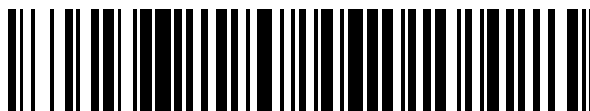


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 788 199**

51 Int. Cl.:

A61K 38/48 (2006.01)

C07K 14/33 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.10.2013** **E 17201198 (3)**

97 Fecha y número de publicación de la concesión europea: **04.03.2020** **EP 3326644**

54 Título: **Neurotoxinas recombinantes de *Clostridium botulinum***

30 Prioridad:

31.10.2012 GB 201219602

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
20.10.2020

73 Titular/es:

IPSEN BIOINNOVATION LIMITED (50.0%)
102 Park Drive, Milton Park
Abingdon, Oxfordshire OX14 4RY, GB y
IPSEN BIOPHARM LIMITED (50.0%)

72 Inventor/es:

COSSINS, AIMEE;
MARKS, PHILIP y
BEARD, MATTHEW

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 788 199 T3

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

DESCRIPCIÓN

Neurotoxinas recombinantes de *Clostridium botulinum*

Esta solicitud de patente reivindica la prioridad de GB 1219602.8 presentada el 31 de octubre de 2012.

La presente invención se refiere a secuencias de ácidos nucleicos que codifican neurotoxinas de *Clostridium botulinum* (*C. botulinum*) del serotipo E (BoNT/E) y a métodos para producir BoNT/E recombinante. La presente descripción también se refiere a los usos médicos correspondientes de una BoNT/E recombinante.

La neurotoxina botulínica es producida por *C. botulinum* en la forma de un gran complejo proteico, que consiste en BoNT complejada con una serie de proteínas accesorias. En la actualidad hay siete clases diferentes de neurotoxina botulínica, a saber: serotipos de neurotoxina botulínica A, B, C₁, D, E, F y G, todos los cuales comparten estructuras y modos de acción similares. Se pueden distinguir diferentes serotipos de BoNT basándose en la inactivación por antisueros neutralizantes específicos, correlacionándose dicha clasificación por serotipo con el porcentaje de identidad de secuencia a nivel de aminoácidos. Las proteínas BoNT de un serotipo dado se dividen además en diferentes subtipos basándose en el porcentaje de identidad de secuencia de aminoácidos.

Las BoNT son las toxinas más potentes conocidas, con valores medios de dosis letal (DL₅₀) para ratones que varían de 0,5 a 5 ng/kg dependiendo del serotipo. Las BoNT se adsorben en el tracto gastrointestinal y, después de entrar en la circulación general, se unen a la membrana presináptica de los terminales nerviosos colinérgicos y previenen la liberación de su neurotransmisor acetilcolina. BoNT/B, BoNT/D, BoNT/F y BoNT/G escinden la proteína de la membrana asociada a sinaptobrevina/vesícula (VAMP); BoNT/C, BoNT/A y BoNT/E escinden la proteína sinaptosómica asociada de 25 kDa (SNAP-25); y BoNT/C escinde la sintaxina.

En la naturaleza, las neurotoxinas clostridiales se sintetizan como un polipéptido de cadena sencilla que se modifica postraduccion mediante un evento de escisión proteolítica para formar dos cadenas polipeptídicas unidas por un enlace disulfuro. La escisión se produce en un sitio de escisión específico, a menudo denominado sitio de activación, que se encuentra entre los restos de cisteína que proporcionan el enlace disulfuro entre cadenas. Es esta forma de doble cadena la que es la forma activa de la toxina. Las dos cadenas se denominan cadena pesada (cadena H), que tiene una masa molecular de aproximadamente 100 kDa, y cadena ligera (cadena L), que tiene una masa molecular de aproximadamente 50 kDa. La cadena H comprende un componente de direccionamiento en el extremo terminal C (dominio H_C) y un componente de translocación en el extremo terminal N (dominio H_N). El sitio de escisión está ubicado entre la cadena L y los componentes de translocación. Después de la unión del dominio H_C a su neurona objetivo y la internalización de la toxina unida a la célula a través de un endosoma, el dominio H_N transloca la cadena L a través de la membrana endosomal y al citosol, y la cadena L proporciona una función proteasa (también conocida como proteasa no citotóxica).

Las proteasas no citotóxicas actúan escindiendo proteolíticamente proteínas de transporte intracelular conocidas como proteínas SNARE (por ejemplo, SNAP-25, VAMP o Syntaxin) -véase Gerald K (2002) "Cell and Molecular Biology" (4^a edición) John Wiley & Sons, Inc-. El acrónimo SNARE se deriva del nombre en inglés del receptor de unión al NSF soluble, donde NSF significa factor sensible a N-etilmaleimida. Las proteínas SNARE son esenciales para la fusión de vesículas intracelulares y, por lo tanto, para la secreción de moléculas a través del transporte de vesículas desde una célula. La función de proteasa es una actividad endopeptidasa dependiente de zinc y exhibe una alta especificidad de sustrato para las proteínas SNARE. En consecuencia, una vez administrada a una célula objetivo deseada, la proteasa no citotóxica es capaz de inhibir la secreción celular de la célula objetivo. Las proteasas de la cadena L de las neurotoxinas clostridiales son proteasas no citotóxicas que escinden las proteínas SNARE.

Las neurotoxinas botulínicas son bien conocidas por su capacidad de causar una parálisis muscular flácida. Dichas propiedades relajantes musculares han llevado a que las neurotoxinas botulínicas (tales como BoNT/A) se empleen en una variedad de procedimientos médicos y cosméticos, incluido el tratamiento de líneas glabellares o líneas faciales hiperkinéticas, dolor de cabeza, espasmo hemifacial, hiperactividad de la vejiga, hiperhidrosis, líneas labiales nasales, distonía cervical, blefaroespasma y espasticidad.

Tradicionalmente, la producción de BoNT se lleva a cabo mediante cultivo de bacterias de *C. botulinum*, seguido del aislamiento y la purificación del complejo de la neurotoxina botulínica. Sin embargo, la producción de BoNT de esta manera es ineficiente y proporciona bajos rendimientos de proteína. Además, las *C. botulinum* son bacterias formadoras de esporas y, por lo tanto, requieren equipos e instalaciones de cultivo especializados, que no son necesarios para el cultivo de bacterias tales como *Escherichia coli* (*E. coli*). El uso cada vez mayor de BoNT ha conducido a la necesidad de métodos alternativos y/o mejorados para producir y purificar BoNT.

El documento US 20080103098 describe un método para producir proteínas BoNT recombinantes en una forma de cadena doble que comprende la expresión de una construcción de ácido nucleico recombinante en una célula huésped de *E. coli*. Sin embargo, dicho método requiere la inserción de una secuencia específica de pentapéptido no nativa (es decir, no clostridial) en un dominio de bucle de la neurotoxina. La secuencia insertada del pentapéptido forma un sitio de escisión de activación que se escinde mediante una proteasa endógena de *E. coli* tras la lisis celular. El método del documento US 20080103098 enseña por lo tanto que para lograr una expresión óptima de

BoNT, la secuencia de BoNT debe modificarse mediante la inserción de un sitio de escisión no nativo.

El documento US 7132259 describe moléculas de ácido nucleico recombinante que codifican proteínas BoNT. Sin embargo, las moléculas de ácido nucleico del documento US 7132259 se modifican para reemplazar el sitio de escisión nativo con un sitio de escisión no nativo. Por lo tanto, el método del documento US 7132259 también enseña que se requiere la inserción de un sitio de escisión no nativo para la expresión óptima de BoNT.

El documento US 6495143 describe moléculas de ácido nucleico recombinante que codifican fragmentos de la cadena pesada (Hc) de una BoNT, para usar en la inducción de respuestas inmunitarias (tales como en la vacunación). Sin embargo, las moléculas de ácido nucleico no codifican secuencias de BoNT de longitud completa. Se puede lograr la expresión en *E. coli* y la purificación de las cadenas H y L individuales de la toxina tetánica y la BoNT; estas cadenas aisladas son, por sí mismas, no tóxicas. Después de la producción separada de estas cadenas peptídicas y en condiciones estrictamente controladas, las subunidades H y L pueden combinarse mediante un enlace disulfuro oxidativo para formar dos cadenas activas. Desafortunadamente, esta estrategia tiene varios inconvenientes. En primer lugar, no es práctico expresar y aislar grandes cantidades de las cadenas individuales; en particular, en ausencia de la cadena H, la cadena L aislada es bastante insoluble en solución acuosa y es altamente susceptible a la degradación proteolítica. En segundo lugar, la oxidación *in vitro* de las cadenas H y L expresadas y purificadas individualmente para producir la cadena doble activa es muy ineficiente y conduce a bajos rendimientos de la toxina activa y a la producción de muchas formas inactivas incorrectamente plegadas u oxidadas. La purificación de la toxina que contiene la cadena H y L correctamente plegada y oxidada es difícil, así como su separación de estas formas inactivas y las cadenas H y L separadas que no han reaccionado. Por lo tanto, el método del documento US 6495143 está asociado con considerables desventajas.

Gilmore et al. (2008) Toxicon, Elmsford, NY, 51 (1): 11-12 se refiere a una proteína BoNT/E marcada con His. El documento WO 2008/008805 se refiere a toxinas clostridiales que comprenden un dominio de direccionamiento alterado. Fischer et al. (2008) J. Biol. Chem. 283 (7): 3997-4003 describe la arquitectura molecular de BoNT/A y BoNT/E.

Por lo tanto, existe una necesidad en la técnica por métodos mejorados para producir BoNT recombinantes, en particular BoNT/E recombinante activada por BoNT de cadena doble.

La presente invención resuelve uno o más de los problemas mencionados anteriormente, proporcionando secuencias de ácido nucleico y métodos como se especifica en las reivindicaciones.

La presente divulgación proporciona una secuencia de ácidos nucleicos que comprende una secuencia de nucleótidos contiguos, donde dicha secuencia de nucleótidos contiguos tiene al menos 80% (por ejemplo, al menos 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99,5, 99,9 o 100%) de identidad de secuencia con la secuencia de ácidos nucleicos de la SEQ ID NO: 1, y en el que dicha secuencia de nucleótidos contiguos codifica una proteína BoNT/E1 de cadena sencilla. De acuerdo con la invención, dicha secuencia de ácidos nucleicos tiene al menos un 90% de identidad con la secuencia de ácidos nucleicos de la SEQ ID NO: 1.

La presente invención proporciona además un método para producir proteína BoNT/E1 de cadena sencilla soluble en una célula huésped de *E. coli*, comprendiendo dicho método: expresar una secuencia de ácidos nucleicos de la invención en un sistema de expresión de *E. coli*.

La invención también proporciona una célula huésped o vector de expresión que comprende una secuencia de ácidos nucleicos de la invención.

El serotipo BoNT/E se divide en ocho subtipos, BoNT/E1 a BoNT/E8, que comparten al menos 90% de identidad de secuencia de aminoácidos; las proteínas BoNT/E dentro de un subtipo dado comparten un mayor porcentaje de identidad de secuencia de aminoácidos (por ejemplo, al menos 95% o más). Como se describió anteriormente, las secuencias de ácidos nucleicos de la invención codifican una proteína BoNT/E1. Un ejemplo de una proteína BoNT/E1 es la proteína codificada por la secuencia de aminoácidos de UniParc UPI000016EA7F. Otro ejemplo de una proteína BoNT/E1 es la proteína codificada por la secuencia de aminoácidos de la SEQ ID NO: 2.

Las secuencias de ácidos nucleicos de la presente invención se han diseñado para proporcionar ventajosamente altos niveles de expresión en células de *E. coli*.

Varios factores influyen en los niveles de expresión de una proteína determinada. Uno de tales factores es la velocidad a la que se traduce la secuencia de ARNm que codifica esa proteína. Este factor en sí mismo se ve afectado por los codones particulares que utiliza el ARNm para especificar cada aminoácido de la proteína. Algunos codones se traducen más rápidamente que otros. La elección del codón para cada aminoácido puede variar debido a que los codones del ARNm son de naturaleza degenerada. Varios codones diferentes pueden todos especificar el mismo aminoácido; por lo tanto, varias secuencias de ARNm diferentes pueden codificar la misma proteína. Los diferentes codones que especifican el mismo aminoácido se llaman codones sinónimos. La mezcla precisa de codones sinónimos en un ARNm particular afecta la velocidad de traducción de la proteína codificada.

Hay una serie de razones diferentes que explican por qué algunos codones se traducen más rápidamente que otros.

Cada codón especifica un aminoácido al reclutar una molécula de ARNt unida a ese aminoácido. La velocidad de traducción se ve afectada por la abundancia relativa de varias moléculas diferentes de ARNt, por la afinidad con la que cada molécula particular de ARNt se une al codón que la recluta y también por otros factores tales como qué tan bien interactúa el par codón-molécula de ARNt con otros elementos de la maquinaria de traducción. Las velocidades aproximadas de traducción de codones pueden estimarse determinando la frecuencia a la que se encuentran diferentes codones en genes altamente expresados. Sin embargo, no todos los codones frecuentes producen una expresión óptima.

Sin desear estar ligados a ninguna teoría particular, los presentes inventores creen que la expresión óptima de las secuencias de ácidos nucleicos de BoNT/E1 se consigue reduciendo la frecuencia (es decir, el número de apariciones en una secuencia) de ciertos codones, en lo sucesivo denominados "codones lentos" y se exponen a continuación. A este respecto, los presentes inventores creen que dichos codones lentos están asociados con velocidades de traducción reducidas.

Aminoácido	Codón lento (ARN)	Codón lento (equivalente de ADN)
Fenilalanina	UUU	TTT
Tirosina	UAU	TAT
Cisteína	UGU	TGT
Histidina	CAU	CAT
Glutamina	CAA	CAA
Prolina	CCA y/o CCG	CCA y/o CCG
Serina	UCA y/o UCG	TCA y/o TCG
Arginina	CGG	CGG
Leucina	UUA y/o CUA	TTA y/o CTA

Los presentes inventores han empleado un proceso de diseño de secuencia racional para producir las secuencias de ácidos nucleicos de la invención. Una forma en que las secuencias de ácidos nucleicos de la invención proporcionan altos niveles de expresión de las proteínas BoNT/E1 codificadas es teniendo un número optimizado de codones lentos (por ejemplo, una reducción en la frecuencia a la que aparecen los codones lentos en la secuencia).

En una realización, la secuencia de ácidos nucleicos tiene un máximo de 160 codones lentos (por ejemplo, un máximo de 160, 150, 140, 130, 120, 110, 100, 90, 95, 94, 93, 92, 91, 90, 89, 88 u 87 codones lentos) como se define en el presente documento.

Por lo tanto, en una realización, la secuencia de ácidos nucleicos tiene entre 0 y 160 codones lentos (por ejemplo, 0-160, 0-150, 0-140, 0-130, 0-120, 0-110, 0-100, 0-90, 0-95, 0-94, 0-93, 0-92, 0-91, 0-90, 0-89, 0-88 o 0-87 codones lentos).

