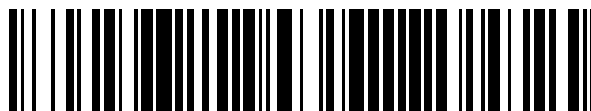


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 380 677**

51 Int. Cl.:
C12N 15/70 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06705826 .3**
96 Fecha de presentación: **20.01.2006**
97 Número de publicación de la solicitud: **1844150**
97 Fecha de publicación de la solicitud: **17.10.2007**

54 Título: **Expresión recombinante de proteínas en una forma bicatenaria, con puente disulfuro**

30 Prioridad:
21.01.2005 DE 102005002978

45 Fecha de publicación de la mención BOPI:
17.05.2012

45 Fecha de la publicación del folleto de la patente:
17.05.2012

73 Titular/es:
**MERZ PHARMA GMBH & CO. KGAA
ECKENHEIMER LANDSTRASSE 100
60318 FRANKFURT AM MAIN, DE**

72 Inventor/es:
SPECHT, Volker

74 Agente/Representante:
Carpintero López, Mario

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).

ES 2 380 677 T3

DESCRIPCIÓN

Expresión recombinante de proteínas en una forma bicatenaria, con puente disulfuro

Un objeto de la presente invención se refiere a un procedimiento para la producción de proteínas en forma bicatenaria por medio de expresión recombinante en células huésped de *E. coli*. Otro aspecto de la presente invención se refiere a proteínas o polipéptidos en forma bicatenaria y biológicamente activa que pueden producirse por medio del procedimiento mencionado.

La ventaja esencial con respecto a proteínas/polipéptidos recombinantes correspondientes que no presentan las características según la invención consiste en que no deben tratarse con una proteasa específica para la escisión dirigida de la cadena de polipéptido, de modo que se simplifica esencialmente el procedimiento de producción. Otros objetos de la presente invención son ácidos nucleicos que codifican los polipéptidos/las proteínas según la invención, vectores que contienen tales ácidos nucleicos o secuencias de ácido nucleico, células huésped que contienen a su vez los vectores mencionados y por último preparaciones farmacéuticas que contienen las proteínas/los polipéptidos bicatenarios y biológicamente activos.

Las neurotoxinas clostridiales son fuertes inhibidores de la secreción de neurotransmisores dependientes del calcio en células neuronales. Tras la ingestión oral de toxinas botulínicas (BoNT), por ejemplo a través de alimentos en mal estado se produce un cuadro clínico denominado como botulismo, que se caracteriza por la parálisis de distintos músculos. Una parálisis de la musculatura respiratoria puede llevar en último lugar a la muerte del afectado. A este respecto se interrumpe la transmisión de señales del nervio al músculo en la placa terminal motora, dado que las neuronas motoras ya no pueden distribuir nada de acetilcolina. Las toxinas botulínicas muestran su efecto inhibitor mediante la escisión proteolítica de proteínas que participan en los procesos de secreción, las denominadas proteínas SNARE. A este respecto las neurotoxinas de distintos serotipos tienen diferentes especificidades con respecto a las proteínas SNARE y los sitios de escisión en las respectivas secuencias de aminoácidos. BoNT(A) y BoNT(E) escinden la proteína SNARE SNAP-25, mientras que BoNT(C) reconoce tanto SNAP-25 como syntaxina-1 como sustrato. Por último, las toxinas de los serotipos B, D, F y G así como la toxina tetánica (TeNT) escinden VAMP-2 (sinaptobrevina-2) (Schiavo y col., 1997).

En el caso de las neurotoxinas clostridiales se trata de los venenos conocidos más potentes. Así, la dosis letal administrada por vía intravenosa, a la que mueren de botulismo la mitad de todos los ratones del grupo de dosis, asciende únicamente a 5 pg. El hecho de que las toxinas de la mayoría de los serotipos actúen también por vía oral, ha de atribuirse a proteínas complejas en las que están encapsuladas y que protegen por lo tanto en el paso gastrointestinal frente a la degradación por enzimas digestivas. A ellas se atribuye también una función en la reabsorción de las toxinas por el epitelio del intestino delgado (Fujinaga, 1997).

En el transcurso de las últimas décadas se han aprovechado terapéuticamente las toxinas botulínicas de los serotipos A y B. Así, mediante la inyección controlada de sólo pequeñas dosis se relajan músculos individuales, crónicamente contraídos. Una ventaja especial es la larga duración de efecto de por ejemplo BoNT(A) y BoNT(B) durante tres a seis meses. Las primeras áreas de indicación han sido, entre otras, distonías tales como tortícolis, blefaroespasmos y estrabismo, han de sumarse otras tales como la hiperhidrosis o tratamientos cosméticos para el suavizado de arrugas. El mercado para la toxina botulínica como agente terapéutico crece rápidamente, no en último término por el desarrollo de otras áreas de indicación y el aprovechamiento intensivo en aplicaciones ya existentes. A este respecto se están haciendo esfuerzos para mejorar las propiedades de las neurotoxinas con respecto a la duración de acción, la fuerza de acción y el potencial antigénico. Ensayos han mostrado que las proteínas complejas que están contenidas en las preparaciones comercialmente disponibles hasta el momento (BOTOX de Allergan y Dysport de Ipsen-Beaufort como preparaciones de BoNT(A) así como Myobloc/Neurobloc de Elan como preparación de BoNT(B)), no tienen ningún efecto positivo sobre la duración y la intensidad de acción, pero debido a la mayor cantidad de proteína en comparación con una preparación de la neurotoxina pura con la misma actividad, pueden contribuir a desencadenar una reacción inmunitaria en el paciente que puede hacer ineficaces inyecciones adicionales.

Es decir, dado que las proteínas complejas no se necesitan en la formulación de principio activo e incluso son desventajosas y algunas modificaciones para la mejora de las propiedades sólo pueden conseguirse mediante rutas de ingeniería genética, existe una gran necesidad de producir las neurotoxinas de manera recombinante, por ejemplo mediante expresión en *Escherichia coli* (las neurotoxinas así generadas están libres de las proteínas complejas mencionadas anteriormente). Además se explotarán nuevas áreas de indicación de manera que se confiera a las toxinas botulínicas una especificidad celular distinta. Para ello se preferirá también la ruta a través de una toxina recombinante o un derivado de toxina.

Las toxinas botulínicas así como la toxina tetánica presentan altas homologías en su secuencia de aminoácidos y se asemejan especialmente en su estructura de dominios. Están compuestas por un dominio de unión a receptor (H_C), un dominio de translocación (H_N) y una subunidad catalítica (L), que en la célula nerviosa provoca la escisión de la proteína SNARE correspondiente. H_C es responsable de la unión específica de las neurotoxinas a las placas terminales motoras, mientras que el dominio de translocación se ocupa de que L llegue desde los endosomas hasta el citoplasma de las neuronas. H_N (extremo N terminal) y H_C (extremo C terminal) forman la cadena pesada de 100

kDa, mientras que L es la cadena ligera y constituye la subunidad catalítica de 50 kDa. Ambas cadenas de polipéptido están unidas entre sí mediante un puente disulfuro. Entre los restos de cisteína implicados se extiende una región de conector o de bucle (sinónimos de ellos también secuencia de conector o secuencia de bucle o de forma simplificada conector o bucle), cuya longitud entre las toxinas botulínicas de los serotipos individuales varía considerablemente. A más tardar durante la liberación de las toxinas de los clostridios en el curso de la lisis celular se escinde el bucle mediante una endopeptidasa clostridial no caracterizada hasta el momento, variando la relación de especies escindidas y no escindidas entre los serotipos. Para la actividad de las neurotoxinas, es esencial la escisión del bucle con respecto a la toxina bicatenaria (Schiavo y col., 1997). Por ejemplo, en el caso de la neurotoxina botulínica A se corta un decapeptido del bucle, es decir, en la secuencia de bucle VRGIITSKTKSLDKGYNKALNDL, que tiene de manera N terminal así como C terminal un resto de cisteína como vecinos inmediatos, no se corta sólo una unión peptídica, sino que se corta dos veces de forma proteolítica. Por lo tanto el peso molecular de la neurotoxina botulínica A biológicamente se encuentra por naturaleza por debajo del de la toxina originalmente traducida por clostridios.

Dado que la proteasa clostridial no está presente en otros organismos huésped tales como *Escherichia coli*, en ellos se expresan toxinas botulínicas recombinantes y sus fragmentos o derivados como polipéptidos monocatenarios. Esto es válido de manera correspondiente también para cualquier otra proteína que ejerza su actividad biológica/bioquímica normal como proteína bicatenaria: en general tales proteínas se obtienen por medio de la tecnología de ADN recombinante como proteínas monocatenarias, su actividad biológica/bioquímica, que ejercen de manera natural como proteínas bicatenarias, no está por lo tanto apenas presente o no lo está en absoluto.

Para generar una proteína activa, especialmente una toxina botulínica activa, es necesaria por lo tanto hasta el momento la incorporación de una secuencia de reconocimiento para una proteasa específica de secuencia tal como por ejemplo trombina, factor Xa o genenasa, de modo que tras la purificación puede realizarse una escisión y con ello una activación mediante la adición de una de estas endoproteasas. Esto se describe por ejemplo en el documento WO 2004/024909. El uso de una de tales endoproteasas tiene esencialmente dos ventajas: por un lado no siempre puede descartarse que existan aún otros sitios de escisión en la secuencia de aminoácidos además del sitio de escisión insertado por ingeniería genética. Asimismo, cuando en estos sitios de escisión secundarios se desea cortar de manera esencialmente ineficiente, tras el tratamiento con proteasa puede generarse una mezcla de distintas variantes de escisión de la toxina que sólo se separaría con dificultad. Por otro lado, en el caso de preparaciones farmacéuticas es considerablemente desventajoso por motivos de la legislación de fármacos (regulatorios), añadir posteriormente una proteína o dejar entrar en contacto la preparación con una proteína adicional, dado que en el tratamiento adicional debe detectarse la eliminación completa de esta proteína y sus impurezas eventualmente presentes, lo que por regla general está relacionado con un gasto considerable.

Una activación mediante escisión proteolítica en un polipéptido bicatenario con puente disulfuro es necesaria también en el caso de otras toxinas bacterianas, tales como por ejemplo la exotoxina de *Pseudomonas* o toxina diftérica, para que el dominio enzimático pueda ejercer el efecto tóxico (por ejemplo mediante ADP-ribosilación de un factor de elongación y con ello la inhibición de la síntesis de proteína). Estas toxinas se usan en la producción de las denominadas inmunotoxinas que se usan especialmente en terapia antitumoral. Para ello se cambia el dominio de unión celular de la toxina por un dominio de proteína que presenta una alta afinidad de unión hacia una proteína de superficie específica de tumor (antígeno de diferenciación o antígeno asociado a tumor). Mientras que en las inmunotoxinas clásicas este dominio de proteína consiste en un anticuerpo monoclonal o un fragmento del mismo, la especificidad por determinadas células tumorales puede conferirse también mediante citocinas, factores de crecimiento así como proteínas mutadas y seleccionadas de la familia de las afilinas, proteínas de repetición de anquirina o anticalinas, por mencionar sólo algunos ejemplos. En el caso de la expresión recombinante de proteínas de fusión de este tipo se obtienen polipéptidos monocatenarios. Mientras que por ejemplo ricina no tiene ningún sitio de procesamiento para proteasas además del de *Ricinus communis* y debe introducirse uno tal, los fragmentos de toxina diftérica y exotoxina de *Pseudomonas* se escinden como componente de las inmunotoxinas tras la internalización en el compartimento endosómico mediante una proteasa de la célula diana. Esto se produce en la región de bucle entre los restos de cisteína que forman un puente disulfuro. No obstante no se procesan de este modo todas las moléculas de inmunotoxina internalizadas, sino sólo una pequeña parte (Ogata y col., 1990).

Para obtener proteínas recombinantes, especialmente polipéptidos de menor tamaño, en cantidades suficientes y en forma soluble, es en muchos casos necesario expresar los mismos en *E. coli* como proteína de fusión o proteína híbrida con por ejemplo glutatión-S-transferasa o proteína de unión a maltosa. Asimismo en el mercado se encuentran numerosos sistemas de expresión, mediante los que se expresa el polipéptido deseado con una marca N terminal o C terminal para la purificación por afinidad tal como la marca de His, marca Strep o marca FLAG. Con frecuencia se encuentra en el plásmido de expresión entre los múltiples sitios de clonación, en los que se inserta la secuencia DNS que codifica por la proteína deseada, y la secuencia codificante para el componente de fusión o la marca de afinidad, una secuencia de reconocimiento de proteasa. Ésta debe posibilitar que tras la expresión y purificación de la proteína de fusión pueda liberarse por la región peptídica adicional la proteína deseada mediante la adición de una endoproteasa específica de secuencia correspondiente (por ejemplo trombina, factor Xa o genenasa). Si los dos componentes de fusión no estuvieran unidos covalentemente entre sí mediante una unión peptídica, sino a través de un puente disulfuro, tras la purificación podría tener lugar una separación de uno de otro mediante una reducción sencilla con sustancias que contienen tiol tales como β -mercaptoetanol, DTT o glutatión

reducido. Por ejemplo, la proteína deseada podría eluirse por una matriz de afinidad tal como por ejemplo Ni-NTA-agarosa o StrepTactin-Sepharose con los agentes de reducción mencionados, mientras que las marcas de afinidad permanecen unidas a la matriz. Por lo tanto podría no tener lugar una etapa de purificación adicional para la separación de la marca de afinidad o de una endoproteasa añadida.

5 Sería deseable por lo tanto un procedimiento para la expresión recombinante de proteínas/polipéptidos en general, especialmente de neurotoxinas así como de fragmentos y derivados de estas neurotoxinas, y de proteínas de fusión o proteínas híbridas, especialmente de inmunotoxinas, que ya tras la lisis de las células huésped existen en su estructura bicatenaria (biológicamente) activa, estando unidas las dos cadenas por puentes disulfuro. Un procedimiento de este tipo para la producción de tales proteínas y polipéptidos se proporciona por la invención descrita en el presente documento.

10 Sorprendentemente el inventor ha descubierto ahora que el fragmento LH_N de BoNT(A) tal como también la neurotoxina A completa, que se obtienen de forma bicatenaria de manera recombinante, ejercen su actividad biológica/bioquímica normal también en forma bicatenaria, con puente disulfuro, se obtienen de manera recombinante en forma bicatenaria, el fragmento LH_N o la toxina completa, preferentemente al nivel de ácido nucleico, se someten al menos a una modificación determinada tal como se representa en la reivindicación 1. Ensayos posteriores del inventor han dado como resultado que lo correspondiente vale también para cualquier otra proteína/polipéptido, siempre que se obtengan en forma monocatenaria según procedimientos recombinantes convencionales, pero ejerzan su actividad biológica en forma bicatenaria, con puente disulfuro.

20 En el caso de la mencionada "al menos una modificación", en el caso de BoNT(A) o en el caso del fragmento LH_N de BoNT(A), se trata de la incorporación (inserción) de una secuencia de pentapéptido denominada en el presente documento como PRS (por *protease recognition site* (sitio de reconocimiento de proteasa)) tal como se representa en la reivindicación 1. En el caso general de la proteína/del polipéptido, una secuencia de pentapéptido, que está presente en la proteína/el polipéptido que va a modificarse, (preferentemente al nivel de ácido nucleico) puede modificarse alternativamente (por ejemplo mediante al menos un cambio de un resto de aminoácido o mediante inserción de sólo algunos restos de aminoácido de PRS o mediante delección de restos de aminoácido), que corresponde a la secuencia de pentapéptido insertada mediante la inserción en la secuencia existente. Del mismo modo puede insertarse también una secuencia de hexa/hepta/octa (etc.) péptido, con o sin que deban deleccionarse uno o dos o tres o varios restos de aminoácido. Según la invención es ventajoso únicamente que el polipéptido expresado en último lugar presente la secuencia (de pentapéptido) PRS su región de bucle, estando definida la región de bucle según la invención como la secuencia de aminoácidos que se encuentra entre los dos restos de cisteína que participan en el puente disulfuro. Si esta secuencia PRS se encuentra en la región de bucle, esto lleva a que en el caso de la escisión del polipéptido monocatenario, en las proximidades a la secuencia de polipéptido PRS (al nivel de aminoácido) se encuentren también en dos cadenas distintas las secuencias que se encuentra de manera natural en dos cadenas distintas. En el caso de la neurotoxina botulínica A (BoNT(A) estas secuencia PRS se introduce en el bucle preferentemente con la delección del pentapéptido Asp₄₄₃ - Asn₄₄₇ de BoNT(A) (véase la figura 3-1). En el caso de otras proteínas/polipéptidos (por ejemplo en el caso de BoNT(B), BoNT(C1), BoNT(D), BoNT(E), en el caso de ricina, en el caso de PE40 de la exotoxina de *Pseudomonas* o en el caso de la toxina diftérica (DT)) se prefiriere por el contrario introducir un bucle modificado de BoNT(A) (véanse las figuras 3-2 a 3-5) en la secuencia de bucle (pudiendo deleccionarse restos de aminoácido de la secuencia de bucle natural o también no). Las secuencias de bucle modificadas en las figuras 3-2 a 3-5 son las secuencias respectivas sin los dos restos de Cys de extremo, pudiendo ser el aminoácido central de la secuencia PRS no sólo R, Y, H o Q, sino cualquiera de los aminoácidos de origen natural. En el caso de las otras proteínas/polipéptidos recién mencionadas se prefiere especialmente introducir sólo una parte del bucle modificado de BoNT(A), especialmente la secuencia GIITSKTSLVXPXGSKALNDL (X = un aminoácido de origen natural), en la secuencia de bucle (pudiendo deleccionarse restos de aminoácido de la secuencia de bucle natural o también no). Las secuencias de bucle modificadas en las figuras 3-2 a 3-5 son las secuencias respectivas sin los dos restos de Cys de extremo.

50 Para el fragmento LH_N de BoNT(A) o para la toxina recombinante completa quiere decir que la modificación de secuencia es una modificación en la región de bucle entre L y H_N que proporciona la presencia de una secuencia PRS. La secuencia PRS es según la invención, y no sólo para BoNT(A), la secuencia de pentapéptido Val-Pro-Xaa-Gly-Ser. Xaa representa cualquier aminoácido de origen natural. No importa si Xaa es Arg u otro aminoácido de origen natural, la secuencia de pentapéptido Val-Pro-Xaa-Gly-Ser se denomina en cada caso como secuencia de pentapéptido PRS. Por el contrario si uno de los otros cuatro restos de aminoácido de la secuencia PRS está cambiado, lo que puede ser de enteramente en el contexto de la invención, especialmente por restos polares o no polares hidrófilos/hidrófobos de manera correspondiente, se habla en este caso y en adelante de una variante de la secuencia de pentapéptido PRS. Las variantes se producen por ejemplo cuando Val se sustituye por Leu, Ile, Ala, Phe, Pro o Gly. Además se producen variantes cuando (también o sólo) prolina en la segunda posición de PRS, visto desde el extremo N terminal, se sustituye por, por ejemplo, Leu, Ile, Ala, Phe, Val o Gly. También la glicina en la cuarta posición de PRS, por último, puede sustituirse por ejemplo por Leu, Ile, Ala, Pro, Phe o Val, lo que lleva a su vez a otras variantes. Y cuando la serina en la quinta posición de PRS se sustituye por ejemplo por Tyr, Trp o Thr, dado el caso también por Cys o Met, se produce un tipo adicional de variantes. Es decir, según la invención se denominan como variantes de la secuencia de pentapéptido aquellas secuencias que en al menos una de las posiciones 1, 2, 4 y 5 de la secuencia PRS contienen un resto de aminoácido que es diferente de Val-1, Pro-2, Gly-4 y/o Ser-5.

Si el fragmento LH_N de BoNT(A) (o la toxina completa) o cualquier otra proteína/polipéptido, que se obtiene normalmente de manera recombinante como proteína/polipéptido monocatenario, está contenido en forma biológicamente/bioquímicamente activa pero (sólo) en forma bicatenaria, la secuencia de pentapéptido Val-Pro-Xaa-Gly-Ser (en la que Xaa representa cualquiera de los 20 aminoácidos de origen natural, y también pueden cambiarse los otros cuatro restos de aminoácido en el sentido del párrafo anterior) se encuentra en el lisado de las células huésped de *E. coli* (por ejemplo K12 de *E. coli*, especialmente células huésped K12 de *E. coli* de las cepas M15[pREP4], XL1-BLUE o UT5600) en la forma bicatenaria, en la que en el caso de BoNT(A) la cadena ligera permanece covalentemente unida mediante un puente disulfuro con H_N o con la cadena pesada completa (figura 7). La escisión de la cadena de polipéptido tiene lugar o bien inmediatamente tras la lisis celular o se concluye en gran parte tras una incubación de varias horas del lisado celular. Puede descartarse una autoproteólisis por la actividad del dominio de proteasa de la toxina o fragmento de toxina, dado que se producen también en la estructura bicatenaria mutantes inactivos por proteasa y modificados de manera correspondiente en la región de bucle tras la expresión y digestión de las células huésped de *E. coli*. Evidentemente una proteasa de la cepa huésped de *E. coli* es responsable de la escisión de la secuencia de pentapéptido PRS.

Una modificación según la invención adicional según el párrafo "Sorprendentemente el inventor ha" cuatro párrafos más arriba (en la página 5) consiste en que de manera N terminal de la secuencia PRS a una distancia de 1 a 20 restos de aminoácido (la dirección de aminoácido del extremo N terminal, que se encuentra inmediatamente adyacente al resto de valina de la secuencia de pentapéptido PRS, en el caso de las figuras 3-2 a 3-5 un resto de leucina, tiene una distancia de 1 resto de aminoácido con respecto a la secuencia PRS), especialmente a una distancia de 3 a 15 restos de aminoácido, especialmente a una distancia de 3 a 10 restos de aminoácido, de manera especialmente preferente a una distancia de 3 a 8 restos de aminoácido, y de manera muy especialmente preferente a una distancia de 3 restos de aminoácido, un resto de aminoácido básico, se encuentra preferentemente un resto de lisina o de arginina, en cuyo extremo C terminal la proteasa de la célula huésped de *E. coli* escinde la secuencia de bucle. Es decir, tras la escisión se obtiene un polipéptido que presenta por ejemplo dos restos de aminoácido (cuando la distancia recién definida asciende a 3 restos de aminoácido) de manera N terminal con respecto al resto de valina de la secuencia PRS. En el presente caso "variación" no significa necesariamente una variación en el sentido real, es decir por ejemplo una inserción o sustitución de un resto de aminoácido, de modo que a continuación de manera N terminal de la secuencia PRS a la distancia indicada anteriormente de 1 a 20 restos de aminoácido se encuentra un resto de aminoácido básico (por ejemplo un resto de lisina). Tan sólo es decisivo que se encuentre un resto de aminoácido básico (tal como un resto de lisina o de arginina) de manera N terminal de la secuencia PRS a la distancia indicada anteriormente.

Una variación adicional, igualmente no necesaria, pero preferida según el párrafo "Sorprendentemente el inventor ha" cinco párrafos más arriba consiste en que la secuencia de bucle, en la que escinde la proteasa de las células huésped de *E. coli*, tienen una longitud de al menos nueve restos de aminoácido. Longitudes preferidas de la secuencia de bucle son al menos doce, al menos 15, al menos 18, al menos 20 y al menos 23 restos de aminoácido. Longitudes especialmente preferidas de la secuencia de bucle son de 1,5 a 22, especialmente de 18 a 22 restos de aminoácido.

Es decir el procedimiento según la invención es muy en general un procedimiento para la producción de proteínas/polipéptidos en forma bicatenaria, en la que las dos cadenas están unidas por puente disulfuro, por medio de expresión recombinante en células huésped de *E. coli*, en el que (i) la proteína/el polipéptido ejerce su actividad biológica como proteína/polipéptido bicatenario, con puente disulfuro, (ii) el resto de aminoácido C terminal de la primera cadena es un resto de Arg o Lys, (iii) la segunda cadena de la proteína/del polipéptido, de manera N terminal con respecto a un resto de cisteína como extremo N terminal presenta de 1 a 20 restos de aminoácido y una secuencia de pentapéptido VPXGS denominada como PRS, en la que X significa cualquier aminoácido de origen natural, en la que V significa Val, Leu, Ile, Ala, Phe, Pro o Gly, en la que P significa Pro, Leu, Ile, Ala, Phe, Val o Gly, en la que G significa Gly, Leu, Ile, Ala, Pro, Phe o Val y en la que S significa Ser, Tyr, Trp o Thr; y (iv) el procedimiento comprende las siguientes etapas: (a) modificar la proteína/del polipéptido, al nivel de ácido nucleico, de modo que la proteína/el polipéptido en su forma modificada en su región de bucle presenta la secuencia de pentapéptido mencionada (VPXGS); (b) introducir el constructo modificado al nivel de ácido nucleico en células de *E. coli*; (c) cultivar y a continuación lisar las células huésped; y (d) aislar la proteína/el polipéptido bicatenario.

