaciones, la lacasa se selecciona entre el grupo que consiste en: isoenzimas de lacasa de *Botryotinia fuckeliana* (*Botrytis cinerea*) (UniProt Q12570, Q96UM2 y Q96WM9); isoenzimas de lacasa WR1 de *Cerrena* sp. (UniProt E7BLQ8, E7BLQ9 y E7BLR0); lacasa de *Cerrena unicolor* (*Daedalea unicolor*) (UniProt B8YQ97); isoenzimas de lacasa de *Ganoderma lucidum* (UniProt Q6RYA2, B5G547, B5G549, B5G550, B5G551 y B5G552); lacasa de *Melanocarpus albomyces* (UniProt Q70KY3); isoenzimas de lacasa de *Pleurotus ostreatus* (UniProt Q12729, Q12739, Q6RYA4, Q6RYA4 y G3FGX5); isoenzimas de lacasa de *Pycnoporus cinnabarinus* (UniProt 059896, Q9UVQ2, D2CSG0, D2CSG1, D2CSG4, D2CSG6 y D2CSG7); isoenzimas de lacasa de *Pycnoporus coccineus* (UniProt D2CSF2, D2CSF5, D2CSF6, D2CSM7 y D7F484); lacasa de *Trametes hirsuta* (*Coriolus hirsutus*) (UniProt Q02497); lacasa de *Trametes maxima* (*Cerrena maxima*) (UniProt D0VWU3); isoenzimas de lacasa de *Trametes versicolor* (*Coriolus versicolor*) (UniProt Q12717, Q12718 y Q12719); e isoenzimas de lacasa de *Trametes villosa* (UniProt Q99044, Q99046, Q99049, Q99055 y Q99056).

Aunque las tres principales enzimas modificadoras de lignina de hongos de pudrición blanca son LiP, MnP y lacasa, otro tipo de peroxidasa, la peroxidasa versátil (VP), se encuentra en varias especies de los géneros *Pleurotus* y *Bjerkandera*. VP es de interés debido a su versatilidad catalítica: puede oxidar sustratos de LiP, alcohol veratrílico, metoxi-bencenos y compuestos modelo de lignina no fenólicos, así como el sustrato de MnP Mn<sup>2+</sup>. La actividad versátil de VP se puede someter a ensayo usando el ensayo con 2,6-DMP de MnP, o el ensayo con alcohol veratrílico de LiP, o el ensayo con ABTS de lacasa (véase más arriba). En algunas realizaciones de la invención, la coexpresión inducible se utiliza para expresar VP; en ciertas realizaciones, la VP se selecciona a partir del grupo que consiste en: VP de *Bjerkandera adusta* (UniProt A5JTV4); isoenzimas de VP de *Pleurotus eryngii* (UniProt 094753, Q9UR19 y Q9UVP6); y VP de *Pleurotus pulmonarius* (UniProt I6TLM2).

#### EJEMPLO 1

Coexpresión inducible de cadenas pesadas y ligeras de IgG1 para producir anticuerpos de longitud completa en células bacterianas

##### A. Construcción de los vectores de expresión

El sistema de coexpresión inducible se utilizó para producir anticuerpos de longitud completa, específicamente anticuerpos de IgG1 de ratón anti-CD19 humano, en células bacterianas. La secuencia que codifica la cadena pesada de IgG1 de ratón anti-CD19 humano ("cadena pesada de IgG1", "IgG1HC", "cadena pesada" o "HC") se proporciona como SEQ ID NO: 1 y es la misma que la de GenBank nº de orden AJ555622.1, y específicamente las bases 13 hasta 1407 de la secuencia de nucleótidos de GenBank AJ555622.1. La secuencia de aminoácidos correspondiente de la cadena pesada de IgG1 de ratón anti-CD19 humano de longitud completa, se proporciona como SEQ ID NO: 2 (y es la misma que la de GenBank nº de orden CAD88275.1). La secuencia que codifica la cadena ligera de IgG1 de ratón anti-CD19 humano ("cadena ligera de IgG1", "IgG1LC", "cadena ligera" o "LC") se proporciona como SEQ ID NO: 3 y es la misma que la de GenBank nº de orden AJ555479.1, y específicamente las bases 23 a 742 de la secuencia de nucleótidos de GenBank AJ555479.1. La secuencia de aminoácidos correspondiente de la cadena ligera de IgG1 de ratón anti-CD19 humano de longitud completa, se proporciona como SEQ ID NO: 4 (y es la misma que la de GenBank nº de orden CAD88204.1).

La síntesis de polinucleótidos que codifican las cadenas pesadas y ligeras de IgG1, y la optimización de la expresión en *E. coli*, se llevó a cabo mediante GenScript (Piscataway, NJ). El sistema de diseño de genes OptimumGene® utiliza un algoritmo, el algoritmo OptimumGene®, que tiene en cuenta una variedad de factores involucrados en diferentes etapas de la expresión de proteínas, tales como el sesgo de uso de codones, el contenido en GC, el contenido en dinucleótido CpG, la estructura prevista de ARNm y varios elementos en *cis* en la transcripción y la traducción. Las secuencias optimizadas para las cadenas pesadas y ligeras de IgG1 se proporcionan como SEQ ID NO: 5 y SEQ ID NO: 6, respectivamente; estas secuencias optimizadas incluyen algunos nucleótidos adicionales aguas arriba y aguas abajo de la secuencia codificante; las secuencias codificadoras son las bases 25 hasta 1419 de SEQ ID NO: 5, y las bases 25 hasta 747 de SEQ ID NO: 6. La secuencia codificante optimizada en SEQ ID NO: 10 para la cadena ligera de IgG1, codifica un aminoácido adicional, un residuo de alanina, en la posición 2 de la secuencia de aminoácidos codificada, en relación con la secuencia de aminoácidos de la cadena ligera de IgG1, SEQ ID NO: 4. Las secuencias optimizadas de la cadena pesada y ligera, contienen ambas codones de parada TAA y una secuencia preferida de unión al ribosoma de *E. coli* (AGGAGG), situada a -14 a -9 bases aguas arriba del codón de iniciación. Además, la secuencia optimizada de la cadena pesada de IgG1 (SEQ ID NO: 5) contiene un sitio de restricción de NcoI que comprende el codón de iniciación ATG, y un sitio de restricción de HindIII inmediatamente aguas abajo del codón de parada TAA, mientras que la secuencia optimizada de la cadena ligera de

IgG1 (SEQ ID NO: 6) contiene un sitio de restricción de NheI inmediatamente aguas arriba de la secuencia de unión al ribosoma, y también un sitio de restricción de HindIII inmediatamente aguas abajo del codón de parada TAA.

Las secuencias codificadoras optimizadas para las cadenas pesadas y ligeras de IgG1 se obtuvieron a partir de GenScript como insertos de polinucleótidos clonados en pUC57. El vector pBAD24 se obtuvo de la American Type Culture Collection (ATCC) como ATCC 87399, y tiene la secuencia de nucleótidos mostrada en la base de datos GenBank n° de orden X81837.1 (25-OCT-1995); el vector pPRO33 se obtuvo de la Universidad de California (Berkeley, California), y tiene la secuencia de nucleótidos mostrada en SEQ ID NO: 7. La secuencia de nucleótidos de pPRO33 se compiló a partir de las secuencias del vector pBAD18 (GenBank n° de orden X81838.1), la secuencia genómica de *E. coli* de la región prpR-<sub>JP</sub>, y el vector pBAD33, como se describe en Guzman *et al.*, "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter", J Bacteriol 1995 Jul; 177(14): 4121-4130, y en el documento de patente de EE.UU. n° 8.178.338 B2; 15 de mayo de 2012; Keasling, Jay. Para clonar la cadena pesada de IgG1 en el vector pBAD24, y la cadena ligera de IgG1 en el vector pPRO33, las estructuras artificiales pUC57-HC y pUC57-LC se transformaron primero en células de *E. coli* BL21 (New England Biolabs (o 'NEB'), Ipswich, Massachusetts) usando el método de choque térmico. A continuación, el ADN del plásmido para cada uno de estos vectores se obtuvo por un método de miniprep. A menos que se indique lo contrario, el crecimiento de las células de *E. coli* en cultivo líquido era a 37°C con agitación rotatoria a 250 rpm. Cepas de células de *E. coli* que contenían plásmidos (BL21 que contenía pUC57-HC, BL21 que contenía pUC57-LC y DH5alpha que contenía pBAD24) se cultivaron cada una en 8 ml de medio LB + ampicilina ("AMP"), y las células de *E. coli* DH10B (Life Technologies, Grand Island, Nueva York) que contenían pPRO33 se cultivaron en 5 ml de medio LB + cloranfenicol ("CAM"), durante 14 horas o durante una noche, a continuación, las células se sedimentaron, se lisaron y el ADN del plásmido se separó por medio de una columna de centrifugación QIAprep® (QIAGEN, Germantown, MD) de acuerdo con el protocolo del fabricante. Para mayores rendimientos, el protocolo de columna de centrifugación QIAprep® se realizó con 100% de etanol añadido al tampón PE, y el material sobrenadante se pasó a través de la columna dos veces, en lugar de una vez.

Para clonar la cadena pesada de IgG1 en pBAD24, los plásmidos purificados pUC57-HC y pBAD24 se digirieron con las enzimas de restricción NcoI y HindIII (NEB), entonces el fragmento NcoI-HindIII de IgG1HC y el plásmido pBAD24 cortado con NcoI-HindIII se separaron de otros fragmentos de polinucleótidos mediante electroforesis en gel, se escindieron del gel y se purificaron usando el kit de purificación de bandas ilustra GFX Gel (GE Healthcare Life Sciences, Piscataway, NJ) según las instrucciones del fabricante. A continuación, se ligó el fragmento NcoI-HindIII de IgG1HC (durante una noche a 16°C) con el plásmido pBAD24 cortado con NcoI-HindIII para formar la estructura artificial de expresión pBAD24-HC. Debido a que el sitio de restricción de NcoI en el fragmento NcoI-HindIII de IgG1HC comprende el codón de iniciación ATG de la secuencia que codifica IgG1HC, cuando se inserta el fragmento NcoI-HindIII de IgG1HC en el vector pBAD24 cortado con NcoI-HindIII, la secuencia que codifica IgG1HC se coloca aguas abajo de una secuencia preferida de *E. coli* de unión al ribosoma, AGGAGG, presente en el vector pBAD24 (véase la Figura 1 de Guzman *et al.*, "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter", J Bacteriol 1995 Jul; 177(14): 4121-4130). En la estructura artificial de expresión resultante pBAD24-HC, la secuencia de unión al ribosoma se encuentra localizada desde -14 a -9 a bases aguas arriba del codón de iniciación de IgG1HC.

Para clonar la cadena ligera de IgG1 en pPRO33, los plásmidos pUC57-LC y pPRO33 purificados se digirieron con las enzimas de restricción NheI y HindIII (NEB). Debido a que el plásmido pPRO33 tiene dos sitios de restricción de HindIII, había dos fragmentos que se tenían que purificar en gel procedentes de la digestión de pPRO33 con NheI-HindIII: un fragmento de 4,4 kb y un fragmento de 1,5 kb. Después de la purificación en gel de todos los fragmentos deseados, utilizando un kit de purificación de bandas ilustra GFX Gel (véase más arriba), el fragmento de pPRO33 de 4,4 kb purificado se trató con fosfatasa alcalina (también denominada fosfatasa intestinal de ternera o CIP) (NEB), de acuerdo con las instrucciones del fabricante. El fragmento de 4,4 kb de pPRO33 tratado con CIP, se purificó de la mezcla de reacción de fosfatasa utilizando un kit de purificación ilustra GFX (véase más arriba), y después se ligó al fragmento de 1,5 kb de pPRO33 (a 16°C durante 10,5 horas). El plásmido resultante pPRO33 cortado con NheI-HindIII se ligó al fragmento NheI-HindIII de IgG1LC (durante una noche a 16°C) para formar la estructura artificial de expresión pPRO33-LC. Debido a que el sitio de restricción de NheI en el fragmento NheI-HindIII de IgG1LC se encuentra inmediatamente aguas arriba de la secuencia de unión al ribosoma AGGAGG en la secuencia optimizada de la cadena ligera (SEQ ID NO: 6), cuando se inserta el fragmento NheI-HindIII de IgG1LC en el vector pPRO33 cortado con NheI-HindIII, la secuencia de IgG1LC conserva su secuencia preferida de unión al ribosoma de *E. coli*.

#### B. Coexpresión inducible de las cadenas pesada y ligera de IgG1 en células bacterianas

[Ejemplo Comparativo]

La estructura artificial de expresión pBAD24-HC y la estructura artificial de expresión pPRO33-LC se co-transformaron en células de *E. coli* BL21 y en células de *E. coli* SHuffle® Express (NEB), utilizando el método de choque térmico. Las células transformadas BL21 (pBAD24-HC/pPRO33-LC) y SHuffle® Express (pBAD24-HC/pPRO33-LC) se cultivaron a 37°C durante una noche en 5 ml de medio LB con 30 microgramos/mL de CAM y 100 microgramos/mL de AMP. Debido a que las células SHuffle® Express parecen crecer más lentamente que las células BL21, 2% (100 microlitros) del cultivo de una noche de SHuffle® Express (pBAD24-HC/pPRO33-LC) y 1% (50

microlitros) del cultivo de una noche de BL21 (pBAD24-HC/pPRO33-LC), se utilizaron para inocular 5 ml de medio mínimo M9 con casaminoácidos y 0,2% de glicerol, más 30 microgramos/mL de CAM y 100 microgramos/mL de AMP. Estos cultivos se dejaron crecer hasta que la DO<sub>600</sub> era de aproximadamente 0,5: 3,1 horas para las células SHuffle<sup>®</sup> Express (pBAD24-HC/pPRO33-LC) y 2,5 horas para las células BL21 (pBAD24-HC/pPRO33-LC). Antes de la inducción, se tomaron muestras de control de cada cultivo que había crecido sin inducción. El resto de cultivos celulares de SHuffle<sup>®</sup> Express (pBAD24-HC/pPRO33-LC) y BL21 (pBAD24-HC/pPRO33-LC) se indujo entonces con la adición de 0,2% de arabinosa y propionato 50 mM, y crecieron en paralelo con las muestras de control no inducidas durante 6 horas. Al final de ese tiempo, se centrifugaron todos los cultivos celulares para sedimentar las células, y se colocaron en un congelador a -80°C. Los sedimentos celulares se descongelaron sobre hielo, y las células se lisaron usando un kit de lisis bacteriana Qproteome (QIAGEN), según las instrucciones del fabricante, con las excepciones de que se añadió un comprimido de mezcla de inhibidores de proteasas Complete Mini (Roche, Indianapolis, Indiana) a 10 mL de tampón de lisis natural, antes de añadir la lisozima y la nucleasa, y que las muestras se centrifugaron a 25°C en lugar de a 4°C.

Los extractos de proteínas solubles procedentes de las células inducidas y los controles no inducidos se separaron por electroforesis en gel de SDS en condiciones reductoras, en un gel de Bis-Tris al 4-12% de NuPAGE<sup>®</sup> (Life Technologies, Grand Island, Nueva York). El gel se tiñó usando RAPIDstain<sup>®</sup> (G-Biosciences, St. Louis, Missouri). Como se muestra en la Figura 3, se observó que una banda de proteína (cadena pesada) migraba a 51 kDa y otra banda de proteína (cadena ligera) a 26 kDa; estas bandas están presentes en las células inducidas, pero no en las células no inducidas. Los mismos extractos de proteínas solubles a partir de células SHuffle<sup>®</sup> Express y BL21 inducidas y no inducidas que contenían tanto los vectores de expresión inducibles pBAD24-HC como pPRO33-LC, se separaron por electroforesis en gel en condiciones naturales (no reductoras) sobre un gel NOVEX<sup>®</sup> de Tris-Glicina al 10-20% (Life Technologies). El gel se tiñó usando RAPIDstain<sup>®</sup> (G-Biosciences). Como se muestra en la Figura 4, una banda de proteína (anticuerpo IgG1 que comprendía las cadenas pesadas y ligeras) migra a 154 kDa; esta banda está presente en las células SHuffle<sup>®</sup> Express inducidas, pero se reduce significativamente o está ausente en las células BL21 inducidas y en las células no inducidas.

