(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 791 364**

(51) Int. CI.:

C12N 15/82 (2006.01)
C07K 14/805 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **11.09.2014 PCT/US2014/055227**

(87) Fecha y número de publicación internacional: **19.03.2015 WO15038796**

(96) Fecha de presentación y número de la solicitud europea: **11.09.2014 E 14844701 (4)**

(97) Fecha y número de publicación de la concesión europea: **19.02.2020 EP 3044320**

(54) Título: **Secreción de polipéptidos que contienen hemo**

(30) Prioridad:

11.09.2013 US 201361876676 P
25.11.2013 US 201361908689 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
04.11.2020

(73) Titular/es:

IMPOSSIBLE FOODS INC. (100.0%)
525 Chesapeake Drive
Redwood City, CA 94063, US

(72) Inventor/es:

FRASER, RACHEL;
DAVIS, SIMON CHRISTOPHER y
BROWN, PATRICK O'RIELLY

(74) Agente/Representante:

DEL VALLE VALIENTE, Sonia

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

Secreción de polipéptidos que contienen hemo

Campo técnico

Esta invención se refiere a métodos y material para producir polipéptidos que contienen hemo, y más particularmente, para producir polipéptidos que contienen hemo en células bacterianas recombinantes tales como células de *Bacillus* o en plantas recombinantes o células de planta.

Antecedentes

Existe la continua necesidad de métodos para producir proteínas a gran escala con fines industriales y alimentarios. Las especies de *Bacillus* pueden usarse en la producción de enzimas industriales tales como lipasas y proteasas. Además, pueden producirse varios aditivos alimentarios tales como glucoamilasa, lipasas y amilasas en estos huéspedes, lo que proporciona un largo historial de uso seguro en la industria alimentaria. La especie *Bacillus* puede secretar altos niveles de proteína en los medios que rodean a las bacterias. También pueden usarse especies de plantas, tales como *Nicotiana tabacum* o *Glycine max*, para la producción de proteínas.

Los polipéptidos que contienen hemo pueden ser difíciles de secretar, ya que el cofactor debe insertarse en el polipéptido y permanecer asociado con el polipéptido durante todo el proceso de secreción en su configuración original. La especie *Bacillus* puede usar dos sistemas diferentes para secretar proteínas (SEC y TAT). La ruta SEC despliega la proteína a medida que pasa a través de la membrana celular. El sistema TAT puede secretar las proteínas en el estado plegado. Sin embargo, no está claro si una hemoproteína recombinante que contiene un grupo hemo unido de manera no covalente puede expresarse, secretarse y plegarse de manera apropiada por el sistema de *Bacillus*, hasta que se realice con éxito.

El documento WO98/12913 A1 da a conocer la modificación por ingeniería genética de plantas para una utilización y asimilación potenciada de oxígeno, en particular plantas transgénicas modificadas por ingeniería genética para expresar proteínas de globina tales como hemoglobina, mioglobina y hemoproteínas.

Sumario

La presente invención se refiere a los elementos tal como se formulan en las reivindicaciones 1-12.

Este documento da a conocer una célula de bacteria recombinante (por ejemplo, una célula de *Bacillus* tal como una célula de *Bacillus subtilis*, *Bacillus megaterium* o *Bacillus licheniformis*) que puede secretar un polipéptido que contiene hemo. La célula incluye al menos un ácido nucleico exógeno, comprendiendo el ácido nucleico exógeno unas secuencias de ácido nucleico primera y segunda, en el que la primera secuencia de ácido nucleico codifica para un péptido señal y la segunda secuencia de ácido nucleico codifica para un polipéptido que contiene hemo, en el que la primera y la segunda secuencias de ácido nucleico se ligan operativamente para producir un polipéptido de fusión que comprende el péptido señal y el polipéptido que contiene hemo. El ácido nucleico exógeno también puede incluir una tercera secuencia de ácido nucleico que codifica para una etiqueta tal como una etiqueta de afinidad. La célula puede secretar el polipéptido que contiene hemo de la célula, y tras la secreción, el péptido señal se retira del polipéptido que contiene hemo. El péptido señal puede comprender o consistir en una secuencia de aminoácidos que tiene al menos el 60% de identidad con respecto a un péptido señal expuesto en la SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91 ó 93. Por ejemplo, el péptido señal puede comprender o consistir en una secuencia de aminoácidos que tiene al menos el 60% de identidad de secuencia de aminoácidos con respecto a la secuencia de aminoácidos expuesta en la SEQ ID NO: 55 o a los residuos 1-52 de la SEQ ID NO: 55

Este documento también presenta un método para producir un polipéptido que contiene hemo. El método incluye cultivar una célula de bacteria recombinante (por ejemplo, una célula de *Bacillus* tal como una célula de *Bacillus subtilis*, *Bacillus megaterium* o *Bacillus licheniformis*) en un medio de cultivo en condiciones que permitan al polipéptido que contiene hemo que se secrete en el medio de cultivo, comprendiendo la célula de bacteria recombinante al menos un ácido nucleico exógeno, comprendiendo el ácido nucleico exógeno unas secuencias de ácido nucleico primera y segunda, en el que la primera secuencia de ácido nucleico codifica para un péptido señal y la segunda secuencia de ácido nucleico codifica para un polipéptido que contiene hemo, en el que las secuencias de ácido nucleico primera y segunda se ligan operativamente para producir un polipéptido de fusión que comprende el péptido señal y el polipéptido que contiene hemo, y en el que, tras la secreción del polipéptido de fusión de la célula en el medio de cultivo, el péptido señal se retira del polipéptido que contiene hemo. El método puede incluir además recuperar el polipéptido que contiene hemo del medio de cultivo. El péptido señal puede comprender o consistir en una secuencia de aminoácidos que tiene al menos el 60% de identidad con respecto a un péptido señal expuesto en la SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91 ó 93. Por ejemplo, el péptido señal puede comprender o consistir en una secuencia de aminoácidos que tiene al menos el 60% de identidad de secuencia de aminoácidos con respecto a la secuencia de aminoácidos

expuesta en la SEQ ID NO: 55 o a los residuos 1-52 de la SEQ ID NO: 55.

Adicionalmente, este documento da a conocer una planta recombinante o célula vegetal (una planta o célula de planta *Glycine max, Zea mays, Hordeum vulgare* o *Arabidopsis thaliana*) que produce un polipéptido que contiene hemo. La planta o célula vegetal puede incluir al menos un ácido nucleico exógeno que codifica para un polipéptido que contiene hemo, en la que la planta o célula vegetal es de una especie distinta de *Nicotiana.* El ácido nucleico exógeno puede incluir además un elemento de control regulador tal como un promotor (por ejemplo, un promotor específico de tejido tal como hojas, raíces, tallos o semillas). El ácido nucleico exógeno también puede codificar para un péptido señal que selecciona como diana el polipéptido que contiene hemo a una ubicación subcelular tal como un esferosoma, una vacuola, un plástido (por ejemplo, cloroplasto) u otro orgánulo.

Este documento también presenta un método de producción de un polipéptido que contiene hemo seleccionado del grupo que consiste en una leghemoglobina, una eritrocruorina, una hemoglobina no simbiótica, una flavohemoglobina, una protoglobina, una cianoglobina, una globina I de Hell's gate, una hemoglobina bacteriana, una mioglobina de ciliado, una protoglobina, globina truncada 2/2, HbN, HbO y Glb3. El método incluye hacer crecer una planta recombinante (por ejemplo, una planta *Glycine max*, *Zea mays*, *Hordeum vulgare* o *Arabidopsis thaliana*), comprendiendo la planta recombinante al menos un ácido nucleico exógeno que codifica para el polipéptido que contiene hemo, en el que la planta es de una especie distinta de *Nicotiana,* y purificar el polipéptido que contiene hemo de un tejido de la planta.

Este documento da a conocer además un vector que incluye una secuencia de polinucleótido que codifica para un polipéptido que contiene hemo; y una secuencia de polinucleótido que codifica para un péptido señal, en la que el péptido señal comprende o consiste en una secuencia de aminoácidos que tiene al menos el 60% de identidad de secuencia de aminoácidos con respecto a un péptido señal enumerado en la tabla 1. Por ejemplo, el péptido señal puede incluir una secuencia de aminoácidos que tiene al menos el 60% de identidad de secuencia de aminoácidos con respecto a la secuencia de aminoácidos expuesta en la SEQ ID NO: 55 o a los residuos 1-52 de la SEQ ID NO: 55. En algunos casos, el péptido señal comprende o consiste en la secuencia de aminoácidos de los residuos 1-52 de la SEQ ID NO: 55. La secuencia de polinucleótido que codifica para el polipéptido que contiene hemo puede ligarse operativamente a un promotor.

Aún más, este documento da a conocer una composición que incluye un polipéptido purificado que contiene hemo; y una célula de *Bacillus* recombinante (por ejemplo, una célula de *Bacillus subtilis*, *Bacillus megaterium* o *Bacillus licheniformis*) o una célula vegetal recombinante distinta de una célula de la planta *Nicotiana* (por ejemplo, un componente que no se produce de manera natural de una célula de *Bacillus* recombinante o una célula vegetal recombinante). En algunos casos, el polipéptido que contiene hemo no se produce de manera natural en la célula huésped. La célula vegetal puede ser, por ejemplo, una célula de la planta *Glycine max*, una célula de planta la *Zea mays* o una célula de la planta *Arabidopsis thaliana*. La composición puede incluir al menos 1 parte por mil millones de dicho componente de la célula o como máximo el 1% (p/p) del componente de la célula.

Este documento también da a conocer un vector que comprende un secuencia de polinucleótido que codifica para un polipéptido que contiene hemo, un péptido señal; y una etiqueta, en el que la expresión de la secuencia de polinucleótido en una célula huésped produce una proteína de fusión que contiene el polipéptido que contiene hemo, el péptido señal y la etiqueta (por ejemplo, una etiqueta de afinidad tal como un etiqueta 6-histidina o una etiqueta detectable), y organismos genéticamente modificados que contienen un vector de este tipo. El vector puede incluir además una secuencia de polinucleótido que codifica para al menos uno de: a) un ligador de aminoácidos entre la secuencia que codifica para la etiqueta y la secuencia que codifica para el polipéptido que contiene hemo; y b) un ligador de aminoácidos entre la secuencia que codifica para el péptido señal y la secuencia que codifica para el polipéptido que contiene hemo.

Este documento da a conocer un método para secretar un polipéptido que contiene hemo de una bacteria (por ejemplo, una célula de *Bacillus* tal como una célula de *Bacillus subtilis*, *Bacillus megaterium* o *Bacillus licheniformis*) que incluye cultivar una bacteria recombinante en condiciones que permitan al polipéptido que contiene hemo que se secrete de la bacteria, comprendiendo la bacteria recombinante un ácido nucleico exógeno que codifica para el polipéptido que contiene hemo, un péptido señal y una etiqueta.

Este documento también da a conocer un polipéptido de fusión purificado que incluye un polipéptido que contiene hemo y una etiqueta. El polipéptido puede incluir además un ligador entre la etiqueta y el polipéptido que contiene hemo. En algunos casos, la etiqueta puede ubicarse en el extremo C-terminal del polipéptido que contiene hemo o bien directamente unido al extremo C-terminal o bien a través de un ligador.

A menos que se especifique lo contrario, en cualquiera de los métodos, las composiciones, las células bacterianas recombinantes, plantas recombinantes o células vegetales, o los vectores, el polipéptido que contiene hemo puede seleccionarse del grupo que consiste en una androglobina, una citoglobina, globina E, globina X, globina Y, una hemoglobina, una mioglobina, una leghemoglobina, una eritrocruorina, una hemoglobina beta, una hemoglobina alfa, una hemoglobina no simbiótica, una flavohemoglobina, una protoglobina, una cianoglobina, una globina I de Hell's gate, una hemoglobina bacteriana, una mioglobina de ciliado, una histoglobina, una neuroglobina, una protoglobina y

una globina truncada (por ejemplo, globina truncada 2/2, HbN, HbO o Glb3). Por ejemplo, el polipéptido que contiene hemo puede tener al menos el 60% de identidad de secuencia con respecto a una secuencia de aminoácidos expuesta en las SEQ ID NO: 1-31.

En el presente documento también se da a conocer un vector que comprende una secuencia de polinucleótido que codifica para un polipéptido que contiene hemo, por ejemplo, una globina, y una secuencia de polinucleótido que codifica para un péptido señal. En algunas variantes, el péptido señal es para una ruta secretora. En algunas de tales variantes, el péptido señal puede denominarse péptido señal o péptido señal de secreción. En algunas variantes, el péptido señal se refiere a dicho polipéptido que contiene hemo, por ejemplo, una globina, en una ruta secretora. En algunas variantes, el péptido señal comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a un péptido señal enumerado en la tabla 1. En algunas variantes, el péptido señal comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a un péptido señal de PhoD (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En alguna variante, el péptido señal es un péptido señal de PhoD. En algunas variantes, la secuencia de polinucleótido que codifica para un polipéptido que contiene hemo, por ejemplo, una globina, se liga operativamente a un promotor. En algunas variantes, el polipéptido que contiene hemo se selecciona del grupo que consiste en: androglobina, citoglobina, globina E, globina X, globina Y, hemoglobina, mioglobina, leghemoglobina, eritrocruorina, hemoglobina beta, hemoglobina alfa, hemoglobina no simbiótica, flavohemoglobina, protoglobina, cianoglobina, hemoglobina no simbiótica, globina I de Hell's gate, hemoglobina bacteriana, mioglobina de ciliado, flavohemoglobina, histoglobina, neuroglobinas, protoglobina, globina truncada 2/2, HbN, HbO y Glb3. En el presente documento también se da a conocer un organismo genéticamente modificado que comprende un vector que comprende una secuencia de polinucleótido que codifica para un polipéptido que contiene hemo, por ejemplo, una globina, y una secuencia de polinucleótido que codifica para un péptido señal. En algunas variantes, el organismo genéticamente modificado es una especie gram positiva de bacterias. En algunas variantes, el organismo genéticamente modificado es de la especie *Bacillus*. En algunas variantes, el organismo genéticamente modificado se selecciona del grupo que consiste en *Bacillus subtilis*, *Bacillus megaterium* y *Bacillus licheniformis.* En algunas variantes, el organismo genéticamente modificado es de la especie *Nicotiana*. En algunas variantes, el organismo genéticamente modificado es *Nicotiana tabacum.*

En un aspecto, la divulgación proporciona una composición que no se produce de manera natural que es un sucedáneo de carne, que comprende un polipéptido purificado que contiene hemo, por ejemplo, una globina, y un componente de una célula huésped que incluye una pared celular, un compartimento subcelular (tal como complejo de Golgi, retículo endoplasmático o núcleo), ácido nucleico, proteína, ADN genómico y/o una membrana plasmática. El polipéptido que contiene hemo es una globina que no se produce de manera natural en dicha célula huésped. La célula huésped se selecciona del grupo que consiste en célula de la planta *Glycine max*, célula de la planta *Zea mays* y célula de la planta *Arabidopsis thaliana*. En algunas realizaciones, la composición comprende al menos 1 parte por mil millones de dicha parte de la célula huésped. En algunas realizaciones, la composición comprende como máximo el 1% (p/p) de dicha parte de una célula huésped. El polipéptido que contiene hemo se selecciona del grupo que consiste en: leghemoglobina, eritrocruorina, hemoglobina no simbiótica, flavohemoglobina, protoglobina, cianoglobina, globina I de Hell's gate, hemoglobina bacteriana, mioglobina de ciliado, protoglobina, globina truncada 2/2, HbN, HbO y Glb3. En algunas realizaciones, el polipéptido que contiene hemo comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a una secuencia de aminoácidos expuesta en las SEQ ID NO: 1-17 ó 21-31 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas realizaciones, el polipéptido que contiene hemo comprende una etiqueta, por ejemplo, se une de manera covalente a una etiqueta, por ejemplo, en el extremo C-terminal o N-terminal. En algunas realizaciones, un producto cárnico consumible comprende un polipéptido que contiene hemo tal como se describe en el presente documento. En algunas realizaciones, un producto cárnico consumible comprende al menos el 0,001% (p/p) del polipéptido que contiene hemo. En algunas realizaciones, el producto cárnico consumible comprende como máximo el 10% (p/p) de la globina. En algunas realizaciones, el producto cárnico consumible comprende una réplica seleccionada del grupo que consiste en: una réplica de grasa, una réplica de tejido conjuntivo y una réplica de músculo, o cualquier combinación de los mismos. En algunas realizaciones, el producto cárnico consumible recapitula con precisión características clave asociadas con el cocinado y el consumo de un producto cárnico equivalente derivado de animales. En algunas realizaciones, el polipéptido que contiene hemo se secreta de dicha célula huésped.

La divulgación describe un método para purificar un polipéptido que contiene hemo, por ejemplo, una globina, de una célula vegetal que comprende insertar un polinucleótido que comprende un polinucleótido que codifica para un polipéptido que contiene hemo, por ejemplo, una globina, en una célula vegetal, y purificar el polipéptido que contiene hemo. En algunas variantes, el polinucleótido comprende además una secuencia que codifica para una etiqueta. En algunas variantes, la célula vegetal es de la especie *Nicotiana*. En algunas variantes, la célula vegetal es de *Nicotiana tabacum.* En algunas variantes, la célula vegetal es de *Glycine max.* En algunas variantes, la célula vegetal se selecciona del grupo que consiste en *Nicotiana tabacum*, *Glycine max*, *Zea mays* y *Arabidopsis thaliana.* En algunas variantes, el polipéptido que contiene hemo se selecciona del grupo que consiste en: androglobina, citoglobina, globina E, globina X, globina Y, hemoglobina, mioglobina, leghemoglobina, eritrocruorina, hemoglobina

beta, hemoglobina alfa, hemoglobina no simbiótica, flavohemoglobina, protoglobina, cianoglobina, hemoglobina no simbiótica, globina I de Hell's gate, hemoglobina bacteriana, mioglobina de ciliado, flavohemoglobina, histoglobina, neuroglobinas, protoglobina, globina truncada 2/2, HbN, HbO y Glb3. En algunas variantes, el polipéptido que contiene hemo es una globina. En algunas variantes, la globina es leghemoglobina. En algunas variantes, la globina es hemoglobina. En algunas variantes, el polipéptido que contiene hemo comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a una secuencia de aminoácidos expuesta en la figura 9 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, el método comprende además combinar un polipéptido que contiene hemo con un producto cárnico consumible. En algunas variantes, el producto cárnico consumible comprende una réplica seleccionada del grupo que consiste en: una réplica de grasa, una réplica de músculo y una réplica de tejido conjuntivo, o cualquier combinación de los mismos. En algunas variantes, el producto cárnico consumible comprende al menos el 0,001% (p/p) de dicho polipéptido que contiene hemo, por ejemplo, una globina. En algunas variantes, el producto cárnico consumible comprende como máximo el 10% (p/p) de dicho polipéptido que contiene hemo. En algunas variantes, el producto cárnico consumible recapitula con precisión características clave asociadas con el cocinado y el consumo de un producto cárnico equivalente derivado de animales.

En el presente documento se da a conocer un método para purificar un polipéptido endógeno que contiene hemo, por ejemplo, una globina, de una planta que comprende alterar los niveles de expresión de un polipéptido endógeno que contiene hemo en una planta, y purificar el polipéptido que contiene hemo de dicha planta. En algunas variantes, la alteración aumenta los niveles de expresión del polipéptido endógeno que contiene hemo. En algunas variantes, la alteración aumenta los niveles de expresión de dicho polipéptido endógeno que contiene hemo en una hoja, una semilla, un grano, o cualquier combinación de los mismos. En algunas variantes, la alteración comprende alterar los niveles de expresión de una proteína en la ruta de producción del polipéptido endógeno que contiene hemo. En algunas variantes, la planta es de la especie *Nicotiana*. En algunas variantes, la célula vegetal es de *Nicotiana tabacum.* En algunas variantes, la célula vegetal es de *Glycine max.* En algunas variantes, la célula vegetal se selecciona del grupo que consiste en: *Nicotiana tabacum*, *Glycine max*, *Zea mays* y *Arabidopsis thaliana.* En algunas variantes, el polipéptido que contiene hemo es una globina seleccionada del grupo que consiste en: hemoglobina, leghemoglobina, hemoglobina no simbiótica y Glb3. En algunas variantes, la globina es leghemoglobina. En algunas variantes, la globina es hemoglobina. En algunas variantes, el polipéptido que contiene hemo comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a una secuencia de aminoácidos enumerada en la figura 9 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, el polipéptido que contiene hemo comprende una etiqueta.

En el presente documento se da a conocer un método para secretar un polipéptido que contiene hemo de una bacteria que comprende insertar un polinucleótido que comprende un polinucleótido que codifica para un polipéptido que contiene hemo y un péptido señal en una bacteria, y secretar dicho polipéptido que contiene hemo de dicha bacteria. En algunas variantes, la bacteria es de la especie *Bacillus.* En algunas variantes, la bacteria es *Bacillus subtilis.* En algunas variantes, el polipéptido que contiene hemo se selecciona del grupo que consiste en: androglobina, citoglobina, globina E, globina X, globina Y, hemoglobina, mioglobina, leghemoglobina, eritrocruorina, hemoglobina beta, hemoglobina alfa, hemoglobina no simbiótica, flavohemoglobina, protoglobina, cianoglobina, globina I de Hell's gate, hemoglobina bacteriana, mioglobina de ciliado, histoglobina, neuroglobinas, globina truncada 2/2, HbN, HbO y Glb3. En algunas variantes, el polipéptido que contiene hemo es una globina. En algunas variantes, la globina es leghemoglobina. En algunas variantes, la globina es hemoglobina. En algunas variantes, el polipéptido que contiene hemo comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a una secuencia de aminoácidos enumerada en la figura 9 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, el péptido señal es un péptido señal para una ruta secretora. En algunas variantes, el péptido señal se refiere a dicho polipéptido que contiene hemo en una ruta secretora. En algunas variantes, el péptido señal comprende al menos aproximadamente el 60% de identidad de secuencia de aminoácidos con respecto a un péptido señal enumerado en la tabla 1 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, la secuencia de polinucleótido codifica para un péptido señal que comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a un péptido señal de PhoD (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, la secuencia de polinucleótido codifica para un péptido señal de PhoD. En algunas variantes, el método comprende además purificar el polipéptido que contiene hemo. En algunas variantes, el método comprende además combinar dicho polipéptido purificado que contiene hemo con un producto cárnico consumible. En algunas variantes, el producto cárnico consumible comprende una réplica de grasa, una réplica de músculo y una réplica de tejido conjuntivo, o cualquier combinación de los mismos. En algunas variantes, el producto cárnico consumible comprende al menos el 0,001% (p/p) de dicho polipéptido purificado que contiene hemo. En algunas variantes, el producto cárnico consumible comprende como máximo el

10% (p/p) de dicho polipéptido purificado que contiene hemo. En algunas variantes, el producto cárnico consumible recapitula con precisión características clave asociadas con el cocinado y el consumo de un producto cárnico equivalente derivado de animales.

En el presente documento se da a conocer un método para secretar un polipéptido que contiene hemo de una bacteria que comprende alterar los niveles de expresión de un polipéptido endógeno que contiene hemo en una bacteria, y purificar dicho polipéptido que contiene hemo de dicha bacteria. En algunas variantes, la alteración aumenta los niveles de expresión de dicho polipéptido endógeno que contiene hemo. En algunas variantes, la alteración comprende alterar los niveles de expresión de una proteína en la ruta de producción de dicho polipéptido endógeno que contiene hemo. En algunas variantes, la bacteria es de la especie *Bacillus*. En algunas variantes, la bacteria es *Bacillus subtilis*. En algunas variantes, el polipéptido que contiene hemo se selecciona del grupo que consiste en: hemoglobina, flavohemoglobina, cianoglobina, globina I de Hell's gate, hemoglobina bacteriana, HbN y HbO. En algunas variantes, el polipéptido que contiene hemo es hemoglobina. En algunas variantes, el polipéptido que contiene hemo comprende al menos el 60% de identidad de secuencia de aminoácidos con respecto a una secuencia de aminoácidos enumerada en la figura 9 (por ejemplo, al menos aproximadamente el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos). En algunas variantes, el polipéptido que contiene hemo comprende una etiqueta. En algunas variantes, el método comprende además purificar el polipéptido que contiene hemo. En algunas variantes, el método comprende además combinar el polipéptido purificado que contiene hemo con un producto cárnico consumible. En algunas variantes, el producto cárnico consumible comprende una réplica de grasa, una réplica de músculo y una réplica de tejido conjuntivo, o cualquier combinación de los mismos.

Breve descripción de los dibujos

Las características novedosas de la divulgación se exponen con particularidad en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención haciendo referencia a la siguiente descripción detallada que expone realizaciones ilustrativas, en las que se utilizan los principios, y los dibujos adjuntos de los que:

La figura 1 contiene tres geles SDS-PAGE de las proteínas después de la purificación por afinidad con Ni-NTA del sedimento celular, que muestra una comparación del efecto de un péptido señal de secreción (*PhoD*) sobre la expresión citosólica de la hemoglobina *Aquifex aeolicus* (*AaHb*). La figura 1A es el control de vector vacío. La figura 1B es sin el péptido señal de secreción y la figura 1C es con el péptido señal de secreción.

La figura 2 es un gráfico que presenta el contenido de hemo de un polipéptido expresado de manera citosólica (*AaHb*). La línea correspondiente al vector vacío es la línea con el pico más bajo. La línea correspondiente al polipéptido sin péptido señal es la línea con el pico más alto. La línea correspondiente al polipéptido con el péptido señal de PhoD es la línea con el segundo pico más alto.

La figura 3 contiene dos geles SDS-PAGE de las proteínas después de la purificación por afinidad con Ni-NTA de los medios, que muestra una comparación del efecto de un péptido señal de secreción (*PhoD*) sobre la expresión secretora de un polipéptido (*AaHb*). La figura 3A es el control de vector vacío. La figura 3B es con el péptido señal de secreción.

La figura 4 es la secuencia del polipéptido de fusión que contiene la secuencia PhoD (texto en negrita)-sitio de escisión de proteasa sintética (texto en cursiva, ASAA)-AaHb (subrayado)-His6 (doble subrayado) (SEQ ID NO: 94). Se muestra el sitio de reconocimiento predicho de peptidasa I señal (SPI) (SEQ ID NO: 95), con el sitio de escisión indicado. La secuenciación N-terminal (SEQ ID NO: 96) del polipéptido secretado presente en los medios después de la expresión citosólica de un polipéptido de fusión *PhoD-AaHb* indicó que el extremo N-terminal correspondía al extremo N-terminal de la proteína AaHb.

La figura 5 representa el contenido de hemo de un polipéptido secretado (después de la expresión del polipéptido de fusión *PhoD-AaHb*). La línea correspondiente al vector vacío es la línea con el pico más bajo. La línea correspondiente al polipéptido secretado después de la expresión del polipéptido de fusión que incluía el péptido señal de PhoD es la línea con el pico más alto.

La figura 6 ilustra 1) detección de dos polipéptidos de fusión a modo de ejemplo (*PhoD-yjbI* e *YwbN-yjbI*) en el sedimento celular tras la expresión de un polipéptido endógeno (*yjbI*) fusionado con uno de los dos péptidos de señalización diferentes (*PhoD* o *YwbN*), y 2) detección de medios de los polipéptidos, que demuestra la escisión apropiada de los péptidos señal.

La figura 7 ilustra 1) detección de dos polipéptidos de fusión a modo de ejemplo (*PhoD-LGB2* y *PhoD-HGbI*) en el sedimento celular tras la expresión de dos polipéptidos heterólogos (*LGB2* y *HGbI)* fusionados con el péptido de señalización (*PhoD*), y 2) detección de medios de los polipéptidos, que demuestra la escisión apropiada del péptido señal.

La figura 8 ilustra la detección de medios de un polipéptido (*AaHb*) después de la fusión con un subconjunto de varios péptidos de señalización de secreción a modo de ejemplo diferentes (*PhoD, TipA, WapA, WprA, YmaC, YolA, YuiC, YwbN, AppB* y *BglS*), expresión y escisión del péptido señal de secreción. Las etiquetas indican el péptido señal de secreción que se fusionó con el extremo 5' del polipéptido antes de la escisión.

La figura 9 contiene las secuencias de aminoácidos de polipéptidos a modo de ejemplo que contienen hemo (SEQ ID NO: 1-31).

Descripción detallada

Polipéptidos

Esta divulgación describe composiciones y métodos para la expresión de un polipéptido en una célula huésped (por ejemplo, bacterias y/o plantas). Un polipéptido puede referirse a subunidades o dominios de un polipéptido. Un polipéptido de la divulgación puede ser un polipéptido que contiene hemo. El término polipéptido que contiene hemo puede referirse a todas las proteínas o subunidades proteína que pueden unirse de manera covalente o no covalente a un resto hemo. Los polipéptidos que contienen hemo pueden transportar o almacenar oxígeno. En algunos casos, el polipéptido de la divulgación puede ser una globina. Los polipéptidos pueden comprender el pliegue de la globina, que puede comprender una serie de ocho hélices alfa. Un polipéptido puede comprender una globina alfa y/o una globina beta. Un polipéptido puede comprender una estructura superior característica (por ejemplo, el "pliegue de mioglobina") generalmente asociada con globinas. Un polipéptido puede ser un oligómero. Los polipéptidos pueden ser monómeros, dímeros, trímeros, tetrámeros y/u oligómeros de orden superior. En algunos casos, un polipéptido puede ser un polipéptido que contiene hierro.

Un polipéptido de la divulgación puede incluir, pero no se limita a, androglobina, citoglobina, globina E, globina X, globina Y, hemoglobina, mioglobina, leghemoglobinas, eritrocruorinas, hemoglobinas beta, hemoglobinas alfa, hemoglobinas no simbióticas, flavohemoglobinas, protoglobinas, cianoglobinas, citoglobina, globina I de Hell's gate, hemoglobinas bacterianas, mioglobinas de ciliado, histoglobinas, neuroglobinas, clorocruorina, eritrocruorina, protoglobina, globina truncada 2/2, HbN, HbO, Glb3 y citocromos, proteínas ribosómicas, actina, hexocinasa, lactato deshidrogenasa, fructosa-bisfosfato aldolasa, fosfofructocinasas, triosa-fosfato isomerasas, fosfoglicerato cinasas, fosfoglicerato mutasas, enolasas, piruvato cinasas, proteasas, lipasas, amilasas, glicoproteínas, lectinas, mucinas, gliceraldehído-3-fosfato deshidrogenasas, piruvato descarboxilasas, actinas, factores de elongación de la traducción, histonas, ribulosa-1,5-bisfosfato carboxilasa oxigenasa (rubisco), ribulosa-1,5-bisfosfato carboxilasa oxigenasa activasa (rubisco activasa), albúminas, glicininas, conglicininas, globulinas, vicilinas, conalbúmina, gliadina, glutelina, gluten, glutenina, hordeína, prolamina, faseolina (proteína), proteinoplasto, secalina, extensinas, gluten de *Triticeae*, colágenos, zeína, cafirina, avenina, deshidrinas, hidrofilinas, proteínas abundantes en embriogénesis tardía, proteínas no plegadas de manera nativa, cualquier proteína de almacenamiento de semillas, oleosinas, caloleosinas, esteroleosinas u otras proteínas con esferosoma, proteína A de almacenamiento vegetativo, proteína B de almacenamiento vegetativo, globulina 8S de almacenamiento de semillas de mungo, globulina, globulinas de guisante y albúminas de guisante. En algunos casos, un polipéptido de la divulgación puede comprender o puede ser un polipéptido enumerado en la figura 9. En algunos casos, puede introducirse un polipéptido en una célula huésped. Por ejemplo, un polipéptido puede expresarse, secretarse y/o purificarse de bacterias tales como de la especie *Bacillus.* Un polipéptido puede expresarse y/o purificarse de una planta.

Un polipéptido enumerado en la figura 9 puede expresarse, pero no puede secretarse y/o plegarse de manera apropiada, usando los métodos de la divulgación. Un polipéptido enumerado en la figura 9 puede expresarse, pero no puede ubicarse correctamente en la célula usando los métodos de la divulgación. Un polipéptido enumerado en la figura 9 puede expresarse, pero no puede mantener niveles de actividad comparables a un polipéptido de tipo natural. Un polipéptido enumerado en la figura 9 puede mantener al menos aproximadamente el 1, el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 100% del nivel de actividad de un polipéptido de tipo natural. Un polipéptido enumerado en la figura 9 puede mantener como máximo aproximadamente el 1, el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% del nivel de actividad de un polipéptido de tipo natural. Un polipéptido que comprende al menos aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido enumerado en la figura 9 puede expresarse, pero no puede secretarse y/o plegarse de manera apropiada usando los métodos de la divulgación. Un polipéptido que comprende como máximo aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido enumerado en la figura 9 puede expresarse, pero no puede secretarse y/o plegarse de manera apropiada usando los métodos de la divulgación. Un polipéptido que comprende como máximo aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido enumerado en la figura 9 puede expresarse, pero no puede mantener la actividad en comparación con un polipéptido de tipo natural. Un polipéptido que comprende como máximo aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido enumerado en la figura 9 puede expresarse, pero puede contener menos cofactor con hemos en comparación con un polipéptido de tipo natural.

En algunos casos, una secuencia de un polipéptido que va a expresarse en una célula huésped puede ser una secuencia que comprende al menos aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de identidad de secuencia de aminoácidos con respecto a una secuencia de polipeptido endógeno, por ejemplo, un polipéptido endógeno que contiene hemo, de la célula huésped. En algunos casos, una secuencia de un polipéptido que va a expresarse en una célula huésped puede ser una secuencia que comprende como máximo aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de identidad de secuencia de aminoácidos con respecto a una secuencia de polipeptido endógeno de la célula huésped. Por ejemplo, un polipéptido puede ser una secuencia polipeptídica encontrada en un animal, un mamífero, un vertebrado, un invertebrado, una planta, un hongo, una bacteria, una levadura, un alga, una arquea, un organismo genéticamente modificado tal como una bacteria o levadura genéticamente modificada. Una secuencia polipeptídica puede sintetizarse químicamente y/o sintetizarse mediante síntesis *in vitro*.

Una secuencia polipeptídica puede ser una secuencia de un polipéptido, por ejemplo, un polipéptido que contiene hemo, encontrado en plantas. Los ejemplos no limitativos de plantas pueden incluir cereales tales como, por ejemplo, maíz, avena, arroz, trigo, cebada, centeno, triticale, teff, semillas oleaginosas incluyendo semilla de algodón, semilla girasol, semilla de cártamo, crambe, camelina, mostaza, colza, verduras de hoja tales como, por ejemplo, lechuga, espinaca, col rizada, berza, grelos, acelga, hojas de mostaza, diente de león, brócoli, repollo, caña de azúcar, árboles, tubérculos tales como mandioca, batata, patata, zanahorias, remolachas, nabos, plantas de la familia de las legumbres, tales como, por ejemplo, trébol, guisantes tales como caupíes, guisantes ingleses, guisantes amarillos, guisantes verdes, judías tales como, por ejemplo, semilla de soja, habas, habas de Lima, judías rojas, garbanzos, semillas de mungo, judías pintas, lentejas, altramuces, mezquite, algarrobo, soja y cacahuetes, coco, veza (*Vicia*), *Stylo* (*Stylosanthes*), *Arachis*, *Indigofera*, *Acacia*, *Leucaena*, *Cyamopsis* y *Sesbania*. Las plantas no consumidas habitualmente por seres humanos, incluyendo cultivos de biomasa, incluyendo, por ejemplo, pastos, miscanto, tabaco, *Arundo donax,* caña energética, sorgo, otras hierbas, alfalfa, mazorca de maíz, alga parda u otras algas marinas. Los polipéptidos que pueden encontrarse en cualquier organismo en el reino vegetal pueden usarse en la presente divulgación. En algunos casos, la planta puede ser soja. En algunos casos, la planta puede ser cebada.

En algunos casos, una secuencia polipeptídica puede ser una secuencia, por ejemplo, una secuencia polipeptídica que contiene hemo, encontrada en metazoos. Por ejemplo, una secuencia polipeptídica de la divulgación puede ser una secuencia polipeptídica encontrada en mamíferos tales como vaca, cerdo, rata, perro o caballo. En algunos casos, la secuencia polipeptídica proviene de la vaca. En algunos casos, la secuencia polipeptídica proviene del cerdo. En algunos casos, una secuencia polipeptídica puede ser una secuencia encontrada en protistas. Por ejemplo, una secuencia polipeptídica de la divulgación puede ser una secuencia polipeptídica encontrada en protistas tales como algas. En algunos casos, una secuencia polipeptídica puede ser una secuencia encontrada en arqueas. Por ejemplo, una secuencia polipeptídica de la divulgación puede ser una secuencia polipeptídica encontrada en arqueas tales como halobacterias o *Pyrococcus*. En algunos casos, una secuencia polipeptídica puede ser una secuencia encontrada en eubacterias. Por ejemplo, una secuencia polipeptídica de la divulgación puede ser una secuencia polipeptídica encontrada en eubacterias tales como *Bacillus*, *Clostridium* o *Escherichia.*

Tal como se usa en el presente documento, el término “proteína que contiene hemo” incluye cualquier polipéptido que puede unirse de manera covalente o no covalente a un resto hemo. En algunas realizaciones, el polipéptido que contiene hemo es una globina y puede incluir un pliegue de globina, que comprende una serie de siete a nueve hélices alfa. Las proteínas de tipo globina pueden ser de cualquier clase (por ejemplo, clase I, clase II o clase III) y, en algunas realizaciones, pueden transportar o almacenar oxígeno. Por ejemplo, un polipéptido que contiene hemo puede ser un tipo no simbiótico de hemoglobina o una leghemoglobina. Un polipéptido que contiene hemo puede ser un monómero, es decir, una cadena polipeptídica individual, o puede ser un dímero, un trímero, tetrámero y/o oligómeros de orden superior. La vida útil del estado de Fe^{2+} oxigenado de un polipéptido que contiene hemo puede ser similar al de mioglobina o puede excederlo en un 10%, 20%, 30%, 40%, 50%, 100% o más.

Los ejemplos no limitativos de polipéptidos que contienen hemo pueden incluir una androglobina, una citoglobina, una globina E, una globina X, una globina Y, una hemoglobina, una mioglobina, una eritrocruorina, una hemoglobina beta, una hemoglobina alfa, una protoglobina, una cianoglobina, una histoglobina, una neuroglobina, una clorocruorina, una hemoglobina truncada (por ejemplo, HbN, HbO, una globina truncada 2/2, una hemoglobina 3 (por ejemplo, Glb3)), un citocromo o una peroxidasa.

Los polipéptidos que contienen hemo pueden ser de mamíferos (por ejemplo, animales de granja tales como vacas, cabras, ovejas, cerdos, bueyes o conejos), aves, plantas, algas, hongos (por ejemplo, levadura u hongos filamentosos), ciliados o bacterias. Por ejemplo, un polipéptido que contiene hemo puede ser de un mamífero tal como un animal de granja (por ejemplo, una vaca, una cabra, una oveja, un cerdo, un buey o un conejo) o un ave tal como un pavo o un pollo. Los polipéptidos que contienen hemo pueden ser de una planta tal como *Nicotiana tabacum* o *Nicotiana silvestris* (tabaco); *Zea mays* (maíz), *Arabidopsis thaliana*, una legumbre tal como *Glycine max* (semilla de soja), *Cicer arietinum* (garbanzo), variedades de *Pisum sativum* (guisante) tales como guisantes de huerta o guisantes de azúcar, variedades de *Phaseolus vulgaris* de judías comunes tales como judías verdes, judías negras, judías blancas, alubias o judías pintas, variedades de *Vigna unguiculata* (caupíes), *Vigna radiate* (semillas de mungo), *Lupinus albus* (altramuz) o *Medicago sativa* (alfalfa); *Brassica napus* (colza); *Triticum sps.* (trigo,

incluyendo granos de trigo, y espelta); *Gossypium hirsutum* (algodón); *Oryza sativa* (arroz); *Zizania sps.* (arroz salvaje); *Helianthus annuus* (girasol); *Beta vulgaris* (remolacha azucarera); *Pennisetum glaucum* (mijo perla); *Chenopodium sp.* (quinoa); *Sesamum sp.* (sésamo); *Linum usitatissimum* (lino); u *Hordeum vulgare* (cebada). Los polipéptidos que contienen hemo pueden aislarse de hongos tales como *Saccharomyces cerevisiae*, *Pichia pastoris*, *Magnaporthe oryzae*, *Fusarium graminearum* o *Fusarium oxysporum.* Los polipéptidos que contienen hemo pueden aislarse de bacterias tales como *Escherichia coli*, *Bacillus subtilis*, *Synechocistis sp.*, *Aquifex aeolicus*, *Methylacidiphilum infernorum* o bacterias termófilas tales como *Thermophilus.*

Se conocen las secuencias y la estructura de numerosos polipéptidos que contienen hemo. Véase, por ejemplo, Reedy *et al.*, Nucleic Acids Research, 2008, vol. 36, ejemplar de base de datos D307-D313 y la base de datos de proteínas con hemo disponible en Internet en la página http://hemeprotein.info/heme.php.

Una hemoglobina no simbiótica puede ser de una planta seleccionada del grupo que consiste en semilla de soja, semilla de soja germinada, alfalfa, lino dorado, judía negra, judía de careta, alubia, garbanzo, semilla de mungo, caupíes, judías pintas, guisantes en vaina, quinoa, sésamo, girasol, granos de trigo, espelta, cebada, arroz salvaje o arroz.