En una realización, la secuencia de ácidos nucleicos tiene 60-160 codones lentos (por ejemplo, 60-160, 60-150, 60-140, 70-150, 70-140, 70-130, 70-120, 70-110, 70-100, 70-90, 80-130, 80-120, 80-110, 80-100 u 80-90 codones lentos).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, hay menos codones lentos en el primer 50% de la secuencia de ácidos nucleicos que en el segundo 50% de la secuencia de ácidos nucleicos. El primer 50% de la secuencia de ácidos nucleicos se define con referencia a la posición de nucleótido número 1 como punto de partida, y por lo tanto comprende el sitio de inicio de la traducción; el segundo 50% de la secuencia de ácidos nucleicos comprende el sitio de terminación de la traducción. A modo de ejemplo, con referencia a la SEQ ID NO: 1 (que tiene una longitud total de 3.759 nucleótidos), la primera mitad de dicha secuencia puede representarse mediante las posiciones de los nucleótidos 1-1.881 (que comprende 627 tripletes de nucleótidos), y la segunda mitad de dicha secuencia puede estar representada por las posiciones 1.882-3.759 (que comprenden 626 tripletes de nucleótidos); alternatively, la primera mitad de dicha secuencia puede estar representada por las posiciones de los nucleótidos 1-1.878 (que comprende 626), y la segunda mitad de dicha secuencia puede estar representada por las posiciones 1879-3759 (que comprenden 627 tripletes de nucleótidos).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 30 (por ejemplo, 30, 25,

20, 15 o 10) codones lentos de fenilalanina (ARN = UUU, ADN = TTT). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 10 codones lentos de fenilalanina (ARN = UUU; ADN = TTT).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 30 (por ejemplo, 30, 25, 20, 19 o 18) codones lentos de tirosina (ARN = UAU; ADN = TAT). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas anteriormente relacionadas con codones lentos de fenilalanina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 18 codones lentos de tirosina (ARN = UAU, ADN = TAT).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina y/o tirosina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 19 (por ejemplo, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6 o 5) codones lentos de leucina (ARN = UUA y/o CUA, ADN = TTA y/o CTA). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina y/o tirosina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 5 codones lentos de leucina (ARN = UUA y/o CUA; ADN = TTA y/o CTA).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina, tirosina y/o leucina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 14 (por ejemplo, 14, 12, 10, 8, 6, 5, 4 o 3) codones lentos de glutamina (ARN = CAA, ADN = CAA). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (incluidas las realizaciones previamente descritas relacionadas con codones lentos de fenilalanina, tirosina y/o leucina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 3 codones lentos de glutamina (ARN = CAA; ADN = CAA).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina y/o glutamina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 20 (por ejemplo, 20, 19, 18, 17 o 16) codones lentos de serina (ARN = UCA y/o UCG, ADN = TCA y/o TCG). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina y/o glutamina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 16 codones lentos de serina (ARN = UCA y/o UCG, ADN = TCA y/o TCG).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina, tirosina, leucina, glutamina y/o serina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 23 (por ejemplo, 23, 22, 21, 20 o 19) codones lentos de prolina (ARN = CCA y/o CCG, ADN = CCA y/o CCG). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina, tirosina, leucina, glutamina y/o serina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 19 codones lentos de prolina (ARN = CCA y/o CCG, ADN = CCA y/o CCG).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con los codones lentos de fenilalanina, tirosina, leucina, glutamina, serina y/o prolina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 3 (por ejemplo 3 o 2) codones lentos de cisteína (ARN = UGU; ADN = TGT). En una realización, opcionalmente en combinación con una o más de las realizaciones anteriores (incluyendo las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina, glutamina, serina y/o prolina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 2 codones lentos de cisteína (ARN = UGU; ADN = TGT).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina, glutamina, serina, prolina y/o cisteína), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 5 (por ejemplo 5, o 4) codones lentos de histidina (ARN = CAU; ADN = CAT). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina, glutamina, serina, prolina y/o cisteína), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 4 codones lentos de histidina (ARN = CAU; ADN = CAT).

En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que

- 5 incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina, glutamina, serina, prolina, cisteína y/o histidina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende de 5 a 10 (por ejemplo, 5, 6, 7, 8, 9 o 10) codones lentos de arginina (ARN = CGG, ADN = CGG). En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores (que incluyen las realizaciones descritas previamente que se relacionan con codones lentos de fenilalanina, tirosina, leucina, glutamina, serina, prolina, cisteína y/o histidina), la secuencia de ácidos nucleicos (como se describió anteriormente) comprende 10 codones lentos de arginina (ARN = CGG, ADN = CGG).
- 10 En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 30 (por ejemplo, 30, 25, 20, 15 o 10; preferiblemente 10) codones lentos de fenilalanina (ARN = UUU; ADN = TTT) y un máximo de 30 (por ejemplo, 30, 25, 20, 19 o 18, preferiblemente 18) codones lentos de tirosina (ARN = UAU; ADN = TAT).
- 15 En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 30 (por ejemplo, 30, 25, 20, 15 o 10; preferiblemente 10) codones lentos de fenilalanina (ARN = UUU; ADN = TTT), un máximo de 30 (por ejemplo, 30, 25, 20, 19 o 18, preferiblemente 18) codones lentos de tirosina (ARN = UAU, ADN = TAT) y un máximo de 19 (por ejemplo, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6 o 5, preferiblemente 5) codones lentos de leucina (ARN = UUA y/o CUA; ADN = TTA y/o CTA).
- 20 En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende un máximo de 30 (por ejemplo, 30, 25, 20, 15 o 10; preferiblemente 10) codones lentos de fenilalanina (ARN = UUU; ADN = TTT), un máximo de 30 (por ejemplo, 30, 25, 20, 19 o 18, preferiblemente 18) codones lentos de tirosina (ARN = UAU; ADN = TAT), un máximo de 19 (por ejemplo, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6 o 5, preferiblemente 5) codones lentos de leucina (ARN = UUA y/o CUA; ADN = TTA y/o CTA), y un máximo de 14 (por ejemplo, 14, 12, 10, 8, 6, 5, 4 o 3, preferiblemente 3) codones lentos de glutamina (ARN = CAA; ADN = CAA).
- 25 En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende:
- un máximo de 10 codones lentos de fenilalanina;
 - un máximo de 18 codones lentos de tirosina;
 - un máximo de 2 codones lentos de cisteína;
 - 30 un máximo de 4 codones lentos de histidina;
 - un máximo de 3 codones lentos de glutamina;
 - un máximo de 19 codones lentos de prolina;
 - un máximo de 16 codones lentos de serina; y
 - un máximo de 5 codones lentos de leucina.
- 35 En una realización, opcionalmente en combinación con una cualquiera o más de las realizaciones anteriores, la secuencia de ácidos nucleicos (como se describió anteriormente) comprende:
- un máximo de 10 codones lentos de fenilalanina;
 - un máximo de 18 codones lentos de tirosina;
 - un máximo de 2 codones lentos de cisteína;
 - 40 un máximo de 4 codones lentos de histidina;
 - un máximo de 3 codones lentos de glutamina;
 - un máximo de 19 codones lentos de prolina;
 - un máximo de 16 codones lentos de serina;
 - un máximo de 5 codones lentos de leucina; y
 - 45 un máximo de 10 codones lentos de arginina.
- En una realización, en la que la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos como se describió anteriormente, dicha proteína BoNT/E1 de cadena sencilla comprende una secuencia de aminoácidos

contiguos, y en la que dicha secuencia de aminoácidos contiguos tiene al menos 95% (por ejemplo, al menos 95, 96, 97, 98, 99, 99,1; 99,2; 99,3; 99,4; 99,5; 99,6; 99,7; 99,8; 99,9; o 100%) de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 2.

5 En una realización, en la que la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos como se describió anteriormente, dicha proteína BoNT/E1 de cadena sencilla comprende un sitio de activación nativo proporcionado por una secuencia de aminoácidos seleccionada de: KGIRK, VKGIRKS, SVKGIRKSI, VSVKGIRKSI, IVSVKGIRKSI, NIVSVKGIRKSI, KNIVSVKGIRKSI, CKNIVSVKGIRKSI, KNIVSVKGIRKSIC y CKNIVSVKGIRKSIC.

10 En una realización, la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos como se describió anteriormente, con la condición de que la BoNT/E1 de cadena sencilla como se describió anteriormente o la secuencia de aminoácidos contiguos como se describió anteriormente incluye uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más u ocho) de los siguientes aminoácidos (en donde la numeración de posición del aminoácido comienza con el resto del aminoácido del extremo terminal N y termina con el resto de aminoácido del extremo terminal C de la proteína BoNT/E1):

15 glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

En una realización, dichos uno o más aminoácidos comprenden (o consisten en) glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; y asparagina en la posición 1.195.

20 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) glicina en la posición 177; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; y asparagina en la posición 1.195.

25 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) glicina en la posición 177, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

En una realización, dichos uno o más aminoácidos comprenden (o consisten en) serina en la posición 198, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; y asparagina en la posición 1.195.

30 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) alanina en la posición 340, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

35 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) leucina en la posición 773, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

40 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) leucina en la posición 963, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

45 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) glutamina en la posición 964, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; alanina en la posición 967; asparagina en la posición 1.195.

En una realización, dichos uno o más aminoácidos comprenden (o consisten en) alanina en la posición 967, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; asparagina en la posición 1.195.

50 En una realización, dichos uno o más aminoácidos comprenden (o consisten en) asparagina en la posición 1.195, y uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de: glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967.

En una realización, la presencia de dichos uno o más aminoácidos, como se describió anteriormente, proporciona

una proteína BoNT/E1 que tiene una solubilidad mejorada en comparación con una proteína BoNT/E1 que carece de dichos aminoácidos. Dicha solubilidad mejorada aumenta el rendimiento de la proteína en un sistema de expresión heterólogo (*E. coli*).

En una realización, cuando la secuencia de ácidos nucleicos es una secuencia de ácidos nucleicos como se describió anteriormente, la secuencia de nucleótidos contiguos tiene al menos 770 (por ejemplo, al menos 770, 775, 780, 785, 790, 795, 800, 810, 820, 830, 840, 850, 860, 870 u 880) codones sinónimos cuando se compara con la secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre (SEQ ID NO: 3). Por lo tanto, en una realización, la secuencia de ácidos nucleicos comprende al menos 770 codones que difieren, pero codifican el mismo aminoácido que el codón correspondiente en la secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre (SEQ ID NO: 3). Preferiblemente, la secuencia de ácidos nucleicos comprende al menos 785 codones sinónimos en comparación con la secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre (SEQ ID NO: 3).

En una realización, la secuencia de ácidos nucleicos (como se describió anteriormente) tiene un contenido de G-C de al menos 41% (por ejemplo, al menos 41 o 42%). En una realización, la secuencia de ácidos nucleicos (como se describió anteriormente) tiene un contenido de G-C del 42%. El concepto de contenido de G-C en el ácido nucleico (también conocido como contenido de GC o contenido de G + C) se refiere a la proporción de nucleótidos de una secuencia dada de ácido nucleico que son G (guanina) o C (citocina). Por lo tanto, en una realización, el contenido de G-C de una secuencia de ácidos nucleicos de la invención se altera (por ejemplo, mediante la sustitución de codones sinónimos) para coincidir más estrechamente con el contenido de GC de ácidos nucleicos expresados preferentemente en células huésped de *E. coli*, mejorando así la expresión de la secuencia y proporcionando mayores rendimientos de proteína.

En un aspecto, la invención proporciona un vector de expresión que codifica una secuencia de ácidos nucleicos como se describió anteriormente. En una realización, el vector de expresión es un vector pET-26b(+).

En un aspecto, la invención proporciona una célula huésped que comprende una secuencia de ácidos nucleicos como se describió anteriormente, o un vector de expresión como se describió anteriormente. En una realización, la célula huésped es una célula de *E. coli*. En una realización, la célula huésped de *E. coli* es una célula BLR de *E. coli* (DE3).

En un aspecto, la invención proporciona un método para producir proteína BoNT/E1 de cadena sencilla soluble en una célula huésped de *E. coli*, comprendiendo dicho método: expresar una secuencia de ácidos nucleicos (como se describió anteriormente) en un sistema de expresión de *E. coli*.

Los métodos y técnicas usados para expresar proteínas heterólogas en sistemas de expresión de *E. coli* (*Escherichia coli*) son bien conocidos en la técnica.

En una realización, dicha proteína BoNT/E1 de cadena sencilla soluble se expresa en el citoplasma de dicha célula huésped de *E. coli*.

En una realización, dicha proteína BoNT/E1 de cadena sencilla soluble se expresa a un nivel de al menos 3 mg/L (por ejemplo, al menos 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45 o 50 mg/L).

En una realización, el método para producir proteína BoNT/E1 de cadena sencilla soluble, como se describió anteriormente, comprende la lisis de la célula huésped de *E. coli* para proporcionar un homogeneizado de células huéspedes de *E. coli* que contiene dicha proteína BoNT/E1 de cadena sencilla soluble. Los métodos y técnicas usados para lisar células huésped, tales como células huésped de *E. coli*, son conocidos en la técnica. Los ejemplos incluyen sonicación y el uso de una prensa francesa.

En un aspecto, la divulgación proporciona un método para producir proteína BoNT/E1 de doble cadena soluble, comprendiendo dicho método: proporcionar una proteína BoNT/E1 de cadena sencilla soluble que comprende una secuencia de aminoácidos contiguos, y en el que dicha secuencia de aminoácidos contiguos tiene al menos 95% (por ejemplo, al menos 95, 96, 97, 98, 99, 99,1; 99,2; 99,3; 99,4; 99,5; 99,6; 99,7; 99,8; 99,9 o 100%) de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 2, y poner en contacto dicha proteína BoNT/E1 con tripsina en solución.

Cuando la proteína BoNT/E1 de cadena sencilla de la divulgación se pone en contacto con tripsina, la acción proteolítica de la tripsina escinde la proteína de cadena sencilla en un sitio entre el componente de la proteasa de la cadena L y el componente de translocación para producir una proteína de doble cadena, donde las dos cadenas están unidas por un puente disulfuro (en forma más detallada, las dos cadenas formadas después de la escisión de BoNT/E1 de cadena sencilla en el sitio de activación son una primera cadena de restos de aminoácidos 1-419 y una segunda cadena de restos de aminoácidos 423-1.252, con los restos 420, 421 y 422 eliminados por el evento de escisión). Por lo tanto, la tripsina puede usarse para activar el polipéptido de cadena sencilla convirtiéndolo en la forma de doble cadena activa. De este modo, ventajosamente, el uso de tripsina significa que no es necesario diseñar un sitio de escisión exógeno (no nativo) en una BoNT/E1 de la divulgación.

En un aspecto, la referencia a tripsina abarca enzimas de tipo tripsina que escinden en el mismo sitio de escisión de proteasa que la tripsina.

La tripsina escinde secuencias de proteína en las que determinados aminoácidos se encuentran en determinadas posiciones a ambos lados del enlace peptídico escindido. Tales secuencias se pueden representar mediante la nomenclatura P4-P3-P2-P1-enlace escindido-P'1-P'2-P'3-P'4; en la que P1 a P4 designan aminoácidos posicionados en las posiciones 1 a 4 en el lado del extremo terminal N del enlace peptídico escindido respectivamente y P'1 a P'4 designan a las posiciones 1 a 4 en el lado del extremo terminal C del enlace peptídico escindido respectivamente.

Lo más importante es que la tripsina escinde secuencias de proteína donde los aminoácidos Arg o Lys ocupan la posición P1. Cuando Lys está en la posición P1, hay tres tipos principales de secuencia que no son sensibles a la tripsina:

(1) Pro en la posición P'1 generalmente reduce la susceptibilidad a la escisión por tripsina (pero no cuando Trp está en la posición P2).

(2) O bien Cys o Asp en la posición P2 junto con Asp en la posición P'1 reduce la susceptibilidad a la escisión por tripsina.

(3) Cys en la posición P2 junto con His o Try en la posición P'1 reduce la susceptibilidad a la escisión por tripsina.

Cuando Arg está en la posición P1 también hay tres tipos principales de secuencia que no son sensibles a la tripsina:

(1) Pro en la posición P'1 generalmente reduce la susceptibilidad a la escisión por la tripsina (pero no cuando Met, o posiblemente Glu, está en la posición P2).

(2) Cys en la posición P2 junto con Lys en la posición P'1 reduce la susceptibilidad a la escisión por tripsina.

(3) Arg en la posición P2 junto con His o Arg en la posición P'1 reduce la susceptibilidad a la escisión por tripsina.

En un aspecto, la divulgación proporciona un método (como se describió anteriormente) para producir proteína BoNT/E1 de doble cadena soluble, con la condición de que dicha secuencia de aminoácidos contiguos incluye uno o más (por ejemplo, uno o más, dos o más, tres o más, cuatro o más, cinco o más, seis o más, o siete) de los siguientes aminoácidos (en donde la numeración de posición de aminoácido comienza con el resto de aminoácido del extremo terminal N y termina con el resto de aminoácido del extremo terminal C de la proteína BoNT/E1): glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

En un aspecto, la presencia de dichos uno o más aminoácidos, como se describió anteriormente (y con referencia a las permutaciones múltiples de dichos uno o más aminoácidos como se describió anteriormente), proporciona una proteína BoNT/E1 que tiene una solubilidad mejorada en comparación con un proteína BoNT/E1 que carece de dichos aminoácidos. Dicha solubilidad mejorada puede aumentar el rendimiento de la proteína en un sistema de expresión heterólogo.