## EJEMPLO 2

Caracterización de las estructuras artificiales de expresión y anticuerpos IgG1 de longitud completa producidos en células bacterianas

### A. Caracterización de los vectores de expresión

La secuencia de las estructuras artificiales de expresión pBAD24-HC y pPRO33-LC se confirma utilizando los cebadores mostrados en la tabla siguiente, para iniciar las reacciones de secuenciación de terminación de cadena didesoxi, junto con otros cebadores diseñados según sea necesario. La secuencia de nucleótidos del promotor *ppBCDE* y la región del vector pPRO24 aguas arriba del MCS, como se muestra en la Figura 1C del documento de Patente de EE.UU. n° 8.178.338 B2, se utiliza para diseñar al menos un cebador directo de oligonucleótidos para secuenciar secuencias clonadas en el MCS de los vectores de expresión pPRO.

Tabla 4. Cebadores de oligonucleótidos

Cebador	SEQ ID NO:	Secuencia	Comentarios
IgG1HC Fc directo	8	5'- TTC ACC ATG GAA GTT TCA TCG GTC TTT ATT TTC CCG -3'	Bases apareadas 780-807 de SEQ ID NO: 5, añadidas al sitio NcoI en el extremo 5'
IgG1HC Fc inverso	9	5'- AGC CAA GCT TTT ATT TAC CCG GCG AGT GGG AC -3'	Emparejamiento en la orientación inversa con las bases 1398-1429 de SEQ ID NO: 6
pBAD24/33 directo	10	5'- CTG TTT CTC CAT ACC CGT T - 3'	Localizado entre el promotor pBAD y el MCS en todos los vectores pBAD
pBAD24/33 inverso n° 1	11	5'- CTC ATC CGC CAA AAC AG -3'	Justo aguas abajo de MCS en los vectores pBAD y pPRO, separado del sitio HindIII de MCS por 2 bases (GG)
pBAD24/33	12	5'- GGC TGA AAA TCT TCT CT - 3'	Aguas abajo del cebador inverso n° 1 en los

inverso nº 2			vectores pBAD y pPRO, y se solapa con el mismo, las dos últimas bases son las dos primeras del nº 1
--------------	--	--	---

#### B. Detección de los anticuerpos de longitud completa de IgG1 de ratón sobre una transferencia Western

Los geles de proteínas utilizados para separar los extractos de proteínas solubles mediante electroforesis (geles Bis-Tris al 4-12% de NuPAGE<sup>®</sup> o geles Tris-Glicina al 10-20% de NOVEX<sup>®</sup>), como se describen en el Ejemplo 1, se colocan en un módulo de transferencia XCell II<sup>®</sup> (Life Technologies, Grand Island, Nueva York). Se aplica corriente al gel de acuerdo con las instrucciones del fabricante, lo que da lugar a la transferencia de proteínas desde el gel a una membrana de nitrocelulosa. La membrana de nitrocelulosa se incuba entonces con un anticuerpo primario, IgG anti-ratón, a 4°C durante una noche. La membrana de nitrocelulosa se lava después para eliminar el anticuerpo no unido, y se incuba con un anticuerpo secundario, IgG de cabra anti-ratón conjugado con fosfatasa alcalina, durante una hora a temperatura ambiente. La membrana de nitrocelulosa se lava entonces para eliminar el anticuerpo secundario no unido, se incuba con una solución que contiene tetrazolio de nitroazul (NBT) y 5-bromo-4-cloro-indolil-fosfato (BCIP) para teñir la o las bandas de proteínas, y después se lava para eliminar el exceso de solución de tinción.

#### EJEMPLO 3

##### Introducción de alteraciones genómicas en células hospedadoras para facilitar la coexpresión

Como se ha descrito anteriormente, ciertos cambios en la expresión génica de la célula hospedadora pueden mejorar la coexpresión del o de los productos génicos deseados. Las siguientes delecciones y alteraciones se realizaron en el genoma de la célula hospedadora de *E. coli* SHuffle<sup>®</sup> Express de Gene Bridges GmbH (Heidelberg, Alemania), utilizando un método de modificación por recombinación, descrito como una delección por contraselección, que elimina perfectamente secuencias genómicas. Una delección del operón *araBAD* de la célula hospedadora se realizó para reducir el catabolismo de la arabinosa a través de la célula hospedadora, de manera que estaba disponible más inductor arabinosa para la inducción de un producto génico coexpresado procedente de una estructura artificial de expresión que comprendía el promotor *araBAD*. Esta delección elimina 4269 pares de bases del operón *araBAD*, correspondientes a la posición 70.135 hasta 65.867 (de la cadena negativa) del genoma de *E. coli* (las posiciones dentro de las secuencias de nucleótidos genómicas se proporcionan como en la Tabla 1), de modo que la mayor parte del promotor natural *araBAD* a lo largo de casi todos los codones de la región que codifica AraD, se elimina. La secuencia de nucleótidos (cadena negativa) alrededor de la unión en la delección (posición 70.136 | posición 65.866) es: TTAT | TACG. Otra delección se realizó dentro del operón *sbm-ygfDGH* (también llamado *scpA-argK-scpBC*), eliminando la función de los genes implicados en la biosíntesis de 2-metilcitrato, para aumentar la sensibilidad del promotor inducible con propionato de las células hospedadoras, hacia el propionato suministrado exógenamente. La delección *sbm-ygfDGH* elimina 5542 pares de bases (posición 3.058.754 hasta 3.064.295 del genoma de *E. coli*), extrayendo el promotor *sbm-ygfDGH* y todo el operón excepto el último codón de la secuencia que codifica *ygfH*, dejando la secuencia adyacente que codifica *ygfI* y el codón de parada intacto. La secuencia de nucleótidos (cadena positiva) alrededor de la unión en la delección (posición 3058753 | posición 3064296) es: ACAA | GGGT. Además de estas delecciones realizadas en el genoma de la célula hospedadora de *E. coli* SHuffle<sup>®</sup> Express, Gene Bridges GmbH introdujo una mutación puntual en la secuencia genómica que codificaba el gen *rpsL*, que se extiende en la hebra negativa desde la posición 3.472.574 hasta 3.472.200, cambiando la A en la posición 3.472.447 por una G, alterando el codón de Lys43 a un codón de Arg, lo que da como resultado un fenotipo resistente a la estreptomycinina cuando se expresa el gen mutante *rpsL*-Arg43. Otra alteración del genoma de la célula hospedadora, que permite una expresión inducible controlada más estrechamente como se ha descrito anteriormente, es hacer que el promotor *araE* sea constitutivo más que sensible a la arabinosa. La mayor parte del promotor *araE* natural, incluyendo los sitios de unión a CRP-cAMP y AraC, se eliminó mediante la delección de 97 pares de bases (posición 2.980.335 hasta 2.980.239 (cadena negativa)) y la sustitución de esa secuencia por la secuencia de 35 pares de bases del promotor J23104 constitutivo, con las secuencias de sitios de unión resultantes: TGAA | TTGA ← promotor J23104 → TAGC | TTCA. Una célula hospedadora de *E. coli*, tal como una célula hospedadora de *E. coli* SHuffle<sup>®</sup> Express, con cualquiera de estas alteraciones genómicas, o cualquier combinación de las mismas, se puede emplear en la coexpresión inducible de productos génicos.

#### EJEMPLO 4

##### Coexpresión inducible de manganeso peroxidasa y proteína disulfuro isomerasa en presencia de grupos hemo

#### 50 A. Construcción de vectores de expresión

La manganeso peroxidasa ("MnP"; también denominada peroxidasa dependiente de manganeso) es una enzima que permite que algunos tipos de hongos degraden la lignina a dióxido de carbono, y median en la oxidación de una amplia variedad de contaminantes orgánicos. Un ejemplo de manganeso peroxidasa es la isoenzima H4 (denominada en este documento "MnP-H4") del hongo de pudrición blanca *Phanerochaete chrysosporium*

(UniProtKB/Swiss-Prot nº de orden P19136). En su forma funcional, MnP-H4 está asociada con un ion de manganeso ( $Mn^{2+}$ ), una molécula de hemo que contiene hierro y dos iones de calcio ( $Ca^{2+}$ ); MnP-H4 también tiene cinco enlaces disulfuro (Sundaramoorthy *et al.*, "The crystal structure of manganese peroxidase from *Phanerochaete chrysosporium* at 2.06-Å resolution", J Biol Chem 1994 Dec 30; 269(52): 32759-32767). La inactivación térmica de la enzima MnP-H4 implica una pérdida de las interacciones entre MnP-H4 y los iones de calcio. La creación de un enlace disulfuro adicional en MnP-H4 mediante la alteración de la secuencia de aminoácidos de MnP-H4 para sustituir una cisteína por otro aminoácido en dos posiciones cerca del sitio de interacción de calcio distal, tal como una variante MnP-H4 A48C/A63C, hace que la enzima resultante MnP-H4 A48C/A63C sea más resistente a la inactivación térmica (Reading y Aust, "Engineering a disulfide bond in recombinant manganese peroxidase results in increased thermostability", Biotechnol Prog 2000 May-Jun; 16(3): 326-333). Para expresar MnP-H4 de una manera que favorezca la formación de los enlaces disulfuro, se creó una estructura artificial de expresión que codificaba una enzima MnP-H4 que carecía del péptido señal, de modo que la proteína permanecía en el entorno citoplasmático oxidante de la célula hospedadora (*E. coli* SHuffle® Express).

La secuencia de aminoácidos de MnP-H4 sin un péptido señal se muestra como SEQ ID NO: 13; esta forma de la proteína MnP-H4 tiene un residuo de metionina inicial fijado a la secuencia de aminoácidos madura prevista, a partir de A14 de la secuencia de aminoácidos de longitud completa, como se describe en Pease *et al.*, "Manganese-dependent peroxidase from *Phanerochaete chrysosporium*. Primary structure deduced from cDNA sequence", J Biol Chem 1989 Aug 15; 264(23): 13531-13535. En otras referencias, como Sundaramoorthy *et al.*, "The crystal structure of manganese peroxidase from *Phanerochaete chrysosporium* at 2.06-Å resolution", J Biol Chem 1994 Dec 30; 269(52): 32759-32767, la secuencia de aminoácidos MnP-H4 madura se indica como a partir de A25 de la secuencia de aminoácidos de longitud completa.

Las proteínas que comprenden los aminoácidos 13 hasta 370 de SEQ ID NO: 13 (que se corresponde a la secuencia de aminoácidos madura más corta) también son producidas por los métodos de la invención, y en ciertas realizaciones tienen un residuo de metionina inicial fijado a los aminoácidos 13 hasta 370 de SEQ ID NO: 13, y por lo tanto, tienen la secuencia de aminoácidos mostrada como SEQ ID NO: 23. La secuencia de nucleótidos que se ha optimizado para la expresión de la forma de SEQ ID NO: 13 de MnP-H4 en *E. coli*, se muestra como SEQ ID NO: 14; estructuras artificiales de expresión que comprenden SEQ ID NO: 14 se usaron en los métodos de la invención para expresar MnP-H4. Los nucleótidos 37 hasta 1110 de SEQ ID NO: 14 codifican la secuencia de aminoácidos madura de MnP-H4 más corta (a partir de A25 de la secuencia de aminoácidos de longitud completa); los polinucleótidos que comprenden los nucleótidos 37 hasta 1110 de SEQ ID NO: 14, se utilizan en algunas realizaciones de la invención para la producción de la proteína MnP-H4, y en ciertas realizaciones comprenden un codón ATG para un residuo de metionina inicial fijado al extremo 5' de la secuencia de nucleótidos desde 37 hasta 1110 de SEQ ID NO: 14.

Una estructura artificial de expresión similar se crea para la expresión de una proteína MnP-H4 correspondiente a la proteína A48C/A63C MnP-H4 y que carece de un péptido señal (secuencia de aminoácidos mostrada como SEQ ID NO: 15; la secuencia codificadora optimizada se muestra como SEQ ID NO: 16). Debido a los doce aminoácidos adicionales en SEQ ID NO: 15, en comparación con la secuencia de aminoácidos de MnP-H4 madura más corta, las alteraciones de alanina a cisteína se indican como A60C/A75C en SEQ ID NO: 15. Ejemplos adicionales de estructuras artificiales de expresión que se utilizan en los métodos de la invención, codifican una proteína que comprende SEQ ID NO: 15, o los aminoácidos 13 hasta 370 de SEQ ID NO: 15, o un residuo de metionina inicial fijado a los aminoácidos 13 hasta 370 de SEQ ID NO: 15. En ciertas realizaciones, las estructuras artificiales de expresión que se utilizan en los métodos de la invención comprenden SEQ ID NO: 16, o los nucleótidos 37 a 1110 de SEQ ID NO: 16, o un codón ATG de un residuo de metionina inicial fijado al extremo 5' de la secuencia de nucleótidos desde 37 hasta 1110 de SEQ ID NO: 16. Otra variación de la secuencia de aminoácidos de MnP-H4 se produce en la posición 105 de la secuencia de aminoácidos de longitud completa, en donde una serina se cambia por una asparagina. Los métodos de la invención se usan para producir proteínas MnP-H4 con esta variación (serina cambiada a asparagina en la posición 93 de SEQ ID NO: 13 o SEQ ID NO: 15), con las estructuras artificiales de expresión que comprenden secuencias de nucleótidos que tienen una G en la posición 278 de SEQ ID NO: 14 o SEQ ID NO: 16 cambiada por una A, alterando el codón AGC de serina a un codón AAC de asparagina.

Además de favorecer la formación de enlaces disulfuro en MnP-H4, para producir una enzima MnP-H4 totalmente activa, la enzima se expresa de forma óptima en presencia de grupos hemo. Para permitir que la célula hospedadora de *E. coli* capte moléculas que contienen grupos hemo, tales como hemina, desde el medio, una secuencia de nucleótidos que codifica el receptor específico de hemina de la membrana externa de *E. coli* O157:H7 ChuA, también se incluye en la estructura artificial de expresión MnP-H4. El polipéptido ChuA tiene la misma secuencia de aminoácidos N-terminal, incluyendo el péptido señal, como la proteína natural ChuA de *E. coli* O157:H7 cepa EC4113 (SEQ ID NO: 17), de manera que se inserta en la membrana externa de la célula hospedadora *E. coli*. La secuencia codificadora optimizada para el polipéptido ChuA (SEQ ID NO: 17) se muestra en las posiciones 68 hasta 2047 de SEQ ID NO: 19. La secuencia de aminoácidos de ChuA mostrada en SEQ ID NO: 17, procedente de *E. coli* O157:H7 cepa EC4113, difiere de la de la secuencia de aminoácidos ChuA de *E. coli* CFT073 (NCBI nº de ID de gen 1037196) por tener una valina (EC4113) en lugar de una isoleucina (CFT073) en la posición 106 de SEQ ID NO: 17. Un cambio V106I en la secuencia de aminoácidos de ChuA puede estar codificado por un cambio en el codón GTG de Val a un codón ATT o ATC de Ile en las posiciones 383-385 de SEQ ID NO: 19. Otras variaciones en la secuencia de aminoácidos de proteínas ChuA empleadas en la expresión de proteínas asociadas a hemo incluyen, por ejemplo, un cambio desde ácido glutámico a glicina en la posición 259 de SEQ ID NO: 17 (codificado por un

cambio en el codón GAG de Glu a un codón GGT o GGC de Gly en las posiciones 842-844 de SEQ ID NO: 19), o un cambio desde ácido glutámico a asparagina en la posición 262 de SEQ ID NO: 17 (codificado por un cambio en el codón GAG de Glu a un codón GAT o GAC de Asp en las posiciones 851-853 de SEQ ID NO: 19).