Cualquiera de los polipéptidos que contienen hemo descritos en el presente documento puede tener al menos el 60% (por ejemplo, al menos el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 97%, el 98%, el 99% o el 100%) de identidad de secuencia con respecto a la secuencia de aminoácidos del correspondiente polipéptido que contiene hemo de tipo natural o fragmentos del mismo que contienen un motivo de unión a hemo. Por ejemplo, un polipéptido que contiene hemo puede tener al menos el 60% de identidad de secuencia con respecto a una secuencia de aminoácidos expuesta en la figura 9, incluyendo una hemoglobina no simbiótica tal como la de *Vigna radiata* (SEQ ID NO: 1), *Hordeum vulgare* (SEQ ID NO: 5), *Zea mays* (SEQ ID NO: 13), *Oryza sativa subsp. japonica* (arroz) (SEQ ID NO: 14) o *Arabidopsis thaliana* (SEQ ID NO: 15), una globina I de Hell's gate tal como la de *Methylacidiphilum infernorum* (SEQ ID NO: 2), una flavohemoproteína tal como de *Aquifex aeolicus* (SEQ ID NO: 3), una leghemoglobina tal como la de *Glycine max* (SEQ ID NO: 4), *Pisum sativum* (SEQ ID NO: 16) o *Vigna unguiculata* (SEQ ID NO: 17), una peroxidasa dependiente de hemo tal como la de *Magnaporthe oryzae* (SEQ ID NO: 6) o *Fusarium oxysporum* (SEQ ID NO: 7), una citocromo c peroxidasa de *Fusarium graminearum* (SEQ ID NO: 8), una hemoglobina truncada de *Chlamydomonas moewusii* (SEQ ID NO: 9), *Tetrahymena pyriformis* (SEQ ID NO: 10, truncada de grupo I), *Paramecium caudatum* (SEQ ID NO: 11, truncada de grupo I), una hemoglobina de *Aspergillus niger* (SEQ ID NO: 12), o una proteína de mioglobina de mamífero tal como la mioglobina de *Bos taurus* (SEQ ID NO: 18), mioglobina de *Sus scrofa* (SEQ ID NO: 19), mioglobina de *Equus caballus* (SEQ ID NO: 20), una hemoglobina truncada de *Synechocystis PCC6803* (SEQ ID NO: 21), una hemoglobina truncada de *Synechococcus sp. PCC 7335* (SEQ ID NO: 22), una hemoglobina de *Nostoc commune* (SEQ ID NO: 23), una hemoglobina de *Vitreoscilla stercoraria* (SEQ ID NO: 24), una hemoglobina de *Corynebacterium glutamicum* (SEQ ID NO: 25), una hemoglobina truncada de *Bacillus subtilis* (SEQ ID NO: 26), una hemoglobina truncada de *Bacillus megaterium* (SEQ ID NO: 27), una flavohemoglobina de *Saccharomyces cerevisiae* (SEQ ID NO: 28), una hemoglobina no simbiótica de *Nicotina tobaccum* (SEQ ID NO: 29), una hemoglobina no simbiótica de *Medicago sativa* (SEQ ID NO: 30) o una hemoglobina no simbiótica de *Glycine max* (SEQ ID NO: 31).

La identidad en porcentaje entre dos secuencias de aminoácidos puede determinarse de la siguiente manera. En primer lugar, se alinean las secuencias de aminoácidos usando el programa BLAST 2 Sequences (Bl2seq) de la versión independiente de BLASTZ que contiene la versión 2.0.14 de BLASTP. Esta versión independiente de BLASTZ puede obtenerse de la página web de Fish & Richardson (por ejemplo, www.fr.com/blast/) o de la página web del Centro Nacional del Gobierno de los Estados Unidos para Información de Biotecnología (www.ncbi.nlm.nih.gov). Pueden encontrarse instrucciones que explican cómo usar el programa Bl2seq en el archivo de instrucciones que acompaña a BLASTZ. Bl2seq realiza una comparación entre dos secuencias de aminoácidos usando el algoritmo de BLASTP. Para comparar dos secuencias de aminoácidos, las opciones de Bl2seq se establecen de la siguiente manera: -i se establece para un archivo que contiene la primera secuencia de aminoácidos que va a compararse (por ejemplo, C:\seq1.txt); -j se establece para un archivo que contiene la segunda secuencia de aminoácidos que va a compararse (por ejemplo, C:\seq2.txt); -p se establece para blastp; -o se establece para cualquier nombre de archivo deseado (por ejemplo, C:\output.txt); y todas las otras opciones se dejan en su valor por defecto. Por ejemplo, pueden usarse los siguientes comandos para generar un archivo de salida que contiene una comparación entre dos secuencias de aminoácidos: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. Si las dos secuencias comparadas comparten homología, entonces el archivo de salida designado presentará aquellas regiones de homología como las secuencias alineadas. Si las dos secuencias comparadas no comparten homología, entonces el archivo de salida designado no presentará secuencias alineadas. Pueden seguirse procedimientos similares para secuencias de ácido nucleico excepto si se usa blastn.

Una vez alineadas, se determina el número de coincidencias contando el número de posiciones en las que está presente un residuo de aminoácido idéntico en ambas secuencias. La identidad en porcentaje se determina dividiendo el número de coincidencias entre la longitud de la secuencia de aminoácidos del polipéptido de cadena completa seguido por multiplicar el valor resultante por 100. Se observa que el valor de identidad en porcentaje se redondea a la décima más cercana. Por ejemplo, 78,11, 78,12, 78,13 y 78,14 se redondean hacia abajo hasta 78,1, mientras que 78,15, 78,16, 78,17, 78,18 y 78,19 se redondean hacia arriba hasta 78,2. También se observa que el

valor de longitud siempre es un número entero.

Se apreciará que varios ácidos nucleicos pueden codificar para un polipéptido que tiene una secuencia de aminoácidos particular. El código degenerado del código genético se conoce bien en la técnica; es decir, para muchos aminoácidos, existe más de un triplete de nucleótidos que sirve como codón para el aminoácido. Por ejemplo, pueden modificarse codones en la secuencia codificante para una enzima dada de modo que se obtiene la expresión óptima en una especie particular (por ejemplo, bacterias u hongo), usando tablas de sesgo de codón apropiadas para esas especies.

Los polipéptidos que contienen hemo pueden extraerse del material fuente (por ejemplo, extraerse de tejido animal, o biomasa de plantas, hongos, algas o bacterias, o del sobrenadante de cultivo para proteínas secretadas) o de una combinación de materiales fuente (por ejemplo, múltiples especies vegetales). La leghemoglobina se encuentra fácilmente disponible como un subproducto no usado de cultivos primarios de legumbres (por ejemplo, semilla de soja, alfalfa o guisante). La cantidad de leghemoglobina en las raíces de estos cultivos en los Estados Unidos excede el contenido de mioglobina de toda la carne roja consumida en los Estados Unidos.

En algunas realizaciones, los extractos de polipéptidos que contienen hemo incluyen uno o más polipéptidos que no contienen hemo del material fuente (por ejemplo, otras proteínas de animales, plantas, hongos, algas o bacterias) o de una combinación de materiales fuente (por ejemplo, diferentes animales, plantas, hongos, algas o bacterias).

Un polipéptido de la divulgación (por ejemplo, una globina, un polipéptido que contiene hemo o una proteína que contiene hierro), puede denominarse polipéptido “purificado”. Un polipéptido de la divulgación puede purificarse de otros componentes del material fuente (por ejemplo, otras proteínas de animales, plantas, hongos, algas o bacterias). Un polipéptido purificado puede referirse a un polipéptido que se ha enriquecido en una composición, se ha manipulado de alguna manera para retirar residuos no deseados (por ejemplo, residuos celulares, ADN genómico y/u otros polipéptidos) y/o se retira de la célula huésped en la que se sintetizó (por ejemplo, transcribió/tradujo) (por ejemplo, lisis celular). Un polipéptido “purificado” puede ser un polipéptido extraído de su célula huésped. En algunas variantes, un polipéptido “purificado” es al menos el 1% puro, por ejemplo, al menos el 2%, el 5%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95% o el 99% puro. Las proteínas pueden separarse en función de su peso molecular, por ejemplo, mediante cromatografía de exclusión molecular, ultrafiltración a través de membranas o centrifugación por densidad. En algunos casos, las proteínas pueden separarse en función de su carga superficial, por ejemplo, mediante precipitación isoeléctrica, cromatografía de intercambio aniónico o cromatografía de intercambio catiónico. Las proteínas también pueden separarse en función de su solubilidad, por ejemplo, mediante precipitación con sulfato de amonio, precipitación isoeléctrica, extracción con tensioactivos, detergentes o disolventes. Las proteínas también pueden separarse por su afinidad con otra molécula usando, por ejemplo, cromatografía de interacción hidrófoba, tintes reactivos o hidroxiapatita. La cromatografía de afinidad también puede incluir el uso de anticuerpos que tienen afinidad de unión específica para el polipéptido que contiene hemo, del anticuerpo a la proteína, níquel-NTA para proteínas recombinantes etiquetadas con His, lectinas para unirse a restos de azúcares en una glucoproteína u otras moléculas que se unen específicamente a la proteína.

Hemoglobina

La hemoglobina (Hb) puede ser el principal constituyente de un eritrocito que puede llevar oxígeno desde los pulmones por todo el cuerpo. Cuando está contenida en glóbulos rojos, la Hb humana puede existir como una estructura tetramérica compuesta de dos αβ-dímeros ligados a oxígeno, teniendo cada uno un peso molecular de aproximadamente 32 kD. Cada subunidad α y β de cada dímero puede tener una cadena proteínica y una molécula con hemo. La hemoglobina o “Hb” puede referirse a (a) un pigmento respiratorio que contiene hierro encontrado en glóbulos rojos de vertebrados que comprende una globina compuesta de cuatro subunidades (un tetrámero) cada una de las cuales se liga a una molécula con hemo, que funciona en el transporte de oxígeno a los tejidos después de su conversión para dar la forma oxigenada en las branquias o los pulmones, y que ayuda en el transporte de dióxido de carbono de vuelta a las branquias o los pulmones después de la entrega de su oxígeno. Una hemoglobina puede referirse a una hemoglobina producida de manera recombinante; αβ-dímeros de hemoglobina, hemoglobina reticulada de manera intermolecular o intramolecular, así como versiones modificadas de las hemoglobinas descritas en la divulgación, que pueden incluir, pero no se limitan a, modificaciones que aumentan o disminuyen la afinidad por el oxígeno de la hemoglobina (por ejemplo, tal como sustituyendo una alanina, valina, leucina o fenilalanina por histidina en la posición E7 (por ejemplo, posición 62 de la SEQ ID NO: 4)). Véase, por ejemplo, Hargrove *et al*., J. Mol. Biol. (1997) 266, 1032-1042. Todas las hemoglobinas pueden ser capaces de unirse a hemo. Una hemoglobina puede ser una variante de hemoglobina. Las variantes de hemoglobina pueden comprender mutaciones, sustituciones, adiciones y/o deleciones de aminoácidos. Las variantes de hemoglobina pueden incluir hemoglobina Kansas, hemoglobina S, hemoglobina C, hemoglobina E, hemoglobina D-Punjab, hemoglobina O-Arab, hemoglobina G-Philadelphia, hemoglobina Hasharon, hemoglobina Lepore y hemoglobina M.

Leghemoglobina

En algunos casos, la secuencia (aminoácido y/o ácido nucleico) de una leghemoglobina puede ser una secuencia de

leghemoglobina vegetal. Diversas especies de legumbres y sus variedades, por ejemplo, semilla de soja, haba, haba de Lima, caupíes, guisantes ingleses, guisantes amarillos, altramuz, judía roja, garbanzos, cacahuete, alfalfa, heno de algarroba, trébol, lespedeza y judía pinta, comprenden nódulos de raíz que fijan nitrógeno en las que la leghemoglobina puede desempeñar un papel clave en controlar las concentraciones de oxígeno. Las leghemoglobinas de diferentes especies pueden ser homólogas y tener propiedades de color similares. Algunas especies vegetales pueden expresar varias isoformas de leghemoglobina (por ejemplo, la semilla de soja tiene cuatro isoformas de leghemoglobina). Variaciones mínimas en una secuencia de aminoácidos precisa pueden modificar la carga global de la proteína a un pH particular y pueden modificar la conformación estructural precisa de un grupo hemo que contiene hierro en la leghemoglobina. En algunos casos, puede sustituirse una alanina, valina, leucina o fenilalanina por una histidina en la posición 62 de la SEQ ID NO: 4. Las diferencias en la conformación estructural del grupo hemo de diferentes leghemoglobinas pueden influir en las velocidades de oxidación y reducción del hierro del hemo. Estas diferencias pueden contribuir en las propiedades de generación de color y sabor de diferentes leghemoglobinas.

En otros casos, la secuencia (aminoácido y/o ácido nucleico) de un polipéptido que contiene hemo puede ser de un organismo no vegetal, tal como de animales (por ejemplo, una vaca, un cerdo, un perro, una rata o un caballo), pescado, arqueas, protistas, bacterias, hongo, eubacterias, metazoos o levadura.

Variantes

Un polipéptido de la divulgación puede ser una variante (por ejemplo, comprende una mutación tal como una sustitución de aminoácido, por ejemplo, una sustitución de aminoácido no conservativa o conservativa, una deleción de aminoácido, una inserción de aminoácido o una secuencia no nativa). En algunos casos, una variante de polipéptido puede ser una variante de un polipéptido enumerado en la figura 9 (véase, por ejemplo, las SEQ ID NO: 1-31). En algunos casos, una variante de polipéptido puede incluir como máximo 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40 ó 50 mutaciones. En algunos casos, una variante de polipéptido comprende al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40 ó 50 o más mutaciones. En algunos casos, puede mutarse al menos el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40 o el 50% de la secuencia de un polipéptido de la divulgación. En algunos casos, puede mutarse como máximo el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40 o el 50% de la secuencia de un polipéptido de la divulgación. En algunos casos, un polipéptido de la divulgación puede comprender al menos aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido que se produce de manera natural de la divulgación. En algunos casos, un polipéptido de la divulgación puede comprender como máximo aproximadamente el 10, el 20, el 30, el 40, el 50, el 60, el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos con respecto a un polipéptido que se produce de manera natural de la divulgación.

En algunos casos, el polipéptido de la divulgación comprende una secuencia no nativa (por ejemplo, una etiqueta o un marcador). Una etiqueta puede unirse de manera covalente a la secuencia polipeptídica del polipéptido. La etiqueta puede unirse al extremo N-terminal o al extremo C-terminal, o a un aminoácido intermedio. La etiqueta puede insertarse en la secuencia polipeptídica (por ejemplo, en un bucle superficial accesible al disolvente). Los ejemplos de etiquetas pueden incluir, pero no se limitan a, etiquetas de afinidad (por ejemplo, myc, proteína de unión a maltosa o 6xhis, péptidos quelantes de metal tales como módulos de histidina-triptófano que permiten la purificación sobre metales inmovilizados, dominios de proteína A que permiten la purificación sobre inmunoglobulina inmovilizada y el dominio utilizado en el sistema de purificación por extensión/afinidad de FLAGS), y etiquetas fluorescentes (por ejemplo, proteína verde fluorescente).

Una etiqueta puede ser una composición detectable mediante medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, químicos u otros físicos. Por ejemplo, las etiquetas adecuadas para su uso en la presente divulgación pueden incluir biotina, digoxigenina o haptenos, así como proteínas que pueden hacerse detectables, tintes fluorescentes (por ejemplo, isotiocianato de fluoresceína, rojo Texas, rodamina, y similares), radiomarcadores (por ejemplo, ^{3}H, ^{125}I, ^{35}S, ^{14}C o ^{32}P), tintes (por ejemplo, alexa, cy3, cy5), conjugados químicos (por ejemplo, puntos cuánticos), enzimas (por ejemplo, peroxidasa del rábano, fosfatasa alcalina y otras usadas habitualmente en un ELISA) y marcadores colorimétricos tales como oro coloidal, vidrio coloreado o perlas de plástico (por ejemplo, poliestireno, polipropileno, látex, etc.).

Una etiqueta puede detectarse. Por ejemplo, cuando la etiqueta es radioactiva, los medios para la detección pueden incluir un contador de centelleo o una película fotográfica, como en autorradiografía. Cuando la etiqueta es una etiqueta fluorescente, puede detectarse excitando el fluorocromo con la longitud de onda apropiada de luz y detectando la fluorescencia resultante. La fluorescencia puede detectarse visualmente mediante el uso de detectores electrónicos tales como dispositivos de carga acoplada (CCD) o fotomultiplicadores, y similares. De manera similar, las etiquetas enzimáticas pueden detectarse proporcionando los sustratos apropiados para la enzima y detectando el producto de reacción resultante. Las etiquetas colorimétricas o quimioluminiscentes pueden detectarse simplemente observando el color asociado con la etiqueta.

En algunos casos, una etiqueta puede ser un péptido señal. Un péptido señal puede ser una secuencia peptídica presente normalmente en el extremo N-terminal de polipéptidos secretores o de membrana recién sintetizados que dirige al polipéptido a través de o hacia una membrana celular de la célula (la membrana plasmática en procariotas o la membrana del retículo endoplasmático en eucariotas). Posteriormente puede retirarse (por ejemplo, mediante una proteasa). En particular, el péptido señal puede ser capaz de dirigir al polipéptido hacia la ruta secretora de una célula. En algunos casos, el péptido señal es un péptido señal de ruta secretora. En algunos de tales casos, el péptido señal puede denominarse péptido señal o péptido señal de secreción.

Los ejemplos de péptidos señal pueden incluir, pero no se limitan a, los péptidos señal enumerados en la tabla 1 (SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75,77, 79, 81, 83, 85, 87, 89, 91 ó 93). Los nucleótidos entre paréntesis en la tabla 1 pueden incluirse o no y, por tanto, el péptido señal puede tener o no los residuos codificados por los nucleótidos. Por ejemplo, en algunos casos, la secuencia de ácido nucleico expuesta en la SEQ ID NO: 42 no incluye los últimos 9 nucleótidos (es decir, sólo contendría los nucleótidos 1-84 de la SEQ ID NO: 42) y, por consiguiente, el péptido señal expuesto en la SEQ ID NO: 43 no incluye los últimos tres residuos (es decir, el péptido señal sólo contendría los aminoácidos 1-28 de la SEQ ID NO: 43). Por ejemplo, en algunos casos, la secuencia de ácido nucleico expuesta en la SEQ ID NO: 44 no incluye los últimos 6 nucleótidos (es decir, sólo contendría los nucleótidos 1-96 de la SEQ ID NO: 44) y, por consiguiente, el péptido señal expuesto en la SEQ ID NO: 45 no incluye los últimos dos residuos (es decir, el péptido señal sólo contendría los aminoácidos 1-32 de la SEQ ID NO: 45). Por ejemplo, en algunos casos, la secuencia de ácido nucleico expuesta en la SEQ ID NO: 54 no incluye los últimos 12 nucleótidos (es decir, sólo contendría los nucleótidos 1-156 de la SEQ ID NO: 54) y, por consiguiente, el péptido señal expuesto en la SEQ ID NO: 55 no incluye los últimos cuatro residuos (es decir, el péptido señal sólo contendría los aminoácidos 1-52 de la SEQ ID NO: 55). En algunos casos, un péptido señal puede comprender PhoD (por ejemplo, SEQ ID NO: 55 o los residuos 1-52 de la SEQ ID NO: 55). De manera similar, los péptidos señal de las SEQ ID NO: 59, 63, 65, 67, 71, 73, 75, 77, 79, 83, 85, 87 ó 93 pueden carecer de uno a cuatro aminoácidos C-terminales.

En algunos casos, un péptido señal puede ser un péptido señal variante. Un péptido señal puede ser una variante de los péptidos señal enumerados en la tabla 1 (SEQ ID NO: 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75,77, 79, 81, 83, 85, 87, 89, 91 ó 93). En algunos casos, un péptido señal variante comprende al menos aproximadamente el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos con respecto a un péptido señal (por ejemplo, un péptido señal enumerado en la tabla 1). Por ejemplo, un péptido señal variante puede tener al menos aproximadamente el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia con respecto a uno cualquiera de los péptidos señal expuestos en la tabla 1. En algunos casos, un péptido señal variante puede comprender como máximo aproximadamente el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 76, el 77, el 78, el 79, el 80, el 81, el 82, el 83, el 84, el 85, el 86, el 87, el 88, el 89, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% de identidad de secuencia de aminoácidos con respecto a un péptido señal (por ejemplo, un péptido señal enumerado en la tabla 1). En algunos casos, el sitio de escisión entre el péptido señal y el polipéptido puede derivarse del péptido señal. En algunos casos, el sitio de escisión puede ser un sitio de escisión de proteasa sintética. La escisión del péptido señal puede dar como resultado un polipéptido de la divulgación que comprende todos, algunos o ninguno de los péptidos señal. La escisión del sitio de escisión de proteasa sintética puede dar como resultado un polipéptido de la divulgación que comprende todos, algunos o ninguno de los sitios de escisión de proteasa sintética. Tabla 1: Péptidos señal a modo de ejemplo

Gen	**Secuencia de ADN**	**SEQ ID NO**	**Secuencia peptídica (sitio de corte predicho subrayado)**	**SEQ ID NO:**	**Uniprot**
AbnA	ATGAAAAAGAAAAAAACATG GAAACGCTTCTTACACTTTTC GAGTGCAGCTCTGGCTGCAG GTTTGATATTCACTTCTGCTG CTCCCGCAGAGGCAG	32	MKKKKTWKRFLHFSSAALAAGLIFTSAAP AEAA	33	P94522

Gen	Secuencia de ADN	SEQ ID NO	Secuencia peptídica (sitio de corte predicho subrayado)	SEQ ID NO:	Uniprot
AlbB	ATGTCACCAGCACAAAGAAG AATTTTACTGTATATCCTTTC ATTTATCTTTGTCATCGGCGC AGTCGTCTATTTTGTCAAAAG CGATTATCTGTTTACGCTGAT TTTCATTGCCATTGCCATTCT GTTCGG	34	MSPAQRRILL YILSFIFVIG AVVYFVKSDY LFTLIFIAIA ILF	35	P71010
AmyX	ATGGTCAGCATCCGCCGCAGC TTCGAAGCGTATGTCGATGAC ATGAATATCATTACTGTTCTG ATTCCTGCTGAACAAAAGGA AATCATG	36	MVSIRRSFEA YVDDMNIITV LIPAEQKEIM	37	C0SPA0
AppB	ATGGCTGCATATATTATCAGA AGAACCTTAATGTCTATCCCG ATTTTATTGGGAATTACGATT TTATCATTTGTTATCATGAAA GCCGCGCCCGGAGAT	38	MAAYIIRRTL MSIPILLGIT ILSFVIMKAA PGD	39	P42062
BglC	ATGAAACGGTCAATCTCTATT TTTATTACGTGTTTATTGATTA CGTTATTGACAATGGGCGGCA TGATAGCTTCGCCGGCATCAG CA	40	MKRSISIFITCLLITLLTMGGMIASPASA	41	P42403
BglS	ATGCCTTATCTGAAACGAGTG TTGCTGCTTCTTGTCACTGGA TTGTTTATGAGTTTGTTTGCA GTCACTGCTACTGCCTCAGCT (CAA ACA GGT)	42	MPYLKRVLLL LVTGLFMSLF AVTATASAQT G	43	P04957
LipA	ATGAAATTTGTAAAAAGAAG GATCATTGCACTTGTAACAAT TTTGATGCTGTCTGTTACATC GCTGTTTGCGTTGCAGCCGTC AGCAAAAGCCGCT (GAA CAC)	44	MKFVKRRIIA LVTILMLSVT SLFALQPSAK AAEH	45	O32129
LytD	ATGAAAAAGAGACTAATCGC ACCTATGCTTCTATCCGCCGC GTCCCTTGCCTTTTTTGCCATG TCTGGTTCTGCCCAGGCAGCC GCGTAT	46	MKKRLIAPML LSAASLAFFA MSGSAQAAAY	47	P39848
OppA	ATGAAAAAACGTTGGTCGATT GTCACGTTGATGCTCATTTTC ACTCTCGTGCTGAGCGCGTGC GGCTTTGGC	48	MKKRWSIVTL MLIFTLVLSA CGFG	49	P24141

Gen	Secuencia de ADN	SEQ ID NO	Secuencia peptídica (sitio de corte predicho subrayado)	SEQ ID NO:	Uniprot
OppB	ATGCTAAAATATATCGGAAG ACGCTTAGTCTATATGATTAT CACACTATTTGTGATTGTAAC TGTGACATTCTTCTTAATGCA AGCAGCACCGGGCGGG	50	MLKYIGRRLV YMIITLFVIV TVTFFLMQAA PGG	51	P24138
PbpX	ATGACAAGCCCAACCCGCAG AAGAACTGCGAAACGCAGAC GGAGAAAACTAAATAAAAGA GGAAAACTGTTGTTTGGTCTT TTAGCAGTGATGGTTTGCATT ACGATTTGGAATGCTCTTCAT CGA	52	MTSPTRRRTA KRRRRKLNKR GKLLFGLLAV MVCITIWNAL HR	53	O31773
PhoD	ATGGCATACGACAGTCGTTTT GATGAATGGGTACAGAAACT GAAAGAGGAAAGCTTTCAAA ACAATACGTTTGACCGCCGCA AATTTATTCAAGGAGCGGGG AAGATTGCAGGACTTTCTCTT GGATTAACGATTGCCCAGTCG GTTGGGGCCTTT(GAA GTA AAT GCT)	54	MAYDSRFDEWVQKLKEESFQNNTFDRRK FIQGAGKIAGLSLGLTIAQSVGAFEVNA	55	P42251
QcrA	ATGGGCGGAAAACATGATAT ATCCAGACGTCAATTTTTGAA TTATACGCTCACAGGCGTAGG AGGTTTTATGGCGGCTAGTAT GCTCATGCCTATGGTTCGCTT CGCACTCGACCCG	56	MGGKHDISRR QFLNYTLTGV GGFMAASMLM PMVRFALDP	57	P46911
SpoIIIJ	ATGTTGTTGAAAAGGAGAAT AGGGTTGCTATTAAGTATGGT TGGCGTATTCATGCTTTTGGC TGGA (TGC TCG AGT GTG)	58	MLLKRRIGLL LSMVGVFMLL AG CSSV	59	Q01625
TipA	ATGAAAAAAACACTCACCAC TATTCGCAGATCATCAATTGC AAGGAGACTTATTATTTCTTT CCTGCTGATCTTAATTGTTCC GATAACCGCCCTTTCGGTTAG CGCTTATCAATCAGCAGTTGC CTCA	60	MKKTLTTIRR SSIARRLIIS FLLILIVPIT ALSVSAYQSA VAS	61	
WapA	ATGAAAAAAAGAAAGAGGCG AAACTTTAAAAGGTTCATTGC AGCATTTTTAGTGTTGGCTTT AATGATTTCATTAGTGCCAGC CGATGTACTAGCA (AAA TCT ACA)	62	MKKRKRRNFK RFIAAFLVLA LMISLVPADV LA KST	63	Q07833

Gen	Secuencia de ADN	SEQ ID NO	Secuencia peptídica (sitio de corte predicho subrayado)	SEQ ID NO:	Uniprot
WprA	ATGAAACGCAGAAAATTCAG CTCGGTTGTGGCGGCAGTGCT TATTTTTGCACTGATTTTCAG CCTTTTTTCTCCGGGAACCAA AGCTGCAGCG (GCC GGC GCG)	64	MKRRKFSSVV AAVLIFALIF SLFSPGTKAA A AGA	65	P54423
YdeJ	ATGAAGAAACGCAGAAAGAT ATGTTATTGCAATACTGCCCT GCTGCTTATGATTTTGCTTGC TGGA (TGT ACG GAC AGT)	66	MKKRRKICYC NTALLLMILL AG CTDS	67	P96667
YesM	ATGAAGAAAAGAGTTGCTGG CTGGTACAGGCGGATGAAGA TTAAGGATAAGCTGTTTGTGT TTCTATCGTTGATTATGGCCG TATCCTTTCTGTTTGTATACA GCGGGGTCCAGTATGCCTTTC ATGTG	68	MKKRVAGWYR RMKIKDKLFV FLSLIMAVSF LFVYSGVQYA FHV	69	O31516
YesW	ATGAGAAGGAGCTGTCTGAT GATTAGACGAAGGAAACGCA TGTTTACCGCTGTTACGTTGC TGGTCTTGTTGGTGATGGGAA CCTCTGTATGTCCTGTGAAAG CTGAAGGG (GCA)	70	MRRSCLMIRR RKRMFTAVTL LVLLVMGTSV CPVKAEG A	71	O31526
YfkN	ATGAGAATACAGAAAAGACG AACACACGTCGAAAACATTCT CCGTATTCTTTTGCCCCCAAT TATGATACTTAGCCTAATCCT CCCAACACCACCCATTCATGC A (GAA GAA AGC)	72	MRIQKRRTHV ENILRILLPP IMILSLILPT PPIHA EES	73	O34313
YhcR	ATGCTGTCTGTCGAAATGATA AGCAGACAAAATCGTTGTCAT TATGTGTATAAGGGAGGAAA TATGATGAGGCGTATTCTGCA TATTGTGTTGATCACGGCATT AATGTTCTTAAATGTAATGTA CACGTTCGAAGCT (GTA AAG GCA)	74	MLSVEMISRQ NRCHYVYKGG NMMRRILHIV LITALMFLNV MYTFEA VKA	75	P54602
YkpC	ATGCTAAGAGATTTAGGAAG AAGAGTAGCGATCGCAGCCA TTTTAAGCGGAATTATTCTTG GAGGCATGAGCATTTCTTTGG CA (AAT ATG CCC)	76	MLRDLGRRVA IAAILSGIIL GGMSISLA NM P	77	Q45492

Gen	Secuencia de ADN	SEQ ID NO	Secuencia peptídica (sitio de corte predicho subrayado)	SEQ ID NO:	Uniprot
YkuE	ATGAAAAAGATGTCCAGAAG ACAATTTCTAAAAGGAATGTT CGGCGCTCTTGCTGCCGGGGC TTTAACGGCCGGCGGGGGAT ATGGCTATGCC (AGG TAT CTC)	78	MKKMSRRQFL KGMFGALAAG ALTAGGGYGY A RYL	79	O34870
YmaC	ATGAGAAGATTTTTACTAAAT GTCATATTAGTCTTAGCCATT GTCTTGTTCTTGAGATATGTT CATTACTCATTGGAACCAGAA	80	MRRFLLNVIL VLAIVLFLRY VHYSLEPE	81	O31789
YmzC	ATGTTTGAAAGTGAAGCAGA ACTGAGACGAATCAGGATTG CACTTGTATGGATAGCTGTCT TTTTACTGTTCGGGGCG (TGC GGG AAT)	82	MFESEAELRR IRIALVWIAV FLLFGA CGN	83	O31797
YolA	ATGAAGAAGAGAATTACATA TTCACTGCTTGCTCTTCTAGC AGTTGTTGCTTTCGCTTTCACT GATTCATCAAAAGCAAAAGC G (GCA GAA GCA)	84	MKKRITYSLL ALLAVVAFAF TDSSKAKA AE A	85	O31994
YubF	ATGCAGAAATATAGACGCAG AAACACGGTTGCCTTTACAGT ACTAGCTTATTTTACTTTTTTT GCGGGAGTATTTTTGTTTAGT ATCGGACTCTATAATGCTGAT AATCTGGAACT	86	MQKYRRRNTV AFTVLAYFTF FAGVFLFSIG LYNADNLE	87	O32082
YuiC	ATGATGTTGAATATGATCAGA CGTTTGCTGATGACCTGTTTA TTTCTGCTTGCATTTGGCACG ACATTTTTATCAGTGTCAGGA ATTGAAGCG (AAG GAC TTG)	88	MMLNMIRRLL MTCLFLLAFG TTFLSVSGIE A KDL	89	O32108
YvhJ	ATGGCTGAACGCGTTAGAGT GCGTGTGCGAAAAAAGAAAA AGAGCAAACGTAGGAAAATT TTAAAAAGAATAATGTTATTG TTCGCCCTTGCACTATTGGTA GTTGTAGGGCTTGGCGGGTAT AAACTTTAT	90	MAERVRVRVR KKKKSKRRKI LKRIMLLFAL ALLVVVGLGG YKLY	91	P96499
YwbN	ATGAGCGATGAACAGAAAAA GCCAGAACAAATTCACAGAC GGGACATTTTAAAATGGGGA GCGATGGCGGGGGCAGCCGT TGCGATCGGTGCCAGCGGTCT CGGCGGTCTCGCTCCGCTTGT TCAGACTGCGGCT (AAG CCA)	92	MSDEQKKPEQ IHRRDILKWG AMAGAAVAIG ASGLGGLAPL VQTAA KP	93	P39597

Protoporfirinas

Un polipéptido puede unirse a un tetrapirrol (por ejemplo, protoporfirina). Un polipéptido puede unirse a una protoporfirina con su porción de unión a protoporfirina (por ejemplo, dominio). Un polipéptido puede unirse a una protoporfirina a medida que el polipéptido está traduciéndose/plegándose. Un polipéptido puede unirse a una protoporfirina después de que el polipéptido se traduzca/pliegue. Un polipéptido puede permanecer unido a una protoporfirina después de que se haya ubicado de manera subcelular (por ejemplo, ubicado en un compartimento subcelular, secretado).

Las protoporfirinas pueden comprender cadenas laterales que incluyen grupos metilo, grupos ácido propiónico y grupos vinilo. Las estructuras de protoporfirina adecuadas pueden incluir, pero no se limitan a, diyododeuteroporfirina, mesoporfirina, metaloporfirinas y protoporfirina IX. En algunos casos, un polipéptido puede unirse a más de una protoporfirina. Un polipéptido puede unirse a una, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más protoporfirinas.

Una protoporfirina puede ser una protoporfirina IX. La protoporfirina IX (PpIX), feoforbida, un fotosensibilizador que se produce de manera natural, puede ser el precursor inmediato de hemo en la ruta biosintética de hemo. La protoporfirina IX puede denominarse hemo. El hemo puede comprender un anillo de protoporfirina y un átomo de hierro, en el que el átomo de hierro se coordina por los miembros del anillo (por ejemplo, el átomo de hierro está dentro del anillo). En algunos casos, la protoporfirina puede ser un hemo A, hemo B, hemo C, hemo D, hemo I, hemo M, hemo O o hemo S. En algunos casos, una protoporfirina puede coordinarse a un átomo distinto de hierro (es decir, metaloporfirina). Otros átomos pueden incluir, por ejemplo, zinc, gadolinio, magnesio, manganeso, cobalto, níquel, estaño y cobre.

Vectores y organismos genéticamente modificados

Ácidos nucleicos exógenos

La divulgación describe un ácido nucleico exógeno que codifica para un polipéptido de la divulgación (por ejemplo, un polipéptido que contiene hemo, una globina). Un ácido nucleico exógeno puede codificar para cualquiera de los polipéptidos que contienen hemo descritos en el presente documento, por ejemplo, un polipéptido que contiene hemo que tiene al menos aproximadamente el 60% de identidad (por ejemplo, al menos el 65%, el 75%, el 80%, el 85%, el 90%, el 95%, el 99% de identidad) con respecto a una de las secuencias de aminoácidos enumeradas en la figura 9). Un ácido nucleico exógeno puede ser ARN o ADN, y puede ser monocatenario, bicatenario y/u optimizado por codón. Una secuencia de ácido nucleico exógeno que codifica para un polipéptido de la divulgación puede transcribirse y/o traducirse. El término “polinucleótido” puede usarse de manera intercambiable en el presente documento con “ácido nucleico exógeno”.

El término “exógeno” tal como se usa en el presente documento con referencia a un ácido nucleico (o una proteína) y un huésped se refiere a un ácido nucleico que no se produce en (y no puede obtenerse de) una célula de ese tipo particular ya que se encuentra en la naturaleza o una proteína codificada por un ácido nucleico de este tipo. Por tanto, se considera que un ácido nucleico que no se produce de manera natural es exógeno con respecto al huésped una vez en el huésped. Es importante destacar que los ácidos nucleicos que no se producen de manera natural pueden contener subsecuencias de ácido nucleico o fragmentos de secuencias de ácido nucleico que se encuentran en la naturaleza siempre que el ácido nucleico como conjunto no exista en la naturaleza. Por ejemplo, una molécula de ácido nucleico que contiene una secuencia de ADNc dentro de un vector de expresión es un ácido nucleico que no se produce de manera natural, y por tanto es exógeno con respecto a una célula huésped una vez se introduce en el huésped, ya que esa molécula de ácido nucleico como conjunto (ADNc más ADN de vector) no existe en la naturaleza. Por tanto, se considera que cualquier vector, plásmido que se replica de manera autónoma o virus (por ejemplo, retrovirus, adenovirus o virus del herpes) que como conjunto no existe en la naturaleza es un ácido nucleico que no se produce de manera natural. Se deduce que los fragmentos de ADN genómico producidos mediante PCR o tratamiento con endonucleasa de restricción así como ADNc se consideran como un ácido nucleico que no se produce de manera natural ya que existen como moléculas independientes no encontradas en la naturaleza. También se deduce que cualquier ácido nucleico que contiene un elemento regulador (por ejemplo, una secuencia de promotor y/o una secuencia señal) y una secuencia que codifica para un polipéptido que contiene hemo (por ejemplo, ADNc o ADN genómico) en una disposición no encontrada en la naturaleza es un ácido nucleico que no se produce de manera natural. Un ácido nucleico que se produce de manera natural puede ser exógeno con respecto a un huésped particular (por ejemplo, bacterias o planta). Por ejemplo, un cromosoma completo aislado de una célula de la planta x es un ácido nucleico exógeno con respecto a una célula de la planta y una vez que se introduce el cromosoma en una célula de la planta y.

En cambio, el término “endógeno” tal como se usa en el presente documento con referencia a un ácido nucleico (por ejemplo, un gen) (o una proteína) y un huésped se refiere a un ácido nucleico (o proteína) que no se produce en (y puede obtenerse de) ese huésped particular ya que se encuentra en la naturaleza. Además, una célula “que expresa de manera endógena” un ácido nucleico (o proteína) expresa ese ácido nucleico (o proteína) tal como lo hace un huésped del mismo tipo particular ya que se encuentra en la naturaleza. Además, un huésped “de producción

endógena" o que "produce de manera endógena" un ácido nucleico, una proteína u otro compuesto produce ese ácido nucleico, proteína o compuesto tal como lo hace un huésped del mismo tipo particular ya que se encuentra en la naturaleza.

El código degenerado del código genético puede permitir variaciones de la secuencia de nucleótidos, mientras que aún produce un polipéptido que tiene la secuencia de aminoácidos idéntica que la del polipéptido codificado por la secuencia de polinucleótido nativa. Las variaciones en la secuencia de polinucleótido pueden personalizarse para cualquier organismo de interés. En algunos casos, un polinucleótido que codifica para un polipéptido puede optimizarse por codón para la expresión en una bacteria (por ejemplo, una bacteria gram positiva tal como *B. subtilis*). En algunos casos, un polinucleótido que codifica para un polipéptido puede optimizarse por codón para la expresión en una planta (por ejemplo, *N. tabacum*).

La frecuencia de codones sinónimos individuales para aminoácidos relacionados varía ampliamente de genoma a genoma entre eucariotas y procariotas. Estas diferencias en patrones de elección de codones pueden contribuir a los niveles de expresión globales de genes individuales modulando las tasas de elongación de péptidos.

Vectores bacterianos

Tal como se describe en el presente documento, un ácido nucleico exógeno puede incluir un ácido nucleico que codifica para un péptido señal y un ácido nucleico que codifica para un polipéptido que contiene hemo. En algunos casos, el ácido nucleico exógeno es un vector. Un vector puede ser adecuado para la expresión en un procariota (por ejemplo, bacterias).

En general, los vectores dados a conocer en el presente documento comprenden secuencias de control reguladoras, (por ejemplo, secuencias de control transcripcionales o traduccionales) requeridas para la expresión del polipéptido. Las secuencias de control reguladoras adecuadas pueden incluir, pero no se limitan a, origen de replicación, promotor, potenciador, regiones de unión al represor, sitios de inicio de la transcripción, sitio de unión al ribosoma, sitios de inicio de la traducción o sitios de terminación para la transcripción y traducción.

En algunas variantes, el vector puede comprender una secuencia de polinucleótido que codifica para dos o más polipéptidos (por ejemplo, polipéptidos que contienen hemo o enzimas de la ruta de biosíntesis de hemo), que pueden estar presentes en el mismo vector. En algunas variantes, cuando están presentes en el mismo vector, los polinucleótidos se disponen de tal manera que forman un operón (es decir, la transcripción de los polinucleótidos generará un ARN mensajero policistrónico). En algunos casos, los dos o más polinucleótidos pueden disponerse en el mismo vector de tal manera que se ligan operativamente a su propio promotor.

El origen de la replicación (generalmente denominado secuencia ori) puede permitir la replicación del vector en una célula huésped. La elección de ori puede depender del tipo de células huésped que se emplean. Cuando las células huésped son procariotas, el vector de expresión puede comprender una replicación autónoma que dirige la ori del vector dentro de las células procariotas. Los ejemplos no limitativos de esta clase de ori incluyen pMBI, pUC, ColE1 así como otros orígenes bacterianos.

La célula huésped puede comprender un polinucleótido que codifica para un polipéptido de la divulgación para producción recombinante, en la que el polipéptido puede ligarse a una secuencia señal funcional (por ejemplo, una secuencia que codifica para un péptido señal). Las secuencias que codifican para péptidos señal pueden incluir, por ejemplo, aquellas derivadas de spA, phoA, proteína de unión a ribosa, pelB, ompA, ompT, dsbA, torA, torT y tolT, los péptidos señal enumerados en la tabla 1 o los péptidos señal de la ruta de secreción de TAT en bacterias. También se incluyen dentro del alcance de la divulgación secuencias señal derivadas de células eucariotas que también funcionan como secuencias señal en células huésped procariotas.

Los vectores pueden comprender un marcador seleccionable tal como un gen que codifica para una proteína necesaria para la supervivencia o el crecimiento de una célula huésped transformada con el vector. Un gen marcador puede transportarse en otra secuencia de polinucleótido introducida conjuntamente en la célula huésped. Sólo aquellas células huésped en las que se ha introducido un gen seleccionable pueden sobrevivir y/o crecer en condiciones selectivas. Los genes de selección típicos pueden codificar proteína(s) que (a) confieren resistencia al tratamiento con antibióticos u otras toxinas (por ejemplo, ampicilina, kanamicina, neomicina, G418, metotrexato, etc.); (b) deficiencias auxotróficas del complemento; o (c) suministro de nutrientes críticos no disponible a partir de medios complejos.

El vector puede comprender una secuencia de polinucleótido que codifica para una etiqueta. Una secuencia de etiqueta puede estar en marco con respecto a la secuencia codificante del polipéptido de la divulgación, de tal manera que tras la traducción de la secuencia, el polipéptido de la divulgación se une de manera covalente, por ejemplo, se fusiona, a la etiqueta (por ejemplo, es una proteína de fusión). La etiqueta puede separarse del polipéptido de la divulgación mediante un ligador, por ejemplo, un ligador polipeptídico. Un vector puede comprender una secuencia de polinucleótido que codifica para un ligador entre la secuencia que codifica para la etiqueta y la secuencia que codifica para el polipéptido de la divulgación.

Los vectores descritos en el presente documento pueden obtenerse usando métodos de clonación recombinante y/o mediante síntesis química. Las técnicas de clonación recombinante pueden incluir PCR, digestión con endonucleasa de restricción y ligación. Pueden usarse datos de secuencia, que pueden localizarse en las bases de datos públicas o privadas, para obtener un vector deseado mediante cualquier medio de síntesis. Adicionalmente, usando técnicas de restricción y ligación, pueden escindirse secuencias apropiadas de diversas fuentes de ADN e integrarse en relación operativa con las secuencias exógenas que van a expresarse según la presente divulgación.