La primera cadena de la proteína/del polipéptido es según la invención de manera preferente la cadena que se codifica por el extremo N terminal del ADN correspondiente, mientras que la segunda cadena de la proteína/del polipéptido es por consiguiente la cadena que se codifica por el extremo C terminal del ADN correspondiente. Es decir, dado que la expresión de 5'-ADN-3' lleva a N-polipéptido-C, quiere decir que para caso de la invención preferido mencionado anteriormente que la expresión puede representarse tal como sigue: 5'-ADN-3' expresa para dar N-primer cadena de polipéptido-C-bucle-N-segunda cadena de polipéptido-C. El bucle se escinde según la presente invención ya *in situ*, de modo que por último se obtiene el polipéptido/la proteína según la invención N-primer cadena de polipéptido-C-N-segunda cadena de polipéptido-C en estructura bicatenaria.

Con la expresión "la segunda cadena de la proteína/del polipéptido presenta de manera N terminal con respecto a un resto de cisteína como extremo N terminal de 1 a 20 restos de aminoácido y una secuencia de pentapéptido VPXGS denominada como PRS" se quiere expresar que el extremo N terminal no se forma por ejemplo por el resto de valina de la secuencia de pentapéptido VPXGS, sino por otro resto de aminoácido (cualquiera). Entre éste y el

resto de valina de PRS pueden encontrarse entonces aún de 1 a 19 restos de aminoácido más, pero el resto de aminoácido N terminal puede unirse por ejemplo con el resto de valina también directamente por medio de unión peptídica, es decir, estar directamente adyacente al resto de valina de PRS.

Las proteínas/polipéptidos según la invención que pueden aislarse en su estructura bicatenaria (biológicamente) activa son proteínas cuyo extremo C terminal de la primera cadena es un resto de aminoácido básico, especialmente un resto de Arg o de Lys, y cuya segunda cadena presenta de manera N terminal de 1 a 20 restos de aminoácido y la secuencia de pentapéptido VPXGS denominada como PRS, en la que X, V, P, G y S se definen tal como anteriormente.

Según de la presente invención, por ejemplo en el caso de las inmunotoxinas que se basan en ricina recombinante, no es necesario un tratamiento mediante una proteasa específica de secuencia tal como trombina o factor Xa para la activación. En el caso de por ejemplo inmunotoxinas a base de toxina diftérica o exotoxina de *Pseudomonas* era de esperar un claro aumento de eficacia, que también se consigue realmente, dado que ya no es necesario el procesamiento mediante una proteasa de la célula diana como etapa determinante de la velocidad para la translocación del dominio enzimático de las toxinas en el citoplasma. Las inmunotoxinas de este tipo, que se encuentran ya como polipéptidos bicatenarios, con puente disulfuro, pueden usarse en pequeñas dosis y no obstante conseguir el mismo efecto citotóxico. Esto baja por un lado los costes de la terapia, por otro lado reduce el riesgo de formación de anticuerpos que podían hacer ineficaz a la inmunotoxina en otras aplicaciones. Un procedimiento para la producción de inmunotoxinas bicatenarias, con puente disulfuro y por lo tanto activadas se proporciona por la presente invención.

Con el procedimiento proporcionado según la invención pueden producirse también proteínas de fusión o proteínas híbridas, es decir por ejemplo proteínas con una marca peptídica para la purificación por afinidad, en una forma bicatenaria, cuyas dos cadenas de polipéptido están unidas covalentemente por un puente disulfuro y pueden separarse una de otra según un método de cromatografía de afinidad u otro método de purificación mediante reducción simple con sustancias que contienen tiol tales como β -mercaptoetanol, DTT o glutatión reducido.

La expresión recombinante de neurotoxinas clostridiales y sus fragmentos (por ejemplo un fragmento LH_N o un derivado de una neurotoxina clostridial, por ejemplo especificidad celular modificada) en cepas de expresión de *E. coli* tales como M15[pREP4] o BL21(DE3) lleva a polipéptidos monocatenarios. Mediante el tratamiento de estos polipéptidos con tripsina tiene lugar una escisión en la región de la secuencia de bucle en la región de transición del dominio de proteasa al dominio de translocación. Dado que en el caso de tripsina no se trata de una proteasa específica de secuencia, es probable una escisión, en la mayoría de los casos indeseada, en otras regiones de los polipéptidos. Así BoNT(A) se escinde mediante tripsina adicionalmente entre H_N y H_C, generándose un fragmento bicatenario LH_N y un fragmento H_C. Para garantizar una escisión selectiva deseada en la mayoría de los casos en la región de bucle, es necesaria la presencia, dado el caso tras su introducción, de una secuencia de reconocimiento para endoproteasas específicas.

La escisión de proteínas de fusión/proteínas híbridas producidas de manera recombinante por medio de endoproteasas específicas de secuencia tales como trombina, factor Xa, genenasa etc. pertenece al espectro de métodos generalmente conocidos. Así es posible separar tras la purificación un componente de fusión que ayuda a la proteína/el polipéptido recombinante a conseguir una solubilidad y/o expresión mejorada o que sirve como marca peptídica para la purificación por afinidad. Para ello se incubaba la disolución de proteína con la endoproteasa adecuada en forma soluble o inmovilizada en una matriz.

Esta técnica puede aprovecharse también en el caso de la expresión de las proteínas/los polipéptidos recombinantes mencionados anteriormente que ejercen su actividad biológica/bioquímica normal como proteínas/polipéptidos bicatenarios, por medio de la tecnología de ADN recombinante pero se obtienen como proteínas/polipéptidos monocatenarios inactivos (por ejemplo en el caso de la expresión de neurotoxinas clostridiales, fragmentos de neurotoxinas clostridiales tales como fragmentos LH_N o de derivados de neurotoxinas clostridiales, por ejemplo con especificidad celular modificada): en el polipéptido, preferentemente al nivel de los ácidos nucleicos, se clona por ejemplo en la región de bucle entre L y H_N una secuencia de reconocimiento para una endoproteasa, y además se clona en el polipéptido en el extremo N o C terminal una secuencia de reconocimiento adicional para la misma u otra endoproteasa, flanqueada por una marca peptídica para la purificación por afinidad. La proteína/el polipéptido expresado de forma monocatenaria se activa entonces mediante tratamiento con la o las endoproteasa(s) correspondientes al mismo tiempo o de forma sucesiva mediante la escisión en la región de bucle entre L y H_N y se libera por la marca peptídica.

A excepción de los costes del uso de tales endoproteasas y las etapas de trabajo adicionales relacionadas con el mismo, su uso en preparaciones farmacéuticas (por ejemplo en el caso de uso de toxinas botulínicas recombinantes o sus derivados) es sumamente problemático por motivos de la legislación de fármacos (regulatorios). Por un lado la pureza de la endoproteasa usada debe estar experimentalmente comprobada, por otro lado su agotamiento completo y especialmente la ausencia de virus de la preparación en el transcurso adicional del protocolo de purificación ha de documentarse con la máxima exactitud, lo que por regla general significa un enorme esfuerzo analítico. Dado que en el futuro se producirán también toxinas botulínicas por ejemplo con propiedades mejoradas o especificidades celulares modificadas en rutas recombinantes, existe una gran necesidad de un procedimiento de

expresión que posibilite la provisión de las proteínas/los polipéptidos recombinantes mencionados anteriormente, que ejercen su actividad biológica/bioquímica normal como proteínas bicatenarias/ polipéptidos, por medio de la tecnología de ADN recombinante pero que se obtienen como proteínas/polipéptidos monocatenarios, especialmente la provisión de toxinas botulínicas o sus derivados, como polipéptidos/proteínas bicatenarios, con puente disulfuro y por lo tanto biológicamente activos, sin que deban usarse endoproteasas.

La invención explicada en detalle en las secciones siguientes proporciona por tanto en el sentido más amplio procedimientos con los que pueden expresarse proteínas tales como neurotoxinas clostridiales así como sus fragmentos y derivados de manera recombinante en células huésped de *E. coli* y pueden aislarse en su forma bicatenaria, con puente disulfuro y por lo tanto biológicamente activa, sin que para su activación se requiera una endoproteasa añadida.

En una primera forma de realización preferida de la invención, la secuencia de aminoácidos de la región de bucle de BoNT(A) está modificada entre los restos de cisteína 430 y 454 (véanse las figuras 3-1 a 3-5) de manera que la toxina expresada, o sus fragmentos/derivados, se encuentran en el lisado de las células huésped de *E. coli* ya como polipéptidos bicatenarios. A este respecto las dos cadenas están unidas covalentemente entre sí a través de puente disulfuro con la participación de los restos de cisteína 430 y 454. En una forma de realización especialmente preferida de la invención, tal como se explica en la figura 3, puede sustituirse el pentapéptido Asp₄₄₃ - Asn₄₄₇ (DKGYN) por Val-Pro-Arg-Gly-Ser (VPRGS). En formas de realización preferidas adicionales de la invención el pentapéptido Asp₄₄₃ - Asn₄₄₇ (DKGYN) se sustituye también por Val-Pro-Tyr-Gly-Ser (VPYGS), Val-Pro-His-Gly-Ser (VPHGS) o Val-Pro-Gln-Gly-Ser (VPQGS). En este caso es también válido que no sólo el resto de aminoácido central puede ser un aminoácido de origen natural cualquiera, sino también, que los otros cuatro restos de aminoácido pueden estar asimismo cambiados, tal como se explicó adicionalmente con todo detalle anteriormente (con el cambio de al menos uno de estos restos se produce una variante de la secuencia PRS en el sentido de la invención). Además para éstos es válido tal como para todas las formas de realización preferidas adicionales, que se describen a continuación, que se prefiere adicionalmente que la secuencia de bucle presente de manera N terminal con respecto a la PRS a una distancia de 1 a 20 aminoácidos un resto de aminoácido básico, especialmente un resto de lisina o de arginina.

Para el experto es sin más evidente que cambios adicionales de restos de aminoácido individuales o varios restos de aminoácido o la inserción o delección de restos de aminoácido adicionales en la región del bucle caracterizado anteriormente de BoNT(A) conducen asimismo al resultado de que la toxina expresada según la invención o los fragmentos/derivados que se derivan de la misma se encuentran el lisado como polipéptidos bicatenarios. Estas variantes posibles son asimismo objeto de la invención.

Pero igualmente para el experto es sin más evidente que el pentapéptido Asp₄₄₃ - Asn₄₄₇ (DKGYN) presente en el tipo natural de BoNT(A) puede sustituirse por un hexapéptido, por un heptapéptido, por un octapéptido, etc., siempre que en el polipéptido/proteína expresado y traducido en forma monocatenaria la secuencia de pentapéptido PRS o una de sus variantes concebibles se encuentre en la región de bucle. Tal como ya se estableció además anteriormente se prefiere que de manera N terminal con respecto al pentapéptido esté presente un resto de aminoácido básico (preferentemente lisina).

El experto apreciará además que en el caso de la forma de realización preferida del pentapéptido (Val-Pro-Arg-Gly-Ser) de PRS se trata de una parte de una secuencia de reconocimiento posible para la proteasa trombina, que desempeña un papel importante en la cascada de coagulación sanguínea y presenta una alta especificidad de secuencia. Se indica expresamente que, 1. ni en el caso de la neurotoxina botulínica de tipo A ni en el caso de otros polipéptidos es necesaria una escisión mediante trombina para obtener la forma bicatenaria, con puente disulfuro deseada, y que 2. si bien la secuencia de reconocimiento de trombina como tal, es decir su forma no modificada, es conveniente para la escisión mediante la actividad proteasa del lisado de *E. coli*, en cambio no es en absoluto necesaria. Las formas de realización de las secuencias de pentapéptido PRS insertadas en los polipéptidos correspondientes (mejor dicho: en sus bucles) o generadas en los mismos, que no contienen el resto de arginina (sino en su lugar otro de los aminoácidos de origen natural), en su extremo C terminal puede escindir trombina, llevan, tal como ya se mencionó, asimismo a la escisión en el bucle. A este respecto la escisión tiene lugar de manera preferente en un resto de lisina del bucle de manera N terminal con respecto al pentapéptido, tal como anteriormente se explicó adicionalmente (véase también el ejemplo 2; figura 3).

Dado que también otros serotipos de las toxinas botulínicas tales como BoNT(B) y BoNT(C1) pueden aprovecharse terapéuticamente como neurotoxina de acción prolongada y BoNT(E) como neurotoxina de acción corta, y otros polipéptidos/proteínas muy diferentes, que se generan en forma monocatenaria de manera recombinante, pero sólo desarrollan actividad biológica en forma bicatenaria, sería deseable que estas neurotoxinas así como fragmentos o derivados de las mismas (y también los polipéptidos/proteínas distintos) puedan obtenerse asimismo como polipéptidos/proteínas bicatenarios, con puente disulfuro, a partir de lisados de *E. coli*. Especialmente en el caso de BoNT(B) una escisión completa de la toxina recombinante en el lisado de *E. coli* en el polipéptido/proteína bicatenario traería una clara ventaja con respecto a la neurotoxina nativa secretada en *Clostridium botulinum*, que por regla general se encuentra en al menos el 40 % como polipéptido monocatenario y por lo tanto inactivo y que no puede separarse de la forma activa bicatenaria. Es sorprendente que las regiones de bucle de las neurotoxinas de los serotipos B, C1 y E son claramente más cortas entre los restos de cisteína que participan en el puente disulfuro

con respecto al bucle de BoNT(A) (figuras 3 y 4). Mientras que en el caso de BoNT(A) son 23 restos de aminoácido (Val₄₃₁ - Leu₄₅₃), en el caso de BoNT(B) se encuentran únicamente 8 (Lys₄₃₈ - Ile₄₄₅), en el caso de BoNT(C1) 15 (His₄₃₈ - Asp₄₅₂) y en el caso de BoNT(E) 13 restos de aminoácido (Lys₄₁₃ - Ile₄₂₅) en esta región. No obstante se mostró, a excepción de BoNT(B), que estas regiones acortadas en comparación son suficientemente largas para permitir la escisión de la cadena y la formación de puentes disulfuro, cuando presentan la secuencia PRS según la invención. Aunque también la BoNT(B) en el caso de un cambio de un pentapéptido en el bucle mediante una secuencia de pentapéptido PRS (con ello la longitud total de la secuencia de bucle es de sólo ocho restos de aminoácido) en el sentido de la invención se escindió en dos cadenas (ligera y pesada), se obtuvieron sin embargo mejores resultados, es decir, según la invención se prefiere tener un bucle de al menos 9, al menos 15, al menos 20 o incluso al menos 22 restos de aminoácido. Una de las últimas formas de realización mencionadas, en la que el bucle presenta 22 restos de aminoácido, puede deducirse a modo de ejemplo de las secuencias en las figuras 4-1 y 4-2 o una comparación entre ambas.

Por lo tanto se comprobó también experimentalmente que se preferiría un cambio de las regiones de bucle en los subtipos B, C1, etc., o partes esenciales de las mismas, frente a la región de bucle de BoNT(A), o partes esenciales de la misma, lo que afecta a la escisión de las neurotoxinas para dar polipéptidos/proteínas bicatenarias, con puente disulfuro, especialmente cuando mediante esta medida el bucle se alarga al menos 9, mejor 15 restos y/o de manera N terminal de PRS podía introducirse un resto de aminoácido básico (por ejemplo y preferentemente un resto de Lys) (siempre que previamente no hubiera tampoco ningún resto de Lys o básico N terminal). Se preferían especialmente variaciones, tal como se representan en la figura 4 (siendo las secuencias PRS en la figura 4 VPRGS, igualmente, y se prefieren de igual modo también las secuencias VPYGS, VPHGS, VPQGS, VPKGS, VPIGS y VPAGS).

En otras formas de realización de la invención las secuencias de aminoácidos y los segmentos génicos que codifican las mismas de las regiones de bucle en las toxinas botulínicas de los serotipos B, C1, D, E, F y G así como de toxina tetánica están modificadas entre los restos de cisteína que participan en el puente disulfuro entre L y H_N de tal manera que las toxinas expresadas o los fragmentos/derivados derivados de las mismas en el lisado de las células huésped de *E. coli* se encuentran ya como polipéptidos bicatenarios, en los que las dos cadenas están unidas covalentemente a través de un puente disulfuro (de manera correspondiente es válido también para cualquier otro polipéptido/proteína que se genere en forma monocatenaria de manera recombinante, pero sólo muestre actividad biológica en forma bicatenaria). En formas de realización preferidas de la invención las regiones de bucle completas (o partes de las mismas) de las neurotoxinas o de los fragmentos/derivados de toxina que se derivan de las mismas pueden cambiarse por la región de bucle completa de BoNT(A), tal como se caracteriza en la figura 3, o partes de la región de bucle de BoNT(A), en las que el pentapéptido Asp₄₄₃ - Asn₄₄₇ se sustituye preferentemente por ejemplo por Val-Pro-Arg-Gly-Ser (VPRGS). En formas de realización preferidas adicionales de la invención el pentapéptido Asp₄₄₃ - Asn₄₄₇ puede sustituirse también por Val-Pro-Tyr-Gly-Ser, Val-Pro-His-Gly-Ser o Val-Pro-Gln-Gly-Ser. En formas de realización especialmente preferidas de la invención las regiones de bucle o partes de las regiones de bucle de las neurotoxinas mencionadas y de los fragmentos/derivados que se derivan de las mismas pueden sustituirse por el oligopéptido Arg/Ser-Gly-Ile-Ile-Thr-Ser-Lys-Thr-Lys-Ser-Leu-Val-Pro-Arg-Gly-Ser-Lys-Ala (18-mero: R/SGIITSKT KSLVPRGSKA). Otros cambios, inserciones o deleciones de restos de aminoácido individuales o de varios restos de aminoácido en la región de la secuencia de bucle caracterizada anteriormente, tal como puede verse por ejemplo en la figura 4, y que asimismo llevan a que la neurotoxina expresada o su fragmento/derivado tras la expresión en *E. coli* (por ejemplo en células huésped K12 de *E. coli* o sus derivados) se encuentre como polipéptido/proteína bicatenario, con puente disulfuro, expresamente con un componente de esta invención (de manera correspondiente es válido también para cualquier otro polipéptido/proteína que se generen en forma monocatenaria de manera recombinante, pero que sólo muestren actividad biológica en forma bicatenaria).

Tal como ya se explicó repetidamente, con el procedimiento según la invención según una forma de realización adicional de la invención pueden producirse también proteínas de fusión o proteínas híbridas que presentan por ejemplo los siguientes componentes A, B y C:

- un dominio efector que, mediante su actividad enzimática puede por ejemplo inhibir la secreción en células diana o destruir las mismas (A);
- una secuencia de bucle que está modificada según la invención, tal como se describió anteriormente, y que presenta la secuencia de pentapéptido PRS VPXGS definida anteriormente (por ejemplo una de las secuencias de bucle de BoNT(A) modificadas, representadas en la figura 3 o variantes de las mismas) y a la que puede añadirse de manera N y/o C terminal un resto de cisteína (B), así como
- un dominio de unión celular que confiere a la proteína de fusión/híbrida una especificidad celular (C).

El componente B (secuencia de bucle) puede ser en cambio en las dos formas de realización mencionadas de manera inmediatamente anterior preferente igualmente (i) una de las secuencias de bucle modificadas representadas en la figura 4, (ii) cualquier secuencia derivada de las mismas, como el resto central de PRS puede ser el resto de un aminoácido cualquiera de origen natural, o (iii) una variante (para la definición de variante véase anteriormente) de (i) o (ii). En la figura 4 se delecionaron las secuencias de bucle de BoNT(B), BoNT(C1) o BoNT(E) respectivas hasta uno o dos restos de aminoácido N terminales y los dos C terminales y los restos de aminoácido delecionados se sustituyeron por el 17-mero GIITSKTKSLVPRGSKA (figuras 4-2 y 4-6) o el 18-mero RGIITSKTKSLVPRGSK A (figura 4-4) a partir de la secuencia de bucle modificada de BoNT(A).

Además de los componentes A, B y C mencionados anteriormente las proteínas de fusión/proteínas híbridas pueden presentar un dominio de translocación (que en el caso de las neurotoxinas botulínicas se encuentra entre la secuencia de bucle y el dominio de unión celular). Este dominio adicional es de ayuda en la inclusión del dominio efector en el citoplasma de la célula diana. La expresión de proteínas de fusión de este tipo en *E. coli* (por ejemplo K12 de *E. coli* o derivados de la misma) lleva a polipéptidos/proteínas bicatenarios en los que un dominio se encuentra en una cadena y los dos otros dominios se encuentran en la segunda cadena (en el caso de las toxinas botulínicas el dominio efector está unido covalentemente en la cadena ligera a través de un puente disulfuro con los otros dos dominios en la cadena pesada).

En el caso de estas proteínas de fusión o proteínas híbridas según la invención puede tratarse de las denominadas inmunotoxinas que se emplean especialmente en la terapia antitumoral. A este respecto se confiere a un dominio de toxina mediante la adición de un dominio de unión celular una especificidad por un tipo celular determinado, por regla general una célula tumoral. Como dominios de toxina se usan sobre todo los dominios enzimáticos de toxina diftérica, exotoxina de *Pseudomonas* y ricina. Estas toxinas pertenecen a las toxinas AB bicatenarias, en las que la cadena A contiene la actividad enzimática, está unida mediante un puente disulfuro covalentemente con la cadena B, que reúne la actividad de translocación y de unión celular. Pueden concebirse también otras toxinas o fragmentos de toxina en inmunotoxinas, siempre que desarrollen el efecto deseado (por ejemplo la destrucción de células tumorales) en las células diana. Mientras que la primera generación de inmunotoxinas se produjo mediante acoplamiento químico de un dominio de toxina tal como por ejemplo la cadena A de ricina con un anticuerpo monoclonal, las inmunotoxinas de la segunda generación se expresaron principalmente en *E. coli* de manera recombinante como toxinas Fab, toxinas Fv monocatenarias (toxinas scFv) o toxinas Fv estabilizadas por disulfuro (toxinas dsFv), pero también como proteínas de fusión con factores de crecimiento o citocinas (Reiter, 2001). En generaciones futuras de inmunotoxinas la especificidad celular puede conferirse también por polipéptidos modificados y seleccionados tras unión de alta afinidad por ejemplo a una proteína de superficie específica de tumor por ejemplo de las familias de proteínas de las afilinas, proteínas de repetición de anquirina o anticalinas.

En todas las variantes concebibles de las inmunotoxinas debe estar garantizado que el dominio enzimático de toxina llega al citoplasma de la célula diana, para poder desplegar allí el efecto tóxico. Dado que las inmunotoxinas se expresan en *E. coli* como polipéptidos monocatenarios, son necesarias una escisión proteolítica así como una reducción de un puente disulfuro para separar en cadenas el dominio enzimático de toxina de la unidad de translocación y el dominio de unión celular. En el caso del fragmento de toxina diftérica recombinante y del fragmento de exotoxina de *Pseudomonas* recombinante tiene lugar una escisión tras la internalización en el compartimento endosómico de la célula diana mediante una proteasa celular tal como furina (Williams y col., 1990). Por el contrario, la ricina no tiene ningún sitio de procesamiento de este tipo y necesita por lo tanto una secuencia de reconocimiento de proteasa introducida artificialmente, para poder aplicarse ya como una inmunotoxina bicatenaria, con puente disulfuro. Pero también en el caso de las inmunotoxinas que se basan en toxina diftérica y exotoxina de *Pseudomonas*, sólo se escinde una pequeña parte de las proteínas de fusión internalizadas, de modo que también sólo una parte igualmente pequeña de los dominios enzimáticos puede llegar al citoplasma (Ogata y col., 1990). Las formas de realización preferidas de la invención presentadas a continuación describen procedimientos y constructos, pudiendo expresarse con los procedimientos, de manera recombinante en células huésped de *E. coli*, variantes de inmunotoxinas, tal como se describieron en las secciones anteriores, y aislarse en su forma activa bicatenaria, con puente disulfuro y por lo tanto biológicamente (enzimáticamente) activa, sin que para su activación necesite una endoproteasa celular o una endoproteasa añadida *in vitro*. Estas inmunotoxinas pueden llevar los dominios enzimáticos de toxina en una forma competente de translocación a la célula diana, de modo que no es necesaria una escisión mediante una proteasa celular y deben usar esencialmente menores dosis de la inmunotoxina para alcanzar los mismos efectos citotóxicos deseados.