5 La optimización para la expresión en *E. coli* y la síntesis de polinucleótidos que corresponden a SEQ ID NOs 18, 19 y 22 se llevó a cabo mediante DNA2.0 (Menlo Park, CA). SEQ ID NO: 18 codifica la proteína MnP-H4, y está diseñada para ser insertada inmediatamente aguas abajo de un promotor, tal como un promotor inducible. La secuencia de nucleótidos de SEQ ID NO: 18 comienza en su extremo 5' con un sitio de restricción GCTAGC de NheI, y tiene un sitio de unión al ribosoma AGGAGG en los nucleótidos 7 hasta 12 de SEQ ID NO: 18, seguido de la secuencia optimizada que codifica MnP-H4 en los nucleótidos 21 hasta 1130 de SEQ ID NO: 18. Aguas abajo del  
10 codón de parada de MnP-H4 está el doble terminador B0015, desde la posición 1142 hasta 1270 de SEQ ID NO: 18, seguido por un sitio de restricción de XbaI, TCTAGA. La secuencia de nucleótidos del doble terminador B0015 se obtuvo a partir de la página web partsregistry.org. SEQ ID NO: 19 codifica ChuA e incluye un promotor constitutivo, por lo que esta estructura artificial de expresión para ChuA se podría colocar en cualquier vector de expresión o en el genoma de la célula hospedadora; debido a que la secuencia que codifica ChuA se ha colocado bajo el control del  
15 promotor constitutivo J23104, su transcripción ya no está sujeta a la represión por parte de Fur. En esta realización, SEQ ID NO: 19 comienza en su extremo 5' con un sitio de restricción de XbaI TCTAGA, y está diseñada para ser colocada dentro de un vector de expresión en 3' de las secuencias de SEQ ID NO: 18. Los nucleótidos 7 a 41 de SEQ ID NO: 19 son el promotor constitutivo J23104 y los nucleótidos 50 hasta 61 de SEQ ID NO: 19 son el sitio de unión al ribosoma de B0034, ambos situados aguas arriba de la secuencia que codifica ChuA en los nucleótidos 68  
20 hasta 2047 de SEQ ID NO: 19; las secuencias de nucleótidos de J23104 y B0034 fueron obtenidas a partir de la página web partsregistry.org. SEQ ID NO: 19 termina con un sitio de restricción de Sall GTCGAC. Los sitios de XbaI en SEQ ID NO: 18 y SEQ ID NO: 19 permiten que estas secuencias de nucleótidos se ligen entre sí en un sitio de XbaI; la secuencia de nucleótidos de la estructura artificial de expresión MnP-H4 ChuA resultante se muestra en SEQ ID NO: 20. Los sitios NheI y Sall en SEQ ID NOs 18 y 19, respectivamente, y en SEQ ID NO: 20, permiten que  
25 la estructura artificial de expresión MnP-H4 ChuA se inserte en un vector de expresión, tal como pPRO33 utilizando los sitios de restricción NheI y Sall en su sitio de clonación múltiple.

Para facilitar la producción de la enzima MnP-H4 correctamente plegada, MnP-H4 se coexpresaba con la proteína chaperona disulfuro isomerasa ('PDI') procedente de *Humicola insolens*, un hifomiceto del suelo termófilo, celulolítico y saprófito (hongo de pudrición blanda). La secuencia de aminoácidos de la PDI que se coexpresaba se muestra  
30 como SEQ ID NO: 21; carece del péptido señal de la proteína natural, de manera que permanece en el citoplasma de la célula hospedadora mientras que se producen los polipéptidos MnP-H4. La secuencia de nucleótidos que codifica PDI también se optimizó para la expresión en *E. coli*; la estructura artificial de expresión de PDI se muestra como SEQ ID NO: 22. SEQ ID NO: 22 contiene un sitio de restricción de NheI GCTAGC en su extremo 5', un sitio de unión al ribosoma AGGAGG en los nucleótidos 7 a 12, la secuencia que codifica PDI en los nucleótidos 21 hasta  
35 1478, y un sitio de restricción de Sall GTCGAC en su extremo 3'. La secuencia de nucleótidos de SEQ ID NO: 22 fue diseñada para ser insertada inmediatamente aguas abajo de un promotor, tal como un promotor inducible. Los sitios de restricción de NheI y Sall en SEQ ID NO: 22 se utilizaron para insertarla en el sitio de clonación múltiple del vector de expresión pBAD24.

Las estructuras artificiales de expresión sintetizadas que comprenden SEQ ID NO: 20 y SEQ ID NO: 22, y los  
40 vectores pPRO33 y pBAD24, se cortaron con las enzimas de restricción NheI y Sall, y los fragmentos de la estructura artificial de expresión sintetizada se ligaron en los vectores para crear pPRO33-MnP-ChuA y pBAD24-PDI, como se ha descrito inmediatamente antes, y en el Ejemplo 1. Células de *E. coli* SHuffle<sup>®</sup> Express (New England Biolabs (o 'NEB'), Ipswich, Massachusetts) se cotransformaron con los vectores de expresión resultantes (pPRO33-MnP-ChuA y pBAD24-PDI) usando el método de choque térmico.

#### 45 B. Coexpresión inducible de manganeso peroxidasa y proteína disulfuro isomerasa en células bacterianas

Las células hospedadoras cotransformadas con los vectores de expresión pPRO33-MnP-ChuA y pBAD24-PDI (células SHuffle<sup>®</sup> Express (pPRO33-MnP-ChuA/pBAD24-PDI)) se utilizaron para inocular cuatro tubos de agitación, en donde cada uno contenía 5 ml de medio LB más 34 microgramos/mL de CAM y 100 microgramos/mL de AMP. Después de la incubación (a 30°C con agitación rotatoria a 250 rpm) durante 16 horas, se centrifugaron las células a  
50 4000 rpm durante 10 minutos a 4°C, el medio LB se separó por decantación, y las células se resuspendieron en 4 x 400 microlitros de medio mínimo M9 con casaminoácidos y 0,2% de glicerol, además de 34 microgramos/mL de CAM y 100 microgramos/mL de AMP ('M9-CA-gly + CAM + AMP'). A continuación, se añadieron 75 microlitros del volumen combinado de 1,6 ml de células resuspendidas a cada uno de los diez tubos de agitación que contenían 5 ml de M9-CA-gly + CAM + AMP, y se determinó que la DO<sub>600</sub> del cultivo resultante era 0,6. La hemina (Sigma-Aldrich, St. Louis, Missouri) se añadió hasta una concentración final de 8 micromolar en todos los tubos, excepto el tubo 1, el control no inducido. Los inductores L-arabinosa y propionato se añadieron a los tubos de 2-10 con las  
55 siguientes concentraciones:

## ES 2 621 320 T3

Tubo 1	No inducido (sin hemina, sin propionato, sin arabinosa)	
Tubo 2	propionato 50 mM	0,002% de arabinosa
Tubo 3	propionato 25 mM	0,002% de arabinosa
Tubo 4	propionato 12,5 mM	0,002% de arabinosa
Tubo 5	propionato 50 mM	0,01% de arabinosa
Tubo 6	propionato 25 mM	0,01% de arabinosa
Tubo 7	propionato 12,5 mM	0,01% de arabinosa
Tubo 8	propionato 50 mM	0,05% de arabinosa
Tubo 9	propionato 25 mM	0,05% de arabinosa
Tubo 10	propionato 12,5 mM	0,05% de arabinosa

Se indujeron las células a 25°C durante 12 horas con agitación por rotación, a continuación, se centrifugaron y se colocaron en un congelador a -80°C. Los sedimentos celulares se descongelaron sobre hielo, y se lisaron las células utilizando un kit de lisis bacteriana Qproteome (QIAGEN), según las instrucciones del fabricante; no se añadió una mezcla de inhibidores de proteasas y las muestras se centrifugaron a 4°C. Los extractos de proteínas solubles procedentes de las células inducidas y el control no inducido se separaron por electroforesis en gel de SDS, en condiciones reductoras sobre un gel de Bis-Tris al 10% (Life Technologies, Grand Island, Nueva York). El gel se tiñó usando RAPIDstain® (G-Biosciences). Como se muestra en la Figura 5, el lisado de las células inducidas contenía proteínas correspondientes a PDI (53 kDa) y MnP-H4 (39 kDa), lo que indicaba que estas proteínas se expresaban como proteínas solubles en las células bacterianas. La mayor cantidad de MnP-H4 soluble se producía por inducción con propionato 50 mM y 0,002% de arabinosa (carril 2).

### C. Coexpresión inducible de una forma madura alternativa de manganeso peroxidasa junto con proteína disulfuro isomerasa, y medición de la actividad de MnP-H4 [Ejemplo Comparativo]

Los vectores de expresión se prepararon para expresar una versión de MnP-H4 truncada más completamente, denominada MnP-H4\_FT, que se corresponde con la versión madura de la proteína MnP-H4 como se describe en Sundaramoorthy *et al.*, 1994 (citado anteriormente). Por tanto, la secuencia de aminoácidos de MnP-H4\_FT tiene un residuo de metionina inicial fijado a los aminoácidos 13 hasta 370 de SEQ ID NO: 13, y se proporciona como SEQ ID NO: 23. En este experimento, la secuencia que codifica MnP-H4\_FT optimizada para la expresión en *E. coli* (obtenida a partir de la secuencia que codifica MnP-H4 optimizada, descrita anteriormente), y la secuencia que codifica ChuA, se expresaron en el vector pBAD24, y la secuencia que codifica PDI (optimizada de manera similar para la expresión en *E. coli*, como se ha descrito anteriormente) se expresó en el vector pPRO33. La estructura artificial de expresión pBAD24-MnP\_FT-ChuA se preparó mediante amplificación con PCR de la secuencia que codificaba MnP-H4 y el terminador a partir de un molde que comprendía la secuencia de nucleótidos de SEQ ID NO: 18, utilizando el cebadore directo e inverso de SEQ ID NOs 24 y 25, respectivamente. El uso del cebador directo de SEQ ID NO: 24 coloca un sitio de restricción de NcoI y un codón ATG inmediatamente aguas arriba de la secuencia que codifica los aminoácidos 13 hasta 370 de SEQ ID NO: 13, y por lo tanto crea una secuencia que codifica la secuencia de aminoácidos de MnP-H4\_FT de SEQ ID NO: 23. Las secuencias de la estructura artificial de expresión de ChuA se amplificaron por PCR a partir de un molde que comprendía la secuencia de nucleótidos de SEQ ID NO: 19, utilizando los cebadores directo e inverso de SEQ ID NOs 26 y 27, respectivamente. Los productos de la PCR de MnP-H4\_FT y ChuA se cortaron con NcoI y Sall, y con Sall y HindIII, respectivamente, y se purificaron en gel y se ligaron juntos en el vector pBAD24, que se había cortado con NcoI y HindIII, tratado con CIP y purificado en gel. Los productos pBAD24-MnP\_FT-ChuA ligados comprenden la secuencia que codifica MnP-H4\_FT expresada a partir del promotor pBAD, seguida por el terminador doble B0015, el promotor constitutivo J23104, el sitio de unión al ribosoma B0034 y la secuencia que codifica la proteína ChuA. Esta estructura artificial de expresión pBAD24-MnP\_FT-ChuA tiene la secuencia de nucleótidos proporcionada en SEQ ID NO: 28.

La estructura artificial de expresión pPRO33-PDI se preparó cortando el vector pPRO33 y la estructura artificial de expresión PDI, optimizada para la expresión en *E. coli* como se ha descrito anteriormente y que comprendía la secuencia de nucleótidos de SEQ ID NO: 22, con NheI y Sall, tratando el vector pPRO33 cortado vector con CIP, purificando en gel los fragmentos y ligándolos entre sí. La estructura artificial de expresión pPRO33-PDI resultante tiene la secuencia de nucleótidos proporcionada en SEQ ID NO: 29. Las estructuras artificiales de expresión pBAD24-MnP\_FT-ChuA y pPRO33-PDI se utilizaron para cotransformar células *E. coli* SHuffle® Express (NEB) a 37 grados C durante una noche, para formar células SHuffle® Express (pBAD24-MnP\_FT-ChuA/pPRO33-PDI).

Se utilizaron células SHuffle<sup>®</sup> Express (pBAD24-MnP\_FT-ChuA/pPRO33-PDI) y control SHuffle<sup>®</sup> Express (pBAD24/pPRO33) para inocular 5 ml de medio LB + CAM (34 microgramos/mililitro) + AMP (100 microgramos/mililitro) y se dejaron crecer durante una noche a 30 grados C con agitación a 250 rpm. Las células se centrifugaron a 4000 rpm durante 10 minutos a 4 grados C, y se resuspendieron, primero en 400 microlitros de medio mínimo M9 (CA\_noGlicerol: con casaminoácidos como una fuente de carbono, pero sin glicerol) + CAM + AMP, y luego 75 microlitros del mismo se añadieron a 5 mililitros de medio mínimo M9 (CA\_noGlicerol) + CAM + AMP. Después que el cultivo, que crecía a 30 grados C con agitación a 250 rpm, alcanzó una DO600 de 0,6, se tomaron diez partes alícuotas de células SHuffle<sup>®</sup> Express (pBAD24-MnP\_FT-ChuA/pPRO33-PDI) y se indujeron con concentraciones variables de hemina y los inductores arabinosa y propionato. La muestra 1, el control, no recibió hemina ni inductores; las otras nueve muestras tenían hemina añadida hasta una concentración final de 8 micromolar; concentraciones de arabinosa de 0,025%, 0,05% o 0,1%; y concentraciones de propionato de 12,5 mM, 25 mM o 50 mM. También se incluyeron células de control SHuffle<sup>®</sup> Express (pBAD24/pPRO33) (sin insertos) en el proceso de inducción, con una muestra "inducida" y una muestra de control no inducida. Las células fueron inducidas durante 12 horas a 25 grados C con agitación a 250 rpm, a continuación, se centrifugaron y se almacenaron a -80 grados C. Para visualizar las proteínas producidas por la coexpresión inducida, los sedimentos celulares congelados se descongelaron y se lisaron usando un kit de lisis bacteriana Qproteome (QIAGEN) de acuerdo con las instrucciones, pero empleando 35 microlitros de tampón de lisis por cada 1 microlitro de cultivo bacteriano. Extractos de proteínas solubles procedentes de las células inducidas SHuffle<sup>®</sup> Express (pBAD24-MnP\_FT-ChuA/pPRO33-PDI) y el control no inducido se separaron por electroforesis en gel de SDS en condiciones reductoras, en un gel de Bis-Tris al 10%; las muestras se calentaron a 70 grados C durante 10 minutos antes de cargarlas en el gel. Después de la tinción del gel con RAPIDstain<sup>®</sup> (gel no mostrado), había bandas visibles correspondientes a MnP-H4\_FT y PDI en todas las muestras inducidas, pero no en el control no inducido, y la densidad aparente de las bandas en el gel indicaba que la combinación de 0,1% de arabinosa y propionato 50 mM producían la mayor parte de proteína MnP-H4\_FT, y que las concentraciones más altas de arabinosa generalmente producían más MnP-H4\_FT, mientras que las concentraciones de arabinosa inferiores producían generalmente más PDI, debido posiblemente a la represión catabólica del promotor de propionato mediante arabinosa.