En algunos casos, el vector puede comprender un elemento regulador (también denominado elemento de control regulador en el presente documento). Un elemento regulador puede incluir, por ejemplo, origen de replicación, promotores, cajas TATA, potenciadores, sitios de unión al ribosoma, regiones de unión al represor, sitios de inicio de la transcripción, sitios de terminación de la transcripción, secuencias no nativas (por ejemplo, etiquetas) y regiones sin traducir. En algunos casos, el elemento regulador puede ser un promotor. Un promotor puede ser constitutivo, inducible y/o específico de tejido. Los promotores a modo de ejemplo pueden incluir tanto promotores constitutivos como promotores inducibles. Un promotor natural puede modificarse mediante reemplazo, sustitución, adición o eliminación de uno o más nucleótidos sin cambiar su función.

En algunas variantes, además de una secuencia promotora, la secuencia de polinucleótido también incluye una región de terminación de la transcripción en el sentido de 3' de la secuencia de ácido nucleico que codifica para el polipéptido para proporcionar una terminación eficiente. En algunas variantes, la región de terminación puede obtenerse a partir del mismo gen que la secuencia promotora, mientras que en otras variantes puede obtenerse a partir de otro gen.

En algunos casos, el vector puede comprender una secuencia de polinucleótido que codifica para un polipéptido de la divulgación y una etiqueta. En algunos casos, la etiqueta puede ser un péptido señal.

En algunas variantes, una vez que se obtiene la forma deseada de un polipéptido, una secuencia de ácido nucleico, un homólogo, una variante o fragmento del mismo, puede modificarse de varias maneras. Cuando la secuencia implica regiones flanqueantes no codificantes, las regiones flanqueantes pueden someterse a resección, mutagénesis, etc. Por tanto, pueden realizarse transiciones, transversiones, deleciones e inserciones en la secuencia que se produce de manera natural.

En algunas variantes particulares, un polinucleótido que codifica para un polipéptido puede incluir la secuencia codificante para al menos un polipéptido (por ejemplo, globina), o variante(s), fragmento(s) o variante(s) de corte y empalme del mismo: (i) en aislamiento; (ii) en combinación con secuencias codificantes adicionales; tales como secuencias codificantes de proteínas de fusión o péptidos señal, en las que la secuencia codificante del polipéptido es la secuencia codificante dominante; y/o (iii) en combinación con secuencias no codificantes, tales como elementos de control, tales como elementos promotores y terminadores o regiones sin traducir en 5' y/o 3', eficaces para la expresión de la secuencia codificante en un huésped adecuado.

En algunas variantes, puede introducirse un polinucleótido que codifica para un polipéptido, junto con secuencias promotoras y de control apropiadas, en células huésped bacterianas para permitir que las células expresen al menos un polipéptido (por ejemplo, globina) o una variante del mismo.

Pueden incorporarse en vectores fragmentos de polinucleótidos naturales o sintéticos que codifican para un polipéptido (por ejemplo, un polipéptido que contiene hemo, una globina), que pueden introducirse en, y replicarse en, una célula bacteriana. Puede usarse cualquier vector siempre que pueda replicarse y sea viable en las células en las que se introduce. Puede insertarse la secuencia de ADN apropiada en un plásmido o vector mediante cualquier método adecuado. En general, la secuencia de ADN se inserta en un sitio o sitios de endonucleasa de restricción apropiado(s) mediante técnicas de biología molecular recombinante.

En algunos casos, la divulgación describe un organismo genéticamente modificado. En algunos casos, un organismo genéticamente modificado puede comprender un polipéptido de la divulgación. Un organismo genéticamente modificado puede comprender un polinucleótido que codifica para un polipéptido que contiene hemo.

Vectores vegetales

La presente divulgación describe vectores para la introducción de ácido nucleico exógeno en un método según la divulgación que pueden encontrar uso en la expresión de un ácido nucleico en una célula vegetal, tejidos vegetales específicos tales como la hoja, la raíz, la semilla o el grano, o un compartimento específico de una célula vegetal.

En el presente documento se dan a conocer células vegetales y plantas transgénicas que comprenden al menos un ácido nucleico exógeno. Tal como se describe en el presente documento, un ácido nucleico exógeno puede incluir un ácido nucleico que codifica para un péptido señal y un ácido nucleico que codifica para un polipéptido que contiene hemo. Una planta o célula vegetal puede transformarse teniendo un constructo integrado en su genoma, es decir, puede transformarse de manera estable. Las células transformadas de manera estable mantienen

normalmente el ácido nucleico introducido con cada división celular. Una planta o célula vegetal también puede transformarse de manera transitoria de tal manera que el constructo no se integra en su genoma. Las células transformadas de manera transitoria pierden normalmente todo o alguna porción del constructo de ácido nucleico introducido con cada división celular de tal manera que el ácido nucleico introducido no puede detectarse en células hijas después de un número suficiente de divisiones celulares. Las plantas y células vegetales transgénicas tanto transformadas de manera transitoria como transformadas de manera estable pueden ser útiles en los métodos descritos en el presente documento.

Normalmente, las células vegetales transgénicas usadas en los métodos descritos en el presente documento constituyen parte o toda de una planta completa. Tales plantas pueden hacerse crecer de una manera adecuada para las especies en consideración, o bien en una cámara de crecimiento, en un invernadero o en un campo. Las plantas transgénicas pueden cultivarse según se desee para un fin particular, por ejemplo, introducir un ácido nucleico recombinante en otras líneas, transferir un ácido nucleico recombinante a otras especies o producir el polipéptido deseado. Alternativamente, las plantas transgénicas pueden propagarse de manera vegetativa para aquellas especies susceptibles a tales técnicas. La progenie incluye descendientes de una planta o línea de plantas particular siempre que la progenie herede el transgén. La progenie de una planta instantánea incluye semillas formadas en plantas de la generación F_1, F_2, F_3, F_4, F_5, F_6 y posteriores, o semillas formadas en plantas de la generación BC_1, BC_2, BC_3 y posteriores, o semillas formadas en plantas de la generación F_1BC_1, F_1BC_2, F_1BC_3 y posteriores. Las semillas producidas por una planta transgénica pueden hacerse crecer y luego autosembrarse (o cruzarse y autosembrarse) para obtener semillas homocigotas para el ácido nucleico exógeno.

Las células vegetales transgénicas que se hacen crecer en cultivo de suspensión, o cultivo de tejidos u órganos, pueden ser útiles para la extracción de polipéptidos. Pueden usarse técnicas de cultivo de tejidos sólidos y/o líquidos. Cuando se usa un medio sólido, pueden colocarse células vegetales transgénicas directamente sobre el medio o pueden colocarse sobre una película de filtro que luego se pone en contacto con el medio. Cuando se usa un medio líquido, pueden colocarse células vegetales transgénicas sobre un dispositivo de flotación, por ejemplo, una membrana porosa que está en contacto con el medio líquido. El medio sólido se elabora normalmente a partir de medio líquido al añadir agar. Por ejemplo, un medio sólido puede ser medio Murashige y Skoog (MS) que contiene agar y una concentración adecuada de una auxina, por ejemplo, ácido 2,4-diclorofenoxiacético (2,4-D), y una concentración adecuada de una citocinina, por ejemplo, cinetina.

En algunas variantes, la divulgación describe un método en el que las células de *Agrobacterium* comprenden un vector que comprende elementos de secuencia, que son esenciales para el mantenimiento y la replicación del plásmido en células de *Escherichia coli* y/o *Agrobacterium*, y para la transferencia de un ADN-T a una célula vegetal, y además una región de ADN-T, que comprende la secuencia codificante de un polipéptido que está bajo control de elementos reguladores funcionales en una planta y, opcionalmente, un gen marcador seleccionable vegetal.

En cualquiera de los métodos de transformación, el vector puede incluir un marcador seleccionable vegetal tal como un marcador seleccionable basado en antibióticos (por ejemplo, espectinomicina, kanamicina, estreptomicina). Un marcador seleccionable vegetal puede ser un marcador de proteína fluorescente y/o un marcador colorimétrico (por ejemplo, LacZ). En algunos casos, un marcador seleccionable vegetal puede comprender una péptido deformilasa. La péptido deformilasa puede hidrolizar el grupo N-formilo en una metionina de inicio. La actividad de la péptido deformilasa puede ser esencial para la viabilidad celular. Una péptido deformilasa puede originarse a partir de varios genes incluyendo DEF1 y DEF2. Las plantas que expresan un marcador seleccionable de péptido deformilasa pueden ser resistentes a los inhibidores de péptido deformilasa (por ejemplo, actinonina). Los genes de marcador seleccionable pueden escindirse. La escisión puede producirse mediante recombinación específica de sitio (por ejemplo, recombinación de Cre, Int), transposones (por ejemplo, transposones de la familia Ac/Ds), meganucleasas (por ejemplo, endonucleasas de alojamiento, I-sceI), recombinación homóloga intracromosómica, las endonucleasas Cas9/CRISPR y/o dedos de zinc. Véase, por ejemplo, la discusión a continuación sobre recombinación homóloga.

En algunos casos, se utiliza la rotura específica de sitio en el genoma de la planta o bien para potenciar en gran medida la mutación/el reemplazo basado en recombinación homóloga dirigida de secuencias endógenas (es decir, para reprogramar un gen de globina) o bien para potenciar en gran medida las tasas de mutación o recombinación en sitios específicos (por ejemplo, promotor de genes de globina o promotor de genes de globina/promotor de un gen altamente expresado en el endospermo para redireccionar la expresión de la globina a semillas u otros tejidos seleccionados como diana). La rotura específica de sitio puede producirse mediante TALENS (endonucleasas derivadas de factores de transcripción de plantas que explotan un sistema simple para diseñar elementos de reconocimiento de ADN para crear endonucleasas sintéticas con especificad suficientemente alta para escindir un solo sitio genómico *in vivo*) o el sistema Cas9/Crispr.

En algunos casos, un vector puede comprender uno o más de los siguientes elementos de ácido nucleico: a) un primer elemento de ácido nucleico que comprende una secuencia de nucleótidos que codifica para un marcador seleccionable (por ejemplo, que puede ser funcional en las especies *Escherichia coli* y/o *Agrobacterium*); b) un segundo elemento de ácido nucleico que comprende una secuencia de nucleótidos de un primer origen de replicación que puede ser funcional en *Escherichia coli;* c) un tercer elemento de ácido nucleico que comprende una secuencia de nucleótidos que codifica para una proteína iniciadora de la replicación; y/o d) un cuarto elemento de

ácido nucleico que comprende una secuencia de nucleótidos de un segundo origen de replicación, que puede ser diferente del primer origen de replicación y que es funcional en *Agrobacterium,* en el que los elementos de ácido nucleico anteriores se proporcionan en una molécula de polinucleótido circular y se separan mediante secuencias de nucleótidos espaciadoras que no tienen función en la replicación, el mantenimiento o la transferencia de ácidos nucleicos, y en el que dichas secuencias de nucleótidos espaciadoras representan menos del 20%, el 25%, el 30%, el 35%, el 40% o el 45% del tamaño total del vector.

La divulgación describe un método, en el que las secuencias reguladoras que funcionan en una planta o una célula vegetal incluyen un promotor que puede dirigir y/o controlar la expresión de un gene de interés. Los promotores adecuados pueden incluir promotores del virus del mosaico de *Mirabilis* (MMV), virus del mosaico de escrofulariáceas (FMV) o caulimovirus del rayado clorótico del cacahuete (PCLSV). Otros ejemplos de promotores adecuados pueden incluir un promotor 35S del virus del mosaico de la coliflor, un promotor 35S del virus del mosaico de la coliflor modificado, un promotor 35S del virus del mosaico de la coliflor doble, un promotor 35S mínimo, promotor de nopalina sintasa, un promotor del virus del mosaico del caupí, un promotor de HT-CPMV, un promotor CPS2p de copalil sintasa del tabaco, un promotor de dihidrinina, un promotor de plastocianina, un promotor de 35S/HT-CPMV, promotores de transcrito de cadena completa (FLt), promotores de transcrito subgenómico y muchos otros promotores que se derivan de virus de ADN pertenecientes a la familia de virus *Caulimoviridae*.

Muchos de tales promotores pueden modificarse ligando múltiples copias, por ejemplo, dos copias, de su secuencia potenciadora en tándem para potenciar la actividad del promotor, tal como, pero sin limitarse a, promotor 35S de CaMV doble (35Sx2), promotor de MMV doble (MMVx2) o promotor de FMV doble (FMVx2). Pueden usarse fragmentos funcionales de estos promotores en el vector de la divulgación. También pueden usarse secuencias de nucleótidos que son al menos el 90%, el 95%, el 96%, el 97%, el 98%, el 99% o el 100% idénticas a estas secuencias promotoras y que son funcionales en permitir la expresión en plantas de la secuencia de nucleótidos ligada operativamente en los vectores de la divulgación.

Los vectores de expresión vegetales que pueden ser funcionales en una célula vegetal y que pueden usarse dentro del método de la presente divulgación con el fin de dirigir y/o controlar la expresión de un gen de interés en una planta también pueden comprender, si se desea, una región reguladora del promotor (por ejemplo, una expresión inducible o constitutiva, regulada por el entorno o el desarrollo, o específica de célula o tejido), un sitio de inicio de la transcripción, un sitio de unión al ribosoma, una señal de procesamiento de ARN, un sitio de terminación de la transcripción y/o una señal de poliadenilación. Los elementos reguladores que van a usarse dentro de los métodos de la divulgación pueden estar presentes en una molécula de vector (por ejemplo, vector binario) ligada operativamente a una secuencia de nucleótidos que codifica para un polipéptido de la divulgación. En algunos casos, un elemento regulador puede estar presente en la región de ADN-T de un vector binario (por ejemplo, un vector binario de tamaño mínimo).

En algunos casos, los promotores que controlan la expresión génica lo hacen de manera dependiente de tejido y según la etapa de desarrollo de la planta. Las secuencias de transgenes de la divulgación impulsadas por este tipo de promotores pueden expresarse en tejidos en los que se desea el producto transgénico, dejando el resto de los tejidos en la planta sin modificar mediante la expresión transgénica. Los promotores específicos de tejido pueden inducirse por factores endógenos o exógenos. Los ejemplos específicos de tejido pueden incluir, pero no se limitan a, promotores del gen de la beta-amilasa o gen de hordeína de la cebada (para la expresión génica en semillas), promotores de los genes pz7 y pz130 del tomate (para la expresión génica en ovarios), promotor del gen RD2 del tabaco (para expresión génica en raíces), promotor de TRX en el plátano y promotor de actina en el melón (para expresión génica en frutos). En algunas variantes, los promotores específicos de tejido pueden elegirse para dirigir la expresión del polipéptido a partes voluminosas y fácilmente cosechables de la planta tales como las semillas, los frutos, los tubérculos y las hojas.

Los vectores de expresión vegetales pueden comprender además una secuencia de nucleótidos que codifica para un péptido señal que puede dirigir la proteína recién expresada a una ubicación subcelular. Los péptidos señal que pueden usarse dentro de tales moléculas de vector pueden ser, por ejemplo, un secuencia de direccionamiento de vacuolas, una secuencia de direccionamiento de cloroplastos, una secuencia de direccionamiento de mitocondrias, una secuencia que induce la formación de cuerpos proteicos en una célula vegetal, una secuencia que induce el direccionamiento específico de la proteína fusionada en los mismos a un orgánulo específico dentro de la planta o célula vegetal, o una secuencia que induce la formación de esferosomas en una célula vegetal.

En algunos casos, la secuencia de direccionamiento puede ser un péptido señal para la exportación de una proteína al espacio extracelular. Los péptidos señal pueden ser péptidos transitorios que se ubican en el extremo N-terminal de una proteína y se escinden de manera cotraduccional durante la translocación a través de una membrana plasmática.

En algunas variantes, el péptido señal puede ser una secuencia que cuando se fusiona a una proteína da como resultado la formación de orgánulos de almacenamiento no secretores en el retículo endoplasmático.

Los ácidos nucleicos endógenos pueden modificarse mediante técnicas de recombinación homóloga. Por ejemplo,

pueden usarse endonucleasas específicas de secuencia (por ejemplo, nucleasas de dedos de zinc (ZFN)) y meganucleasas para estimular la recombinación homóloga en genes vegetales endógenos. Véanse, por ejemplo, Townsend *et al*., Nature 459:442-445 (2009); Tovkach *et al*., Plant J., 57:747-757 25 (2009); y Lloyd *et al*., Proc. Natl. Acad. Sci. USA, 102:2232-2237 (2005). También pueden usarse técnicas de edición del genoma de CRISPR Cas (Xie y Yang, Mol. Plant 2013, 6:1975-1983) y TALEN (Zhang *et al*., Plant Physiology 2013 161:20-27) para reemplazar un ácido nucleico endógeno.

En algunas variantes, el vector puede comprender además en la región de ADN-T un sitio de recombinación específica de sitio para la recombinación específica de sitio. En algunas variantes, el sitio de recombinación específica de sitio puede ubicarse en el sentido de 3' del elemento regulador de la planta. En algunas variantes, el sitio de recombinación específica de sitio puede ubicarse en el sentido de 5' del elemento regulador de la planta. En algunas variantes, el sitio de recombinación puede ser un sitio LoxP y una parte de un sistema de recombinación específica de sitio Cre-Lox. El sistema de recombinación específico de sitio Cre-Lox puede usar una recombinasa cíclica (Cre), que puede catalizar la recombinación entre sitios específicos (LoxP) que comprenden sitios de unión específicos para Cre.

En algunas variantes, el sitio de recombinación puede ser un sitio de destino Gateway. Por ejemplo, pueden clonarse polinucleótidos en un "vector de entrada" disponible comercialmente y recombinarse posteriormente en un "vector de destino". El vector de destino puede usarse para el análisis de la actividad promotora de una secuencia de ácido nucleico dada o de varias secuencias, para el análisis de la función, para la ubicación de proteínas, para la interacción proteína-proteína, para el silenciamiento de un gen dado o para experimentos de purificación por afinidad. La tecnología de clonación Gateway puede adquirirse de Invitrogen Inc., EE.UU.

En algunas variantes, se crean lesiones dirigidas mediante recombinación homóloga u otras técnicas de edición génica en la proximidad de un ácido nucleico endógeno que codifica para una hemoproteína o globina y en la proximidad de promotores específicos de tejido deseados y se seleccionan mutantes para la expresión de la hemoproteína o globina endógena deseada en tejidos abundantes y accesibles. En algunas variantes, se crean lesiones dirigidas en la proximidad de promotores específicos de tejido deseados y se seleccionan mutantes para la expresión de la hemoproteína o globina endógen deseada en tejidos abundantes y accesibles. Los ejemplos de métodos para crear estas lesiones dirigidas incluyen, pero no se limitan a, TALENS y el sistema cas9/crispr tal como se comentó anteriormente.

En algunas variantes, las plantas resultantes pueden contener más proteína diana que la planta original. Por ejemplo, la planta resultante puede expresar más de 2X el nivel de la proteína diana nativa en comparación con la planta original. En algunas variantes, la planta resultante puede expresar la proteína diana nativa en un tejido, tal como la semilla o la hoja, en la que no se encuentra normalmente. La proteína diana nativa puede expresarse en un tejido que no se encuentra normalmente 1 vez, 2 veces, 3 veces, 4, veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces o más mayor o menor que los niveles de la proteína diana nativa en el tejido en el que se encuentra normalmente. En algunas variantes, la proteína diana es una hemoglobina tal como la leghemoglobina de *Glycine max* (por ejemplo, SEQ ID NO: 4) o la hemoglobina no simbiótica de la cebada (SEQ ID NO: 5) producidas en plantas de soja o cebada modificadas por ingeniería genética, respectivamente.

En algunas variantes, la planta resultante no contiene ADN foráneo. En algunos casos, pueden usarse lesiones en la proximidad del polipéptido endógeno de la divulgación y/o promotores específicos de tejido para la modificación por ingeniería genética de la expresión del polipéptido de la divulgación. Los métodos para evaluar este tipo de modificación por ingeniería genética pueden incluir la selección para la producción del polipéptido endógeno.

Métodos de expresión en bacterias

La divulgación describe métodos para la expresión de un polipéptido (por ejemplo, globina) en una célula huésped (por ejemplo, bacterias). Las bacterias adecuadas para la expresión de un polipéptido de la divulgación pueden ser bacterias gram negativas o gram positivas. Por ejemplo, las bacterias pueden ser de una especie de *Escherichia* (por ejemplo, *E. coli*) o de una especie de *Bacillus* (por ejemplo, *B. subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. clausii, B. coagulans, B. circulans, B. lautus, B. megaterium* o *B. thuringiensis*). En algunos casos, las bacterias adecuadas para la expresión de un polipéptido de la divulgación pueden ser *B. subtilis.*

En algunos casos, la expresión de un polipéptido puede incluir la introducción de un vector que comprende una secuencia de polinucleótido que codifica para el polipéptido en la célula huésped, y la inducción de la expresión del polipéptido.

En algunas variantes, los métodos de la divulgación describen una célula huésped que comprende una secuencia integrada de manera estable de interés (es decir, ácido nucleico que codifica para un polipéptido). Sin embargo, en casos alternativos, los métodos de la presente divulgación describen el mantenimiento de un vector de transformación extracromosómico de autorreplicación.

Los métodos de introducción del polinucleótido en células para la expresión de la secuencia de polinucleótido pueden incluir, pero no se limitan a, electroporación, transformación, transducción, bombardeo de alta velocidad con microproyectiles recubiertos de ADN, infección con ácidos nucleicos virales modificados (por ejemplo, fago); transformación mediada químicamente o competencia. En algunos casos, los polinucleótidos que codifican para un polipéptido de la divulgación pueden transcribirse *in vitro*, y el ARN resultante puede introducirse en la célula huésped.

Tras la introducción de un polinucleótido que comprende la secuencia codificante para un polipéptido de la divulgación, la célula huésped puede cultivarse en medios de nutrientes convencionales modificados según sea apropiado para activar promotores, seleccionar transformantes y/o amplificar la expresión de un polinucleótido que codifica para un polipéptido. Las condiciones de cultivo, tales como temperatura, pH y similares, pueden ser aquellas usadas previamente para la célula huésped seleccionada para la expresión. Puede considerarse que la progenie de células en las que se han introducido tales constructos de polinucleótido comprende el polinucleótido que codifica para un polipéptido.

En algunas variantes, el polipéptido o la variante del mismo puede expresarse como una proteína de fusión por la célula bacteriana huésped. Aunque la escisión del polipéptido de fusión para liberar la proteína deseada a menudo puede ser útil, no es necesaria. Los polipéptidos y las variantes de los mismos expresados y secretados como proteínas de fusión pueden mantener su función.

La expresión de un polipéptido de la divulgación puede comprender expresión transitoria y/o expresión constitutiva (por ejemplo, desarrollo de una línea celular estable).

Expresión de polipéptidos recombinantes

La expresión de un polipéptido puede comprender la inducción de la célula huésped para transcribir y/o traducir el polipéptido codificado en el polinucleótido introducido en la célula huésped. La inducción puede producirse después de que la célula huésped se haya cultivado durante al menos aproximadamente 0,1, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 ó 20 o más horas. La inducción puede producirse después de que el cultivo de células huésped tenga una densidad óptica (D.O.) de al menos aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4 ó 1,5 o más. La inducción puede producirse después de que el cultivo de células huésped tenga una densidad óptica (D.O.) de como máximo aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 2, 3, 5, 10 ó 20 o más. La inducción puede provocarse por la adición de sustancias químicas tales como IPTG, arabinosa o en respuesta a un nutriente limitante tal como nitrógeno, fósforo, glucosa u oxígeno. En algunos casos, el polipéptido se liga a un promotor, tal como aprE, liaG, lepA, cry3Aa o gsiB que conduce a la expresión constitutiva del polipéptido.

En algunos casos, pueden añadirse agentes químicos a los medios. En algunos casos, los agentes químicos pueden ayudar en la estabilidad, el contenido de hemo y/o la capacidad de plegamiento de proteínas del polipéptido expresado. Un agente químico puede comprender una molécula pequeña tal como un metal. Los ejemplos de metales adecuados para la adición a los medios pueden incluir fluoruros de hierro (difluoruro de hierro, trifluoruro de hierro), dicloruro de hierro, tricloruro de hierro, dibromuro de hierro, tribromuro de hierro, diyoduro de hierro, triyoduro de hierro, óxido de hierro, trióxido de dihierro, tetraóxido de trihierro, sulfuro de hierro, persulfuro de hierro, seleniuro de hierro, teluluro de hierro, nitruro de dihierro, pentacarbonilo de hierro, nonacarbonilo de dihierro, dodecacarbonilo de trihierro, dicloruro de hierro dihidratado, trifluoruro de hierro trihidratado, dibromuro de hierro hexahidratado, dicloruro de hierro tetrahidratado, nitrato de hierro hexahidratado, tricloruro de hierro hexahidratado, difluoruro de hierro tetrahidratado, sulfato de hierro heptahidratado, trinitrato de hierro nonahidratado, trisulfato de dihierro nonahidratado, cromato de hierro, citrato de hierro, gluconato de hierro, hexahidruro de hierro y magnesio, lactato de hierro, fosfato de hierro, pentacarbonilo de hierro, sulfato de hierro y amonio, citrato férrico y de amonio, oxalato férrico y difosfato de trihierro octahidratado.

Puede añadirse un agente químico a los medios a una concentración final de al menos aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 5,0, 10,0 o más milimolar. Puede añadirse un agente químico a los medios a una concentración final de como máximo aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 5,0, 10,0 o más milimolar.

En algunos casos, un agente químico puede ser un derivado de hemo. Un derivado de hemo puede aumentar el contenido de hemo del polipéptido expresado (por ejemplo, aumentar el número de moléculas de globina que comprenden un hemo). Los derivados de hemo adecuados pueden incluir ácido delta-aminolevulínico, derivados de hemo A, derivados de hemo B, derivados de hemo C, derivados de hemo O, precursores de hemo, derivados de hemo I, derivados de hemo M, derivados de hemo D y derivados de hemo S. Puede añadirse un derivado de hemo a los medios a una concentración final de al menos aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 5,0, 10,0 o más milimolar. Puede añadirse un derivado de hemo a los medios a una concentración final de como máximo aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4, 1,5, 1,6, 1,7, 1,8, 1,9, 2,0, 5,0, 10,0 o más milimolar. En algunos casos, no se añade derivado de hemo a los medios.

Después de la inducción del polipéptido, la célula huésped puede cultivarse durante un periodo de tiempo que favorezca los niveles de expresión máximos del polipéptido. Por ejemplo, un polipéptido puede expresarse durante al menos aproximadamente 0,1 horas, 0,5 horas, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 11 horas, 12 horas, 13 horas, 14 horas, 15 horas, 16 horas, 17 horas, 18 horas, 19 horas, 20 horas, 21 horas, 22 horas, 23 horas, 24 horas, 1 día, 2 días, 3 días, 4 días, 5 días, 6 días, 7 días, 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas, 10 o más semanas. Un polipéptido puede expresarse en una célula huésped durante como máximo aproximadamente al menos aproximadamente 0,1 horas, 0,5 horas, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 11 horas, 12 horas, 13 horas, 14 horas, 15 horas, 16 horas, 17 horas, 18 horas, 19 horas, 20 horas, 21 horas, 22 horas, 23 horas, 24 horas, 1 día, 2 días, 3 días, 4 días, 5 días, 6 días, 7 días, 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas, o 10 o más semanas.

Un polipéptido puede expresarse a una variedad de temperaturas. Un polipéptido puede expresarse a una temperatura de al menos aproximadamente 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49 ó 50ºC. Un polipéptido puede expresarse a una temperatura de como máximo aproximadamente 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49 ó 50ºC.

Las proteínas auxiliares tales como tiol-disulfuro oxidorreductasas o chaperonas pueden ser beneficiosas para ayudar al pliegue del polipéptido en su conformación activa. Las tiol-disulfuro oxidorreductasas y las proteína disulfuro isomerasas pueden catalizar la formación de los enlaces disulfuro correctos en la proteína. Se ha demostrado que la expresión del operón bdbDC en *B. subtilis* es beneficiosa para la producción de un polipéptido con enlaces disulfuro. Las chaperonas pueden ayudar al pliegue de la proteína secretora uniéndose a las regiones hidrófobas expuestas en los estados no plegados y evitando interacciones no favorables, y las prolilpeptidil *cis-trans* isomerasas ayudan en la formación de la conformación apropiada de la cadena peptídica adyacente a residuos de prolina. En algunas variantes, las células huésped pueden transformarse con un vector de expresión que codifica para al menos una tiol-disulfuro oxidorreductasa o chaperona.

En algunos casos, puede aumentarse la fracción de polipéptido plegado de manera adecuada mediante la adición de sustancia químicas al medio de crecimiento que reducen/oxidan los enlaces disulfuro y/o alteran el potencial rédox general, y/o sustancias químicas que alteran las propiedades del disolvente afectando de ese modo a la conformación y agregación de las proteínas. En algunos casos, un reactivo que reduce los enlaces disulfuro, tal como 2-mercaptoetanol, puede aumentar la fracción de proteína plegada correctamente. En algunos casos y dependiendo del medio usado, otros agentes de reducción u oxidación de disulfuro (por ejemplo, DTT, TCEP, glutatión reducido y oxidado, cisteína, cistina, cisteamina, tioglicolato, $S_2O_3^{2}$, $S_2O_4^{2}$, $S_2O_5^{2}$, SO_3^{2}, $S_2O_7^{2}$, Cu^{+}, etc.), usados o bien solos o bien en combinación, pueden encontrar uso en la presente divulgación. Puede contemplarse que otros adyuvantes que alteran las propiedades del disolvente (por ejemplo, urea, DMSO, TWEEN®-80, etc.), añadidos al medio de crecimiento o bien solos o bien preferiblemente en combinación con agentes de reducción/oxidación de disulfuro, tales como βME, también pueden aumentar la fracción de proteína secretora plegada correctamente y encontrar uso en diversos casos de la presente divulgación. En algunos casos, puede usarse βME a concentraciones que oscilan entre 0,5 y 4 mM, o entre aproximadamente 0,1 mM y 10 mM.

El polipéptido puede recuperarse del cultivo (por ejemplo, mediante centrifugación, purificación, etc.), tal como se describe a continuación y en el presente documento.

Parámetros de fermentación

La presente divulgación describe procedimientos de fermentación para cultivar especies bacterianas. El cultivo puede lograrse en un medio de crecimiento que comprende un medio acuoso con sales minerales, factores de crecimiento orgánicos, el carbono y la fuente de material energético, oxígeno molecular (para bacterias aerobias y facultativas), y, por supuesto, un inóculo de partida de una o más especies de microorganismos particulares que van a emplearse.

Además del carbono y la fuente energética, oxígeno, nitrógeno asimilable y un inóculo del microorganismo, puede ser necesaria suministrar cantidades adecuadas en proporciones apropiadas de nutrientes minerales para garantizar el crecimiento de microorganismos apropiado, maximizar la asimilación del carbono y la fuente energética por las células en el procedimiento de conversión microbiana y lograr rendimientos celulares máximos con una densidad celular máxima en el medio de fermentación.

Pueden usarse diversos medios de cultivo. Pueden usarse medios de cultivo bacterianos convencionales. Los medios pueden incluir, además de nitrógeno, cantidades adecuadas de fósforo, magnesio, calcio, potasio, azufre y sodio, en formas iónicas y combinadas asimilables solubles adecuadas, y también pueden comprender determinados elementos traza tales como cobre, manganeso, molibdeno, zinc, hierro, boro y yodo, y otros, de nuevo en forma asimilable soluble adecuada.

En casos, la reacción de fermentación puede implicar un procedimiento aerobio en el que el oxígeno molecular necesario puede suministrarse mediante un gas que contiene oxígeno molecular tale como aire, aire enriquecido en oxígeno o incluso oxígeno molecular sustancialmente puro, siempre que mantenga el contenido del recipiente de fermentación con una presión parcial de oxígeno adecuada eficaz en ayudar al crecimiento de las especies de microorganismos de manera próspera. En efecto, mediante el uso de un sustrato hidrocarbonado oxigenado, puede reducirse el requerimiento de oxígeno para el crecimiento del microorganismo. Sin embargo, puede suministrarse oxígeno molecular para el crecimiento de organismos aerobios y, en menor medida, facultativos.

Aunque la tasa de aireación puede variar a lo largo de un intervalo considerable, la aireación puede llevarse a cabo generalmente a una tasa que está en el intervalo de desde aproximadamente 0,5 hasta 10 o desde aproximadamente 0,5 hasta 7 volúmenes (a la presión empleada y a 25ºC.) de gas que contiene oxígeno por volumen de líquido en el fermentador por minuto. Esta cantidad puede basarse en el aire del contenido de oxígeno normal que se suministra al reactor, y en cuanto a oxígeno puro, los intervalos respectivos pueden ser de desde aproximadamente 0,1 hasta 1,7, o desde aproximadamente 0,1 hasta 1,3 volúmenes (a la presión empleada y a 25ºC) de oxígeno por volumen de líquido en el fermentador por minuto.

La presión empleada para el procedimiento de conversión microbiana puede variar ampliamente. Las presiones pueden estar generalmente dentro del intervalo de aproximadamente 0 a 50 psig, desde aproximadamente 0 hasta 30 psig o al menos ligeramente por encima de la presión atmosférica, como un equilibrio de coste de equipos y de funcionamiento frente a la solubilidad de oxígeno lograda. Las presiones mayores de la atmosférica pueden aumentar la concentración de oxígeno disuelto en el fermento acuoso, lo que a su vez puede ayudar a aumentar las tasas de crecimiento celular.

La temperatura de fermentación puede variar un poco, pero para la mayoría de especies bacterianas huésped usadas en el presente documento, la temperatura puede estar generalmente dentro del intervalo de desde aproximadamente 20ºC hasta 40ºC, o en el intervalo de desde aproximadamente 28ºC hasta 37ºC, dependiendo de la cepa de microorganismo elegida.

En algunos casos, los microorganismos pueden requerir una fuente de nitrógeno asimilable. La fuente de nitrógeno asimilable puede ser cualquier compuesto o compuestos que contiene(n) nitrógeno que puede(n) liberar nitrógeno en una forma adecuada para la utilización metabólica por el microorganismo. Aunque pueden emplearse una variedad de compuestos de fuente de nitrógeno orgánico, tales como hidrolizados de proteína, habitualmente pueden utilizarse compuestos que contienen nitrógeno baratos tales como amoniaco, hidróxido de amonio, urea y diversas sales de amonio tales como fosfato de amonio, sulfato de amonio, pirofosfato de amonio, cloruro de amonio o diversos otros compuestos de amonio. El propio gas de amoniaco puede ser conveniente para operaciones a gran escala, y puede emplearse burbujeándose a través del fermento acuoso (medio de fermentación) en cantidades adecuadas. Al mismo tiempo, también puede emplearse tal amoniaco para ayudar en el control del pH.

El intervalo de pH en el fermento microbiano acuoso (mezcla de fermentación) puede estar en el intervalo a modo de ejemplo de desde aproximadamente 2,0 hasta 8,0. Sin embargo, el intervalo de pH óptimo para determinados microorganismos puede depender de los medios empleados en cierta medida, así como el microorganismo particular, y por tanto cambiar un poco con el cambio en los medios.

El tiempo de retención promedio de la mezcla de fermentación en el fermentador puede variar considerablemente, dependiendo en parte de la temperatura de fermentación y el cultivo empleado. En algunas variantes, la fermentación puede llevarse a cabo de tal manera que el sustrato que contiene carbono puede controlarse como factor limitante, proporcionando de ese modo una buena conversión del sustrato que contiene carbono para dar células y evitando la contaminación de las células con una cantidad sustancial de sustrato sin convertir. El último puede no ser un problema con sustratos solubles en agua, dado que puede retirarse fácilmente cualquier traza restante. Sin embargo, puede ser un problema en el caso de sustratos no solubles en agua, y pueden usarse etapas de tratamiento del producto adicionales tales como etapas de lavado adecuadas. El tiempo necesario para lograr este nivel de sustrato limitante puede variar con el microorganismo particular y el procedimiento de fermentación que va a llevarse a cabo. La fermentación puede llevarse a cabo como una operación discontinua o continua, la operación alimentada discontinua puede usarse para facilitar el control, la producción de cantidades uniformes de productos y los usos más económicos de todos los equipos.

Si se desea, puede añadirse parte o todo el carbono y la fuente de material energético y/o parte de la fuente de nitrógeno asimilable tal como amoniaco al medio mineral acuoso antes de la alimentación del medio mineral acuoso en el fermentador. De hecho, cada una de las corrientes introducidas en el reactor puede controlarse a una tasa predeterminada, o en respuesta a una necesidad que puede determinarse mediante la monitorización de, tal como concentración del carbono y el sustrato energético, pH, oxígeno disuelto, oxígeno o dióxido de carbono en los efluentes gaseosos del fermentador, la densidad celular que puede medirse mediante la transmitancia de luz, o similares. Las tasas de alimentación de los diversos materiales pueden variarse para obtener una tasa de crecimiento celular tan rápida como sea posible, consistente con la utilización eficiente del carbono y la fuente energética, para obtener un rendimiento de células de microorganismo con respecto a la carga de sustrato tan alto como sea posible, y para obtener la mayor producción de la proteína deseada por volumen unitario.

En un lote, pueden esterilizarse inicialmente los equipos, el reactor o el medio de fermentación, el recipiente o receptáculo, las tuberías, los dispositivos de enfriamiento o circulación concomitantes, y similares, normalmente empleando vapor a aproximadamente 121ºC durante al menos aproximadamente 15 minutos. El reactor esterilizado puede inocularse con un cultivo del microorganismo seleccionado en presencia de todos los nutrientes requeridos, incluyendo oxígeno, y el sustrato que contiene carbono.

Métodos de secreción en bacterias

En algunos casos, puede secretarse un polipéptido expresado de una célula huésped (por ejemplo, bacterias). La secreción de un polipéptido puede comprender la liberación del polipéptido de una célula o un compartimento subcelular en una célula (por ejemplo, núcleo, pared celular, membrana plasmática). La secreción puede producirse a través de las membranas plasmáticas, que pueden rodear células y/o compartimentos subcelulares. En algunos casos, la secreción puede referirse a la liberación de un polipéptido a la envoltura celular. En algunos casos, la secreción puede referirse a la liberación de un polipéptido al espacio extracelular (por ejemplo, en los medios de cultivo).

Una célula huésped de la divulgación puede comprender rutas secretoras, que pueden comprender un número de proteínas que funcionan juntas para secretar una proteína. En algunos casos, la célula huésped puede comprender una ruta secretora de translocación de doble arginina (TAT). En algunos casos, un organismo puede comprender una ruta secretora de SEC. La ruta secretora de TAT puede comprender la secreción de polipéptidos (por ejemplo, globinas) en un estado plegado. La ruta secretora de TAT puede transportar proteínas a través de una membrana plasmática (por ejemplo, capa lipídica, es decir, bicapa lipídica).

La divulgación describe factores de secreción y métodos que pueden usarse en células huésped para mejorar el obstáculo para la secreción de proteínas y la producción de proteínas en forma secretada, en particular cuando los polipéptidos se introducen de manera recombinante y se expresan por la célula huésped. La presente divulgación describe los factores de secreción TatC y TatA derivados de *Bacillus subtilis.* En particular, el péptido TatAdCd y TatAyCy, así como los genes que codifican para ellos también son factores de secreción adecuados. El PhoD de *B. subtilis* puede secretarse mediante la ruta de translocación de doble arginina. TatAdCd es de mayor importancia para la secreción de PhoD, mientras que TatAyCy puede no requerirse para este procedimiento.

Expresión de polipéptidos en plantas

Un polipéptido puede expresarse en plantas monocotiledóneas y/o plantas dicotiledóneas. Las técnicas para la introducción de ácidos nucleicos en plantas se conocen en la técnica, e incluyen, sin limitación, transformación mediada por *Agrobacterium*, transformación mediada por vectores virales, electroporación y transformación mediante bombardeo de partículas (también denominada transformación mediante biolística). Véanse, por ejemplo, las patentes estadounidenses n.^{os} 5.538.880; 5.204.253; 6.329.571; y 6.013.863; Richards *et al*., Plant Cell. Rep. 20:48-20 54 (2001); Somleva *et al*., Crop Sci. 42:2080-2087 (2002); Sinagawa-Garcia *et al*., Plant Mol Biol (2009) 70:487-498; y Lutz *et al*., Plant Physiol., 2007, vol. 145, págs. 1201-1210. En algunos casos, puede usarse la transformación intergénica de plástidos como método de introducción de un polinucleótido en una célula vegetal. En algunos casos, el método de introducción de un polinucleótido en una planta comprende la transformación de cloroplastos. En algunos casos, las hojas y/o los tallos pueden ser el tejido diana del polinucleótido introducido. Si se usa una célula o un tejido cultivado como tejido receptor para la transformación, las plantas pueden regenerarse a partir de los cultivos transformados si se desea, mediante técnicas conocidos por los expertos en la técnica.

Otros métodos adecuados para introducir polinucleótidos incluyen electroporación de protoplastos, administración mediada por polietilenglicol de ADN desnudo en protoplastos vegetales, transformación génica directa a través de imbibición (por ejemplo, introducción de un polinucleótido en una planta deshidratada), transformación para dar protoplastos (que puede comprender la transferencia de un polinucleótido a través de descargas osmóticas o eléctricas), transformación química (que puede comprender el uso de una composición de polibreno-espermidina), microinyección, transformación de la ruta del tubo de polen (que puede comprender la administración de un polinucleótido al óvulo de la planta), transformación mediante liposomas, método de transformación del ápice del brote (que puede comprender la introducción de un polinucleótido en el brote y la regeneración del brote), método de transformación de transformación de *Agrobacterium* asistida por sonicación (SAAT), infiltración (que puede comprender una inmersión floral, o una inyección mediante jeringa en una parte particular de la planta (por ejemplo, hoja)), transformación mediada por carburo de silicio (SCMT) (que puede comprender la adición de fibras de carburo de silicio al tejido de la planta y el polinucleótido de interés), electroporación y electroforesis.