Una forma de realización preferida adicional de la invención comprende por consiguiente adicionalmente una proteína de fusión o proteína híbrida que presenta los siguientes componentes A, B y C:

- un dominio de toxina o su fragmento/derivado (A),
- una secuencia de bucle, que según la invención, tal como se describió anteriormente, está modificada y presenta la secuencia de pentapéptido PRS VPXGS definida anteriormente (por ejemplo una de las secuencias de bucle de BoNT(A) modificadas representadas en la figura 3 o variantes de las mismas) y puede añadirse de manera N y/o C terminal un resto de cisteína (B), así como
- un dominio de unión celular que puede derivarse de un representante de las familias de proteínas de los anticuerpos monoclonales, de sus fragmentos, de las afilinas, de las proteínas de repetición de anquirina, de las anticalinas, de los factores de crecimiento (por ejemplo TGF-alfa, FGF, VEGF o IGF-1) o de las citocinas (por ejemplo IL2, IL4 o IL6) (C).

El componente B (secuencia de bucle) puede ser según esta última forma de realización preferida igualmente (i) una de las secuencias de bucle modificadas representadas en la figura 4, (ii) cualquier secuencia derivada de las mismas, como el resto central de PRS puede ser el resto de un aminoácido de origen natural, o (iii) una variante (para la definición de variante véase anteriormente) de (i) o (ii).

El dominio de toxina puede ser la cadena A de ricina, un fragmento de la exotoxina de *Pseudomonas* tal como PE40 o PE38 (dominios II y III con o sin comprender el dominio Ib; figura 2) o un fragmento de la toxina diftérica. Los

dominios efector, de toxina y de unión celular mencionados han de entenderse como ejemplos. La invención incluye todas las proteínas o fragmentos de proteína, que por un lado confieren a la proteína de fusión/proteína híbrida una actividad de unión específica en un antígeno superficial de una célula diana, por ejemplo de una célula tumoral, y por otro lado en una célula diana tras la internalización ejercen un efecto determinado, por ejemplo la destrucción de la célula, llevando la expresión de tales proteínas de fusión/proteínas híbridas según la invención en *E. coli* a polipéptidos/proteínas bicatenarios, en los que el dominio de toxina o derivados del mismo está covalentemente unido mediante un puente disulfuro con el dominio de unión celular.

Para mejorar la eficacia y especificidad de inmunotoxinas a base de exotoxina de *Pseudomonas* se seleccionaron en el pasado distintos planteamientos. Así se cambió por ejemplo el dominio de unión a receptor (dominio Ia con los restos de aminoácido 1 - 152) por fragmentos de un anticuerpo monoclonal y al mismo tiempo se modificó la región de bucle (figura 2 y 5) en el dominio de translocación (dominio II) entre los restos de cisteína 13 y 35 (numeración con respecto al dominio II) de tal manera que éste ya no era sensible a la escisión de la proteasa celular ubicua furina, sino que en su lugar a proteasas especiales que sólo se expresan por células tumorales determinadas en mayor medida y se secretan parcialmente (patente estadounidense 6.426.075). Esta sensibilidad a proteasa modificada debería conferir a la inmunotoxina además del dominio de unión a receptor cambiado una especificidad celular elevada. No obstante no se puede suponer que mediante otras proteasas celulares se produce una escisión aumentada en el bucle y con ello una eficacia de translocación mejorada del dominio enzimático III.

En un planteamiento adicional para una inmunotoxina se eliminan el dominio de unión a receptor y la región N terminal del dominio de translocación hasta el resto de arginina 27 en la región de bucle. La especificidad celular necesaria se confería a una inmunotoxina de este tipo por ejemplo mediante inserción de un dominio V_H de un anticuerpo monoclonal, al que estaba unido a través de un puente disulfuro el dominio V_L , en el sitio del dominio Ib entre los dominios II y III o mediante inserción en el extremo C terminal del dominio III (patente estadounidense 5.980.895). En tales constructos ya no es necesaria una activación mediante una proteasa, lo que por un lado provocaría una eficacia de translocación claramente elevada. No obstante ha de temerse que por otro lado la translocación se impida mediante los dominios de unión a receptor que se encuentran de manera N o C terminal del dominio enzimático III tal como el dominio V_H de un anticuerpo monoclonal o TGF- α . Dado que estos dominios de unión a receptor no se separan del dominio enzimático, ha de contarse también con efectos negativos sobre la actividad enzimática y por lo tanto la toxicidad en las células diana. Un máximo relativo en actividad citotóxica se obtiene con una inmunotoxina a base de exotoxina de *Pseudomonas*, cuando por un lado el bucle entre los restos de cisteína 13 y 35 se encuentra ya durante la aplicación en la forma escindida, con puente disulfuro y por lo tanto no es necesaria una activación mediante una proteasa celular, por otro lado el dominio de unión a receptor en el sitio del dominio I de la exotoxinas está fusionado al extremo N terminal del dominio de translocación, de modo que se separa de los dominios de toxina tras la reducción en el citoplasma y por lo tanto no puede verse afectada la actividad enzimática del dominio III.

Una forma de realización especialmente preferida de la invención comprende por lo tanto una proteína de fusión/proteína híbrida que comprende un dominio de unión celular que puede tomarse de un representante de las familias de proteínas de los anticuerpos monoclonales, de sus fragmentos, de las afilinas, de las proteínas de repetición de anquirina, de las anticalinas, de los factores de crecimiento (por ejemplo TGF- α , FGF, VEGF o IGF-1) o de las citocinas (por ejemplo IL2, IL4 o IL6), al que está fusionado de manera C terminal un fragmento PE38 modificado que puede portar en C terminal la señal de retención para el retículo endoplasmático, Lys-Asp-Glu-Leu, o variantes de la misma. La modificación del fragmento PE38 consiste en que la secuencia de bucle completa (o también sólo una parte de la misma) entre los restos de cisteína 13 y 35 se cambió por la secuencia de pentapéptido PRS VPXGS, preferentemente por la secuencia de bucle de BoNT(A) modificada, representada en la figura 3 o de la misma, especialmente por la secuencia peptídica Arg-Gly-Ile-Ile-Thr-Ser-Lys-Thr-Lys-Ser-Leu-Val-Pro-Arg-Gly-Ser-Lys-Ala (figura 5) (con respecto a la definición de las variantes, véase anteriormente). Preferentemente también en esta forma de realización se ocupa de que de manera N terminal con respecto a PRS a una distancia de 1 a 20 restos de aminoácido se encuentre un resto de aminoácido básico, tal como en la secuencia representada en la figura 5. Un fragmento PE-38 modificado de manera correspondiente así como proteínas de fusión/proteínas híbridas, que contiene este fragmento modificado, se encuentran en el lisado de las células huésped de *E. coli* (por ejemplo M15 [pREP4]) en la forma bicatenaria, con puente disulfuro.

A diferencia del caso de la exotoxina de *Pseudomonas* el dominio enzimático de la toxina diftérica, la cadena A, se encuentra en el extremo N terminal. En la cadena B C terminal se encuentran los dominios de translocación y de unión a receptor. Ambas cadenas están unidas mediante una secuencia de bucle, en la que en el resto de arginina 193 durante la secreción a partir de células de *Corynebacterium diphtheriae* mediante una proteasa tiene lugar una escisión proteolítica (Collier, 2001). Las dos cadenas permanecen unidas entre sí covalentemente tras la escisión mediante un puente disulfuro entre los restos de cisteína 186 y 201. En este sentido la toxina diftérica se parece en su estructura de dominios a las toxinas botulínicas y a la toxina tetánica.

Para la producción de inmunotoxinas recombinantes se cambió el dominio de unión a receptor o una parte del mismo por ejemplo por VEGF o IL2 (Arora y col., 1999; Williams y col., 1990), para conferir a la proteína de fusión una nueva especificidad celular. Para que la cadena A pueda llegar al citoplasma de las células diana, por un lado debe escindirse la cadena de polipéptido de la inmunotoxina expresada en forma monocatenaria en *E. coli* en la región del bucle entre la cadena A y B, por otro lado reducirse el puente disulfuro. Mientras que esto último tiene

lugar en el transcurso del proceso de translocación, la escisión proteolítica mediante una proteasa celular es incompleta, de modo que sólo puede liberarse un pequeño porcentaje de la cadena A en el citoplasma (Williams y col., 1990). Si la inmunotoxina se encontrara ya durante la aplicación en la forma bicatenaria, con puente disulfuro, cabría esperar un claro aumento de eficacia, dado que todas las cadenas A se proporcionarían en una forma competente de translocación.

Una forma de realización especialmente preferida adicional de la invención comprende por lo tanto una proteína de fusión o proteína híbrida que comprende un dominio de unión celular que puede tomarse de un representante de las familias de proteínas de los anticuerpos monoclonales, de sus fragmentos, de las afilinas, de las proteínas de repetición de anquirina, de las anticalinas, de los factores de crecimiento (por ejemplo TGF-alfa, FGF, VEGF o IGF-1) o de las citocinas (por ejemplo IL2, IL4 o IL6), al que está fusionado de manera N terminal un fragmento de toxina diftérica modificado. Este fragmento de toxina puede comprender la cadena A así como al menos el dominio de translocación de la cadena B (Gly₁ - Phe₃₈₉ o Gly₁ - Asn₄₈₆). La modificación del fragmento de toxina diftérica consiste en que la secuencia de bucle completa (o también sólo una parte de la misma) entre los restos de cisteína 186 y 201 se cambió por la secuencia de bucle de BoNT(A) modificada representada en la figura 3 o variantes de la misma, especialmente por la secuencia peptídica Arg-Gly-Ile-Ile-Thr-Ser-Lys-Thr-Lys-Ser-Leu-Val-Pro-Arg-Gly-Ser-Lys-Ala (figura 5) (con respecto a la definición de las variantes, véase anteriormente). Un fragmento de toxina diftérica modificado de manera correspondiente así como proteínas de fusión que contienen este fragmento modificado, se encuentran en el lisado de las células huésped de *E. coli* tal como por ejemplo M15[pREP4] en la forma bicatenaria, con puente disulfuro.

Las inmunotoxinas a base de ricina de la primera generación se producían mediante la unión de la cadena A de la ricina con un anticuerpo monoclonal. Esto sucedía hasta el momento mediante derivatización del anticuerpo con una molécula de conector químico, que con la función tiol del resto de cisteína que se encuentra en el extremo C terminal de la cadena A formaba un puente disulfuro. Los conjugados de este tipo eran heterogéneos debido a la derivatización no controlada del anticuerpo. La eficacia contra tumores resultó no ser suficiente, no en último lugar, debido al tamaño del conjugado y a la ausencia del dominio de translocación localizado en la cadena B. Si la cadena B en la forma nativa es también componente de la inmunotoxina, es decir se aumenta la toxicidad claramente, se produce en cambio debido a las propiedades de unión celular de tipo lectina de la cadena B para la captación no específica también en otras distintas de las células diana deseadas. Este conflicto de diana se encontró con una estrategia en la que la cadena B se modificó de tal manera que si bien se mantenía la actividad de translocación, en cambio se aumentaba la afinidad de unión para glicoestructuras en las superficies celulares (solicitud de patente WO 89/04839). Las inmunotoxinas expresadas de manera recombinante que contienen una cadena B modificada de este tipo, están en cambio en forma monocatenaria, de modo que debido a la ausencia de secuencia de reconocimiento para una proteasa celular en el péptido conector entre la cadena A y la cadena B no es posible o sólo de manera muy ineficaz una liberación y translocación de la cadena A en caso de la captación de la inmunotoxina en la célula diana. En la patente estadounidense 6.593.132 se documentan modificaciones de este péptido conector nativo, que representan las secuencias de reconocimiento para proteasas diferentes, específicas de células. Las variantes de ricina con modificaciones de este tipo presentarán una especificidad celular correspondiente, siempre que la proteasa respectiva, que puede escindir proteolíticamente el péptido conector modificado, sólo se expresa en las células diana correspondientes en comparación con otros tipos celulares en cantidades significativamente elevadas. No obstante puede suponerse que la escisión tiene lugar sólo en una fracción de la molécula de toxina internalizada y por lo tanto también sólo puede translocarse una cantidad correspondientemente pequeña de cadenas A en el citoplasma. Serían deseables inmunotoxinas bicatenarias a base de ricina, en las que la cadena A está unida a través de un puente disulfuro con una cadena B modificada, en la que se mantiene la actividad de translocación, que contienen en cambio las propiedades de unión celular inespecífica, de tipo lectina, y que están fusionadas en su extremo C terminal con un dominio de unión celular específico. Las inmunotoxinas de este tipo reunirían especificidad celular y alta toxicidad.

Aún una forma de realización especialmente preferida adicional de la invención comprende por lo tanto una proteína de fusión que presenta los siguientes componentes A, B y C:

- la cadena A de la ricina (A),
- una secuencia de bucle, que según la invención, tal como se describió anteriormente, está modificada y que presenta la secuencia de pentapéptido PRS VPXGS definida anteriormente (por ejemplo una de las secuencias de bucle de BoNT(A) modificadas, representadas en la figura 3, o variantes de las mismas) y a la que puede añadirse de manera N y/o C terminal un resto de cisteína (B), así como
- un dominio de unión celular que puede tomarse de un representante de las familias de proteínas de los anticuerpos monoclonales, de sus fragmentos, de las afilinas, de las proteínas de repetición de anquirina, de las anticalinas, de los factores de crecimiento (por ejemplo TGF-alfa, FGF, VEGF o IGF-1) o de las citocinas (por ejemplo IL2, IL4 o IL6) (C).

El componente B puede ser según esta última forma de realización preferida igualmente (i) una de las secuencias de bucle modificadas representadas en la figura 4, (ii) cualquier secuencia derivada de las mismas, como el resto central de PRS puede ser el resto de un aminoácido de origen natural cualquiera, o (iii) una variante (para la definición de variante véase anteriormente) de (i) o (ii).

Especialmente la secuencia de bucle puede presentar la secuencia peptídica Ala-Pro-Pro-Arg-Gly-Ile-Ile-Thr-Ser-Lys-Thr-Lys-Ser-Leu-Val-Pro-Arg-Gly-Ser-Lys-Ala-Asp-Val (figura 5-6), es decir puede ser un bucle modificado de la cadena A de la ricina. En la secuencia de bucle se coloca de manera preferente adicionalmente de manera terminal un resto de cisteína. En la secuencia PRS Val-Pro-Arg-Gly-Ser contenida en la misma Arg puede ser sin embargo también cualquier otro aminoácido natural Xaa. A ambos lados la secuencia de bucle puede ampliarse mediante restos de aminoácido adicionales (por ejemplo restos de glicina y de serina). Además la cadena A de la ricina puede estar unida con la cadena B completa o partes o variantes de la misma a través de una secuencia de bucle, que sustituye completa o parcialmente los restos de aminoácido entre los restos de cisteína 259 y 283 de la secuencia de tipo natural de la Pro-ricina y comprende al menos la región del bucle de BoNT(A) modificado, descrito en la figura 3 o variantes de la misma. A este respecto se forma un puente disulfuro por los restos de cisteína 259 y 283 (con respecto a la secuencia de tipo natural de la Pro-ricina). En el extremo C terminal de la cadena B está fusionado un dominio de unión celular que puede derivarse de las familias de polipéptidos mencionados anteriormente. Las proteínas de fusión/proteínas híbridas correspondientes se encuentran en el lisado de las células huésped de *E. coli*, por ejemplo de células de la cepa M15[pREP4], en la forma bicatenaria, con puente disulfuro.

Una forma de realización adicional de la invención se refiere a proteínas de fusión recombinantes que presentan los siguientes componentes A, B y C:

- una proteína o un oligopéptido que confiere a la proteína de fusión una mejor solubilidad, provoca una mayor tasa de expresión y/o posibilita una purificación por afinidad (por ejemplo glutatión-S-transferasa (GST), proteína de unión a maltosa (MBP), marca de His, marca Strep, marca FLAG) (A),
- una secuencia de bucle, que según la invención, tal como se describió anteriormente, está modificada y que presenta la secuencia de pentapéptido PRS VPXGS definida anteriormente (por ejemplo una de las secuencias de bucle de BoNT(A) modificadas, representadas en la figura 3, o variantes de las mismas) y a la que puede añadirse de manera N y/o C terminal un resto de cisteína, así como
- un polipéptido cualquiera (C).

El componente B (secuencia de bucle) puede ser según esta última forma de realización preferida igualmente (i) una de las secuencias de bucle modificadas representadas en la figura 4, (ii) cualquier secuencia derivada de las mismas, como el resto central de PRS puede ser el resto de un aminoácido de origen natural cualquiera, o (iii) una variante (para la definición de variante véase anteriormente) de (i) o (ii).

Especialmente el bucle puede presentar la secuencia peptídica Val-Arg-Gly-Ile-Ile-Thr-Ser-Lys-Thr-Lys-Ser-Leu-Val-Pro-Arg-Gly-Ser-Lys-Ala-Leu-Asn-Asp-Leu, en la que Arg en el centro de PRS ha de entenderse de nuevo como Xaa. A ambos lados puede ampliarse mediante otros restos de aminoácido (por ejemplo restos de glicina y de serina). La expresión de tales proteínas de fusión en *E. coli* lleva a polipéptidos/proteínas bicatenarios, cuyas dos cadenas están unidas covalentemente mediante un puente disulfuro y tras tener lugar la purificación pueden separarse una de otra sin la adición de una proteasa tras la reducción simple mediante sustancias que contienen tiol (por ejemplo β -mercaptoetanol, DTT o glutatión reducido). Un sistema de expresión de este tipo es especialmente adecuado para proteínas recombinantes en las que se introducirá en uno de los dos extremos terminales un resto cisteína, para tras la purificación y separación del componente de fusión con el grupo tiol reactivo tener un punto de partida por ejemplo para reacciones de acoplamiento con moléculas de conector reactivas con tiol o modificaciones con por ejemplo polietilenglicol.

Además se proporcionan todos los ácidos nucleicos que codifican los polipéptidos según la invención descritos en las secciones anteriores considerando las distintas posibilidades de la utilización de codones. Además se proporcionan plásmidos de expresión y de clonación contruidos individualmente o comercialmente disponibles que contienen las secuencias ADN codificantes para los polipéptidos según la invención respectivos, así como cepas de expresión y de clonación adecuadas de *E. coli*, que se han transformado con los plásmidos de expresión correspondientes y que pueden expresar los polipéptidos según la invención respectivos en su forma activa, bicatenaria y con puente disulfuro. Un ejemplo de un sistema de expresión de este tipo es un plásmido de expresión de la serie pQE en conexión con la cepa huésped de *E. coli* M15[pREP4].

Para el experto en la técnica que se dedica especialmente al desarrollo de polipéptidos/proteínas que van a usarse farmacéuticamente, son claramente evidentes las ventajas que están relacionadas con que para la activación de estos polipéptidos/proteínas no debe añadirse ninguna endoproteasa. La mayor parte de los polipéptidos/proteínas según la invención descritos en las secciones anteriores son especialmente específicos para el uso farmacéutico. Por lo tanto también son parte componente de la invención preparaciones farmacéuticas que contienen uno de los polipéptidos/proteínas según la invención o una mezcla de los polipéptidos/proteínas según la invención como componente de principio activo así como aditivos útiles, que confieren a la preparación una estabilidad suficiente y cuya composición está adaptada a la presentación farmacéutica deseada.

Las figuras y secuencias adjuntas del protocolo de secuencias se describen tal como sigue:

La figura 1 muestra una representación esquemática de la liberación de neurotoxina botulínica de tipo A con bucle de tipo natural o bucle modificado según la invención de *Clostridium botulinum* o K12 de *Escherichia coli*.
A: Durante la lisis de células de *Clostridium botulinum* se escinde la neurotoxina en la región de bucle entre

cadena ligera (L) y cadena pesada (H) mediante una endoproteasa clostridial. Ambas cadenas están unidas entre sí mediante un puente disulfuro. B: Tras la expresión de una neurotoxina recombinante con bucle de tipo natural en *E. coli* y lisis de las células ésta se encuentra en la forma monocatenaria. C: durante la liberación de una neurotoxina recombinante con bucle modificado según la invención a partir de células de *E. coli* tiene lugar una escisión en la región de bucle mediante una endoproteasa.

La figura 2 muestra una representación esquemática de distintas toxinas recombinantes con regiones de bucle de tipo natural así como regiones de bucle modificadas según la invención en comparación tras su liberación a partir de células de *E. coli*. A: neurotoxina botulínica; B: exotoxina de *Pseudomonas*; C: toxina diftérica.

La figura 3 muestra una comparación del bucle de tipo natural con una selección de secuencias de bucle de BoNT(A) modificadas según la invención. Están representadas las secuencias de nucleótidos y las secuencias de aminoácidos derivadas que incluyen los restos de cisteínas limitantes de la cadena ligera y pesada. La flecha marca el sitio de escisión para la endoproteasa en el lisado de *E. coli*.

La figura 4 muestra una comparación de los bucles de tipo natural con en cada caso una secuencia de bucle modificada según la invención, a modo de ejemplo, de las neurotoxinas botulínicas de los serotipos B, C1 y E. Están representadas las secuencias de nucleótidos y las secuencias de aminoácidos derivadas que incluyen los restos de cisteína limitantes de la cadena ligera y pesada. La flecha marca el sitio de escisión para la endoproteasa en el lisado de *E. coli*.

La figura 5 muestra una comparación de los bucles de tipo natural con en cada caso una secuencia de bucle modificada según la invención, a modo de ejemplo, del fragmento PE40 de la exotoxina de *Pseudomonas*, toxina diftérica (DT) y ricina. Están representadas las secuencias de nucleótidos y las secuencias de aminoácidos derivadas que incluyen los restos de cisteína limitantes. La flecha marca el sitio de escisión para la endoproteasa en el lisado de *E. coli*.

La figura 6 muestra una lista de los oligonucleótidos que se usaron en las clonaciones de las toxinas recombinantes y fragmentos de toxina. Las secuencias de reconocimiento para endonucleasas de restricción están subrayadas.

La figura 7 muestra un análisis del fragmento LH_N recombinante de BoNT(A) con secuencia de bucle modificada según la invención en gel de SDS-poliacrilamida. La expresión del fragmento LH_N tuvo lugar en células de M15[pREP4], que se habían transformado con el plásmido pQE-BoNT(A)-L_{mod1}H_N. Carriles 2 y 5: fragmento LH_N purificado sobre Ni-NTA; Carriles 1 y 4: fragmento LH_N tras incubación con trombina; Carril 3: marcador de peso molecular. Aplicación de muestra en condiciones reductoras (carriles 1 y 2) o condiciones no reductoras (carriles 4 y 5).

La figura 8 muestra un análisis del fragmento LH_N recombinante de BoNT(B) con secuencia de bucle modificada según la invención en gel de SDS-poliacrilamida. La expresión del fragmento LH_N tuvo lugar en células de M15[pREP4], que se habían transformado con el plásmido pQE-BoNT(B)-L_{mod1}H_N. Carriles 1 y 4: fragmento LH_N purificado sobre Ni-NTA-agarosa; Carril 2: marcador de peso molecular; Carril 3: ninguna aplicación. Aplicación de muestras en condiciones reductoras (carril 1) o condiciones no reductoras (carril 4).

La figura 9 muestra un análisis de BoNT(C1) recombinante con secuencia de bucle modificada según la invención en gel de SDS-poliacrilamida. La expresión de la toxina tuvo lugar en células de M15[pREP4] que se habían transformado con el plásmido pQE-BoNT (C1)-L_{mod1}H_NH_C. Carriles 1 y 4: toxina purificada sobre Ni-NTA-agarosa; Carril 2: marcador de peso molecular; Carril 3: ninguna aplicación. Aplicación de muestras en condiciones reductoras (carril 1) o condiciones no reductoras (carril 4).