Para determinar los niveles de actividad de MnP-H4\_FT producida por coexpresión inducible, 0,5 mililitros de la fracción de lisis celular soluble, obtenidos a partir de la coexpresión de la proteína MnP-H4\_FT en 0,1% de arabinosa y propionato 50 mM, se dializó frente a acetato de sodio 10 milimolar, CaCl<sub>2</sub> 5 milimolar, pH 4,5, utilizando un Slide-A-Lyzer<sup>®</sup> de 0,5 mililitros, MWCO 20000 (corte de peso molecular) (Thermo Fisher Scientific Inc., Waltham MA). La muestra se dializó a 4 grados C durante 2 horas, el tampón se cambió por tampón de nuevo aporte y la diálisis continuó a 4 grados C durante una noche. No hubo precipitación de proteínas al final de la diálisis, pero la proteína MnP-H4\_FT todavía era soluble y activa, como se mostró por electroforesis en gel, y con un ensayo de actividad enzimática colorimétrica. La muestra de proteína dializada que contenía MnP-H4\_FT, junto con las fracciones de proteínas solubles obtenidas a partir de células SHuffle<sup>®</sup> Express de control no inducidas (pBAD24-MnP\_FT-ChuA/pPRO33-PDI), y a partir de células SHuffle<sup>®</sup> Express de control inducidas y no inducidas sin inserto (pBAD24/pPRO33), se separaron por electroforesis en gel de SDS en condiciones reductoras en un gel de Bis-Tris al 10%; las muestras se calentaron a 70 grados C durante 10 minutos antes de cargarlas en el gel. Después de la tinción del gel con RAPIDstain<sup>®</sup>, como se muestra en la Figura 6, las bandas correspondientes a PDI (54 kDa) y MnP-H4\_FT (38 kDa) son claramente visibles en la muestra de proteína dializada, producida por la coexpresión en 0,1% de arabinosa y propionato 50 mM (Figura 6, carril 2), y no son visibles en ninguna de las muestras de control (carril 3 - SHuffle<sup>®</sup> Express (pBAD24-MnP\_FT-ChuA/pPRO33-PDI) control no inducido; carril 4 - control inducido sin inserto; carril 5 - control no inducido sin inserto). La actividad enzimática de la proteína MnP-H4\_FT producida por la coexpresión inducible se sometió a ensayo usando el ensayo colorimétrico de manganeso peroxidasa descrito en la sección E del Ejemplo 4 a continuación, y se observó que la proteína MnP-H4\_FT coexpresada tenía actividad manganeso peroxidasa comparable a la de una muestra de MnP de control positivo.

#### D. Producción y purificación de manganeso peroxidasa

Las células hospedadoras que han sido cotransformadas con vectores de expresión MnP-ChuA y PDI se sembraron en placas LB con cloranfenicol y ampicilina. Se recogen las colonias individuales y cada una se usa para inocular un tubo de agitación que contiene 15 ml de medio LB + AMP + CAM. Después de la incubación (a 30-37°C con agitación rotatoria a 250 rpm) durante un periodo de tiempo adecuado para generar cultivos en fase estacionaria, se emplean 5 ml de cada tubo con crecimiento adecuado para inocular un matraz de agitación Erlenmeyer que contiene 100 ml de medio LB + CAM + AMP. Después de una incubación adicional (a 30 - 37°C con agitación rotatoria a 250 rpm) suficiente para generar cultivos en fase estacionaria, en una muestra de cada matraz de agitación se comprueba la densidad celular adecuada. Un volumen apropiado de un matraz de agitación se introduce en un biorreactor esterilizado con el pH calibrado y se cultiva a 30 - 37°C con agitación; después de un período de crecimiento, las células se cultivan en medio que contiene hemina u otra fuente de grupos hemo, hasta que las células alcanzan una DO<sub>600</sub> de aproximadamente 0,5. Las células se inducen después mediante la adición de arabinosa y propionato; por ejemplo, 0,02% de arabinosa y propionato 50 mM, o como se determina por los métodos de valoración del Ejemplo 7, y continúa el crecimiento a 25-30°C con agitación. Después de la incubación, las células se recuperan del medio de cultivo, se lisan y se recoge el material sobrenadante de la lisis (extracto de proteína soluble). La proteína MnP-H4 se purifica a partir del extracto de proteína soluble, utilizando métodos tales

como cromatografía líquida rápida de proteínas (FPLC) con una columna de exclusión por tamaño o una columna de intercambio de iones.

#### E. Ensayo de la actividad manganeso peroxidasa

La cantidad de actividad manganeso peroxidasa en una muestra y, por lo tanto, la concentración de manganeso peroxidasa activa, se puede determinar mediante un ensayo de la capacidad para oxidar 2,6-dimetoxifenol (2,6-DMP) en coerulignona, en presencia de manganeso y peróxido de hidrógeno. El siguiente ensayo es para una muestra de 10 microlitros; cantidades alternativas para una muestra de 1 microlitro se proporcionan entre paréntesis. Se añade lo siguiente a una cubeta espectrofotométrica antes de la adición de la muestra: 0,590 ml de dH<sub>2</sub>O (0,599 ml de dH<sub>2</sub>O); 0,1 ml de solución de malonato monohidrato de sal disódica (MDSH) (5 g de MDSH en 60 ml de dH<sub>2</sub>O, pH 4,5); y 0,1 ml de solución MnSO<sub>4</sub> · H<sub>2</sub>O (0,06 g de MnSO<sub>4</sub> · H<sub>2</sub>O en 90 ml de dH<sub>2</sub>O). Inmediatamente antes de medir, 0,1 ml de solución 2,6-DMP (0,014 g de 2,6-dimetoxifenol en 90 ml de dH<sub>2</sub>O) y la muestra de 10 microlitros (1 microlitro) se añade a la cubeta. La cubeta se coloca en el espectrofotómetro y se hace una puesta a cero a 469 nm y, a continuación, se añaden 0,1 ml de H<sub>2</sub>O<sub>2</sub> de nuevo aporte (3,4 microlitros de H<sub>2</sub>O<sub>2</sub> al 30% en 30 ml de dH<sub>2</sub>O) a la cubeta. El contenido de la cubeta se mezcla pipeteando arriba y abajo tres veces. Un minuto después de la adición de H<sub>2</sub>O<sub>2</sub>, se mide la DO a 469 nm. Si la DO es mayor que 0,2, el tamaño de la muestra se reduce a 1 microlitro, o la muestra que se va a medir se diluye; si la DO es inferior a 0,005, el tamaño de la muestra se incrementa a 0,1 ml y el volumen de H<sub>2</sub>O se reduce a 0,500 ml, permaneciendo todos los demás volúmenes iguales; a continuación, se repite el ensayo.

La absorbancia medida se utiliza para calcular la concentración de enzima (CE) en Unidades/L, de acuerdo con la siguiente ecuación:

$$CE [U/L] = \frac{\text{(Absorbancia)} (\text{Volumen del ensayo [ml]} (10^6 \mu\text{mol/mol})(1 \text{ cm}))}{(\epsilon_{1 \text{ cm}}) * (\text{Volumen de la muestra [ml]})}$$

en donde  $\epsilon_{1 \text{ cm}} = 49600$  absorbancia/M·cm·min y la longitud de la vía = 1 cm. La concentración de enzima como se calculó anteriormente se convierte a mg/L, de acuerdo con la siguiente ecuación: CE [mg/L] = concentración de la enzima [U/L] / actividad enzimática específica [U/mg], en donde una actividad enzimática específica estándar para MnP es 160 U/mg. La actividad enzimática específica se calcula para cualquier muestra de MnP, midiendo independientemente la concentración de proteína en la muestra en mg/L, por ejemplo, mediante un análisis espectrofotométrico a 280 nm. Cuando la concentración enzimática (CE) se determina utilizando el ensayo de oxidación de 2,6-DMP anterior, la actividad específica en U/mg se calcula como: CE [U/L] / concentración [mg/L] = actividad específica [U/mg].

#### EJEMPLO 5

##### Coexpresión inducible de infliximab [Ejemplo Comparativo]

Infliximab es un anticuerpo monoclonal quimérico que se une a TNF-alfa, una citocina inflamatoria, y se utiliza en el tratamiento de afecciones que implican TNF-alfa, tales como enfermedades autoinmunes (enfermedad de Crohn, artritis reumatoide, psoriasis, etc.). Infliximab se forma a partir de una cadena pesada (secuencia de aminoácidos mostrada como SEQ ID NO: 30) y una cadena ligera (secuencia de aminoácidos mostrada como SEQ ID NO: 31); cada una de estas cadenas tiene una secuencia de dominio variable obtenida a partir de anticuerpos anti-TNF-alfa de ratón, y un dominio constante humano. La optimización de codones para la expresión en *E. coli* y la síntesis de polinucleótidos que codifican SEQ ID NOs 30 y 31 se realizó mediante DNA2.0 (Menlo Park, CA). La estructura artificial de expresión formada por ligación de la secuencia codificadora optimizada para la cadena pesada de infliximab en el sitio de clonación múltiple (MCS) del vector de expresión pBAD24, es pBAD24-Infliximab\_HC, que tiene la secuencia de nucleótidos mostrada como SEQ ID NO: 32. La estructura artificial de expresión formada por ligación de la secuencia codificadora optimizada para la cadena ligera de infliximab en el MCS del vector de expresión pPRO33, es pPRO33-Infliximab\_LC, que tiene la secuencia de nucleótidos mostrada como SEQ ID NO: 33. Las estructuras artificiales de expresión pBAD24-Infliximab\_HC y pPRO33-Infliximab\_LC se usan para transformar células de *E. coli* SHuffle<sup>®</sup> Express (NEB) a 37 grados C durante una noche, creando células SHuffle<sup>®</sup> Express (pBAD24- Infliximab\_HC / pPRO33-Infliximab\_LC). Estas células se cultivan generalmente como se describe en los Ejemplos 1 y 4, incluyendo la adición de compuestos selectivos tales como ampicilina y/o cloranfenicol, según sea necesario, con la excepción de que el hierro se añade preferiblemente, aunque opcionalmente al medio de crecimiento LB, por ejemplo, en forma de FeSO<sub>4</sub>·7H<sub>2</sub>O a una concentración de 3,0 miligramos por litro. Las células se centrifugan y se resuspenden en medio M9, preferiblemente, pero opcionalmente con acetato de etilo (por ejemplo, acetato de sodio al 0,27% (C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>Na)) que se añade como fuente de carbono junto con casaminoácidos (que proporcionan aminoácidos esenciales, además de ser una fuente de carbono), y también preferiblemente, pero opcionalmente con la administración de suplementos de hierro a los medios, como se ha descrito anteriormente, y se cultivan hasta que la densidad óptica (DO600) del cultivo alcanza 0,8. (Véase, Paliy y Gunasekera, "Growth of *E. coli* BL21 in minimal media with different gluconeogenic carbon sources and salt contents", Appl Microbiol Biotechnol 2007 Jan; 73(5): 1169-1172; Epub 2006 Aug 30; Errata en: Appl Microbiol Biotechnol 2006 Dec; 73(4): 968). En ese momento, las células se inducen por la adición de arabinosa (inicialmente

a concentraciones que incluyen 0,1%) y propionato de etilo (inicialmente a concentraciones que incluyen 50 mM). El ajuste de las concentraciones de arabinosa y propionato se puede realizar como se describe en el Ejemplo 7, y los anticuerpos de infliximab producidos se purifican y se caracterizan como se describe en el Ejemplo 9, hasta el Ejemplo 11.

## 5 EJEMPLO 6

Coexpresión inducible de proteínas en células de levadura [Ejemplo Comparativo]

Estructuras artificiales de expresión para la coexpresión inducible en células hospedadoras de levadura de MnP-H4\_FT con PDI y las cadenas pesadas y ligeras de anticuerpo IgG1 de ratón anti-CD19 humano, han sido creadas por el siguiente método llevado a cabo por GenWay Biotech Inc. (San Diego CA). Los vectores pJ1231-03C y pJ1234-03C (DNA2.0, Menlo Park, CA) se utilizaron como la parte de la estructura principal de las estructuras artificiales de expresión de levadura, ya que contienen elementos necesarios para el mantenimiento del plásmido en *E. coli* (origen de replicación de pUC) o *Saccharomyces cerevisiae* (origen de replicación de plásmido circular de 2 micras, junto con marcadores seleccionables útiles en cualquier hospedador: KanR (*E. coli*) y Leu2 (levadura) en pJ1231-03C; AmpR (*E. coli*) y Ura3 (levadura) en pJ1234-03C. Las secuencias de nucleótidos de los vectores pJ1231-03C y pJ1234-03C se muestran como SEQ ID NOs 34 y 35, respectivamente. Los vectores pJ1231-03C y pJ1234-03C fueron tratados con la endonucleasa de restricción BfuA1 (NEB, n° de catálogo R0701S); los fragmentos de los vectores que no fueron retenidos después de la digestión con BfuA1 comprenden un promotor de la expresión, la secuencia que codifica DasherGFP y la secuencia terminadora de CYC1. La PCR se realizó en cuatro estructuras artificiales de expresión que contienen secuencias que codifican los polipéptidos que se van a expresar, para crear secuencias que a continuación, se ligan en los vectores pJ1231-03C y pJ1234-03C cortados con BfuA1. Todas las reacciones de PCR se realizaron utilizando ADN polimerasa Pfx Platinum® (Life Technologies, Grand Island, Nueva York, n° de catálogo 11708021) y el kit de purificación de ADN QIAEX II (QIAGEN, n° de catálogo 20051) se utilizó para la purificación del producto de la PCR y fragmentos de vector, de acuerdo con las instrucciones del fabricante.

La estructura artificial de expresión pBAD24-MnP\_FT-ChuA (SEQ ID NO: 28), que contiene la secuencia que codifica AraC, el promotor inducible con arabinosa pBAD y las secuencias que codifican MnP-H4\_FT y ChuA, se utiliza como molde en dos reacciones de PCR distintas, con el fin de agregar un marcador 6xHis en los extremos C-terminales de la secuencia que codifica MnP-H4\_FT y la secuencia que codifica ChuA, y sitios de restricción de Bsal para la clonación en los vectores. (Preferiblemente, pero opcionalmente, la secuencia que codifica ChuA no está alterada para añadir un marcador His en su extremo C-terminal; esto se puede lograr mediante el uso de un cebador con una secuencia similar a la de SEQ ID NO: 39, pero sin las bases 21 hasta 38 de SEQ ID NO: 39). Los cebadores usados en estas dos reacciones de PCR son (1) el cebador Bsal-AraC-MnP-H4\_FT (SEQ ID NO: 36) y el cebador MnP-H4\_FT-6xHis-inverso (SEQ ID NO: 37), y (2) el cebador MnP-H4\_FT-6xHis-directo (SEQ ID NO: 38) y el cebador ChuA-6xHis-Bsal (SEQ ID NO: 39). Una reacción de PCR adicional se lleva a cabo a continuación, sobre los dos productos de PCR purificados, utilizando los cebadores Bsal-AraC-MnP-H4\_FT y ChuA-6xHis-Bsal, para crear un único producto (AraC-MnP-H4\_FT-ChuA, SEQ ID NO: 40) que se purifica, se corta con Bsal (NEB, n° de catálogo R0535S), y se liga en el vector pJ1231-03C cortado con BfuA1, utilizando ADN ligasa de T4 (Life Technologies, n° de catálogo 15224025) para crear la estructura artificial de expresión pJ1231-AraC-MnP-H4\_FT-ChuA. (Preferiblemente, el fragmento AraC-MnP-H4\_FT-ChuA cortado con Bsal también se liga en el vector pJ1234-03C cortado con BfuA1, para crear la estructura artificial de expresión pJ1234-AraC-MnP-H4\_FT-ChuA).

La estructura artificial de expresión pPRO33-PDI (SEQ ID NO: 29), que contiene la secuencia que codifica PrpR, el promotor de pPRO inducible con propionato y la secuencia que codifica PDI, se utilizó como molde en una reacción PCR, que añadía un marcador 5xHis en el extremo C-terminal de la secuencia que codifica PDI, y sitios de restricción Bsal para la clonación en los vectores. (Preferiblemente, pero opcionalmente, la secuencia que codifica PDI no está alterada para añadir un marcador His en su extremo C-terminal; esto se puede lograr mediante el uso de un cebador con una secuencia semejante a la de SEQ ID NO: 42, pero sin las bases 21 hasta 35 de SEQ ID NO: 42). Los cebadores utilizados en esta reacción de PCR eran el cebador Bsal-PrpR-PDI (SEQ ID NO: 41) y el cebador PDI-5xHis-Bsal (SEQ ID NO: 42), creando un producto de PCR (PrpR-PDI, SEQ ID NO: 43) que se purificó, se cortó con Bsal y se ligó en el vector pJ1231-03C cortado con BfuA1, utilizando ADN ligasa de T4 para crear la estructura artificial de expresión pJ1231-PrpR-PDI. (Preferiblemente, el fragmento PrpR-PDI cortado con Bsal también se liga en el vector pJ1234-03C cortado con Bsal para crear la estructura artificial de expresión pJ1234-PrpR-PDI).