Un polipéptido puede expresarse en muchas especies de plantas diferentes, incluyendo, por ejemplo, granos tales como, por ejemplo, maíz, avena, arroz, trigo, cebada, centeno, triticale, teff, semillas oleaginosas incluyendo semilla de algodón, semilla de girasol, semilla de cártamo, crambe, camelina, mostaza, colza, verduras de hoja tales como, por ejemplo, lechuga, espinaca, col rizada, berza, grelos, acelga, hojas de mostaza, diente de león, brócoli, repollo, caña de azúcar, árboles, tubérculos tales como mandioca, batata, patata, zanahorias, remolachas, nabos, plantas de la familia de las legumbres, tales como, por ejemplo, trébol, guisantes tales como caupíes, guisantes ingleses,

guisantes amarillos, guisantes verdes, judías tales como, por ejemplo, semilla de soja, habas, habas de Lima, judías rojas, garbanzos, semillas de mungo, judías pintas, lentejas, altramuces, mezquite, algarrobo, soja y cacahuetes, coco, veza (*Vicia*), *Stylo* (*Stylosanthes*), *Arachis*, *Indigofera*, *Acacia*, *Leucaena*, *Cyamopsis* y *Sesbania*. Las plantas no consumidas habitualmente por seres humanos, incluyendo cultivos de biomasa, incluyendo, por ejemplo, pastos, miscanto, tabaco, *Arundo donax,* caña energética, sorgo, otras hierbas, alfalfa, mazorca de maíz, alga parda u otras algas marinas. Los polipéptidos que pueden encontrarse en cualquier organismo en el reino vegetal pueden usarse en la presente divulgación. En algunos casos, la planta puede ser soja (*Glycine max*). En algunos casos, la planta puede ser cebada (*Hordeum vulgare*). En algunos casos, la planta puede ser *Nicotiana tabacum.* En algunos casos, la planta o célula vegetal no es una planta o célula de la planta *Nicotiana*.

En algunos casos, la variedad, línea de cultivo o variedad de cultivo de *Nicotiana tabacum* puede ser el registro de *N. tabacum* PM016, PM021, PM92, PM102, PM132, PM204, PM205, PM215, PM216 o PM217 depositado en NCIMB, Aberdeen, Escocia, o DAC Mata Fina, P02, BY-64, AS44, RG17, RG8, HB04P, Basma Xanthi BX 2A, Coker 319, Hicks, McNair 944 (MN 944), Burley 21, K149, Yaka JB 125/3, Kasturi Mawar, NC 297, Coker 371 Gold, P02, Wislica, Simmaba, Turkish Samsun, AA37-1, B13P, F4 del cruce BU21 x Hoja Parado line 97, Samsun NN, Izmir, Xanthi NN, Karabalgar, Denizli y P01.

La expresión de un polipéptido de la divulgación puede comprender expresión transitoria y/o expresión constitutiva (por ejemplo, desarrollo de una línea celular estable).

Especies y cepas de Agrobacterium

La divulgación describe cepas de *Agrobacterium* para su uso en métodos para producir un polipéptido mediante la expresión de una secuencia que puede expresarse (por ejemplo, una secuencia que codifica para un polipéptido de la divulgación). Puede usarse una de las cepas de *Agrobacterium* para infiltrar una variedad de planta preseleccionada con el fin de optimizar el rendimiento del polipéptido. En determinados casos, la especie *Agrobacterium* que puede usarse en el método según la divulgación puede incluir, pero no se limita a, *Agrobacterium tumefaciens*, *Agrobacterium rhizogenes*, *Agrobacterium radiobacter, Agrobacterium rubi, Argobacterium vitis.* En algunas variantes, al menos una cepa de Agrobacterium comprende *Agrobacterium tumefaciens.* La especie *Agrobacterium* usada puede ser de tipo natural (por ejemplo, virulenta) o una cepa desarmada. Las cepas de *Agrobacterium* adecuadas pueden incluir cepas de tipo natural (por ejemplo, tales como *Agrobacterium tumefaciens*) o cepas en las que se muta uno o más genes para aumentar la eficiencia de transformación (por ejemplo, tales como cepas de *Agrobacterium* en las que la expresión génica de vir y/o la inducción del mismo se altera debido a la presencia de genes virA o virG mutantes o quiméricos), cepas de *Agrobacterium* que comprenden una copia extra del gen virG, tales como el supergén virG derivado de pTiBo542, ligado a un plásmido de múltiples copias. Otras cepas adecuadas pueden incluir, pero no se limitan a: *A. tumefaciens* C58C1, A136; LBA4011, LBA4404; EHA101; EHA105; AGL1; y A281. En algunos casos, la cepa de *Agrobacterium* seleccionada puede ser AGL1, EHA105, GV2260, GV3101 o Chry5.

En algunos casos, pueden usarse múltiples suspensiones de células de *Agrobacterium*, cada una expresando diferentes genes, para producir el polipéptido, o para potenciar el nivel de expresión de un polipéptido de la divulgación. En tales casos, si se contempla que las células de *Agrobacterium* en las diferentes suspensiones de células de *Agrobacterium* pueden ser la misma cepa o cepas diferentes. Alternativa o adicionalmente, una sola cepa de *Agrobacterium* puede comprender una pluralidad de secuencias que comprenden diferentes polinucleótidos, particularmente polinucleótidos que codifican para polipéptidos de la divulgación. Los diferentes genes pueden estar comprendidos dentro de una sola molécula de ácido nucleico (por ejemplo, un solo vector) o pueden proporcionarse en diferentes vectores. Un ejemplo no limitativo de un segundo gen que puede expresarse en la planta huésped es un gen que codifica para un supresor del silenciamiento, de origen viral.

La expresión de un polipéptido en una célula huésped (por ejemplo, célula vegetal) puede producirse durante al menos aproximadamente 0,1 horas, 0,5 horas, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 11 horas, 12 horas, 13 horas, 14 horas, 15 horas, 16 horas, 17 horas, 18 horas, 19 horas, 20 horas, 21 horas, 22 horas, 23 horas, 24 horas, 1 día, 2 días, 3 días, 4 días, 5 días, 6 días, 7 días, 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas, 10 semanas, 15 semanas, 20 semanas, 30 o más semanas. Un polipéptido puede expresarse en una célula huésped durante como máximo aproximadamente al menos aproximadamente 0,1 horas, 0,5 horas, 1 hora, 2 horas, 3 horas, 4 horas, 5 horas, 6 horas, 7 horas, 8 horas, 9 horas, 10 horas, 11 horas, 12 horas, 13 horas, 14 horas, 15 horas, 16 horas, 17 horas, 18 horas, 19 horas, 20 horas, 21 horas, 22 horas, 23 horas, 24 horas, 1 día, 2 días, 3 días, 4 días, 5 días, 6 días, 7 días, 1 semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 9 semanas o 10 semanas, 15 semanas, 20 semanas, 30 o más semanas.

Un polipéptido puede expresarse a una variedad de temperaturas. Un polipéptido puede expresarse a una temperatura de al menos aproximadamente 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49 ó 50ºC. Un polipéptido puede expresarse a una temperatura de como máximo aproximadamente 4, 10, 16, 18, 21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 42, 43, 44, 45, 46, 47, 48, 49 ó 50ºC.

Potenciamiento de polipéptidos endógenos

La divulgación describe la producción potenciada de un polipéptido endógeno (por ejemplo, un polipéptido endógeno que contiene hemo). La producción potenciada de un polipéptido endógeno puede lograrse mediante la modulación de la ruta que produce el polipéptido endógeno. La modulación puede referirse a la modulación de la transcripción, traducción, ubicación subcelular, ubicación para diferentes tejidos, tiempo de expresión, plegamiento, afinidad por parejas de unión, y similares. La modulación puede producirse a nivel de ADN (por ejemplo, inactivación del gen, inserción de un elemento potenciador/promotor). La modulación puede producirse a nivel de ARN (por ejemplo, silenciamiento del gen mediante interferencia de ARN). La modulación puede producirse a nivel de proteína (por ejemplo, modulación mediante inhibidores alostéricos, grupos de unión de moléculas pequeñas).

En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles del polipéptido endógeno en al menos aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 50 veces o 100 veces o más mayor o menor con respecto a los niveles de tipo natural del polipéptido endógeno. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles del polipéptido endógeno en como máximo aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 50 veces o 100 veces o más mayor o menor con respecto a los niveles de tipo natural del polipéptido endógeno. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles del polipéptido endógeno en al menos aproximadamente el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de los niveles de tipo natural del polipéptido endógeno. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles del polipéptido endógeno en al menos aproximadamente el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de los niveles de tipo natural del polipéptido endógeno.

En algunos casos, pueden modularse los polipéptidos en la ruta de biosíntesis de hemo que pueden producir el cofactor de hemo. Véase, por ejemplo, Tanaka y Tanaka, Annu. Rev. Plant Biol. 2007. 58:321-46, que describe modificaciones para la ruta de biosíntesis de tetrapirrol en plantas. La modulación de polipéptidos en la ruta de biosíntesis de hemo puede ser a nivel de ADN, ARN o proteína. En algunos casos, la modulación de otros polipéptidos en la ruta puede referirse al aumento de los niveles y/o la actividad de un activador de la ruta. En algunos casos, la modulación de otros polipéptidos en la ruta puede referirse a la disminución de los niveles y/o la actividad de un supresor de la ruta.

En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles de polipéptidos en la ruta de biosíntesis de hemo (incluyendo el cofactor de hemo que se asocia con un polipéptido que contiene hemo de la divulgación) en al menos aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 50 veces o 100 veces o más mayor o menor con respecto a los niveles de tipo natural del polipéptido en la ruta. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles de polipéptidos en la ruta de biosíntesis de hemo (incluyendo el cofactor de hemo) en como máximo aproximadamente 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 20 veces, 50 veces o 100 veces o más mayor o menor con respecto a los niveles de tipo natural del polipéptido en la ruta. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles de polipéptidos en la ruta de biosíntesis de hemo (incluyendo el cofactor de hemo) en al menos aproximadamente el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de los niveles de tipo natural del polipéptido en la ruta. En algunos casos, la modulación puede referirse a la alteración de la actividad y/o los niveles de polipéptidos en la ruta de biosíntesis de hemo (incluyendo el cofactor de hemo) en al menos aproximadamente el 5, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90 o el 100% de los niveles de tipo natural del polipéptido en la ruta.

En algunos casos, la modulación puede referirse a la alteración de los niveles de expresión de un polipéptido endógeno en una ubicación específica de la célula huésped. Por ejemplo, los niveles de expresión de un polipéptido endógeno pueden alterarse en una hoja, una semilla, un grano, un tallo, un xilema, un estambre y un pétalo, o cualquier combinación de los mismos.

Métodos de purificación

Un polipéptido expresado y/o secretado de la divulgación puede recuperarse (por ejemplo, del medio de cultivo o de un tejido vegetal). Por ejemplo, cuando el polipéptido expresado que contiene hemo se secreta de las células bacterianas, el polipéptido puede purificarse del medio de cultivo. En algunos casos, las células huésped que expresan el polipéptido pueden retirarse de los medios antes de la purificación del polipéptido (por ejemplo, mediante centrifugación).

Cuando el polipéptido recombinante expresado deseado no se secreta de una célula huésped, puede romperse la célula huésped y liberarse el polipéptido en un “extracto” acuoso que puede ser la primera etapa de purificación. Las células huésped de expresión pueden recogerse de los medios antes de la ruptura celular. La ruptura celular puede realizarse usando cualquier medio adecuado, tal como mediante digestión con lisozimas o beta-glucanasa, trituración, sonicación, homogeneización, molienda o forzando las células a alta presión.

Un polipéptido recuperado puede purificarse. La purificación puede lograrse por medio de lavado/fraccionamiento con una sal (por ejemplo, sulfato de amonio) o a pH bajo (normalmente menos de 3) o procedimientos cromatográficos (por ejemplo, cromatografía de intercambio iónico, cromatografía de afinidad, cromatografía de interacción hidrófoba, cromatografía de inducción de carga hidrófoba, cromatografía de exclusión molecular, etc.). Durante la purificación, la abundancia acumulada en masa de componentes proteicos distintos de la proteína especificada, que pueden ser una sola especie de proteína monomérica o multimérica, puede reducirse en un factor de 2 o más, 3 o más, 5 o más, 10 o más, 20 o más, 50 o más, 100 o más, o 1000 o más con respecto a la materia prima de la que se aisló la proteína especificada.

En algunos casos, un polipéptido puede recuperarse de un fermentador. El caldo de fermentación puede comprender generalmente residuos celulares, incluyendo células, diversos sólidos suspendidos y otros contaminantes de biomasa, así como el producto proteico deseado, que pueden retirarse del caldo de fermentación. Los procedimientos adecuados para tal retirada pueden incluir técnicas de separación sólido-líquido convencionales (por ejemplo, centrifugación, filtración, diálisis, microfiltración, filtración rotaria a vacío u otros procedimientos conocidos) para producir un filtrado libre de células. En algunas variantes, puede ser aceptable para concentrar adicionalmente el caldo de fermentación o el filtrado libre de células antes del procedimiento de purificación y/o cristalización usando técnicas tales como ultrafiltración, evaporación y/o precipitación. En algunos casos, el polipéptido se purifica adicionalmente para reducir la abundancia acumulada en masa de componentes proteicos distintos de la proteína especificada, que pueden ser una sola especie de proteína monomérica o multimérica, en un factor de 2 o más, 3 o más, 5 o más, 10 o más, 20 o más, 50 o más, 100 o más, o 1000 o más con respecto a la materia prima de la que se aisló la proteína especificada. La purificación puede lograrse por medio de lavado/fraccionamiento con una sal (por ejemplo, sulfato de amonio) o a pH bajo (normalmente menos de 3) o procedimientos cromatográficos (por ejemplo, cromatografía de intercambio iónico, cromatografía de afinidad, cromatografía de interacción hidrófoba y/o cromatografía de inducción de carga hidrófoba, etc.).

Caracterización de un polipéptido

Un polipéptido purificado puede caracterizarse para determinar su pureza, contenido de hemo, estado de oligomerización, estabilidad, degradación, afinidad de unión, y similares. Para algunas aplicaciones, los polipéptidos (por ejemplo, globinas) producidos usando la presente divulgación puede ser muy puros (por ejemplo, teniendo una pureza de más del 99%). Un polipéptido purificado puede caracterizarse para determinar su olor, sabor y color.

El polipéptido purificado puede ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% puro. El polipéptido purificado puede ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% puro. El polipéptido purificado puede comprender al menos aproximadamente el 0,01, el 0,05, el 0,1, el 0,2, el 0,3, el 0,4, el 0,5, el 0,6, el 0,7, el 0,8, el 0,9, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98 o el 99% de impurezas. El polipéptido purificado puede comprender como máximo aproximadamente el 0,01, el 0,05, el 0,1, el 0,2, el 0,3, el 0,4, el 0,5, el 0,6, el 0,7, el 0,8, el 0,9, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98 o el 99% de impurezas. El polipéptido purificado puede comprender al menos aproximadamente 0,001, 0,005, 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de impurezas. El polipéptido purificado puede comprender como máximo aproximadamente 0,001, 0,005, 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de impurezas. El polipéptido purificado puede comprender al menos aproximadamente 0,001, 0,005, 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de impurezas. El polipéptido purificado puede comprender como máximo aproximadamente 0,001, 0,005, 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de impurezas.

En algunos casos, la globina purificada puede someterse a prueba para determinar su actividad, estado de oligomerización, plegamiento de proteína apropiado, estabilidad, estructura secundaria y/o contenido de hemo. La actividad, el estado de oligomerización, el plegamiento de proteína y/o la estabilidad pueden determinarse mediante varios métodos incluyendo espectroscopía, ELISA, ensayos de unión, ultracentrifugación analítica, dicroísmo circular, cristalografía de rayos X, resonancia de plasmón superficial, espectrometría de masa o RMN.

Un polipéptido de esta divulgación puede tener propiedades similares a una mioglobina aislada de tejidos animales. En una variante, puede pedirse a un grupo de personas que califique una mioglobina aislada de un tejido animal, según las propiedades que describen la mioglobina. Estas calificaciones pueden usarse como indicación de las propiedades de la mioglobina derivada de tejido animal. El polipéptido de la presente divulgación puede compararse entonces con la globina derivada de animal para determinar cómo de similar es el polipéptido de esta divulgación con respecto a la mioglobina derivada de tejido animal. Así, en algunos casos, el polipéptido se califica como similar con respecto a la mioglobina derivada de tejido animal según la evaluación humana. En algunas variantes, el polipéptido es indistinguible de la mioglobina derivada de tejido animal para un ser humano.

En algunos casos, los polipéptidos de esta divulgación se comparan con mioglobina derivada de tejido animal

basándose en lecturas de olfatómetro. En diversas variantes, el olfatómetro puede usarse para evaluar la concentración de olor y los umbrales de olor, los superumbrales de olor con respecto a un gas de referencia, puntuaciones de escala hedónica para determinar el grado de apreciación o la intensidad relativa de olores. En algunas variantes, el olfatómetro permite el entrenamiento y la evaluación automática de paneles de expertos. En algunos casos, el producto consumible es un producto que provoca lecturas de olfatómetro similares o idénticas. En algunas variantes, la similitud es suficiente para superar el umbral de detección de la percepción humana.

La cromatografía de gases-espectrometría de masas (CG-EM) es un método que combina las características de la cromatografía de gases-líquidos y la espectrometría de masas para separar e identificar sustancias diferentes dentro de una muestra de prueba. En algunas variantes, la CG-EM puede usarse para evaluar las propiedades de polipéptidos de esta divulgación. Por ejemplo, pueden aislarse sustancias químicas volátiles del espacio de cabeza alrededor de la mioglobina derivada de tejido animal. Estas sustancias químicas pueden identificarse usando CG-EM. Por tanto, se crea un perfil de las sustancias químicas volátiles en el espacio de cabeza alrededor de la mioglobina derivada de tejido animal. En algunos casos, puede evaluarse adicionalmente cada pico de la CG-EM. Por ejemplo, un ser humano podría calificar la experiencia de oler la sustancia química responsable de un determinado pico. Esta información podría usarse para refinar adicionalmente el perfil. La CG-EM podría usarse entonces para evaluar las propiedades de un polipéptido de la divulgación. El perfil de CG-EM podría usarse para refinar el polipéptido.

El contenido de hemo puede referirse al porcentaje de moléculas de polipéptido que comprenden la cantidad correcta de restos hemo. Por ejemplo, si un polipéptido de la divulgación se une a un resto hemo, entonces el contenido de hemo puede referirse al número de polipéptidos que se unen al resto hemo. El contenido de hemo de un polipéptido puede ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 75, el 80, el 85, el 90, el 95 o el 100%. El contenido de hemo de un polipéptido puede ser como máximo aproximadamente 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 75, el 80, el 85, el 90, el 95 o el 100%. En algunos casos, el contenido de hemo puede expresarse como razón molar de concentración de polipéptido con respecto a concentración de hemo. El contenido de hemo de la razón molar puede ser al menos aproximadamente 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 1:30 ó 1:40 o menos. El contenido de hemo de la razón molar puede ser como máximo aproximadamente 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:20, 1:30 ó 1:40 o menos. El contenido de hemo puede ser 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 15 veces, 20 veces, 30 veces o 40 veces o más menor que el contenido de hemo de un polipéptido completamente ocupado (por ejemplo, el polipéptido está ocupado al 100% con hemo). El contenido de hemo puede ser 1 vez, 2 veces, 3 veces, 4 veces, 5 veces, 10 veces, 15 veces, 20 veces, 30 veces o 40 veces o más mayor que el contenido de hemo de un polipéptido completamente desocupado (por ejemplo, el polipéptido está ocupado al 0% con hemo). El contenido de hemo puede determinarse mediante un número de métodos incluyendo espectroscopía (Raman, UV-Vis), resonancia paramagnética electrónica (RPE), ensayos de desnaturalización de proteínas, ensayos de robo de hemo y ensayos de reducción de hemo.

Métodos para el uso de un polipéptido en un producto cárnico consumible

La divulgación describe métodos para el uso de un polipéptido de la divulgación en un producto cárnico consumible. Los productos consumibles pueden competir con, complementar o reemplazar alimentos de origen animal. Por ejemplo, los productos consumibles pueden ser réplicas de carne elaboradas completamente a partir de fuentes vegetales. Los productos consumibles pueden elaborarse para imitar el corte o aspecto de carne tal como se vende actualmente. Por ejemplo, un producto consumible puede ser visualmente similar a o indistinguible de carne de vacuno picada o un corte particular de carne de vacuno. Alternativamente, los productos consumibles pueden elaborarse con una apariencia o un aspecto único. Por ejemplo, el producto consumible puede contener patrones o marcados que se basan en la estructura del producto consumible. En algunos casos, los productos consumibles pueden parecer similares a productos cárnicos tradicionales después de prepararse. Por ejemplo, puede producirse un producto consumible que es más grande que un corte tradicional de carne de vacuno pero que, después de que el producto consumible se trocee y cocine parece el mismo que una carne cocinada tradicional. En algunos casos, el producto consumible puede parecerse a una forma de carne tradicional en dos dimensiones, pero no en una tercera. Por ejemplo, el producto consumible puede parecerse a un corte de carne en dos dimensiones (por ejemplo, cuando se observa desde la parte superior), pero puede ser mucho más grande (o grueso) que el corte tradicional. Un producto cárnico consumible (por ejemplo, sucedáneo) puede tener características físicas similares que la carne tradicional (sabor, textura, fuerza, nutrientes). En algunos casos, un producto cárnico consumible puede comprender una característica de pérdida de cocinado similar a la de la carne. En algunas variantes, un producto cárnico consumible puede comprender un contenido de grasas y proteínas similar ya que la carne de vacuno picada tiene la misma reducción en tamaño cuando se cocina como carne de vacuno picada real. Los métodos de producción de un producto cárnico consumible se describen en el documento PCT/US2012/046560.

En algunos casos, un producto cárnico consumible puede comprender un polipéptido de la divulgación. Un polipéptido de la divulgación puede usarse como colorante o indicador de cocinado del producto cárnico consumible.

En algunos casos, la divulgación describe un método para expresar un polipéptido (por ejemplo, globina) en una célula huésped, secretar el polipéptido de la célula huésped, purificar el polipéptido secretado y mezclar el

polipéptido purificado con grasas y lípidos para producir un sucedáneo de carne.

En algunos casos, la divulgación describe un método para potenciar la expresión de un polipéptido endógeno (por ejemplo, globina) en una célula huésped, purificar el polipéptido de la célula y mezclar el polipéptido purificado con grasas y lípidos para producir un sucedáneo de carne.

En algunos casos, la divulgación describe un método para expresar un polipéptido (por ejemplo, globina) en una célula huésped (por ejemplo, una planta), purificar el polipéptido secretado y mezclar el polipéptido purificado con grasas y lípidos para producir un sucedáneo de carne.

Composiciones

En algunos casos, la divulgación describe una composición que comprende un polipéptido de la divulgación. Una composición puede comprender medios en los que se secretó el polipéptido. Una composición puede comprender al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% (p/p) o más de medios. Una composición puede comprender como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% o más de medios. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de medios. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de medios. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de medios. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de medios.

Una composición puede comprender una célula huésped (por ejemplo, una bacteria). Una célula huésped de la composición puede ser la célula huésped de la que se expresa y/o secreta el polipéptido. Una composición puede comprender al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% (p/p) o más de células huésped. Una composición puede comprender como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% de células huésped. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de células huésped. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de células huésped. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de células huésped. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de células huésped. Una composición puede ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de una célula huésped. Una composición puede ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de una célula huésped.

Una composición puede comprender un componente de una célula huésped recombinante (por ejemplo, una célula bacteriana o célula vegetal recombinante). Un componente de una célula huésped puede incluir, por ejemplo, una pared celular, un compartimento subcelular (por ejemplo, complejo de Golgi, retículo endoplasmático o núcleo), ácido nucleico, proteína, ADN genómico y/ o una membrana plasmática. Para una célula bacteriana, un componente también puede incluir flagelos, y para una planta o célula vegetal, un componente puede ser cualquier parte de la planta tal como un brote, un tallo, una semilla, un grano, una hoja, tejido de xilema, una roseta, una raíz. Un componente de una célula huésped puede ser un componente de una célula huésped de la que se expresa y/o secreta el polipéptido. Por ejemplo, una composición puede incluir un componente que no se produce de manera natural de una célula huésped recombinante tal como una proteína de fusión o un ácido nucleico exógeno. Una composición puede comprender al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% (p/p) o más de un componente de una célula huésped. Una composición puede comprender como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% de un componente de una célula huésped. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón del componente de una bacteria. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de un componente de una célula huésped. Una composición puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de un componente de una célula huésped. Una composición puede comprender como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de un componente de una célula huésped. Una composición puede ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped. Una composición puede ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 75, el 80, el 85, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped.

Composiciones de producto cárnico consumible

En algunos casos, una composición de la divulgación puede comprender un producto cárnico consumible y una célula huésped (por ejemplo, bacteria, una parte de una bacteria y/o una parte de una planta). En algunos casos, una composición puede comprender además un polipéptido de la divulgación. En algunos casos, una composición comprende un polipéptido y un producto cárnico consumible (es decir, sucedáneo de carne) (tal como se describe en el documento PCT/US2012/046560).

Un producto cárnico consumible puede referirse a un producto similar a carne (por ejemplo, un sucedáneo de carne) que no está elaborado de carne. Un producto cárnico consumible puede referirse a un sucedáneo de carne que está elaborado de productos no animales (por ejemplo, una planta). Un producto cárnico consumible puede ser réplicas de carne elaboradas completamente a partir de fuentes de origen vegetal. Los productos consumibles también pueden elaborarse a partir de una combinación de fuentes de origen vegetal y fuentes de origen animal. Los productos consumibles pueden elaborarse para imitar el corte o aspecto de carne tal como se vende actualmente. Por ejemplo, un producto consumible puede ser visualmente similar a o indistinguible de carne de vacuno picada o un corte particular de carne de vacuno. En algunos casos, los productos consumibles parecen similares a productos cárnicos tradicionales después de prepararse. El producto cárnico consumible puede estar sustancialmente o completamente compuesto por componentes derivados de fuentes de origen no animal, que aún recapitula características clave asociadas con el cocinado y el consumo de un producto cárnico equivalente derivado de animales.

Una composición puede comprender un producto cárnico consumible y un polipéptido de la divulgación. Un producto cárnico consumible puede comprender al menos aproximadamente el 1, el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 80, el 85, el 90, el 95 o el 100% (p/p) de uno o más polipéptidos de la divulgación. En algunos casos, un producto cárnico consumible puede comprender como máximo aproximadamente el 1, el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 80, el 85, el 90, el 95 o el 100% de uno o más polipéptidos de la divulgación. Un producto cárnico consumible puede comprender al menos aproximadamente el 1, el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 80, el 85, el 90, el 95 o el 100% peso/volumen de uno o más polipéptidos de la divulgación. En algunos casos, un producto cárnico consumible puede comprender como máximo aproximadamente el 1, el 5, el 10, el 15, el 20, el 25, el 30, el 35, el 40, el 45, el 50, el 55, el 60, el 65, el 70, el 75, el 80, el 85, el 90, el 95 o el 100% peso/volumen de uno o más polipéptidos de la divulgación.

Una composición puede comprender un producto cárnico consumible y una célula huésped (por ejemplo, bacteria). Una célula huésped de la composición puede ser la célula huésped de la que se expresa y/o secreta el polipéptido. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% (p/p) o más de células huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% de células huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de célula huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de célula huésped. Una composición puede comprender un producto cárnico consumible y ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de una célula huésped. Una composición puede comprender un producto cárnico consumible y ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de una célula huésped.

Una composición puede comprender una parte de un producto cárnico consumible y un componente de una célula huésped (por ejemplo, una parte de una bacteria). Un componente de una célula huésped puede incluir una pared celular, un compartimento subcelular (por ejemplo, complejo de Golgi, retículo endoplasmático, núcleo), un flagelo, ácido nucleico, proteína, ADN genómico o una membrana plasmática. Un componente de una célula huésped puede ser una parte de una bacteria de la que se expresa y/o secreta el polipéptido. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% o más de una parte de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de una parte de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de un componente de una célula huésped. Una composición puede

comprender un producto cárnico consumible y ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped.

Una composición puede comprender un producto cárnico consumible y un componente de una célula huésped (por ejemplo, planta, por ejemplo, una planta de tabaco, es decir, la especie *Nicotiana tabacum*, o una planta de soja, es decir, la especie *Glycine max*). En algunos casos, la célula huésped no es una planta *Nicotiana*. Una parte de una célula huésped puede incluir una pared celular, un compartimento subcelular (por ejemplo, complejo de Golgi, retículo endoplasmático, núcleo), un brote, un tallo, una hoja, una semilla, un grano, un xilema, una roseta, una raíz, ácido nucleico, proteína, ADN genómico y una membrana plasmática. Un componente de una célula huésped puede ser una parte de una planta de la que se expresa y/o secreta el polipéptido. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% o más de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente el 0,001, el 0,005, el 0,01, el 0,05, el 0,1, el 0,5, el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9 o el 10% de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por millón de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por millón de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más partes por mil millones de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y como máximo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 partes por mil millones de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y ser al menos aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped. Una composición puede comprender un producto cárnico consumible y ser como máximo aproximadamente el 1, el 2, el 3, el 4, el 5, el 6, el 7, el 8, el 9, el 10, el 20, el 30, el 40, el 50, el 60, el 70, el 80, el 90, el 91, el 92, el 93, el 94, el 95, el 96, el 97, el 98, el 99 o el 100% libre de un componente de una célula huésped.

En algunos casos, la divulgación describe un producto consumible que puede estar sustancialmente o completamente compuesto por componentes derivados de fuentes de origen no animal, que aún recapitula características clave asociadas con el cocinado y el consumo de un producto cárnico equivalente derivado de animales. El producto cárnico equivalente puede ser una carne blanca o una carne oscura. El producto cárnico equivalente puede derivarse de cualquier animal. Los ejemplos no limitativos de animales usados para derivar el producto cárnico equivalente incluyen animales de granja tales como, por ejemplo, ganado vacuno, oveja, cerdo, pollo, pavo, oca, pato, caballo, perro o animales de caza (ya sean salvajes o de granja) tal como, por ejemplo, conejo, ciervo, bisonte, búfalo, jabalí, serpiente, faisán, codorniz, oso, alce, antílope, pichón, paloma, urogallo, zorro, jabalí, cabra, canguro, emú, caimán, cocodrilo, tortuga, marmota canadiense, marmota, zarigüeya, perdiz, ardilla, mapache, ballena, foca, avestruz, carpincho, nutria, cobaya, rata, ratones, topillo, cualquier variedad de insecto u otro artrópodo, marisco tal como, por ejemplo, pescado, cangrejo, langosta, ostra, mejillón, vieira, oreja de mar, calamar, pulpo, erizo de mar, tunicado y otros. Muchos productos cárnicos se derivan normalmente de músculos esqueléticos de un animal, pero se entiende que la carne también puede proceder de otros músculos u órganos del animal. En algunos casos, el producto cárnico equivalente es un corte de carne derivada de músculo esquelético. En algunas variantes, el producto cárnico equivalente es un órgano tal como, por ejemplo, un riñón, corazón, hígado, vesícula biliar, intestino, estómago, médula ósea, cerebro, timo, pulmón, lengua. Por consiguiente, en algunos casos, las composiciones del presente documento son productos consumibles similares a órganos o músculos esqueléticos.

En algunos casos, la divulgación describe productos de sucedáneo de carne que comprenden una o más de una primera composición que comprende una réplica de tejido muscular, una segunda composición que comprende una réplica de tejido adiposo y/o una tercera composición que comprende una réplica de tejido conjuntivo, en los que la una o más composiciones se combinan de una manera que recapitula la organización física de la carne. En otros casos, la presente divulgación describe composiciones para una réplica de tejido muscular (denominada en el presente documento “réplica de músculo”), una réplica de tejido adiposo (denominada en el presente documento “réplica de grasa”) y una réplica de tejido conjuntivo (denominada en el presente documento “réplica de tejido conjuntivo”). En algunos casos, las composiciones y los productos de sucedáneo de carne están compuestos principalmente o completamente por componentes derivados de fuentes de origen no animal. En variantes alternativas, la réplica de músculo, grasa y/o tejido conjuntivo, o los productos de sucedáneo de carne que comprenden una o más de dichas réplicas, se derivan parcialmente de fuentes de origen animal pero se complementan con componentes derivados de fuentes de origen no animal.

En algunos casos, los productos cárnicos pueden derivarse sustancialmente de fuentes de origen animal, pero que se complementan con una o más de una réplica de tejido muscular, una réplica de grasa y/o una réplica de tejido

conjuntivo, en los que las réplicas pueden derivarse sustancialmente o completamente de fuentes de origen no animal. Un ejemplo no limitativo de un producto cárnico de este tipo es un producto de carne de vacuno picada ultramagra complementada con una réplica de grasa de origen no animal que puede mejorar la textura y la sensación en la boca a la vez que conserva los beneficios de salud de un producto consumible bajo en grasa animal. Tales variantes alternativas dan como resultado productos con propiedades que recapitulan más estrechamente características clave asociadas con la preparación y el consumo de carne, pero que son menos costosos y están asociados con un menor impacto medioambiental, menos impacto en el bienestar animal o beneficios de salud mejorados para el consumidor.

La organización física del producto de sucedáneo de carne puede manipularse controlando la ubicación, la organización, el ensamblaje o la orientación de las réplicas de músculo, grasa y/o tejido conjuntivo descritas en el presente documento. En algunos casos, el producto se diseña de tal manera que las réplicas descritas en el presente documento se asocian entre sí como en la carne. En algunos casos, el producto consumible se diseña de manera que, después del cocinado, las réplicas descritas en el presente documento se asocian entre sí como en la carne cocinada. En algunas variantes, una o más de las réplicas de músculo, grasa y/o tejido conjuntivo se combinan de tal manera que recapitula la organización física de diferentes cortes o preparaciones de carne. En una variante de ejemplo, las réplicas se combinan de tal manera que se aproxima a la organización física de carne picada natural. En otras variantes, las réplicas se combinan de tal manera que se aproxima a diferentes cortes de carne de vacuno, tales como, por ejemplo, lomo alto, filete miñón, asado londinense, entre otros.

Indicadores para cocinado de carne

En algunos casos, puede usarse un polipéptido de la divulgación en una composición de la divulgación como indicador para cocinado de carne. La liberación de odorizantes tras el cocinado es un aspecto importante del consumo de carne. En algunas variantes, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocinan generan un aroma reconocible por los seres humanos como típico del cocinado de carne de vacuno. En algunos casos, el producto consumible cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de carne de cerdo. En algunos casos, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de beicon. En algunos casos, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de pollo. En algunas variantes, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de cordero. En algunos casos, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de pescado. En algunos casos, el producto consumible es una réplica de la carne que se compone totalmente de productos de origen no animal que cuando se cocina genera un aroma reconocible por los seres humanos como típico del cocinado de pavo. En algunos casos el producto consumible es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, con un odorizante que se libera tras el cocinado. En algunas variantes, el producto consumible es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, con un odorizante que se produce por reacciones químicas que tienen lugar tras el cocinado. En algunos casos, el producto consumible es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, que comprende un polipéptido de la divulgación y mezclas de proteínas, péptidos, aminoácidos, nucleótidos, azúcares y polisacáridos y grasas en combinaciones y disposiciones espaciales que permiten que estos compuestos experimenten reacciones químicas durante el cocinado para producir odorizantes y compuestos productores de sabor. En algunas variantes, el producto consumible es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal (por ejemplo, un polipéptido de la divulgación), con un odorizante volátil o lábil que se libera tras el cocinado. En algunos casos, el producto consumible es un método para preparar una réplica de la carne en el que se calientan las réplicas de la carne compuestas principal o totalmente de componentes derivados de fuentes de origen no animal para liberar un odorizante volátil o lábil.

Los odorizantes liberados durante el cocinado de carne se generan por reacciones que pueden implicar como reactantes grasas, proteína, aminoácidos, péptidos, nucleótidos, ácidos orgánicos, compuestos de azufre, azúcares y otros hidratos de carbono. En algunos casos, un reactante puede ser un polipéptido de la divulgación (por ejemplo, una globina, una globina secretada). En algunos casos, los odorizantes que se combinan durante el cocinado de carne se identifican y ubican cerca entre sí en el producto consumible, de manera que tras el cocinado del producto consumible se combinan los odorizantes. En algunas variantes, los componentes de sabor y aroma característicos se producen durante el procedimiento de cocinado por reacciones químicas que implican aminoácidos, grasas y azúcares encontrados en plantas, así como en carne. En algunos casos, los componentes de sabor y aroma característicos se producen la mayoría durante el procedimiento de cocinado por reacciones químicas que implican uno o más aminoácidos, grasas, péptidos, nucleótidos, ácidos orgánicos, compuestos de azufre, azúcares y otros hidratos de carbono encontrados en plantas, así como en carne.

Algunas reacciones que generan odorizantes liberados durante el cocinado de carne pueden catalizarse por hierro, en particular el hierro del hemo, que puede estar compuesto (por ejemplo, unido) por un polipéptido de la divulgación. Por tanto, en algunos casos, algunos de los componentes de sabor y aroma característicos se producen durante el procedimiento de cocinado por reacciones químicas catalizadas por hierro. En algunas variantes, algunos de los componentes de sabor y aroma característicos se producen durante el procedimiento de cocinado por reacciones químicas catalizadas por hemo. En algunos casos, algunos de los componentes de sabor y aroma característicos se producen durante el procedimiento de cocinado por reacciones químicas catalizadas por el hierro del hemo en leghemoglobina. En algunos casos, algunos de los componentes de sabor y aroma característicos se producen durante el procedimiento de cocinado por reacciones químicas catalizadas por el hierro del hemo en una hemoproteína (por ejemplo, los polipéptidos enumerados en la figura 9, hemoglobina, mioglobina, neuroglobina, citoglobina, leghemoglobina, hemoglobina no simbiótica, globina I de Hell's gate, hemoglobinas bacterianas, mioglobinas de ciliado, flavohemoglobinas, androglobina, citoglobina, globina E, globina X, globina Y, mioglobina, leghemoglobinas, eritrocruorinas, hemoglobinas beta, hemoglobinas alfa, hemoglobinas no simbióticas, protoglobinas, cianoglobinas, globina I de Hell's gate, hemoglobinas bacterianas, mioglobinas de ciliado, histoglobinas y neuroglobinas, etc.).

Indicadores del color

El color de la carne es una parte importante de la experiencia de cocinado y consumo de la carne. Por ejemplo, los cortes de carne de vacuno son de un color rojo característico en estado crudo y pasan gradualmente a un color marrón durante el cocinado. Como otro ejemplo, las carnes blancas tales como el pollo o la carne de cerdo tienen un color rosa característico en su estado crudo y pasan gradualmente a un color blanco o amarronado durante el cocinado. La cantidad de la transición de color se usa para indicar la progresión del cocinado de carne de vacuno y valorar el tiempo y la temperatura de cocinado para producir el estado deseado de acabado. La divulgación describe un producto sucedáneo de carne no basado en carne que proporciona un indicador visual de progresión del cocinado. En algunos casos, el indicador visual es un indicador de color que experimenta una transición de color durante el cocinado. En casos particulares, el indicador de color recapitula la transición de color de un corte de carne a medida que la carne progresa de un estado crudo a un estado cocinado. En algunas variantes, el indicador de color proporciona un color rojo al producto sucedáneo de carne antes del cocinado para indicar un estado crudo y produce que el producto sucedáneo de carne pase a un color marrón durante la progresión del cocinado. En algunos casos, el indicador de color proporciona un color rosa al producto sucedáneo de carne antes del cocinado para indicar un estado crudo y produce que el producto sucedáneo de carne pase a un color blanco o marrón durante la progresión del cocinado.

El determinante principal de la definición nutricional del color de la carne es la concentración de proteínas que portan hierro en la carne. En el componente de músculo esquelético de productos cárnicos, una de las principales proteínas que portan hierro es la mioglobina. Así, en algunos casos, la composición es un producto cárnico consumible (por ejemplo, una réplica) que comprende una proteína que porta hierro. En algunas variantes, la composición comprende aproximadamente el 0,05%, aproximadamente el 0,1%, aproximadamente el 0,2%, aproximadamente el 0,3%, aproximadamente el 0,4%, aproximadamente el 0,5%, aproximadamente el 0,6%, aproximadamente el 0,7%, aproximadamente el 0,8%, aproximadamente el 0,9%, aproximadamente el 1%, aproximadamente el 1,1%, aproximadamente el 1,2%, aproximadamente el 1,3%, aproximadamente el 1,4%, aproximadamente el 1,5%, aproximadamente el 1,6%, aproximadamente el 1,7%, aproximadamente el 1,8%, aproximadamente el 1,9%, aproximadamente el 2%, o más de aproximadamente el 2% de una proteína que porta hierro en peso seco o peso total. En algunos casos, la composición comprende al menos aproximadamente el 10% de un polipéptido de la divulgación. En algunos casos, la composición comprende como máximo aproximadamente el 10% de un polipéptido de la divulgación. En algunos casos, la proteína que porta hierro se ha aislado y purificado de una fuente. En otros casos, la proteína que porta hierro no se ha aislado ni purificado. En algunos casos, la fuente de la proteína que porta hierro es una fuente de origen animal, o una fuente de origen no animal tal como una planta, un hongo u organismos modificados genéticamente tales como, por ejemplo, bacterias o levadura. En algunos casos, la proteína que porta hierro es mioglobina. En algunos casos, la composición comprende un producto consumible que es una réplica de la carne a base de vegetales que tiene mioglobina animal añadida. Así, por ejemplo, una réplica de carne de ternera puede tener aproximadamente el 0,4-1% de mioglobina. En algunos casos, la proteína que porta hierro es leghemoglobina. En algunas variantes, la composición comprende un producto consumible que es una réplica de la carne a base de vegetales que tiene leghemoglobina añadida. Así, por ejemplo, una réplica de carne de ternera puede tener aproximadamente el 0,4-1% leghemoglobina. En algunos casos, la proteína que porta hierro es un citocromo. En algunas variantes, la composición comprende un producto consumible que es una réplica de la carne a base de vegetales que tiene un citocromo añadido. Así, por ejemplo, una réplica de carne de ternera puede tener aproximadamente el 0,4-1% de un citocromo. En algunos casos, el producto consumible es una réplica de la carne a base de vegetales que contiene hemoglobina. En algunos casos, la proteína que porta hierro es un polipéptido de la divulgación (por ejemplo, una globina).

Las proteínas que contienen hierro adicionales existen en la naturaleza. En algunos casos, la composición (por ejemplo, producto consumible) comprende una proteína que contiene hierro que no es mioglobina. En algunas variantes, la composición (por ejemplo, producto consumible) no contiene mioglobina. En algunos casos, las composiciones (por ejemplo, producto consumible) no contienen hemoglobina. En algunos casos, el producto

consumible es una réplica de la carne que comprende una proteína que contiene hierro distinta de mioglobina o hemoglobina (por ejemplo, las globinas enumeradas en la figura 9, y descritas en el presente documento, por ejemplo, hemoglobina, mioglobina, neuroglobina, citoglobina, leghemoglobina, hemoglobina no simbiótica, globina I de Hell's gate, hemoglobinas bacterianas, mioglobinas de ciliado, flavohemoglobinas).