SEC ID N°:1 es un ejemplo de un ácido nucleico (ADN) que codifica por una neurotoxina botulínica de tipo A recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(A)-mod1).

SEC ID N°:2 es un ejemplo de una neurotoxina botulínica de tipo A recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(A)-mod1).

SEC ID N°:3 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento LH_N recombinante de la neurotoxina botulínica de tipo A con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(A)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:1, en la que están delecionados los nucleótidos 2620 - 3888.

SEC ID N°:4 es un ejemplo de un fragmento LH_N recombinante de la neurotoxina botulínica de tipo A con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(A)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:2, en la que están delecionados los restos de aminoácido 874 - 1296.

SEC ID N°:5 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento LH_NH_{CN} recombinante de la neurotoxina botulínica de tipo A con secuencia de bucle modificada según la invención y marca de

hexahistidina C-terminal (rBoNT(A)-L_{mod1}H_NH_{CN}). La secuencia corresponde a la SEC ID N°:1, en la que están delecionados los nucleótidos 3286 - 3888.

SEC ID N°:6 es un ejemplo de un fragmento LH_NH_{CN} recombinante de la neurotoxina botulínica de tipo A con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(A)-L_{mod1}H_NH_{CN}). La secuencia corresponde a la SEC ID N°:2, en la que están delecionados los restos de aminoácido 109 - 1296.

SEC ID N°:7 es un ejemplo de un ácido nucleico (ADN) que codifica por una neurotoxina botulínica de tipo B recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(B)-mod1).

SEC ID N°:8 es un ejemplo de una neurotoxina botulínica de tipo B recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(B)-mod1).

SEC ID N°:9 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento LH_N recombinante de la neurotoxina botulínica de tipo B con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(B)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:7, en la que están delecionados los nucleótidos 2623 - 3915.

SEC ID N°:10 es un ejemplo de un fragmento LH_N recombinante de la neurotoxina botulínica de tipo B con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(B)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:8, en la que están delecionados los restos de aminoácido 875 - 1305.

SEC ID N°:11 es un ejemplo de un ácido nucleico (ADN) que codifica por una neurotoxina botulínica de tipo C1 recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(C1)-mod1).

SEC ID N°:12 es un ejemplo de una neurotoxina botulínica de tipo C1 recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(C1)-mod1).

SEC ID N°: 13 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento LH_N recombinante de la neurotoxina botulínica de tipo C1 con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(C1)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:11, en la que están delecionados los nucleótidos 2599 - 3858.

SEC ID N°:14 es un ejemplo de un fragmento LH_N recombinante de la neurotoxina botulínica de tipo C1 con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(C1)-L_{mod1}H_N). La secuencia corresponde a la SEC ID N°:12, en la que están delecionados los restos de aminoácido 867 - 1286.

SEC ID N°:15 es un ejemplo de un ácido nucleico (ADN) que codifica por una neurotoxina botulínica de tipo E recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(E)-mod1).

SEC ID N°:16 es un ejemplo de una neurotoxina botulínica de tipo E recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rBoNT(E)-mod1).

SEC ID N°:17 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento de 40 kDa, recombinante, que comprende los dominios II, Ib y III de la exotoxina de Pseudomonas con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (PE40-mod1).

SEC ID N°:18 es un ejemplo de un fragmento de 40 kDa, recombinante, que comprende los dominios II, Ib y III de la exotoxina de Pseudomonas con secuencia de bucle modificada según la invención y marca de hexahistidina C terminal (PE40-mod1).

SEC ID N°: 19 es un ejemplo de un ácido nucleico (ADN) que codifica por un fragmento recombinante, que comprende la cadena A y un fragmento N terminal de la cadena B de la toxina diftérica con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (DT389-mod1).

SEC ID N°:20 es un ejemplo de un fragmento recombinante, que comprende la cadena A y un fragmento N terminal de la cadena B de la toxina diftérica con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (DT389-mod1).

SEC ID N°:21 es un ejemplo de un ácido nucleico (ADN) que codifica por una toxina ricina recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rRicin-mod1).

SEC ID N°:22 es un ejemplo de una toxina ricina recombinante con secuencia de bucle modificada según la invención y marca de hexahistidina C-terminal (rRicin-mod1).

Ejemplos

Ejemplo 1: Clonación y expresión del fragmento LH_N de neurotoxina botulínica de tipo A con bucle modificado

Para la clonación de las secuencias de ADN de la cadena ligera así como del dominio de translocación se aisló ADN cromosómico a partir de un cultivo de *Clostridium botulinum* tipo A (cepa ATCC 3502). Mediante amplificación por PCR con los cebadores N° 1 y N° 2 (figura 6) se obtuvo un fragmento génico que codifica por la cadena ligera de BoNT(A) con secuencia de bucle modificada y marca de His C terminal. El amplificado de PCR se clonó en el plásmido de expresión pQE- 60 a través de los sitios de restricción para Nco I y Sal I, mediante lo cual resultó el plásmido pQE-BoNT(A)-L_{mod1}. Mediante amplificación por PCR con los cebadores N° 3 y N° 4 (figura 6) se generó el fragmento génico que codifica por el dominio de translocación de BoNT(A). A través de los sitios de restricción para Stu I y Xho I se clonó entre la secuencia de bucle y la secuencia para la marca de His en pQE-BoNT(A)-L_{mod1} (plásmido pQE-BoNT(A)-L_{mod1}H_N; secuencia N° 2; figura 3, N° 2). La cepa de expresión de *E. coli* M15[pREP4] (Qiafen) se transformó con el plásmido pQE-BoNT(A)-L_{mod1}H_N. La expresión del fragmento LH_N modificado se realizó mediante una inducción escalonada con 500 µM de concentración final de IPTG a 25 °C durante la noche. Las células se disgregaron en un tampón fosfato 50 mM a pH 8,0 con NaCl 300 mM mediante tratamiento con lisozima y ultrasonidos. El lisado centrifugado se sometió a cromatografía a través de una columna de Ni-NTA-agarosa. Un análisis en gel de SDS-poliacrilamida dio como resultado que mediante Coomassie en condiciones reductoras se tiñeron dos bandas a aproximadamente 50 kDa así como una banda a 100 kDa, mientras que en condiciones no reductoras sólo podía observarse la banda a aproximadamente 100 kDa (figura 7). Con ello se muestra claramente que el fragmento LH_N se liberó en > 75 % como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí mediante un puente disulfuro. El tratamiento posterior con trombina llevó por un lado a la escisión de la forma monocatenaria, por otro lado a un acortamiento del dominio de translocación en el polipéptido bicatenario (figura 7). Una incubación de dos horas del lisado de *E. coli* antes de la purificación del fragmento LH_N llevó al polipéptido bicatenario en el caso de la escisión completa.

Un fragmento LH_N expresado y purificado de manera correspondiente con la secuencia de bucle nativa (figura 3, N° 1) mostró en gel de SDS-poliacrilamida tanto en condiciones no reductoras como en condiciones reductoras una banda a 100 kDa. El péptido monocatenario sólo pudo transformarse mediante escisión con tripsina en el fragmento LH_N bicatenario, con puente disulfuro.

Ejemplo 2: Clonación y expresión del fragmento LH_NH_{CN} de neurotoxina botulínica de tipo A con bucle modificado y caracterización del sitio de escisión

El fragmento H_NH_{CN} (dominio de translocación con mitad N terminal del dominio de unión a receptor de BoNT(A)) se generó mediante amplificación por PCR con los cebadores N° 3 y N° 5 (figura 6) y se clonó a través de los sitios de restricción para Stu I y Xho I en el plásmido pQE-BoNT(A)-L_{mod1} (plásmido pQE-BoNT(A)-L_{mod1}H_NH_{CN}; secuencia N° 3). La expresión y la purificación tuvieron lugar según el esquema descrito en el ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado además de una banda más débil que correspondía al polipéptido monocatenario, y otras bandas no definidas, una banda a 50 kDa así como una a 75 kDa, que correspondían a la cadena ligera o al fragmento H_NH_{CN}. La secuenciación N terminal de los cuatro primeros restos de aminoácido del fragmento H_NH_{CN} dio como resultado la secuencia Ser-Leu-Val-Pro. La escisión mediante la actividad proteasa en el lisado de *E. coli* tuvo lugar por lo tanto detrás de Lys₄₄₀ y con ello de manera N terminal con respecto al pentapéptido Val-Pro-Arg-Gly-Ser introducido en el bucle.

Ejemplo 3: Clonación y expresión del fragmento LH_N de neurotoxina botulínica de tipo B con bucle modificado

Para la clonación de las secuencias ADN de la cadena ligera así como del dominio de translocación se aisló ADN cromosómico a partir de un cultivo de *Clostridium botulinum* tipo B (cepa Okra). Mediante amplificación por PCR con los cebadores N° 6 y N° 7 (figura 6) se generó un fragmento génico que codifica por la cadena ligera de BoNT(B) con secuencia de bucle modificada de BoNT(A). Con los cebadores N° 8 y N° 9 (figura 6) se generó el fragmento génico que codifica por el dominio de translocación de BoNT(B). La clonación en el plásmido de expresión pQE-60 se realizó en primer lugar mediante cambio del fragmento génico de BoNT(A)-L en pQE-BoNT(A)-L_{mod1} por el amplificado de BoNT(B)-L_{mod1} a través de los sitios de restricción para Nco I y Stu I. A continuación se clonó detrás del amplificado de BoNT(B)-H_N a través de los sitios de restricción para Stu I y Xho I, mediante lo cual resultó el plásmido pQE-BoNT(B)-L_{mod1}H_N (secuencia N° 5). La expresión en la cepa huésped M15[pREP4] y la purificación del fragmento LH_N se realizaron de manera análoga al ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado que se tiñeron mediante Coomassie en condiciones reductoras dos bandas a aproximadamente 50 kDa y 55 kDa, mientras que en condiciones no reductoras pudo observarse una banda a aproximadamente 105 kDa (figura 8). Con ello se muestra claramente que el fragmento LH_N se liberó en su mayor parte como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí en > 80 % mediante un puente disulfuro.

Ejemplo 4: Clonación y expresión del fragmento LH_N de neurotoxina botulínica de tipo C1 con bucle modificado y caracterización del sitio de escisión

Para la clonación de las secuencias ADN de la cadena ligera así como del dominio de translocación se preparó ADN cromosómico a partir de un cultivo de *Clostridium botulinum* tipo C1 (cepa C205). Mediante amplificación por PCR con los cebadores N° 10 y N° 11 (figura 6) se generó un fragmento génico que codifica por la cadena ligera de BoNT(C1) con secuencia de bucle modificada de BoNT(A). Con los cebadores N° 12 y N° 13 (figura 6) se generó el fragmento génico que codifica por el dominio de translocación de BoNT(C1). La clonación en el plásmido de expresión pQE-60 se realizó en primer lugar mediante cambio del fragmento génico de BoNT(A)-L en pQE-BoNT(A)-L_{mod1} por el amplificado de BoNT(C1)-L_{mod1} a través de los sitios de restricción para *Nco* I y *Stu* I. A continuación se clonó detrás del amplificado de BoNT(C1)-H_N a través de los sitios de restricción para *Stu* I y *Xho* I, mediante lo cual resultó el plásmido pQE-BoNT(C1)-L_{mod1}H_N (secuencia N° 7). La expresión en la cepa huésped M15[pREP4] y la purificación del fragmento LH_N se realizaron de manera análoga al ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado que en condiciones reductoras se tiñeron mediante Coomassie dos bandas a aproximadamente 50 kDa y 55 kDa, mientras que en condiciones no reductoras pudo observarse una banda a aproximadamente 105 kDa. Con ello se muestra claramente que el fragmento LH_N se liberó en > 90 % como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí mediante un puente disulfuro. La secuenciación N terminal de los cuatro primeros restos de aminoácido del fragmento H_N dio como resultado la secuencia Ser-Leu-Val-Pro. La escisión mediante la actividad proteasa en el lisado de *E. coli* tuvo lugar por lo tanto detrás de Lys₄₄₇ y con ello de manera N terminal con respecto al pentapéptido Val-Pro-Arg-Gly-Ser introducido en el bucle de BoNT(A). Mediante mutagénesis dirigida se cambió el resto de arginina del pentapéptido introducido por histidina, tirosina o glutamina. Los fragmentos LH_N mutagenizados, expresados de la misma manera se encontraban tras una incubación de dos horas del lisado celular de *E. coli* en > 90 % en la forma bicatenaria, con puente disulfuro, siendo la eficacia de la escisión ligeramente inferior a la del fragmento LH_N que contiene el bucle de BoNT(A) modificado con el pentapéptido Val-Pro-Arg-Gly-Ser.

Ejemplo 5: Clonación y expresión de una neurotoxina botulínica de tipo C1 recombinante con bucle modificado

Usando ADN cromosómico de la cepa C205 de *Clostridium botulinum* se amplificó por medio de PCR con los cebadores N° 12 y N° 14 (figura 6) el fragmento génico que codifica por la cadena pesada. A través de los sitios de restricción para *Stu* I y *Xho* I se clonó entre el fragmento de secuencia que codifica por la cadena ligera y la secuencia para la marca de His en el plásmido BoNT(C1)-L_{mod1}H_N (plásmido pQE-BoNT(C1)-L_{mod1}H_NHc; secuencia N° 6). La cepa de expresión de *E. coli* M15[pREP4] (Qiagen) se transformó con el plásmido de expresión correspondiente. La expresión en la cepa huésped M15[pREP4] y la purificación tuvieron lugar de manera análoga al ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado que se tiñeron mediante Coomassie en condiciones reductoras dos bandas a aproximadamente 50 kDa y 105 kDa, mientras que en condiciones no reductoras pudo observarse una banda a aproximadamente 155 kDa (figura 9). Con ello se muestra claramente que la neurotoxina recombinante se liberó en > 90 % como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí mediante un puente disulfuro. Una prueba de actividad en el ensayo de hemidiafragma dio como resultado una toxicidad elevada en comparación tal como la de la neurotoxina de tipo 1 nativa, aislada a partir de *Clostridium botulinum*. La modificación de la región de bucle entre la cadena ligera y el dominio de translocación no tenía por lo tanto ningún efecto sobre la toxicidad.

Ejemplo 6: Clonación y expresión de un fragmento recombinante de la exotoxina de *Pseudomonas* (PE40) con bucle modificado

Usando ADN cromosómico de la cepa *Pseudomonas aeruginosa* 103 se amplificó por medio de PCR con los cebadores N° 17 y N° 18 (figura 6) un fragmento génico que codifica por la región del dominio II, que se encuentra de manera C terminal con respecto al bucle entre los restos de cisteína 13 y 36, así como para el dominio III. El amplificado se clonó a través de *Nco* I y *Mlu* I en el cambio por el fragmento génico BoNT(A)-L_{mod1} en el plásmido pQE-BoNT(A)-L_{mod1} (plásmido pQE-PEII3-III). El segmento de secuencia para la región del dominio II de manera N terminal con respecto al bucle se introdujo mediante hibridación de los oligonucleótidos N° 15 y N° 16 (figura 6) y clonación a través de los sitios de restricción para *Nco* I y *Kpn* I en el plásmido pQE-PEII3-III (plásmido pQE-PEIImodIII; secuencia N° 9). La cepa de expresión de *E. coli* M15 [pREP4] (Qiagen) se transformó con el plásmido de expresión correspondiente. La expresión en la cepa huésped M15[pREP4] y la purificación se realizaron de manera análoga al ejemplo 1. Un análisis en gel de SDS en condiciones reductoras dio como resultado una banda más débil a 40 kDa así como una más intensa a 37 kDa. En condiciones no reductoras pudo observarse por el contrario sólo una banda a 40 kDa. Si se incubaba el lisado celular antes de la purificación por cromatografía de afinidad durante al menos dos horas a temperatura ambiente, ya no podía detectarse la banda de 40 kDa en condiciones reductoras. Mediante el cambio de la región de bucle entre los restos de cisteína 13 y 36 en el dominio II del fragmento PE40 por un bucle de BoNT(A) modificado tuvo lugar por lo tanto una escisión de la cadena de polipéptido, formando los restos de cisteína mencionados un puente disulfuro. El fragmento N terminal, de aproximadamente 3 kDa de tamaño no podía detectarse tras la reducción en gel de SDS al 12 %.

Ejemplo 7: Clonación y expresión de un fragmento recombinante de la toxina diftérica (DT389) con bucle modificado

Usando ADN cromosómico de la cepa NCTC 13129 de *Corynebacterium diphtheriae* se amplificó por medio de PCR con los cebadores N° 19 y N° 20 (figura 6) el fragmento génico que codifica por la cadena A de la toxina diftérica. A través de los sitios de restricción para *Nco* I y *Stu* I se clonó (plásmido pQE-DT-A_{mod1}) el amplificado en el plásmido pQE-BoNT(A)-L_{mod1} (véase el ejemplo 1). De la misma manera se amplificó el fragmento génico que codifica por un fragmento N terminal de la cadena B con los cebadores N° 21 y N° 22 (figura 6) y se clonó a través de los sitios de restricción para *Stu* I y *Xho* I en pQE-DT-A_{mod1} (plásmido pQE-DT389_{mod1}; secuencia N° 10). La cepa de expresión de *E. coli* M15[pREP4] (Qiagen) se transformó con el plásmido de expresión correspondiente. La expresión en la cepa huésped M15 [pREP4] y la purificación se realizaron de manera análoga al ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado que se tiñeron mediante Coomassie en condiciones reductoras dos bandas a aproximadamente 22 kDa, mientras que en condiciones no reductoras pudo observarse una banda a aproximadamente 43 kDa. Con ello se muestra claramente que el fragmento de toxina diftérica recombinante se liberó en > 90 % como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí mediante un puente disulfuro.

Ejemplo 8: Clonación y expresión de ricina recombinante con bucle modificado

Usando ARNm de semillas de *Ricinus communis* se amplificó por medio de RT-PCR con los cebadores N° 23 y N° 24 (figura 6) el fragmento génico que codifica por la cadena A de ricina. A través de los sitios de restricción para *Nco* I y *Xho* I se clonó (plásmido pQE-ricina-A) en el plásmido pQE-BoNT(A)-L_{mod1} (véase el ejemplo 1). De la misma manera se amplificó el fragmento génico que codifica por la cadena B con los cebadores N° 25 y N° 26 (figura 6) y se clonó a través de los sitios de restricción para *Kpn* I y *Xho* I en pQE-ricina-A (plásmido pQE-ricina-mod1; secuencia N° 11). La cepa de expresión de *E. coli* M15[pREP4] (Qiagen) se transformó con el plásmido de expresión correspondiente. La expresión en la cepa huésped M15[pREP4] y la purificación de la parte soluble de la ricina expresada se realizaron de manera análoga al ejemplo 1. Un análisis en gel de SDS-poliacrilamida dio como resultado que se tiñeron mediante Coomassie en condiciones reductoras dos bandas a aproximadamente 19 kDa y 42 kDa, mientras que en condiciones no reductoras pudo observarse una banda a aproximadamente 62 kDa. Con ello se muestra claramente que la parte soluble de la ricina recombinante se liberó en > 90 % como polipéptido bicatenario a partir de las bacterias, en el que las dos cadenas estaban unidas covalentemente entre sí mediante un puente disulfuro.

Bibliografía científica

- Arora y col. (1999), Cancer Res. 59: 183-8
 Collier (2001), Toxicon 39(11): 1793-803
 Fujinaga (1997), Microbiology 143: 3841-47
 Ogata y col. (1990), J. Biol. Chem. 265(33): 20678-85
 Reiter (2001), Adv. Cancer Res. 81: 93-124
 Schiavo y Montecucco (1997), The Clostridia: Molecular Biology and Pathogenesis, Academic Press, San Diego: 295 – 322
 Williams y col., (1990), J. Biol. Chem. 265(33): 20673-77

Bibliografía de patentes

- Borgford, patente estadounidense 6.593.132
 Brown y Jones, documento WO 89/04839
 Fitzgerald y col., patente estadounidense 6.426.075
 Pastan y col., patente estadounidense 5.980.895
 Health Protection Agency, documento WO2004/024909

LISTA DE SECUNCIAS

- <110> BioteCon Therapeutics GmbH
 <120> Expresión recombinante de proteínas en una forma bicatenaria con puente disulfuro
 <130> BIO-008 PCT
 <140> desconocido
 <141> 20-01-2006
 <150> 10 2005 002978.7
 <151> 21-01-2005
 <160> 22
 <170> PatentIn versión 3.3

<210> 1
 <211> 3915
 <212> ADN
 <213> *Clostridium botulinum*

5 <400> 1

atggcatttg	ttaataaaca	atttaattat	aaagatcctg	taaatgggtg	tgatattgct	60
tatataaaaa	ttccaaatgc	aggacaaatg	caaccagtaa	aagcttttaa	aattcataat	120
aaaatatggg	ttattccaga	aagagataca	tttacaaatc	ctgaagaagg	agattttaat	180
ccaccaccag	aagcaaaaca	agttccagtt	tcatattatg	attcaacata	tttaagtaca	240
gataatgaaa	aagataatta	tttaaaggga	gttacaaaat	tatttgagag	aatttattca	300
actgatcttg	gaagaatggt	gttaacatca	atagtaaggg	gaataccatt	ttgggggtgga	360
agtacaatag	atacagaatt	aaaagttatt	gatactaatt	gtattaatgt	gatacaacca	420
gatggtagtt	atagatcaga	agaacttaat	ctagtaataa	taggaccctc	agctgatatt	480
atacagtttg	aatgtaaaag	ctttggacat	gaagttttga	atcttacgcg	aaatggttat	540
ggctctactc	aatacattag	atttagccca	gattttacat	ttggttttga	ggagtcactt	600
gaagttgata	caaactctct	tttaggtgca	ggcaaatttg	ctacagatcc	agcagtaaca	660
ttagcacatg	aacttatata	tgctggacat	agattatatg	gaatagcaat	taatccaaat	720
aggggttttta	aagtaaatac	taatgcctat	tatgaaatga	gtgggttaga	agtaagcttt	780
gaggaactta	gaacatttgg	gggacatgat	gcaaagttta	tagatagttt	acaggaaaac	840
gaatttcgtc	tatattatta	taataagttt	aaagatatag	caagtacact	taataaagct	900
aatcaatag	taggtactac	tgcttcatta	cagtatatga	aaaatgtttt	taaagagaaa	960
tatctcctat	ctgaagatac	atctggaaaa	ttttcggtag	ataaattaaa	atttgataag	1020
ttatacaaaa	tgттаacaga	gatttacaca	gaggataatt	ttgttaagtt	ttttaagta	1080
cttaacagaa	aaacatatatt	gaattttgat	aaagccgtat	ttaagataaa	tatagtacct	1140
aaggtaaatt	acacaatata	tgatggattt	aattttaagaa	atacaaattt	agcagcaaac	1200
tttaatggtc	aaaatacaga	aattaataat	atgaatttta	ctaaactaaa	aaattttact	1260
ggattgtttg	aattttataa	gttgctatgt	gtaagaggga	taataacttc	taaaactaaa	1320
tcattagtgc	cgcggtggatc	caaggcctta	aatgatttat	gtatcaaagt	taataattgg	1380
gacttggttt	ttagtctctc	agaagataat	tttactaatg	atctaaataa	aggagaagaa	1440
attacatctg	atactaatat	agaagcagca	gaagaaaata	ttagtttaga	tttaatacaa	1500
caatattatt	taacctttta	ttttgataat	gaacctgaaa	atatttcaat	agaaaaatctt	1560
tcaagtgaca	ttataggcca	attagaactt	atgcctaata	tagaaagatt	tcctaattgga	1620
aaaaagtatg	agttagataa	atatactatg	ttccattatc	ttcgtgctca	agaatttgaa	1680
catggtaaatt	ctaggattgc	tttaacaaat	tctgttaacg	aagcattatt	aaatcctagt	1740
cgtgtttata	cattttttttc	ttcagactat	gtaaagaaag	ttaataaagc	tacggaggca	1800
gctatgtttt	taggctgggt	agaacaatta	gtatatgatt	ttaccgatga	aactagcgaa	1860
gtaagtacta	cggataaaaat	tgcggtatata	actataatta	ttccatatat	aggacctgct	1920
ttaaatatag	gtaatatggt	atataaagat	gattttgtag	gtgctttaat	attttcagga	1980
gctgttatte	tgttagaatt	tataccagag	attgcaatac	ctgtattagg	tacttttgca	2040
cttgatcat	atattgcgaa	taagggttcta	accgttcaaa	caatagataa	tgctttaagt	2100
aaaagaaatg	aaaaatggga	tgaggtctat	aaatatatag	taacaaattg	gttagcaaag	2160
gttaatacac	agattgatct	aataagaaaa	aaaatgaaag	aagcttttaga	aatcaagca	2220
gaagcaacaa	aggctataat	aaactatcag	tataatcaat	atactgagga	agagaaaaat	2280
aatattaatt	ttaatatatga	tgattttaagt	tcgaaactta	atgagtctat	aaataaagct	2340