Las estructuras artificiales de expresión que codifican la cadena pesada de IgG1 de ratón anti-CD19 humano y la cadena ligera de IgG1 de ratón anti-CD19 humano - pBAD24-HC y pPRO33-LC, respectivamente - se utilizaron como moldes en dos reacciones de PCR distintas, cada una con el fin de eliminar la secuencia señal de cada secuencia codificante y para añadir sitios de restricción Bsal para la clonación en los vectores. Para pBAD24-HC, los cebadores usados en estas dos reacciones de PCR eran (1) el cebador Bsal-AraC-HC-directo (SEQ ID NO: 44) y el cebador HC-inverso (SEQ ID NO: 45), y (2) el cebador HC-directo (SEQ ID NO: 46) y el cebador HC-Bsal-inverso (SEQ ID NO: 47). Para pPRO33-LC, los cebadores usados en estas dos reacciones de PCR eran (1) el cebador Bsal-PrpR-LC-directo (SEQ ID NO: 48) y el cebador LC-inverso (SEQ ID NO: 49), y (2) el cebador LC-directo (SEQ ID NO: 50) y el cebador LC-Bsal-inverso (SEQ ID NO: 51). El uso del cebador Bsal-AraC-HC-directo (SEQ ID NO:

44) y el cebador Bsal-PrpR-LC-directo (SEQ ID NO: 48) en estas reacciones da lugar a secuencias que codifican HC y LC que carecen de secuencia señal, a las que se hace referencia como HC\_NS (sin señal, del inglés "no signal") y LC\_NS: la secuencia que codifica HC\_NS modificada da como resultado un polipéptido HC\_NS que tiene un residuo de metionina inicial seguido por los aminoácidos 20 hasta 464 de SEQ ID NO: 2; la secuencia que codifica LC\_NS modificada da como resultado un polipéptido LC\_NS que tiene un residuo de metionina inicial seguido de un residuo de alanina y después los aminoácidos 21 hasta 239 de SEQ ID NO: 4. A continuación, se llevó a cabo una reacción de PCR adicional en cada conjunto de dos productos de PCR purificados, utilizando los cebadores Bsal-AraC-HC-directo y HC-Bsal-inverso, y los cebadores Bsal-PrpR-LC-directo y LC-Bsal-inverso, respectivamente, para crear dos productos individuales (AraC-HC\_NS, SEQ ID NO: 52 y PrpR-LC\_NS, SEQ ID NO: 53) que se purificaron cada uno, se cortaron con Bsal y se ligaron en el vector pJ1231-03C cortado con BfuA1 (PrpR-LC\_NS) o el vector pJ1234-03C cortado con BfuA1 (AraC-HC\_NS) utilizando la ligasa de ADN de T4, para crear las estructuras artificiales de expresión pJ1231-PrpR-LC\_NS y pJ1234-AraC-HC\_NS.

Las mezclas de ligasa se transformaron en células de *E. coli* DH5alfa y se sembraron en placas de agar LB con cantidades suficientes de kanamicina o ampicilina para mantener los plásmidos en las células DH5alfa. La preparación de ADN plasmídico se realizó de acuerdo con métodos convencionales. (De forma opcional, pero preferiblemente, el ADN plasmídico preparado se utiliza en reacciones de secuenciación para confirmar las secuencias de los insertos de plásmidos).

Células INVSc-1 de *S. cerevisiae* competentes se prepararon usando el kit de transformación S.c. EasyComp<sup>®</sup> (Life Technologies, n° de catálogo K5050-01) según las instrucciones del fabricante. La cepa INVSc-1 de *S. cerevisiae* tiene el siguiente genotipo (MATa his3Δ1 leu2 trp1-289 ura3-52/MATα his3Δ1 leu2 trp1-289 ura3-52) y el fenotipo: His<sup>-</sup>, Leu<sup>-</sup>, Trp<sup>-</sup>, Ura<sup>-</sup>. Resumiendo, una sola colonia de la cepa INVSc-1 se inoculó en 10 ml de medio YPD (contiene, por litro: 10 g de extracto de levadura, 20 g de peptona, 20 g de glucosa) y se dejó crecer durante una noche a 30 grados C en una incubadora con agitación a 250 rpm. Al día siguiente, el cultivo de una noche se diluyó en 10 mililitros de medio YPD de nuevo aporte hasta una DO<sub>600</sub> de 0,3 y se dejó crecer hasta que la DO<sub>600</sub> alcanzó 0,8. Las células se recogieron por centrifugación a 1500 rpm durante 5 minutos a temperatura ambiente. Después de eso, las células se resuspendieron en 10 ml de Solución de 1 (solución de lavado) y se recogieron mediante centrifugación a 1500 rpm durante 5 minutos a temperatura ambiente. El material sobrenadante se desechó y las células se resuspendieron en 1 mililitro de Solución 2 (solución de resuspensión), se dividió en partes alícuotas de 50 microlitros y se almacenó a -80 grados C. Para transformar las células de levadura competentes con las estructuras artificiales de expresión, 50 microlitros de las células INVSc-1 competentes se mezclaron con 1,2 microgramos de cada vector de expresión o vector de control y 500 microlitros de Solución 3 (solución de transformación). Luego, las células se mezclaron vigorosamente y se incubaron en un baño de agua a 30 grados C durante 1 hora y se agitaron en vórtice durante 10 segundos cada 15 minutos. Partes alícuotas de 100 y 400 microlitros de las mezclas de transformación se sembraron en placas de agar mínimo SC, en ausencia de reactivo selectivo apropiado. El medio mínimo SC contiene, por litro: 6,7 g de base nitrogenada de levadura, 20 g de glucosa, 0,05 g de ácido aspártico, 0,05 g de histidina, 0,05 g de isoleucina, 0,05 g de metionina, 0,05 g de fenilalanina, 0,05 g de prolina, 0,05 g de serina, 0,05 g de tirosina, 0,05 g de valina, 0,1 g de adenina, 0,1 g de arginina, 0,1 g de cisteína, 0,1 g de leucina (omitida en los medios selectivos -Leu), 0,1 g de lisina, 0,1 g de treonina, 0,1 g de triptófano y 0,1 g de uracilo (omitido en los medios selectivos -Ura). Células INVSc-1 transformadas con pJ1231-PrpR-PDI, con pJ1231-PrpR-LC\_NS y con el vector de control pJ1231-03C, se seleccionaron sobre placas sin leucina a 30 grados C durante 48-72 horas. Células INVSc-1 transformadas con pJ1234-AraC-HC\_NS y con el vector de control pJ1234-03C, se cultivaron en las mismas condiciones en placas de agar mínimo SC sin uracilo. (Preferiblemente, las células INVSc-1 transformadas con pJ1231-AraC-MnP-H4\_FT-ChuA, con pJ1234-AraC-MnP-H4\_FT-ChuA y con pJ1234-PrpR-PDI también se cultivan en las mismas condiciones en placas de agar mínimo SC sin uracilo). Células INVSc-1 cotransformadas simultáneamente con pJ1234-AraC-HC\_NS y pJ1231-PrpR-LC\_NS fueron seleccionadas en las mismas condiciones en placas de agar mínimo SC sin leucina o uracilo. (Preferiblemente, las células INVSc-1 cotransformadas de forma simultánea con pJ1231-AraC-MnP-H4\_FT-ChuA y pJ1234-PrpR-PDI, o con pJ1234-AraC-MnP-H4\_FT-ChuA y pJ1231-PrpR-PDI, se seleccionan en las mismas condiciones en placas de agar mínimo SC sin leucina o uracilo).

Las células de todas las colonias de cada transformación se rasparon y se resuspendieron en 4 ml de medio mínimo SC líquido, en ausencia de reactivo selectivo apropiado y fuente de carbono. La DO<sub>600</sub> en cada cultivo se midió y se normalizó a 0,4 unidades ópticas. La expresión de proteínas se indujo mediante la adición de 2% de arabinosa filtrada de forma estéril (Sigma-Aldrich, St. Louis, Missouri, n° de catálogo A3256-25G) a cultivos transformados con pJ1234-AraC-HC\_NS. En las células transformadas con pJ1231-PrpR-PDI y con pJ1231-PrpR-LC\_NS, la expresión de proteínas se indujo con la adición de 2% de propionato filtrado de forma estéril (Sigma-Aldrich, número de catálogo P188-100G). Y la coexpresión de pJ1234-AraC-HC\_NS y pJ1231-PrpR-LC\_NS en el cultivo correspondiente fue inducida por la adición de 1% de arabinosa y 1% de propionato. (Preferiblemente, el medio de inducción para la expresión de proteínas a través de las células transformadas con pJ1231-AraC-MnP-H4\_FT-ChuA, y para las células cotransformadas con pJ1231-AraC-MnP-H4\_FT-ChuA y pJ1234-PrpR-PDI, o con pJ1234-AraC-MnP-H4\_FT-ChuA y pJ1231-PrpR-PDI, también contiene hemina (Sigma-Aldrich, n° de catálogo H9039 o 51280) añadida hasta una concentración final de 8 micromolar). El curso de tiempo para la expresión de proteínas era de 24 horas, a 30 grados C en una incubadora con agitación a 250 rpm. Las células procedentes de 0,5 mililitros de cada cultivo inducido previamente, se recogieron mediante centrifugación, se lavaron con 1 mililitro de agua desionizada y

se almacenaron a -80 grados C. Las muestras procedentes de cultivos inducidos posteriormente, se prepararon de la misma manera. Los extractos de proteínas totales se prepararon a partir de cultivos inducidos previa y posteriormente, se resolvieron en 4-20% de SDS-PAGE y se transfirieron a una membrana de PDVF. El nivel de expresión de las proteínas diana se analizó mediante transferencia Western, utilizando marcador anti-6xHis y anticuerpos anti-IgG humana.

#### EJEMPLO 7

Valoración de la coexpresión mediante la variación de la concentración de inductor

Para optimizar la producción de un producto multímero utilizando los sistemas de coexpresión inducible de la invención, es posible ajustar de forma independiente o valorar la concentración de los inductores. Las células hospedadoras que contienen estructuras artificiales de expresión inducibles con L-arabinosa y propionato se cultivan hasta la densidad deseada (tal como una  $DO_{600}$  de aproximadamente 0,5) en medio mínimo M9 que contiene los antibióticos apropiados, a continuación, las células se dividen en partes alícuotas en volúmenes pequeños de medio mínimo M9, preparado opcionalmente sin ninguna fuente de carbono tal como glicerol, y con los antibióticos apropiados y concentraciones variables de cada inductor. La concentración de L-arabinosa necesaria para inducir la expresión es normalmente menor que 2%. En un experimento de valoración, las concentraciones sometidas a ensayo de L-arabinosa pueden variar de 2% a 1,5%, 1%, 0,5%, 0,2%, 0,1%, 0,05%, 0,04%, 0,03%, 0,02%, 0,01%, 0,005%, 0,002% y 0,001%. Las concentraciones de L-ramnosa o D-xilosa necesarias para inducir la expresión de los promotores inducibles con L-ramnosa o D-xilosa se analizan de manera similar, con las concentraciones analizadas en el intervalo desde 5% a 0,01%. Para cada concentración 'x' de L-arabinosa (o L-ramnosa o D-xilosa) que se somete a ensayo, la concentración de un inductor diferente, tal como propionato, añadido a cada uno de los tubos que contienen la concentración 'x' del primer inductor, varía en cada serie de muestras, que en el caso del propionato está en el intervalo desde 1 M a 750 mM, 500 mM, 250 mM, 100 mM, 75 mM, 50 mM, 25 mM, 10 mM, 5 mM y 1 mM. Alternativamente, los experimentos de valoración pueden comenzar con una combinación "estándar" de las concentraciones de inductor, que es 0,002% de cualquiera entre L-arabinosa, L-ramnosa o D-xilosa, y/o propionato 50 mM, y someter a ensayo nuevas combinaciones de concentraciones de inductor que varían de la combinación "estándar". Experimentos de valoración similares se pueden realizar con cualquier combinación de inductores utilizados en un sistema de coexpresión inducible de la invención, incluyendo pero no limitados a, L-arabinosa, propionato, L-ramnosa y D-xilosa. Después del crecimiento en presencia de inductores durante 6 horas, las células se sedimentan, el producto deseado se extrae de las células y el rendimiento de producto por valor de la masa de las células se determina con un ensayo inmunológico cuantitativo, tal como ELISA, o mediante la purificación del producto y cuantificación por absorbancia de UV a 280 nm.

También es posible valorar las concentraciones de inductor utilizando un ensayo de alto rendimiento, en el que las proteínas que se van a expresar, han sido diseñadas para incluir un resto de proteína fluorescente, tal como la proporcionada por la proteína fluorescente de color rojo mKate2 (Evrogen, Moscú, Rusia), o las proteínas fluorescentes verdes mejoradas de *Aequorea victoria* y *Bacillus cereus*. Otro enfoque para determinar la cantidad y la actividad de los productos génicos producidos con diferentes concentraciones de inductores en un experimento de valoración de alto rendimiento, es el uso de un sensor capaz de medir las interacciones de unión biomolecular, tal como un sensor que detecta la resonancia de plasmón superficial, o un sensor que emplea interferometría de biocapa (BLI) (por ejemplo, un sistema Octet<sup>®</sup> QK de forteBIO, Menlo Park, CA).

#### EJEMPLO 8

Medición de la fuerza de los promotores y la homogeneidad de la expresión inducible

La fuerza de un promotor se mide como la cantidad de transcripción de un producto génico, iniciada por ese promotor, respecto a un control adecuado. Para promotores constitutivos que dirigen la expresión de un producto génico en una estructura artificial de expresión, un control adecuado podría utilizar la misma estructura artificial de expresión, excepto que se emplee la versión "de tipo silvestre" del promotor, o un promotor de un gen de "mantenimiento", en lugar del promotor que se va a someter a ensayo. Para los promotores inducibles, la expresión del producto génico desde el promotor se puede comparar en condiciones inductoras y no inductoras.

A. Medición de la fuerza del promotor usando PCR cuantitativa para determinar los niveles de ARN transcrito desde el promotor

El método de De Mey *et al.* ("Promoter knock-in: a novel rational method for the fine tuning of genes", BMC Biotechnol 2010 Mar 24; 10: 26) se utiliza para determinar la fuerza relativa de los promotores en células hospedadoras que pueden crecer en un cultivo. Las células hospedadoras que contienen una estructura artificial de expresión con el promotor que se va a someter a ensayo, y células hospedadoras de control que contienen una estructura artificial de expresión de control, se dejan crecer en cultivo por triplicado. Se recogen muestras de un ml a una  $DO_{600} = 1,0$  para el ARNm y la recogida de proteínas. La extracción de ARN total se consigue mediante un mini kit RNeasy (QIAGEN, Países Bajos). La pureza del ARN se verifica en un gel de agarosa FA según lo recomendado por QIAGEN y la concentración de ARN se determina midiendo la absorbancia a 260 nm. Dos microgramos de ARN se utilizan para sintetizar ADNc, utilizando un cebador aleatorio y transcriptasa inversa RevertAid H Minus M-MuLV

(Fermentas, Glen Burnie, Maryland). La fuerza del promotor se determina por RT-qPCR realizada en un iCycler IQ® (Bio-Rad, Eke, Bélgica) utilizando cebadores directos e inversos, diseñados para amplificar el ADNc correspondiente al transcrito producido desde el promotor. (Para este propósito, los autores De Mey *et al.* utilizaron los cebadores Fw-ppc-qPCR y Rv-ppc-qPCR, y los cebadores Fw-rpoB-qPCR y Rv-rpoB-qPCR desde el gen de mantenimiento de control *rpoB*). La supermezcla SYBR Greener qPCR (Life Technologies, Grand Island, Nueva York) se utiliza para realizar una breve incubación con UDG (uracil ADN glicosilasa) (50°C durante 2 min) seguida inmediatamente por la amplificación con PCR (95°C durante 8,5 min; 40 ciclos a 95°C durante 15 s y 60°C durante 1 min) y el análisis de la curva de fusión (95°C durante 1 min, 55°C durante 1 min y 80 ciclos de 55°C + 0,5°C/ciclos durante 10 s) para identificar la presencia de dímeros de cebadores y analizar la especificidad de la reacción. Esta etapa de incubación con UDG antes de la PCR cíclica destruye cualquier producto contaminante que contiene dU de las reacciones anteriores. La UDG se inactiva después mediante temperaturas elevadas durante el ciclo normal de PCR, lo que permite la amplificación de secuencias diana auténticas. Cada muestra se realiza por triplicado. Las proporciones de expresión relativa se calculan utilizando el "método delta-delta ct" de PE Applied Biosystems (Perkin Elmer, Foster City, California).