En un aspecto, en la que la divulgación proporciona un método (como se describió anteriormente) para producir proteína BoNT/E1 de cadena sencilla soluble, la proteína BoNT/E1 de cadena sencilla soluble se proporciona mediante un método como se describió anteriormente para producir proteína BoNT/E1 de cadena sencilla soluble en una célula huésped de *E. coli*.

En un aspecto en el que la divulgación proporciona un método (como se describe anteriormente) para producir proteína BoNT/E1 soluble de cadena doble, el método comprende separar la proteína BoNT/E1 soluble de la tripsina poniendo en contacto la solución que contiene la proteína BoNT/E1 soluble y la tripsina con una superficie hidrófoba, donde la proteína BoNT/E1 soluble se une preferentemente a la superficie hidrófoba.

Los presentes inventores han descubierto que pueden obtenerse altos rendimientos de la proteína BoNT/E1 de doble cadena activada usando un proceso de purificación hidrófoba para separar el polipéptido de doble cadena activado de la tripsina. Sorprendentemente, este proceso proporciona una purificación superior a la purificación estándar usando cromatografía de intercambio iónico, que los presentes inventores han encontrado que es ineficaz para separar el polipéptido de doble cadena activado de la tripsina. Además, el proceso proporciona ventajosamente una proteína BoNT/E1 de doble cadena activada que está libre de la proteasa activadora, como parte de un proceso de purificación general.

La producción de BoNT/E1 recombinante activa requiere una etapa proteolítica que escinde la molécula en la forma de doble cadena activa. Esta escisión se puede lograr mediante una etapa de activación *in vitro* usando la proteasa, tripsina. Después de la etapa de activación, es importante eliminar la proteasa del producto final, lo que también evita cualquier división adicional no específica de BoNT/E1.

Los puntos isoeléctricos (pI) de tripsina y BoNT/E1 son 9,32 y 6,2 respectivamente, lo que indica que la separación de las dos proteínas se debe lograr mediante cromatografía de intercambio iónico (IEX), que explota la diferencia de carga entre las dos moléculas. La carga neta de una proteína se ve afectada por el pH del entorno que la rodea y se volverá más positiva o negativamente cargada, dependiendo de si gana o pierde protones. El pI es el valor de pH al

cual una molécula no porta carga eléctrica y por lo tanto no interactuará con un medio IEX cargado. Esto significa que, si una proteína está a un pH por encima de su pI, entonces tendrá una carga neta negativa y se unirá a un medio con carga positiva, como un intercambiador de aniones. Del mismo modo, si el pH del regulador está por debajo del pI, entonces la proteína tendrá una carga neta positiva y no se unirá a un intercambiador de aniones.

5 Con base en este principio a pH 8, se esperaría que BoNT/E (que tiene un pI de 6,2) se unirá a una columna de intercambio aniónico, mientras que la tripsina con un pI de 9,32 no lo haría, permitiendo que las dos proteínas se separen. El IEX es un método de cromatografía simple y económico, ya que no requiere que la proteína cargada en la columna esté en un regulador con alto contenido de sal, lo que puede provocar pérdidas de proteína por precipitación.

10 Los presentes inventores han probado una variedad de columnas de intercambio aniónico, utilizando grupos funcionales fuertes y débiles unidos a perlas de agarosa entrecruzada, a pH 8. En cada caso, se encontró que una gran proporción de tripsina no se unía a la columna como se había predicho y estaba presente en el flujo continuo. Sin embargo, cuando las columnas se eluyeron con un gradiente lineal de fuerza iónica creciente, la tripsina se eluyó de la columna indicando que una proporción de la tripsina era capaz de unirse a las columnas. Cuando se comparó con la elución de BoNT/E1, se descubrió que, inesperadamente, la tripsina se eluía con una fuerza iónica similar (Tabla 1; Figura 1) lo que indica que la tripsina no se separó como se predijo y estaría presente en el producto BoNT/E1 purificado final con la posibilidad adicional de una mayor degradación de BoNT/E1.

Tabla 1: Fracciones de elución de columnas de intercambio aniónico en las que se evaluó la separación de tripsina de BoNT/E1. Los picos se indican en el número de volúmenes de columna (CV) F/T: Flujo que pasa a través de la columna, FF: resina de flujo rápido.

Columna	Tripsina		BoNT/E1	
	Pico principal	Pico menor	Pico principal	Pico menor
ANX	F/T	8,8; 11,3; 12,3	10,7	17,3
QHP	F/T	9,0; 10,6	-	-
DEAE	F/T	10,5	9,8	13,2
Q FF	F/T	9,2; 10,9	16,1	10,7

Los presentes inventores han resuelto el problema anterior. En forma más detallada, los inventores han identificado sorprendentemente que la separación óptima de tripsina-BoNT/E1 se logra mediante el uso de una superficie de separación hidrófoba (por ejemplo, mediante cromatografía de interacción hidrófoba (HIC), que separa proteínas de acuerdo con su hidrofobicidad superficial utilizando una interacción reversible entre estas proteínas y la superficie hidrófoba de un medio de HIC).

En un aspecto, la superficie hidrófoba es una matriz inerte a la que está unido un ligando que consiste en grupos arilo o alquilo.

El término "arilo" se refiere a grupos aromáticos, por ejemplo, fenilo, naftilo, tienilo e indolilo.

30 El término "alquilo" se refiere a grupos alifáticos que incluyen grupos cíclicos de cadena lineal, cadena ramificada y combinaciones de los mismos. Un grupo alquilo puede tener de 1 a 12 átomos de carbono. Ejemplos de grupos alquilo incluyen, pero sin limitación, grupos tales como metilo, etilo, propilo (por ejemplo, n-propilo, isopropilo), butilo (por ejemplo, n-butilo, isobutilo, sec-butilo, t-butilo), pentilo, hexilo, heptilo y octilo.

En un aspecto, la superficie hidrófoba se selecciona del grupo que consiste en: ligandos de butilo, fenilo u octilo.

35 En un aspecto, la superficie hidrófoba comprende ligandos de butilo. En una realización, la superficie hidrófoba comprende ligandos de fenilo. En un aspecto, la superficie hidrófoba comprende ligandos de octilo.

Los presentes inventores han descubierto que se obtienen resultados particularmente preferibles para separar tripsina de BoNT/E con HIC usando resinas de cromatografía que contienen grupos alquilo o arilo, por ejemplo, ligandos de butilo, fenilo y octilo, acoplados a una matriz inerte, tal como agarosa entrecruzada o perlas de poliestireno (Tabla 2, Figura 2).

Tabla 2: Fracciones de elución a través de columnas de interacción hidrófobas comerciales en las que se evaluó la separación de tripsina de BoNT/E. Los picos se expresan en cantidad de volúmenes de columna (CV) F/T: Flujo continuo a través de la columna, FF: resina de flujo rápido, HP: resina de alto rendimiento, (LS): baja sustitución de grupos hidrófobos, (HS): alta sustitución de grupos hidrófobos.

Columna	Tripsina		BoNT/E	
	Pico principal	Pico menor	Pico principal	Pico menor
Fenil (HS) FF	23,3	-	32,2	-
Fenil (LS) FF	F/T	-	21,4	-
Fenil HP	F/T	16,1	24,4	-
Butil FF	F/T	16,8	23,7	-
Butil HP	18	lavado	27,6	-
Octil FF	F/T	-	27,1	-

En un aspecto, el proceso de purificación hidrófoba para separar la proteína BoNT/E1 de doble cadena activada de tripsina reduce la concentración de tripsina al menos 100 veces, al menos 150 veces, al menos 200 veces, al menos 250 veces, al menos 300 veces, al menos 350 veces, al menos 400 veces, al menos 450 veces o al menos 500 veces. En un aspecto preferido, el proceso de purificación hidrófoba para separar la proteína BoNT/E1 de doble cadena activada de tripsina reduce la concentración de tripsina al menos 350 veces.

En otro aspecto, la divulgación proporciona una proteína BoNT/E1 de doble cadena activa, en la que la primera cadena comprende una secuencia de aminoácidos contiguos, y en la que dicha secuencia de aminoácidos contiguos tiene al menos 95% (por ejemplo, al menos 95, 96, 97, 98, 99; 99,1; 99,2; 99,3; 99,4; 99,5; 99,6; 99,7; 99,8; 99,9 o 100%) de identidad de secuencia con la secuencia de aminoácidos de las posiciones 1-419 de la SEQ ID NO: 2; en la que la segunda cadena comprende una secuencia de aminoácidos contiguos, y en la que dicha secuencia de aminoácidos contiguos tiene al menos 95% (por ejemplo, al menos 95, 96, 97, 98, 99; 99,1; 99,2; 99,3; 99,4; 99,5; 99,6; 99,7; 99,8; 99,9 o 100%) de identidad de secuencia con la secuencia de aminoácidos de las posiciones 423-1.252 de la SEQ ID NO: 2; en la que las cadenas primera y segunda se unen entre sí mediante un enlace disulfuro entre la cisteína 412 en la primera cadena y la cisteína 426 en la segunda cadena; con la condición de que dicha secuencia de aminoácidos contiguos incluye uno o más (por ejemplo, dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, u ocho) de los siguientes aminoácidos (en el que la numeración de posición del aminoácido comienza con el resto de aminoácido del extremo terminal N y termina con el resto de aminoácido del extremo terminal C de la proteína BoNT/E1): glicina en la posición 177; serina en la posición 198; alanina en la posición 340; leucina en la posición 773; leucina en la posición 963; glutamina en la posición 964; alanina en la posición 967; asparagina en la posición 1.195.

En un aspecto relacionado, la divulgación proporciona una proteína BoNT/E1 de doble cadena activa que se puede obtener mediante un método (como se describió anteriormente) para producir proteína BoNT/E1 de doble cadena soluble.

En un aspecto, la divulgación proporciona una composición que comprende una proteína BoNT/E1 de doble cadena activa (como se describió anteriormente), en la que dicha composición está sustancialmente libre de tripsina.

Por lo tanto, la composición está, ventajosamente, libre sustancialmente de tripsina proteasa (utilizada para activar el polipéptido de cadena sencilla convirtiéndolo en la forma de doble cadena activa), evitando así la escisión inespecífica no deseada de la proteína BoNT/E1.

En un aspecto en el que la composición (como se describió anteriormente) está sustancialmente libre de tripsina, la composición contiene menos de 100 picogramos (pg) de tripsina por 100 nanogramos (ng) de proteína BoNT/E1; por ejemplo, menos de 50, 20, 10, 9, 8, 7, 6 o 5 pg de tripsina por 100 ng de proteína BoNT/E1. En un aspecto, la composición (como se describió anteriormente) contiene menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 5 pg de tripsina por 100 ng de la proteína BoNT/E1. En un aspecto preferido, la composición (como se describió anteriormente) contiene menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1.

Por lo tanto, en un aspecto, la expresión "sustancialmente libre de tripsina" significa menos de 100 pg de tripsina por 100 ng de proteína BoNT/E1; por ejemplo, menos de 50, 20, 10, 9, 8, 7, 6 o 5 pg de tripsina por 100 ng de proteína BoNT/E1, preferiblemente menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg tripsina por 100 ng de proteína BoNT/E1.

Los métodos para determinar la concentración de tripsina en una composición son conocidos en la técnica. A modo de ejemplo, la concentración de tripsina en una composición de la divulgación se puede determinar usando un ELISA tipo sándwich (ensayo por inmunoabsorción ligado a enzimas).

En un aspecto adicional, la presente divulgación proporciona una composición farmacéutica sólida o líquida que comprende:

- (a) una proteína BoNT/E1 de doble cadena activa como se describió anteriormente, y
- (b) un agente estabilizante.

5 En un aspecto, la composición (como se describió anteriormente) está sustancialmente libre de tripsina. En un aspecto, la composición contiene menos de 100 pg de tripsina por 100 ng de proteína BoNT/E1, por ejemplo, menos de 50, 20, 10, 9, 8, 7, 6 o 5 pg de tripsina por 100 ng de proteína BoNT/E1. En un aspecto, la composición contiene menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1.

10 Los agentes estabilizantes que se pueden utilizar en composiciones de acuerdo con la divulgación incluyen estabilizantes de proteína, tales como albúmina, en particular albúmina sérica humana (HSA) y estabilizadores no proteicos.

Los agentes estabilizantes no proteicos que se pueden utilizar en la composición de acuerdo con la divulgación incluyen tensioactivos, en particular tensioactivos no iónicos. Los ejemplos de tensioactivos no iónicos incluyen

15 polisorbato, tales como polisorbato 20 o polisorbato 80, y copolímeros en bloque tales como poloxámeros (es decir, copolímeros de polietileno y propilenglicol).

En un aspecto particular, la composición no comprende una proteína como agente estabilizante.

De acuerdo con un aspecto particular de la divulgación, la composición farmacéutica es una composición farmacéutica líquida que comprende:

- 20 (a) una proteína BoNT/E1 de doble cadena activa, como se describió anteriormente;
- (b) un agente estabilizante no proteico que es un tensioactivo; y
- (c) agua;

en el que dicha composición farmacéutica líquida no comprende un agente estabilizante de proteínas; y

25 en el que dicha composición farmacéutica líquida está sustancialmente libre de tripsina (por ejemplo, dicha composición farmacéutica líquida contiene menos de 100 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 5 pg de tripsina por 100 ng de proteína BoNT/E1; preferiblemente en el que dicha composición farmacéutica líquida contiene menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1.

30 En un aspecto, la proteína BoNT/E1 de doble cadena activa está presente en la composición (como se describió anteriormente) a una concentración de 1-100 ng/mL. En un aspecto, la proteína BoNT/E1 de doble cadena activa está presente en la composición (como se describió anteriormente) a una concentración de 5-50 ng/mL, por ejemplo, aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 ng/mL. En un aspecto preferido, la proteína BoNT/E1 de doble cadena activa está presente a una concentración de aproximadamente 20 ng/mL.

35 En un aspecto, el tensioactivo (como se describió anteriormente) es un polisorbato, tal como un polisorbato que tiene un grado medio de polimerización que varía de 20 a 100 unidades monoméricas, y puede ser, por ejemplo, polisorbato 80. En un aspecto preferido, el polisorbato se deriva de una fuente vegetal. La concentración del tensioactivo es preferiblemente inferior al 1% v/v, por ejemplo, de aproximadamente 0,005% a 0,02% v/v en el caso del polisorbato 80.

40 La composición farmacéutica de acuerdo con la divulgación también puede comprender un agente cristalino.

Por agente cristalino se entiende un agente que, entre otros, mantiene una estructura de torta mecánicamente resistente para un complejo de neurotoxina botulínica liofilizada (tipo A, B, C, D, E, F o G) o una neurotoxina botulínica de alta pureza (tipo A, B, C, D, E, F o G). Cuando se incluyen en formulaciones sólidas, los agentes

45 cristalinos también tienen un efecto de aumento de volumen. Los agentes cristalinos incluyen notablemente cloruro de sodio. La concentración de agente cristalino puede ser, por ejemplo, de 0,1 a 0,5 M, preferiblemente de 0,1 a 0,4 M, notablemente de aproximadamente 0,15 a 0,3 M.

La composición farmacéutica de acuerdo con la divulgación también puede comprender un regulador para mantener el pH a un nivel comprendido entre 5,5 y 7,5, o entre 6,0 y 7,0. El regulador puede ser cualquier regulador capaz de mantener el pH adecuado. Por ejemplo, el regulador para las composiciones de acuerdo con la divulgación se puede

50 elegir del grupo que consiste en succinato, fosfato disódico/ácido cítrico y un aminoácido tal como histidina. La concentración del regulador puede ser, por ejemplo, de 1 a 50 mM, preferiblemente de 5 a 20 mM, preferiblemente aproximadamente de 10 mM.

La composición farmacéutica de acuerdo con la divulgación también puede comprender un disacárido.

El disacárido usado en las composiciones de acuerdo con la divulgación se puede elegir del grupo que consiste en sacarosa, trehalosa, manitol y lactosa. En un aspecto específico, el disacárido es sacarosa. La concentración del disacárido puede ser, por ejemplo, de 5 a 50 mM, preferiblemente de 5 a 25 mM, más preferiblemente de 10 a 20 mM, y lo más preferiblemente de aproximadamente 11,7 mM.