En algunos casos, la composición comprende un producto consumible que es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, incluyendo una réplica de tejido muscular, una réplica de tejido adiposo, una réplica de tejido conjuntivo y leghemoglobina. En algunas variantes, la composición comprende un producto consumible que es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, que contiene una hemoproteína. En algunos casos, la composición comprende un producto consumible que es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, que contiene una leghemoglobina. En algunos casos, la composición comprende un producto consumible que es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, que contiene un miembro de la familia de proteínas globina. En algunos casos, la composición comprende un producto consumible que es una réplica de la carne que se compone principal o totalmente de componentes derivados de fuentes de origen no animal, con un alto contenido de hierro de una hemoproteína. En algunas variantes, el contenido de hierro es similar al de la carne. En algunos casos, el producto consumible tiene el color rojo distintivo de la carne, proporcionado tal color por la leghemoglobina.

Se usa leghemoglobina, en algunos casos, como un indicador de que el producto consumible está acabado de cocinar. En algunas variantes de la divulgación, se da a conocer un método para cocinar un producto consumible que comprende la detección de leghemoglobina que ha migrado desde el interior del producto consumible hasta la superficie cuando el producto se cocina. En algunas variantes de la divulgación, se describe un método para cocinar un producto consumible que comprende la detección del cambio de color de rojo a marrón cuando se cocina el producto.

El estado de oxidación del ion de hierro en la leghemoglobina puede ser importante para su color. La leghemoglobina con el hierro del hemo en el estado de oxidación +2 puede parecer de un color rojo intenso, mientras que la leghemoglobina con el hierro del hemo en el estado de oxidación +3 puede parecer rojo amarronado. Por tanto, usando leghemoglobina como fuente de color rojo en una réplica de la carne, por ejemplo, puede ser deseable reducir el hierro del hemo del estado +3 al estado +2. El hierro del hemo en leghemoglobina puede cambiarse del estado oxidado (+3) al estado reducido (+2) con reactivos reductores.

Una hemoproteína puede usarse, en algunos casos, como un indicador de que el producto consumible está acabado de cocinar. En algunos, se da a conocer un método para cocinar un producto consumible que comprende la detección de leghemoglobina que ha migrado desde el interior del producto consumible hasta la superficie cuando se cocina el producto. En algunos casos, se da a conocer un método para cocinar un producto consumible que comprende la detección del cambio de color de rojo a marrón cuando se cocina el producto.

Una hemoproteína (por ejemplo, hemoglobina, mioglobina, neuroglobina, citoglobina, leghemoglobina, hemoglobina no simbiótica, globina I de Hell's gate, hemoglobinas bacterianas, mioglobinas de ciliado, flavohemoglobinas) puede usarse, en algunos casos, como un indicador de que el producto consumible está acabado de cocinar. Así, en algunos casos, la divulgación describe un método para cocinar un producto consumible que comprende la detección de leghemoglobina que ha migrado desde el interior del producto consumible hasta la superficie cuando se cocina el producto. La divulgación describe un método para cocinar un producto consumible que comprende la detección del cambio de color de rojo a marrón cuando se cocina el producto.

Productos alimenticios que comprenden polipéptido purificado

En algunos casos, un polipéptido de la divulgación (por ejemplo, un polipéptido que contiene hemo, una globina tal como leghemoglobina) se añade a la carne para mejorar las propiedades de la carne. Véanse, por ejemplo, los documentos WO 2014/110532, WO 2014/110539 y WO 2013/010042. Por ejemplo, una disolución que contiene polipéptidos puede inyectarse en carne cruda o cocinada. En otro ejemplo, una disolución que comprende un polipéptido de la divulgación se hace gotear sobre la carne o un producto consumible para mejorar el aspecto. En algunos casos, la publicidad, fotografía o videografía de productos alimenticios tales como carne o un sucedáneo de carne se mejora con leghemoglobina.

Por ejemplo, pueden combinarse polipéptidos, leghemoglobina y hemoglobina, con otros componentes de la réplica de la carne a base de vegetales. En algunos casos, los polipéptidos se capturan en un gel que contiene otros componentes, por ejemplo, lípidos y/o proteínas. En algunos casos, se combinan múltiples geles con hemoproteínas no basadas en geles. En algunos casos, la combinación de los polipéptidos y los otros compuestos del producto consumible se realiza para garantizar que las hemoproteínas puedan difundirse a través del producto consumible. En algunas variantes, el producto consumible comprende una disolución que contiene hemoproteínas, por ejemplo, una disolución de leghemoglobina. En algunos casos, el producto consumible se remoja en una disolución que contiene hemoproteínas, por ejemplo, una disolución de leghemoglobina durante 1, 5, 10, 15, 20 ó 30 horas. En

algunas variantes, el producto consumible se remoja en una disolución que contiene hemo, por ejemplo, una disolución de leghemoglobina durante 1, 5, 10, 15, 30 ó 45 minutos.

Réplicas de músculo

Un gran número de productos cárnicos comprende una alta proporción de músculo esquelético. Por consiguiente, la presente divulgación describe una composición derivada de fuentes de origen no animal que replica o se aproxima a características clave de músculo esquelético animal. La presente divulgación también describe un producto sucedáneo de carne que comprende una composición derivada de fuentes de origen no animal que replica o se aproxima a músculo esquelético animal. Una composición de este tipo se denominará en el presente documento "réplica de músculo". En algunos casos, la réplica de músculo y/o el producto sucedáneo de carne que comprende la réplica de músculo se deriva parcialmente de fuentes de origen animal. En algunos casos, la réplica de músculo y/o el producto sucedáneo de carne que comprende la réplica de músculo se deriva totalmente de fuentes de origen no animal.

Muchos productos cárnicos comprenden una alta proporción de músculo esquelético estriado en el que las fibras musculares individuales se organizan principalmente de una manera isotrópica. Por consiguiente, en algunas variantes, la réplica de músculo comprende fibras que están en cierta medida organizadas isotrópicamente. En algunos casos, las fibras comprenden un componente proteico. En algunos casos, las fibras comprenden aproximadamente el 1%, aproximadamente el 2%, aproximadamente el 5%, aproximadamente el 10%, aproximadamente el 15%, aproximadamente el 20%, aproximadamente el 30%, aproximadamente el 40%, aproximadamente el 50%, aproximadamente el 60%, aproximadamente el 70%, aproximadamente el 80%, aproximadamente el 90%, aproximadamente el 95%, aproximadamente el 99% o más de un componente proteico. El músculo esquelético animal contiene normalmente alrededor del 1% de mioglobina, pero puede ser como mucho el 7% de masa muscular en algunos músculos de ballena. En algunas variantes, la réplica de músculo comprende hemoglobinas de esta divulgación.

En algunos casos, el componente proteico comprende una o más proteínas aisladas purificadas. Por ejemplo, la una o más proteínas aisladas purificadas puede comprender la globulina 8S de semillas de soja verde, o la fracción de albúmina o globulina de semillas de guisante. Estas proteínas proporcionan ejemplos de proteínas con propiedades favorables para construir réplicas de la carne debido a su capacidad para formar geles con texturas similares a tejido adiposo o músculo animal. Los ejemplos de la una o más proteínas aisladas purificadas se describen en el presente documento. La lista de posibles candidatas es, en el presente documento, esencialmente abierta y puede incluir rubisco, cualquier proteína principal de almacenamiento de semillas, proteínas aisladas de hongos, bacterias, arqueas, virus o microorganismos modificados por ingeniería genética, o sintetizada *in vitro*. Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de tejido muscular animal. Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de tejido muscular animal. En algunas variantes, una o más proteínas aisladas purificadas representan aproximadamente el 0,1%, el 0,2%, el 0,5%, el 1%, el 2%, el 3%, el 4%, el 5%, el 6%, el 7%, el 8%, el 9%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 99% o más del componente proteico en peso.

El músculo esquelético de animales, tales como ganado vacuno de carne, contiene normalmente cantidades sustanciales de glucógeno, que puede comprender del orden de >1% de la masa del tejido muscular en el momento de sacrificio. Después del sacrificio, una fracción de este glucógeno continúa metabolizándose produciendo productos incluyendo ácido láctico, que contribuye a reducir el pH del tejido muscular, una cualidad deseable en la carne. El glucógeno es un polímero ramificado de glucosa unido entre sí por enlaces alfa (1->4) glucosídicos en cadenas lineales, con puntos de ramificación que comprenden enlaces alfa (1->6) glucosídicos. Los almidones de plantas, particularmente amilopectinas, son también polímeros ramificados de glucosa unidos entre sí por enlaces alfa (1->4) glucosídicos en cadenas lineales, con puntos de ramificación que comprenden enlaces alfa (1->6) glucosídicos y, por tanto, pueden usarse como un análogo de glucógeno en la construcción de réplicas de la carne. Por tanto, en algunos casos, el músculo o la réplica de la carne incluyen un almidón o una pectina.

Los componentes adicionales de tejido muscular animal incluyen sodio, potasio, calcio, magnesio, otros iones metálicos, ácido láctico, otros ácidos orgánicos, aminoácidos libres, péptidos, nucleótidos y compuestos de azufre. Por tanto, en algunos casos, la réplica de músculo puede incluir sodio, potasio, calcio, magnesio, otros iones metálicos, ácido láctico, otros ácidos orgánicos, aminoácidos libres, péptidos, nucleótidos y compuestos de azufre. En algunos casos, la concentración de sodio, potasio, calcio, magnesio, otros iones metálicos, ácido láctico, otros ácidos orgánicos, aminoácidos libres, péptidos, nucleótidos y/o compuestos de azufre en la réplica de músculo o el producto consumible están dentro de >10% de las concentraciones encontradas en un músculo o una carne que va a replicarse.

La divulgación describe métodos para elaborar una réplica de músculo. En algunos casos, la composición se conforma en fibras asimétricas antes de la incorporación en el producto consumible. En algunos casos, estas fibras replican fibras musculares. En algunos casos, las fibras son fibras tejidas. En otras variantes, las fibras son fibras extruidas. Por consiguiente, la presente divulgación describe métodos para producir fibras proteicas asimétricas o

tejidas. En algunos casos, las fibras se forman mediante extrusión del componente proteico a través de una extrusora.

En algunos casos, la extrusión puede realizarse usando una extrusora de doble husillo MPF19 (APV Baker, Grand Rapids, Mich.) con una boquilla de enfriamiento. La boquilla de enfriamiento puede enfriar el extruido antes de que el extruido vuelva a presión atmosférica, por tanto, se impide sustancialmente la expansión o el inflado del producto final. En el aparato MPF19, pueden añadirse por separado alimentación seca y líquido y mezclarse en el tambor giratorio. Los parámetros de extrusión pueden ser, por ejemplo: velocidad de husillo de 200 rpm, temperatura de producto en la boquilla de 150ºC, velocidad de alimentación de 23 g/min, y velocidad de flujo de agua de 11 g/min. La temperatura de producto puede medirse durante la extrusión mediante un termopar en el extremo del tambor giratorio de extrusión. Pueden realizarse observaciones sobre el color, la opacidad, la estructura y la textura para cada muestra recogida. Las muestras recogidas pueden secarse opcionalmente a temperatura ambiente durante la noche, luego molerse hasta dar un polvo fino (con un tamaño de <60 de malla) usando una trituradora de alimentos Braun. El pH de las muestras puede medirse por duplicado usando suspensiones al 10% (p/v) de muestra en polvo en agua destilada.

Réplica de grasa

La grasa animal es importante para la experiencia de consumo de carne cocinada. Por consiguiente, la presente divulgación describe una composición derivada de fuentes de origen no animal que recapitula características clave de la grasa animal. Además, la presente divulgación describe un producto sucedáneo de carne que comprende una composición derivada de fuentes de origen no animal que recapitula grasa animal. Una composición de este tipo se denominará en el presente documento “réplica de grasa”. En algunos casos, la réplica de grasa y/o el producto sucedáneo de carne que comprende la réplica de grasa se derivan parcialmente de fuentes de origen animal.

En algunos casos, el producto sucedáneo de carne tiene un componente de grasa. En algunos casos, el contenido de grasa del producto consumible es del 1%, el 5%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55% o el 60% de grasa. En algunas variantes, la réplica de grasa comprende un gel con gotitas de grasa suspendidas en el mismo. En algunos casos, el gel es un gel suave elástico que comprende proteínas y opcionalmente hidratos de carbono. En variantes particulares, las proteínas usadas en el gel son proteínas vegetales o microbianas. En algunos casos, las proteínas usadas en la réplica de grasa pueden incluir rubisco, cualquier proteína principal de almacenamiento de semillas, proteínas aisladas de hongos, bacterias, arqueas, virus o microorganismos modificados por ingeniería genética, o sintetizada *in vitro*. Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de la grasa animal. Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de la grasa animal.

Las gotitas de grasa usadas en algunas variantes de la presente divulgación pueden ser de una variedad de fuentes. En algunos casos, las fuentes son fuentes de origen no animal. En casos particulares, las fuentes son fuentes de origen vegetal. Los ejemplos no limitativos de aceites incluyen aceite de maíz, aceite de oliva, aceite de soja, aceite de cacahuete, aceite de nuez, aceite de almendra, aceite de sésamo, aceite de semilla de algodón, aceite de colza, aceite de canola, aceite de cártamo, aceite de girasol, aceite de linaza, aceite de algas, aceite de palma, aceite de palmiste, aceite de coco, aceite de babasú, manteca de karité, manteca de mango, manteca de cacao, aceite de germen de trigo, aceite de afrecho de arroz, aceites producidos por bacterias, algas, arqueas u hongos o bacterias, algas, arqueas u hongos modificados por ingeniería genética, triglicéridos, monoglicéridos, diglicéridos, esfingósidos, glicolípidos, lecitina, lisolecitina, ácidos fosfatídicos, ácidos lisofosfatídicos, ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosanoico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, ácido oleico conjugado, o ésteres de: ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosanoico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2 o ácido oleico conjugado, o ésteres de glicerol de ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosanoico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, o ácido oleico conjugado, o derivados de triglicéridos de ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosanoico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2 o ácido oleico conjugado.

En algunos casos, las gotitas de grasa se derivan de aceite de pulpa o semillas. En otros casos, la fuente puede ser levadura o moho. Por ejemplo, en una variante, las gotitas de grasa comprenden triglicéridos derivados de *Mortierella isabellina.*

En algunas variantes, los aceites vegetales se modifican para parecerse a grasas animales. Los aceites vegetales pueden modificarse con aromatizante u otros agentes para recapitular el sabor y el olor de la carne durante y después del cocinado. Por consiguiente, algunas variantes de la divulgación implican métodos para someter a

prueba la semejanza cualitativa entre las propiedades de cocinado de grasa animal y las propiedades de cocinado de aceites vegetales en el producto consumible.

En algunos casos, la réplica de grasa comprende un componente proteico que comprende una o más proteínas aisladas purificadas. Las proteínas purificadas contribuyen al sabor y la textura de la réplica de la carne. En algunas variantes, las proteínas purificadas pueden estabilizar grasas emulsionadas. En algunos casos, las proteínas purificadas pueden formar geles tras la desnaturalización o reticulación enzimática, que replican el aspecto y la textura de la grasa animal. Los ejemplos y las variantes de las una o más proteínas aisladas purificadas se describen en el presente documento. En casos particulares, la una o más proteínas aisladas comprenden una proteína aislada de la familia de las legumbres de plantas. Los ejemplos no limitativos de plantas leguminosas se describen en el presente documento, aunque son posibles variaciones con otras legumbres. En algunos casos, la planta leguminosa es una planta de guisantes. En algunos casos, las proteínas aisladas purificadas estabilizan emulsiones. En algunos casos, las proteínas aisladas purificadas forman geles tras la reticulación o reticulación enzimática. En algunas variantes, las proteínas aisladas purificadas comprenden proteínas de almacenamiento de semillas. En algunos casos, las proteínas aisladas purificadas comprenden albúmina. En algunos casos, las proteínas aisladas purificadas comprenden globulina. En una variante particular, la proteína aislada purificada es una proteína de albúmina de guisante purificada. En otra variante particular, la proteína aislada purificada es una proteína de globulina de guisante purificada. En otra variante particular, la proteína aislada purificada es una globulina 8S de soja verde. En otra variante particular, la proteína aislada purificada es una oleosina. En otra variante particular, la proteína aislada purificada es una caloleosina. En otra variante particular, la proteína aislada purificada es rubisco. En algunas variantes, el componente proteico comprende aproximadamente el 0,1%, el 0,5%, el 1%, el 2%, el 5%, el 10%, el 15%, el 20%, el 25%, el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90% o más de la réplica de grasa en peso seco o peso total. En algunas variantes, el componente proteico comprende aproximadamente el 0,1-5%, aproximadamente el 0,5-10%, aproximadamente el 1-20%, aproximadamente el 5-30%, aproximadamente el 10-50%, aproximadamente el 20-70%, o aproximadamente el 30-90% o más de la réplica de grasa en peso seco o peso total. En algunas variantes, el componente proteico comprende una disolución que contiene una o más proteínas aisladas purificadas.

En algunos casos, la réplica de grasa comprende enzimas de reticulación que catalizan reacciones que conducen a reticulaciones covalentes entre proteínas. Pueden usarse enzimas de reticulación para crear o estabilizar la estructura y textura deseadas de la réplica de tejido adiposo, para imitar la textura deseada de una grasa animal deseada equivalente. Los ejemplos no limitativos de enzimas de reticulación incluyen, por ejemplo, transglutaminasa, lisil oxidasas u otras amino oxidasas (por ejemplo, lisil oxidasa de *Pichia pastoris*). En algunos casos, las enzimas de reticulación se aíslan y purifican de una fuente de origen no animal, cuyos ejemplos y versiones se describen en el presente documento. En algunos casos, la réplica de grasa comprende al menos el 0,0001%, o al menos el 0,001%, o al menos el 0,01%, o al menos el 0,1%, o al menos el 1% (p/v) de una enzima de reticulación. En variantes particulares, la enzima de reticulación es transglutaminasa.

La divulgación describe además métodos para elaborar una réplica de grasa. En algunos casos, las gotitas de grasa se suspenden en un gel. En algunas variantes, la presente divulgación describe métodos para producir gotitas de grasa suspendidas en el gel. La grasa puede aislarse y homogeneizarse. Por ejemplo, puede usarse una mezcla de disolvente orgánico para ayudar a mezclar un lípido. Luego puede eliminarse el disolvente. En este punto, el lípido puede congelarse, liofilizarse o almacenarse. Así, en algunas variantes, la divulgación describe un método para aislar y almacenar un lípido que se ha seleccionado para tener características similares a la grasa animal. Luego puede hidratarse la película o torta lipídica. La hidratación puede utilizar agitación o cambios de temperatura. La hidratación puede producirse en una disolución precursora para dar un gel. Después de la hidratación, puede tratarse la suspensión de lípidos por ultrasonidos o extruirse para alterar adicionalmente las propiedades del lípido en la disolución.

En algunos casos, la réplica de grasa se agrupa para aproximarse a la organización del tejido adiposo en la carne. En algunos casos, algunos o todos los componentes de la réplica de grasa se suspenden en un gel. En diversas variantes, el gel puede ser un gel proteínico, un hidrogel, un organogel o un xerogel. En algunas variantes, el gel puede espesarse hasta la consistencia deseada usando un agente basado en polisacáridos o proteínas. Por ejemplo, pueden usarse fécula, arrurruz, almidón de maíz, almidón de katakuri, almidón de patata, palmera de sagú, tapioca, alginina, goma guar, goma de algarrobo, goma de xantano, colágeno, claras de huevo, furcelarán, gelatina, agar, carragenano, celulosa, metilcelulosa, hidroximetilcelulosa, goma acacia, konjac, almidón, pectina, amilopectina o proteínas derivadas de legumbres, cereales, nueces, otras semillas, hojas, algas, bacterias u hongos, solos o en combinación, para espesar el gel, formando una arquitectura o estructura para el producto consumible.

En variantes particulares, la réplica de grasa es una emulsión que comprende una disolución de una o más proteínas y una o más grasas suspendidas en la misma como gotitas. En algunos casos, la emulsión se estabiliza por una o más enzimas de reticulación en un gel. En algunos casos, la una o más proteínas en disolución son proteínas aisladas purificadas. En algunos casos, las proteínas aisladas purificadas comprenden una fracción enriquecida en albúmina de guisante purificada. En algunas variantes, las proteínas aisladas purificadas comprenden una fracción enriquecida en globulina de guisante purificada. En algunos casos, las proteínas aisladas purificadas comprenden una fracción enriquecida en globulina 8S de soja verde purificada. En algunos casos, las

proteínas aisladas purificadas comprenden una fracción enriquecida en rubisco. En algunos casos, la una o más grasas se derivan de aceites a base de vegetales. En algunos casos, la una o más grasas se derivan de unos o más de: aceite de maíz, aceite de oliva, aceite de soja, aceite de cacahuete, aceite de nuez, aceite de almendra, aceite de sésamo, aceite de semillas de algodón, aceite de colza, aceite de canola, aceite de cártamo, aceite de girasol, aceite de linaza, aceite de algas, aceite de palma, aceite de palmiste, aceite de coco, aceite de babasú, manteca de karité, manteca de mango, manteca de cacao, aceite de germen de trigo, aceite de afrecho de arroz, aceites producidos por bacterias, algas, arqueas u hongos o bacterias, algas, arqueas u hongos modificados por ingeniería genética, triglicéridos, monoglicéridos, diglicéridos, esfingósidos, glicolípidos, lecitina, lisolecitina, ácidos fosfatídicos, ácidos lisofosfatídicos, ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosaonico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, ácido oleico conjugado, o ésteres de: ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosaonico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, o ácido oleico conjugado, o ésteres de glicerol de ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosaonico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, o ácido oleico conjugado, o derivados de triglicéridos de ácido oleico, ácido palmitoleico, ácido palmítico, ácido mirístico, ácido láurico, ácido miristoleico, ácido caproico, ácido cáprico, ácido caprílico, ácido pelargónico, ácido undecanoico, ácido linoleico, ácido eicosaonico 20:1, ácido araquidónico, ácido eicosapentanoico, ácido docosahexanoico, ácido linoleico conjugado 18:2, o ácido oleico conjugado. En aún variantes incluso más particulares, la una o más grasas es un aceite de afrecho de arroz. En otra variante particular, la una o más grasas es un aceite de canola. En algunas variantes, la enzima de reticulación es transglutaminasa, lisil oxidasa u otra amino oxidasa. En algunas variantes, la enzima de reticulación es transglutaminasa. En variantes particulares, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de albúmina de guisante purificada emulsionada con el 40-80% de aceite de afrecho de arroz, estabilizada con el 0,5-5% (p/v) de transglutaminasa para dar un gel. En algunas variantes, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de globulina 8S de soja verde parcialmente purificada emulsionada con el 40-80% de aceite de afrecho de arroz, estabilizada con el 0,5-5% (p/v) de transglutaminasa para dar un gel. En algunas variantes, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de globulina 8S de soja verde parcialmente purificada emulsionada con el 40-80% de aceite de canola, estabilizada con el 0,5-5% (p/v) de transglutaminasa para dar un gel. En algunas variantes, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de albúmina de guisante purificada emulsionada con >40-80% de aceite de afrecho de arroz, estabilizada con el 0,0001-1% (p/v) de transglutaminasa para dar un gel. En algunas variantes, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de globulina 8S de soja verde parcialmente purificada emulsionada con el 40-80% de aceite de afrecho de arroz, estabilizada con el 0,0001 - 1% (p/v) de transglutaminasa para dar un gel. En algunas variantes, la réplica de grasa es una emulsión con alto contenido de grasa que comprende una disolución de proteínas de globulina 8S de soja verde parcialmente purificada emulsionada con el 40-80% de aceite de canola, estabilizada con el 0,0001 -1% (p/v) de transglutaminasa para dar un gel.

Réplica de tejido conjuntivo

El tejido conjuntivo animal proporciona características clave de textura que son un importante componente de la experiencia de consumo de carne. Por consiguiente, la presente divulgación describe una composición derivada de fuentes de origen no animal que recapitula características clave de tejido conjuntivo animal. La presente divulgación da a conocer además un producto sucedáneo de carne que comprende una composición derivada de fuentes de origen no animal que recapitula importantes características de textura y visuales de tejido conjuntivo animal. Una composición de este tipo se denominará en el presente documento “réplica de tejido conjuntivo”. En algunas variantes, la réplica de tejido conjuntivo y/o el producto sucedáneo de carne que comprende la réplica de tejido conjuntivo se derivan parcialmente de fuentes de origen animal.

El tejido conjuntivo animal puede dividirse generalmente en tejido de tipo fascia y de tipo cartílago. El tejido de tipo fascia es muy fibroso, resistente contra la extensión (tiene alto módulo de elasticidad), y tiene un alto contenido de proteínas, un contenido de agua moderado (aprox. el 50%) y un contenido de grasas y polisacáridos nulo o bajo. Por consiguiente, la presente divulgación describe una réplica de tejido conjuntivo que recapitula características clave del tejido de tipo fascia. En algunos casos, la réplica de tejido conjuntivo comprende aproximadamente el 50% de proteína en peso total, aproximadamente el 50% en peso líquido, y tiene un bajo componente de grasas y polisacáridos.

El contenido de proteínas de la mayoría del tejido conjuntivo de tipo fascia está compuesto principalmente de colágeno. El colágeno se caracteriza por una alta fracción de prolina y alanina, y también se agrupa para dar fibrillas alargadas o estructuras flexibles similares a varillas características. Las prolaminas son una familia de proteínas encontradas en fuentes de origen no animal, tales como fuentes de origen vegetal. Las prolaminas son muy abundantes en plantas y son similares en composición de aminoácidos al colágeno. Entre las proteínas que se

sometieron a prueba para este fin, las prolaminas eran particularmente favorables debido a su bajo coste y su capacidad para formar fácilmente fibras o láminas cuando se hilan o extruyen. Los ejemplos no limitativos de proteínas de la familia de la prolamina incluyen, por ejemplo, zeína (encontrada en el maíz), estas incluyen hordeína de cebada, gliadina de trigo, secalina, extensinas de centeno, kafirina de sorgo, avenina de avena. En el tejido conjuntivo de tipo fascia, la familia de proteínas prolamina, de manera individual o combinaciones de las mismas, demuestra idoneidad para el componente proteico ya que son muy abundantes, similares en composición global de aminoácidos al colágeno (alta fracción de prolina y alanina) y susceptibles de procesarse para dar películas y fibras. Además de la zeína (encontrada en el maíz), estas incluyen hordeína de cebada, gliadina de trigo, secalina, extensinas de centeno, kafirina de sorgo, avenina de avena. Pueden ser necesarias otras proteínas para complementar las prolaminas con el fin de lograr especificaciones objetivo para propiedades fisicoquímicas y nutricionales. La lista de posibles candidatas es, en el presente documento, esencialmente abierta y puede incluir rubisco, cualquier proteína principal de almacenamiento de semillas, proteínas aisladas de hongos, bacterias, arqueas, virus o microorganismos modificados por ingeniería genética, o sintetizada *in vitro*. Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de tejido conjuntivo animal, colágeno de origen animal o recombinante, extensinas (glicoproteínas ricas en hidroxiprolina abundantes en las paredes celulares, por ejemplo, *Arabidopsis thaliana*, cuyos monómeros son moléculas flexibles similares a varillas “de tipo colágeno”). Las proteínas pueden diseñarse de manera artificial para emular propiedades físicas de tejido conjuntivo animal.

Los métodos para formar tejido conjuntivo de tipo fascia serán los practicados en la técnica con una preferencia hacia los métodos que producen estructuras fibrosas o de tipo fibroso por medios biológicos, químicos o físicos, de manera individual o en combinación, en serie o en paralelo, antes del conformado final. Estos métodos pueden incluir extrusión o hilado.

El tejido de tipo cartílago puede ser macroscópicamente homogéneo, resistente a la compresión, tiene mayor contenido de agua (hasta el 80%), menor contenido de proteínas (colágeno) y mayor contenido de polisacáridos (proteoglicanos) (aprox. el 10% de cada uno).

Con respecto a la composición, el tejido conjuntivo de tipo cartílago puede ser muy similar al tejido de tipo fascia con las razones relativas de cada uno ajustado a imitar más estrechamente el tejido conjuntivo de la “carne”.

Los métodos para formar tejido conjuntivo de tipo cartílago pueden ser similares a los del tejido conjuntivo de tipo fascia, pero con una preferencia hacia métodos que producen estructuras isotrópicamente homogéneas.

La grasa puede suspenderse en un gel. La presente divulgación describe métodos para producir gotitas de grasa suspendidas en el gel proteínico. La grasa puede aislarse de tejidos vegetales y emulsionarse. La emulsificación puede utilizar mezclado de alta velocidad, homogeneización, agitación o cambios de temperatura. La suspensión de lípidos puede tratarse por ultrasonidos o extruirse para alterar adicionalmente las propiedades del lípido en la disolución. En este punto, en algunos casos se añaden otros componentes del producto consumible a la disolución seguido por un agente de gelificación. En algunos casos, se añaden agentes de reticulación (por ejemplo, transglutaminasa o lisil oxidasa) para unir los componentes del producto consumible. En otras variantes, se añade el agente de gelificación y posteriormente se combina la suspensión de lípido/gel con componentes adicionales del producto consumible. En el tejido conjuntivo de tipo fascia, la familia de proteínas prolamina, individualmente o combinaciones de las mismas, demuestra idoneidad para el componente proteico ya que son muy abundantes, similares en composición global de aminoácidos al colágeno (alta fracción de prolina y alanina) y susceptibles a procesarse para dar películas. Además de la zeína (encontrada en el maíz), estas incluyen hordeína de cebada, gliadina de trigo, secalina, extensinas de centeno, kafirina de sorgo, avenina de avena. Pueden ser necesarias otras proteínas para complementar las prolaminas para lograr especificaciones objetivo para propiedades fisicoquímicas y nutricionales. La lista de posibles candidatas es, en el presente documento, esencialmente abierta y puede incluir cualquier proteína principal de almacenamiento de semillas, colágeno de origen animal o recombinante, extensinas (glicoproteínas ricas en hidroxiprolina abundantes en los monómeros de paredes celulares, por ejemplo, *Arabidopsis thaliana*, cuyos monómeros son moléculas flexibles similares a varillas “de tipo colágeno”).

En algunos casos, algunos o todos los componentes del producto consumible se suspenden en un gel. En diversos casos, el gel puede ser un hidrogel, un organogel o un xerogel. El gel puede espesarse usando un agente basado en polisacáridos o proteínas. Por ejemplo, pueden usarse fécula, arrurruz, almidón de maíz, almidón de katakuri, almidón de patata, palmera de sagú, tapioca, alginina, goma guar, goma de algarrobo, goma de xantano, colágeno, claras de huevo, furcelarán, gelatina, agar, carragenano, celulosa, metilcelulosa, hidroximetilcelulosa, goma acacia, konjac, almidón, pectina, amilopectina o proteínas derivadas de legumbres, cereales, nueces, otras semillas, hojas, algas, bacterias u hongos, solas o en combinación, para espesar el gel, formando una arquitectura o estructura para el producto consumible. Las enzimas que catalizan las reacciones que conducen a reticulaciones covalentes entre proteínas pueden usarse solas o en combinación para formar una arquitectura o estructura para el producto consumible. Por ejemplo, pueden usarse transglutaminasa, lisil oxidasas u otras amino oxidasas (por ejemplo, lisil oxidasa de *Pichia pastoris* (PPLO)), solas o en combinación, para formar una arquitectura o estructura para el producto consumible. En algunos casos, se combinan múltiples geles con diferentes componentes para formar el producto consumible. Por ejemplo, un gel que contiene una proteína con base vegetal puede asociarse con un gel que contiene una grasa de base vegetal. En algunos casos, las fibras o los aguijones de proteínas se orientan

paralelos entre sí y luego se mantienen en su lugar por la aplicación de un gel que contiene grasas de base vegetal.

Las composiciones de la divulgación pueden inflarse o expandirse mediante calentamiento, tal como friendo, horneando, calentando en microondas, calentando en un sistema de aire forzado, calentando en un túnel de aire, y similares.

En algunos casos, se combinan múltiples geles con diferentes componentes para formar el producto consumible. Por ejemplo, un gel que contiene una proteína con base vegetal puede asociarse con un gel que contiene una grasa de base vegetal. En algunas versiones, las fibras o los aguijones de proteínas se orientan paralelos entre sí y luego se mantienen en su lugar por la aplicación de un gel que contiene grasas de base vegetal.

En algunos casos la réplica de la carne no contiene productos animales, menos del 1% de gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene aislado de proteína de soja, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene concentrado de proteína de soja, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene proteína de soja, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene tofu, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene tofu ni gluten de trigo. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene proteína de soja ni gluten de trigo. En algunas variantes, la réplica de la carne no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo, no contiene harina de konjac, no contiene goma arábiga ni goma acacia. En algunas variantes, la réplica de la carne no contiene productos animales ni menos del 5% de hidratos de carbono.

En algunas variantes, la réplica de la carne no contiene productos animales, no contiene proteína de soja, no contiene gluten de trigo, no contiene metilcelulosa, no contiene carragenano, no contiene color caramelo ni harina de konjac, no contiene goma arábiga ni goma acacia ni menos del 5% de hidratos de carbono. En algunas variantes, la réplica de la carne no contiene productos animales, ni menos del 1% de celulosa. En algunas variantes, la réplica de la carne no contiene productos animales, ni menos del 5% de hidratos de carbono insolubles. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene proteína de soja, ni menos del 1% de celulosa. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene proteína de soja, ni menos del 5% de hidratos de carbono insolubles. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene gluten de trigo, ni menos del 1% de celulosa. En algunas variantes, la réplica de la carne no contiene productos animales, no contiene gluten de trigo, ni menos del 5% de hidratos de carbono insolubles.

El porcentaje de los diferentes componentes puede controlarse. Por ejemplo, pueden combinarse sustitutos con base no animal para músculo, tejido adiposo, tejido conjuntivo y componentes sanguíneos en diferentes razones y organizaciones físicas para aproximarse mejor al aspecto y tacto de la carne. Los diversos componentes pueden disponerse para garantizar la consistencia entre bocados del producto consumible. Los componentes pueden disponerse para garantizar que no se generan residuos del producto consumible. Por ejemplo, mientras que un corte tradicional de carne puede tener partes que no se comen normalmente, una réplica de la carne puede mejorarse por encima de la carne no incluyendo estas partes no comestibles. Una mejora de este tipo permite que todo el producto elaborado o enviado se consuma, lo que reduce los residuos y los costes de envío. Alternativamente, una réplica de la carne puede incluir partes no comestibles para imitar la experiencia de consumo de carne. Tales partes pueden incluir hueso, cartílago, tejido conjuntivo u otros materiales denominados habitualmente ternilla, o materiales incluidos que simulan estos componentes. En algunos casos, el producto consumible puede contener partes no comestibles simuladas de productos cárnicos que se diseñan para cumplir funciones secundarias. Por ejemplo, puede diseñarse un hueso simulado para dispersar calor durante el cocinado, haciendo que el cocinado del producto consumible sea más rápido o más uniforme que el de la carne. En otros casos, un hueso simulado también puede servir para mantener el producto consumible a una temperatura constante durante el envío. En otros casos, las partes no comestibles simuladas pueden ser biodegradables.

En algunos casos, las composiciones de sucedáneo de carne no contienen proteína animal, que comprende entre el 10-30% de proteína, entre el 5-80% de agua, entre el 5-70% de grasa, que comprenden una o más proteínas aisladas purificadas. En variantes particulares, las composiciones de sucedáneo de carne comprenden transglutaminasa. En algunos casos, el producto consumible contiene componentes para replicar los componentes de la carne. El componente principal de la carne es normalmente músculo esquelético. El músculo esquelético consiste normalmente en aproximadamente el 75% de agua, el 19% de proteína, el 2,5% de grasa intramuscular, el

1,2% de hidratos de carbono y el 2,3% de otras sustancias no proteicas solubles. Estas incluyen ácidos orgánicos, compuestos de azufre, compuestos nitrogenados, tales como aminoácidos y nucleótidos, y sustancias inorgánicas tales como minerales.

Por consiguiente, algunas variantes de la presente divulgación describen aproximaciones de replicación de esta composición para el producto consumible. Por ejemplo, en algunos casos, el producto consumible es una réplica de la carne a base de vegetales que puede comprender aproximadamente el 75% de agua, >19% de proteína, el 2,5% de grasa, el 1,2% de hidratos de carbono; y el 2,3% de otras sustancias no proteicas solubles. En algunos casos, el producto consumible es una réplica de la carne a base de vegetales que comprende entre el 60-90% de agua, el 10-30% de proteína, el 1-20% de grasa, el 0,1-5% de hidratos de carbono; y el 1-10% de otras sustancias no proteicas solubles. En algunos casos, el producto consumible es una réplica de la carne a base de vegetales que comprende entre el 60-90% de agua, el 5-10% de proteína, el 1-20% de grasa, el 0,1-5% de hidratos de carbono; y el 1-10% de otras sustancias no proteicas solubles. En algunos casos, el producto consumible es una réplica de la carne a base de vegetales que comprende entre >0-50% de agua, el 5-30% de proteína, >20-80% de grasa, el 0,1-5% de hidratos de carbono; y el 1-10% de otras sustancias no proteicas solubles. En algunos casos, la réplica contiene entre el 0,01% y el 5% en peso de una hemoproteína. En algunos casos, la réplica contiene entre el 0,01% y el 5% en peso de leghemoglobina. Alguna carne también contiene mioglobina, una hemoproteína, que representa la mayoría del color rojo y el contenido de hierro de alguna carne. En algunos casos, la réplica contiene entre el 0,01% y el 5% en peso de una hemoproteína. En algunos casos, la réplica contiene entre el 0,01% y el 5% en peso de leghemoglobina. Se entiende que estos porcentajes pueden variar en la carne, y pueden producirse las réplicas de la carne para aproximarse a la variación natural en la carne. Adicionalmente, en algunos casos, la presente divulgación describe réplicas de la carne mejoradas, que comprenden estos componentes en porcentajes normalmente no naturales. Por ejemplo, una réplica de la carne puede producirse con un contenido mayor que el contenido de grasa promedio típico. Los porcentajes de estos componentes también pueden alterarse para aumentar otras propiedades deseables.

En algunos casos, una réplica de la carne se diseña de manera que, cuando se concina, los porcentajes de componentes son similares a los de la carne cocinada. Así, en algunos casos, el producto consumible no cocinado tiene diferentes porcentajes de componentes que la carne no cocinada, pero cuando se cocina, el producto consumible es similar a la carne cocinada. Por ejemplo, una réplica de la carne puede elaborarse con un contenido mayor que el contenido de agua típico para carne cruda, pero cuando se cocina en un microondas, el producto resultante tiene porcentajes de componentes similares a la carne cocinada sobre un fuego.

En algunas variantes, el producto consumible es una réplica de la carne con un contenido menor que el contenido de agua típico para la carne. En algunos casos, las divulgaciones describen métodos para hidratar una réplica de la carne para producir la réplica de la carne que tiene un contenido similar a la carne. Por ejemplo, una réplica de la carne con un contenido de agua que podría ser bajo para la carne, por ejemplo, el 1%, el 10%, el 20%, el 30%, el 40% o el 50% de agua, se hidrata hasta aproximadamente el 75% de agua. Luego, una vez hidratada, en algunos casos, la réplica de la carne se cocina para consumo humano.

Aunque se han mostrado y descrito realizaciones preferidas de la presente invención en el presente documento, resultará evidente para los expertos en la técnica que tales realizaciones se proporcionan sólo a modo de ejemplo. Numerosas variaciones, cambios y sustituciones se les ocurrirán ahora a los expertos en la técnica sin apartarse de la presente divulgación. Debe entenderse que pueden emplearse diversas alternativas a las realizaciones de la invención descritas en el presente documento en la práctica de la invención. Se pretende que las siguientes reivindicaciones definan el alcance de la divulgación y que los métodos y las estructuras dentro del alcance de estas reivindicaciones y sus equivalentes estén cubiertos de ese modo.

Ejemplos

Ejemplo 1: método de producción y caracterización de un polipéptido que contiene hemo

Se produjo un plásmido que comprendía la secuencia de polinucleótido de *Aquifex aeolicus* que codifica para hemoglobina (*AaHb*), en el que la secuencia de nucleótidos estaba optimizada por codón para *E. coli*. Se subclonó el plásmido en un vector de expresión pBE-S proporcionado en el sistema de expresión de proteínas secretoras de *B.subtilis* (Takara Bio). Este vector contenía la secuencia promotora de *aprE* para fomentar la expresión constitutiva de AaHb en *B. subtilis* y una etiqueta de 6-histidina C-terminal. Se sintetizó un polinucleótido que codifica para el péptido señal de secreción a partir de la proteína PhoD de *B. subtilis* dependiente de la translocación de doble arginina (Tat) y se clonó en marco en el extremo 5' de *AaHb,* reemplazando el péptido señal de secreción de *aprE* dentro de la estructura principal *de pBE-S*. Para genera un constructo de expresión citosólico de *AaHb*, se delecionó el péptido señal de secreción de *aprE* del extremo 5' del marco de lectura abierto de *AaHb* usando PCR inversa seguido por ligación.

Los plásmidos de expresión se transformaron en la cepa de *B.subtilis RIK1285.* La expresión AaHb se monitorizó haciendo crecer cepas transformadas en medio LB, 10 µg/ml de kanamicina, $FeCl_3$ 0,1 mM y 20 µg/ml de ácido d-aminolevulínico. La expresión se llevó a cabo a 37ºC, con agitación a 200 RPM durante 24 horas. Después de la

expresión, se recogió el cultivo y se separó el polipéptido secretado de las bacterias.

Se monitorizó la expresión citosólica y secretada de AaHb mediante purificación por afinidad con Ni-NTA del sedimento celular y el sobrenadante de los medios seguido por SDS-PAGE y tinción con Coomassie de las fracciones de elución. Se monitorizó la carga de hemo mediante análisis de UV-Vis de fracciones que contienen AaHb purificadas.

Tal como se muestra en la figura 1, se comparó la expresión citosólica de AaHb en *B.subtilis* con (figura 1C) y sin (figura 1B) un péptido señal de secreción. El péptido de secreción de PhoD no interrumpió la expresión citosólica del polipéptido de AaHb. Se sometió a prueba AaHb citosólico para determinar el contenido de hemo usando espectroscopía de UV-Vis. Tal como se muestra en la figura 2, la adición del péptido señal de PhoD no interfiere con la unión del hemo de AaHb en el citosol.