atgattaata	taaataaatt	tttgaatcaa	tgctctgttt	catattttaat	gaattctatg	2400
atccccttatg	gtgttaaacg	gttagaagat	tttgatgcta	gtcttaaaga	tgcatatta	2460
aagtatatat	atgataatag	aggaacttta	attgggtcaag	tagatagatt	aaaagataaa	2520
gttaataata	cacttagtac	agatatacct	tttcagcttt	ccaaatacgt	agataatcaa	2580
agattattat	ctacatttac	tgaatatatt	aagaatatta	ttaatacttc	tatattgaat	2640
ttaagatatg	aaagtaatca	tttaatagac	ttatctaggt	atgcatcaaa	aataaatatt	2700
ggtagtaaa	taaattttga	tccaatagat	aaaaatcaaa	ttcaattatt	taatttagaa	2760
agtagtaaaa	ttgaggtaat	tttaaaaaat	gctattgtat	ataatagtat	gtatgaaaat	2820
tttagtacta	gcttttggat	aagaattcct	aagtatttta	acagtataag	tctaaataat	2880
gaatatacaa	taataaattg	tatggaaaat	aattcaggat	ggaaagtatc	acttaattat	2940
ggtgaaataa	tctggacttt	acaggatact	caggaaataa	aacaaagagt	agtttttaaa	3000
tacagtcaaa	tgattaatat	atcagattat	ataaacagat	ggatttttgt	aactatcact	3060
aataatagat	taaataactc	taaaatttat	ataaatggaa	gattaataga	tcaaaaacca	3120
atttcaaaat	taggtaatat	tcatgctagt	aataatataa	tgtttaaatt	agatggttgt	3180
agagatacac	atagatatat	ttggataaaa	tatttttaatc	tttttgataa	ggaattaaat	3240
gaaaaagaaa	tcaaagattt	atatgataat	caatcaaatt	caggatattt	aaaagacttt	3300
tggggtgatt	atttacaata	tgataaacca	tactatatgt	taaattttata	tgatccaaat	3360
aaatatgtcg	atgtaaataa	tgtaggtatt	agaggttata	tgtatcttaa	agggcctaga	3420
ggtagcgtaa	tgactacaaa	catttattta	aattcaagtt	tgtatagggg	gacaaaattt	3480
attataaaaa	aatatgcttc	tggaaataaa	gataatattg	ttagaaataa	tgatcgtgta	3540
tatattaatg	tagtagttaa	aaataaagaa	tataggttag	ctactaatgc	atcacaggca	3600
ggcgtagaaa	aaataactaag	tgcattagaa	atacctgatg	taggaaatct	aagtcaagta	3660
gtagtaatga	agtcaaaaaa	tgatcaagga	ataacaaata	aatgcaaaaat	gaattttacaa	3720
gataataatg	ggaatgatat	aggctttata	ggatttcatc	agtttaataa	tatagctaaa	3780
ctagtagcaa	gtaattggta	taatagacaa	atagaaagat	ctagtaggac	tttgggttgc	3840
tcatgggaat	ttattcctgt	agatgatgga	tggggagaaa	ggccactgct	cgagcaccac	3900
caccaccacc	actaa					3915

<210> 2

<211> 1304

<212> PRT

5 <213> *Clostridium botulinum*

<400> 2

ES 2 380 677 T3

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

ES 2 380 677 T3

Arg	Ser	Glu	Glu	Leu	Asn	Leu	Val	Ile	Ile	Gly	Pro	Ser	Ala	Asp	Ile	145	150	155	160
Ile	Gln	Phe	Glu	Cys	Lys	Ser	Phe	Gly	His	Glu	Val	Leu	Asn	Leu	Thr	165	170	175	
Arg	Asn	Gly	Tyr	Gly	Ser	Thr	Gln	Tyr	Ile	Arg	Phe	Ser	Pro	Asp	Phe	180	185	190	
Thr	Phe	Gly	Phe	Glu	Glu	Ser	Leu	Glu	Val	Asp	Thr	Asn	Pro	Leu	Leu	195	200	205	
Gly	Ala	Gly	Lys	Phe	Ala	Thr	Asp	Pro	Ala	Val	Thr	Leu	Ala	His	Glu	210	215	220	
Leu	Ile	His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	Ile	Asn	Pro	Asn	225	230	235	240
Arg	Val	Phe	Lys	Val	Asn	Thr	Asn	Ala	Tyr	Tyr	Glu	Met	Ser	Gly	Leu	245	250	255	
Glu	Val	Ser	Phe	Glu	Glu	Leu	Arg	Thr	Phe	Gly	Gly	His	Asp	Ala	Lys	260	265	270	
Phe	Ile	Asp	Ser	Leu	Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr	Tyr	Asn	275	280	285	
Lys	Phe	Lys	Asp	Ile	Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys	Ser	Ile	Val	290	295	300	
Gly	Thr	Thr	Ala	Ser	Leu	Gln	Tyr	Met	Lys	Asn	Val	Phe	Lys	Glu	Lys	305	310	315	320
Tyr	Leu	Leu	Ser	Glu	Asp	Thr	Ser	Gly	Lys	Phe	Ser	Val	Asp	Lys	Leu	325	330	335	
Lys	Phe	Asp	Lys	Leu	Tyr	Lys	Met	Leu	Thr	Glu	Ile	Tyr	Thr	Glu	Asp	340	345	350	
Asn	Phe	Val	Lys	Phe	Phe	Lys	Val	Leu	Asn	Arg	Lys	Thr	Tyr	Leu	Asn	355	360	365	
Phe	Asp	Lys	Ala	Val	Phe	Lys	Ile	Asn	Ile	Val	Pro	Lys	Val	Asn	Tyr	370	375	380	
Thr	Ile	Tyr	Asp	Gly	Phe	Asn	Leu	Arg	Asn	Thr	Asn	Leu	Ala	Ala	Asn	385	390	395	400
Phe	Asn	Gly	Gln	Asn	Thr	Glu	Ile	Asn	Asn	Met	Asn	Phe	Thr	Lys	Leu	405	410	415	
Lys	Asn	Phe	Thr	Gly	Leu	Phe	Glu	Phe	Tyr	Lys	Leu	Leu	Cys	Val	Arg	420	425	430	
Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser	Leu	Val	Pro	Arg	Gly	Ser	Lys	435	440	445	
Ala	Leu	Asn	Asp	Leu	Cys	Ile	Lys	Val	Asn	Asn	Trp	Asp	Leu	Phe	Phe	450	455	460	
Ser	Pro	Ser	Glu	Asp	Asn	Phe	Thr	Asn	Asp	Leu	Asn	Lys	Gly	Glu	Glu	465	470	475	480

ES 2 380 677 T3

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu
 485 490 495
 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro
 500 505 510
 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu
 515 520 525
 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
 530 535 540
 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu
 545 550 555 560
 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685
 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700
 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720
 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735
 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750
 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765
 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780
 Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800
 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815

ES 2 380 677 T3

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830
 Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845
 Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860
 Thr Phe Thr Glu Tyr Ile Lys Asn Ile Ile Asn Thr Ser Ile Leu Asn
 865 870 875 880
 Leu Arg Tyr Glu Ser Asn His Leu Ile Asp Leu Ser Arg Tyr Ala Ser
 885 890 895
 Lys Ile Asn Ile Gly Ser Lys Val Asn Phe Asp Pro Ile Asp Lys Asn
 900 905 910
 Gln Ile Gln Leu Phe Asn Leu Glu Ser Ser Lys Ile Glu Val Ile Leu
 915 920 925
 Lys Asn Ala Ile Val Tyr Asn Ser Met Tyr Glu Asn Phe Ser Thr Ser
 930 935 940
 Phe Trp Ile Arg Ile Pro Lys Tyr Phe Asn Ser Ile Ser Leu Asn Asn
 945 950 955 960
 Glu Tyr Thr Ile Ile Asn Cys Met Glu Asn Asn Ser Gly Trp Lys Val
 965 970 975
 Ser Leu Asn Tyr Gly Glu Ile Ile Trp Thr Leu Gln Asp Thr Gln Glu
 980 985 990
 Ile Lys Gln Arg Val Val Phe Lys Tyr Ser Gln Met Ile Asn Ile Se
 995 1000 1005
 Asp Tyr Ile Asn Arg Trp Ile Phe Val Thr Ile Thr Asn Asn Arg
 1010 1015 1020
 Leu Asn Asn Ser Lys Ile Tyr Ile Asn Gly Arg Leu Ile Asp Gln
 1025 1030 1035
 Lys Pro Ile Ser Asn Leu Gly Asn Ile His Ala Ser Asn Asn Ile
 1040 1045 1050
 Met Phe Lys Leu Asp Gly Cys Arg Asp Thr His Arg Tyr Ile Trp
 1055 1060 1065
 Ile Lys Tyr Phe Asn Leu Phe Asp Lys Glu Leu Asn Glu Lys Glu
 1070 1075 1080
 Ile Lys Asp Leu Tyr Asp Asn Gln Ser Asn Ser Gly Ile Leu Lys
 1085 1090 1095
 Asp Phe Trp Gly Asp Tyr Leu Gln Tyr Asp Lys Pro Tyr Tyr Met
 1100 1105 1110
 Leu Asn Leu Tyr Asp Pro Asn Lys Tyr Val Asp Val Asn Asn Val
 1115 1120 1125
 Gly Ile Arg Gly Tyr Met Tyr Leu Lys Gly Pro Arg Gly Ser Val
 1130 1135 1140

ES 2 380 677 T3

Met	Thr	Thr	Asn	Ile	Tyr	Leu	Asn	Ser	Ser	Leu	Tyr	Arg	Gly	Thr
1145						1150					1155			
Lys	Phe	Ile	Ile	Lys	Lys	Tyr	Ala	Ser	Gly	Asn	Lys	Asp	Asn	Ile
1160						1165					1170			
Val	Arg	Asn	Asn	Asp	Arg	Val	Tyr	Ile	Asn	Val	Val	Val	Lys	Asn
1175						1180					1185			
Lys	Glu	Tyr	Arg	Leu	Ala	Thr	Asn	Ala	Ser	Gln	Ala	Gly	Val	Glu
1190						1195					1200			
Lys	Ile	Leu	Ser	Ala	Leu	Glu	Ile	Pro	Asp	Val	Gly	Asn	Leu	Ser
1205						1210					1215			
Gln	Val	Val	Val	Met	Lys	Ser	Lys	Asn	Asp	Gln	Gly	Ile	Thr	Asn
1220						1225					1230			
Lys	Cys	Lys	Met	Asn	Leu	Gln	Asp	Asn	Asn	Gly	Asn	Asp	Ile	Gly
1235						1240					1245			
Phe	Ile	Gly	Phe	His	Gln	Phe	Asn	Asn	Ile	Ala	Lys	Leu	Val	Ala
1250						1255					1260			
Ser	Asn	Trp	Tyr	Asn	Arg	Gln	Ile	Glu	Arg	Ser	Ser	Arg	Thr	Leu
1265						1270					1275			
Gly	Cys	Ser	Trp	Glu	Phe	Ile	Pro	Val	Asp	Asp	Gly	Trp	Gly	Glu
1280						1285					1290			
Arg	Pro	Leu	Leu	Glu	His	His	His	His	His	His				
1295						1300								

<210> 3

<211> 2646

<212> ADN

5 <213> *Clostridium botulinum*

<400> 3

atggcatttg	ttaataaaca	atttaattat	aaagatcctg	taaatgggtg	tgatattgct	60
tatataaaaa	ttccaaatgc	aggacaaatg	caaccagtaa	aagcttttaa	aattcataat	120
aaaatatggg	ttattccaga	aagagataca	tttacaatc	ctgaagaagg	agatttaa	180
ccaccaccag	aagcaaaaca	agttccagtt	tcatattatg	attcaacata	tttaagtaca	240
gataatgaaa	aagataatta	tttaaaggga	gttacaaaat	tatttgagag	aatttattca	300
actgatcttg	gaagaatggt	gttaacatca	atagtaaggg	gaataccatt	ttgggggtgga	360
agtacaatag	atacagaatt	aaaagttatt	gatactaatt	gtattaatgt	gatacaacca	420
gatggtagtt	atagatcaga	agaactta	ctagtaataa	taggaccctc	agctgatatt	480
atacagtttg	aatgtaaaag	ctttggacat	gaagttttga	atcttacgcg	aaatgggttat	540
ggctctactc	aatacattag	atttagccca	gattttacat	ttgggtttga	ggagtcactt	600
gaagttgata	caaactcctc	tttaggtgca	ggcaaatctg	ctacagatcc	agcagtaaca	660
ttagcacatg	aacttatata	tgctggacat	agatttatatg	gaatagcaat	taatccaaat	720
aggggttttta	aagtaaatac	taatgcctat	tatgaaatga	gtgggttaga	agtaagcttt	780
gaggaactta	gaacattttg	gggacatgat	gcaaagttta	tagatagttt	acaggaaaaac	840
gaatttcgtc	tatattatta	taataagttt	aaagatatag	caagtacact	taataaagct	900
aaatcaatag	taggtactac	tgcttcatta	cagtatatga	aaaatgtttt	taaagagaaa	960
tatctcctat	ctgaagatac	atctggaaaa	ttttcggtag	ataaattaaa	atttgataag	1020
ttatacaaaa	tgtaacaga	gatttacaca	gaggataatt	ttgttaagtt	ttttaagta	1080
cttaacagaa	aaacatat	gaattttgat	aaagccgtat	ttaagataaa	tatagtacct	1140
aaggtaaat	acacaatata	tgatggattt	aattttaagaa	atacaaat	agcagcaaac	1200
tttaatggtc	aaaatacaga	aattaataat	atgaatttta	ctaaactaaa	aaattttact	1260
ggattgtttg	aattttataa	gttgctatgt	gtaagaggga	taataacttc	taaaactaaa	1320

tcattagttc	cgcggtggatc	caaggcctta	aatgattttat	gtatcaaagt	taataattgg	1380
gacttgtttt	ttagtccttc	agaagataat	tttactaatg	atctaaataa	aggagaagaa	1440
attacatctg	atactaatat	agaagcagca	gaagaaaata	ttagtttaga	tttaatacaa	1500
caatattatt	taacctttta	ttttgataat	gaacctgaaa	atattttcaat	agaaaatcct	1560
tcaagtgaca	ttataggcca	attagaactt	atgcctaata	tagaaagatt	tcctaattgga	1620
aaaaagtatg	agttagataa	atatactatg	ttccattatc	ttcgtgctca	agaatttgaa	1680
catggtaaat	ctaggattgc	tttaacaaat	tctgttaacg	aagcattatt	aaatcctagt	1740
cgtgtttata	cattttttttc	ttcagactat	gtaaagaaag	ttaataaagc	tacggaggca	1800
gctatgtttt	taggctgggt	agaacaatta	gtatatgatt	ttaccgatga	aactagcgaa	1860
gtaagtacta	cggataaaaat	tgcggtatata	actataatta	ttccatatata	aggacctgct	1920
ttaaatatag	gtaatatggt	atataaagat	gattttgttag	gtgctttaat	atttttcagga	1980
gctgttatcc	tgttagaatt	tataccagag	attgcaatac	ctgtattagg	tacttttgca	2040
cttgatcat	atattgcgaa	taagggttcta	accgttcaaa	caatagataa	tgctttaagt	2100
aaaagaaatg	aaaaatggga	tgaggtctat	aaatatatag	taacaaattg	gttagcaaaag	2160
gttaatacac	agattgatct	aataagaaaa	aaaatgaaag	aagctttaga	aaatcaagca	2220
gaagcaacaa	aggctataat	aaactatcag	tataatcaat	atactgagga	agagaaaaat	2280
aatattaatt	ttaatataga	tgatttaagt	tcgaaactta	atgagtctat	aaataaagct	2340
atgattaata	taaataaatt	tttgaatcaa	tgctctggtt	catattta	gaattctatg	2400
atcccttatg	gtgttaaacg	gttagaagat	tttgatgcta	gtcttaaaaga	tgcatattata	2460
aagtatatat	atgataatag	aggaaacttta	attgggtcaag	tagatagatt	aaaagataaa	2520
gttaataata	cacttagtac	agatatacct	tttcagcttt	ccaaatacgt	agataatcaa	2580
agattattat	ctacattttac	tgaatatatt	aagaatat	tcgagcacca	ccaccaccac	2640
cactaa						2646

<210> 4

<211> 881

<212> PRT

5 <213> *Clostridium botulinum*

<400> 4

ES 2 380 677 T3

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

ES 2 380 677 T3

165										170					175				
Arg	Asn	Gly	Tyr	Gly	Ser	Thr	Gln	Tyr	Ile	Arg	Phe	Ser	Pro	Asp	Phe				
			180					185					190						
Thr	Phe	Gly	Phe	Glu	Glu	Ser	Leu	Glu	Val	Asp	Thr	Asn	Pro	Leu	Leu				
		195					200					205							
Gly	Ala	Gly	Lys	Phe	Ala	Thr	Asp	Pro	Ala	Val	Thr	Leu	Ala	His	Glu				
	210					215					220								
Leu	Ile	His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	Ile	Asn	Pro	Asn				
225					230					235					240				
Arg	Val	Phe	Lys	Val	Asn	Thr	Asn	Ala	Tyr	Tyr	Glu	Met	Ser	Gly	Leu				
				245					250					255					
Glu	Val	Ser	Phe	Glu	Glu	Leu	Arg	Thr	Phe	Gly	Gly	His	Asp	Ala	Lys				
			260					265					270						
Phe	Ile	Asp	Ser	Leu	Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr	Tyr	Asn				
		275					280					285							
Lys	Phe	Lys	Asp	Ile	Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys	Ser	Ile	Val				
	290					295					300								
Gly	Thr	Thr	Ala	Ser	Leu	Gln	Tyr	Met	Lys	Asn	Val	Phe	Lys	Glu	Lys				
305					310					315					320				
Tyr	Leu	Leu	Ser	Glu	Asp	Thr	Ser	Gly	Lys	Phe	Ser	Val	Asp	Lys	Leu				
				325					330					335					
Lys	Phe	Asp	Lys	Leu	Tyr	Lys	Met	Leu	Thr	Glu	Ile	Tyr	Thr	Glu	Asp				
			340					345					350						
Asn	Phe	Val	Lys	Phe	Phe	Lys	Val	Leu	Asn	Arg	Lys	Thr	Tyr	Leu	Asn				
		355					360					365							
Phe	Asp	Lys	Ala	Val	Phe	Lys	Ile	Asn	Ile	Val	Pro	Lys	Val	Asn	Tyr				
	370					375					380								
Thr	Ile	Tyr	Asp	Gly	Phe	Asn	Leu	Arg	Asn	Thr	Asn	Leu	Ala	Ala	Asn				
385					390					395					400				
Phe	Asn	Gly	Gln	Asn	Thr	Glu	Ile	Asn	Asn	Met	Asn	Phe	Thr	Lys	Leu				
				405					410					415					
Lys	Asn	Phe	Thr	Gly	Leu	Phe	Glu	Phe	Tyr	Lys	Leu	Leu	Cys	Val	Arg				
			420						425				430						
Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser	Leu	Val	Pro	Arg	Gly	Ser	Lys				
		435					440					445							
Ala	Leu	Asn	Asp	Leu	Cys	Ile	Lys	Val	Asn	Asn	Trp	Asp	Leu	Phe	Phe				
	450					455					460								
Ser	Pro	Ser	Glu	Asp	Asn	Phe	Thr	Asn	Asp	Leu	Asn	Lys	Gly	Glu	Glu				
465					470					475					480				
Ile	Thr	Ser	Asp	Thr	Asn	Ile	Glu	Ala	Ala	Glu	Glu	Asn	Ile	Ser	Leu				
				485					490					495					
Asp	Leu	Ile	Gln	Gln	Tyr	Tyr	Leu	Thr	Phe	Asn	Phe	Asp	Asn	Glu	Pro				

ES 2 380 677 T3

500										505					510				
Glu	Asn	Ile	Ser	Ile	Glu	Asn	Leu	Ser	Ser	Asp	Ile	Ile	Gly	Gln	Leu				
		515					520					525							
Glu	Leu	Met	Pro	Asn	Ile	Glu	Arg	Phe	Pro	Asn	Gly	Lys	Lys	Tyr	Glu				
	530					535					540								
Leu	Asp	Lys	Tyr	Thr	Met	Phe	His	Tyr	Leu	Arg	Ala	Gln	Glu	Phe	Glu				
545					550					555					560				
His	Gly	Lys	Ser	Arg	Ile	Ala	Leu	Thr	Asn	Ser	Val	Asn	Glu	Ala	Leu				
				565					570						575				
Leu	Asn	Pro	Ser	Arg	Val	Tyr	Thr	Phe	Phe	Ser	Ser	Asp	Tyr	Val	Lys				
			580					585					590						
Lys	Val	Asn	Lys	Ala	Thr	Glu	Ala	Ala	Met	Phe	Leu	Gly	Trp	Val	Glu				
		595					600					605							
Gln	Leu	Val	Tyr	Asp	Phe	Thr	Asp	Glu	Thr	Ser	Glu	Val	Ser	Thr	Thr				
	610					615					620								
Asp	Lys	Ile	Ala	Asp	Ile	Thr	Ile	Ile	Ile	Pro	Tyr	Ile	Gly	Pro	Ala				
625					630					635					640				
Leu	Asn	Ile	Gly	Asn	Met	Leu	Tyr	Lys	Asp	Asp	Phe	Val	Gly	Ala	Leu				
				645					650					655					
Ile	Phe	Ser	Gly	Ala	Val	Ile	Leu	Leu	Glu	Phe	Ile	Pro	Glu	Ile	Ala				
			660					665					670						
Ile	Pro	Val	Leu	Gly	Thr	Phe	Ala	Leu	Val	Ser	Tyr	Ile	Ala	Asn	Lys				
		675					680					685							
Val	Leu	Thr	Val	Gln	Thr	Ile	Asp	Asn	Ala	Leu	Ser	Lys	Arg	Asn	Glu				
	690					695					700								
Lys	Trp	Asp	Glu	Val	Tyr	Lys	Tyr	Ile	Val	Thr	Asn	Trp	Leu	Ala	Lys				
705					710					715					720				
Val	Asn	Thr	Gln	Ile	Asp	Leu	Ile	Arg	Lys	Lys	Met	Lys	Glu	Ala	Leu				
				725					730					735					
Glu	Asn	Gln	Ala	Glu	Ala	Thr	Lys	Ala	Ile	Ile	Asn	Tyr	Gln	Tyr	Asn				
			740					745					750						
Gln	Tyr	Thr	Glu	Glu	Glu	Lys	Asn	Asn	Ile	Asn	Phe	Asn	Ile	Asp	Asp				
		755					760					765							
Leu	Ser	Ser	Lys	Leu	Asn	Glu	Ser	Ile	Asn	Lys	Ala	Met	Ile	Asn	Ile				
	770					775					780								
Asn	Lys	Phe	Leu	Asn	Gln	Cys	Ser	Val	Ser	Tyr	Leu	Met	Asn	Ser	Met				
785					790					795					800				
Ile	Pro	Tyr	Gly	Val	Lys	Arg	Leu	Glu	Asp	Phe	Asp	Ala	Ser	Leu	Lys				
				805					810					815					
Asp	Ala	Leu	Leu	Lys	Tyr	Ile	Tyr	Asp	Asn	Arg	Gly	Thr	Leu	Ile	Gly				
			820					825					830						
Gln	Val	Asp	Arg	Leu	Lys	Asp	Lys	Val	Asn	Asn	Thr	Leu	Ser	Thr	Asp				

ES 2 380 677 T3

	835					840					845						
Ile	Pro	Phe	Gln	Leu	Ser	Lys	Tyr	Val	Asp	Asn	Gln	Arg	Leu	Leu	Ser		
	850					855					860						
Thr	Phe	Thr	Glu	Tyr	Ile	Lys	Asn	Ile	Leu	Glu	His	His	His	His	His		
865						870					875					880	
His																	