En un aspecto particular, la composición farmacéutica es una composición farmacéutica líquida que comprende:

(a) una proteína BoNT/E1 de doble cadena activa, como se describió anteriormente,

(b) un agente estabilizante no proteico que es un tensioactivo,

(c) cloruro de sodio,

(c) un regulador para mantener el pH entre 5,5 y 7,5,

(e) un disacárido, y

(f) agua estéril,

en la que dicha composición farmacéutica líquida no comprende un agente estabilizante de proteína; y

en la que dicha composición farmacéutica líquida está sustancialmente libre de tripsina (por ejemplo, dicha composición farmacéutica líquida contiene menos de 100 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 5 pg de tripsina por 100 ng de proteína BoNT/E1; preferiblemente en la que dicha composición farmacéutica líquida contiene menos de 10 pg de tripsina por 100 ng de proteína BoNT/E1, o menos de 7 pg de tripsina por 100 ng de proteína BoNT/E1.

De acuerdo con un aspecto específico, la composición farmacéutica de acuerdo con la divulgación en forma líquida se sella en un vial o en un dispositivo listo para usar, tal como una jeringa, sin interfase líquida/gaseosa, y es estable durante al menos tres meses o al menos seis meses a 23 a 27°C y durante al menos doce meses a 2-8°C.

En un aspecto, la divulgación proporciona una proteína BoNT/E1 de doble cadena activa como se describe anteriormente, o una proteína BoNT/E1 de doble cadena activa que se puede obtener por escisión proteolítica de la proteína BoNT/E1 de cadena sencilla como se describió anteriormente, o una composición como se describe anteriormente, o una composición farmacéutica líquida como se describe anteriormente, para uso en terapia.

Los presentes inventores han identificado que las proteínas BoNT/E1 de doble cadena activas de la divulgación, y las composiciones y composiciones farmacéuticas líquidas de las mismas, se pueden usar en terapia. Las terapias adecuadas pueden incluir tratamientos cosméticos y métodos de tratamiento médico.

Clave para las SEQ ID NOs

SEQ ID NO: 1 Secuencia optimizada de ácidos nucleicos de BoNT/E1

SEQ ID NO: 2 Secuencia de aminoácidos de BoNT/E1

SEQ ID NO: 3 Secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre

SEQ ID NO: 1 Secuencia optimizada de ácidos nucleicos de BoNT/E1

ATGCCGAAAAATCAACTCTTTCAACTACAACGACCCGGTTAACGACCGTACCATCCTGTAT
 ATCAAACCGGGTGGTTGCCAGGAGTTCTACAAATCTTTCAACATCATGAAAAACATCTGG
 ATCATCCCGGAACGTAACGTTATCGGTACCACCCCGCAGGACTTCCACCCGCCGACCTCT
 CTGAAAAACGGTGACTCTTCTTACTACGACCCGAACTACCTCCAGTCTGACGAAGAAAAA
 GACCGTTTCTGAAAATCGTTACCAAAATCTTCAACCGTATCAACAACAACCTGTCTGGT
 GGTATCCTGCTGGAAGAACTGTCTAAAGCTAACCCGTACCTGGGTAACGACAACACCCCG
 GACAACCAGTTCCACATCGGTGACGCTTCTGCTGTTGAAATCAAATTCTCTAACGGTTCT
 CAGGACATCCTGCTGCCGAACGTTATCATCATGGGTGCTGAACCGGACCTGTTGAAACC
 AACTCTTCTAACATCTCTCTGCGTAACAACCTACATGCCGTCTAACACGGTTTCGGTTCT
 ATCGCTATCGTTACCTTCTCTCCGGAATACTCTTTCCGTTTCAACGACAACAGCATGAAC
 GAGTTCATCCAGGACCCGGCTCTGACCCTGATGCACGAACTGATCCACTCTCTGCACGGT
 CTGTACGGTGCTAAAGGTATCACCACCAAATACACCATCACCCAGAAACAGAACCCGCTG
 ATCACCACATCCGTGGTACCAACATCGAAGAGTTCCTGACCTTCGGTGGTACCGACCTG
 AACATCATCACCTCTGCTCAGTCTAACGACATCTACACCAACCTGCTGGCTGACTACAAA
 AAAATCGCTTCTAAACTGTCTAAAGTTCAGGTTTCTAACCCGCTGCTGAACCCGTACAAA
 GACGTTTTCTGAAGCTAAATACGGTCTGGACAAAGACGCTTCTGGTATCTACTCTGTTAAC
 ATCAACAAATTCACGACATCTTCAAAAACTGTACTCTTTCACCGAGTTCGACCTGGCG
 ACCAAATTCAGGTTAAATGCCGTGACACCTACATCGGTGAGTACAAATACTTCAAACCTG
 TCTAACCTGCTGAACGACTCTATCTACAACATCTCTGAAGGTTACAACATCAACAACCTG
 AAAGTTAACTTCCGTGGTCAGAACGCTAACCTGAACCCGCGTATCATCACCCGATCACC
 GGTGCTGGTCTGGTTAAAAAATCATCCGTTTCTGCAAGAATATTGTAAGCGTTAAAGGA
 ATAAGAAAAAGTATCTGCATCGAAATCAACAACGCTGAACTGTTCTTCTGCTTCTGAA
 AACTCTTACAACGACGACAACATCAACACCCCGAAAGAAATCGACGACACCGTTACCTCT
 AACAACAACCTACGAAAACGACCTGGACCAGGTTATCCTGAACCTCAACTCTGAATCTGCT
 CCGGGTCTGTCTGACGAAAAACTGAACCTGACCATCCAGAACGACGCTTACATCCCGAAA
 TACGACTCTAACGGTACCTCTGACATCGAACAGCACGACGTTAACGAACCTGAACGTTTTCT
 TTCTACCTGGACGCTCAGAAAGTTCCGGAAGGTGAAAACAACGTTAACCTGACCTCTTCT
 ATCGACACCGCTCTGCTGGAACAGCCGAAAAATCTACACCTTCTTCTCTTCTGAGTTCATC
 AACAACGTTAACAAACCGGTTTCAAGGCTGCTCTGTTTCGTTTCTTGGATTTCAGCAGGTTCTG
 GTTGACTTCACCACCGAAGCTAACAGAAATCTACCGTTGACAAAATCGCTGACATCTCT
 ATCGTTGTTCCGTACATCGGTCTGGCTCTGAACATCGGTAACGAAGCTCAGAAAGGTAAC
 TTCAAAGACGCTCTGGAACCTGCTGGGTGCTGGTATCCTGCTGGAGTTTCGAACCGGAACCTG
 CTGATCCCGACCATCCTGGTTTTTACCATCAAATCTTTCCTGGGTCTTCTGACAACAAA
 AACAAAGTTATCAAAGCTATCAACAACGCTCTGAAAGAACGTGACGAAAAATGGAAAGAA
 GTTTACTCTTTCATCGTTTCTAACTGGATGACCAAAATCAACACCCAGTTCAACAAACGT
 AAAGAACAGATGTACCAGGCTCTCCAGAACCAGGTTAACGCTATCAAACCATCATCGAA
 TCTAAATACAACCTTACACCCTGGAAGAAAAAACGAACTGACCAACAAATACGACATC

AAACAGATCGAAAACGAACTGAACCAGAAAGTTTCTATCGCTATGAACAACATCGACCGT
 TTCCTGACCGAATCTTCTATCTCTTACCTGATGAAACTCATCAACGAAGTTAAATCAAC
 AAAGTGCCTGAATACGACGAAAACGTTAAACCTACCTGCTGAACTACATCATCCAGCAC
 GGTTCTATCCTGGGTGAATCTCAGCAGGAAGTGAAGTCTATGGTTACCGACACCCCTGAAC
 AACTCTATCCCGTTCAAACGTCTTCTTACACCGACGACAAAATCCTGATCTCTTACTTC
 AACAAATTCTTTAAACGCATTAAGAGTTTCATCGGTTCTGAATATGCGGTACAAAAATGAT
 AAATATGTCGATACTTCTGGATATGATAGCAATATCAACATTAACGGCGACGTGTATAAA
 TATCCGACAAATAAAAAACAGTTTGGGATATATAACGACAAGCTGTGCGGAGGTCAATATT
 TCTCAAAACGACTATATCATTTACGATAATAAATATAAAAACTTTAGCATTAGTTTTTGG
 GTTCGTATACCTAATTATGACAATAAAATTGTAAATGTGAATAACGAGTATACCATTTATA
 AACTGTATGCGCGACAATAACAGTGGTTGGAAGGTATCGCTGAACCATAATGAGATTATC
 TGGACCCCTGCAGGATAATGCAGGTATAAACAGAACTGGCTTTTAACTATGGAAACGCA
 AATGGGATCTCAGATTACATTAATAAATGGATTGTTTGTGTTACCATACGACGATCGCTTA
 GGCGACTCAAAACTTTATATTAATGGCAATCTGATAGATCAGAAATCAATCTTAAATTTG
 GGCAATATTCTGTCTCTGATAACATCTTGTTCAGATCGTTAATTGCAGTTACACTCGT
 TATATTGGCATTCTGTTACTTTAATATCTTCGATAAAGAACTGGACGAGACGGAAATCCAG
 ACTCTGTATTCAAACGAGCCCAATACTAATATATTGAAAGATTTTTGGGGTAACTATCTT
 TTATATGATAAAGAATACTATCTCCTGAATGTATTGAAGCCAAACAATTTATAGATAGA
 CGCAAGGATAGCACATTAAGTATCAACAATATCAGATCTACTATACTGTTAGCAAATCGC
 CTCTACTCCGGTATTAAAGTGAAGATTCAGCGGGTTAATAACTCCAGTACCAATGATAAT
 CTGGTCCGTAAGAACGATCAGGTATACATCAATTTCTGTCGCGAGCAAACTCATCTCTTC
 CCGCTTTACGCGGATACAGCTACGACAAACAAGGAAAAAACCATAAAAAATTTCCAGCTCC
 GGAAACAGATTCAATCAAGTAGTTGTAATGAACTCTGTGGGTAAATAATTGTACGATGAAC
 TTTAAGAATAACAATGGGAACAATATTGGACTTTTGGGCTTCAAAGCCGACACAGTGGTG
 GCGTCCACCTGGTATTACACGCACATGCGGGACCATAACGAATTCGAACGGTTGCTTCTGG
 AACTTTATCTCGGAAGAACACGGGTGGCAAGAAAAATAA

SEQ ID NO: 2 Secuencia de aminoácidos de BoNT/E1

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS
 LKNGDSSYYDPNYLQSDDEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP
 DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPNSHGFSG
 IAIVTFSPESYFRFNDNSMNEFIQDPALTLMHLEIHSLSHGLYGAIGITTKYITITQKQNPL
 ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLNPNYK
 DVFEAKYGLDKDASGIYSVNINKFNDFIKKLYSFTEFDLTKFQVKCRQTYIGQYKYFKL
 SNLLNDSIYNISEGYNNLNKVNFRGNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG
 IRKSICIEINNGELFFVASSENSYNDNINPTKEIDDTVTSNNNYENDLDQVILNFNSESA
 PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQVPEGENNVNLTSS
 IDTALLEQPKIYTFESSEFINNVNKPQVQALFVSWIQQVLVDFTTEANQKSTVDKIADIS
 IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK
 NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNVNAIKTIIIE
 SKYNSYTLLEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN
 KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNIPFKLSSYDDKILISYF
 NKFFKRIKSSSVLNMRYKNDKYVDTSYGSNININGDVYKYPTNKNQFGIYNDKLSEVNI
 SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLSLNHNEII
 WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL
 GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNLIKDFWGNLYL
 LYDKEYYLLNVLPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNSSTNDN
 LVRKNDQVYINFAVASKTHLFLYADTATTNKEKTIKISSSGNRFNQVVMNSVGNNCTMN
 FKNNGNNGNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHWQEK

SEQ ID NO: 3 Secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre

ATGCCAAAAATTAATAGTTTTAATTATAATGATCCTGTAAATGATAGAACAAATTTTATAT
 ATTAAACCAGGCGGTTGTCAAGAATTTTATAAATCATTTAATATTATGAAAAATATTTGG
 ATAATTCCAGAGAGAAATGTAATTGGTACAACCCCCCAAGATTTTCATCCGCTACTTCA
 TTAaaaaatGGAGATAGTAGTTATTATGACCCTAATTATTTACAAAGTGATGAAGAAAAG
 GATAGATTTTTTAAAAATAGTCACAAAAATATTTAATAGAATAAATAAATCTTTCAGGA
 GGGATTTTATTAGAGAAGCTGTCAAAAGCTAATCCATATTTAGGGAATGATAATACTCCA
 GATAATCAATTCCATATTGGTGATGCATCAGCAGTTGAGATTAAATTTCTCAAAATGGTAGC
 CAAGACATACTATTACCTAATGTTATTATAATGGGAGCAGAGCCTGATTATTTGAAACT
 AACAGTTCCAATATTTCTCTAAGAAATAATTATATGCCAAGCAATCACGGTTTTGGATCA
 ATAGCTATAGTAACATTCTCACCTGAATATTCTTTTAGATTTAATGATAATAGTATGAAT
 GAATTTATTCAAGATCCTGCTCTTACATTAATGCATGAATTAATACATTCATTACATGGA
 CTATATGGGGCTAAAGGGATTACTACAAAGTATACTATAACACAAAAACAAAATCCCCTA
 ATAACAAATATAAGAGGTACAAATATTGAAGAATTCTTAACTTTTGGAGGTACTGATTTA
 AACATTATTACTAGTGCTCAGTCCAATGATATCTATACTAATCTTCTAGCTGATTATAAA
 AAAATAGCGTCTAAACTTAGCAAAAGTACAAGTATCTAATCCACTACTTAATCCTTATAAA
 GATGTTTTTTGAAGCAAAGTATGGATTAGATAAAGATGCTAGCGGAATTTATTCGGTAAAT
 ATAAACAAATTTAATGATATTTTTTAAAAAATTATACAGCTTTACGGAATTTGATTTAGCA
 ACTAAATTTCAAGTTAAATGTAGGCAAACTTATATTGGACAGTATAAATACTTCAAACCTT
 TCAAACCTTGTAAATGATTCTATTTATAATATATCAGAAGGCTATAATATAAATAATTTA
 AAGGTAAATTTTAGAGGACAGAATGCAAAATTTAAATCCTAGAATTATTACACCAATTACA
 GGTAGAGGACTAGTAAAAAAAATCATTAGATTTTGTAAAAATATTGTTTCTGTAAAAGGC
 ATAAGGAAATCAATATGTATCGAAATAAATAATGGTGAGTTATTTTTTGTGGCTTCCGAG
 AATAGTTATAATGATGATAATATAAATACTCCTAAAGAAATTGACGATACAGTAACTTCA
 AATAATAATTATGAAAATGATTTAGATCAGGTTATTTTAAATTTTAATAGTGAATCAGCA
 CCTGGACTTTTCAGATGAAAAATTAAATTTAACTATCCAAAATGATGCTTATATACCAAAA
 TATGATTCTAATGGAACAAGTGATATAGAACAACATGATGTTAATGAACCTAATGTATTT
 TTCTATTTAGATGCACAGAAAGTGCCCGAAGGTGAAAATAATGTCAATCTCACCTCTTCA
 ATTGATACAGCATTATTAGAACAACCTAAAAATATATACATTTTTTTTCATCAGAATTTATT
 AATAATGTCAATAAACCTGTGCAAGCAGCATTATTTGTAAGCTGGATACAACAAGTGTTA
 GTAGATTTTACTACTGAAGCTAACCAAAAAAGTACTGTTGATAAAATTGCAGATATTTCT
 ATAGTTGTTCCATATATAGGTCTTGCTTTAAATATAGGAAATGAAGCACAAAAAGGAAAT
 TTTAAAGATGCACTTGAATTATTAGGAGCAGGTATTTTATTAGAATTTGAACCCGAGCTT
 TTAATTCCTACAATTTTAGTATTACAGATAAAATCTTTTTTAGGTTTCATCTGATAATAAA
 AATAAAGTTATTAAAGCAATAAATAATGCATTGAAAGAAAGAGATGAAAAATGGAAAGAA
 GTATATAGTTTTTATAGTATCGAATTGGATGACTAAAAATTAATACACAATTTAATAAAAGA
 AAAGAACAATGTATCAAGCTTTACAAAATCAAGTAAATGCAATTAAAAACAATAATAGAA
 TCTAAGTATAATAGTTATACTTTAGAGGAAAAAAAATGAGCTTACAAAATAAATATGATATT
 AAGCAATAGAAAATGAACTTAATCAAAAGGTTTCTATAGCAATGAATAATATAGACAGG
 TTCTTAACTGAAAGTTCTATATCCTATTTAATGAAATTAATAAATGAAGTAAAAATTAAT
 AAATTAAGAGAATATGATGAGAATGTCAAAACGTATTTATTGAATTATATTATACAACAT
 GGATCAATCTTGGGAGAGAGTCAGCAAGAAGTAAATTTCTATGGTAACTGATACCTTAAAT
 AATAGTATTCCTTTTAAGCTTTCTTCTTATACAGATGATAAAATTTTAATTTTCATATTTT
 AATAAATTCCTTAAGAGAATTAAGAGTAGTTTCAGTTTTAAATATGAGATATAAAAAATGAT
 AAATACGTAGATACTTCAGGATATGATTCAAATATAAATATTAATGGAGATGTATATAAA
 TATCCAACATAAAAAATCAATTTGGAATATATAATGATAAACTTAGTGAAGTTAATATA
 TCTCAAAATGATTACATTATATATGATAATAAATATAAAAAATTTTAGTATTAGTTTTTGG
 GTAAGAATTCCTAACTATGATAATAAGATAGTAAATGTTAATAATGAATACACTATAATA
 AATTGTATGAGAGATAATAATTGAGGATGGAAAGTATCTCTTAATCATAATGAAATAATT
 TGGACATTGCAAGATAATGCAGGAATTAATCAAAAATTAGCATTAACTATGGTAACGCA
 AATGGTATTTCTGATTATATAAATAAGTGGATTTTTGTAACTATACTAATGATAGATTA