#### 15 B. Medición de la fuerza del promotor inducible y la homogeneidad de la inducción usando un gen informador fluorescente

Estos experimentos se llevan a cabo utilizando los métodos de Khlebnikov *et al.* ("Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture", J Bacteriol 2000 Dec; 182(24): 7029-7034). Los experimentos que miden la inducción de un promotor inducible se realizan en medio C complementado con 3,4% de glicerol como fuente de carbono (Helmstetter, "DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r", J Mol Biol 1968 Feb 14; 31(3): 507-518). Cepas de *E. coli* que contienen estructuras artificiales de expresión que comprenden al menos un promotor inducible que controla la expresión de un gen informador fluorescente, se hacen crecer a 37°C con selección con antibióticos a una densidad óptica a 600 nm (DO600) desde 0,6 a 0,8. Las células se recogen por centrifugación (15.000 x g), se lavan en medio C sin una fuente de carbono, se resuspenden en medio C de nuevo aporte que contiene antibióticos, glicerol y/o inductor (para la inducción de la expresión génica) a una DO600 desde 0,1 a 0,2, y se incuban durante 6 h. Las muestras se toman de manera rutinaria durante el período de crecimiento para el análisis. La fluorescencia del cultivo se mide en un fluorómetro Versafluor (Bio-Rad Inc., Hercules, California) con excitación a una longitud de onda de 360/40 nm y filtros de emisión con una longitud de onda de 520/10 nm. La fuerza de la expresión de un promotor inducible después de la inducción se puede expresar como la relación de la fluorescencia de población máxima promediada (proporción de fluorescencia/DO) de las células inducidas, con respecto a la de las células de control (como células no inducidas). Para determinar la homogeneidad de la inducción dentro de la población celular, se realiza una citometría de flujo en un citómetro de flujo de Beckman-Coulter EPICS XL (Beckman Instruments Inc., Palo Alto, California) equipado con un láser de argón (emisión a una longitud de onda de 488 nm y 15 mW) y un filtro de paso de banda con una longitud de onda de 525 nm. Antes del análisis, las células de la muestra se lavan con solución salina tamponada con fosfato que se había filtrado (tamaño de poro del filtro, 0,22 micrómetros), se diluyen hasta una DO600 de 0,05 y se colocan sobre hielo. Para cada muestra, se recogen 30.000 eventos con una tasa de entre 500 y 1.000 eventos/s. El porcentaje de células inducidas (fluorescentes) en cada muestra se puede calcular a partir de los datos de la citometría de flujo.

#### 40 EJEMPLO 9

##### Purificación de anticuerpos

Los anticuerpos producidos por los sistemas de coexpresión inducible de la invención se purifican por centrifugación de las muestras de células hospedadoras lisadas, a 10.000 x g durante 10 minutos, para eliminar cualquier célula y residuo. El material sobrenadante se filtra a través de un filtro de 0,45 micrómetros. Una columna de 1 ml de proteína G recombinante - Sepharose® 4B (Life Technologies, Grand Island, Nueva York) está configurada para alcanzar caudales de 1 ml/min, y se utiliza con los siguientes tampones: tampón de unión: fosfato de sodio 0,02 M, pH 7,0; tampón de elución: glicina-HCl 0,1 M, pH 2,7; y tampón de neutralización: Tris-HCl 1 M, pH 9,0. La columna se equilibra con 5 volúmenes de columna (5 ml) de tampón de unión, y después la muestra se aplica a la columna. La columna se lava con 5-10 volúmenes de columna de tampón de unión para eliminar las impurezas y el material no unido, continuando hasta que no se detecta ninguna proteína en el eluyente (determinado por absorbancia de UV a 280 nm). La columna se eluye a continuación con 5 volúmenes de columna de tampón de elución, y la columna se vuelve a equilibrar inmediatamente con 5-10 volúmenes de columna de tampón de unión.

#### EJEMPLO 10

##### Medición de la afinidad de unión del anticuerpo

55 La afinidad de unión del anticuerpo, expresada como "Kd" o "valor de Kd", se mide mediante un ensayo de unión a antígeno radiomarcado (RIA) realizado con la versión Fab de un anticuerpo de interés y su antígeno, tal como se describe mediante el siguiente ensayo. La producción de la versión Fab de un anticuerpo de longitud completa es bien conocida en la técnica. La afinidad de unión en solución de los Fabs hacia un antígeno se mide equilibrando un Fab con una concentración mínima de antígeno marcado con (<sup>125</sup>I) en presencia de una serie de valoraciones de

antígeno no marcado, a continuación, se captura el antígeno unido con una placa recubierta de anticuerpo anti-Fab (véase, por ejemplo, Chen *et al.*, "Selection and analysis of an optimized anti-VEGF antibody: crystal structure of an affinity-matured Fab in complex with antigen", J Mol Biol 1999 Nov 5; 293(4): 865-881). Para establecer las condiciones del ensayo, placas de microtitulación (Dyner Technologies, Inc., Chantilly, Virginia) se recubren durante una noche con 5 microgramos/mL de un anticuerpo anti-Fab de captura (Cappel Labs, West Chester, Pensilvania) en carbonato de sodio 50 mM (pH 9,6), y posteriormente se bloquean con 2% (p/v) de albúmina de suero bovino en PBS durante dos a cinco horas a temperatura ambiente (aproximadamente 23°C). En una placa no adsorbente (Nunc nº 269620; Thermo Scientific, Rochester, Nueva York), se mezcla antígeno (<sup>125</sup>I) 100 pM o 26 pM con diluciones en serie de un Fab de interés (por ejemplo, en consonancia con la determinación del anticuerpo anti-VEGF, Fab-12, en Presta *et al.*, "Humanization of an anti-vascular endothelial growth factor monoclonal antibody for the therapy of solid tumors and other disorders", Cancer Res 1997 Oct 15; 57(20): 4593-4599). El Fab de interés se incubaba después durante una noche; sin embargo, la incubación puede continuar durante un período más largo (por ejemplo, aproximadamente 65 horas) para asegurar que se alcanza el equilibrio. A continuación, las mezclas se transfieren a la placa de captura para incubarla a temperatura ambiente (por ejemplo, durante una hora). A continuación, se retira la solución y se lava la placa ocho veces con 0,1% de tensioactivo Tween-20® en PBS. Cuando las placas se han secado, se añaden 150 microlitros/pocillo de líquido de centelleo (Microscint-20® PerkinElmer, Waltham, Massachusetts), y se hace un recuento de las placas en un contador gamma TOPCOUNT® (PerkinElmer) durante diez minutos. Las concentraciones de cada Fab que proporcionan menos o igual a un 20% de unión máxima, se eligen para uso en ensayos de unión competitiva.

Alternativamente, la K<sub>d</sub> o el valor de K<sub>d</sub> se mide utilizando ensayos de resonancia de plasmón de superficie, usando un instrumento BIAcore®-2000 o BIAcore®-3000 (BIAcore, Inc., Piscataway, New Jersey) a 25°C con chips con antígeno CM5 inmovilizado, a ~10 unidades de respuesta (UR). Brevemente, chips de biosensores con dextrano carboximetilado (CM5, BIAcore Inc) se activan con clorhidrato de N-etil-N'-(3-dimetilamino-propil)carbodiimida (EDC) y N-hidroxisuccinimida (NHS) de acuerdo con las instrucciones del proveedor. El antígeno se diluye con acetato de sodio 10 mM, pH 4,8, a 5 microgramos/mL (~0,2 micromolar) antes de la inyección con un caudal de 5 microlitros/minuto, para conseguir aproximadamente 10 UR de proteína acoplada. Después de la inyección de antígeno, se inyecta etanolamina 1 M para bloquear los grupos que no han reaccionado. Para las mediciones de cinética, se inyectan diluciones dobles en serie de Fab (0,78 nM a 500 nM) en PBS con 0,05% de tensioactivo Tween 20® (PBST) a 25°C con un caudal de aproximadamente 25 microlitros/min. Las tasas de asociación (k<sub>on</sub>) y las tasas de disociación (k<sub>off</sub>) se calculan usando un modelo simple de unión a uno-a-uno de Langmuir (programa informático de evaluación de BIAcore®, versión 3.2) mediante el ajuste simultáneo de los sensogramas de asociación y disociación. La constante de disociación en equilibrio (K<sub>d</sub>) se calcula como la relación k<sub>off</sub>/k<sub>on</sub>. Si la tasa de asociación excede a 10<sup>6</sup> M<sup>-1</sup> s<sup>-1</sup> en el ensayo de resonancia de plasmón superficial anterior, entonces la tasa de asociación se puede determinar mediante el uso de una técnica de extinción de la fluorescencia que mide el incremento o la disminución de la intensidad de la emisión de fluorescencia (excitación = 295 nm; emisión = 340 nm, banda de paso de 16 nm) a 25°C de un anticuerpo anti-antígeno 20 nM (forma Fab) en PBS, pH 7,2, en presencia de concentraciones crecientes de antígeno, tal como se mide en un espectrómetro, como un espectrofotómetro equipado con flujo interrumpido (Aviv Instruments) o un espectrómetro de 8000 series SLM-AMINCO® (ThermoSpectronic) con una cubeta agitada.

#### 40 EJEMPLO 11

##### Caracterización de los enlaces disulfuro presentes en los productos de coexpresión

La cantidad y la ubicación de los enlaces disulfuro en los productos de proteínas coexpresadas se pueden determinar mediante la digestión de la proteína con una proteasa, tal como tripsina, en condiciones no reductoras, y someter los fragmentos peptídicos resultantes a espectrometría de masas (MS) combinando etapas de MS de disociación de transferencia de electrones secuencial (ETD) y disociación inducida por colisión (CID) (MS2, MS3) (Nili *et al.*, "Defining the disulfide bonds of insulin-like growth factor-binding protein-5 by tandem mass spectrometry with electron transfer dissociation and collision-induced dissociation", J Biol Chem 2012 Jan 6; 287(2): 1510-1519; Epub 2011 Nov 22).

Digestión de proteínas coexpresadas. Para evitar los reordenamientos de los enlaces disulfuro, cualquier residuo de cisteína libre se bloquea primero mediante alquilación: la proteína coexpresada se incubaba protegida de la luz con el agente alquilante yodoacetamida (5 mM) con agitación durante 30 minutos a 20°C en tampón con urea 4 M, y a continuación, se separa mediante SDS-PAGE no reductor, usando geles prefabricados. Alternativamente, la proteína coexpresada se incubaba en el gel después de la electroforesis con yodoacetamida, o sin ella como control. Las bandas de proteínas se tiñen, se destiñen con agua doblemente desionizada, se escinden y se incuban dos veces en 500 microlitros de bicarbonato de amonio 50 mM, 50% (v/v) de acetonitrilo, mientras que se agita durante 30 minutos a 20°C. Las muestras de proteína se deshidratan en 100% de acetonitrilo durante 2 minutos, se secan por centrifugación a vacío y se rehidratan con 10 mg/mL de tripsina o quimotripsina en tampón que contiene bicarbonato de amonio 50 mM y cloruro de calcio 5 mM durante 15 minutos sobre hielo. El exceso de tampón se elimina y se reemplaza con 50 microlitros del mismo tampón sin enzima, seguido de incubación durante 16 horas a 37°C o 20°C, para tripsina y quimotripsina, respectivamente, con agitación. Las digestiones se detienen mediante la adición de 3 microlitros de ácido fórmico al 88%, y después de una breve agitación en vórtice, se retira el material sobrenadante y se almacenan a -20°C hasta el análisis.

Localización de los enlaces disulfuro por espectrometría de masas. Los péptidos se inyectan en una columna de atrapamiento de 1 mm x 8 mm (Michrom BioResources, Inc., Auburn, CA) a 20 microlitros/minuto en una fase móvil que contiene ácido fórmico al 0,1%. El cartucho de atrapamiento se coloca entonces en línea con una columna de 0,5 mm x 250 mm que contiene 5 mm de fase estacionaria Zorbax SB-C18 (Agilent Technologies, Santa Clara, CA), y los péptidos se separan por un gradiente de acetonitrilo de 2-30% durante 90 minutos a 10 microlitros/minuto, con HPLC capilar en serie 1100 (Agilent Technologies). Los péptidos se analizan utilizando una trampa de iones lineal LTQ Velos con una fuente de ETD (Thermo Scientific, San Jose, CA). La ionización por electrospray se realiza usando una fuente de Captive Spray (Michrom BioResources, Inc.). Las exploraciones MS de vigilancia están seguidas por siete exploraciones dependientes de los datos que consisten en exploraciones MS2 CID y ETD en el ion más intenso en la exploración de vigilancia, seguidas por cinco exploraciones MS3 CID en el primer hasta el quinto ion más intenso en la exploración MS2 ETD. Las exploraciones CID utilizan energía de colisión normalizada de 35, y las exploraciones ETD utilizan un tiempo de activación de 100 ms con activación suplementaria permitida. Las señales mínimas para iniciar las exploraciones MS2 CID y ETD son 10.000, las señales mínimas para el inicio de exploraciones MS3 CID son 1000, y las anchuras de aislamiento para todas las exploraciones MS2 y MS3 son 3,0 *m/z*. La característica dinámica de exclusión del programa informático está habilitada con un número de repetición de 1, un tamaño de exclusión de la lista de 100 y una duración de la exclusión de 30 s. Se emplean listas de inclusión para especies diana reticuladas específicas para la recogida de las exploraciones MS2 ETD. Los archivos de datos separados para las exploraciones de MS2 y MS3 están creados por Bioworks 3.3 (Thermo Scientific) utilizando un análisis de estado de carga de ZSA. La adecuación de las exploraciones MS2 y MS3 a las secuencias de péptidos se realiza mediante Sequest (V27, Rev 12, Thermo Scientific). El análisis se realiza sin especificidad de la enzima, una tolerancia de masa de ion primario de 2,5, una tolerancia de masa de fragmento de 1,0, y una masa variable de +16 para los residuos de metionina oxidados. Los resultados se analizan después usando el programa de Scaffold (V3\_00\_08, Proteoma Software, Portland, OR) empleando probabilidades mínimas de péptidos y proteínas de 95 y 99%. Los péptidos de los resultados de MS3 se ordenan por el número de exploración, y los péptidos que contienen cisteína se identifican a partir de grupos de exploraciones MS3 producidas a partir de los cinco iones más intensos observados en las exploraciones MS2 ETD. Las identidades de los péptidos con cisteína que participan en las especies unidas por puentes disulfuro, se confirman adicionalmente por un examen manual de las masas de iones primarias, observadas en la exploración de vigilancia y la exploración MS2 ETD.

## 30 EJEMPLO 12

Aislamiento de productos de coexpresión procedentes del periplasma de células bacterianas, de esferoplastos y de células completas

El sistema de coexpresión inducible de la invención se puede utilizar para expresar productos génicos que se acumulan en diferentes compartimentos de la célula, tales como el citoplasma o el periplasma. Las células hospedadoras tales como *E. coli* o *S. cerevisiae* tienen una membrana celular externa o pared celular, y pueden formar esferoplastos cuando se retira la membrana o pared externa. Las proteínas coexpresadas producidas en tales hospedadores se pueden purificar específicamente a partir del periplasma o de esferoplastos o de células completas, utilizando el siguiente método (Schoenfeld, "Convenient, rapid enrichment of periplasmic and spheroplasmic protein fractions using the new PeriPreps™ Periplasting Kit", Epicentre Forum 1998 5(1): 5; véase [www.epibio.com/newsletter/f5\\_l/f5\\_lpp.asp](http://www.epibio.com/newsletter/f5_l/f5_lpp.asp)). Este método, que emplea el kit PeriPreps® Periplasting (Epicentre® Biotechnologies, Madison WI; protocolo disponible en [www.epibio.com/pdftechlit/107p10612.pdf](http://www.epibio.com/pdftechlit/107p10612.pdf)), está diseñado para *E. coli* y otras bacterias Gram-negativas, pero el enfoque general se puede modificar para otras células hospedadoras tales como *S. cerevisiae*.