<210> 5

<211> 3312

<212> ADN

5 <213> *Clostridium botulinum*

<400> 5

ES 2 380 677 T3

atggcatttg	ttaataaaca	atttaattat	aaagatcctg	taaatgggtg	tgatattgct	60
tatataaaaa	ttccaaatgc	aggacaaatg	caaccagtaa	aagcttttaa	aatttcataat	120
aaaatatggg	ttattccaga	aagagataca	tttaciaaat	ctgaagaagg	agattttaa	180
ccaccaccag	aagcaaaaaca	agttccagtt	tcatattatg	attcaacata	tttaagtaca	240
gataatgaaa	aagataatta	tttaaaggga	gttaciaaaa	tatttgagag	aattttattca	300
actgatcttg	gaagaatggt	gttaacatca	atagtaaggg	gaataccatt	ttgggggtgga	360
agtacaatag	atacagaatt	aaaagttatt	gatactaatt	gtattaatgt	gatacaacca	420
gatggtagtt	atagatcaga	agaacttaat	ctagtaataa	taggaccctc	agctgatatt	480
atacagtttg	aatgtaaaag	ctttggacat	gaagttttga	atcttacgag	aaatggttat	540
ggctctactc	aatacattag	atttagccca	gattttacat	ttgggttttg	ggagtcactt	600
gaagttgata	caaatcctct	tttaggtgca	ggcaaatttg	ctacagatcc	agcagtaaca	660
ttagcacatg	aacttatata	tgctggacat	agatttatatg	gaatagcaat	taatccaaat	720
aggggttttta	aagtaaatat	taatgcctat	tatgaaatga	gtgggttaga	agtaagcttt	780
gaggaactta	gaacatttgg	gggacatgat	gcaaagttta	tagatagttt	acaggaaaac	840
gaattttcgtc	tatatattta	taataagttt	aaagatatag	caagtacact	agataaagct	900
aaatcaatag	taggtactac	tgcttcatta	cagtatatga	aaaatgtttt	taaagagaaa	960
tatctcctat	ctgaagatac	atctggaaaa	ttttcggtag	ataaattaaa	atttgataag	1020
ttatacaaaa	tgtaaacaga	gatttacaca	gaggataatt	ttgttaagtt	ttttaagta	1080
cttaacagaa	aaacatattt	gaattttgat	aaagccgtat	ttaagataaa	tatagtacct	1140
aaggtaaat	acacaatata	tgatggattt	aatttaagaa	atacaaat	agcagcaaac	1200
tttaaatggtc	aaaatacaga	aattaataat	atgaatttta	ctaaactaaa	aaattttact	1260
ggattgtttg	aattttataa	gttgctatgt	gtaagaggga	taataaactc	taaaactaaa	1320
tcattagttc	cgctgtggat	caaggcctta	aatgattttat	gtatcaaagt	taataattgg	1380
gacttgtttt	ttagtccctc	agaagataat	tttactaatg	atctaaataa	aggagaagaa	1440
attacatctg	atactaata	agaagcagca	gaagaaaaata	ttagttaga	tttaatacaa	1500
caatattatt	taacctttta	tttgataaat	gaacctgaaa	atatttcaat	agaaaatctt	1560
tcaagtgaca	ttataggcca	attagaactt	atgcctaata	tagaaagatt	tcctaattgga	1620
aaaaagtatg	agttagataa	atatactatg	ttccattatc	ttcgtgctca	agaatttgaa	1680
catggtaaat	ctaggattgc	tttaacaaat	tctgttaacg	aagcattatt	aaatcctagt	1740
cgtgtttata	catttttttc	ttcagactat	gtaaagaaag	ttaataaagc	tacggaggca	1800
gctatgtttt	taggctgggt	agaacaatta	gtatatgatt	ttaccgatga	aactagcgaa	1860
gtaagtacta	cggataaaat	tgcgatata	actataatta	ttccatatat	aggacctgct	1920
ttaaatatag	gtaatatggt	atataaagat	gattttgtag	gtgctttaat	atttttcagga	1980
gctgttattc	tgttagaatt	tataccagag	attgcaatac	ctgtattagg	tacttttgca	2040
cttgatcat	atattgcgaa	taagggttcta	accgttcaaa	caatagataa	tgctttaagt	2100
aaaagaaatg	aaaaatggga	tgagggtctat	aaatatatag	taacaaattg	gttagcaaag	2160
gttaatacac	agattgatct	aataagaaaa	aaaatgaaag	aagctttaga	aaatcaagca	2220
gaagcaacaa	aggctataat	aaactatcag	tataatcaat	atactgagga	agagaaaaat	2280
aatatttaatt	ttaatattga	tgattttaagt	tcgaaaactta	atgagtctat	aaataaagct	2340
atgattaata	taaataaatt	tttgaatcaa	tgctctgttt	catattttaat	gaattctatg	2400
atcccttatg	gtgttaaacg	gttagaagat	tttgatgcta	gtcttaagaa	tgcatatta	2460
aagtatatat	atgataatag	aggaacttta	attgggtcaag	tagatagatt	aaaagataaa	2520
gttaataata	cacttagtac	agataacact	tttcagcttt	ccaaatacgt	agataatcaa	2580
agattattat	ctacatttac	tgaatatatt	aagaatatta	ttaataacttc	tatattgaa	2640
ttaaagatatg	aaagtaatca	tttaatatagac	ttatctaggt	atgcatcaaa	aataaatatt	2700
ggtagtaaaag	taaattttga	tccaatagat	aaaaatcaaa	ttcaattatt	taatttagaa	2760
agtagtaaaa	ttgaggtaat	tttaaaaaat	gctattgtat	ataatagtat	gtatgaaaat	2820
tttagtacta	gcttttggt	aagaattcct	aagtatttta	acagtataag	tctaataaat	2880
gaatatacaa	taataaattg	tatggaaaat	aattcaggat	ggaaagtatc	acttaattat	2940
ggtgaaataa	tctggacttt	acaggatact	caggaaataa	aacaaagagt	agtttttaaa	3000
tacagtcaaa	tgattaatat	atcagattat	ataaacagat	ggatttttgt	aactatcact	3060
aataatagat	taaataactc	taaaatttat	ataaatggaa	gattaataga	tcaaaaacca	3120
atttcaaatt	taggtaatat	tcatgctagt	aataatataa	tgtttaaat	agatggttgt	3180
agagatacac	atagatatat	ttggataaaa	tatttttaatc	tttttgataa	ggaattaaat	3240
gaaaaagaaa	tcaaagattt	atatgataat	caatcaaatt	cagggtctcga	gcaccaccac	3300
caccaccact	aa					3312

ES 2 380 677 T3

<210> 6
 <211> 1103
 <212> PRT
 <213> *Clostridium botulinum*

5 <400> 6

```

Met Ala Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1          5          10          15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Val Gly Gln Met Gln Pro
          20          25          30

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

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50          55          60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65          70          75          80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
          85          90          95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val
          100          105          110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
          115          120          125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130          135          140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145          150          155          160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr
          165          170          175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe
          180          185          190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu
          195          200          205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu
210          215          220
  
```

ES 2 380 677 T3

Leu	Ile	His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	Ile	Asn	Pro	Asn	
225					230					235					240	
Arg	Val	Phe	Lys	Val	Asn	Thr	Asn	Ala	Tyr	Tyr	Glu	Met	Ser	Gly	Leu	
				245					250					255		
Glu	Val	Ser	Phe	Glu	Glu	Leu	Arg	Thr	Phe	Gly	Gly	His	Asp	Ala	Lys	
			260					265					270			
Phe	Ile	Asp	Ser	Leu	Gln	Glu	Asn	Glu	Phe	Arg	Leu	Tyr	Tyr	Tyr	Asn	
		275					280					285				
Lys	Phe	Lys	Asp	Ile	Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys	Ser	Ile	Val	
	290						295				300					
Gly	Thr	Thr	Ala	Ser	Leu	Gln	Tyr	Met	Lys	Asn	Val	Phe	Lys	Glu	Lys	
305					310					315					320	
Tyr	Leu	Leu	Ser	Glu	Asp	Thr	Ser	Gly	Lys	Phe	Ser	Val	Asp	Lys	Leu	
				325					330					335		
Lys	Phe	Asp	Lys	Leu	Tyr	Lys	Met	Leu	Thr	Glu	Ile	Tyr	Thr	Glu	Asp	
			340					345					350			
Asn	Phe	Val	Lys	Phe	Phe	Lys	Val	Leu	Asn	Arg	Lys	Thr	Tyr	Leu	Asn	
		355					360					365				
Phe	Asp	Lys	Ala	Val	Phe	Lys	Ile	Asn	Ile	Val	Pro	Lys	Val	Asn	Tyr	
	370					375					380					
Thr	Ile	Tyr	Asp	Gly	Phe	Asn	Leu	Arg	Asn	Thr	Asn	Leu	Ala	Ala	Asn	
385					390					395					400	
Phe	Asn	Gly	Gln	Asn	Thr	Glu	Ile	Asn	Asn	Met	Asn	Phe	Thr	Lys	Leu	
				405					410					415		
Lys	Asn	Phe	Thr	Gly	Leu	Phe	Glu	Phe	Tyr	Lys	Leu	Leu	Cys	Val	Arg	
			420					425					430			
Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser	Leu	Val	Pro	Arg	Gly	Ser	Lys	
		435					440					445				
Ala	Leu	Asn	Asp	Leu	Cys	Ile	Lys	Val	Asn	Asn	Trp	Asp	Leu	Phe	Phe	
	450					455					460					
Ser	Pro	Ser	Glu	Asp	Asn	Phe	Thr	Asn	Asp	Leu	Asn	Lys	Gly	Glu	Glu	
465					470					475					480	
Ile	Thr	Ser	Asp	Thr	Asn	Ile	Glu	Ala	Ala	Glu	Glu	Asn	Ile	Ser	Leu	
				485					490					495		
Asp	Leu	Ile	Gln	Gln	Tyr	Tyr	Leu	Thr	Phe	Asn	Phe	Asp	Asn	Glu	Pro	
			500					505					510			
Glu	Asn	Ile	Ser	Ile	Glu	Asn	Leu	Ser	Ser	Asp	Ile	Ile	Gly	Gln	Leu	
		515					520					525				
Glu	Leu	Met	Pro	Asn	Ile	Glu	Arg	Phe	Pro	Asn	Gly	Lys	Lys	Tyr	Glu	
	530					535					540					
Leu	Asp	Lys	Tyr	Thr	Met	Phe	His	Tyr	Leu	Arg	Ala	Gln	Glu	Phe	Glu	
545					550					555					560	

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu
 565 570 575
 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
 580 585 590
 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
 595 600 605
 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr
 610 615 620
 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala
 625 630 635 640
 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu
 645 650 655
 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
 660 665 670
 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
 675 680 685
 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu
 690 695 700
 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
 705 710 715 720
 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu
 725 730 735
 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn
 740 745 750
 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
 755 760 765
 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile
 770 775 780
 Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met
 785 790 795 800
 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys
 805 810 815
 Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly
 820 825 830
 Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp
 835 840 845
 Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser
 850 855 860
 Thr Phe Thr Glu Tyr Ile Lys Asn Ile Ile Asn Thr Ser Ile Leu Asn
 865 870 875 880
 Leu Arg Tyr Glu Ser Asn His Leu Ile Asp Leu Ser Arg Tyr Ala Ser
 885 890 895

ES 2 380 677 T3

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

<210> 7

<211> 3942

<212> ADN

5 <213> *Clostridium botulinum*

<400> 7

ES 2 380 677 T3

atggcagtta	caataaataa	ttttaactat	aatgatacta	ttgataataa	taatattatt	60
atgatggagc	ctccatttgc	gagaggtacg	gggagatatt	ataaagcttt	taaaatcaca	120
gatcgtattt	ggataatacc	ggaaagatat	acttttggat	ataaacctga	ggatttttaat	180
aaaagttccg	gtatttttta	tagagatggt	tgtgaatatt	atgatccaga	ttactttaat	240
actaatgata	aaaagaatat	attttttaca	acaatgatca	agttatttaa	tagaatcaaa	300
tcaaaaccat	tgggtgaaaa	gttatttagag	atgattataa	atggatatac	ttatcttgga	360
gatagacgtg	ttccactcga	agagtttaac	acaaacattg	ctagtgtaac	tgtaataaaa	420
ttaatcagta	atccaggaga	agtggagcga	aaaaaaggta	ttttcgcaaa	tttaataata	480
tttggacctg	ggccagtttt	aaatgaaaat	gagactatag	atataggtat	acaaaatcat	540
tttgcacaa	gggaaggcct	cgggggtata	atgcaaataa	agttttgccc	agaatatgta	600
agcgtattta	ataatgttca	agaaaacaaa	ggcgcaagta	tatttaataa	acgtggatat	660
ttttcagatc	cagccttgat	attaatgcac	gaacttatat	atgtttttaca	tggattatat	720
ggcattaaag	tagatgattt	accaattgta	ccaaatgaaa	aaaaattttt	tatgcaatct	780

acagatgcta	tacaggcaga	agaactatat	acatttggag	gacaagatcc	cagcatcata	840
actccttcta	cggataaaaag	tatctatgat	aaagttttgc	aaaatttttag	agggatagtt	900
gatagactta	acaagggtttt	agtttgcata	tcagatcccta	acattaatat	taatataatat	960
aaaaataaat	ttaaagataa	atataaatct	gttgaagatt	ctgagggaaa	atatagtata	1020
gatgtagaaa	gttttgataa	attatataaa	agcttaaatgt	ttggttttac	agaaactaat	1080
atagcagaaa	attataaaaat	aaaaactaga	gcttcttatt	ttagtatttc	cttaccacca	1140
gtaaaaataa	aaaattttatt	agataatgaa	atctatacta	tagagggaagg	gtttaatata	1200
tctgataaag	atatggaaaa	agaatataga	ggtcagaata	aagctataaaa	taaaacaagct	1260
tatgaagaaa	ttagcaagga	gcatttggct	gtatataaga	tacaaatgtg	taaaagtggg	1320
ataataactt	ctaaaactaa	atcattgggt	ccacgtggat	ccaaggcctc	aggaatatgt	1380
attgatgttg	ataatgaaga	tttgttcttt	atagctgata	aaaatagttt	ttcagatgat	1440
ttatctaaaa	acgaaagaat	agaatataat	acacagagta	attatataga	aatgacttc	1500
cctataaatg	aattaattttt	agatactgat	ttaataagta	aaatagaatt	accaagtga	1560
aatacagaat	cacttactga	ttttaatgta	gatgttccag	tatatgaaaa	acaaccgcct	1620
ataaaaaaaaa	tttttacaga	tgaaaataacc	atctttcaat	atttatactc	tcagacattt	1680
cctctagata	taagagatat	aagtttaaca	tcttcatttg	atgatgcatt	attattttct	1740
aacaaagttt	attcattttt	ttctatggat	tatattaaaa	ctgctaataa	agtggtagaa	1800
gcaggattat	ttgcagggtg	ggtgaaacag	atgtaaaatg	attttgtaat	cgaagcgaat	1860
aaaagcaata	ctatggataa	aattgcagat	atatctctaa	ttgttcctta	tataggatta	1920
gctttaaatg	taggaaatga	aacagctaaa	ggaaattttg	aaaatgcttt	tgagattgca	1980
ggagccagta	ttctactaga	atttatacca	gaacttttaa	tacctgtagt	tggagccttt	2040
ttattagaat	catatattga	caataaaaaat	aaaattatta	aaacaataga	taatgcttta	2100
actaaaagaa	atgaaaaatg	gagtgatatg	tacggattaa	tagtagcgca	atggctctca	2160
acagttaata	ctcaatttta	tacaataaaa	gagggaaatgt	ataaggcctt	aaattatcaa	2220
gcacaagcat	tgggaagaaat	aataaaaatac	agatataata	tatatctctga	aaaaagaaaag	2280
tcaaatatta	acatcgattt	taatgatata	aattctaaac	ttaatgaggg	tattaaccaa	2340
gctatagata	atataaataa	ttttataaat	ggatgttctg	tatcatattt	aatgaaaaaa	2400
atgattccat	tagctgtaga	aaaattacta	gactttgata	atactctcaa	aaaaaatttg	2460
ttaaattata	tagatgaaaa	taaattatat	ttgattggaa	gtgcagaata	tgaaaaatca	2520
aaagtaataa	ataacttgaa	aaccattatg	cogtttgatc	tttcaatata	taccaatgat	2580
acaatactaa	tagaaatggt	taataaatat	aatagcgaaa	ttttaaataa	tattatctta	2640
aattttaagat	ataaggataa	taatttaata	gatttatcag	gatatggggc	aaaggtagag	2700
gtatatgarg	gagtcgagct	taatgataaa	aatcaattta	aattaactag	ttcagcaaat	2760
agtaagatta	gagtgactca	aaatcagaat	atcatattta	atagtgtgtt	ccttgatttt	2820
agcgttagct	tttgataag	aatacctaaa	tataagaatg	atggatataca	aaattatatt	2880
cataatgaat	atacaataat	taattgtatg	aaaaataatt	cgggctggaa	aatatctatt	2940
aggggtaata	ggataaatatg	gactttaatt	gatataaatg	gaaaaaccaa	atcgggtattt	3000
tttgaatata	acataagaga	agatatatca	gagtatataa	atagatgggt	ttttgtaact	3060
attactaata	atttgaataa	cgctaaaatt	tatattaatg	gtaagctaga	atcaaataca	3120
gatattaaag	atataagaga	agttattgct	aatgggtgaa	taatatttta	attagatggg	3180
gatatagata	gaacacaatt	tatttggatg	aatattttca	gtatttttaa	tacggaatta	3240
agtcaatcaa	atattgaaga	agatatataa	attcaatcat	atagcgaata	tttaaaagat	3300
ttttggggaa	atcctttta	gtacaataaaa	gaatattata	tgtttaaatgc	ggggaaataaa	3360
aattcatata	ttaaactaaa	gaaagattca	cctgtagggtg	aaatttttaac	acgtagcaaa	3420
tataatcaaa	attctaaata	tataaattat	agagatttat	atattggaga	aaaaatttatt	3480
ataagaagaa	agtcaaattc	tcaatctata	aatgatgata	tagttagaaa	agaagattat	3540
atatacttag	atttttttaa	tttaaatcaa	gagtgagag	tatataccta	taaatatttt	3600
aagaaagagg	aagaaaaatt	gttttttagct	cctataagtg	attctgatga	gtttttacaat	3660
actatacaaa	taaaagaata	tgatgaacag	ccaacatata	gttgctcagtt	gcttttttaa	3720
aaagatgaag	aaagtactga	tgagatagga	ttgattggta	ttcatcggtt	ctacgaatct	3780
ggaattgtat	ttgaagagta	taaagattat	ttttgtataa	gtaaatggta	cttaaaagag	3840
gtaaaaagga	aaccatataa	tttaaaattg	ggatgtaatt	ggcagtttat	tcctaaagat	3900
gaaggggtga	ctgaactcga	gcaccaccac	caccaccact	aa		3942

<210> 8

<211> 1313

<212> PRT

5 <213> *Clostridium botulinum*

<400> 8

ES 2 380 677 T3

Met Ala Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn

ES 2 380 677 T3

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

ES 2 380 677 T3

340										345					350				
Met	Phe	Gly	Phe	Thr	Glu	Thr	Asn	Ile	Ala	Glu	Asn	Tyr	Lys	Ile	Lys				
		355						360					365						
Thr	Arg	Ala	Ser	Tyr	Phe	Ser	Asp	Ser	Leu	Pro	Pro	Val	Lys	Ile	Lys				
	370						375				380								
Asn	Leu	Leu	Asp	Asn	Glu	Ile	Tyr	Thr	Ile	Glu	Glu	Gly	Phe	Asn	Ile				
385					390					395					400				
Ser	Asp	Lys	Asp	Met	Glu	Lys	Glu	Tyr	Arg	Gly	Gln	Asn	Lys	Ala	Ile				
				405					410					415					
Asn	Lys	Gln	Ala	Tyr	Glu	Glu	Ile	Ser	Lys	Glu	His	Leu	Ala	Val	Tyr				
			420					425					430						
Lys	Ile	Gln	Met	Cys	Lys	Ser	Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser				
		435					440						445						
Leu	Val	Pro	Arg	Gly	Ser	Lys	Ala	Ser	Gly	Ile	Cys	Ile	Asp	Val	Asp				
	450					455					460								
Asn	Glu	Asp	Leu	Phe	Phe	Ile	Ala	Asp	Lys	Asn	Ser	Phe	Ser	Asp	Asp				
465					470					475					480				
Leu	Ser	Lys	Asn	Glu	Arg	Ile	Glu	Tyr	Asn	Thr	Gln	Ser	Asn	Tyr	Ile				
				485					490					495					
Glu	Asn	Asp	Phe	Pro	Ile	Asn	Glu	Leu	Ile	Leu	Asp	Thr	Asp	Leu	Ile				
			500					505					510						
Ser	Lys	Ile	Glu	Leu	Pro	Ser	Glu	Asn	Thr	Glu	Ser	Leu	Thr	Asp	Phe				
		515						520					525						
Asn	Val	Asp	Val	Pro	Val	Tyr	Glu	Lys	Gln	Pro	Ala	Ile	Lys	Lys	Ile				
	530					535					540								
Phe	Thr	Asp	Glu	Asn	Thr	Ile	Phe	Gln	Tyr	Leu	Tyr	Ser	Gln	Thr	Phe				
545					550					555					560				
Pro	Leu	Asp	Ile	Arg	Asp	Ile	Ser	Leu	Thr	Ser	Ser	Phe	Asp	Asp	Ala				
				565					570					575					
Leu	Leu	Phe	Ser	Asn	Lys	Val	Tyr	Ser	Phe	Phe	Ser	Met	Asp	Tyr	Ile				
			580					585					590						
Lys	Thr	Ala	Asn	Lys	Val	Val	Glu	Ala	Gly	Leu	Phe	Ala	Gly	Trp	Val				
		595					600					605							
Lys	Gln	Ile	Val	Asn	Asp	Phe	Val	Ile	Glu	Ala	Asn	Lys	Ser	Asn	Thr				
	610					615					620								
Met	Asp	Lys	Ile	Ala	Asp	Ile	Ser	Leu	Ile	Val	Pro	Tyr	Ile	Gly	Leu				
625					630					635					640				
Ala	Leu	Asn	Val	Gly	Asn	Glu	Thr	Ala	Lys	Gly	Asn	Phe	Glu	Asn	Ala				
				645					650					655					
Phe	Glu	Ile	Ala	Gly	Ala	Ser	Ile	Leu	Leu	Glu	Phe	Ile	Pro	Glu	Leu				
			660					665					670						
Leu	Ile	Pro	Val	Val	Gly	Ala	Phe	Leu	Leu	Glu	Ser	Tyr	Ile	Asp	Asn				