GGAGATTCTAACTTTATATTAATGGAAATTTAATAGATCAAAAATCAATTTTAAATTTA
 GGTAAATATTCATGTTAGTGACAATATATTATTTAAAATAGTTAATTGTAGTTATACAAGA
 TATATTGGTATTAGATATTTTAATATTTTTGATAAAGAATTAGATGAAACAGAAATTCAA
 ACTTTATATAGCAATGAACCTAATACAAATATTTTGAAGGATTTTGGGGAAATTATTTG
 CTTTATGACAAAGAATACTATTTTAAATGTGTTAAACCAAATAACTTTATTGATAGG
 AGAAAAGATTCTACTTTAAGCATTAAATAATAAGAAGCACTATTCTTTTAGCTAATAGA
 TTATATAGTGGAATAAAAGTTAAATACAAAGAGTTAATAATAGTAGTACTAACGATAAT
 CTTGTTAGAAAGAATGATCAGGTATATATTAATTTTGTAGCCAGCAAAACTCACTTATTT
 CCATTATATGCTGATACAGCTACCACAAATAAAGAGAAAAACAATAAAAAATATCATCATCT
 GGCAATAGATTTAATCAAGTAGTAGTTATGAATTCAGTAGGAAATAATTGTACAATGAAT
 TTTAAAAATAATAATGGAAATAATATTGGGTTGTTAGGTTTCAAGGCAGATACTGTAGTT
 GCTAGTACTTGGTATTATACACATATGAGAGATCATACAAACAGCAATGGATGTTTTTGG
 AACTTTATTTCTGAAGAACATGGATGGCAAGAAAAATAA

Lista de Figuras

Figura 1: Fracciones de elución de columnas de intercambio aniónico en las que se evaluó la separación de tripsina de BoNT/E1. El pico de tripsina, BoNT/E1 y el gradiente de sal están marcados. Fig. 1A: Q-Sefarosa HP; Fig. 1B: DEAE Sefarosa.

Figura 2: Fracciones de elución de columnas de interacción hidrófoba en las que se evaluó la separación de tripsina de BoNT/E1. El pico de tripsina, BoNT/E1 y el gradiente de sal están marcados. Fig. 2A: Fenil Sefarosa HP; Fig. 2B: Butil Sefarosa HP; Fig. 2C: Octil Sefarosa FF.

Figura 3: Nivel de expresión soluble del cultivo de rBoNT/E1 determinado por transferencia Western, en comparación con BoNT/E1 comercial.

Figura 4: SDS-PAGE de rBoNT/E1 en condiciones no reductoras y reductoras que confirman la formación de la estructura de doble cadena.

Ejemplos

Ejemplo 1

Construcción de una secuencia optimizada de ácido nucleico de BoNT/E1

La secuencia de ADN se diseñó inicialmente por traducción inversa de la secuencia de aminoácidos de BoNT/E1 (SEQ ID NO: 2). Se añadió una secuencia de restricción (PstI) al extremo del terminal N y un codón de terminación y secuencias de restricción adicionales, XbaI-codón de detención-HindIII, al extremo del terminal C. La secuencia de ADN se optimizó luego para la expresión en función del número y la ubicación de los codones lentos (como se definió anteriormente).

La secuencia se optimizó para seleccionar contra codones lentos. Esto se aplicó particularmente al comienzo de la secuencia para obtener un buen inicio y comenzar la traducción. Cuando se incluyeron codones lentos (para permitir el uso de acuerdo con el sesgo de expresión del codón huésped), estos se encontraban hacia el final de la secuencia (donde el comienzo de la secuencia se define como el lugar donde se inicia la traducción).

Una vez que se había diseñado la secuencia, se sintetizó la secuencia de ADN optimizada en dos partes usando un sitio PstI único/nativo para su posterior ensamblaje en el gen de la toxina de longitud completa. La secuencia del primer gen incluía un sitio NdeI en el extremo terminal amino y un sitio PstI en el extremo terminal carboxilo. Esta parte del gen tenía una longitud de 2.895 pb, que codifica la LC de BoNT/E1 y la porción amino de HC. La secuencia del segundo gen incluía un sitio PstI en el extremo terminal N y el sitio HindIII en el extremo terminal carboxilo, tenía una longitud de 882 pb y codificaba la porción carboxilo de la HC de BoNT/E1.

Ejemplo 2

Construcción de la secuencia de ácidos nucleicos de BoNT/E1 del vector de expresión

Se empleó un vector de expresión basado en el vector pET-26b(+) (Novagen), que incluye los sitios de restricción de clonación NdeI y HindIII localizados al comienzo y al final del ADN que codifica el ORF de BoNT/E1. El vector pET-26b(+) era de movilización deficiente pero podría mobilizarse si era corresponsable con otros plásmidos que pueden ser movilizados. El vector pET-26b(+) se modificó para eliminar los genes de movilidad y hacerlos inmovilizables.

El vector de expresión se digirió con NdeI y PstI y la cadena principal del vector purificado se ligó con el primer fragmento de ADN de BoNT/E1 que se había digerido con las mismas enzimas de restricción para crear un producto intermedio. En la segunda etapa de clonación, el ADN de BoNT/E1 del segundo fragmento que había sido digerido con PstI y HindIII se ligó en el producto intermedio de la etapa uno (que también se había digerido con las mismas enzimas de restricción). Esto condujo a la creación del producto final de ADN de BoNT/E1 en el vector de expresión.

Ejemplo 3

Inserción del vector de expresión de BoNT/E1 en el huésped

Este ejemplo se basa en el uso de células BLR de *E. coli* (DE3), aunque los procedimientos y métodos son igualmente aplicables a cualquier otra cepa de expresión de *E. coli*. Las células competentes BLR de *E. coli* (DE3) se almacenaron por debajo de -70°C hasta que fueron requeridas. La transformación de las células se llevó a cabo utilizando una adaptación del protocolo del fabricante. Las células se descongelaron en hielo y se prepararon alícuotas secundarias de 10 µL. Se transformó una alícuota usando un choque térmico a 42°C durante 80 segundos con 1 µL de ADN plasmídico. Después de recuperación en hielo durante 5 minutos, se añadieron 90 µL de caldo SOC libre de productos animales a las transformaciones que luego se transfirieron a incubadoras con agitación y se incubaron durante 1 hora a 37°C y 250 rpm. Después de la incubación, se transfirieron 90 µL de cada transformación y se extendieron en placas de agar LB libres de productos animales suplementadas con 50 µg/mL de kanamicina. Las placas se incubaron a 37°C durante 16 horas.

Ejemplo 4

Cultivo del huésped y expresión de la proteína rBoNT/E1 soluble

Se usó una sola colonia de BoNT/E1 transformada en células BLR (DE3) para inocular un matraz cónico de 250 mL que contenía 100 mL de Caldo Terrific modificado (mTB) suplementado con 0,2% de glucosamina y 30 µg/mL de kanamicina. Este método sería igualmente aplicable cuando se usa una perla Microbank o un patrón de glicerol (10-100 µl) para inocular el matraz.

El matraz se incubó durante 16 horas a 37°C con agitación a 250 RPM. Se usaron 10 mL de este cultivo de partida para inocular matraces cónicos de 2 L que contenían cada uno 1 L suplementado con glucosamina al 0,2% y 30 µg/mL de kanamicina. Las células se cultivaron a 37°C durante ~2 horas a 225 RPM hasta que se alcanzó una DO₆₀₀ de 0,5. En este punto, la temperatura de cultivo se redujo a 16°C. Después de 1 hora, se indujeron las células para expresar BoNT/E1 mediante la adición de IPTG 1 mM durante 20 horas. Las células se recogieron por centrifugación durante 20 minutos a 4°C, se pesaron y luego se almacenaron a -20°C.

Ejemplo 5

Extracción de la proteína BoNT/E1 del huésped y análisis del nivel de expresión

Las pastas celulares de expresión de rBoNT/E1 se descongelaron a temperatura ambiente y se resuspendieron pipeteando en 3 mL de regulador de resuspensión Tris-NaCl por gramo de células complementado con 10 µl de benzonasa. Las células se sometieron a lisis mediante sonicación a una amplitud de 4 µm - 10x 30 s en +> 45 s de desactivación. El lisado se centrifugó a 4.000 g durante 1 h a 4°C para obtener la rBoNT/E1 soluble en el sobrenadante.

Ensayo de Bradford para determinar la concentración de proteína total de lisados preparados

Se añadió una muestra (50 µL) de lisado de rBoNT/E1 diluido o estándar de BSA a cubetas de plástico de 1 mL. Se añadieron 450 µL de reactivo de ensayo Coomassie Bradford a cada cubeta y se dejó incubar a temperatura ambiente durante 10 minutos antes de leer una DO₆₀₀. Los valores obtenidos para los estándares de BSA se usaron para determinar la cantidad de proteína en las muestras de lisado.

Preparación de muestras de lisado para análisis semicuantitativo de transferencia Western

Se utilizó una muestra comercial de proteína BoNT/E1 adquirida de Metabionics para elaborar los estándares de SDS-PAGE. A continuación, se prepararon muestras de SDS-PAGE a partir de las muestras de lisado de los cultivos celulares expresados hasta una concentración de proteína total conocida.

Transferencia Western

Los geles se cargaron y corrieron a 200 V durante 55 minutos y se transfirieron a 0,4 mA durante 1 hora sobre membrana de nitrocelulosa en regulador de transferencia libre de metanol. Las transferencias de nitrocelulosa se bloquearon durante 1 hora con BSA al 0,5% en PBS-Tween 20 al 0,1% y luego se sondearon con un anticuerpo para BoNT/E1 durante 1 hora. Las transferencias se detectaron con anticuerpo secundario conjugado con HRP desarrollado con el sustrato SuperSignal DuraWest. Se obtuvieron imágenes de las transferencias desarrolladas con un instrumento de generación de imágenes Syngene (Figura 3).

Ejemplo 6 (ejemplo de referencia)

Purificación inicial y activación de la proteína BoNT/E1 objetivo en forma de doble cadena

Este ejemplo se basó en una combinación de etapas de captura y columna intermedia, aunque la combinación podría alterarse o invertirse para usar las mismas propiedades en un orden diferente. El sobrenadante clarificado se

llevó a una alta concentración de sal y se cargó en una columna de captura hidrófoba (butil sefarosa). La rBoNT/E1 unida se eluyó de la columna usando un gradiente de regulador Tris bajo en sal. La proteína eluida se purificó adicionalmente usando una columna de intercambio iónico tal como Q-sefarosa, eluyendo un gradiente de regulador Tris con alto contenido de sal. Luego se añadió tripsina a la muestra de rBoNT/E1 eluida hasta una concentración final de 2,5 µg/mL y se incubó a 37°C durante 40 minutos. Esto cortó el bucle de activación de BoNT/E1 y formó la estructura de doble cadena BoNT/E1 final, como se confirmó al reducir la SDS-PAGE (Figura 4).

Ejemplo 7 (ejemplo de referencia)

Purificación final de la proteína BoNT/E1 objetivo libre de activación de proteasa

La muestra de rBoNT/E1 activada se cargó inmediatamente en un regulador de alto contenido de sal en una columna hidrófoba (butil sefarosa). La columna se lavó con regulador con alto contenido en sal para eliminar la tripsina débilmente asociada, antes de aplicar un gradiente de regulador Tris bajo en sal para eliminar adicionalmente la tripsina de la columna y la proteína rBoNT/E1 unida. La proteína rBoNT/E1 se eluyó luego tarde en el gradiente, lejos de la tripsina.

Ensayo para determinar los niveles de tripsina

Se desarrolló un ELISA de tripsina para determinar los niveles presentes en las fracciones de columna y en la muestra final de BoNT/E1. Se recubrió un anticuerpo de captura anti-tripsina con placas de microtitulación durante 1 hora a 37°C. Los patrones de tripsina y las muestras de prueba se añadieron a la placa (100 µL/pozo) y se incubaron durante 1 hora a 37°C antes de su detección con un segundo anticuerpo anti-tripsina. La cantidad de tripsina en cada fracción de muestra/columna se interpoló luego a partir de los patrones y se superpuso sobre el cromatograma de purificación para confirmar la separación de la tripsina de la BoNT/E1 (Figura 2B).

Ejemplo 8 (ejemplo de referencia)

Formulación que comprende BoNT/E1 de doble cadena activa sustancialmente libre de tripsina

Se prepararon las siguientes seis composiciones líquidas que comprenden BoNT/E1 de doble cadena activa (Tabla 3).

	1	2	3	4	5	6
Polisorbato 80	0,10 mg/mL	0,10 mg/mL	0,10 mg/mL	0,10 mg/mL	-	-
Poloxámero	-	-	-	-	0,04 mg/mL	0,04 mg/mL
Sacarosa	4,0 mg/mL	-	4,0 mg/mL	-	4,0 mg/mL	-
Manitol	-	4,0 mg/mL	-	4,0 mg/mL	-	4,0 mg/mL
Cloruro de sodio	8,76 mg/mL	8,76 mg/mL	8,76 mg/mL	8,76 mg/mL	8,76 mg/mL	8,76 mg/mL
pH	6,5	6,5	6,5	6,5	6,5	6,5
Regulador	L-Histidina/ ácido clorhídrico	L-Histidina/ ácido clorhídrico	Fosfato disódico/ ácido cítrico anhidro	Fosfato disódico/ ácido cítrico anhidro	L-Histidina/ ácido clorhídrico	L-Histidina/ ácido clorhídrico
BoNT/E1 de doble cadena	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL
Agua MilliQ	suficiente hasta 1 mL	suficiente hasta 1 mL	suficiente hasta 1 mL	suficiente hasta 1 mL	suficiente hasta 1 mL	suficiente hasta 1 mL

Las seis composiciones se almacenaron a 25°C durante 12 semanas. La estabilidad de la función de la proteasa de BoNT/E1 de doble cadena se evaluó durante ese período usando un ensayo de endopeptidasa libre de células. Las tasas de degradación mensual para las seis formulaciones estuvieron por debajo del 5% por mes durante las 12 semanas, lo que muestra que la función de la proteasa BoNT/E1 de doble cadena de las seis composiciones permanece estable a 25°C durante al menos 12 semanas.