1. El cultivo de células hospedadoras bacterianas se deja crecer solo hasta la fase logarítmica tardía, ya que los cultivos celulares más antiguos en fase estacionaria, muestran comúnmente una cierta resistencia al tratamiento con lisozima. Si la expresión de la proteína recombinante es excesiva, las células se pueden lizar prematuramente; por lo tanto, los cultivos celulares no se cultivan en medio rico o a temperaturas de crecimiento más elevadas que puedan inducir una síntesis de proteínas en exceso. A continuación, se induce la expresión de proteínas; las células deben estar en fase logarítmica o en la fase estacionaria temprana.

2. El cultivo celular se sedimenta por centrifugación a un mínimo de 1.000 x g durante 10 minutos a temperatura ambiente. Nota: las células deben ser frescas, no congeladas. El peso húmedo del sedimento celular se determina con el fin de calcular la cantidad de reactivos necesarios para este protocolo.

3. Las células se vuelven a suspender a fondo en un mínimo de 2 ml de tampón PeriPreps Periplasting (Tris-HCl 200 mM pH 7,5, 20% de sacarosa, EDTA 1 mM y 30 U/microlitro de lisozima Ready-Lyse) por cada gramo de células, ya sea por mezclado en vórtice o pipeteando la suspensión celular hasta que sea homogénea. Nota: una agitación excesiva puede causar una lisis prematura de los esferoplastos dando lugar a una contaminación de la fracción periplásmica con proteínas citoplasmáticas.

4. Incubar durante cinco minutos a temperatura ambiente. La lisozima Ready-Lyse está activa de forma óptima a temperatura ambiente. La lisis a temperaturas más bajas (0°C-4°C) requiere un tiempo de incubación adicional; a tales temperaturas, los tiempos de incubación se extienden de 2 a 4 veces.

5. Añadir 3 ml de agua purificada a 4°C por cada gramo de peso de sedimento celular original (Etapa 2) y mezclar mediante inversión.

6. Incubar durante 10 minutos en hielo.

5 7. Las células lisadas se sedimentan por centrifugación a un mínimo de 4.000 x g durante 15 minutos a temperatura ambiente.

8. El material sobrenadante que contiene la fracción periplásmica se transfiere a un tubo limpio.

9. Para degradar los ácidos nucleicos contaminantes, se añade opcionalmente la endonucleasa OmniCleave al tampón de lisis PeriPreps. La inclusión de una nucleasa generalmente mejora el rendimiento de la proteína y la facilidad de manejo de los lisados, pero la adición de una nucleasa no es deseable en algunos casos: por ejemplo, el uso de una nucleasa se debe evitar si la actividad nucleasa residual o una exposición transitoria al cofactor de magnesio va a interferir con ensayos o usos posteriores de la proteína purificada. La adición de EDTA al lisado para inactivar la endonucleasa OmniCleave, del mismo modo, puede interferir con el ensayo o el uso posterior de la proteína purificada. Si se va a añadir nucleasa, se diluyen 2 microlitros de endonucleasa OmniCleave y 10 microlitros de MgCl<sub>2</sub> 1,0 M hasta 1 ml con el tampón de lisis PeriPreps (Tris-HCl 10 mM pH 7,5, KCl 50 mM, EDTA 1 mM y 0,1% de desoxicolato) por cada mililitro de tampón de lisis necesario en la Etapa 10.

10 10. El sedimento se resuspende en 5 ml de tampón de lisis PeriPreps por cada gramo de peso de sedimento celular original.

11. El sedimento se incuba a temperatura ambiente durante 10 minutos (si se incluye, la actividad de la endonucleasa OmniCleave causará una disminución significativa de la viscosidad; la incubación continúa hasta que la suspensión celular tiene la consistencia del agua).

12. Los restos celulares se sedimentan por centrifugación a un mínimo de 4.000 x g durante 15 minutos a 4°C.

13. El material sobrenadante que contiene la fracción de esferoplastos se transfiere a un tubo limpio.

14. Si se había añadido endonucleasa OmniCleave al tampón de lisis PeriPreps, se añaden 20 microlitros de EDTA 500 mM por cada mililitro de la fracción esferoplástica resultante, para quelar el magnesio (la concentración final de EDTA en el lisado es 10 mM). Después de la hidrólisis de los ácidos nucleicos con endonucleasa OmniCleave, los lisados pueden contener cantidades sustanciales de monooligonucleótidos u oligonucleótidos. La presencia de estos productos de degradación puede afectar adicionalmente al procesamiento del lisado: por ejemplo, los nucleótidos pueden disminuir la capacidad de unión de resinas de intercambio aniónico mediante una interacción con la resina.

El protocolo anterior se puede utilizar para preparar proteína celular total con las siguientes modificaciones. Las células sedimentadas en la Etapa 2 pueden ser frescas o congeladas; en la Etapa 4, las células se incuban durante 15 minutos; las Etapas 5 a 8 se omiten; en la Etapa 10, se añaden 3 ml de tampón de lisis PeriPreps por cada gramo de peso de sedimento celular original.

Después de la preparación de muestras de proteínas periplásmicas o esferoplásticas o de células completas, las muestras se pueden analizar mediante cualquiera de una variedad de métodos de caracterización y/o cuantificación de proteínas. En un ejemplo, el fraccionamiento con éxito de proteínas periplásmicas y esferoplásticas se confirma mediante el análisis de una parte alícuota de ambas fracciones periplásmicas y esferoplásticas mediante SDS-PAGE (dos microlitros de cada fracción es generalmente suficiente para la visualización con una tinción con Coomassie Brilliant Blue). La presencia de proteínas únicas o el enriquecimiento de proteínas específicas en una fracción dada, indica un fraccionamiento exitoso. Por ejemplo, si la célula hospedadora contiene un plásmido con un número elevado de copias con el marcador de resistencia a la ampicilina, a continuación, la presencia de β-lactamasa (31,5 kDa) principalmente en la fracción periplásmica, indica un fraccionamiento exitoso. Otras proteínas de *E. coli* que se encuentran en el espacio periplásmico incluyen fosfatasa alcalina (50 kDa) y factor de elongación Tu (43 kDa). La cantidad de proteína que se encuentra en una fracción dada, se puede cuantificar utilizando cualquiera entre una serie de métodos (por ejemplo, SDS-PAGE y análisis densitométrico de bandas de proteínas teñidas o marcadas, recuento de centelleo de las proteínas radiomarcadas, ensayo de inmunoabsorción ligado a enzimas (ELISA), o ensayo de centelleo de proximidad, entre otros métodos). La comparación de las cantidades de una proteína que se encuentra en la fracción periplásmica en comparación con la fracción esferoplástica, indica el grado en que se ha exportado la proteína desde el citoplasma al periplasma.

### EJEMPLO 13

50 Determinación de la similitud de la secuencia de polinucleótidos o de aminoácidos

El porcentaje de identidad de la secuencia de polinucleótidos o de la secuencia de aminoácidos se define como el número de símbolos alineados, es decir, nucleótidos o aminoácidos, que son idénticos en las dos secuencias alineadas, dividido por el número total de símbolos en la alineación de las dos secuencias, incluyendo los huecos. El grado de similitud (porcentaje de identidad) entre dos secuencias se puede determinar alineando las secuencias

- utilizando el método de alineamiento global de Needleman y Wunsch (J. Mol. Biol. 48: 443, 1970), implementado por el Centro Nacional de Información sobre Biotecnología (NCBI) en la herramienta de alineación de secuencias global Needleman-Wunsch, disponible a través de la página web [blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi). En una realización, los parámetros de alineación de Needleman y Wunsch se establecen para los valores por defecto (Puntuaciones de Emparejamiento/No emparejamiento de 2 y -3, respectivamente, y los costes por huecos por su existencia y extensión de 5 y 2, respectivamente). Otros programas utilizados por los expertos en la técnica de comparación de secuencias, también se pueden utilizar para alinear secuencias, tales como, por ejemplo, la herramienta de búsqueda de alineamiento local básico o programa Blast® (Altschul *et al.*, "Basic local alignment search tool", J Mol Biol 1990 Oct 5; 215(3): 403-410), tal y como se implementa en el NCBI, usando los ajustes de parámetros por defecto que se describen en el sitio web [blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi). El algoritmo BLAST tiene varios parámetros opcionales, incluyendo dos que se pueden utilizar de la siguiente manera: (A) inclusión de un filtro para enmascarar segmentos de la secuencia de consulta que tienen baja complejidad composicional o segmentos que consisten en repeticiones internas de corta periodicidad, que preferiblemente no se utilizan o están desactivados, y (B) un umbral de significación estadística para informar de emparejamientos frente a secuencias de bases de datos, llamado el "Expect" o puntuación E (del inglés "E-score") (la probabilidad esperada de emparejamientos que se encuentran simplemente por casualidad, si la significación estadística atribuida a un emparejamiento es superior a este umbral de puntuación E, no se informará del emparejamiento). Si este valor 'Expect' o puntuación E se ajusta desde el valor predeterminado (10), los valores umbrales preferidos son 0,5, o con el fin de aumentar la preferencia, 0,25, 0,1, 0,05, 0,01, 0,001, 0,0001, 0,00001 y 0,000001.
- En la práctica de la presente invención, se utilizan opcionalmente muchas técnicas convencionales en biología molecular, microbiología y tecnología de ADN recombinante. Tales técnicas convencionales se refieren a vectores, células hospedadoras y métodos recombinantes. Estas técnicas son bien conocidas y se explican en, por ejemplo, Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Mc, San Diego, CA; Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (3ª Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2000; y Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, compiladores, Current Protocols, a joint venture between Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (complementado hasta el 2006). Otras referencias útiles, por ejemplo, para el aislamiento y cultivo celular y para un aislamiento posterior de ácidos nucleicos o proteínas, incluyen Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, tercera edición, Wiley-Liss, New York y las referencias citadas en el mismo; Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. Nueva York, NY; Gamborg y Phillips (compiladores) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York); y Atlas y Parks (compiladores) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL. Métodos para preparar ácidos nucleicos (por ejemplo, mediante amplificación *in vitro*, purificación a partir de células o síntesis química), métodos para la manipulación de ácidos nucleicos (por ejemplo, mediante mutagénesis dirigida al sitio, digestión con enzimas de restricción, ligación, etc.), y diversos vectores, líneas celulares y similares, útiles en la manipulación y la preparación de ácidos nucleicos, se describen en las referencias anteriores. Además, esencialmente, cualquier polinucleótido (incluyendo polinucleótidos marcados o biotinilados) se puede ordenar de forma personalizada o convencional a partir de cualquiera entre una variedad de fuentes comerciales.
- La presente invención se ha descrito en términos de realizaciones particulares, encontradas o propuestas para comprender ciertos modos para la práctica de la invención. Los expertos en la técnica apreciarán que, a de cara a la presente descripción, se pueden realizar numerosas modificaciones y cambios en las realizaciones particulares ejemplificadas, sin apartarse del alcance pretendido de la invención.

#### SECUENCIAS PRESENTADAS EN EL LISTADO DE SECUENCIAS

SEQ ID NO:	Longitud:	Tipo:	Organismo:	Descripción; "Otra información"
1	1526	ADN	<i>Mus musculus</i>	Secuencia codificadora para la cadena pesada de IgG1 de ratón anti-CD19 humano
2	464	PRT	<i>Mus musculus</i>	Secuencia de aminoácidos de longitud completa de la cadena pesada de IgG1 de ratón anti-CD19 humano
3	958	ADN	<i>Mus musculus</i>	Secuencia codificadora para la cadena ligera de IgG1 de ratón anti-CD19 humano
4	239	PRT	<i>Mus musculus</i>	Secuencia de aminoácidos de longitud completa de la cadena ligera de IgG1 de ratón anti-CD19 humano
5	1429	ADN	Secuencia Artificial	Secuencia codificadora optimizada para la cadena pesada de IgG1 de ratón anti-CD19

ES 2 621 320 T3

SEQ ID NO:	Longitud:	Tipo:	Organismo:	Descripción; "Otra información"
				humano
6	757	ADN	Secuencia Artificial	Secuencia codificadora optimizada para la cadena ligera de IgG1 de ratón anti-CD19 humano
7	5874	ADN	Secuencia Artificial	vector pPRO33
8	36	ADN	Secuencia Artificial	cebador directo IgG1HC Fc
9	32	ADN	Secuencia Artificial	cebador inverso IgG1HC Fc
10	19	ADN	Secuencia Artificial	cebador directo pBAD24/33
11	17	ADN	Secuencia Artificial	cebador inverso nº 1 pBAD24/33
12	17	ADN	Secuencia Artificial	cebador inverso nº 2 pBAD24/33
13	370	PRT	Secuencia Artificial	Secuencia de aminoácidos de MnP-H4 de <i>P. chrysosporium</i> , sin péptido señal
14	1113	ADN	Secuencia Artificial	Secuencia codificadora optimizada para MnP-H4 de <i>P. chrysosporium</i> , sin péptido señal
15	370	PRT	Secuencia Artificial	Secuencia de aminoácidos de MnP-H4 A60C/A75C de <i>P. chrysosporium</i> , sin péptido señal
16	1113	ADN	Secuencia Artificial	Secuencia codificadora optimizada para MnP-H4 A60C/A75C de <i>P. chrysosporium</i> , sin péptido señal
17	660	PRT	<i>E. coli</i> O157:H7	Secuencia de aminoácidos de O157:H7 cepa EC4113 ChuA de <i>E. coli</i>
18	1276	ADN	Secuencia Artificial	Estructura artificial de expresión MnP-H4
19	2056	ADN	Secuencia Artificial	Estructura artificial de expresión ChuA
20	3326	ADN	Secuencia Artificial	Estructura artificial de expresión MnP-H4 ChuA
21	486	PRT	Secuencia Artificial	Secuencia de aminoácidos de PDI de <i>Humicola insolens</i> , sin péptido señal
22	1487	ADN	Secuencia Artificial	Estructura artificial de expresión PDI
23	359	PRT	Secuencia Artificial	Secuencia de aminoácidos de MnP-H4 de <i>P. chrysosporium</i> , madura y "totalmente truncada"
24	25	ADN	Secuencia Artificial	Cebador directo MnP-H4_FT NcoI
25	27	ADN	Secuencia Artificial	Cebador inverso MnP-H4_FT Sall
26	24	ADN	Secuencia Artificial	Cebador directo ChuA Sall
27	31	ADN	Secuencia Artificial	Cebador inverso ChuA HindIII
28	7768	ADN	Secuencia Artificial	Estructura artificial de expresión pBAD24-MnP_FT-ChuA
29	7316	ADN	Secuencia Artificial	Estructura artificial de expresión pPRO33-PDI
30	451	PRT	Secuencia Artificial	Cadena pesada de Infliximab quimérico