ES 2 380 677 T3

675					680					685					
Lys	Asn	Lys	Ile	Ile	Lys	Thr	Ile	Asp	Asn	Ala	Leu	Thr	Lys	Arg	Asn
690					695					700					
Glu	Lys	Trp	Ser	Asp	Met	Tyr	Gly	Leu	Ile	Val	Ala	Gln	Trp	Leu	Ser
705					710					715					720
Thr	Val	Asn	Thr	Gln	Phe	Tyr	Thr	Ile	Lys	Glu	Gly	Met	Tyr	Lys	Ala
				725					730					735	
Leu	Asn	Tyr	Gln	Ala	Gln	Ala	Leu	Glu	Glu	Ile	Ile	Lys	Tyr	Arg	Tyr
			740					745					750		
Asn	Ile	Tyr	Ser	Glu	Lys	Glu	Lys	Ser	Asn	Ile	Asn	Ile	Asp	Phe	Asn
		755					760						765		
Asp	Ile	Asn	Ser	Lys	Leu	Asn	Glu	Gly	Ile	Asn	Gln	Ala	Ile	Asp	Asn
		770					775					780			
Ile	Asn	Asn	Phe	Ile	Asn	Gly	Cys	Ser	Val	Ser	Tyr	Leu	Met	Lys	Lys
785					790					795					800
Met	Ile	Pro	Leu	Ala	Val	Glu	Lys	Leu	Leu	Asp	Phe	Asp	Asn	Thr	Leu
			805						810					815	
Lys	Lys	Asn	Leu	Leu	Asn	Tyr	Ile	Asp	Glu	Asn	Lys	Leu	Tyr	Leu	Ile
		820						825					830		
Gly	Ser	Ala	Glu	Tyr	Glu	Lys	Ser	Lys	Val	Asn	Lys	Tyr	Leu	Lys	Thr
		835					840					845			
Ile	Met	Pro	Phe	Asp	Leu	Ser	Ile	Tyr	Thr	Asn	Asp	Thr	Ile	Leu	Ile
		850					855				860				
Glu	Met	Phe	Asn	Lys	Tyr	Asn	Ser	Glu	Ile	Leu	Asn	Asn	Ile	Ile	Leu
865					870					875					880
Asn	Leu	Arg	Tyr	Lys	Asp	Asn	Asn	Leu	Ile	Asp	Leu	Ser	Gly	Tyr	Gly
				885				890						895	
Ala	Lys	Val	Glu	Val	Tyr	Asp	Gly	Val	Glu	Leu	Asn	Asp	Lys	Asn	Gln
			900				905						910		
Phe	Lys	Leu	Thr	Ser	Ser	Ala	Asn	Ser	Lys	Ile	Arg	Val	Thr	Gln	Asn
		915					920					925			
Gln	Asn	Ile	Ile	Phe	Asn	Ser	Val	Phe	Leu	Asp	Phe	Ser	Val	Ser	Phe
		930					935				940				
Trp	Ile	Arg	Ile	Pro	Lys	Tyr	Lys	Asn	Asp	Gly	Ile	Gln	Asn	Tyr	Ile
945					950					955					960
His	Asn	Glu	Tyr	Thr	Ile	Ile	Asn	Cys	Met	Lys	Asn	Asn	Ser	Gly	Trp
			965						970					975	
Lys	Ile	Ser	Ile	Arg	Gly	Asn	Arg	Ile	Ile	Trp	Thr	Leu	Ile	Asp	Ile
			980				985						990		
Asn	Gly	Lys	Thr	Lys	Ser	Val	Phe	Phe	Glu	Tyr	Asn	Ile	Arg	Glu	Asp
		995					1000					1005			
Ile	Ser	Glu	Tyr	Ile	Asn	Arg	Trp	Phe	Phe	Val	Thr	Ile	Thr	Asn	

ES 2 380 677 T3

1010	1015	1020
Asn Leu 1025	Asn Asn Ala Lys Ile 1030	Tyr Ile Asn Gly Lys Leu Glu Ser 1035
Asn Thr 1040	Asp Ile Lys Asp Ile 1045	Arg Glu Val Ile Ala Asn Gly Glu 1050
Ile Ile 1055	Phe Lys Leu Asp Gly 1060	Asp Ile Asp Arg Thr Gln Phe Ile 1065
Trp Met 1070	Lys Tyr Phe Ser Ile 1075	Phe Asn Thr Glu Leu Ser Gln Ser 1080
Asn Ile 1085	Glu Glu Arg Tyr Lys 1090	Ile Gln Ser Tyr Ser Glu Tyr Leu 1095
Lys Asp 1100	Phe Trp Gly Asn Pro 1105	Leu Met Tyr Asn Lys Glu Tyr Tyr 1110
Met Phe 1115	Asn Ala Gly Asn Lys 1120	Asn Ser Tyr Ile Lys Leu Lys Lys 1125
Asp Ser 1130	Pro Val Gly Glu Ile 1135	Leu Thr Arg Ser Lys Tyr Asn Gln 1140
Asn Ser 1145	Lys Tyr Ile Asn Tyr 1150	Arg Asp Leu Tyr Ile Gly Glu Lys 1155
Phe Ile 1160	Ile Arg Arg Lys Ser 1165	Asn Ser Gln Ser Ile Asn Asp Asp 1170
Ile Val 1175	Arg Lys Glu Asp Tyr 1180	Ile Tyr Leu Asp Phe Phe Asn Leu 1185
Asn Gln 1190	Glu Trp Arg Val Tyr 1195	Thr Tyr Lys Tyr Phe Lys Lys Glu 1200
Glu Glu 1205	Lys Leu Phe Leu Ala 1210	Pro Ile Ser Asp Ser Asp Glu Phe 1215
Tyr Asn 1220	Thr Ile Gln Ile Lys 1225	Glu Tyr Asp Glu Gln Pro Thr Tyr 1230
Ser Cys 1235	Gln Leu Leu Phe Lys 1240	Lys Asp Glu Glu Ser Thr Asp Glu 1245
Ile Gly 1250	Leu Ile Gly Ile His 1255	Arg Phe Tyr Glu Ser Gly Ile Val 1260
Phe Glu 1265	Glu Tyr Lys Asp Tyr 1270	Phe Cys Ile Ser Lys Trp Tyr Leu 1275
Lys Glu 1280	Val Lys Arg Lys Pro 1285	Tyr Asn Leu Lys Leu Gly Cys Asn 1290
Trp Gln 1295	Phe Ile Pro Lys Asp 1300	Glu Gly Trp Thr Glu Leu Glu His 1305
His His 1310	His His His	

<210> 9
 <211> 2649
 <212> ADN
 <213> *Clostridium botulinum*

5 <400> 9

```

atggcagtta caataaataa ttttaattat aatgatccta ttgataataa taatattatt      60
atgatggagc ctccatttgc gagaggtagc gggagatatt ataaagcttt taaaatcaca      120
gatcgtattt ggataaatacc ggaaagatat acttttggat ataaacctga ggattttaat : 180
aaaagtcccg gtatttttaa tagagatggt tgtgaatatt atgatccaga ttacttaaat      240
actaatgata aaaagaatat atttttacaa acaatgatca agttatttaa tagaatcaaa      300
tcaaaaccat tgggtgaaaa gttatttagag atgattataa atgggtatacc ttatcttggg      360
gatagacgtg ttccactcga agagtttaac acaaacattg ctagtgtaac tgttaataaa      420
ttaatcagta atccaggaga agtggagcga aaaaaaggta ttttcgcaaa ttttaataa      480
tttggacctg ggccagtttt aaatgaaaat gagactatag atataggtat acaaaatcat      540
tttgcatacaa ggaaggcctt cgggggtata atgcaaatga agttttgccc agtttttgta      600
agcgtatttta ataagtgtca agaaaacaaa ggcgcaagta tatttaatag acgtggatat      660
ttttcagatc cagccttgat attaatgcat gaacttatac atgtttttaca tggattatat      720
ggcattaaag tagatgattt accaattgta ccaaatgaaa aaaaattttt tatgcaatct      780
acagatgcta tacaggcaga agaactatat acatttggag gacaagatcc cagcatcata      840
actccttcta cggataaaaag tatctatgat aaagttttgc aaaatttttag agggatagtt      900
gatagactta acaagggtttt agtttgcata tcagatccta acattaatat taatatatat      960
aaaaataaat ttaaagataa atataaattc gttgaagatt ctgagggaaa atatagtata     1020
gatgtagaaa gttttgataa attatataaa agcttaatgt ttggttttac agaaactaat     1080
atagcagaaa attataaaat aaaaactaga gcttcttatt ttagtgattc cttaccacca     1140
gtaaaaataa aaaattttatt agataatgaa atctatacta tagaggaagg gtttaataa     1200
tctgataaaag atatggaaaa agaatataga ggtcagaata aagctataaa taaacaagct     1260
tatgaagaaa tttagcaagga gcatttggct gtatataaga tacaagtgtg taaaagtggg     1320
ataataactt ctaaaactaa atcattgggt ccacgtggat ccaaggcctc aggaatatgt     1380
attgatgttg ataatgaaga tttgttcttt atagctgata aaaatagttt ttcagatgat     1440
ttatctaaaa acgaaagaat agaataaat acacagagta attatataga aaatgacttc     1500
cctataaatg aattaatttt agatactgat ttaataagta aaatagaatt accaagtga     1560
aatacagaat cacttactga ttttaatgta gatgttccag tatatgaaaa acaaccgct     1620
ataaaaaaaa tttttacaga tgaataatcc atctttcaat atttatactc tcagacattt     1680
cctctagata taagagatat aagtttaaca tcttcatttg atgatgcatt attattttct     1740
aacaaggttt attcattttt ttctatggat tatattaaaa ctgctaataa agtggttagaa     1800
gcaggattat ttgcagggtg ggtgaaacag atagtaaattg attttgtaat cgaagctaat     1860
aaaagcaata ctatggataa aattgcagat atatctctaa ttgttcctta tataggatta     1920
gcttttaaatg taggaaatga aacagctaaa ggaaattttg aaaatgcttt tgagattgca     1980
ggagccagta ttctactaga atttatacca gaacttttaa tacctgtagt tggagccttt     2040
ttattagaat catatattga caataaaaaa aaaattatta aaacaataga taatgcttta     2100
actaaaagaa atgaaaaatg gagtgatatg tacggattaa tagtagcgca atggctctca     2160
acagttaata ctcaatttta tacaataaaa gagggaaatgt ataaggcttt aaattatcaa     2220
gcacaagcat tggaagaaat aataaaatcc agatataata tatattctga aaaagaaaag     2280
tcaaatatta acatcgattt taatgatata aattctaaac ttaatgaggg tattaaccaa     2340
gctatagata atataaataa ttttataaat ggatgttctg tatcatattt aatgaaaaaa     2400
atgattccat tagctgtaga aaaattacta gactttgata atactctcaa aaaaaatttg     2460
ttaaattata tagatgaaaa taaattatat ttgattggaa gtgcagaata tgaaaaatca     2520
aaagtaataa aatacttgaa aaccattatg ccgtttgatc tttcaatata taccaatgat     2580
acaatactaa tagaaatggt taataaatat aatagcgaaa ttctcgagca ccaccaccac     2640
caccactaa                                     .                               2649
    
```

<210> 10
 <211> 882
 <212> PRT
 <213> *Clostridium botulinum*

10

<400> 10

ES 2 380 677 T3

Met	Ala	Val	Thr	Ile	Asn	Asn	Phe	Asn	Tyr	Asn	Asp	Pro	Ile	Asp	Asn
1				5					10					15	

ES 2 380 677 T3

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

ES 2 380 677 T3

Met	Phe	Gly	Phe	Thr	Glu	Thr	Asn	Ile	Ala	Glu	Asn	Tyr	Lys	Ile	Lys	355	360	365
Thr	Arg	Ala	Ser	Tyr	Phe	Ser	Asp	Ser	Leu	Pro	Pro	Val	Lys	Ile	Lys	370	375	380
Asn	Leu	Leu	Asp	Asn	Glu	Ile	Tyr	Thr	Ile	Glu	Glu	Gly	Phe	Asn	Ile	385	390	395
Ser	Asp	Lys	Asp	Met	Glu	Lys	Glu	Tyr	Arg	Gly	Gln	Asn	Lys	Ala	Ile	405	410	415
Asn	Lys	Gln	Ala	Tyr	Glu	Glu	Ile	Ser	Lys	Glu	His	Leu	Ala	Val	Tyr	420	425	430
Lys	Ile	Gln	Met	Cys	Lys	Ser	Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser	435	440	445
Leu	Val	Pro	Arg	Gly	Ser	Lys	Ala	Ser	Gly	Ile	Cys	Ile	Asp	Val	Asp	450	455	460
Asn	Glu	Asp	Leu	Phe	Phe	Ile	Ala	Asp	Lys	Asn	Ser	Phe	Ser	Asp	Asp	465	470	475
Leu	Ser	Lys	Asn	Glu	Arg	Ile	Glu	Tyr	Asn	Thr	Gln	Ser	Asn	Tyr	Ile	485	490	495
Glu	Asn	Asp	Phe	Pro	Ile	Asn	Glu	Leu	Ile	Leu	Asp	Thr	Asp	Leu	Ile	500	505	510
Ser	Lys	Ile	Glu	Leu	Pro	Ser	Glu	Asn	Thr	Glu	Ser	Leu	Thr	Asp	Phe	515	520	525
Asn	Val	Asp	Val	Pro	Val	Tyr	Glu	Lys	Gln	Pro	Ala	Ile	Lys	Lys	Ile	530	535	540
Phe	Thr	Asp	Glu	Asn	Thr	Ile	Phe	Gln	Tyr	Leu	Tyr	Ser	Gln	Thr	Phe	545	550	555
Pro	Leu	Asp	Ile	Arg	Asp	Ile	Ser	Leu	Thr	Ser	Ser	Phe	Asp	Asp	Ala	565	570	575
Leu	Leu	Phe	Ser	Asn	Lys	Val	Tyr	Ser	Phe	Phe	Ser	Met	Asp	Tyr	Ile	580	585	590
Lys	Thr	Ala	Asn	Lys	Val	Val	Glu	Ala	Gly	Leu	Phe	Ala	Gly	Trp	Val	595	600	605
Lys	Gln	Ile	Val	Asn	Asp	Phe	Val	Ile	Glu	Ala	Asn	Lys	Ser	Asn	Thr	610	615	620
Met	Asp	Lys	Ile	Ala	Asp	Ile	Ser	Leu	Ile	Val	Pro	Tyr	Ile	Gly	Leu	625	630	635
Ala	Leu	Asn	Val	Gly	Asn	Glu	Thr	Ala	Lys	Gly	Asn	Phe	Glu	Asn	Ala	645	650	655
Phe	Glu	Ile	Ala	Gly	Ala	Ser	Ile	Leu	Leu	Glu	Phe	Ile	Pro	Glu	Leu	660	665	670
Leu	Ile	Pro	Val	Val	Gly	Ala	Phe	Leu	Leu	Glu	Ser	Tyr	Ile	Asp	Asn	675	680	685

ES 2 380 677 T3

Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn
 690 695 700
 Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser
 705 710 715 720
 Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala
 725 730 735
 Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr
 740 745 750
 Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn
 755 760 765
 Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn
 770 775 780
 Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys
 785 790 795 800
 Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu
 805 810 815
 Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile
 820 825 830
 Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr
 835 840 845
 Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile
 850 855 860
 Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Glu His His His His
 865 870 875 880
 His His

<210> 11

<211> 3885

<212> ADN

5 <213> *Clostridium botulinum*

<400> 11

ES 2 380 677 T3

atggcaataa	caattaacaa	ctttaattat	tcagatcctg	ttgataataa	aaatatttta	60
tatttagata	ctcatttaaa	taçatttagct	aatgagcctg	aaaaagcctt	tcgcattata	120
gggaatatat	gggtaâtacc	cgatagattt	tcaagagatt	ctaataccaaa	tttaaataaaa	180
cctccacgag	ttacaagccc	taaaagtgg	tattatgatc	ctaattattt	gagtactgat	240
tctgaaaaag	atacattttt	aaaagaaatt	ataaagttat	ttaaaagaat	taactctaga	300
gaaataggag	aagaattaat	atatagactt	gcaacagaca	taccctttcc	tggaataaac	360
aatactccaa	ttaatacttt	tgattttgat	gtagatttta	acagtgttga	tgtaaaact	420
agacaaggta	acaactgggt	taaaactgg	agtataaatc	ctagtgttat	aataactgga	480
cctagagaaa	acattataga	cccagaaact	tctacgttta	aattaactaa	caatactttt	540
gcggcacaag	aaggatttgg	tgctttatca	ataatttcaa	tatcacctag	atttatgcta	600
acatatagta	atgcaactaa	taatgtagga	gagggtagat	tttctaagtc	tgaattttgc	660
atggatccaa	tactaatttt	aatgcatgaa	cttaatcatg	caatgcataa	tttatatgga	720
atagctatac	caaatgatca	aagaatttca	tctgtactaa	gtaatatattt	ttattctcaa	780
tataatgtga	aattagagta	tgcaaaaata	tatgcatttg	gaggtccaac	tatagacctt	840
attcctaaaa	gtgcaaggaa	atattttgag	gaaaaggcat	tggattatta	tagatccata	900

gctaaaagac	ttaatagtat	aactactgca	aatccttcaa	gctttaataa	atatatagga	960
gaatataaac	agaaacttat	tagaaagtat	agattcgtag	tagaatcttc	agggtgaagtt	1020
gcagtagatc	gtaataagtt	tgctgagtta	tataaagaac	ttacacaaat	atttacagaa	1080
tttaactacg	ctaaaaatata	taatgtacaa	aataggaaaa	tatatctttc	aaatgtatat	1140
actccggtta	cggcaaatat	attagacgat	aatggtttatg	atatacaaaa	tggttttaac	1200
atacctaaaa	gtaattttaa	tgtactattt	atgggtcaaa	atztatctcg	aaatccagca	1260
ttaagaaaag	tcaatcctga	aaatatgctt	tattttattta	caaaattttg	tcatagaggg	1320
ataataactt	ctaaaaactaa	atcattagtt	ccgcgtggat	ccaaggcctt	agattgtaga	1380
gagcttttag	ttaaaaatac	tgacttacct	tttataggtg	atattagtga	tatcaaaact	1440
gatatatttt	taagcaaaga	tattaatgaa	gaaactgaag	ttatagacta	tccggacaat	1500
gtttcagtg	atcaagttat	tctcagtaag	aatacctcag	aacatggaca	actagattta	1560
ttatacccta	ttattgaagg	tgagagtcaa	gtattaccgg	gagagaatca	agtcttttat	1620
gataatagaa	ctcaaaatgt	tgattatttg	aattcttatt	attacctaga	atctcaaaaa	1680
ctaagtgata	atgttgaaga	ttttactttt	acgacatcaa	ttgaggaagc	tttggataat	1740
agtggaaaag	tatatactta	ctttcctaaa	ctagctgata	aagtaaatac	gggtgttcaa	1800
ggtggtttat	ttttaatgtg	ggcaaatgat	gtagttgaag	attttactac	aaatattcta	1860
agaaaagata	cattagataa	aatatcagat	gtatcagcta	ttattcccta	tataggacct	1920
gcattaaata	taagtaattc	tgtaagaagg	ggaaatttta	ctgaagcatt	tcaggttacc	1980
ggtgtaacta	ttttattaga	agcgtttcaa	gaattttaca	tacctgcact	tggtgcattt	2040
gtgatttata	gtaaggttca	agaaagaaac	gagattatta	aaactataga	taattgttta	2100
gaacaaaagg	ttaaaagatg	gaaagattca	tatgaatgga	tgataggaac	gtgggttatcc	2160
aggattacta	ctcaatttta	taatataagt	tatcaaatgt	atgattcttt	aaattatcag	2220
gcagatgcaa	tcaaagataa	aatagattta	gaatataaaa	aatactcagg	aagtgataaa	2280
gaaaatataa	aaagtcaggt	tgaaaattta	aaaaatagtt	tagatataaa	aatctcggaa	2340
gcaatgaata	atataaataa	atttatacga	gaatgttctg	taacatactt	atttaaaaat	2400
atgctcccta	aagtaattga	tgaattaaat	aagtttgatt	taaaaactaa	aacagaatta	2460
attaacttta	tagatagtca	taatatattt	ctagttgggt	aagtagatag	attaaaagca	2520
aaaataaatg	agagttttga	aaatacaata	ccctttaata	ttttttcata	tactaataat	2580
tctttattaa	aagatatact	cgaggaatat	ttcaatagta	ttaatgattc	aaaaattttg	2640
agcttacaaa	acaaaaaaaa	tgcttttagtg	gatacatcag	gatataatgc	agaagtgagg	2700
ctagaagggt	atgttcaagt	taatacgata	tatacaaatg	attttaaat	aagtagttca	2760
ggagataaaa	ttatagtaaa	tttaataaat	aatattttat	atagcgctat	ttatgagaac	2820
tctagtgtta	gtttttggat	taagatatct	aaagatttaa	ctaattctca	taatgaatat	2880
actataatta	atagtataaa	acaaaattct	gggtggaaat	tatgtattag	aaatggcaat	2940
atagaatgga	ttttacaaga	tattaataga	aagtataaaa	gtttaatttt	tgattatagt	3000
gaatcattaa	gtcatacagg	atatacaaat	aaatggtttt	ttgttactat	aactaataat	3060
ataatggggt	atatgaaact	ttatataaat	ggagaattaa	agcagagtga	aagaattgaa	3120
gatttagatg	aggttaagtt	agataaaacc	atagtatttg	gaatagatga	gaatagatag	3180
gagaatcaga	tgctttggat	tagagatttt	aatatttttt	ctaaagaatt	aagcaatgaa	3240
gataattaata	ttgtatatga	gggacaaata	ttaaagaaatg	ttattaaaga	ttattgggga	3300
aatcctttga	agtttgatac	agaatattat	atgattaatt	ataattatat	agataggtat	3360
atagcaccta	aaaataatat	acttgtactt	gttcagtatt	cagatatatc	taaattatat	3420
actaaaaatc	ctattactat	taaatcagca	gctaataaga	atccttatag	tagaatttta	3480
aatggagatg	atataatggt	tcatatgtta	tatgatagta	gggaatatat	gataataaga	3540
gatactgata	caatatatgc	aacacaagga	ggacagtgtt	caaaaaattg	tgtatatgca	3600
ttaaaattac	agagtaattt	aggtaattat	ggtataggta	tatttagtat	aaaaaatatt	3660
gtatctcaaa	ataaatattg	tagtcaaatt	ttctctagtt	ttatgaaaaa	tacaatgctt	3720
ctagcagata	tatataaacc	ttggagattt	tcttttgaaa	atgcatacac	gccagttgca	3780
gtaactaatt	atgagacaaa	actattatca	acttcatctt	tttggaaatt	tatttctagg	3840
gatccaggat	gggtagagct	cgagcaccac	caccaccacc	actaa		3885

<210> 12

<211> 1294

<212> PRT

5 <213> *Clostridium botulinum*

<400> 12

ES 2 380 677 T3

Met	Ala	Ile	Thr	Ile	Asn	Asn	Phe	Asn	Tyr	Ser	Asp	Pro	Val	Asp	Asn
1				5					10					15	
Lys	Asn	Ile	Leu	Tyr	Leu	Asp	Thr	His	Leu	Asn	Thr	Leu	Ala	Asn	Glu

ES 2 380 677 T3

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

ES 2 380 677 T3

355		360		365
Val Gln Asn Arg Lys Ile Tyr Leu Ser Asn Val Tyr Thr Pro Val Thr				
370		375		380
Ala Asn Ile Leu Asp Asp Asn Val Tyr Asp Ile Gln Asn Gly Phe Asn				
385		390		395
Ile Pro Lys Ser Asn Leu Asn Val Leu Phe Met Gly Gln Asn Leu Ser				
	405		410	415
Arg Asn Pro Ala Leu Arg Lys Val Asn Pro Glu Asn Met Leu Tyr Leu				
	420		425	430
Phe Thr Lys Phe Cys His Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser				
	435		440	445
Leu Val Pro Arg Gly Ser Lys Ala Leu Asp Cys Arg Glu Leu Leu Val				
	450		455	460
Lys Asn Thr Asp Leu Pro Phe Ile Gly Asp Ile Ser Asp Ile Lys Thr				
465		470		475
Asp Ile Phe Leu Ser Lys Asp Ile Asn Glu Glu Thr Glu Val Ile Asp				
	485		490	495
Tyr Pro Asp Asn Val Ser Val Asp Gln Val Ile Leu Ser Lys Asn Thr				
	500		505	510
Ser Glu His Gly Gln Leu Asp Leu Leu Tyr Pro Ile Ile Glu Gly Glu				
	515		520	525
Ser Gln Val Leu Pro Gly Glu Asn Gln Val Phe Tyr Asp Asn Arg Thr				
	530		535	540
Gln Asn Val Asp Tyr Leu Asn Ser Tyr Tyr Tyr Leu Glu Ser Gln Lys				
545		550		555
Leu Ser Asp Asn Val Glu Asp Phe Thr Phe Thr Thr Ser Ile Glu Glu				
	565		570	575
Ala Leu Asp Asn Ser Gly Lys Val Tyr Thr Tyr Phe Pro Lys Leu Ala				
	580		585	590
Asp Lys Val Asn Thr Gly Val Gln Gly Gly Leu Phe Leu Met Trp Ala				
	595		600	605
Asn Asp Val Val Glu Asp Phe Thr Thr Asn Ile Leu Arg Lys Asp Thr				
	610		615	620
Leu Asp Lys Ile Ser Asp Val Ser Ala Ile Ile Pro Tyr Ile Gly Pro				
625		630		635
Ala Leu Asn Ile Ser Asn Ser Val Arg Arg Gly Asn Phe Thr Glu Ala				
	645		650	655
Phe Ala Val Thr Gly Val Thr Ile Leu Leu Glu Ala Phe Gln Glu Phe				
	660		665	670
Thr Ile Pro Ala Leu Gly Ala Phe Val Ile Tyr Ser Lys Val Gln Glu				
	675		680	685
Arg Asn Glu Ile Ile Lys Thr Ile Asp Asn Cys Leu Glu Gln Arg Ile				