Lista de secuencias

- <110> IPSEN BIOINNOVATION LIMITED
IPSEN BIOPHARM LIMITED
- 5 <120> Neurotoxinas recombinantes de *Clostridium botulinum*
- <130> P39711EP-D1-PCT
- 10 <150> GB 1219602.8
<151> 31-10-2012
- <160> 3
- 15 <170> PatentIn versión 3.5
- <210> 1
<211> 3759
<212> ADN
20 <213> Secuencia Artificial
- <220>
<223> Secuencia optimizada de ácido nucleico de BoNT/E1
- 25 <400> 1

atgccgaaaa tcaactcttt caactacaac gacccgggta acgaccgtac catcctgtat	60
atcaaaccgg gtggttgcca ggagttctac aaatctttca acatcatgaa aaacatctgg	120
atcatcccgg aacgtaacgt tatcggtacc accccgcagg acttccaccc gccgacctct	180
ctgaaaaacg gtgactcttc ttactacgac ccgaactacc tccagtctga cgaagaaaaa	240
gaccgtttcc tgaaaatcgt taccaaaatc ttcaaccgta tcaacaacaa cctgtctgggt	300
gggtatcctgc tggaagaact gtctaaagct aaccgcgtacc tgggtaacga caacaccccg	360
gacaaccagt tccacatcgg tgacgcttct gctgttgaaa tcaaattctc taacggttct	420
caggacatcc tgctgccgaa cggttatcatc atgggtgctg aaccggacct gttcgaaacc	480
aactcttcta acatctctct gcgtaacaac tacatgccgt ctaaccacgg tttcggttct	540
atcgctatcg ttaccttctc tccggaatac tctttccggt tcaacgacaa cagcatgaac	600
gagttcatcc aggacccggc tctgaccctg atgcacgaac tgatccactc tctgcacgggt	660
ctgtacgggtg ctaaagggtat caccacaaaa tacaccatca cccagaaaca gaacccgctg	720
atcaccaaca tccgtggtac caacatcgaa gagttcctga ccttcgggtgg taccgacctg	780
aacatcatca cctctgctca gtctaacgac atctacacca acctgctggc tgactacaaa	840
aaaatcgctt ctaaactgtc taaagttcag gtttctaacc cgctgctgaa cccgtacaaa	900
gacgttttctg aagctaaata cggctctggac aaagacgctt ctggtatcta ctctgttaac	960
atcaacaaat tcaacgacat cttcaaaaaa ctgtactctt tcaccgagtt cgacctggcg	1020
accaaattcc aggttaaagt ccgtcagacc tacatcgggtc agtacaataa cttcaaaactg	1080
tctaacctgc tgaacgactc tatctacaac atctctgaag gttacaacat caacaacctg	1140

ES 2 788 199 T3

aaagttaact tccgtggtca gaacgctaac ctgaacccgc gtatcatcac cccgatcacc	1200
ggtcgtgggc tgggttaaaaa aatcatccgt ttctgcaaga atattgtaag cggttaaagga	1260
ataagaaaaa gtatctgcat cgaaatcaac aacggtgaac tgttcttcgt tgcttctgaa	1320
aactcttaca acgacgacaa catcaacacc ccgaaagaaa tcgacgacac cgttacctct	1380
aacaacaact acgaaaacga cctggaccag gttatcctga acttcaactc tgaatctgct	1440
ccgggtctgt ctgacgaaaa actgaacctg accatccaga acgacgctta catcccgaaa	1500
tacgactcta acggtacctc tgacatcgaa cagcacgacg ttaacgaact gaacgttttc	1560
ttctacctgg acgctcagaa agttccggaa ggtgaaaaca acgttaacct gacctcttct	1620
atcgacaccg ctctgctgga acagccgaaa atctacacct tcttctcttc tgagttcatc	1680
aacaacgtta acaaacccgt tcaggctgct ctgttcgttt ctgggattca gcaggttctg	1740
gttgacttca ccaccgaagc taaccagaaa tctaccgttg acaaaatcgc tgacatctct	1800
atcggtgttc cgtacatcgg tctggctctg aacatcggta acgaagctca gaaaggtaac	1860
ttcaaagacg ctctggaact gctgggtgct ggtatcctgc tggagttcga accggaactg	1920
ctgatcccgga ccctcctggt ttccaccatc aaatctttcc tgggttcttc tgacaacaaa	1980
aacaaagtta tcaaagctat caacaacgct ctgaaagaac gtgacgaaaa atggaaagaa	2040
gtttactott tcatcgtttc taactggatg accaaaatca acaccagtt caacaaacgt	2100
aaagaacaga tgtaccaggc tctccagaac cagggttaacg ctatcaaaac catcatcgaa	2160
tctaaatata actcttacac cctggaagaa aaaaacgaac tgaccaacaa atacgacatc	2220
aaacagatcg aaaaacgaact gaaccagaaa gtttctatcg ctatgaacaa catcgaccgt	2280
ttctgacccg aatcttttat ctcttacctg atgaaactca tcaacgaagt taaaatcaac	2340
aaactgctg aatacgacga aaacgttaaa acctacctgc tgaactacat catccagcac	2400
ggttctatcc tgggtgaatc tcagcaggaa ctgaactcta tggttaccga cacctgaac	2460
aactctatcc cgttcaaact gtcttcttac accgacgaca aaatcctgat ctcttacttc	2520
aacaaattct ttaaacgcat taagagttca tcggttctga atatgcggta caaaaatgat	2580
aaatatgtcg atacttctgg atatgatagc aatatcaaca ttaacggcga cgtgtataaa	2640
tatccgacaa ataaaaacca gtttgggata tataacgaca agctgtcgga ggtcaatatt	2700
tctcaaaacg actatatcat ttacgataat aaatataaaa acttttagcat tagtttttgg	2760
gttcgtatac ctaattatga caataaaatt gtaaatgtga ataacgagta taccattata	2820
aactgtatgc ggcacaataa cagtggttgg aagggtatcgc tgaaccataa tgagattatc	2880
tggaccctgc aggataatgc aggtataaac cagaaactgg cttttaacta tggaaacgca	2940
aatgggatct cagattacat taataaatgg atttttgtta ccattacgaa cgatcgctta	3000
ggcgactcaa aactttatat taatggcaat ctgatagatc agaaatcaat cttaaatttg	3060

ES 2 788 199 T3

```

ggcaatattc atgtctctga taacatcttg ttcaagatcg ttaattgcag ttacactcgt      3120
tatattggca ttcgttactt taatatcttc gataaagaac tggacgagac ggaaatccag      3180
actctgtatt caaacgagcc caatactaata atattgaaag atttttgggg taactatctt      3240
ttatatgata aagaatacta tctcctgaat gtattgaagc caaacaattt catagataga      3300
cgcaaggata gcacattaag tatcaacaat atcagatcta ctatactgtt agcaaatcgc      3360
ctctactccg gtattaaagt gaagattcag cgggttaata actccagtac caatgataat      3420
ctgggtccgta agaacgatca ggtatacatc aatttcgtcg cgagcaaaac tcatctcttc      3480
ccgctttacg ccgatacagc tacgacaaac aaggaaaaaa ccataaaaat ttccagctcc      3540
ggaaacagat tcaatcaagt agttgtaatg aactctgtgg gtaataattg tacgatgaac      3600
ttaaagaata acaatgggaa caatattgga cttttgggct tcaaagccga cacagtgggtg      3660
gcgtccacct ggtattacac gcacatgcgg gaccatacga attogaacgg ttgcttctgg      3720
aactttatct cggaagaaca cgggtggcaa gaaaaataa      3759

```

<210> 2

<211> 1252

<212> PRT

<213> *Clostridium botulinum*

<400> 2

```

Met Pro Lys Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asp Arg
1          5          10          15

Thr Ile Leu Tyr Ile Lys Pro Gly Gly Cys Gln Glu Phe Tyr Lys Ser
20          25          30

Phe Asn Ile Met Lys Asn Ile Trp Ile Ile Pro Glu Arg Asn Val Ile
35          40          45

Gly Thr Thr Pro Gln Asp Phe His Pro Pro Thr Ser Leu Lys Asn Gly
50          55          60

Asp Ser Ser Tyr Tyr Asp Pro Asn Tyr Leu Gln Ser Asp Glu Glu Lys
65          70          75          80

Asp Arg Phe Leu Lys Ile Val Thr Lys Ile Phe Asn Arg Ile Asn Asn
85          90          95

Asn Leu Ser Gly Gly Ile Leu Leu Glu Glu Leu Ser Lys Ala Asn Pro
100         105         110

Tyr Leu Gly Asn Asp Asn Thr Pro Asp Asn Gln Phe His Ile Gly Asp
115         120         125

```

ES 2 788 199 T3

Ala	Ser	Ala	Val	Glu	Ile	Lys	Phe	Ser	Asn	Gly	Ser	Gln	Asp	Ile	Leu
130						135					140				
Leu	Pro	Asn	Val	Ile	Ile	Met	Gly	Ala	Glu	Pro	Asp	Leu	Phe	Glu	Thr
145					150					155					160
Asn	Ser	Ser	Asn	Ile	Ser	Leu	Arg	Asn	Asn	Tyr	Met	Pro	Ser	Asn	His
				165					170					175	
Gly	Phe	Gly	Ser	Ile	Ala	Ile	Val	Thr	Phe	Ser	Pro	Glu	Tyr	Ser	Phe
			180					185					190		
Arg	Phe	Asn	Asp	Asn	Ser	Met	Asn	Glu	Phe	Ile	Gln	Asp	Pro	Ala	Leu
		195					200					205			
Thr	Leu	Met	His	Glu	Leu	Ile	His	Ser	Leu	His	Gly	Leu	Tyr	Gly	Ala
	210					215					220				
Lys	Gly	Ile	Thr	Thr	Lys	Tyr	Thr	Ile	Thr	Gln	Lys	Gln	Asn	Pro	Leu
225					230					235					240
Ile	Thr	Asn	Ile	Arg	Gly	Thr	Asn	Ile	Glu	Glu	Phe	Leu	Thr	Phe	Gly
				245					250					255	
Gly	Thr	Asp	Leu	Asn	Ile	Ile	Thr	Ser	Ala	Gln	Ser	Asn	Asp	Ile	Tyr
			260					265					270		
Thr	Asn	Leu	Leu	Ala	Asp	Tyr	Lys	Lys	Ile	Ala	Ser	Lys	Leu	Ser	Lys
		275					280					285			
Val	Gln	Val	Ser	Asn	Pro	Leu	Leu	Asn	Pro	Tyr	Lys	Asp	Val	Phe	Glu
	290					295					300				
Ala	Lys	Tyr	Gly	Leu	Asp	Lys	Asp	Ala	Ser	Gly	Ile	Tyr	Ser	Val	Asn
305					310					315					320
Ile	Asn	Lys	Phe	Asn	Asp	Ile	Phe	Lys	Lys	Leu	Tyr	Ser	Phe	Thr	Glu
				325					330					335	
Phe	Asp	Leu	Ala	Thr	Lys	Phe	Gln	Val	Lys	Cys	Arg	Gln	Thr	Tyr	Ile
			340					345					350		
Gly	Gln	Tyr	Lys	Tyr	Phe	Lys	Leu	Ser	Asn	Leu	Leu	Asn	Asp	Ser	Ile
		355					360					365			
Tyr	Asn	Ile	Ser	Glu	Gly	Tyr	Asn	Ile	Asn	Asn	Leu	Lys	Val	Asn	Phe

ES 2 788 199 T3

370		375		380
Arg Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr				
385		390		395 400
Gly Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val				
	405		410	415
Ser Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly				
	420		425	430
Glu Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile				
	435		440	445
Asn Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr				
	450		455	460
Glu Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala				
	465		470	475 480
Pro Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala				
	485		490	495
Tyr Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His				
	500		505	510
Asp Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val				
	515		520	525
Pro Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala				
	530		535	540
Leu Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile				
	545		550	555 560
Asn Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile				
	565		570	575
Gln Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr				
	580		585	590
Val Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu				
	595		600	605
Ala Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala				
	610		615	620

ES 2 788 199 T3

Leu Glu Leu Leu Gly Ala Gly Ile Leu Leu Glu Phe Glu Pro Glu Leu
 625 630 635 640

Leu Ile Pro Thr Ile Leu Val Phe Thr Ile Lys Ser Phe Leu Gly Ser
 645 650 655

Ser Asp Asn Lys Asn Lys Val Ile Lys Ala Ile Asn Asn Ala Leu Lys
 660 665 670

Glu Arg Asp Glu Lys Trp Lys Glu Val Tyr Ser Phe Ile Val Ser Asn
 675 680 685

Trp Met Thr Lys Ile Asn Thr Gln Phe Asn Lys Arg Lys Glu Gln Met
 690 695 700

Tyr Gln Ala Leu Gln Asn Gln Val Asn Ala Ile Lys Thr Ile Ile Glu
 705 710 715 720

Ser Lys Tyr Asn Ser Tyr Thr Leu Glu Glu Lys Asn Glu Leu Thr Asn
 725 730 735

Lys Tyr Asp Ile Lys Gln Ile Glu Asn Glu Leu Asn Gln Lys Val Ser
 740 745 750

Ile Ala Met Asn Asn Ile Asp Arg Phe Leu Thr Glu Ser Ser Ile Ser
 755 760 765

Tyr Leu Met Lys Leu Ile Asn Glu Val Lys Ile Asn Lys Leu Arg Glu
 770 775 780

Tyr Asp Glu Asn Val Lys Thr Tyr Leu Leu Asn Tyr Ile Ile Gln His
 785 790 795 800

Gly Ser Ile Leu Gly Glu Ser Gln Gln Glu Leu Asn Ser Met Val Thr
 805 810 815

Asp Thr Leu Asn Asn Ser Ile Pro Phe Lys Leu Ser Ser Tyr Thr Asp
 820 825 830

Asp Lys Ile Leu Ile Ser Tyr Phe Asn Lys Phe Phe Lys Arg Ile Lys
 835 840 845

Ser Ser Ser Val Leu Asn Met Arg Tyr Lys Asn Asp Lys Tyr Val Asp
 850 855 860

Thr Ser Gly Tyr Asp Ser Asn Ile Asn Ile Asn Gly Asp Val Tyr Lys
 865 870 875 880

ES 2 788 199 T3

Tyr Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser
 885 890 895
 Glu Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr
 900 905 910
 Lys Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn
 915 920 925
 Lys Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg
 930 935 940
 Asp Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile
 945 950 955 960
 Trp Thr Leu Gln Asp Asn Ala Gly Ile Asn Gln Lys Leu Ala Phe Asn
 965 970 975
 Tyr Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe
 980 985 990
 Val Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn
 995 1000 1005
 Gly Asn Leu Ile Asp Gln Lys Ser Ile Leu Asn Leu Gly Asn Ile
 1010 1015 1020
 His Val Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr
 1025 1030 1035
 Thr Arg Tyr Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu
 1040 1045 1050
 Leu Asp Glu Thr Glu Ile Gln Thr Leu Tyr Ser Asn Glu Pro Asn
 1055 1060 1065
 Thr Asn Ile Leu Lys Asp Phe Trp Gly Asn Tyr Leu Leu Tyr Asp
 1070 1075 1080
 Lys Glu Tyr Tyr Leu Leu Asn Val Leu Lys Pro Asn Asn Phe Ile
 1085 1090 1095
 Asp Arg Arg Lys Asp Ser Thr Leu Ser Ile Asn Asn Ile Arg Ser
 1100 1105 1110
 Thr Ile Leu Leu Ala Asn Arg Leu Tyr Ser Gly Ile Lys Val Lys
 1115 1120 1125

ES 2 788 199 T3

Ile	Gln	Arg	Val	Asn	Asn	Ser	Ser	Thr	Asn	Asp	Asn	Leu	Val	Arg
1130						1135					1140			
Lys	Asn	Asp	Gln	Val	Tyr	Ile	Asn	Phe	Val	Ala	Ser	Lys	Thr	His
1145						1150					1155			
Leu	Phe	Pro	Leu	Tyr	Ala	Asp	Thr	Ala	Thr	Thr	Asn	Lys	Glu	Lys
1160						1165					1170			
Thr	Ile	Lys	Ile	Ser	Ser	Ser	Gly	Asn	Arg	Phe	Asn	Gln	Val	Val
1175						1180					1185			
Val	Met	Asn	Ser	Val	Gly	Asn	Asn	Cys	Thr	Met	Asn	Phe	Lys	Asn
1190						1195					1200			
Asn	Asn	Gly	Asn	Asn	Ile	Gly	Leu	Leu	Gly	Phe	Lys	Ala	Asp	Thr
1205						1210					1215			
Val	Val	Ala	Ser	Thr	Trp	Tyr	Tyr	Thr	His	Met	Arg	Asp	His	Thr
1220						1225					1230			
Asn	Ser	Asn	Gly	Cys	Phe	Trp	Asn	Phe	Ile	Ser	Glu	Glu	His	Gly
1235						1240					1245			
Trp	Gln	Glu	Lys											
1250														