Tal como se muestra en la figura 3, cuando se fusionó el polipéptido de AaHb al péptido de secreción de PhoD, la proteína de fusión se secretó de manera eficaz fuera de la célula huésped (figura 3B, carriles A, B). Adicionalmente, se demostró que el péptido de secreción de PhoD se escindió de manera apropiada dado que ambas formas del polipéptido (escindida y no escindida) se ubicaron en la fracción de sedimento celular (por ejemplo, dentro de la célula huésped) (figura 3B, carril “sedimento celular”). Sin embargo, sólo se secretó la versión escindida del polipéptido de fusión de AaHb, lo que indicó la función apropiada del péptido señal de PhoD.

En apoyo de la efectividad del péptido señal de PhoD, se realizó la secuenciación de proteína N-terminal en el polipéptido secretado. La figura 4 representa la secuencia de proteínas de la secuencia de la proteína de fusión PhoD-AaHb (que también comprende una etiqueta de His6). Se retiró una banda de proteínas del gel para el análisis de secuenciación de proteína N-terminal. Esta banda correspondía a la proteína secretada debido a su abundancia y el tamaño en el gel. Los resultados de la secuenciación N-terminal indicaron que el péptido de PhoD se escindió de manera eficaz (por ejemplo, el tamaño no era debido a degradación no específica) en el sitio correcto dado que el extremo N-terminal de los resultados de secuenciación coincidía con el sitio de corte predicho de la proteasa.

El AaHb secretado se purificó adicionalmente para determinar el contenido de hemo monitorizando la absorbancia de UV-Vis de AaHb purificado a partir de los medios. La figura 5 ilustra los efectos del péptido señal de PhoD sobre el contenido de hemo del AaHb secretado. El AaHb secretado estaba unido a hemo, tal como se evidenció por un pico de absorción pronunciado a aproximadamente 415 nm.

Tomados en conjunto, estos resultados sugieren que el péptido señal de PhoD no interfiere con la secreción citosólica de un polipéptido, se escinde de manera apropiada, mejora la secreción de un polipéptido y mantiene el contenido de hemo del polipéptido secretado.

Ejemplo profético 1: método de uso de un polipéptido en un producto cárnico consumible

En algunos casos, un polipéptido en esta divulgación se expresará y purificará tal como se describe en el ejemplo 1.

En algunos casos, un análogo de tejido muscular se construirá con el polipéptido y la proteína vicilina de guisante usando reticulación con transglutaminasa.

En algunos casos, un análogo de tejido muscular se construirá mediante formación con calor/frío de geles de proteína vicilina de guisantes purificada. El polipéptido que contiene hemo se mezclará por completo con el tejido muscular parcialmente gelificado tras el enfriamiento hasta temperatura ambiente.

En algunos casos, un análogo de tejido muscular se construirá extruyendo conjuntamente polipéptido que contiene hemo con proteína vicilina de guisantes purificada.

En algunos casos, un análogo de tejido adiposo se construirá emulsionado proteínas de albúmina de guisante, aceite de coco y lecitina a través de homogeneización de alta presión seguido por un tratamiento con calor/frío. El polipéptido que contiene hemo se mezclará por completo con el tejido adiposo parcialmente gelificado tras el enfriamiento hasta temperatura ambiente.

En algunos casos, un análogo de tejido conjuntivo se preparará con una fuente de proteína zeína mediante extrusión o electrohilado.

En algunos casos, una réplica de carne de vacuno picada (por ejemplo, producto cárnico consumible) se preparará combinando el análogo de músculo que comprende un polipéptido purificado de la divulgación con cantidades variables del análogo de tejido adiposo y el análogo de tejido conjuntivo. En algunos casos, los diferentes tejidos pueden combinarse usando una picadora de carne. El producto cárnico consumible resultante puede cocinarse antes del consumo. El procedimiento de cocinado inducirá que el color rojo del producto cárnico consumible cambie a color marrón, indicando el cocinado. En algunos casos, el color rojo del producto cárnico consumible es debido al polipéptido que contiene hemo purificado de la divulgación. El procedimiento de cocinado catalizará la liberación de

sabores y aromas de la carne. En algunos casos, los sabores y aromas del producto cárnico consumible son debidos al polipéptido que contiene hemo purificado de esta divulgación.

Ejemplo 2: método para modular la expresión y secreción de un polipéptido que contiene hemo endógeno

Se produjeron, transformaron y cultivaron dos plásmidos tal como se describe en el ejemplo 1. Ambos plásmidos incluyen la secuencia de polinucleótido de *B. subtilis* que codifica para el gen de hemoglobina truncada endógeno, *yjbI* (figura 9, SEQ ID NO: 26). En el caso del primer plásmido, se sintetizó un polinucleótido que codifica para el péptido señal de secreción de la proteína *PhoD* de *B. subtilis* dependiente de la translocación de doble arginina (Tat) (tabla 1) y se clonó en marco en el extremo 5' del gen *yjbI* (*PhoD-yjbI*). El segundo plásmido incluía el péptido señal *YwbN* (tabla 1), que también se sintetizó y se clonó en marco en el extremo 5' del gen *yjbI* (*YwbN-yjbI*), tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 6, después de 24 horas de crecimiento, se monitorizó la expresión citosólica y secretada de *yjbI*. Se separaron las fracciones de sedimento celular y de medios mediante SDS-PAGE, se transfirieron a una membrana de PVDF y se sondearon con un anticuerpo anti-His6 (Abcam). Específicamente, la figura muestra la detección de los polipéptidos fusionados (*PhoD-yjbI* y *YbwNyjbI*) en el sedimento celular tras la expresión del polipéptido endógeno del ácido nucleico exógeno y la detección de medios del polipéptido, que indicaron una escisión apropiada del péptido señal.

Ejemplo 3: método para modular la expresión y secreción de un polipéptido que contiene hemo heterólogo

Se produjeron, transformaron y cultivaron dos plásmidos tal como se describe en el ejemplo 1. El primer plásmido incluía la secuencia de polinucleótido de *Glycine max* (soja) que codifica para *LGB2* (figura 9, SEQ ID NO: 4). El segundo plásmido incluía la secuencia de polinucleótido que codifica para proteínas de *M. infernorum, HGbI* (figura 9, SEQ ID NO: 2). En ambos plásmidos, se sintetizó un polinucleótido que codifica para el péptido señal de secreción de la proteína *PhoD* de *B. subtilis* dependiente de la translocación de doble arginina (Tat) (tabla 1) y se clonó en marco en el extremo 5' de cada secuencia de polinucleótido (*PhoD-LGB2* y *PhoD-HGbI*).

Tal como se muestra en la figura 7, después de 24 horas de crecimiento, se monitorizó la expresión citosólica y secretada de los dos polipéptidos. Se separaron las fracciones de sedimento celular y de medios mediante SDS-PAGE, se transfirieron a una membrana de PVDF y se sondearon con un anticuerpo anti-His6 (Abcam). Específicamente, la figura muestra la detección de los polipéptidos fusionados (*PhoDLGB2* y *PhoD-HGbI)* en el sedimento celular tras la expresión de cada polipéptido heterólogo y la detección de medios de cada polipéptido heterólogo, que indicaron una escisión apropiada del péptido señal.

Ejemplo 4: método de producción y caracterización de un polipéptido que contiene hemo fusionado con diferentes péptidos señal

Se produjeron, transformaron y cultivaron una serie de plásmidos tal como se describe en el ejemplo 1. Cada plásmido incluía una secuencia de polinucleótido de *Aquifex aeolicus* que codifica para hemoglobina (*AaHb*). Al extremo 5' de la secuencia de *AaHb*, se fusionó una secuencia de ácido nucleico que codifica para un péptido de señalización de secreción seleccionado de *AbnA*, *AlbB*, *AppB*, *BglS*, *LipA*, *OppA*, *SpoIIIJ*, *TipA*, *WapA*, *WprA*, *YkpC*, *YmaC*, *YolA*, *YuiC* o *YwbN* (tabla 1) tal como se describió previamente.

Tal como se muestra en la figura 8, después de 24 horas de crecimiento, se separaron los sobrenadantes de los medios de un subconjunto de los polipéptidos de fusión mediante SDS-PAGE, se transfirieron a una membrana de PVDF y se sometieron a transferencia de tipo Western con un anticuerpo anti-His6 (Abcam). La figura ilustra la detección de medios del polipéptido endógeno (*AaHb*) después de la fusión con un subconjunto de varios péptidos de señalización de secreción a modo de ejemplo diferentes (*PhoD*, *TipA*, *WapA*, *WprA*, *YmaC*, *YolA*, *YuiC*, *YwbN*, *AppB*, y *BglS*), la expresión y la escisión del péptido señal de secreción.

Lista de secuencias

<110> Impossible Foods Inc.

<120> SECRECIÓN DE POLIPÉPTIDOS QUE CONTIENEN HEMO

<130> 38767-0009WO1

<150> documento US 61/908.689
<151> 25-11-2013

<150> documento US 61/876.676
<151> 11-09-2013

<160> 96

<170> FastSEQ para Windows versión 4.0

<210> 1
<211> 161
<212> PRT
<213> *Vigna radiata*

<400> 1

```
Met Thr Thr Thr Leu Glu Arg Gly Phe Thr Glu Glu Gln Glu Ala Leu
 1               5                  10                  15
Val Val Lys Ser Trp Asn Val Met Lys Lys Asn Ser Gly Glu Leu Gly
            20                  25                  30
Leu Lys Phe Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys
        35                  40                  45
Leu Phe Ser Phe Leu Arg Asp Ser Thr Val Pro Leu Glu Gln Asn Pro
    50                  55                  60
Lys Leu Lys Pro His Ala Val Ser Val Phe Val Met Thr Cys Asp Ser
65                  70                  75                  80
Ala Val Gln Leu Arg Lys Ala Gly Lys Val Thr Val Arg Glu Ser Asn
                85                  90                  95
Leu Lys Lys Leu Gly Ala Thr His Phe Arg Thr Gly Val Ala Asn Glu
            100                 105                 110
His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala
        115                 120                 125
Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala
    130                 135                 140
Tyr Asp Gln Leu Val Asp Ala Ile Lys Tyr Glu Met Lys Pro Pro Ser
145                 150                 155                 160
Ser
```

<210> 2
<211> 133
<212> PRT
<213> *Methylacidiphilum infernorum*

<400> 2

```
Met Ile Asp Gln Lys Glu Lys Glu Leu Ile Lys Glu Ser Trp Lys Arg
 1               5                  10                  15
Ile Glu Pro Asn Lys Asn Glu Ile Gly Leu Leu Phe Tyr Ala Asn Leu
            20                  25                  30
Phe Lys Glu Glu Pro Thr Val Ser Val Leu Phe Gln Asn Pro Ile Ser
        35                  40                  45
Ser Gln Ser Arg Lys Leu Met Gln Val Leu Gly Ile Leu Val Gln Gly
    50                  55                  60
Ile Asp Asn Leu Glu Gly Leu Ile Pro Thr Leu Gln Asp Leu Gly Arg
65                  70                  75                  80
Arg His Lys Gln Tyr Gly Val Val Asp Ser His Tyr Pro Leu Val Gly
                85                  90                  95
Asp Cys Leu Leu Lys Ser Ile Gln Glu Tyr Leu Gly Gln Gly Phe Thr
            100                 105                 110
Glu Glu Ala Lys Ala Ala Trp Thr Lys Val Tyr Gly Ile Ala Ala Gln
        115                 120                 125
Val Met Thr Ala Glu
    130
```

<210> 3
<211> 139
<212> PRT
<213> *Aquifex aeolicus*

<400> 3

```
Met Leu Ser Glu Glu Thr Ile Arg Val Ile Lys Ser Thr Val Pro Leu
 1               5                  10                  15
Leu Lys Glu His Gly Thr Glu Ile Thr Ala Arg Met Tyr Glu Leu Leu
            20                  25                  30
Phe Ser Lys Tyr Pro Lys Thr Lys Glu Leu Phe Ala Gly Ala Ser Glu
        35                  40                  45
Glu Gln Pro Lys Lys Leu Ala Asn Ala Ile Ile Ala Tyr Ala Thr Tyr
    50                  55                  60
Ile Asp Arg Leu Glu Glu Leu Asp Asn Ala Ile Ser Thr Ile Ala Arg
65                  70                  75                  80
Ser His Val Arg Arg Asn Val Lys Pro Glu His Tyr Pro Leu Val Lys
                85                  90                  95
Glu Cys Leu Leu Gln Ala Ile Glu Glu Val Leu Asn Pro Gly Glu Glu
            100                 105                 110
Val Leu Lys Ala Trp Glu Glu Ala Tyr Asp Phe Leu Ala Lys Thr Leu
        115                 120                 125
Ile Thr Leu Glu Lys Lys Leu Tyr Ser Gln Pro
    130                 135
```

<210> 4
<211> 145
<212> PRT
<213> *Glycine max*

<400> 4

```
Met Gly Ala Phe Thr Glu Lys Gln Glu Ala Leu Val Ser Ser Ser Phe
 1               5                  10                  15
Glu Ala Phe Lys Ala Asn Ile Pro Gln Tyr Ser Val Val Phe Tyr Thr
            20                  25                  30
Ser Ile Leu Glu Lys Ala Pro Ala Ala Lys Asp Leu Phe Ser Phe Leu
        35                  40                  45
Ser Asn Gly Val Asp Pro Ser Asn Pro Lys Leu Thr Gly His Ala Glu
    50                  55                  60
Lys Leu Phe Gly Leu Val Arg Asp Ser Ala Gly Gln Leu Lys Ala Asn
65                  70                  75                  80
Gly Thr Val Val Ala Asp Ala Ala Leu Gly Ser Ile His Ala Gln Lys
                85                  90                  95
Ala Ile Thr Asp Pro Gln Phe Val Val Val Lys Glu Ala Leu Leu Lys
            100                 105                 110
Thr Ile Lys Glu Ala Val Gly Asp Lys Trp Ser Asp Glu Leu Ser Ser
        115                 120                 125
Ala Trp Glu Val Ala Tyr Asp Glu Leu Ala Ala Ala Ile Lys Lys Ala
    130                 135                 140
Phe
145
```

<210> 5
<211> 162
<212> PRT
<213> *Hordeum vulgare*

<400> 5

```
Met Ser Ala Ala Glu Gly Ala Val Val Phe Ser Glu Glu Lys Glu Ala
 1               5                  10                  15
Leu Val Leu Lys Ser Trp Ala Ile Met Lys Lys Asp Ser Ala Asn Leu
            20                  25                  30
Gly Leu Arg Phe Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Arg
        35                  40                  45
Gln Met Phe Pro Phe Leu Arg Asp Ser Asp Val Pro Leu Glu Thr Asn
    50                  55                  60
Pro Lys Leu Lys Thr His Ala Val Ser Val Phe Val Met Thr Cys Glu
65                  70                  75                  80
Ala Ala Ala Gln Leu Arg Lys Ala Gly Lys Ile Thr Val Arg Glu Thr
                85                  90                  95
Thr Leu Lys Arg Leu Gly Gly Thr His Leu Lys Tyr Gly Val Ala Asp
            100                 105                 110
Gly His Phe Glu Val Thr Arg Phe Ala Leu Leu Glu Thr Ile Lys Glu
        115                 120                 125
Ala Leu Pro Ala Asp Met Trp Gly Pro Glu Met Arg Asn Ala Trp Gly
    130                 135                 140
Glu Ala Tyr Asp Gln Leu Val Ala Ala Ile Lys Gln Glu Met Lys Pro
145                 150                 155                 160
Ala Glu
```

<210> 6
<211> 1153
<212> PRT
<213> *Magnaporthe oryzae*

<400> 6

```
Met Asp Gly Ala Val Arg Leu Asp Trp Thr Gly Leu Asp Leu Thr Gly
 1               5                  10                  15
His Glu Ile His Asp Gly Val Pro Ile Ala Ser Arg Val Gln Val Met
            20                  25                  30
Val Ser Phe Pro Leu Phe Lys Asp Gln His Ile Ile Met Ser Ser Lys
        35                  40                  45
Glu Ser Pro Ser Arg Lys Ser Ser Thr Ile Gly Gln Ser Thr Arg Asn
    50                  55                  60
Gly Ser Cys Gln Ala Asp Thr Gln Lys Gly Gln Leu Pro Pro Val Gly
65                  70                  75                  80
Glu Lys Pro Lys Pro Val Lys Glu Asn Pro Met Lys Lys Leu Lys Glu
                85                  90                  95
Met Ser Gln Arg Pro Leu Pro Thr Gln His Gly Asp Gly Thr Tyr Pro
            100                 105                 110
Thr Glu Lys Lys Leu Thr Gly Ile Gly Glu Asp Leu Lys His Ile Arg
        115                 120                 125
Gly Tyr Asp Val Lys Thr Leu Leu Ala Met Val Lys Ser Lys Leu Lys
    130                 135                 140
Gly Glu Lys Leu Lys Asp Asp Lys Thr Met Leu Met Glu Arg Val Met
145                 150                 155                 160
Gln Leu Val Ala Arg Leu Pro Thr Glu Ser Lys Lys Arg Ala Glu Leu
                165                 170                 175
Thr Asp Ser Leu Ile Asn Glu Leu Trp Glu Ser Leu Asp His Pro Pro
            180                 185                 190
```

```
Leu Asn Tyr Leu Gly Pro Glu His Ser Tyr Arg Thr Pro Asp Gly Ser
        195                 200                 205
Tyr Asn His Pro Phe Asn Pro Gln Leu Gly Ala Ala Gly Ser Arg Tyr
    210                 215                 220
Ala Arg Ser Val Ile Pro Thr Val Thr Pro Pro Gly Ala Leu Pro Asp
225                 230                 235                 240
Pro Gly Leu Ile Phe Asp Ser Ile Met Gly Arg Thr Pro Asn Ser Tyr
                245                 250                 255
Arg Lys His Pro Asn Asn Val Ser Ser Ile Leu Trp Tyr Trp Ala Thr
            260                 265                 270
Ile Ile Ile His Asp Ile Phe Trp Thr Asp Pro Arg Asp Ile Asn Thr
        275                 280                 285
Asn Lys Ser Ser Ser Tyr Leu Asp Leu Ala Pro Leu Tyr Gly Asn Ser
    290                 295                 300
Gln Glu Met Gln Asp Ser Ile Arg Thr Phe Lys Asp Gly Arg Met Lys
305                 310                 315                 320
Pro Asp Cys Tyr Ala Asp Lys Arg Leu Ala Gly Met Pro Pro Gly Val
                325                 330                 335
Ser Val Leu Leu Ile Met Phe Asn Arg Phe His Asn His Val Ala Glu
            340                 345                 350
Asn Leu Ala Leu Ile Asn Glu Gly Gly Arg Phe Asn Lys Pro Ser Asp
        355                 360                 365
Leu Leu Glu Gly Glu Ala Arg Glu Ala Ala Trp Lys Lys Tyr Asp Asn
    370                 375                 380
Asp Leu Phe Gln Val Ala Arg Leu Val Thr Ser Gly Leu Tyr Ile Asn
385                 390                 395                 400
Ile Thr Leu Val Asp Tyr Val Arg Asn Ile Val Asn Leu Asn Arg Val
                405                 410                 415
Asp Thr Thr Trp Thr Leu Asp Pro Arg Gln Asp Ala Gly Ala His Val
            420                 425                 430
Gly Thr Ala Asp Gly Ala Glu Arg Gly Thr Gly Asn Ala Val Ser Ala
        435                 440                 445
Glu Phe Asn Leu Cys Tyr Arg Trp His Ser Cys Ile Ser Glu Lys Asp
    450                 455                 460
Ser Lys Phe Val Glu Ala Gln Phe Gln Asn Ile Phe Gly Lys Pro Ala
465                 470                 475                 480
Ser Glu Val Arg Pro Asp Glu Met Trp Lys Gly Phe Ala Lys Met Glu
                485                 490                 495
Gln Asn Thr Pro Ala Asp Pro Gly Gln Arg Thr Phe Gly Gly Phe Lys
            500                 505                 510
Arg Gly Pro Asp Gly Lys Phe Asp Asp Asp Asp Leu Val Arg Cys Ile
        515                 520                 525
Ser Glu Ala Val Glu Asp Val Ala Gly Ala Phe Gly Ala Arg Asn Val
    530                 535                 540
Pro Gln Ala Met Lys Val Val Glu Thr Met Gly Ile Ile Gln Gly Arg
545                 550                 555                 560
Lys Trp Asn Val Ala Gly Leu Asn Glu Phe Arg Lys His Phe His Leu
                565                 570                 575
Lys Pro Tyr Ser Thr Phe Glu Asp Ile Asn Ser Asp Pro Gly Val Ala
            580                 585                 590
Glu Ala Leu Arg Arg Leu Tyr Asp His Pro Asp Asn Val Glu Leu Tyr
        595                 600                 605
Pro Gly Leu Val Ala Glu Glu Asp Lys Gln Pro Met Val Pro Gly Val
    610                 615                 620
Gly Ile Ala Pro Thr Tyr Thr Ile Ser Arg Val Val Leu Ser Asp Ala
625                 630                 635                 640
Val Cys Leu Val Arg Gly Asp Arg Phe Tyr Thr Thr Asp Phe Thr Pro
                645                 650                 655
Arg Asn Leu Thr Asn Trp Gly Tyr Lys Glu Val Asp Tyr Asp Leu Ser
            660                 665                 670
Val Asn His Gly Cys Val Phe Tyr Lys Leu Phe Ile Arg Ala Phe Pro
        675                 680                 685
Asn His Phe Lys Gln Asn Ser Val Tyr Ala His Tyr Pro Met Val Val
```

```
    690                 695                 700
Pro Ser Glu Asn Lys Arg Ile Leu Glu Ala Leu Gly Arg Ala Asp Leu
705                 710                 715                 720
Phe Asp Phe Glu Ala Pro Lys Tyr Ile Pro Pro Arg Val Asn Ile Thr
                725                 730                 735
Ser Tyr Gly Gly Ala Glu Tyr Ile Leu Glu Thr Gln Glu Lys Tyr Lys
            740                 745                 750
Val Thr Trp His Glu Gly Leu Gly Phe Leu Met Gly Glu Gly Gly Leu
        755                 760                 765
Lys Phe Met Leu Ser Gly Asp Asp Pro Leu His Ala Gln Gln Arg Lys
    770                 775                 780
Cys Met Ala Ala Gln Leu Tyr Lys Asp Gly Trp Thr Glu Ala Val Lys
785                 790                 795                 800
Ala Phe Tyr Ala Gly Met Met Glu Glu Leu Leu Val Ser Lys Ser Tyr
                805                 810                 815
Phe Leu Gly Asn Asn Lys His Arg His Val Asp Ile Ile Arg Asp Val
            820                 825                 830
Gly Asn Met Val His Val His Phe Ala Ser Gln Val Phe Gly Leu Pro
        835                 840                 845
Leu Lys Thr Ala Lys Asn Pro Thr Gly Val Phe Thr Glu Gln Glu Met
    850                 855                 860
Tyr Gly Ile Leu Ala Ala Ile Phe Thr Thr Ile Phe Phe Asp Leu Asp
865                 870                 875                 880
Pro Ser Lys Ser Phe Pro Leu Arg Thr Lys Thr Arg Glu Val Cys Gln
                885                 890                 895
Lys Leu Ala Lys Leu Val Glu Ala Asn Val Lys Leu Ile Asn Lys Ile
            900                 905                 910
Pro Trp Ser Arg Gly Met Phe Val Gly Lys Pro Ala Lys Asp Glu Pro
        915                 920                 925
Leu Ser Ile Tyr Gly Lys Thr Met Ile Lys Gly Leu Lys Ala His Gly
    930                 935                 940
Leu Ser Asp Tyr Asp Ile Ala Trp Ser His Val Val Pro Thr Ser Gly
945                 950                 955                 960
Ala Met Val Pro Asn Gln Ala Gln Val Phe Ala Gln Ala Val Asp Tyr
                965                 970                 975
Tyr Leu Ser Pro Ala Gly Met His Tyr Ile Pro Glu Ile His Met Val
            980                 985                 990
Ala Leu Gln Pro Ser Thr Pro Glu Thr Asp Ala Leu Leu Leu Gly Tyr
        995                 1000                1005
Ala Met Glu Gly Ile Arg Leu Ala Gly Thr Phe Gly Ser Tyr Arg Glu
    1010                1015                1020
Ala Ala Val Asp Asp Val Val Lys Glu Asp Asn Gly Arg Gln Val Pro
1025                1030                1035                1040
Val Lys Ala Gly Asp Arg Val Phe Val Ser Phe Val Asp Ala Ala Arg
                1045                1050                1055
Asp Pro Lys His Phe Pro Asp Pro Glu Val Val Asn Pro Arg Arg Pro
            1060                1065                1070
Ala Lys Lys Tyr Ile His Tyr Gly Val Gly Pro His Ala Cys Leu Gly
        1075                1080                1085
Arg Asp Ala Ser Gln Ile Ala Ile Thr Glu Met Phe Arg Cys Leu Phe
    1090                1095                1100
Arg Arg Arg Asn Val Arg Arg Val Pro Gly Pro Gln Gly Glu Leu Lys
1105                1110                1115                1120
Lys Val Pro Arg Pro Gly Gly Phe Tyr Val Tyr Met Arg Glu Asp Trp
                1125                1130                1135
Gly Gly Leu Phe Pro Phe Pro Val Thr Met Arg Val Met Trp Asp Asp
            1140                1145                1150
Glu
```

<210> 7
<211> 530
<212> PRT

<213> *Fusarium oxysporum*

<400> 7

```
Met Lys Gly Ser Ala Thr Leu Ala Phe Ala Leu Val Gln Phe Ser Ala
 1               5                  10                  15
Ala Ser Gln Leu Val Trp Pro Ser Lys Trp Asp Glu Val Glu Asp Leu
            20                  25                  30
Leu Tyr Met Gln Gly Gly Phe Asn Lys Arg Gly Phe Ala Asp Ala Leu
        35                  40                  45
Arg Thr Cys Glu Phe Gly Ser Asn Val Pro Gly Thr Gln Asn Thr Ala
    50                  55                  60
Glu Trp Leu Arg Thr Ala Phe His Asp Ala Ile Thr His Asp Ala Lys
65                  70                  75                  80
Ala Gly Thr Gly Gly Leu Asp Ala Ser Ile Tyr Trp Glu Ser Ser Arg
                85                  90                  95
Pro Glu Asn Pro Gly Lys Ala Phe Asn Asn Thr Phe Gly Phe Phe Ser
            100                 105                 110
Gly Phe His Asn Pro Arg Ala Thr Ala Ser Asp Leu Thr Ala Leu Gly
        115                 120                 125
Thr Val Leu Ala Val Gly Ala Cys Asn Gly Pro Arg Ile Pro Phe Arg
    130                 135                 140
Ala Gly Arg Ile Asp Ala Tyr Lys Ala Gly Pro Ala Gly Val Pro Glu
145                 150                 155                 160
Pro Ser Thr Asn Leu Lys Asp Thr Phe Ala Ala Phe Thr Lys Ala Gly
                165                 170                 175
Phe Thr Lys Glu Glu Met Thr Ala Met Val Ala Cys Gly His Ala Ile
            180                 185                 190
Gly Gly Val His Ser Val Asp Phe Pro Glu Ile Val Gly Ile Lys Ala
        195                 200                 205
Asp Pro Asn Asn Asp Thr Asn Val Pro Phe Gln Lys Asp Val Ser Ser
    210                 215                 220
Phe His Asn Gly Ile Val Thr Glu Tyr Leu Ala Gly Thr Ser Lys Asn
225                 230                 235                 240
Pro Leu Val Ala Ser Lys Asn Ala Thr Phe His Ser Asp Lys Arg Ile
                245                 250                 255
Phe Asp Asn Asp Lys Ala Thr Met Lys Lys Leu Ser Thr Lys Ala Gly
            260                 265                 270
Phe Asn Ser Met Cys Ala Asp Ile Leu Thr Arg Met Ile Asp Thr Val
        275                 280                 285
Pro Lys Ser Val Gln Leu Thr Pro Val Leu Glu Ala Tyr Asp Val Arg
    290                 295                 300
Pro Tyr Ile Thr Glu Leu Ser Leu Asn Asn Lys Asn Lys Ile His Phe
305                 310                 315                 320
Thr Gly Ser Val Arg Val Arg Ile Thr Asn Asn Ile Arg Asp Asn Asn
                325                 330                 335
Asp Leu Ala Ile Asn Leu Ile Tyr Val Gly Arg Asp Gly Lys Lys Val
            340                 345                 350
Thr Val Pro Thr Gln Gln Val Thr Phe Gln Gly Gly Thr Ser Phe Gly
        355                 360                 365
Ala Gly Glu Val Phe Ala Asn Phe Glu Phe Asp Thr Thr Met Asp Ala
    370                 375                 380
Lys Asn Gly Ile Thr Lys Phe Phe Ile Gln Glu Val Lys Pro Ser Thr
385                 390                 395                 400
Lys Ala Thr Val Thr His Asp Asn Gln Lys Thr Gly Gly Tyr Lys Val
                405                 410                 415
Asp Asp Thr Val Leu Tyr Gln Leu Gln Gln Ser Cys Ala Val Leu Glu
            420                 425                 430
Lys Leu Pro Asn Ala Pro Leu Val Val Thr Ala Met Val Arg Asp Ala
        435                 440                 445
Arg Ala Lys Asp Ala Leu Thr Leu Arg Val Ala His Lys Lys Pro Val
    450                 455                 460
Lys Gly Ser Ile Val Pro Arg Phe Gln Thr Ala Ile Thr Asn Phe Lys
```

```
465                 470                 475                 480
Ala Thr Gly Lys Lys Ser Ser Gly Tyr Thr Gly Phe Gln Ala Lys Thr
                485                 490                 495
Met Phe Glu Glu Gln Ser Thr Tyr Phe Asp Ile Val Leu Gly Gly Ser
            500                 505                 510
Pro Ala Ser Gly Val Gln Phe Leu Thr Ser Gln Ala Met Pro Ser Gln
        515                 520                 525
Cys Ser
    530
```

<210> 8
<211> 358
<212> PRT
<213> *Fusarium graminearum*

<400> 8

```
Met Ala Ser Ala Thr Arg Gln Phe Ala Arg Ala Ala Thr Arg Ala Thr
 1               5                  10                  15
Arg Asn Gly Phe Ala Ile Ala Pro Arg Gln Val Ile Arg Gln Gln Gly
            20                  25                  30
Arg Arg Tyr Tyr Ser Ser Glu Pro Ala Gln Lys Ser Ser Ser Ala Trp
        35                  40                  45
Ile Trp Leu Thr Gly Ala Ala Val Ala Gly Gly Ala Gly Tyr Tyr Phe
    50                  55                  60
Tyr Gly Asn Ser Ala Ser Ser Ala Thr Ala Lys Val Phe Asn Pro Ser
65                  70                  75                  80
Lys Glu Asp Tyr Gln Lys Val Tyr Asn Glu Ile Ala Ala Arg Leu Glu
                85                  90                  95
Glu Lys Asp Asp Tyr Asp Asp Gly Ser Tyr Gly Pro Val Leu Val Arg
            100                 105                 110
Leu Ala Trp His Ala Ser Gly Thr Tyr Asp Lys Glu Thr Gly Thr Gly
        115                 120                 125
Gly Ser Asn Gly Ala Thr Met Arg Phe Ala Pro Glu Ser Asp His Gly
    130                 135                 140
Ala Asn Ala Gly Leu Ala Ala Ala Arg Asp Phe Leu Gln Pro Val Lys
145                 150                 155                 160
Glu Lys Phe Pro Trp Ile Thr Tyr Ser Asp Leu Trp Ile Leu Ala Gly
                165                 170                 175
Val Cys Ala Ile Gln Glu Met Leu Gly Pro Ala Ile Pro Tyr Arg Pro
            180                 185                 190
Gly Arg Ser Asp Arg Asp Val Ser Gly Cys Thr Pro Asp Gly Arg Leu
        195                 200                 205
Pro Asp Ala Ser Lys Arg Gln Asp His Leu Arg Gly Ile Phe Gly Arg
    210                 215                 220
Met Gly Phe Asn Asp Gln Glu Ile Val Ala Leu Ser Gly Ala His Ala
225                 230                 235                 240
Leu Gly Arg Cys His Thr Asp Arg Ser Gly Tyr Ser Gly Pro Trp Thr
                245                 250                 255
Phe Ser Pro Thr Val Leu Thr Asn Asp Tyr Phe Arg Leu Leu Val Glu
            260                 265                 270
Glu Lys Trp Gln Trp Lys Lys Trp Asn Gly Pro Ala Gln Tyr Glu Asp
        275                 280                 285
Lys Ser Thr Lys Ser Leu Met Met Leu Pro Ser Asp Ile Ala Leu Ile
    290                 295                 300
Glu Asp Lys Lys Phe Lys Pro Trp Val Glu Lys Tyr Ala Lys Asp Asn
305                 310                 315                 320
Asp Ala Phe Phe Lys Asp Phe Ser Asn Val Val Leu Arg Leu Phe Glu
                325                 330                 335
Leu Gly Val Pro Phe Ala Gln Gly Thr Glu Asn Gln Arg Trp Thr Phe
            340                 345                 350
Lys Pro Thr His Gln Glu
        355
```

<210> 9
<211> 122
<212> PRT
<213> *Chlamydomonas eugametos*

<400> 9

```
Met Ser Leu Phe Ala Lys Leu Gly Gly Arg Glu Ala Val Glu Ala Ala
 1               5                  10                  15
Val Asp Lys Phe Tyr Asn Lys Ile Val Ala Asp Pro Thr Val Ser Thr
            20                  25                  30
Tyr Phe Ser Asn Thr Asp Met Lys Val Gln Arg Ser Lys Gln Phe Ala
        35                  40                  45
Phe Leu Ala Tyr Ala Leu Gly Gly Ala Ser Glu Trp Lys Gly Lys Asp
    50                  55                  60
Met Arg Thr Ala His Lys Asp Leu Val Pro His Leu Ser Asp Val His
65                  70                  75                  80
Phe Gln Ala Val Ala Arg His Leu Ser Asp Thr Leu Thr Glu Leu Gly
                85                  90                  95
Val Pro Pro Glu Asp Ile Thr Asp Ala Met Ala Val Val Ala Ser Thr
            100                 105                 110
Arg Thr Glu Val Leu Asn Met Pro Gln Gln
        115                 120
```

<210> 10
<211> 121
<212> PRT
<213> *Tetrahymena pyriformis*

<400> 10

```
Met Asn Lys Pro Gln Thr Ile Tyr Glu Lys Leu Gly Gly Glu Asn Ala
 1               5                  10                  15
Met Lys Ala Ala Val Pro Leu Phe Tyr Lys Lys Val Leu Ala Asp Glu
            20                  25                  30
Arg Val Lys His Phe Phe Lys Asn Thr Asp Met Asp His Gln Thr Lys
        35                  40                  45
Gln Gln Thr Asp Phe Leu Thr Met Leu Leu Gly Gly Pro Asn His Tyr
    50                  55                  60
Lys Gly Lys Asn Met Thr Glu Ala His Lys Gly Met Asn Leu Gln Asn
65                  70                  75                  80
Leu His Phe Asp Ala Ile Ile Glu Asn Leu Ala Ala Thr Leu Lys Glu
                85                  90                  95
Leu Gly Val Thr Asp Ala Val Ile Asn Glu Ala Ala Lys Val Ile Glu
            100                 105                 110
His Thr Arg Lys Asp Met Leu Gly Lys
        115                 120
```

<210> 11
<211> 117
<212> PRT
<213> *Paramecium caudatum*

<400> 11

```
Met Ser Leu Phe Glu Gln Leu Gly Gly Gln Ala Ala Val Gln Ala Val
 1               5                  10                  15
Thr Ala Gln Phe Tyr Ala Asn Ile Gln Ala Asp Ala Thr Val Ala Thr
            20                  25                  30
Phe Phe Asn Gly Ile Asp Met Pro Asn Gln Thr Asn Lys Thr Ala Ala
        35                  40                  45
Phe Leu Cys Ala Ala Leu Gly Gly Pro Asn Ala Trp Thr Gly Arg Asn
```

```
     50                  55                  60
Leu Lys Glu Val His Ala Asn Met Gly Val Ser Asn Ala Gln Phe Thr
65                  70                  75                  80
Thr Val Ile Gly His Leu Arg Ser Ala Leu Thr Gly Ala Gly Val Ala
                85                  90                  95
Ala Ala Leu Val Glu Gln Thr Val Ala Val Ala Glu Thr Val Arg Gly
            100                 105                 110
Asp Val Val Thr Val
        115
```

<210> 12
<211> 147
<212> PRT
<213> *Aspergillus niger*

<400> 12

```
Met Pro Leu Thr Pro Glu Gln Ile Lys Ile Ile Lys Ala Thr Val Pro
 1               5                  10                  15
Val Leu Gln Glu Tyr Gly Thr Lys Ile Thr Thr Ala Phe Tyr Met Asn
            20                  25                  30
Met Ser Thr Val His Pro Glu Leu Asn Ala Val Phe Asn Thr Ala Asn
        35                  40                  45
Gln Val Lys Gly His Gln Ala Arg Ala Leu Ala Gly Ala Leu Phe Ala
    50                  55                  60
Tyr Ala Ser His Ile Asp Asp Leu Gly Ala Leu Gly Pro Ala Val Glu
65                  70                  75                  80
Leu Ile Cys Asn Lys His Ala Ser Leu Tyr Ile Gln Ala Asp Glu Tyr
                85                  90                  95
Lys Ile Val Gly Lys Tyr Leu Leu Glu Ala Met Lys Glu Val Leu Gly
            100                 105                 110
Asp Ala Cys Thr Asp Asp Ile Leu Asp Ala Trp Gly Ala Ala Tyr Trp
        115                 120                 125
Ala Leu Ala Asp Ile Met Ile Asn Arg Glu Ala Ala Leu Tyr Lys Gln
    130                 135                 140
Ser Gln Gly
145
```

<210> 13
<211> 165
<212> PRT
<213> *Zea mays*

<400> 13

```
Met Ala Leu Ala Glu Ala Asp Asp Gly Ala Val Val Phe Gly Glu Glu
 1               5                  10                  15
Gln Glu Ala Leu Val Leu Lys Ser Trp Ala Val Met Lys Lys Asp Ala
            20                  25                  30
Ala Asn Leu Gly Leu Arg Phe Phe Leu Lys Val Phe Glu Ile Ala Pro
        35                  40                  45
Ser Ala Glu Gln Met Phe Ser Phe Leu Arg Asp Ser Asp Val Pro Leu
    50                  55                  60
Glu Lys Asn Pro Lys Leu Lys Thr His Ala Met Ser Val Phe Val Met
65                  70                  75                  80
Thr Cys Glu Ala Ala Ala Gln Leu Arg Lys Ala Gly Lys Val Thr Val
                85                  90                  95
Arg Glu Thr Thr Leu Lys Arg Leu Gly Ala Thr His Leu Arg Tyr Gly
            100                 105                 110
Val Ala Asp Gly His Phe Glu Val Thr Gly Phe Ala Leu Leu Glu Thr
        115                 120                 125
Ile Lys Glu Ala Leu Pro Ala Asp Met Trp Ser Leu Glu Met Lys Lys
    130                 135                 140
```

```
Ala Trp Ala Glu Ala Tyr Ser Gln Leu Val Ala Ala Ile Lys Arg Glu
145                 150                 155                 160
Met Lys Pro Asp Ala
                165
```

<210> 14
<211> 169
<212> PRT
<213> *Oryza sativa subsp. japonica*

<400> 14

```
Met Ala Leu Val Glu Gly Asn Asn Gly Val Ser Gly Gly Ala Val Ser
 1               5                  10                  15
Phe Ser Glu Glu Gln Glu Ala Leu Val Leu Lys Ser Trp Ala Ile Met
            20                  25                  30
Lys Lys Asp Ser Ala Asn Ile Gly Leu Arg Phe Phe Leu Lys Ile Phe
        35                  40                  45
Glu Val Ala Pro Ser Ala Ser Gln Met Phe Ser Phe Leu Arg Asn Ser
    50                  55                  60
Asp Val Pro Leu Glu Lys Asn Pro Lys Leu Lys Thr His Ala Met Ser
65                  70                  75                  80
Val Phe Val Met Thr Cys Glu Ala Ala Ala Gln Leu Arg Lys Ala Gly
                85                  90                  95
Lys Val Thr Val Arg Asp Thr Thr Leu Lys Arg Leu Gly Ala Thr His
            100                 105                 110
Phe Lys Tyr Gly Val Gly Asp Ala His Phe Glu Val Thr Arg Phe Ala
        115                 120                 125
Leu Leu Glu Thr Ile Lys Glu Ala Val Pro Val Asp Met Trp Ser Pro
    130                 135                 140
Ala Met Lys Ser Ala Trp Ser Glu Ala Tyr Asn Gln Leu Val Ala Ala
145                 150                 155                 160
Ile Lys Gln Glu Met Lys Pro Ala Glu
                165
```

<210> 15
<211> 160
<212> PRT
<213> *Arabidopsis thaliana*

<400> 15

```
Met Glu Ser Glu Gly Lys Ile Val Phe Thr Glu Glu Gln Glu Ala Leu
 1               5                  10                  15
Val Val Lys Ser Trp Ser Val Met Lys Lys Asn Ser Ala Glu Leu Gly
            20                  25                  30
Leu Lys Leu Phe Ile Lys Ile Phe Glu Ile Ala Pro Thr Thr Lys Lys
        35                  40                  45
Met Phe Ser Phe Leu Arg Asp Ser Pro Ile Pro Ala Glu Gln Asn Pro
    50                  55                  60
Lys Leu Lys Pro His Ala Met Ser Val Phe Val Met Cys Cys Glu Ser
65                  70                  75                  80
Ala Val Gln Leu Arg Lys Thr Gly Lys Val Thr Val Arg Glu Thr Thr
                85                  90                  95
Leu Lys Arg Leu Gly Ala Ser His Ser Lys Tyr Gly Val Val Asp Glu
            100                 105                 110
His Phe Glu Val Ala Lys Tyr Ala Leu Leu Glu Thr Ile Lys Glu Ala
        115                 120                 125
Val Pro Glu Met Trp Ser Pro Glu Met Lys Val Ala Trp Gly Gln Ala
    130                 135                 140
Tyr Asp His Leu Val Ala Ala Ile Lys Ala Glu Met Asn Leu Ser Asn
145                 150                 155                 160
```