ES 2 621 320 T3

SEQ ID NO:	Longitud:	Tipo:	Organismo:	Descripción; "Otra información"
				(dominio variable murino, dominio constante humano)
31	215	PRT	Secuencia Artificial	Cadena ligera de Infliximab quimérico (dominio variable murino, dominio constante humano)
32	5875	ADN	Secuencia Artificial	Estructura artificial de expresión pBAD24-Infliximab_HC
33	6503	ADN	Secuencia Artificial	Estructura artificial de expresión pPRO33-Infliximab_LC
34	6545	ADN	Secuencia Artificial	Plásmido pJ1231-03C
35	6009	ADN	Secuencia Artificial	Plásmido pJ1234-03C
36	42	ADN	Secuencia Artificial	Cebador Bsal-AraC-MnP-H4_FT
37	63	ADN	Secuencia Artificial	Cebador inverso MnP-H4_FT-6xHis
38	38	ADN	Secuencia Artificial	Cebador directo MnP-H4_FT-6xHis
39	66	ADN	Secuencia Artificial	Cebador ChuA-6xHis-Bsal
40	4569	ADN	Secuencia Artificial	Producto de PCR AraC-MnP-H4_FT-ChuA
41	42	ADN	Secuencia Artificial	Cebador Bsal-PrpR-PDI
42	55	ADN	Secuencia Artificial	Cebador PDI-5xHis-Bsal
43	3326	ADN	Secuencia Artificial	Producto de PCR PrpR-PDI
44	44	ADN	Secuencia Artificial	Cebador directo Bsal-AraC-HC
45	46	ADN	Secuencia Artificial	Cebador inverso HC
46	46	ADN	Secuencia Artificial	Cebador directo HC
47	45	ADN	Secuencia Artificial	Cebador inverso HC-Bsal
48	43	ADN	Secuencia Artificial	Cebador directo Bsal-PrpR-LC
49	50	ADN	Secuencia Artificial	Cebador inverso LC
50	50	ADN	Secuencia Artificial	Cebador directo LC
51	43	ADN	Secuencia Artificial	Cebador inverso LC-Bsal
52	2615	ADN	Secuencia Artificial	Producto de PCR AraC-HC_NS
53	2584	ADN	Secuencia Artificial	Producto de PCR PrpR-LC_NS

**REIVINDICACIONES**

1. Una célula hospedadora que comprende dos o más tipos de estructuras artificiales de expresión,  
en donde la estructura artificial de expresión de cada tipo comprende un promotor inducible y una secuencia de  
polinucleótidos que codifica un producto génico que se va a transcribir desde el promotor inducible; y
- 5 al menos uno de dichos promotores inducibles es sensible a un inductor que es diferente del inductor de otro de  
dichos promotores inducibles; y en donde cada promotor inducible no es un promotor inducible con lactosa; y
- en donde la célula hospedadora tiene un nivel reducido de función génica de al menos un gen que codifica una  
proteína que metaboliza un inductor de al menos uno de dichos promotores inducibles; y
- 10 al menos uno de dichos productos génicos se selecciona a partir del grupo que consiste en un producto génico que  
forma un multímero con otro de dichos productos génicos, y un polipéptido que forma al menos dos enlaces  
disulfuro.
2. La célula hospedadora según la reivindicación 1, en donde al menos un promotor inducible es un promotor  
inducible con propionato.
- 15 3. La célula hospedadora según la reivindicación 1, en donde al menos un promotor inducible es un promotor  
inducible con L-arabinosa.
4. La célula hospedadora según la reivindicación 1, en donde al menos un promotor inducible se selecciona a partir  
del grupo que consiste en: el promotor *araBAD*, el promotor *prpBCDE*, el promotor *rhaSR* y el promotor *xyIA*.
5. La célula hospedadora según la reivindicación 1, en donde al menos un producto génico se selecciona a partir del  
grupo que consiste en: (a) una cadena pesada de inmunoglobulina; (b) una cadena ligera de inmunoglobulina; y (c)  
20 un fragmento de cualquiera de (a) - (b).
6. La célula hospedadora según la reivindicación 1, en donde al menos un producto génico es un polipéptido que  
carece de un péptido señal.
7. La célula hospedadora según la reivindicación 1, en donde al menos un producto génico es un polipéptido que  
carece de un péptido señal y forma al menos tres enlaces disulfuro.
- 25 8. La célula hospedadora según la reivindicación 1, en donde el gen que codifica una proteína que metaboliza un  
inductor de al menos uno de dichos promotores inducibles se selecciona a partir del grupo que consiste en *araA*,  
*araB*, *araD*, *prpB*, *prpD*, *rhaA*, *rhaB*, *rhaD*, *xyIA* y *xyIB*.
9. La célula hospedadora según la reivindicación 1, en donde al menos uno de dichos productos génicos comprende  
una cadena polipeptídica de insulina madura.
- 30 10. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora tiene una alteración de la  
función génica de al menos un gen que codifica una proteína transportadora para un inductor de al menos uno de  
dichos promotores inducibles.
11. La célula hospedadora según la reivindicación 10, en donde el gen que codifica una proteína transportadora se  
selecciona a partir del grupo que consiste en *araE*, *araF*, *araG*, *araH*, *rhaT*, *xyIF*, *xyIG* y *xyIH*.
- 35 12. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora tiene un nivel reducido de  
función génica de al menos un gen que codifica una proteína implicada en la biosíntesis de un inductor de al menos  
uno de dichos promotores inducibles.
13. La célula hospedadora según la reivindicación 12, en donde el gen que codifica una proteína implicada en la  
biosíntesis de un inductor de al menos uno de dichos promotores inducibles se selecciona a partir del grupo que  
40 consiste en: *scpA/sbm*, *argK/ygfD*, *scpB/ygfG*, *scpC/ygfH*, *rmlA*, *rmlB*, *rmlC* y *rmlD*.
14. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora tiene una función génica  
alterada de un gen que afecta al entorno de reducción/oxidación del citoplasma de la célula hospedadora.
15. La célula hospedadora según la reivindicación 14, en donde el gen que afecta al entorno de reducción/oxidación  
del citoplasma de la célula hospedadora se selecciona a partir del grupo que consiste en *gor* y *gshB*.
- 45 16. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora tiene un nivel reducido de  
función génica de un gen que codifica una reductasa.
17. La célula hospedadora según la reivindicación 16, en donde el gen que codifica una reductasa en *trxB*.

18. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora comprende al menos una estructura artificial de expresión que codifica al menos una proteína isomerasa de enlace disulfuro.
19. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora comprende al menos un polinucleótido que codifica una forma de DsbC que carece de un péptido señal.
- 5 20. La célula hospedadora según la reivindicación 1, en donde la célula hospedadora comprende al menos un polinucleótido que codifica Erv1p.
21. La célula hospedadora según cualquier reivindicación precedente, en donde la célula hospedadora es una célula procariota.
22. La célula hospedadora según la reivindicación 21, en donde la célula hospedadora es *E. coli*.
- 10 23. Un método para producir un producto génico, comprendiendo el método hacer crecer un cultivo de la célula hospedadora según cualquier reivindicación precedente y añadir un inductor de al menos un promotor inducible al cultivo.
24. Un kit que comprende la célula hospedadora según la reivindicación 1.

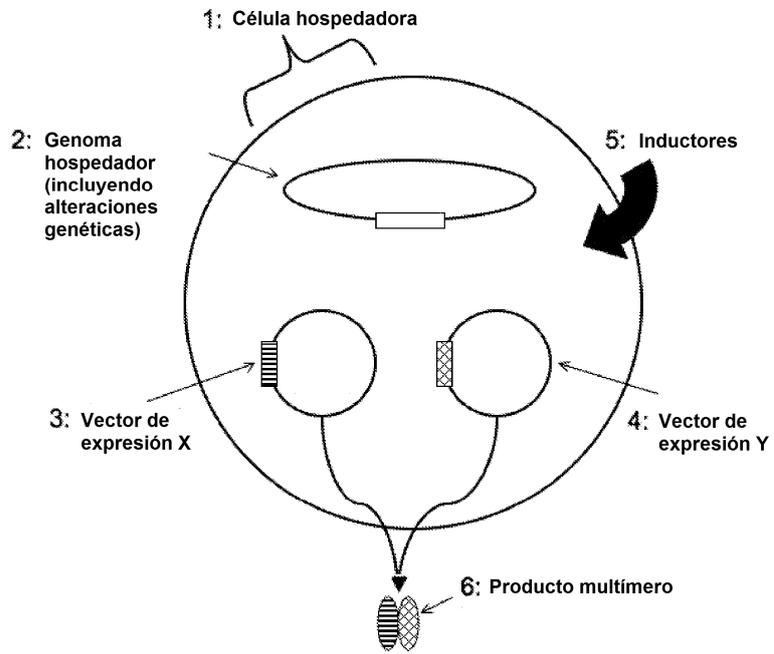


FIG. 1

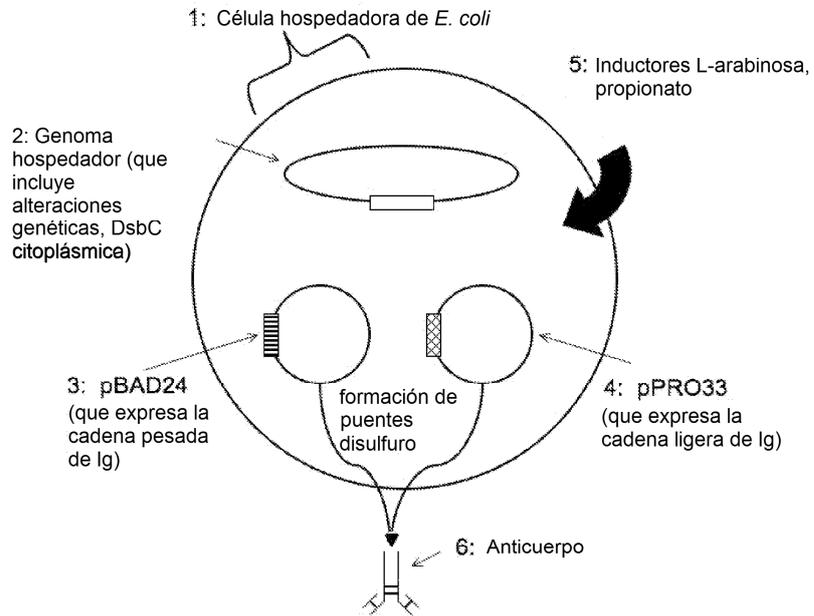


FIG. 2

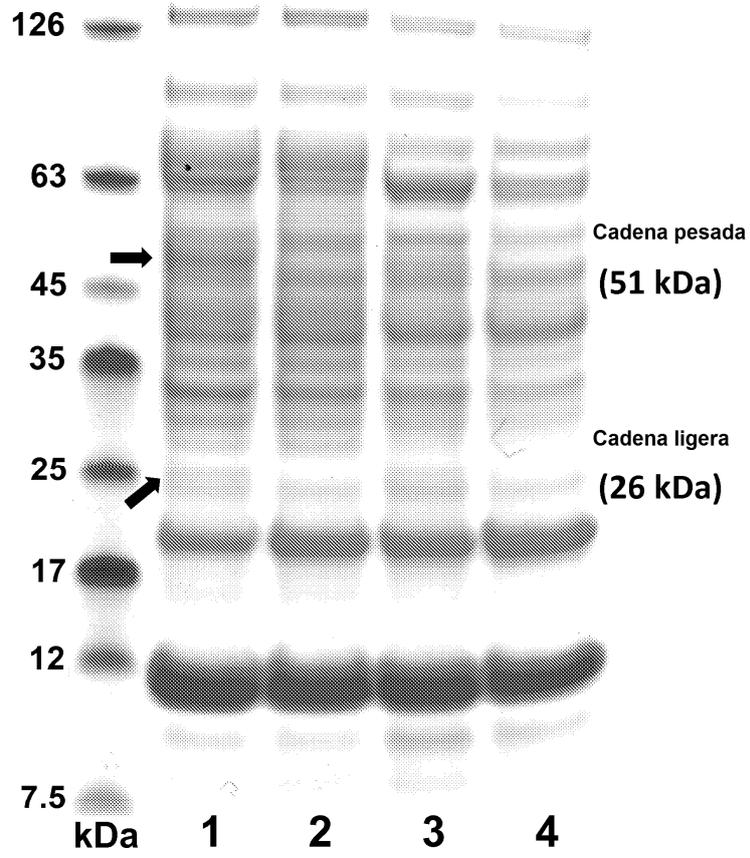


FIG. 3

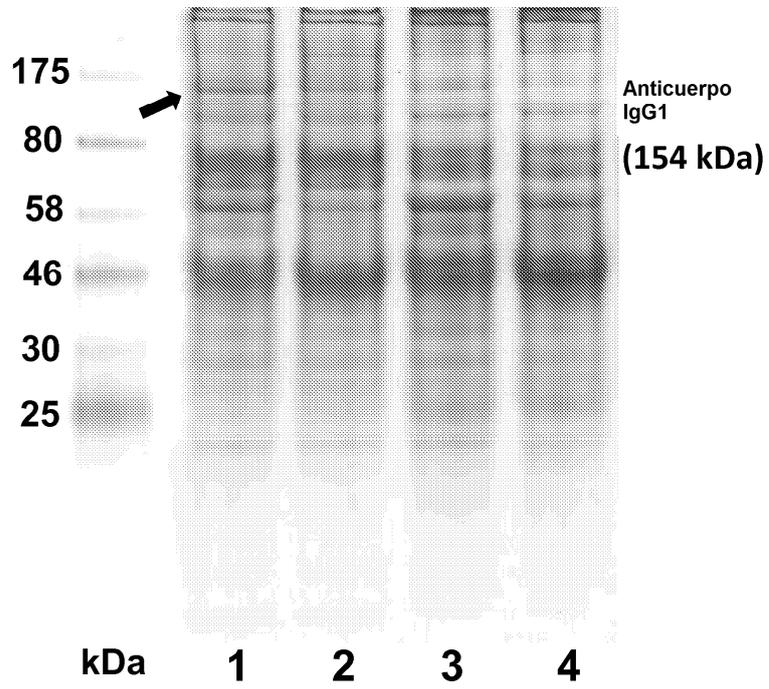


FIG. 4

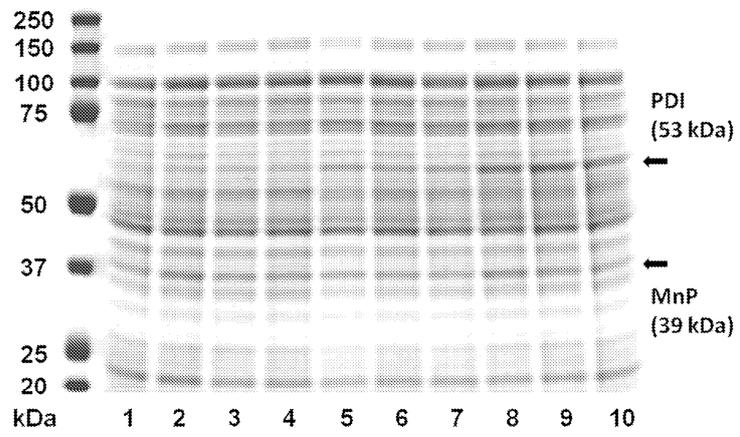


FIG. 5

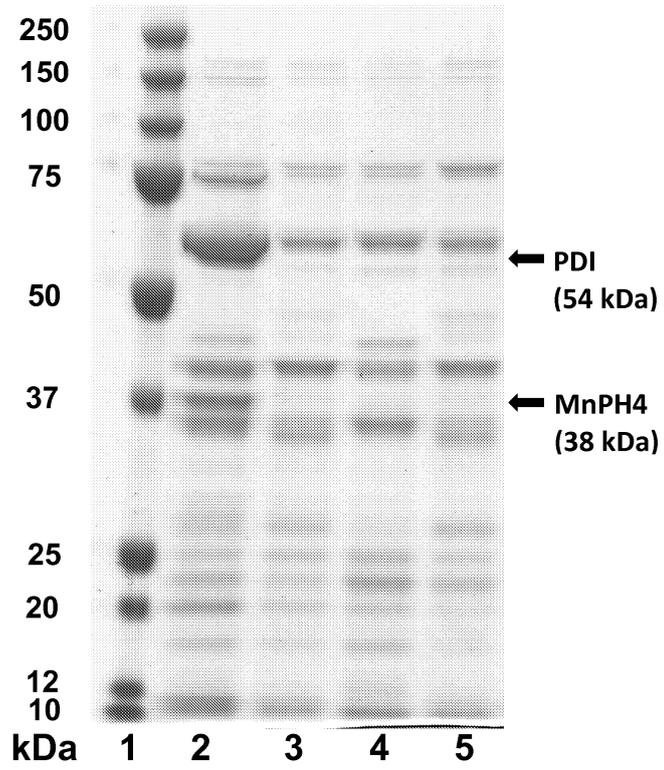


FIG. 6