<210> 3

<211> 3759

5 <212> ADN

<213> *Clostridium botulinum*

<400> 3

atgccaaaaa ttaatagttt taattataat gatcctgtta atgatagaac aattttatat	60
attaaaccag gcggttgtca agaattttat aaatcattta atattatgaa aaatatttgg	120
ataattccag agagaaatgt aattggtaca acccccacag attttcatcc gcctacttca	180
ttaaaaaatg gagatagtag ttattatgac cctaattatt tacaaagtga tgaagaaaag	240
gatagatttt taaaaatagt cacaaaaata tttaatagaa taaataataa tctttcagga	300
gggatTTTTat tagaagaact gtcaaaagct aatccatatt tagggaatga taatactcca	360
gataatcaat tccatattgg tgatgcatca gcagttgaga ttaaattctc aaatggtagc	420
caagacatac tattacctaa tgttattata atgggagcag agcctgattt atttgaaact	480
aacagttcca atatttctct aagaaataat tatatgccaa gcaatcacgg ttttgatca	540
atagctatag taacattctc acctgaatat tcttttagat ttaatgataa tagtatgaat	600

gaattttattc aagatcctgc tcttacatta atgcatgaat taatacattc attacatgga	660
ctatatgggg ctaaagggat tactacaaag tatactataa cacaaaaaca aaatccccta	720
ataacaaata taagaggtag aatatgtga gaattcttaa cttttggagg tactgattta	780
aacattatta ctagtgtca gtccaatgat atctatacta atcttctagc tgattataaa	840
aaaatagcgt ctaaacttag caaagtacaa gtatctaata cactacttaa tccttataaa	900
gatgtttttg aagcaaagta tggattagat aaagatgcta gcggaattta ttcggtaaat	960
ataaaciaaat ttaatgatat ttttaaaaaa ttatacagct ttacggaatt tgatttagca	1020
actaaatttc aagttaaatg taggcaaact tatattggac agtataaata cttcaaactt	1080
tcaaacttgt taaatgattc tatttataat atatacagaag gctataatat aaataattta	1140
aaggtaaat ttagaggaca gaatgcaaat ttaaatccta gaattattac accaattaca	1200
ggtagaggac tagtaaaaaa aatcattaga ttttgtaaaa atattgtttc tgtaaaaggc	1260
ataaggaaat caatatgtat cgaaataaat aatggtgagt tattttttgt ggcttccgag	1320
aatagttata atgatgataa tataaatact cctaaagaaa ttgacgatac agtaacttca	1380
aataataatt atgaaaatga tttagatcag gttattttta attttaatag tgaatcagca	1440
cctggacttt cagatgaaaa attaaattta actatccaaa atgatgctta tataccaaaa	1500
tatgattcta atggaacaag tgatatagaa caacatgatg ttaatgaact taatgtattt	1560
ttctatttag atgcacagaa agtgcccgaa ggtgaaaata atgtcaatct cacctcttca	1620
attgatacag cattattaga acaacctaaa atatatacat ttttttcatc agaatttatt	1680
aataatgtca ataaacctgt gcaagcagca ttatttgtaa gctggatata acaagtgtta	1740
gtagatttta ctactgaagc taaccaaaaa agtactgttg ataaaattgc agatatttct	1800
atagttgttc catatatagg tcttgcttta aatataggaa atgaagcaca aaaaggaaat	1860
tttaagatg cacttgaatt attaggagca ggtattttat tagaatttga acccgagctt	1920
ttaattccta caattttagt attcacgata aaatcttttt taggttcatc tgataataaa	1980
aataaagtta ttaaagcaat aaataatgca ttgaaagaaa gagatgaaaa atggaaagaa	2040
gtatatagtt ttatagtatc gaattggatg actaaaatta atacacaatt taataaaaaga	2100
aaagaacaaa tgtatcaagc ttacaaaaa caagtaaatg caattaaaac aataatagaa	2160
tctaagtata atagttatac tttagaggaa aaaaatgagc ttacaaataa atatgatatt	2220
aagcaaatag aaaatgaact taatcaaaag gtttctatag caatgaataa tatagacagg	2280
ttcttaactg aaagtcttat atcctattta atgaaattaa taaatgaagt aaaaattaat	2340
aaattaagag aatatgatga gaatgtcaaa acgtatttat tgaattatat tataacaacat	2400
ggatcaatct tgggagagag tcagcaagaa ctaaaattcta tggtaactga taccctaaat	2460

ES 2 788 199 T3

aatagtattc cttttaagct ttcttcttat acagatgata aaattttaat ttcataat	2520
aataaattct ttaagagaat taaaagtagt tcagttttaa atatgagata taaaaatgat	2580
aaatacgtag atacttcagg atatgattca aatataaata ttaatggaga tgtatataaa	2640
tatccaacta ataaaaatca atttggaaata tataatgata aacttagtga agttaatata	2700
tctcaaaatg attacattat atatgataat aaatataaaa attttagtat tagtttttgg	2760
gtaagaattc ctaactatga taataagata gtaaatgtta ataatagaata cactataata	2820
aattgtatga gagataataa ttcaggatgg aaagtatctc ttaatcataa tgaaataatt	2880
tggaattgc aagataatgc aggaattaat caaaaattag catttaacta tggtaacgca	2940
aatgggtattt ctgattatat aaataagtgg atttttgtaa ctataactaa tgatagatta	3000
ggagattcta aactttatat taatggaaat ttaatagatc aaaaatcaat tttaaattta	3060
ggtaatatcc atgttagtga caatatatta tttaaaatag ttaattgtag ttatacaaga	3120
tatatggta ttagatattt taatatattt gataaagaat tagatgaac agaaattcaa	3180
actttatata gcaatgaacc taatacaaat attttgaagg atttttgggg aaattatttg	3240
ctttatgaca aagaataacta tttattaaat gtgttaaaac caaataactt tattgatagg	3300
agaaaagatt ctactttaag cattaataat ataagaagca ctattctttt agctaataga	3360
ttatatagtg gaataaaagt taaaatacaa agagttaata atagtagtac taacgataat	3420
cttggttagaa agaattgatca ggtatatatt aattttgtag ccagcaaaac tcacttattt	3480
ccatttatatg ctgatacagc taccacaaat aaagagaaaa caataaaaat atcatcatct	3540
ggcaatagat ttaatcaagt agtagttatg aattcagtag gaaataattg tacaatgaat	3600
tttaaaaaata ataattgaaa taatatggg ttgttaggtt tcaaggcaga tactgtagtt	3660
gctagtactt ggtattatac acatatgaga gatcatacaa acagcaatgg atgtttttgg	3720
aactttattt ctgaagaaca tggatggcaa gaaaaataa	3759

REIVINDICACIONES

1. Una secuencia de ácidos nucleicos que comprende una secuencia de nucleótidos contiguos, en el que dicha secuencia de nucleótidos contiguos tiene al menos 90% de identidad de secuencia con la secuencia de ácidos nucleicos de la SEQ ID NO: 1,
- 5 y en el que dicha secuencia de nucleótidos contiguos codifica una proteína BoNT/E1 de cadena sencilla.
2. La secuencia de ácidos nucleicos de acuerdo con la reivindicación 1, en el que el ácido nucleico tiene un máximo de 160 codones lentos seleccionados de TTT, TAT, TGT, CAT, CAA, CCA, CCG, TCA, TCG, CGG, TTA y CTA.
3. La secuencia de ácidos nucleicos de acuerdo con la reivindicación 1 o la reivindicación 2, en el que dicha proteína BoNT/E1 de cadena sencilla comprende una secuencia de aminoácidos contiguos, y en el que dicha secuencia de aminoácidos contiguos tiene al menos un 95% de identidad de secuencia con la secuencia de aminoácidos de la SEQ ID NO: 2.
- 10 4. La secuencia de ácidos nucleicos de acuerdo con una cualquiera de las reivindicaciones 1-3, con la condición de que la BoNT/E1 de cadena sencilla de acuerdo con la reivindicación 1 o la reivindicación 2 o dicha secuencia de aminoácidos contiguos de acuerdo con la reivindicación 3 incluye uno o más de los siguientes aminoácidos (en los que la numeración de la posición del aminoácido comienza con el resto del aminoácido del extremo terminal N y termina con el resto de aminoácido del extremo terminal C de la proteína BoNT/E1):
glicina en la posición 177
serina en la posición 198
alanina en la posición 340
20 leucina en la posición 773
leucina en la posición 963
glutamina en la posición 964
alanina en la posición 967
asparagina en la posición 1195.
- 25 5. La secuencia de ácidos nucleicos de acuerdo con cualquier reivindicación precedente, en la que dicha secuencia de nucleótidos contiguos tiene al menos 785 codones sinónimos cuando se compara con la secuencia de ácidos nucleicos de BoNT/E1 de tipo silvestre (SEQ ID NO: 3).
6. Un método para producir proteína BoNT/E1 de cadena sencilla soluble en una célula huésped de *E. coli*, comprendiendo dicho método:
30 expresar una secuencia de ácidos nucleicos de acuerdo con cualquier reivindicación precedente en un sistema de expresión de *E. coli*.
7. El método de acuerdo con la reivindicación 6, en el que dicha proteína BoNT/E1 de cadena sencilla soluble se expresa en el citoplasma de dicha célula huésped de *E. coli*.
8. El método de acuerdo con la reivindicación 6 o la reivindicación 7, en el que dicha proteína BoNT/E1 de cadena sencilla soluble se expresa a un nivel de al menos 5 mg/l.
- 35 9. El método de acuerdo con cualquiera de las reivindicaciones 6-8, que comprende la lisis de la célula huésped de *E. coli* para proporcionar un homogeneizado de la célula huésped de *E. coli* que contiene dicha proteína BoNT/E1 de cadena sencilla soluble.
10. Una célula huésped que comprende una secuencia de ácidos nucleicos como se define en una cualquiera de las reivindicaciones 1-5.
- 40 11. Una célula huésped de acuerdo con la reivindicación 10, que es una célula de *E. coli*.
12. Una célula huésped de acuerdo con la reivindicación 10 o la reivindicación 11, que es una célula BLR (DE3) de *E. coli*.
13. Un vector de expresión que comprende una secuencia de ácidos nucleicos como se define en una cualquiera de las reivindicaciones 1-5.
- 45