<210> 16
<211> 147

<212> PRT
<213> *Pisum sativum*

<400> 16

```
Met Gly Phe Thr Asp Lys Gln Glu Ala Leu Val Asn Ser Ser Trp Glu
 1               5                  10                  15
Ser Phe Lys Gln Asn Leu Ser Gly Asn Ser Ile Leu Phe Tyr Thr Ile
            20                  25                  30
Ile Leu Glu Lys Ala Pro Ala Ala Lys Gly Leu Phe Ser Phe Leu Lys
        35                  40                  45
Asp Thr Ala Gly Val Glu Asp Ser Pro Lys Leu Gln Ala His Ala Glu
    50                  55                  60
Gln Val Phe Gly Leu Val Arg Asp Ser Ala Ala Gln Leu Arg Thr Lys
65                  70                  75                  80
Gly Glu Val Val Leu Gly Asn Ala Thr Leu Gly Ala Ile His Val Gln
                85                  90                  95
Arg Gly Val Thr Asp Pro His Phe Val Val Val Lys Glu Ala Leu Leu
            100                 105                 110
Gln Thr Ile Lys Lys Ala Ser Gly Asn Asn Trp Ser Glu Glu Leu Asn
        115                 120                 125
Thr Ala Trp Glu Val Ala Tyr Asp Gly Leu Ala Thr Ala Ile Lys Lys
    130                 135                 140
Ala Met Thr
145
```

<210> 17
<211> 145
<212> PRT
<213> *Vigna unguiculata*

<400> 17

```
Met Val Ala Phe Ser Asp Lys Gln Glu Ala Leu Val Asn Gly Ala Tyr
 1               5                  10                  15
Glu Ala Phe Lys Ala Asn Ile Pro Lys Tyr Ser Val Val Phe Tyr Thr
            20                  25                  30
Thr Ile Leu Glu Lys Ala Pro Ala Ala Lys Asn Leu Phe Ser Phe Leu
        35                  40                  45
Ala Asn Gly Val Asp Ala Thr Asn Pro Lys Leu Thr Gly His Ala Glu
    50                  55                  60
Lys Leu Phe Gly Leu Val Arg Asp Ser Ala Ala Gln Leu Arg Ala Ser
65                  70                  75                  80
Gly Gly Val Val Ala Asp Ala Ala Leu Gly Ala Val His Ser Gln Lys
                85                  90                  95
Ala Val Asn Asp Ala Gln Phe Val Val Val Lys Glu Ala Leu Val Lys
            100                 105                 110
Thr Leu Lys Glu Ala Val Gly Asp Lys Trp Ser Asp Glu Leu Gly Thr
        115                 120                 125
Ala Val Glu Leu Ala Tyr Asp Glu Leu Ala Ala Ala Ile Lys Lys Ala
    130                 135                 140
Tyr
145
```

<210> 18
<211> 154
<212> PRT
<213> *Bos taurus*

<400> 18

```
Met Gly Leu Ser Asp Gly Glu Trp Gln Leu Val Leu Asn Ala Trp Gly
 1               5                  10                  15
Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Glu Val Leu Ile Arg
            20                  25                  30
Leu Phe Thr Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys
        35                  40                  45
His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp Leu Lys Lys
    50                  55                  60
His Gly Asn Thr Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys
65                  70                  75                  80
Gly His His Glu Ala Glu Val Lys His Leu Ala Glu Ser His Ala Asn
                85                  90                  95
Lys His Lys Ile Pro Val Lys Tyr Leu Glu Phe Ile Ser Asp Ala Ile
            100                 105                 110
Ile His Val Leu His Ala Lys His Pro Ser Asp Phe Gly Ala Asp Ala
        115                 120                 125
Gln Ala Ala Met Ser Lys Ala Leu Glu Leu Phe Arg Asn Asp Met Ala
    130                 135                 140
Ala Gln Tyr Lys Val Leu Gly Phe His Gly
145                 150
```

<210> 19
<211> 154
<212> PRT
<213> *Sus scrofa*

<400> 19

```
Met Gly Leu Ser Asp Gly Glu Trp Gln Leu Val Leu Asn Val Trp Gly
 1               5                  10                  15
Lys Val Glu Ala Asp Val Ala Gly His Gly Gln Glu Val Leu Ile Arg
            20                  25                  30
Leu Phe Lys Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys
        35                  40                  45
His Leu Lys Ser Glu Asp Glu Met Lys Ala Ser Glu Asp Leu Lys Lys
    50                  55                  60
His Gly Asn Thr Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys
65                  70                  75                  80
Gly His His Glu Ala Glu Leu Thr Pro Leu Ala Gln Ser His Ala Thr
                85                  90                  95
Lys His Lys Ile Pro Val Lys Tyr Leu Glu Phe Ile Ser Glu Ala Ile
            100                 105                 110
Ile Gln Val Leu Gln Ser Lys His Pro Gly Asp Phe Gly Ala Asp Ala
        115                 120                 125
Gln Gly Ala Met Ser Lys Ala Leu Glu Leu Phe Arg Asn Asp Met Ala
    130                 135                 140
Ala Lys Tyr Lys Glu Leu Gly Phe Gln Gly
145                 150
```

<210> 20
<211> 154
<212> PRT
<213> *Equus caballus*

<400> 20

```
Met Gly Leu Ser Asp Gly Glu Trp Gln Gln Val Leu Asn Val Trp Gly
 1               5                  10                  15
Lys Val Glu Ala Asp Ile Ala Gly His Gly Gln Glu Val Leu Ile Arg
            20                  25                  30
Leu Phe Thr Gly His Pro Glu Thr Leu Glu Lys Phe Asp Lys Phe Lys
        35                  40                  45
His Leu Lys Thr Glu Ala Glu Met Lys Ala Ser Glu Asp Leu Lys Lys
```

```
     50                  55                  60
His Gly Thr Val Val Leu Thr Ala Leu Gly Gly Ile Leu Lys Lys Lys
65                  70                  75                  80
Gly His His Glu Ala Glu Leu Lys Pro Leu Ala Gln Ser His Ala Thr
                85                  90                  95
Lys His Lys Ile Pro Ile Lys Tyr Leu Glu Phe Ile Ser Asp Ala Ile
            100                 105                 110
Ile His Val Leu His Ser Lys His Pro Gly Asp Phe Gly Ala Asp Ala
        115                 120                 125
Gln Gly Ala Met Thr Lys Ala Leu Glu Leu Phe Arg Asn Asp Ile Ala
    130                 135                 140
Ala Lys Tyr Lys Glu Leu Gly Phe Gln Gly
145                 150
```

<210> 21
<211> 124
<212> PRT
<213> *Synechocystis PCC6803*

<400> 21

```
Met Ser Thr Leu Tyr Glu Lys Leu Gly Gly Thr Thr Ala Val Asp Leu
 1               5                  10                  15
Ala Val Asp Lys Phe Tyr Glu Arg Val Leu Gln Asp Asp Arg Ile Lys
            20                  25                  30
His Phe Phe Ala Asp Val Asp Met Ala Lys Gln Arg Ala His Gln Lys
        35                  40                  45
Ala Phe Leu Thr Tyr Ala Phe Gly Gly Thr Asp Lys Tyr Asp Gly Arg
    50                  55                  60
Tyr Met Arg Glu Ala His Lys Glu Leu Val Glu Asn His Gly Leu Asn
65                  70                  75                  80
Gly Glu His Phe Asp Ala Val Ala Glu Asp Leu Leu Ala Thr Leu Lys
                85                  90                  95
Glu Met Gly Val Pro Glu Asp Leu Ile Ala Glu Val Ala Ala Val Ala
            100                 105                 110
Gly Ala Pro Ala His Lys Arg Asp Val Leu Asn Gln
        115                 120
```

<210> 22
<211> 183
<212> PRT
<213> *Synechococcus sp. PCC 7335*

<400> 22

```
Met Asp Val Ala Leu Leu Glu Lys Ser Phe Glu Gln Ile Ser Pro Arg
 1               5                  10                  15
Ala Ile Glu Phe Ser Ala Ser Phe Tyr Gln Asn Leu Phe His His His
            20                  25                  30
Pro Glu Leu Lys Pro Leu Phe Ala Glu Thr Ser Gln Thr Ile Gln Glu
        35                  40                  45
Lys Lys Leu Ile Phe Ser Leu Ala Ala Ile Ile Glu Asn Leu Arg Asn
    50                  55                  60
Pro Asp Ile Leu Gln Pro Ala Leu Lys Ser Leu Gly Ala Arg His Ala
65                  70                  75                  80
Glu Val Gly Thr Ile Lys Ser His Tyr Pro Leu Val Gly Gln Ala Leu
                85                  90                  95
Ile Glu Thr Phe Ala Glu Tyr Leu Ala Ala Asp Trp Thr Glu Gln Leu
            100                 105                 110
Ala Thr Ala Trp Val Glu Ala Tyr Asp Val Ile Ala Ser Thr Met Ile
        115                 120                 125
Glu Gly Ala Asp Asn Pro Ala Ala Tyr Leu Glu Pro Glu Leu Thr Phe
    130                 135                 140
```

```
Tyr Glu Trp Leu Asp Leu Tyr Gly Glu Glu Ser Pro Lys Val Arg Asn
145                 150                 155                 160
Ala Ile Ala Thr Leu Thr His Phe His Tyr Gly Glu Asp Pro Gln Asp
                165                 170                 175
Val Gln Arg Asp Ser Arg Gly
            180
```

<210> 23
<211> 118
<212> PRT
<213> *Nostoc commune*

<400> 23

```
Met Ser Thr Leu Tyr Asp Asn Ile Gly Gly Gln Pro Ala Ile Glu Gln
 1               5                  10                  15
Val Val Asp Glu Leu His Lys Arg Ile Ala Thr Asp Ser Leu Leu Ala
            20                  25                  30
Pro Val Phe Ala Gly Thr Asp Met Val Lys Gln Arg Asn His Leu Val
        35                  40                  45
Ala Phe Leu Ala Gln Ile Phe Glu Gly Pro Lys Gln Tyr Gly Gly Arg
    50                  55                  60
Pro Met Asp Lys Thr His Ala Gly Leu Asn Leu Gln Gln Pro His Phe
65                  70                  75                  80
Asp Ala Ile Ala Lys His Leu Gly Glu Arg Met Ala Val Arg Gly Val
                85                  90                  95
Ser Ala Glu Asn Thr Lys Ala Ala Leu Asp Arg Val Thr Asn Met Lys
            100                 105                 110
Gly Ala Ile Leu Asn Lys
        115
```

<210> 24
<211> 146
<212> PRT
<213> *Vitreoscilla stercoraria*

<400> 24

```
Met Leu Asp Gln Gln Thr Ile Asn Ile Ile Lys Ala Thr Val Pro Val
 1               5                  10                  15
Leu Lys Glu His Gly Val Thr Ile Thr Thr Thr Phe Tyr Lys Asn Leu
            20                  25                  30
Phe Ala Lys His Pro Glu Val Arg Pro Leu Phe Asp Met Gly Arg Gln
        35                  40                  45
Glu Ser Leu Glu Gln Pro Lys Ala Leu Ala Met Thr Val Leu Ala Ala
    50                  55                  60
Ala Gln Asn Ile Glu Asn Leu Pro Ala Ile Leu Pro Ala Val Lys Lys
65                  70                  75                  80
Ile Ala Val Lys His Cys Gln Ala Gly Val Ala Ala Ala His Tyr Pro
                85                  90                  95
Ile Val Gly Gln Glu Leu Leu Gly Ala Ile Lys Glu Val Leu Gly Asp
            100                 105                 110
Ala Ala Thr Asp Asp Ile Leu Asp Ala Trp Gly Lys Ala Tyr Gly Val
        115                 120                 125
Ile Ala Asp Val Phe Ile Gln Val Glu Ala Asp Leu Tyr Ala Gln Ala
    130                 135                 140
Val Glu
145
```

<210> 25
<211> 131
<212> PRT
<213> *Corynebacterium glutamicum*

<400> 25

```
Met Thr Thr Ser Glu Asn Phe Tyr Asp Ser Val Gly Gly Glu Glu Thr
 1               5                  10                  15
Phe Ser Leu Ile Val His Arg Phe Tyr Glu Gln Val Pro Asn Asp Asp
            20                  25                  30
Ile Leu Gly Pro Met Tyr Pro Pro Asp Asp Phe Glu Gly Ala Glu Gln
        35                  40                  45
Arg Leu Lys Met Phe Leu Ser Gln Tyr Trp Gly Gly Pro Lys Asp Tyr
    50                  55                  60
Gln Glu Gln Arg Gly His Pro Arg Leu Arg Met Arg His Val Asn Tyr
65                  70                  75                  80
Pro Ile Gly Val Thr Ala Ala Glu Arg Trp Leu Gln Leu Met Ser Asn
                85                  90                  95
Ala Leu Asp Gly Val Asp Leu Thr Ala Glu Gln Arg Glu Ala Ile Trp
            100                 105                 110
Glu His Met Val Arg Ala Ala Asp Met Leu Ile Asn Ser Asn Pro Asp
        115                 120                 125
Pro His Ala
    130
```

<210> 26
<211> 132
<212> PRT
<213> *Bacillus subtilis*

<400> 26

```
Met Gly Gln Ser Phe Asn Ala Pro Tyr Glu Ala Ile Gly Glu Glu Leu
 1               5                  10                  15
Leu Ser Gln Leu Val Asp Thr Phe Tyr Glu Arg Val Ala Ser His Pro
            20                  25                  30
Leu Leu Lys Pro Ile Phe Pro Ser Asp Leu Thr Glu Thr Ala Arg Lys
        35                  40                  45
Gln Lys Gln Phe Leu Thr Gln Tyr Leu Gly Gly Pro Pro Leu Tyr Thr
    50                  55                  60
Glu Glu His Gly His Pro Met Leu Arg Ala Arg His Leu Pro Phe Pro
65                  70                  75                  80
Ile Thr Asn Glu Arg Ala Asp Ala Trp Leu Ser Cys Met Lys Asp Ala
                85                  90                  95
Met Asp His Val Gly Leu Glu Gly Glu Ile Arg Glu Phe Leu Phe Gly
            100                 105                 110
Arg Leu Glu Leu Thr Ala Arg His Met Val Asn Gln Thr Glu Ala Glu
        115                 120                 125
Asp Arg Ser Ser
    130
```

<210> 27
<211> 136
<212> PRT
<213> *Bacillus megaterium*

<400> 27

```
Met Arg Glu Lys Ile His Ser Pro Tyr Glu Leu Leu Gly Gly Glu His
 1               5                  10                  15
Thr Ile Ser Lys Leu Val Asp Ala Phe Tyr Thr Arg Val Gly Gln His
            20                  25                  30
Pro Glu Leu Ala Pro Ile Phe Pro Asp Asn Leu Thr Glu Thr Ala Arg
        35                  40                  45
Lys Gln Lys Gln Phe Leu Thr Gln Tyr Leu Gly Gly Pro Ser Leu Tyr
    50                  55                  60
```

```
Thr Glu Glu His Gly His Pro Met Leu Arg Ala Arg His Leu Pro Phe
65                  70                  75                  80
Glu Ile Thr Pro Ser Arg Ala Lys Ala Trp Leu Thr Cys Met His Glu
                85                  90                  95
Ala Met Asp Glu Ile Asn Leu Glu Gly Pro Glu Arg Asp Glu Leu Tyr
            100                 105                 110
His Arg Leu Ile Leu Thr Ala Gln His Met Ile Asn Ser Pro Glu Gln
        115                 120                 125
Thr Asp Glu Lys Gly Phe Ser His
    130                 135
```

<210> 28
<211> 399
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 28

```
Met Leu Ala Glu Lys Thr Arg Ser Ile Ile Lys Ala Thr Val Pro Val
 1               5                  10                  15
Leu Glu Gln Gln Gly Thr Val Ile Thr Arg Thr Phe Tyr Lys Asn Met
            20                  25                  30
Leu Thr Glu His Thr Glu Leu Leu Asn Ile Phe Asn Arg Thr Asn Gln
        35                  40                  45
Lys Val Gly Ala Gln Pro Asn Ala Leu Ala Thr Thr Val Leu Ala Ala
    50                  55                  60
Ala Lys Asn Ile Asp Asp Leu Ser Val Leu Met Asp His Val Lys Gln
65                  70                  75                  80
Ile Gly His Lys His Arg Ala Leu Gln Ile Lys Pro Glu His Tyr Pro
                85                  90                  95
Ile Val Gly Glu Tyr Leu Leu Lys Ala Ile Lys Glu Val Leu Gly Asp
            100                 105                 110
Ala Ala Thr Pro Glu Ile Ile Asn Ala Trp Gly Glu Ala Tyr Gln Ala
        115                 120                 125
Ile Ala Asp Ile Phe Ile Thr Val Glu Lys Lys Met Tyr Glu Glu Ala
    130                 135                 140
Leu Trp Pro Gly Trp Lys Pro Phe Asp Ile Thr Ala Lys Glu Tyr Val
145                 150                 155                 160
Ala Ser Asp Ile Val Glu Phe Thr Val Lys Pro Lys Phe Gly Ser Gly
                165                 170                 175
Ile Glu Leu Glu Ser Leu Pro Ile Thr Pro Gly Gln Tyr Ile Thr Val
            180                 185                 190
Asn Thr His Pro Ile Arg Gln Glu Asn Gln Tyr Asp Ala Leu Arg His
        195                 200                 205
Tyr Ser Leu Cys Ser Ala Ser Thr Lys Asn Gly Leu Arg Phe Ala Val
    210                 215                 220
Lys Met Glu Ala Ala Arg Glu Asn Phe Pro Ala Gly Leu Val Ser Glu
225                 230                 235                 240
Tyr Leu His Lys Asp Ala Lys Val Gly Asp Glu Ile Lys Leu Ser Ala
                245                 250                 255
Pro Ala Gly Asp Phe Ala Ile Asn Lys Glu Leu Ile His Gln Asn Glu
            260                 265                 270
Val Pro Leu Val Leu Leu Ser Ser Gly Val Gly Val Thr Pro Leu Leu
        275                 280                 285
Ala Met Leu Glu Glu Gln Val Lys Cys Asn Pro Asn Arg Pro Ile Tyr
    290                 295                 300
Trp Ile Gln Ser Ser Tyr Asp Glu Lys Thr Gln Ala Phe Lys Lys His
305                 310                 315                 320
Val Asp Glu Leu Leu Ala Glu Cys Ala Asn Val Asp Lys Ile Ile Val
                325                 330                 335
His Thr Asp Thr Glu Pro Leu Ile Asn Ala Ala Phe Leu Lys Glu Lys
            340                 345                 350
Ser Pro Ala His Ala Asp Val Tyr Thr Cys Gly Ser Leu Ala Phe Met
```

```
        355                 360                 365
Gln Ala Met Ile Gly His Leu Lys Glu Leu Glu His Arg Asp Asp Met
    370                 375                 380
Ile His Tyr Glu Pro Phe Gly Pro Lys Met Ser Thr Val Gln Val
385                 390                 395
```

<210> 29
<211> 152
<212> PRT
<213> *Nicotina tobaccum*

<400> 29

```
Met Ser Ser Phe Ser Glu Glu Gln Glu Ala Leu Val Leu Lys Ser Trp
 1               5                  10                  15
Asp Ser Met Lys Lys Asn Ala Gly Glu Trp Gly Leu Lys Leu Phe Leu
            20                  25                  30
Lys Ile Phe Glu Ile Ala Pro Ser Ala Lys Lys Leu Phe Ser Phe Leu
        35                  40                  45
Lys Asp Ser Asn Val Pro Leu Glu Gln Asn Ala Lys Leu Lys Pro His
    50                  55                  60
Ala Lys Ser Val Phe Val Met Thr Cys Glu Ala Ala Val Gln Leu Arg
65                  70                  75                  80
Lys Ala Gly Lys Val Val Val Arg Asp Ser Thr Leu Lys Lys Leu Gly
                85                  90                  95
Ala Ala His Phe Lys Tyr Gly Val Ala Asp Glu His Phe Glu Val Thr
            100                 105                 110
Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala Val Pro Asp Met Trp
        115                 120                 125
Ser Val Asp Met Lys Asn Ala Trp Gly Glu Ala Phe Asp Gln Leu Val
    130                 135                 140
Asn Ala Ile Lys Thr Glu Met Lys
145                 150
```

<210> 30
<211> 160
<212> PRT
<213> *Medicago sativa*

<400> 30

```
Met Gly Thr Leu Asp Thr Lys Gly Phe Thr Glu Glu Gln Glu Ala Leu
 1               5                  10                  15
Val Val Lys Ser Trp Asn Ala Met Lys Lys Asn Ser Ala Glu Leu Gly
            20                  25                  30
Leu Lys Leu Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys
        35                  40                  45
Leu Phe Ser Phe Leu Lys Asp Ser Lys Val Pro Leu Glu Gln Asn Thr
    50                  55                  60
Lys Leu Lys Pro His Ala Met Ser Val Phe Leu Met Thr Cys Glu Ser
65                  70                  75                  80
Ala Val Gln Leu Arg Lys Ser Gly Lys Val Thr Val Arg Glu Ser Ser
                85                  90                  95
Leu Lys Lys Leu Gly Ala Asn His Phe Lys Tyr Gly Val Val Asp Glu
            100                 105                 110
His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala
        115                 120                 125
Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala
    130                 135                 140
Tyr Asp Gln Leu Val Asn Ala Ile Lys Ser Glu Met Lys Pro Ser Ser
145                 150                 155                 160
```

<210> 31
<211> 161
<212> PRT

<213> *Glycine max*

<400> 31

```
Met Thr Thr Thr Leu Glu Arg Gly Phe Ser Glu Glu Gln Glu Ala Leu
 1               5                  10                  15
Val Val Lys Ser Trp Asn Val Met Lys Lys Asn Ser Gly Glu Leu Gly
            20                  25                  30
Leu Lys Phe Phe Leu Lys Ile Phe Glu Ile Ala Pro Ser Ala Gln Lys
        35                  40                  45
Leu Phe Ser Phe Leu Arg Asp Ser Thr Val Pro Leu Glu Gln Asn Pro
    50                  55                  60
Lys Leu Lys Pro His Ala Val Ser Val Phe Val Met Thr Cys Asp Ser
65                  70                  75                  80
Ala Val Gln Leu Arg Lys Ala Gly Lys Val Thr Val Arg Glu Ser Asn
                85                  90                  95
Leu Lys Lys Leu Gly Ala Thr His Phe Arg Thr Gly Val Ala Asn Glu
            100                 105                 110
His Phe Glu Val Thr Lys Phe Ala Leu Leu Glu Thr Ile Lys Glu Ala
        115                 120                 125
Val Pro Glu Met Trp Ser Pro Ala Met Lys Asn Ala Trp Gly Glu Ala
    130                 135                 140
Tyr Asp Gln Leu Val Asp Ala Ile Lys Ser Glu Met Lys Pro Pro Ser
145                 150                 155                 160
Ser
```

<210> 32
<211> 97
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 32

```
atgaaaaaga aaaaaacatg gaaacgcttc ttacactttt cgagtgcagc tctggctgca 60
ggtttgatat tcacttctgc tgctcccgca gaggcag                           97
```

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

<220>
<223> péptido señal

<400> 33

```
Met Lys Lys Lys Lys Thr Trp Lys Arg Phe Leu His Phe Ser Ser Ala
 1               5                  10                  15
Ala Leu Ala Ala Gly Leu Ile Phe Thr Ser Ala Ala Pro Ala Glu Ala
            20                  25                  30
Ala
```

<210> 34
<211> 131
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 34

```
atgtcaccag cacaaagaag aattttactg tatatccttt catttatctt tgtcatcggc 60
gcagtcgtct attttgtcaa aagcgattat ctgtttacgc tgattttcat tgccattgcc 120
attctgttcg g                                                      131
```

<210> 35
<211> 43
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 35
Met Ser Pro Ala Gln Arg Arg Ile Leu Leu Tyr Ile Leu Ser Phe Ile
1 5 10 15
Phe Val Ile Gly Ala Val Val Tyr Phe Val Lys Ser Asp Tyr Leu Phe
20 25 30
Thr Leu Ile Phe Ile Ala Ile Ala Ile Leu Phe
35 40

<210> 36
<211> 90
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 36
atggtcagca tccgccgcag cttcgaagcg tatgtcgatg acatgaatat cattactgtt 60
ctgattcctg ctgaacaaaa ggaaatcatg 90

<210> 37
<211> 30
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 37
Met Val Ser Ile Arg Arg Ser Phe Glu Ala Tyr Val Asp Asp Met Asn
1 5 10 15
Ile Ile Thr Val Leu Ile Pro Ala Glu Gln Lys Glu Ile Met
20 25 30

<210> 38
<211> 99
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 38
atggctgcat atattatcag aagaacctta atgtctatcc cgattttatt gggaattacg 60
attttatcat ttgttatcat gaaagccgcg cccggagat 99

<210> 39
<211> 33
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 39

Met Ala Ala Tyr Ile Ile Arg Arg Thr Leu Met Ser Ile Pro Ile Leu
1 5 10 15
Leu Gly Ile Thr Ile Leu Ser Phe Val Ile Met Lys Ala Ala Pro Gly
20 25 30
Asp

<210> 40
<211> 87
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 40
atgaaacggt caatctctat ttttattacg tgtttattga ttacgttatt gacaatgggc 60
ggcatgatag cttcgccggc atcagca 87

<210> 41
<211> 29
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 41
Met Lys Arg Ser Ile Ser Ile Phe Ile Thr Cys Leu Leu Ile Thr Leu
1 5 10 15
Leu Thr Met Gly Gly Met Ile Ala Ser Pro Ala Ser Ala
20 25

<210> 42
<211> 93
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 42
atgccttatc tgaaacgagt gttgctgctt cttgtcactg gattgtttat gagtttgttt 60
gcagtcactg ctactgcctc agctcaaaca ggt 93

<210> 43
<211> 31
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 43
Met Pro Tyr Leu Lys Arg Val Leu Leu Leu Val Thr Gly Leu Phe
1 5 10 15
Met Ser Leu Phe Ala Val Thr Ala Thr Ala Ser Ala Gln Thr Gly
20 25 30

<210> 44
<211> 102
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 44
atgaaatttg taaaaagaag gatcattgca cttgtaacaa ttttgatgct gtctgttaca 60
tcgctgtttg cgttgcagcc gtcagcaaaa gccgctgaac ac 102

<210> 45
<211> 34
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 45
Met Lys Phe Val Lys Arg Arg Ile Ile Ala Leu Val Thr Ile Leu Met
1 5 10 15
Leu Ser Val Thr Ser Leu Phe Ala Leu Gln Pro Ser Ala Lys Ala Ala
20 25 30
Glu His

<210> 46
<211> 90
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 46
atgaaaaaga gactaatcgc acctatgctt ctatccgccg cgtcccttgc ctttttgcc 60
atgtctggtt ctgcccaggc agccgcgtat 90

<210> 47
<211> 30
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 47
Met Lys Lys Arg Leu Ile Ala Pro Met Leu Leu Ser Ala Ala Ser Leu
1 5 10 15
Ala Phe Phe Ala Met Ser Gly Ser Ala Gln Ala Ala Ala Tyr
20 25 30

<210> 48
<211> 72
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 48
atgaaaaaac gttggtcgat tgtcacgttg atgctcattt tcactctcgt gctgagcgcg 60
tgcggctttg gc 72

<210> 49
<211> 24
<212> PRT
<213> Secuencia artificial

<220>

<223> péptido señal

<400> 49

```
Met Lys Lys Arg Trp Ser Ile Val Thr Leu Met Leu Ile Phe Thr Leu
 1               5                  10                  15
Val Leu Ser Ala Cys Gly Phe Gly
            20
```

<210> 50
<211> 99
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 50

```
atgctaaaat atatcggaag acgcttagtc tatatgatta tcacactatt tgtgattgta 60
actgtgacat tcttcttaat gcaagcagca ccgggcggg                         99
```

<210> 51
<211> 33
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 51

```
Met Leu Lys Tyr Ile Gly Arg Arg Leu Val Tyr Met Ile Ile Thr Leu
 1               5                  10                  15
Phe Val Ile Val Thr Val Thr Phe Phe Leu Met Gln Ala Ala Pro Gly
            20                  25                  30
Gly
```

<210> 52
<211> 126
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 52

```
atgacaagcc caacccgcag aagaactgcg aaacgcagac ggagaaaact aaataaaaga 60
ggaaaactgt tgtttggtct tttagcagtg atggtttgca ttacgatttg gaatgctctt 120
catcga                                                            126
```

<210> 53
<211> 42
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 53

```
Met Thr Ser Pro Thr Arg Arg Arg Thr Ala Lys Arg Arg Arg Arg Lys
 1               5                  10                  15
Leu Asn Lys Arg Gly Lys Leu Leu Phe Gly Leu Leu Ala Val Met Val
            20                  25                  30
Cys Ile Thr Ile Trp Asn Ala Leu His Arg
        35                  40
```

<210> 54
<211> 168
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 54
atggcatacg acagtcgttt tgatgaatgg gtacagaaac tgaaagagga aagctttcaa 60
aacaatacgt ttgaccgccg caaatttatt caaggagcgg ggaagattgc aggactttct 120
cttggattaa cgattgccca gtcggttggg gcctttgaag taaatgct 168

<210> 55
<211> 56
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 55
Met Ala Tyr Asp Ser Arg Phe Asp Glu Trp Val Gln Lys Leu Lys Glu
1 5 10 15
Glu Ser Phe Gln Asn Asn Thr Phe Asp Arg Arg Lys Phe Ile Gln Gly
20 25 30
Ala Gly Lys Ile Ala Gly Leu Ser Leu Gly Leu Thr Ile Ala Gln Ser
35 40 45
Val Gly Ala Phe Glu Val Asn Ala
50 55

<210> 56
<211> 117
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 56
atgggcggaa aacatgatat atccagacgt caatttttga ttatacgct cacaggcgta 60
ggaggtttta tggcggctag tatgctcatg cctatggttc gcttcgcact cgacccg 117

<210> 57
<211> 39
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 57
Met Gly Gly Lys His Asp Ile Ser Arg Arg Gln Phe Leu Asn Tyr Thr
1 5 10 15
Leu Thr Gly Val Gly Gly Phe Met Ala Ala Ser Met Leu Met Pro Met
20 25 30
Val Arg Phe Ala Leu Asp Pro
35

<210> 58
<211> 78
<212> ADN
<213> Secuencia artificial

<220>

<223> codifica para péptido señal

<400> 58
atgttgttga aaaggagaat agggttgcta ttaagtatgg ttggcgtatt catgcttttg 60
gctggatgct cgagtgtg 78

<210> 59
<211> 26
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 59
Met Leu Leu Lys Arg Arg Ile Gly Leu Leu Leu Ser Met Val Gly Val
1 5 10 15
Phe Met Leu Leu Ala Gly Cys Ser Ser Val
20 25

<210> 60
<211> 129
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 60
atgaaaaaaa cactcaccac tattcgcaga tcatcaattg caaggagact tattatttct 60
ttcctgctga tcttaattgt tccgataacc gccctttcgg ttagcgctta tcaatcagca 120
gttgcctca 129

<210> 61
<211> 43
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 61
Met Lys Lys Thr Leu Thr Thr Ile Arg Arg Ser Ser Ile Ala Arg Arg
1 5 10 15
Leu Ile Ile Ser Phe Leu Leu Ile Leu Ile Val Pro Ile Thr Ala Leu
20 25 30
Ser Val Ser Ala Tyr Gln Ser Ala Val Ala Ser
35 40

<210> 62
<211> 105
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 62
atgaaaaaaa gaaagaggcg aaactttaaa aggttcattg cagcattttt agtgttggct 60
ttaatgattt cattagtgcc agccgatgta ctagcaaaat ctaca 105

<210> 63
<211> 35
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 63
Met Lys Lys Arg Lys Arg Arg Asn Phe Lys Arg Phe Ile Ala Ala Phe
1 5 10 15
Leu Val Leu Ala Leu Met Ile Ser Leu Val Pro Ala Asp Val Leu Ala
20 25 30
Lys Ser Thr
35

<210> 64
<211> 102
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 64
atgaaacgca gaaaattcag ctcggttgtg gcggcagtgc ttatttttgc actgatttc 60
agccttttt ctccgggaac caaagctgca gcggccggcg cg 102

<210> 65
<211> 34
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 65
Met Lys Arg Arg Lys Phe Ser Ser Val Val Ala Ala Val Leu Ile Phe
1 5 10 15
Ala Leu Ile Phe Ser Leu Phe Ser Pro Gly Thr Lys Ala Ala Ala Ala
20 25 30
Gly Ala

<210> 66
<211> 78
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 66
atgaagaaac gcagaaagat atgttattgc aatactgccc tgctgcttat gattttgctt 60
gctggatgta cggacagt 78

<210> 67
<211> 26
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 67
Met Lys Lys Arg Arg Lys Ile Cys Tyr Cys Asn Thr Ala Leu Leu Leu
1 5 10 15
Met Ile Leu Leu Ala Gly Cys Thr Asp Ser
20 25

<210> 68
<211> 129
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 68
atgaagaaaa gagttgctgg ctggtacagg cggatgaaga ttaaggataa gctgtttgtg 60
tttctatcgt tgattatggc cgtatccttt ctgtttgtat acagcggggt ccagtatgcc 120
tttcatgtg 129

<210> 69
<211> 43
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 69
Met Lys Lys Arg Val Ala Gly Trp Tyr Arg Arg Met Lys Ile Lys Asp
1 5 10 15
Lys Leu Phe Val Phe Leu Ser Leu Ile Met Ala Val Ser Phe Leu Phe
20 25 30
Val Tyr Ser Gly Val Gln Tyr Ala Phe His Val
35 40

<210> 70
<211> 114
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 70
atgagaagga gctgtctgat gattagacga aggaaacgca tgtttaccgc tgttacgttg 60
ctggtcttgt tggtgatggg aacctctgta tgtcctgtga aagctgaagg ggca 114

<210> 71
<211> 38
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 71
Met Arg Arg Ser Cys Leu Met Ile Arg Arg Arg Lys Arg Met Phe Thr
1 5 10 15
Ala Val Thr Leu Leu Val Leu Leu Val Met Gly Thr Ser Val Cys Pro
20 25 30
Val Lys Ala Glu Gly Ala
35

<210> 72
<211> 114
<212> ADN
<213> Secuencia artificial

<220>

<223> codifica para péptido señal

<400> 72
atgagaatac agaaaagacg aacacacgtc gaaaacattc tccgtattct tttgccccca 60
attatgatac ttagcctaat cctcccaaca ccacccattc atgcagaaga aagc 114

<210> 73
<211> 38
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 73
Met Arg Ile Gln Lys Arg Arg Thr His Val Glu Asn Ile Leu Arg Ile
1 5 10 15
Leu Leu Pro Pro Ile Met Ile Leu Ser Leu Ile Leu Pro Thr Pro Pro
20 25 30
Ile His Ala Glu Glu Ser
35

<210> 74
<211> 147
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 74
atgctgtctg tcgaaatgat aagcagacaa aatcgttgtc attatgtgta taagggagga 60
aatatgatga ggcgtattct gcatattgtg ttgatcacgg cattaatgtt cttaaatgta 120
atgtacacgt tcgaagctgt aaaggca 147

<210> 75
<211> 49
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 75
Met Leu Ser Val Glu Met Ile Ser Arg Gln Asn Arg Cys His Tyr Val
1 5 10 15
Tyr Lys Gly Gly Asn Met Met Arg Arg Ile Leu His Ile Val Leu Ile
20 25 30
Thr Ala Leu Met Phe Leu Asn Val Met Tyr Thr Phe Glu Ala Val Lys
35 40 45
Ala

<210> 76
<211> 93
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 76
atgctaagag atttaggaag aagagtagcg atcgcagcca ttttaagcgg aattattctt 60
ggaggcatga gcatttcttt ggcaaatatg ccc 93

<210> 77
<211> 31
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 77
Met Leu Arg Asp Leu Gly Arg Arg Val Ala Ile Ala Ala Ile Leu Ser
1 5 10 15
Gly Ile Ile Leu Gly Gly Met Ser Ile Ser Leu Ala Asn Met Pro
20 25 30

<210> 78
<211> 102
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 78
atgaaaaaga tgtccagaag acaatttcta aaaggaatgt tcggcgctct tgctgccggg 60
gctttaacgg ccggcggggg atatggctat gccaggtatc tc 102

<210> 79
<211> 34
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 79
Met Lys Lys Met Ser Arg Arg Gln Phe Leu Lys Gly Met Phe Gly Ala
1 5 10 15
Leu Ala Ala Gly Ala Leu Thr Ala Gly Gly Gly Tyr Gly Tyr Ala Arg
20 25 30
Tyr Leu

<210> 80
<211> 84
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 80
atgagaagat ttttactaaa tgtcatatta gtcttagcca ttgtcttgtt cttgagatat 60
gttcattact cattggaacc agaa 84

<210> 81
<211> 28
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 81

Met Arg Arg Phe Leu Leu Asn Val Ile Leu Val Leu Ala Ile Val Leu
1 5 10 15
Phe Leu Arg Tyr Val His Tyr Ser Leu Glu Pro Glu
20 25

<210> 82
<211> 87
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 82
atgtttgaaa gtgaagcaga actgagacga atcaggattg cacttgtatg gatagctgtc 60
tttttactgt tcggggcgtg cgggaat 87

<210> 83
<211> 29
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 83
Met Phe Glu Ser Glu Ala Glu Leu Arg Arg Ile Arg Ile Ala Leu Val
1 5 10 15
Trp Ile Ala Val Phe Leu Leu Phe Gly Ala Cys Gly Asn
20 25

<210> 84
<211> 93
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 84
atgaagaaga gaattacata ttcactgctt gctcttctag cagttgttgc tttcgctttc 60
actgattcat caaaagcaaa agcggcagaa gca 93

<210> 85
<211> 31
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 85
Met Lys Lys Arg Ile Thr Tyr Ser Leu Leu Ala Leu Leu Ala Val Val
1 5 10 15
Ala Phe Ala Phe Thr Asp Ser Ser Lys Ala Lys Ala Ala Glu Ala
20 25 30

<210> 86
<211> 116
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 86
atgcagaaat atagacgcag aaacacggtt gcctttacag tactagctta ttttactttt 60
tttgcgggag tatttttgtt tagtatcgga ctctataatg ctgataatct ggaact 116

<210> 87
<211> 38
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 87
Met Gln Lys Tyr Arg Arg Arg Asn Thr Val Ala Phe Thr Val Leu Ala
1 5 10 15
Tyr Phe Thr Phe Phe Ala Gly Val Phe Leu Phe Ser Ile Gly Leu Tyr
20 25 30
Asn Ala Asp Asn Leu Glu
35

<210> 88
<211> 102
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 88
atgatgttga atatgatcag acgtttgctg atgacctgtt tatttctgct tgcatttggc 60
acgacatttt tatcagtgtc aggaattgaa gcgaaggact tg 102

<210> 89
<211> 34
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 89
Met Met Leu Asn Met Ile Arg Arg Leu Leu Met Thr Cys Leu Phe Leu
1 5 10 15
Leu Ala Phe Gly Thr Thr Phe Leu Ser Val Ser Gly Ile Glu Ala Lys
20 25 30
Asp Leu

<210> 90
<211> 132
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 90
atggctgaac gcgttagagt gcgtgtgcga aaaaagaaaa agagcaaacg taggaaaatt 60
ttaaaaagaa taatgttatt gttcgccctt gcactattgg tagttgtagg gcttggcggg 120
tataaacttt at 132

<210> 91
<211> 44
<212> PRT

<213> Secuencia artificial

<220>
<223> péptido señal

<400> 91

```
Met Ala Glu Arg Val Arg Val Arg Val Arg Lys Lys Lys Lys Ser Lys
 1               5                  10                  15
Arg Arg Lys Ile Leu Lys Arg Ile Met Leu Leu Phe Ala Leu Ala Leu
            20                  25                  30
Leu Val Val Val Gly Leu Gly Gly Tyr Lys Leu Tyr
        35                  40
```

<210> 92
<211> 141
<212> ADN
<213> Secuencia artificial

<220>
<223> codifica para péptido señal

<400> 92

```
atgagcgatg aacagaaaaa gccagaacaa attcacagac gggacatttt aaaatgggga 60
gcgatggcgg gggcagccgt tgcgatcggt gccagcggtc tcggcggtct cgctccgctt 120
gttcagactg cggctaagcc a                                           141
```

<210> 93
<211> 47
<212> PRT
<213> Secuencia artificial

<220>
<223> péptido señal

<400> 93

```
Met Ser Asp Glu Gln Lys Lys Pro Glu Gln Ile His Arg Arg Asp Ile
 1               5                  10                  15
Leu Lys Trp Gly Ala Met Ala Gly Ala Ala Val Ala Ile Gly Ala Ser
            20                  25                  30
Gly Leu Gly Gly Leu Ala Pro Leu Val Gln Thr Ala Ala Lys Pro
        35                  40                  45
```

<210> 94
<211> 204
<212> PRT
<213> Secuencia artificial

<220>
<223> proteína de fusión

<400> 94

```
Met Ala Tyr Asp Ser Arg Phe Asp Glu Trp Val Gln Lys Leu Lys Glu
 1               5                  10                  15
Glu Ser Phe Gln Asn Asn Thr Phe Asp Arg Arg Lys Phe Ile Gln Gly
            20                  25                  30
Ala Gly Lys Ile Ala Gly Leu Ser Leu Gly Leu Thr Ile Ala Gln Ser
        35                  40                  45
Ala Ser Ala Ala Gly Ala His Met Glu Leu Gly Leu Ser Glu Glu Thr
    50                  55                  60
Ile Arg Val Ile Lys Ser Thr Val Pro Leu Leu Lys Glu His Gly Thr
65                  70                  75                  80
Glu Ile Thr Ala Arg Met Tyr Glu Leu Leu Phe Ser Lys Tyr Pro Lys
                85                  90                  95
Thr Lys Glu Leu Phe Ala Gly Ala Ser Glu Glu Gln Pro Lys Lys Leu
            100                 105                 110
Ala Asn Ala Ile Ile Ala Tyr Ala Thr Tyr Ile Asp Arg Leu Glu Glu
        115                 120                 125
Leu Asp Asn Ala Ile Ser Thr Ile Ala Arg Ser His Val Arg Arg Asn
    130                 135                 140
Val Lys Pro Glu His Tyr Pro Leu Val Lys Glu Cys Leu Leu Gln Ala
145                 150                 155                 160
Ile Glu Glu Val Leu Asn Pro Gly Glu Glu Val Leu Lys Ala Trp Glu
                165                 170                 175
Glu Ala Tyr Asp Phe Leu Ala Lys Thr Leu Ile Thr Leu Glu Lys Lys
            180                 185                 190
Leu Tyr Ser Gln Pro Arg His His His His His His
        195                 200
```

<210> 95
<211> 156
<212> PRT
<213> Secuencia artificial

<220>
<223> sitio de reconocimiento de peptidasa I señal

<400> 95

```
Ala Ser Ala Ala Gly Ala His Met Glu Leu Gly Leu Ser Glu Glu Thr
 1               5                  10                  15
Ile Arg Val Ile Lys Ser Thr Val Pro Leu Leu Lys Glu His Gly Thr
            20                  25                  30
Glu Ile Thr Ala Arg Met Tyr Glu Leu Leu Phe Ser Lys Tyr Pro Lys
        35                  40                  45
Thr Lys Glu Leu Phe Ala Gly Ala Ser Glu Glu Gln Pro Lys Lys Leu
    50                  55                  60
Ala Asn Ala Ile Ile Ala Tyr Ala Thr Tyr Ile Asp Arg Leu Glu Glu
65                  70                  75                  80
Leu Asp Asn Ala Ile Ser Thr Ile Ala Arg Ser His Val Arg Arg Asn
                85                  90                  95
Val Lys Pro Glu His Tyr Pro Leu Val Lys Glu Cys Leu Leu Gln Ala
            100                 105                 110
Ile Glu Glu Val Leu Asn Pro Gly Glu Glu Val Leu Lys Ala Trp Glu
        115                 120                 125
Glu Ala Tyr Asp Phe Leu Ala Lys Thr Leu Ile Thr Leu Glu Lys Lys
    130                 135                 140
Leu Tyr Ser Gln Pro Arg His His His His His His
145                 150                 155
```

<210> 96
<211> 10
<212> PRT
<213> Secuencia artificial

<220>

<223> péptido N-terminal

<400> 96

Gly Ala His Met Glu Leu Gly Leu Ser Glu
1 5 10

REIVINDICACIONES

1. Método de producción de un polipéptido que contiene hemo seleccionado del grupo que consiste en una leghemoglobina, una eritrocruorina, una hemoglobina no simbiótica, una flavohemoglobina, una protoglobina, una cianoglobina, una globina I de Hell's gate, una hemoglobina bacteriana, una mioglobina de ciliado, una protoglobina, globina truncada 2/2, HbN, HbO y Glb3, comprendiendo dicho método hacer crecer una planta recombinante, comprendiendo dicha planta recombinante al menos un ácido nucleico exógeno que codifica para dicho polipéptido que contiene hemo, en el que dicha planta es de una especie distinta de *Nicotiana,* y purificar dicho polipéptido que contiene hemo de un tejido de dicha planta.