Fig. 1A

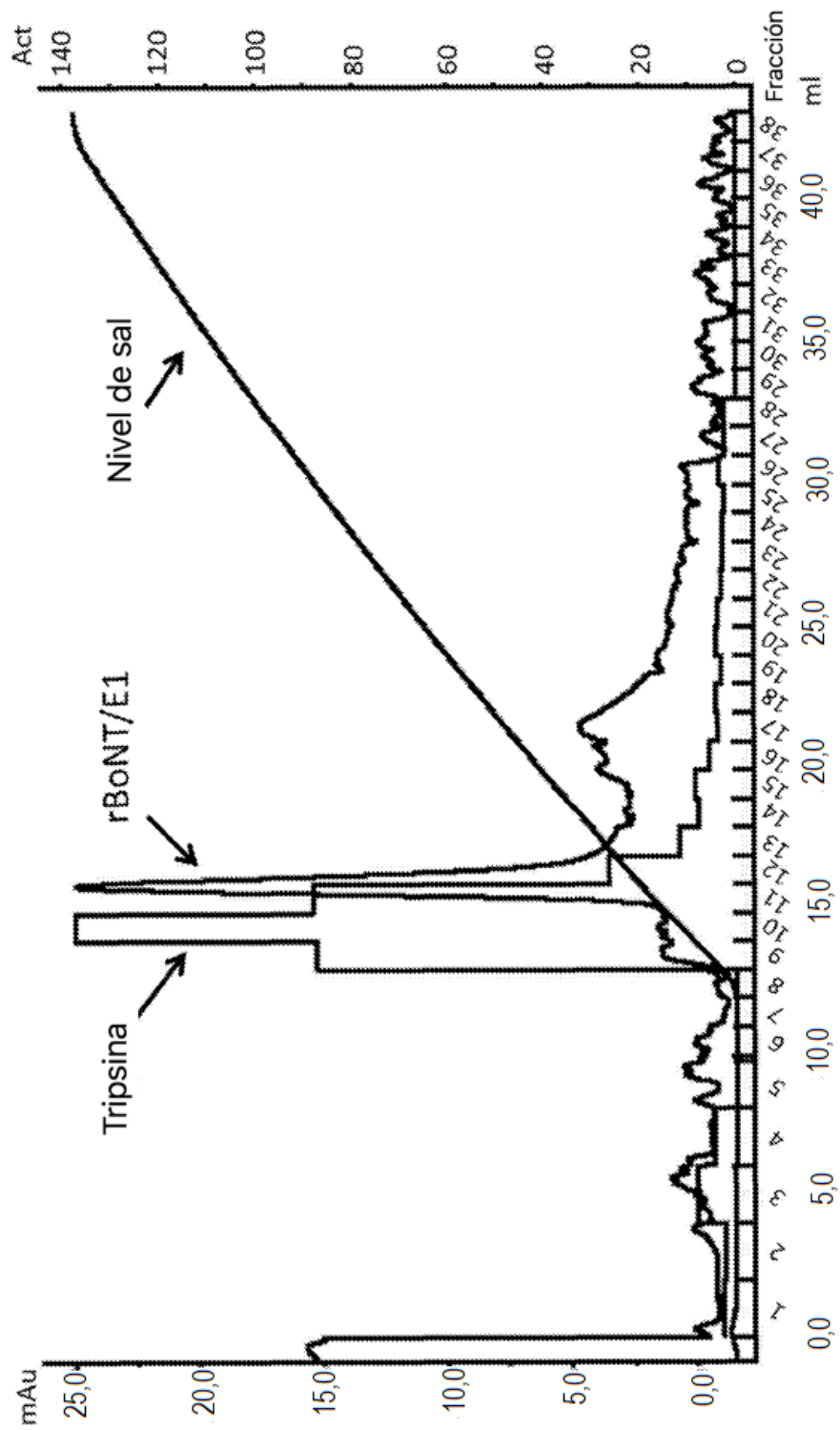


Fig. 1B

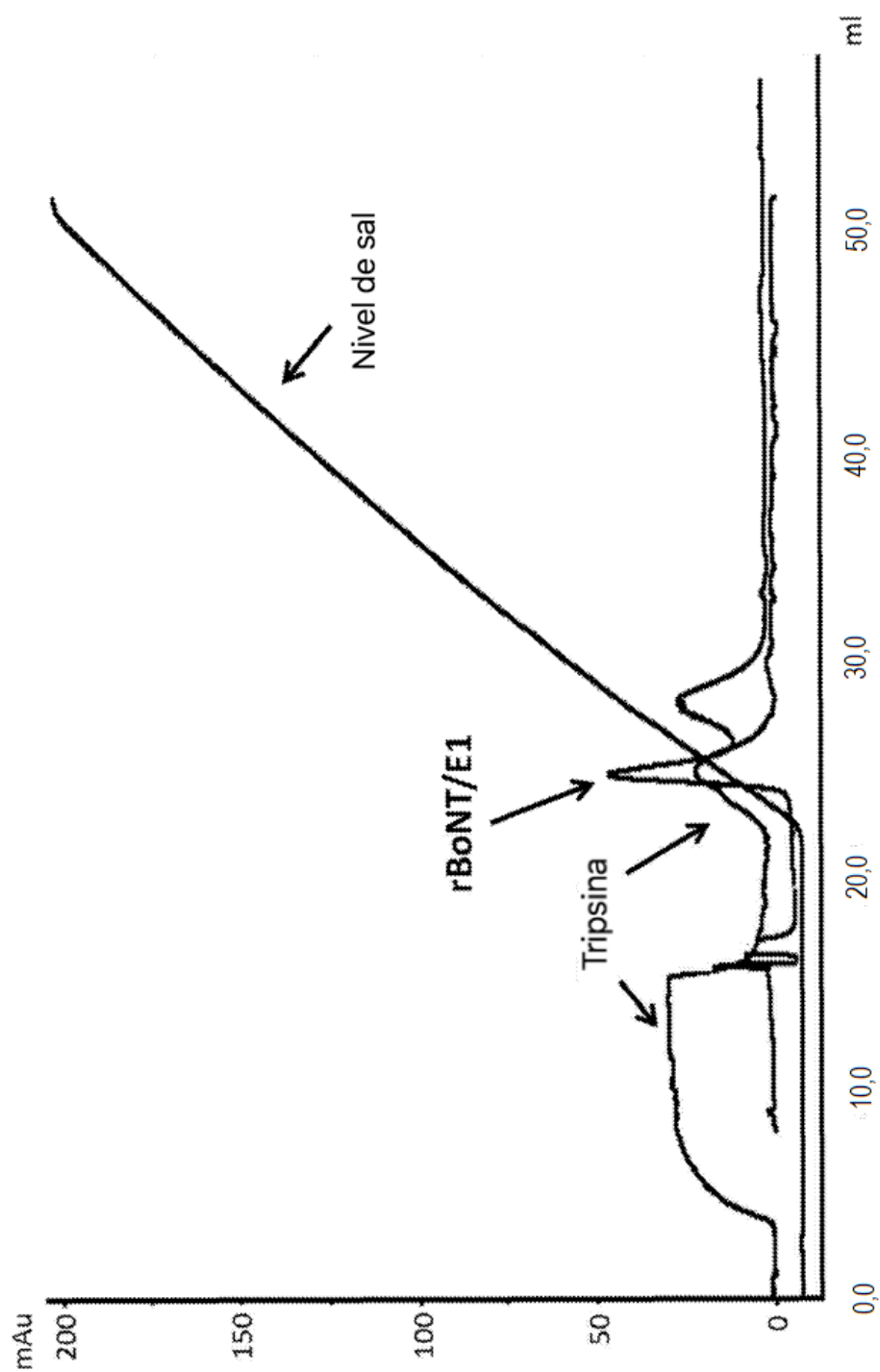


Fig. 2A

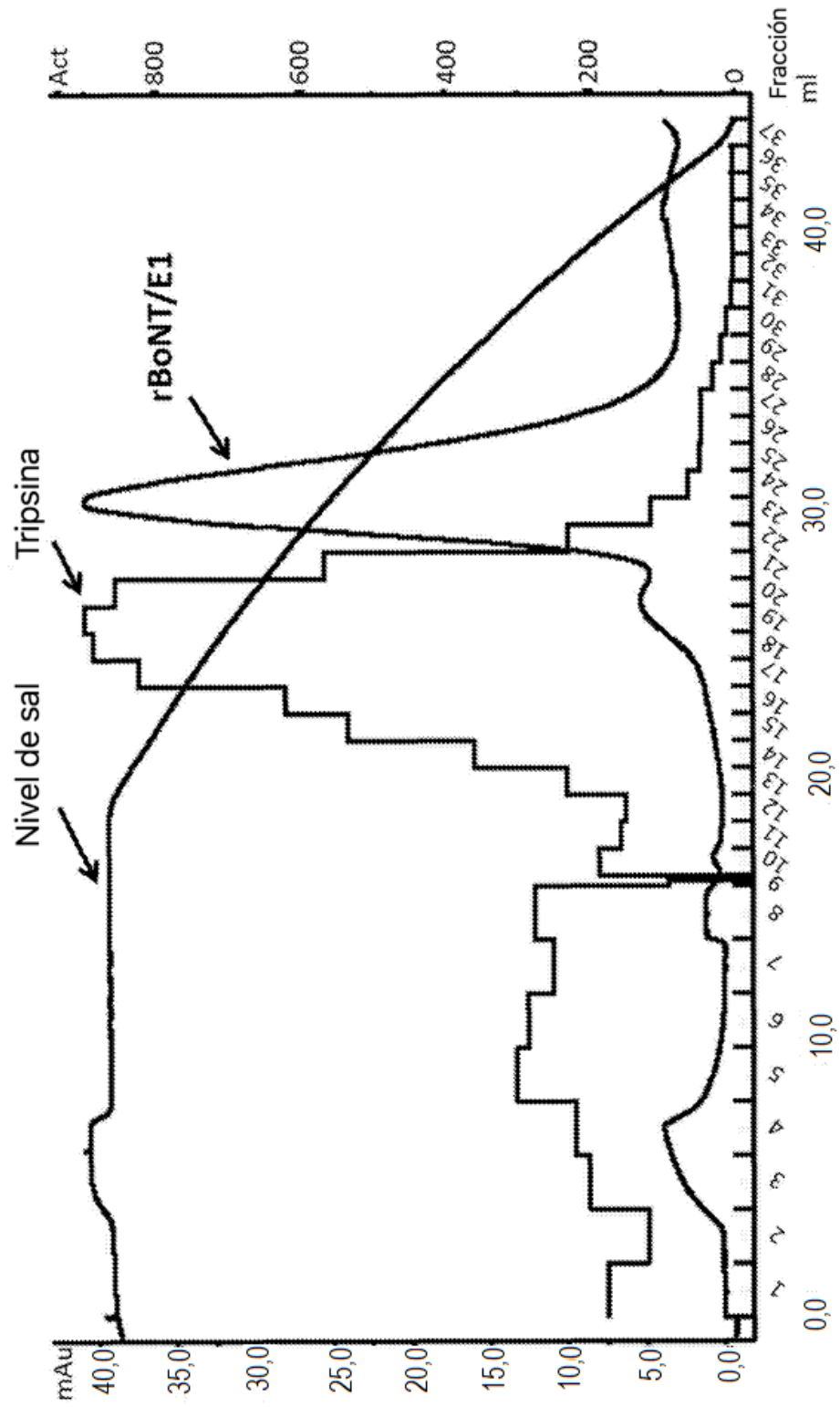


Fig. 2B

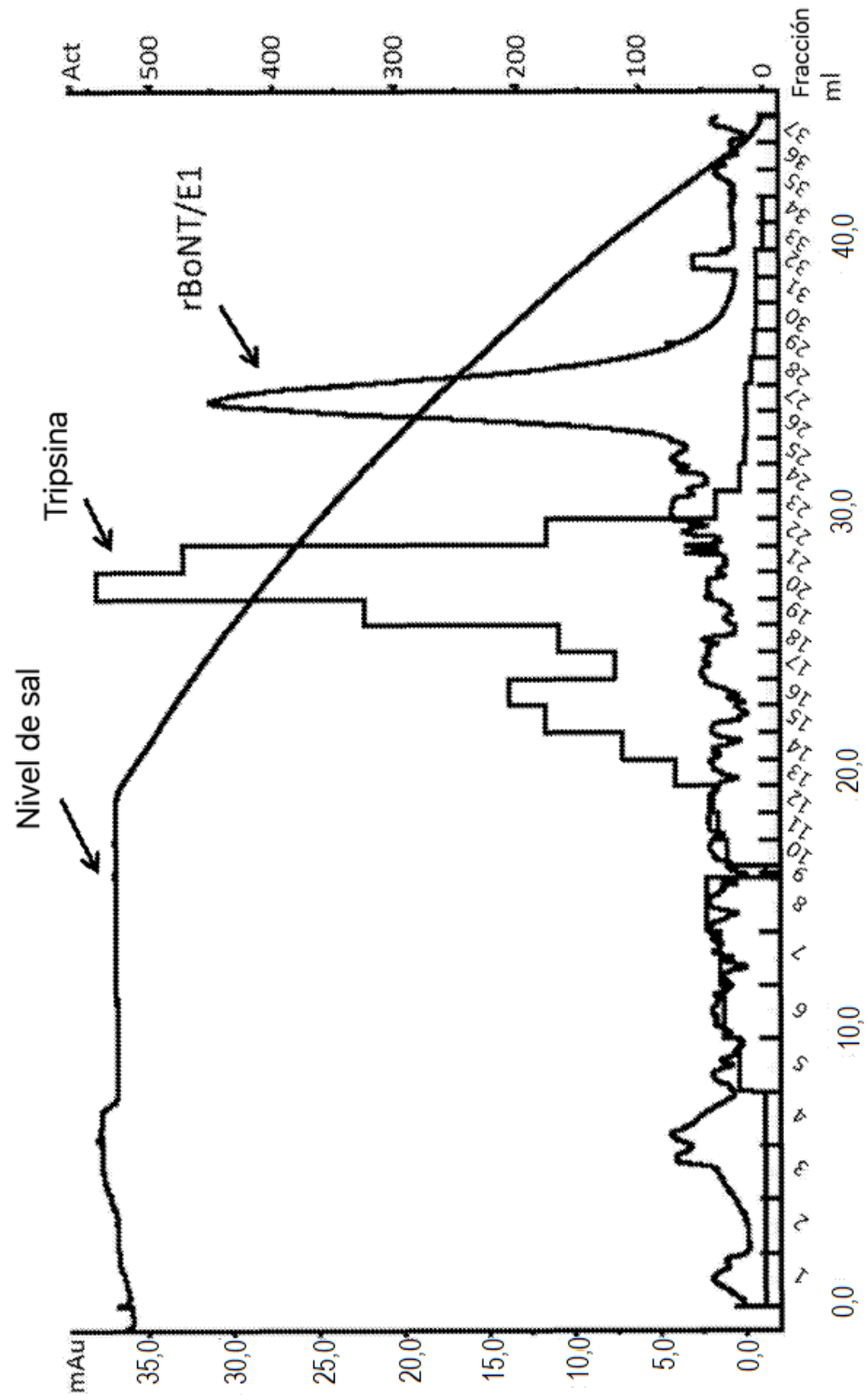


Fig. 2C

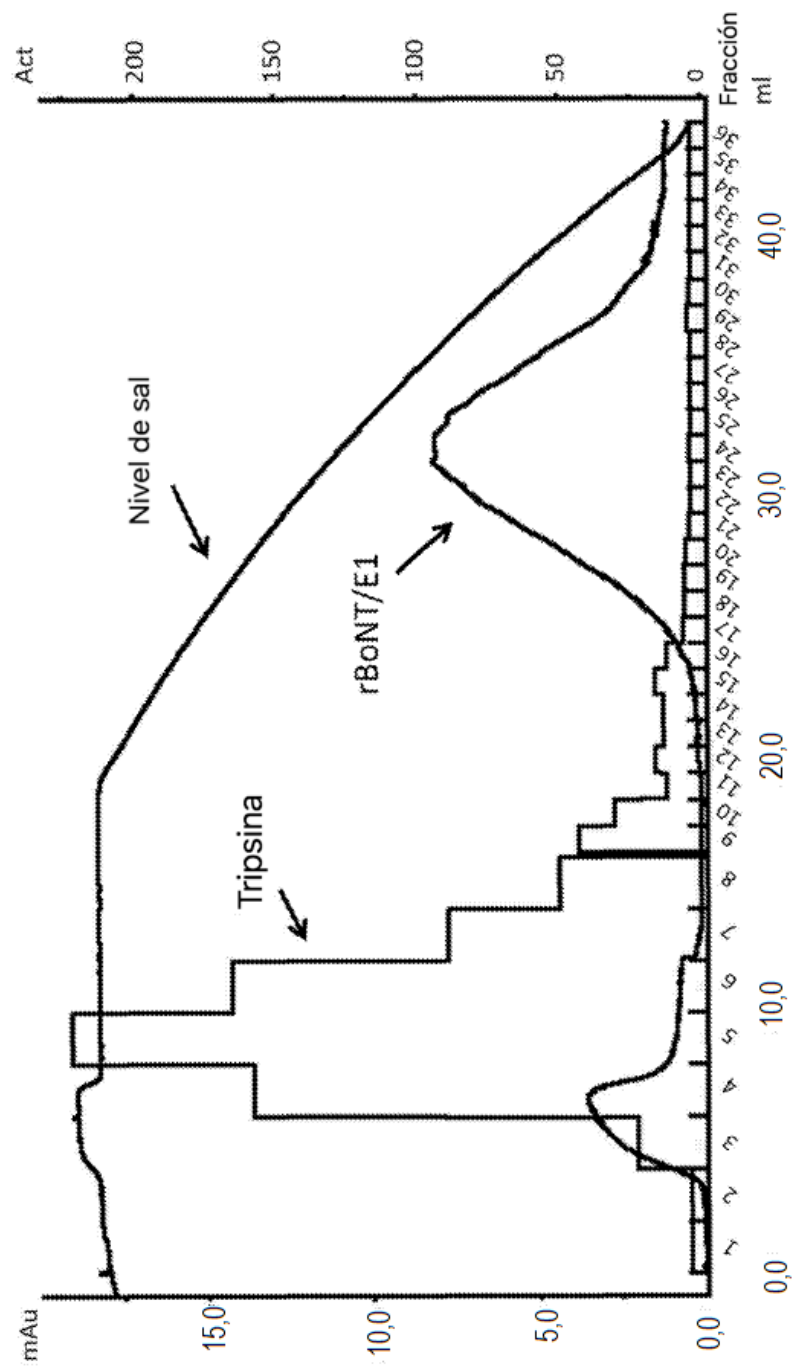


Fig. 3

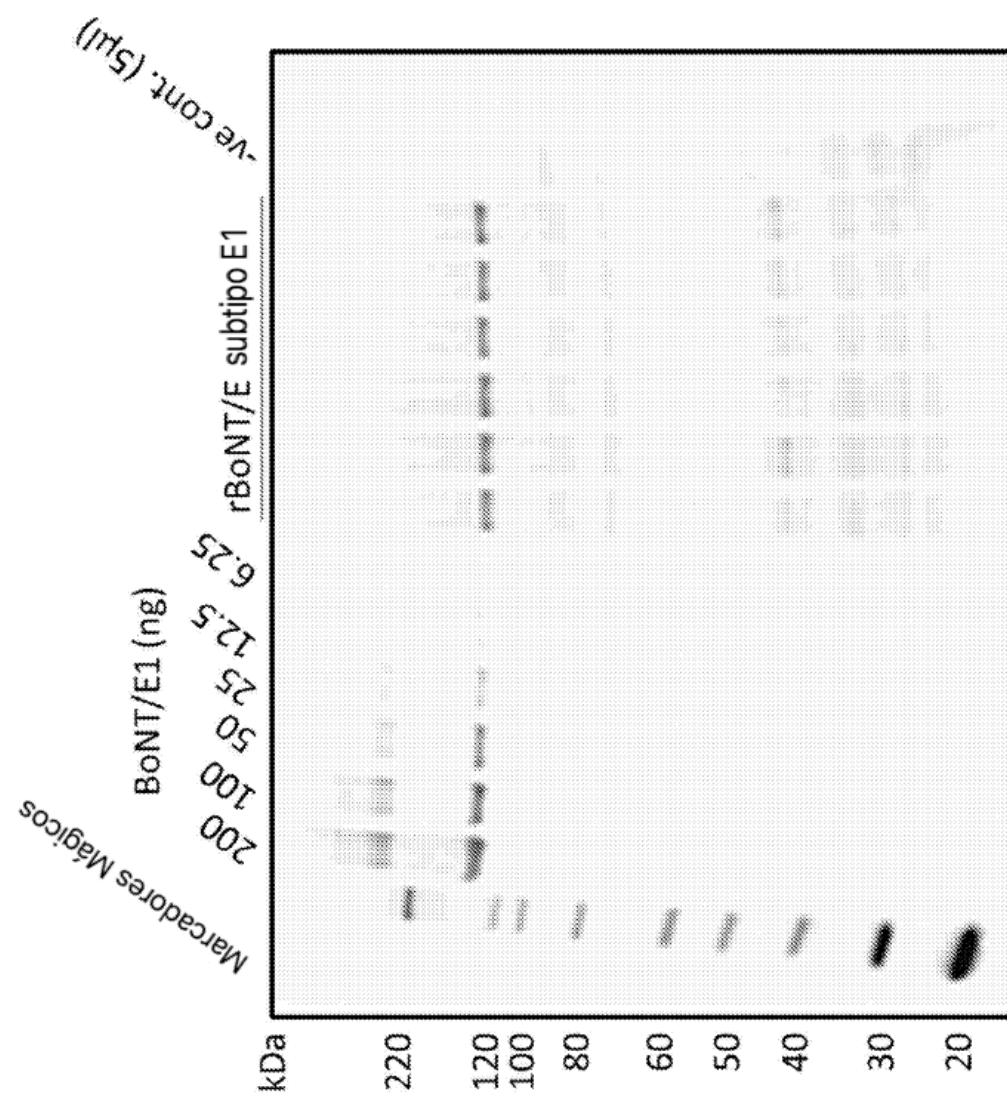


Fig. 4