2. Método según la reivindicación 1, en el que dicho tejido comprende hojas, raíces, tallos o semillas.

3. Método según una cualquiera de las reivindicaciones 1-2, en el que dicho polipéptido que contiene hemo tiene al menos el 60% de identidad de secuencia con respecto a una secuencia de aminoácidos expuesta en las SEQ ID NO: 1-17 ó 21-31.

4. Método según una cualquiera de las reivindicaciones 1-3, en el que dicha planta es una planta *Glycine max, Zea mays* o *Arabidopsis thaliana*.

5. Método según cualquiera de las reivindicaciones 1-4, en el que el método es para producir un sucedáneo de carne, y comprende además mezclar el polipéptido purificado con grasas y/o lípidos para producir un sucedáneo de carne.

6. Composición que comprende:

   (a) un polipéptido purificado que contiene hemo seleccionado del grupo que consiste en una leghemoglobina, una eritrocruorina, una hemoglobina no simbiótica, una flavohemoglobina, una protoglobina, una cianoglobina, una globina I de Hell's gate, una hemoglobina bacteriana, una mioglobina de ciliado, una protoglobina, globina truncada 2/2, HbN, HbO y Glb3; y
   (b) un componente de una célula huésped de planta recombinante de la que se expresa y/o secreta el polipéptido, en el que dicha célula huésped es una célula de la planta *Glycine max*, una célula de la planta *Zea mays* o una célula de la planta *Arabidopsis thaliana*,
   en la que el polipéptido que contiene hemo era exógeno a dicha célula;
   en la que el componente incluye una pared celular, un compartimento subcelular (tal como complejo de Golgi, retículo endoplasmático o núcleo), ácido nucleico, proteína, ADN genómico y/o una membrana plasmática;
   en la que la composición es un sucedáneo de carne.

7. Composición según una cualquiera de las reivindicaciones 6, en la que dicha composición comprende al menos 1 parte por mil millones de dicho componente de dicha célula.

8. Composición según una cualquiera de las reivindicaciones 6-7, en la que dicha composición comprende como máximo el 1% (p/p) de dicho componente de dicha célula.

9. Composición según una cualquiera de las reivindicaciones 6-8, en la que dicho polipéptido que contiene hemo tiene al menos el 60% de identidad de secuencia con respecto a una secuencia de aminoácidos expuesta en las SEQ ID NO: 1-17 ó 21-31.

10. Método según la reivindicación 1, en el que la planta recombinante se elaboró usando TALENS.

11. Método según la reivindicación 1, en el que la planta recombinante se elaboró usando un sistema Cas9/Crispr.

12. Método según la reivindicación 1, en el que la planta recombinante se elaboró usando transformación mediada por *Agrobacterium*.

Figura 1

Purificación con Ni-NTA del sedimento celular

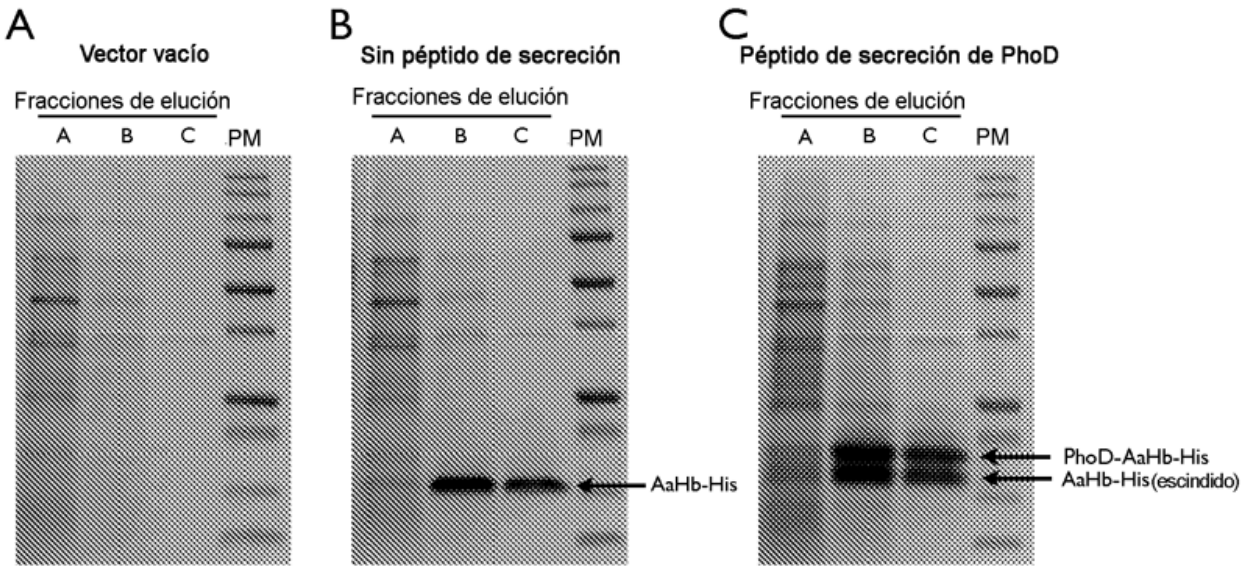

Figura 2

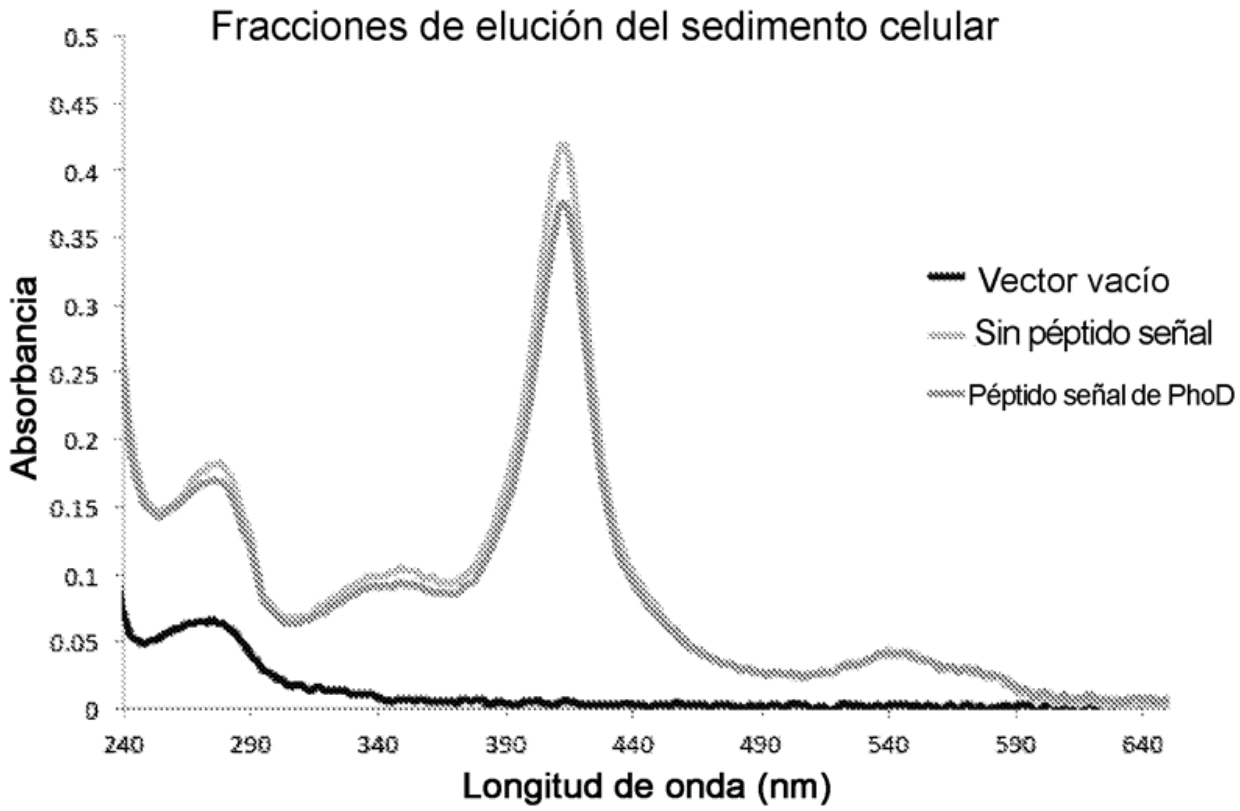
Fracciones de elución del sedimento celular
Absorbancia
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
240
290
340
390
440
490
540
590
640
Longitud de onda (nm)
Vector vacío
Sin péptido señal
Péptido señal de PhoD

Figura 3

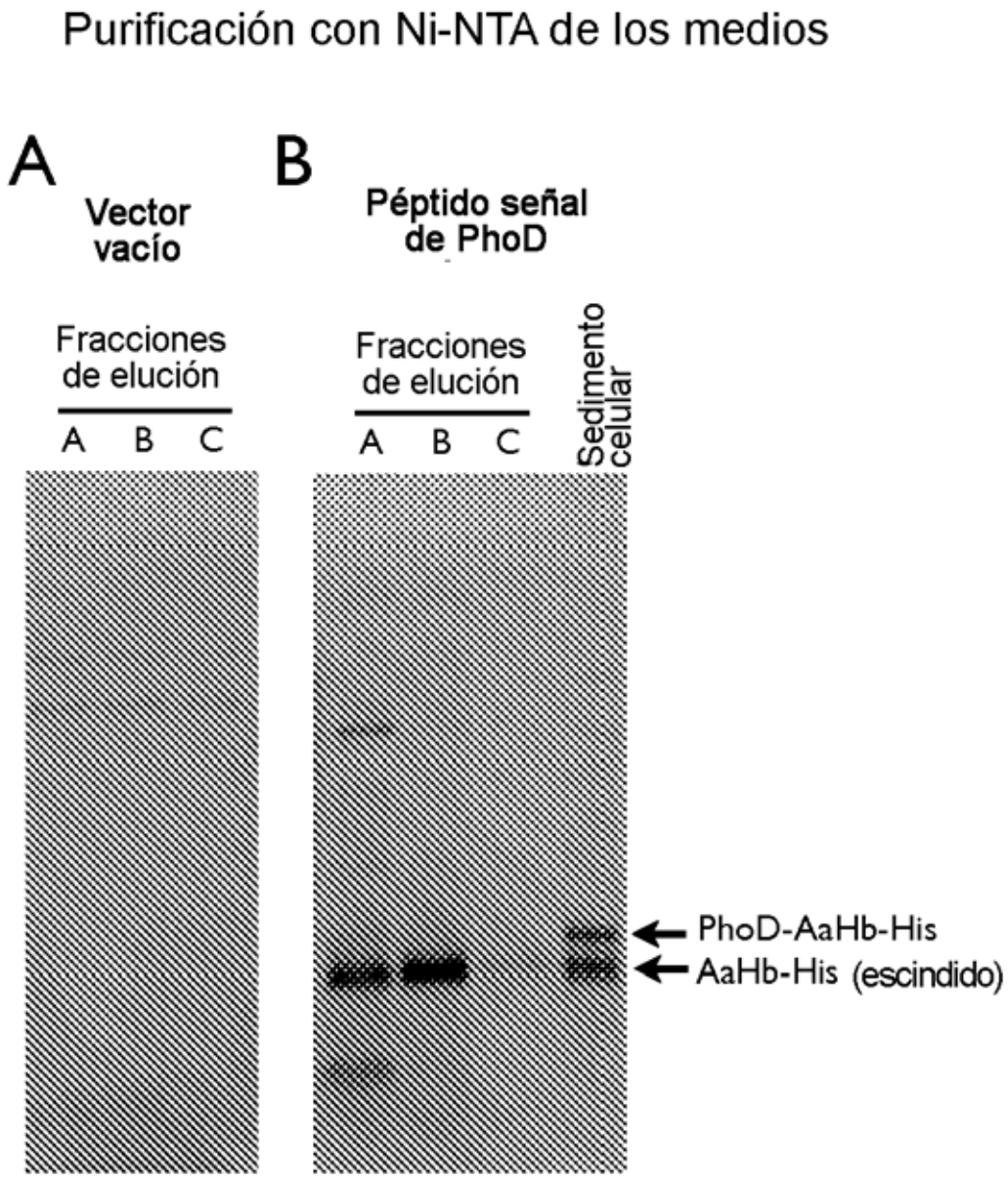
Purificación con Ni-NTA de los medios
A
Vector vacío
Fracciones de elución
A B C
B
Péptido señal de PhoD
Fracciones de elución
A B C
Sedimento celular
PhoD-AaHb-His
AaHb-His (escindido)

Figura 4

Secuencia de longitud completa de **PhoD**-*sitio de escisión de proteasa sintética*-AaHb-His6:

MAYDSRFDEWVQKLKEESFQNNTFDRRKFIQGAGKIAGLSLGLTIAQS*ASAA*GAHMELGLS
EETIRVIKSTVPLLKEHGTEITARMYELLFSKYPKTKELFAGASEEQPKKLANAIIAYATYIDRLEE
LDNAISTIARSHVRRNVKPEHYPLVKECLLQAIEEVLNPGEEVLKAWEEAYDFLAKTLITLEKKL
YSQPRHHHHHH

Sitio de reconocimiento predicho de peptidasa I señal (SPI): ASA/AGAHMELGLS...

Resultados de secuenciación N-terminal:

(N-term) GAHMELGLSE (C-term)

Figura 5

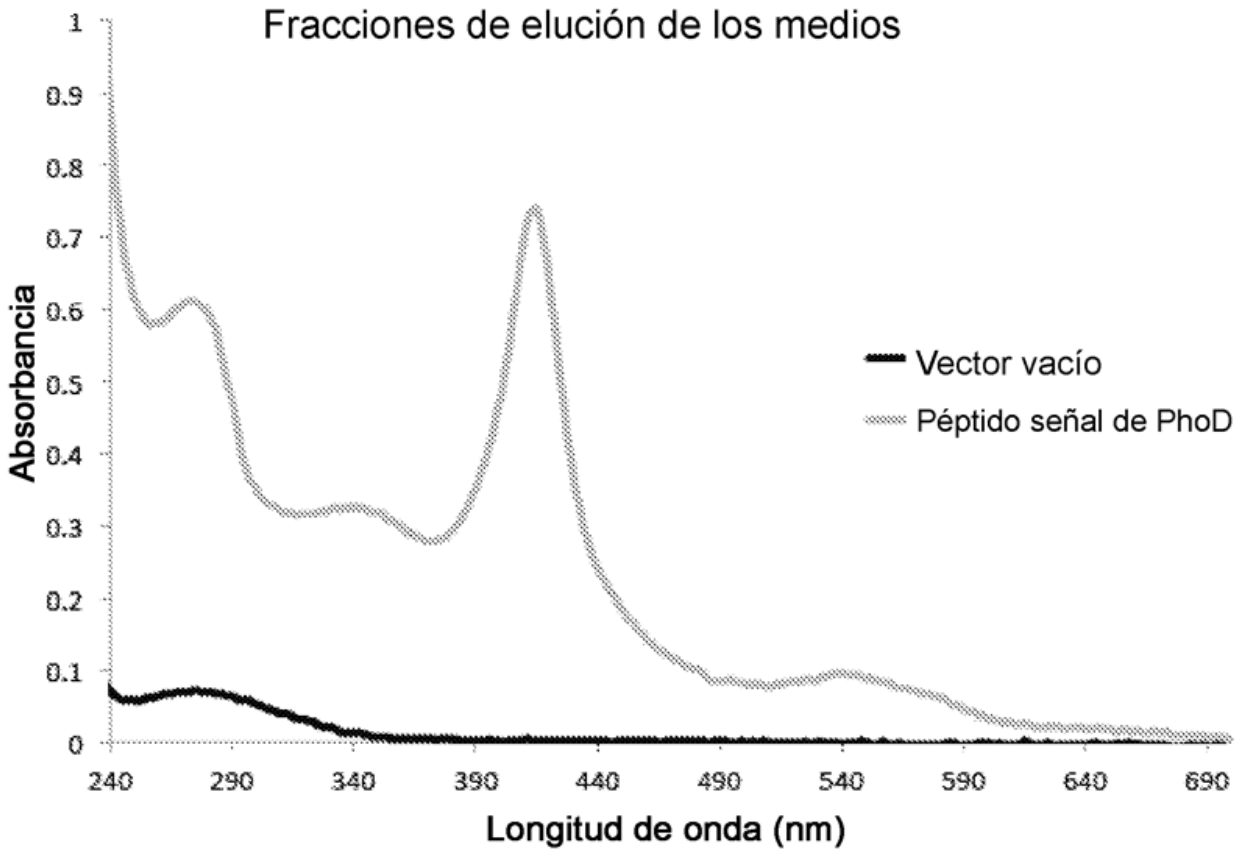
Fracciones de elución de los medios
Absorbancia
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
240
290
340
390
440
490
540
590
640
690
Longitud de onda (nm)
Vector vacío
Péptido señal de PhoD

Figura 6

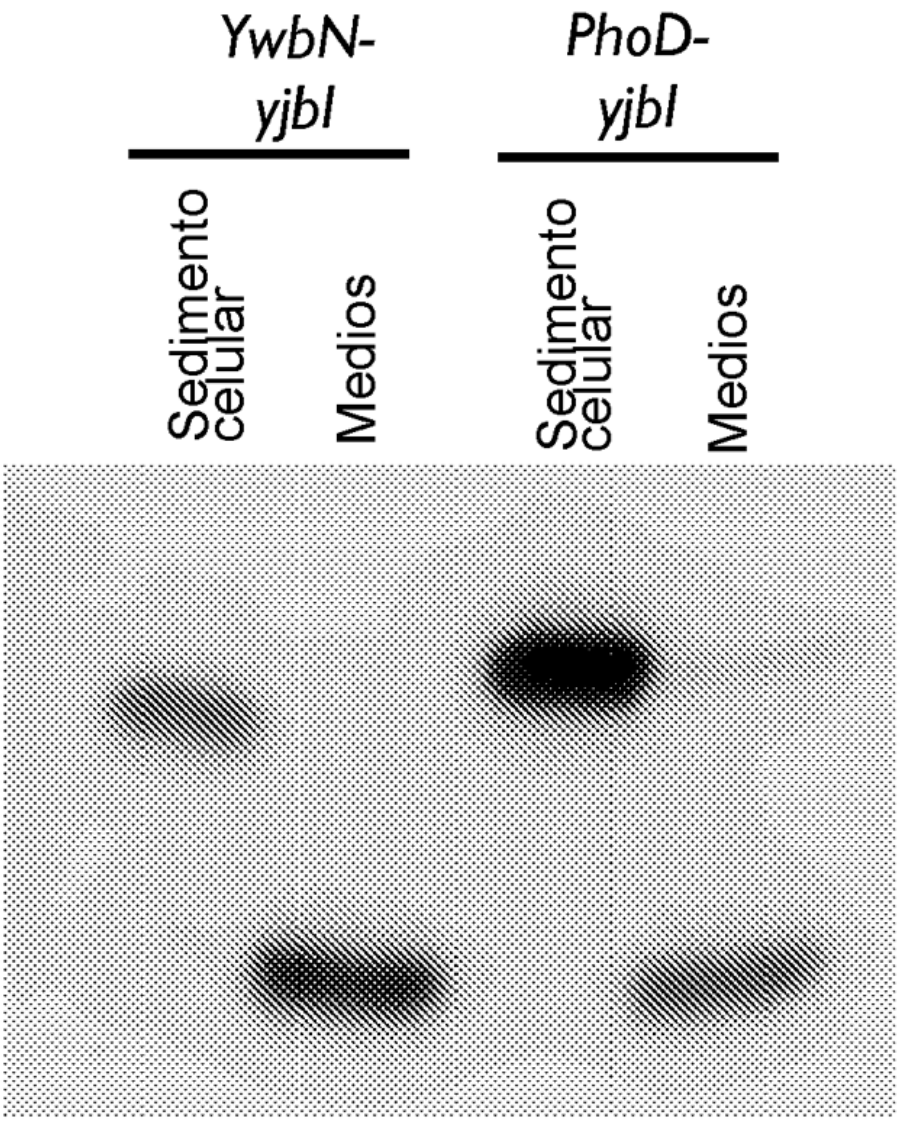
YwbN-yjbI
PhoD-yjbI
Sedimento celular
Medios
Sedimento celular
Medios

Figura 7

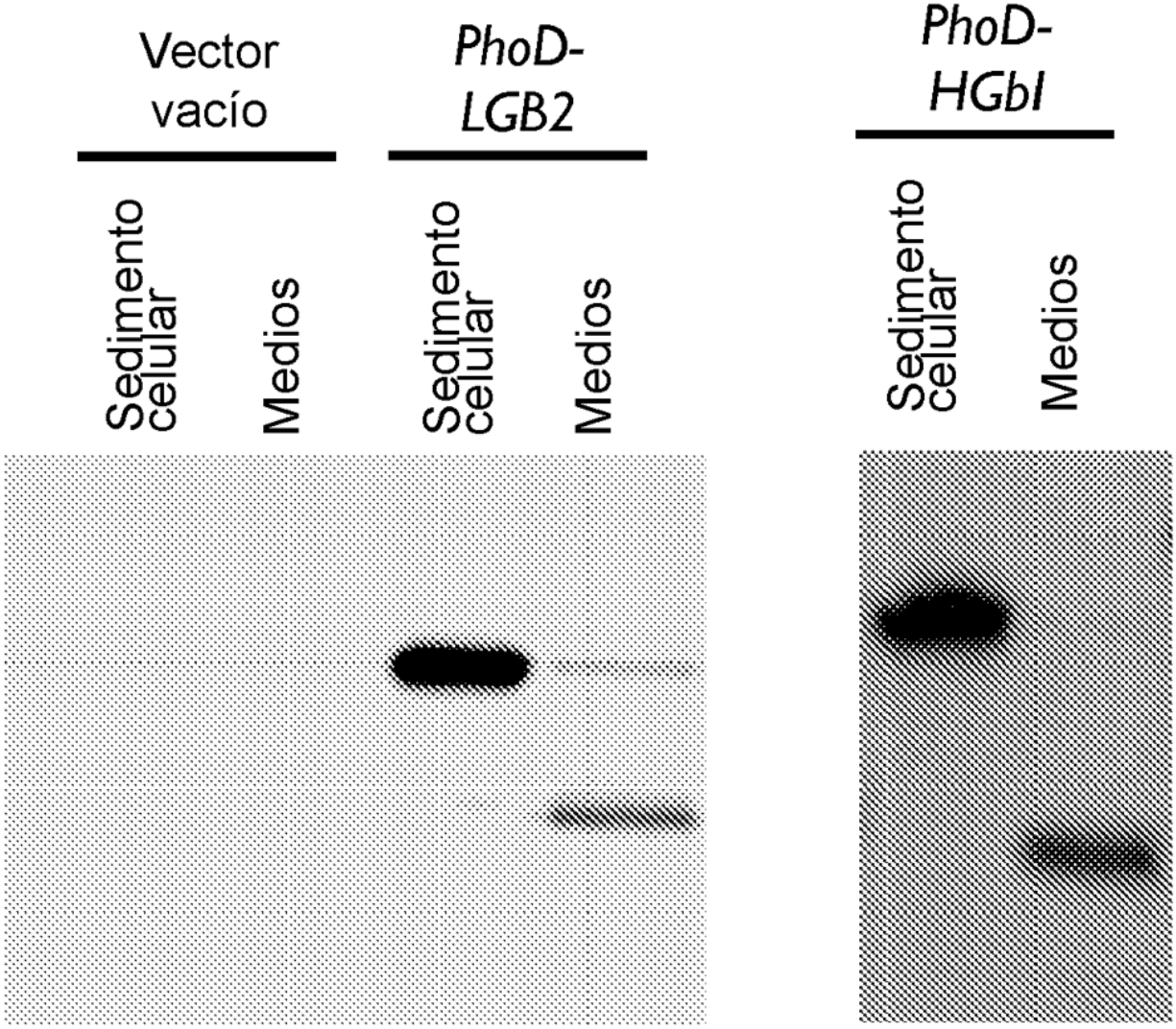
Vector vacío
PhoD-LGB2
PhoD-HGbI
Sedimento celular
Medios
Sedimento celular
Medios
Sedimento celular
Medios

Figura 8

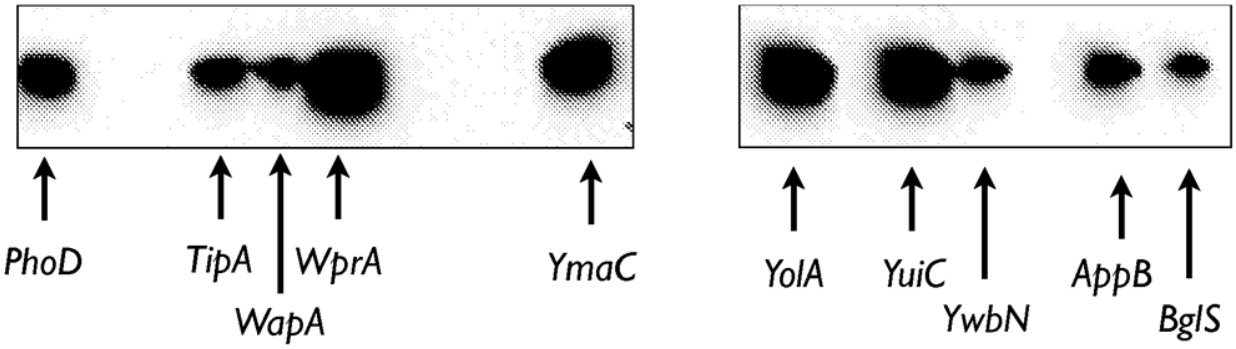

FIGURA 9

SEQ ID NO:1 *Vigna radiata*

MTTTLERGFTEEQEALVVKSWNVMKKNSGELGLKFFLKIFEIAPSAQKLFSFLRDSTVPLEQNPKLKPH
AVSVFVMTCDSAVQLRKAGKVTVRESNLKKLGATHFRTGVANEHFEVTKFALLETIKEAVPEMWSPAMK
NAWGEAYDQLVDAIKYEMKPPSS

SEQ ID NO:2 *Methylacidiphilum infernorum*

MIDQKEKELIKESWKRIEPNKNEIGLLFYANLFKEEPTVSVLFQNPISSQSRKLMQVLGILVQGIDNLE
GLIPTLQDLGRRHKQYGVVDSHYPLVGDCLLKSIQEYLGQGFTEEAKAAWTKVYGIAAQVMTAE

SEQ ID NO:3 *Aquifex aeolicus*

MLSEETIRVIKSTVPLLKEHGTEITARMYELLFSKYPKTKELFAGASEEQPKKLANAIIAYATYIDRLE
ELDNAISTIARSHVRRNVKPEHYPLVKECLLQAIEEVLNPGEEVLKAWEEAYDFLAKTLITLEKKLYSQ
P

SEQ ID NO:4 *Glycine max*

MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLTGHAEKLFGL
VRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGDKWSDELSSAWEVAYDELA
AAIKKAF

SEQ ID NO:5 *Hordeum vulgare*

MSAAEGAVVFSEEKEALVLKSWAIMKKDSANLGLRFFLKIFEIAPSARQMFPFLRDSDVPLETNPKLKT
HAVSVFVMTCEAAAQLRKAGKITVRETTLKRLGGTHLKYGVADGHFEVTRFALLETIKEALPADMWGPE
MRNAWGEAYDQLVAAIKQEMKPAE

Magnaporthe oryzae, (SEQ ID NO:6)

```
   1 mdgavrldwt gldltgheih dgvpiasrvq vmvsfplfkd qhiimsskes psrksstigq
  61 strngscqad tqkgqlppvg ekpkpvkenp mkklkemsqr plptqhgdgt yptekkltgi
 121 gedlkhirgy dvktllamvk sklkgeklkd dktmlmervm qlvarlptes kkraeltdsl
 181 inelwesldh pplnylgpeh syrtpdgsyn hpfnpqlgaa gsryarsvip tvtppgalpd
 241 pglifdsimg rtpnsyrkhp nnvssilwyw atiiihdifw tdprdintnk sssyldlapl
 301 ygnsqemqds irtfkdgrmk pdcyadkrla gmppgvsvll imfnrfhnhv aenlalineg
 361 grfnkpsdll egeareaawk kydndlfqva rlvtsglyin itlvdyvrni vnlnrvdttw
 421 tldprqdaga hvgtadgaer gtgnavsaef nlcyrwhsci sekdskfvea qfqnifgkpa
 481 sevrpdemwk gfakmeqntp adpgqrtfgg fkrgpdgkfd dddlvrcise avedvagafg
 541 arnvpqamkv vetmgiiqgr kwnvaglnef rkhfhlkpys tfedinsdpg vaealrrlyd
 601 hpdnvelypg lvaeedkqpm vpgvqiapty tisrvvlsda vclvrgdrfy ttdftprnlt
 661 nwgykevdyd lsvnhgcvfy klfirafpnh fkqnsvyahy pmvvpsenkr ilealgradl
 721 fdfeapkyip prvnitsygg aeyiletqek ykvtwheglg flmgegglkf mlsgddplha
 781 qqrkcmaaql ykdgwteavk afyagmmeel lvsksyflgn nkhrhvdiir dvgnmvhvhf
 841 asqvfglplk taknptgvft eqemygilaa ifttiffdld psksfplrtk trevcqklak
```

```
 901 lveanvklin kipwsrgmfv gkpakdepls iygktmikgl kahglsdydi awshvvptsg
 961 amvpnqaqvf aqavdyylsp agmhyipeih mvalqpstpe tdalllgyam egirlagtfg
1021 syreaavddv vkedngrqvp vkagdrvfvs fvdaardpkh fpdpevvnpr rpakkyihyg
1081 vgphaclgrd asqiaitemf rclfrrrnvr rvpgpqqgelk kvprpggfyv ymredwgglf
1141 pfpvtmrvmw dde
```

Fusarium oxysporum (SEQ ID NO:7)

```
   1 mkgsatlafa lvqfsaasql vwpskwdeve dllymqggfn krgfadalrt cefgsnvpgt
  61 qntaewlrta fhdaithdak aqtqqldasi ywessrpenp qkafnntfqf fsqfhnprat
 121 asdltalgtv lavgacngpr ipfragrida ykagpagvpe pstnlkdtfa aftkagftke
 181 emtamvacgh aiggvhsvdf peivgikadp nndtnvpfqk dvssfhngiv teylagtskn
 241 plvasknatf hsdkrifdnd katmkklstk agfnsmcadi ltrmidtvpk svqltpvlea
 301 ydvrpyitel slnnknkihf tgsvrvritn nirdnndlai nliyvgrdgk kvtvptqqvt
 361 fqggtsfgag evfanfefdt tmdaknqitk ffiqevkpst katvthdnqk tggykvddtv
 421 lyqlqqscav leklpnaplv vtamvrdara kdaltlrvah kkpvkgsivp rfqtaitnfk
 481 atgkkssgyt gfqaktmfee qstyfdivlg gspasgvqfl tsqampsqcs
```

Fusarium graminearum (SEQ ID NO:8)

```
   1 masatrqfar aatratrngf aiaprqvirq qgrryyssep aqksssawiw ltgaavagga
  61 gyyfygnsas satakvfnps kedyqkvyne iaarleekdd yddgsygpvl vrlawhasgt
 121 ydketgtggs ngatmrfape sdhganagla aardflqpvk ekfpwitysd lwilagvcai
 181 qemlgpaipy rpgrsdrdvs gctpdgrlpd askrqdhlrg ifgrmgfndq eivalsgaha
 241 lgrchtdrsg ysgpwtfspt vltndyfrll veekwqwkkw ngpaqyedks tkslmmlpsd
 301 ialiedkkfk pwvekyakdn daffkdfsnv vlrlfelgvp faqgtenqrw tfkpthqe
```

SEQ ID NO: 9 *Chlamydomonas eugametos*

MSLFAKLGGREAVEAAVDKFYNKIVADPTVSTYFSNTDMKVQRSKQFAFLAYALGGASEWKGKDMRTAH
KDLVPHLSDVHFQAVARHLSDTLTELGVPPEDITDAMAVVASTRTEVLNMPQQ

SEQ ID NO:10 *Tetrahymena pyriformis*

MNKPQTIYEKLGGENAMKAAVPLFYKKVLADERVKHFFKNTDMDHQTKQQTDFLTMLLGGPNHYKGKNM
TEAHKGMNLQNLHFDAIIENLAATLKELGVTDAVINEAAKVIEHTRKDMLGK

SEQ ID NO:11 *Paramecium caudatum*

MSLFEQLGGQAAVQAVTAQFYANIQADATVATFFNGIDMPNQTNKTAAFLCAALGGPNAWTGRNLKEVH
ANMGVSNAQFTTVIGHLRSALTGAGVAAALVEQTVAVAETVRGDVVTV

SEQ ID NO:12 *Aspergillus niger*

MPLTPEQIKIIKATVPVLQEYGTKITTAFYMNMSTVHPELNAVFNTANQVKGHQARALAGALFAYASHI
DDLGALGPAVELICNKHASLYIQADEYKIVGKYLLEAMKEVLGDACTDDILDAWGAAYWALADIMINRE
AALYKQSQG

SEQ ID NO:13 *Zea mays*

MALAEADDGAVVFGEEQEALVLKSWAVMKKDAANLGLRFFLKVFEIAPSAEQMFSFLRDSDVPLEKNPK
LKTHAMSVFVMTCEAAAQLRKAGKVTVRETTLKRLGATHLRYGVADGHFEVTGFALLETIKEALPADMW
SLEMKKAWAEAYSQLVAAIKREMKPDA

SEQ ID NO:14 *Oryza sativa subsp. japonica*

MALVEGNNGVSGGAVSFSEEQEALVLKSWAIMKKDSANIGLRFFLKIFEVAPSASQMFSFLRNSDVPLE
KNPKLKTHAMSVFVMTCEAAAQLRKAGKVTVRDTTLKRLGATHFKYGVGDAHFEVTRFALLETIKEAVP
VDMWSPAMKSAWSEAYNQLVAAIKQEMKPAE

SEQ ID NO:15 *Arabidopsis thaliana*

MESEGKIVFTEEQEALVVKSWSVMKKNSAELGLKLFIKIFEIAPTTKKMFSFLRDSPIPAEQNPKLKPH
AMSVFVMCCESAVQLRKTGKVTVRETTLKRLGASHSKYGVVDEHFEVAKYALLETIKEAVPEMWSPEMK
VAWGQAYDHLVAAIKAEMNLSN

SEQ ID NO:16 *Pisum sativum*

MGFTDKQEALVNSSWESFKQNLSGNSILFYTIILEKAPAAKGLFSFLKDTAGVEDSPKLQAHAEQVFGL
VRDSAAQLRTKGEVVLGNATLGAIHVQRGVTDPHFVVVKEALLQTIKKASGNNWSEELNTAWEVAYDGL
ATAIKKAMT

SEQ ID NO:17 *Vigna unguiculata*

MVAFSDKQEALVNGAYEAFKANIPKYSVVFYTTILEKAPAAKNLFSFLANGVDATNPKLTGHAEKLFGL
VRDSAAQLRASGGVVADAALGAVHSQKAVNDAQFVVVKEALVKTLKEAVGDKWSDELGTAVELAYDELA
AAIKKAY

SEQ ID NO:18 Bos taurus

MGLSDGEWQLVLNAWGKVEADVAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTV
LTALGGILKKKGHHEAEVKHLAESHANKHKIPVKYLEFISDAIIHVLHAKHPSDFGADAQAAMSKALEL
FRNDMAAQYKVLGFHG

SEQ ID NO:19 *Sus scrofa*
MGLSDGEWQLVLNVWGKVEADVAGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGNTV
LTALGGILKKKGHHEAELTPLAQSHATKHKIPVKYLEFISEAIIQVLQSKH
PGDFGADAQGAMSKALELFRNDMAAKYKELGFQG

SEQ ID NO: 20 *Equus caballus*

MGLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVV
LTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKH
PGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

SEQ ID NO: 21 *Synechocystis PCC6803*
MSTLYEKLGGTTAVDLAVDKFYERVLQDDRIKHFFADVDMAKQRAHQKAFLTYAFGGTDKY
DGRYMREAHKELVENHGLNGEHFDAVAEDLLATLKEMGVPEDLIAEVAAVAGAPAHKRDVL
NQ

SEQ ID NO: 22: *Synechococcus sp. PCC 7335*
MDVALLEKSFEQISPRAIEFSASFYQNLFHHHPELKPLFAETSQTIQEKKLIFSLAAIIENLRNPDIL
QPALKSLGARHAEVGTIKSHYPLVGQALIETFAEYLAADWTEQLATAWVEAYDVIASTMIEGA
DNPAAYLEPELTFYEWLDLYGEESPKVRNAIATLTHFHYGEDPQDVQRDSRG

SEQ ID NO: 23 *Nostoc commune*
MSTLYDNIGGQPAIEQVVDELHKRIATDSLLAPVFAGTDMVKQRNHLVAFLAQIFEGPKQYGG
RPMDKTHAGLNLQQPHFDAIAKHLGERMAVRGVSAENTKAALDRVTNMKGAILNK

SEQ ID NO: 24: *Vitreoscilla stercoraria*
MLDQQTINII KATVPVLKEH GVTITTTFYK NLFAKHPEVR PLFDMGRQES LEQPKALAMT VLAAAQNIEN
LPAILPAVKK IAVKHCQAGV AAAHYPIVGQ ELLGAIKEVL GDAATDDILD AWGKAYGVIA DVFIQVEADL YAQAVE

SEQ ID NO: 25 *Corynebacterium glutamicum ATCC 13032*
MTTSENFYDSVGGEETFSLIVHRFYEQVPNDDILGPMYPPDDFEGAEQRLKMFLSQYWGGPKD
YQEQRGHPRLRMRHVNYPIGVTAAERWLQLMSNALDGVDLTAEQREAIWEHMVRAADMLIN
SNPDPHA

SEQ ID NO: 26 *Bacillus subtilis*
MGQSFNAPYEAIGEELLSQLVDTFYERVASHPLLKPIFPSDLTETARKQKQFLTQYLGGPPLYTE
EHGHPMLRARHLPFPITNERADAWLSCMKDAMDHVGLEGEIREFLFGRLELTARHMVNQTEA
EDRSS

SEQ ID NO: 27 *Bacillus megaterium*
MREKIHSPYELLGGEHTISKLVDAFYTRVGQHPELAPIFPDNLTETARKQKQFLTQYLGGPSLY
TEEHGHPMLRARHLPFEITPSRAKAWLTCMHEAMDEINLEGPERDELYHRLILTAQHMINSPE
QTDEKGFSH

SEQ ID NO: 28: *Saccharomyces cerevisiae*
MLAEKTRSIIKATVPVLEQQGTVITRTFYKNMLTEHTELLNIFNRTNQKVGAQPNALATTVLAAAKNIDD
LSVLMDHVKQIGHKHRALQIKPEHYPIVGEYLLKAIKEVLGDAATPEIINAWGEAYQAIADIFITVEKKM
YEEALWPGWKPFDITAKEYVASDIVEFTVKPKFGSGIELESLPITPGQYITVNTHPIRQENQYDALRHYS
LCSASTKNGLRFAVKMEAARENFPAGLVSEYLHKDAKVGDEIKLSAPAGDFAINKELIHQNEVPLVLLSS
GVGVTPLLAMLEEQVKCNPNRPIYWIQSSYDEKTQAFKKHVDELLAECANVDKIIVHTDTEPLINAAFLK
EKSPAHADVYTCGSLAFMQAMIGHLKELEHRDDMIHYEPFGPKMSTVQV

SEQ ID NO: 29 *Nicotina tobaccum*
MSSFSEEQEALVLKSWDSMKKNAGEWGLKLFLKIFEIAPSAKKLFSFLKDSNVPLEQNAKLKPHAKSVFVMTCEAAVQ
LRKAGKVVVRDSTLKKLGAAHFKYGVADEHFEVTKFALLETIKEAVPDMWSVDMKNAWGEAFDQLVNAIKTEMK

SEQ ID NO: 30 *Medicago sativa*
mgtldtkgft eeqealvvks wnamkknsae lglklflkif eiapsaqklf sflkdskvpl eqntklkpha msvflmtces avqlrksgkv tvresslkkl ganhfkygvv dehfevtkfa lletikeavp emwspamkna wgeaydqlvn aiksemkpss

SEQ ID NO 31: *Glycine max*

MTTTLERGFSEEQEALVVKSWNVMKKNSGELGLKFFLKIFEIAPSAQKLFSFLRDSTVPLEQNPKLKPHAVSVFVMTCD
SAVQLRKAGKVTVRESNLKKLGATHFRTGVANEHFEVTKFALLETIKEAVPEMWSPAMKNAWGEAYDQLVDAIKSE
MKPPSS