ES 2 380 677 T3

690	695	700
Lys Arg Trp Lys Asp Ser Tyr Glu Trp Met Ile Gly Thr Trp Leu Ser 705 710 715 720		
Arg Ile Thr Thr Gln Phe Asn Asn Ile Ser Tyr Gln Met Tyr Asp Ser 725 730 735		
Leu Asn Tyr Gln Ala Asp Ala Ile Lys Asp Lys Ile Asp Leu Glu Tyr 740 745 750		
Lys Lys Tyr Ser Gly Ser Asp Lys Glu Asn Ile Lys Ser Gln Val Glu 755 760 765		
Asn Leu Lys Asn Ser Leu Asp Ile Lys Ile Ser Glu Ala Met Asn Asn 770 775 780		
Ile Asn Lys Phe Ile Arg Glu Cys Ser Val Thr Tyr Leu Phe Lys Asn 785 790 795 800		
Met Leu Pro Lys Val Ile Asp Glu Leu Asn Lys Phe Asp Leu Lys Thr 805 810 815		
Lys Thr Glu Leu Ile Asn Leu Ile Asp Ser His Asn Ile Ile Leu Val 820 825 830		
Gly Glu Val Asp Arg Leu Lys Ala Lys Ile Asn Glu Ser Phe Glu Asn 835 840 845		
Thr Ile Pro Phe Asn Ile Phe Ser Tyr Thr Asn Asn Ser Leu Leu Lys 850 855 860		
Asp Ile Leu Glu Glu Tyr Phe Asn Ser Ile Asn Asp Ser Lys Ile Leu 865 870 875 880		
Ser Leu Gln Asn Lys Lys Asn Ala Leu Val Asp Thr Ser Gly Tyr Asn 885 890 895		
Ala Glu Val Arg Leu Glu Gly Asp Val Gln Val Asn Thr Ile Tyr Thr 900 905 910		
Asn Asp Phe Lys Leu Ser Ser Ser Gly Asp Lys Ile Ile Val Asn Leu 915 920 925		
Asn Asn Asn Ile Leu Tyr Ser Ala Ile Tyr Glu Asn Ser Ser Val Ser 930 935 940		
Phe Trp Ile Lys Ile Ser Lys Asp Leu Thr Asn Ser His Asn Glu Tyr 945 950 955 960		
Thr Ile Ile Asn Ser Ile Lys Gln Asn Ser Gly Trp Lys Leu Cys Ile 965 970 975		
Arg Asn Gly Asn Ile Glu Trp Ile Leu Gln Asp Ile Asn Arg Lys Tyr 980 985 990		
Lys Ser Leu Ile Phe Asp Tyr Ser Glu Ser Leu Ser His Thr Gly Tyr 995 1000 1005		
Thr Asn Lys Trp Phe Phe Val Thr Ile Thr Asn Asn Ile Met Gly 1010 1015 1020		
Tyr Met Lys Leu Tyr Ile Asn Gly Glu Leu Lys Gln Ser Glu Arg		

ES 2 380 677 T3

1025					1030					1035				
Ile	Glu	Asp	Leu	Asp	Glu	Val	Lys	Leu	Asp	Lys	Thr	Ile	Val	Phe
1040						1045					1050			
Gly	Ile	Asp	Glu	Asn	Ile	Asp	Glu	Asn	Gln	Met	Leu	Trp	Ile	Arg
1055						1060					1065			
Asp	Phe	Asn	Ile	Phe	Ser	Lys	Glu	Leu	Ser	Asn	Glu	Asp	Ile	Asn
1070						1075					1080			
Ile	Val	Tyr	Glu	Gly	Gln	Ile	Leu	Arg	Asn	Val	Ile	Lys	Asp	Tyr
1085						1090					1095			
Trp	Gly	Asn	Pro	Leu	Lys	Phe	Asp	Thr	Glu	Tyr	Tyr	Met	Ile	Asn
1100						1105					1110			
Tyr	Asn	Tyr	Ile	Asp	Arg	Tyr	Ile	Ala	Pro	Lys	Asn	Asn	Ile	Leu
1115						1120					1125			
Val	Leu	Val	Gln	Tyr	Ser	Asp	Ile	Ser	Lys	Leu	Tyr	Thr	Lys	Asn
1130						1135					1140			
Pro	Ile	Thr	Ile	Lys	Ser	Ala	Ala	Asn	Lys	Asn	Pro	Tyr	Ser	Arg
1145						1150					1155			
Ile	Leu	Asn	Gly	Asp	Asp	Ile	Met	Phe	His	Met	Leu	Tyr	Asp	Ser
1160						1165					1170			
Arg	Glu	Tyr	Met	Ile	Ile	Arg	Asp	Thr	Asp	Thr	Ile	Tyr	Ala	Thr
1175						1180					1185			
Gln	Gly	Gly	Gln	Cys	Ser	Lys	Asn	Cys	Val	Tyr	Ala	Leu	Lys	Leu
1190						1195					1200			
Gln	Ser	Asn	Leu	Gly	Asn	Tyr	Gly	Ile	Gly	Ile	Phe	Ser	Ile	Lys
1205						1210					1215			
Asn	Ile	Val	Ser	Gln	Asn	Lys	Tyr	Cys	Ser	Gln	Ile	Phe	Ser	Ser
1220						1225					1230			
Phe	Met	Lys	Asn	Thr	Met	Leu	Leu	Ala	Asp	Ile	Tyr	Lys	Pro	Trp
1235						1240					1245			
Arg	Phe	Ser	Phe	Glu	Asn	Ala	Tyr	Thr	Pro	Val	Ala	Val	Thr	Asn
1250						1255					1260			
Tyr	Glu	Thr	Lys	Leu	Leu	Ser	Thr	Ser	Ser	Phe	Trp	Lys	Phe	Ile
1265						1270					1275			
Ser	Arg	Asp	Pro	Gly	Trp	Val	Glu	Leu	Glu	His	His	His	His	His
1280						1285					1290			

His

<210> 13

<211> 2625

<212> ADN

5 <213> *Clostridium botulinum*

<400> 13

atggcaataa	caattaacaa	ctttaattat	tcagatcctg	ttgataataa	aaatatattta	60
tatttagata	ctcattttaa	tacattagct	aatgagcctg	aaaaagcctt	tcgcattata	120
gggaatatat	gggtaatacc	cgatagattt	tcaagagatt	ctaattccaa	tttaaataaaa	180
cctccacgag	ttacaagccc	taaaagtggg	tattatgatc	ctaattattt	gagtactgat	240
tctgaaaaag	atacatTTTT	aaaagaaatt	ataaagttat	ttaaaagaat	taactctaga	300
gaaataggag	aagaattaat	atatagactt	gcaacagaca	taccctttcc	tggaataaac	360
aatactccaa	ttaatacttt	tgattttgat	gtagatttta	acagtgttga	tgttaaaact	420
agacaaggta	acaactgggt	taaaactggg	agtataaatc	ctagtgttat	aataactgga	480
cctagagaaa	acattataga	cccagaaact	tctacgttta	aattaactaa	caatactttt	540
gcggcacaag	aaggatttgg	tgctttatca	ataatttcaa	tatcacctag	atztatgcta	600
acatatagta	atgcaactaa	taatgtagga	gagggtagat	tttctaagtc	tgaattttgc	660
atggatccaa	tactaatTTT	aatgcatgaa	cttaatcatg	caatgcataa	tttatatgga	720
atagctatac	caaatgatca	aagaatttca	tctgtaacta	gtaatatTTT	ttattctcaa	780
tataatgtga	aattagagta	tgacagaaata	tatgcatttg	gaggtccaac	tatagacctt	840
attcctaata	gtgcaaggaa	atattttgag	gaaaaggcat	tggtattatta	tagatocata	900
gctaaaagac	ttaatagtat	aactactgca	aatccttcaa	gctttaataa	atatatagga	960
gaatataaac	agaaacttat	tagaaagtat	agattcgtag	tagaatcttc	aggtgaagtt	1020
gcagtagatc	gtaataagtt	tgctgagtta	tataaagaac	ttacacaaat	atttacagaa	1080
tttaactacg	ctaaaatata	taatgtacaa	aataggaaaa	tatatctttc	aaatgtatat	1140
actccggtta	cggcacaatat	attagacgat	aatgtttatg	atatacaaaa	tggtatttaac	1200
atacctaata	gtaattttaa	tgtactattt	atgggtcaaa	atztatctcg	aaatccagca	1260
ttaagaaaag	tcaatcctga	aaatatgctt	tattttttta	caaaattttg	tcatagaggg	1320
taataaactt	ctaaaactaa	atcattagtt	ccgcgtggat	cgaaggcctt	agattgtaga	1380
gagcttttag	ttaaaaatac	tgacttacct	tttatagggtg	atattagtga	tatcaaaaact	1440
gatataTTTT	taagcaaaga	tattaatgaa	gaaactgaag	ttatagacta	tccggacaat	1500
gtttcagtg	atcaagttat	tctcagtaag	aatacctcag	aacatggaca	actagatttta	1560
ttatacccta	ttattgaagg	tgagagtcaa	gtattaccgg	gagagaatca	agtcttttat	1620
gataatagaa	ctcaaaatgt	tgattatttg	aattcttatt	attacctaga	atctcaaaaa	1680
ctaagtgata	atgttgaaaga	ttttactttt	acgacatcaa	ttgaggaagc	tttgagataat	1740
agtggaaaag	tataacttta	ctttcctaaa	ctagctgata	aagtaaatac	gggtgttcaa	1800
ggtgggttat	ttttaatgtg	ggcaaatgat	gtagttgaag	attttactac	aaatattcta	1860
agaaaagata	cattagataa	aatatcagat	gtatcagcta	ttattcccta	tataggacct	1920
gcattaaata	taagtaattc	tgtaagaagg	ggaaatttta	ctgaagcatt	tgagtttacc	1980
ggtgtaacta	ttttattaga	agcgtttcaa	gaattttcaa	tacctgcact	tggtgcattt	2040
gtgatttata	gtaaggttca	agaaagaaac	gagattatta	aaactataga	taattgttta	2100
gaacaaagga	ttaaaagatg	gaaagattca	tatgaatgga	tgataggaac	gtggttatcc	2160
aggattacta	ctcaatttta	taatataagt	tatcaaatgt	atgattcttt	aaattatcag	2220
gcagatgcaa	tcaaagataa	aatagattta	gaatataaaa	aatactcagg	aagtgataaa	2280
gaaaatataa	aaagtcaagt	tgaaaattta	aaaaatagtt	tagatataaa	aatctcgga	2340
gcaatgaata	atataaataa	atztatacga	gaatgttctg	taacatactt	atttaaaaat	2400
atgctcccta	aagtaattga	tgaattaaat	aagtttgatt	taaaaactaa	aacagaatta	2460
attaatctta	tagatagtca	taatattatt	ctagttgggtg	aagtagatag	attaaaagca	2520
aaaataaatg	agagttttga	aaatacaata	ccctttaata	ttttttcata	tactaataat	2580
tctttattaa	aagatatact	cgagcaccac	caccaccacc	actaa		2625

<210> 14

<211> 874

5 <212> PRT

<213> *Clostridium botulinum*

<400> 14

Met	Ala	Ile	Thr	Ile	Asn	Asn	Phe	Asn	Tyr	Ser	Asp	Pro	Val	Asp	Asn
1				5					10					15	

Lys	Asn	Ile	Leu	Tyr	Leu	Asp	Thr	His	Leu	Asn	Thr	Leu	Ala	Asn	Glu
			20					25					30		

Pro	Glu	Lys	Ala	Phe	Arg	Ile	Ile	Gly	Asn	Ile	Trp	Val	Ile	Pro	Asp
		35					40					45			

ES 2 380 677 T3

Arg Phe Ser Arg Asp Ser Asn Pro Asn Leu Asn Lys Pro Pro Arg Val
 50 55 60
 Thr Ser Pro Lys Ser Gly Tyr Tyr Asp Pro Asn Tyr Leu Ser Thr Asp
 65 70 75 80
 Ser Glu Lys Asp Thr Phe Leu Lys Glu Ile Ile Lys Leu Phe Lys Arg
 85 90 95
 Ile Asn Ser Arg Glu Ile Gly Glu Glu Leu Ile Tyr Arg Leu Ala Thr
 100 105 110
 Asp Ile Pro Phe Pro Gly Asn Asn Asn Thr Pro Ile Asn Thr Phe Asp
 115 120 125
 Phe Asp Val Asp Phe Asn Ser Val Asp Val Lys Thr Arg Gln Gly Asn
 130 135 140
 Asn Trp Val Lys Thr Gly Ser Ile Asn Pro Ser Val Ile Ile Thr Gly
 145 150 155 160
 Pro Arg Glu Asn Ile Ile Asp Pro Glu Thr Ser Thr Phe Lys Leu Thr
 165 170 175
 Asn Asn Thr Phe Ala Ala Gln Glu Gly Phe Gly Ala Leu Ser Ile Ile
 180 185 190
 Ser Ile Ser Pro Arg Phe Met Leu Thr Tyr Ser Asn Ala Thr Asn Asn
 195 200 205
 Val Gly Glu Gly Arg Phe Ser Lys Ser Glu Phe Cys Met Asp Pro Ile
 210 215 220
 Leu Ile Leu Met His Glu Leu Asn His Ala Met His Asn Leu Tyr Gly
 225 230 235 240
 Ile Ala Ile Pro Asn Asp Gln Arg Ile Ser Ser Val Thr Ser Asn Ile
 245 250 255
 Phe Tyr Ser Gln Tyr Asn Val Lys Leu Glu Tyr Ala Glu Ile Tyr Ala
 260 265 270
 Phe Gly Gly Pro Thr Ile Asp Leu Ile Pro Lys Ser Ala Arg Lys Tyr
 275 280 285
 Phe Glu Glu Lys Ala Leu Asp Tyr Tyr Arg Ser Ile Ala Lys Arg Leu
 290 295 300
 Asn Ser Ile Thr Thr Ala Asn Pro Ser Ser Phe Asn Lys Tyr Ile Gly
 305 310 315 320
 Glu Tyr Lys Gln Lys Leu Ile Arg Lys Tyr Arg Phe Val Val Glu Ser
 325 330 335
 Ser Gly Glu Val Ala Val Asp Arg Asn Lys Phe Ala Glu Leu Tyr Lys
 340 345 350
 Glu Leu Thr Gln Ile Phe Thr Glu Phe Asn Tyr Ala Lys Ile Tyr Asn
 355 360 365
 Val Gln Asn Arg Lys Ile Tyr Leu Ser Asn Val Tyr Thr Pro Val Thr
 370 375 380

ES 2 380 677 T3

Ala	Asn	Ile	Leu	Asp	Asp	Asn	Val	Tyr	Asp	Ile	Gln	Asn	Gly	Phe	Asn	
385					390					395					400	
Ile	Pro	Lys	Ser	Asn	Leu	Asn	Val	Leu	Phe	Met	Gly	Gln	Asn	Leu	Ser	
				405					410					415		
Arg	Asn	Pro	Ala	Leu	Arg	Lys	Val	Asn	Pro	Glu	Asn	Met	Leu	Tyr	Leu	
			420					425					430			
Phe	Thr	Lys	Phe	Cys	His	Arg	Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser	
		435					440					445				
Leu	Val	Pro	Arg	Gly	Ser	Lys	Ala	Leu	Asp	Cys	Arg	Glu	Leu	Leu	Val	
	450					455					460					
Lys	Asn	Thr	Asp	Leu	Pro	Phe	Ile	Gly	Asp	Ile	Ser	Asp	Ile	Lys	Thr	
465					470					475					480	
Asp	Ile	Phe	Leu	Ser	Lys	Asp	Ile	Asn	Glu	Glu	Thr	Glu	Val	Ile	Asp	
				485					490					495		
Tyr	Pro	Asp	Asn	Val	Ser	Val	Asp	Gln	Val	Ile	Leu	Ser	Lys	Asn	Thr	
			500					505					510			
Ser	Glu	His	Gly	Gln	Leu	Asp	Leu	Leu	Tyr	Pro	Ile	Ile	Glu	Gly	Glu	
		515					520						525			
Ser	Gln	Val	Leu	Pro	Gly	Glu	Asn	Gln	Val	Phe	Tyr	Asp	Asn	Arg	Thr	
	530					535					540					
Gln	Asn	Val	Asp	Tyr	Leu	Asn	Ser	Tyr	Tyr	Tyr	Leu	Glu	Ser	Gln	Lys	
545					550					555					560	
Leu	Ser	Asp	Asn	Val	Glu	Asp	Phe	Thr	Phe	Thr	Thr	Ser	Ile	Glu	Glu	
			565						570					575		
Ala	Leu	Asp	Asn	Ser	Gly	Lys	Val	Tyr	Thr	Tyr	Phe	Pro	Lys	Leu	Ala	
			580					585					590			
Asp	Lys	Val	Asn	Thr	Gly	Val	Gln	Gly	Gly	Leu	Phe	Leu	Met	Trp	Ala	
		595					600					605				
Asn	Asp	Val	Val	Glu	Asp	Phe	Thr	Thr	Asn	Ile	Leu	Arg	Lys	Asp	Thr	
	610					615					620					
Leu	Asp	Lys	Ile	Ser	Asp	Val	Ser	Ala	Ile	Ile	Pro	Tyr	Ile	Gly	Pro	
625					630					635					640	
Ala	Leu	Asn	Ile	Ser	Asn	Ser	Val	Arg	Arg	Gly	Asn	Phe	Thr	Glu	Ala	
			645						650					655		
Phe	Ala	Val	Thr	Gly	Val	Thr	Ile	Leu	Leu	Glu	Ala	Phe	Gln	Glu	Phe	
			660					665					670			
Thr	Ile	Pro	Ala	Leu	Gly	Ala	Phe	Val	Ile	Tyr	Ser	Lys	Val	Gln	Glu	
		675				680						685				
Arg	Asn	Glu	Ile	Ile	Lys	Thr	Ile	Asp	Asn	Cys	Leu	Glu	Gln	Arg	Ile	
	690					695					700					
Lys	Arg	Trp	Lys	Asp	Ser	Tyr	Glu	Trp	Met	Ile	Gly	Thr	Trp	Leu	Ser	
705					710					715				720		

ES 2 380 677 T3

Arg	Ile	Thr	Thr	Gln	Phe	Asn	Asn	Ile	Ser	Tyr	Gln	Met	Tyr	Asp	Ser
				725					730					735	
Leu	Asn	Tyr	Gln	Ala	Asp	Ala	Ile	Lys	Asp	Lys	Ile	Asp	Leu	Glu	Tyr
			740					745					750		
Lys	Lys	Tyr	Ser	Gly	Ser	Asp	Lys	Glu	Asn	Ile	Lys	Ser	Gln	Val	Glu
		755					760					765			
Asn	Leu	Lys	Asn	Ser	Leu	Asp	Ile	Lys	Ile	Ser	Glu	Ala	Met	Asn	Asn
	770					775					780				
Ile	Asn	Lys	Phe	Ile	Arg	Glu	Cys	Ser	Val	Thr	Tyr	Leu	Phe	Lys	Asn
785					790					795					800
Met	Leu	Pro	Lys	Val	Ile	Asp	Glu	Leu	Asn	Lys	Phe	Asp	Leu	Lys	Thr
				805					810					815	
Lys	Thr	Glu	Leu	Ile	Asn	Leu	Ile	Asp	Ser	His	Asn	Ile	Ile	Leu	Val
			820					825					830		
Gly	Glu	Val	Asp	Arg	Leu	Lys	Ala	Lys	Ile	Asn	Glu	Ser	Phe	Glu	Asn
		835					840					845			
Thr	Ile	Pro	Phe	Asn	Ile	Phe	Ser	Tyr	Thr	Asn	Asn	Ser	Leu	Leu	Lys
	850					855					860				
Asp	Ile	Leu	Glu	His	His	His	His	His	His						
865					870										

<210> 15

<211> 3810

<212> ADN

5 <213> *Clostridium botulinum*

<400> 15

ES 2 380 677 T3

atggcaacaa	ttaatagttt	taattataat	gatcctgtta	ataatagaac	aattttatat	60
attaaccag	gcggttgta	acaattttat	aatcattta	atattatgaa	aaatatttgg	120
ataattccag	agagaaatgt	aattggtaca	attccccaag	attttcttcc	gcctacttca	180
ttgaaaaatg	gagatagtag	ttattatgac	cctaattatt	tacaaagtga	tcaagaaaag	240
gataaatttt	taaaaatagt	cacaaaaata	tttaatagaa	taaatgataa	tctttcagga	300
aggattttat	tagaagaact	gtcaaaagct	aatccatatt	taggaaatga	taatactcca	360
gatggtgact	tcattattaa	tgatgcatca	gcagttccaa	ttcaattctc	aaatggtagc	420
caaagcatac	tattacctaa	tgttattata	atggggagcag	agcctgattt	atttgaaact	480
aacagttcca	atatttctct	aagaaataat	tatatgccaa	gcaatcacgg	ttttggatca	540
atagctatag	taacattctc	acctgaatat	tcttttagat	ttaaagataa	tagtatgaat	600
gaattttattc	aagatcctgc	tcttacatta	atgcatgaat	taatacattc	attacatgga	660
ctatatgggg	ctaaagggat	tactacaaag	tatactataa	cacaaaaaca	aaatccccta	720
ataacaaata	taagaggtag	aaatattgaa	gaattcttaa	cttttgagg	tactgattta	780
aacattatta	ctagtgtc	gtccaatgat	atctatacta	atcttctagc	tgattataaa	840
aaaatagcgt	ctaaacttag	caaagtacaa	gtatctaata	cactacttaa	tccttataaa	900
gatgtttttg	aagcaaagta	tggattagat	aaagatgcta	gcggaattta	ttcggtaaat	960
ataaacaat	ttaatgat	ttttaaaaa	ttatacagct	ttacggaatt	tgatttagca	1020
actaaatttc	aagttaaag	taggcaaaact	tatattggac	agtataaata	cttcaaactt	1080
tcaaacttgt	taaatgattc	tatttataat	atatcagaag	gctataatat	aaataattta	1140
aaggtaaatt	ttagaggaca	gaatgcaa	ttaaatccta	gaattattac	accaattaca	1200
ggtagaggac	tagtaaaaa	aatcattaga	ttttgtaaaa	gtgggataat	aacttctaaa	1260
actaaatcat	tggttccacg	tggatccaag	gcctcaatat	gtatcgaaat	aaataatggt	1320
gagttatttt	ttgtggcttc	cgagaatagt	tataatgatg	ataatataaa	tactcctaaa	1380
gaaattgacg	atcacgtaac	ttcaaataat	aattatgaaa	atgatttaga	tcagggttatt	1440
ttaaattttta	atagtgaatc	agcacctgga	ctttcagatg	aaaaattaaa	tttaactatc	1500

caaaatgatg	cttatatacc	aaaatatgat	tctaattggaa	caagtgatat	agaacaacat	1560
gatgttaatg	aacttaatgt	attttttctat	ttagatgcac	agaaagtgcc	cgaaggtgaa	1620
aataatgtca	atctcacctc	ttcaattgat	acagcattat	tagaacaacc	taaaatatat	1680
acattttttt	catcagaatt	tattaataat	gtcaataaac	ctgtgcaagc	agcattat	1740
gtaggctgga	tacaacaagt	attagtagat	tttactactg	aagctaacca	aaaaagtact	1800
gttgataaaa	ttgcagatat	ttctatagtt	gttccatata	taggtccttg	tttaaatata	1860
ggaaatgaag	cacaaaaagg	aaatttttaa	gatgcacttg	aattattagg	agcaggtatt	1920
ttattagaat	ttgaaccoga	gcttttaatt	cctacaattt	tagtattcac	gataaaatct	1980
tttttaggtt	catctgataa	taaaaataaa	gttatttaaag	caataaataa	tgcatgaaa	2040
gaaagagatg	aaaaatggaa	agaagtatat	agtttttatag	tatcgaattg	gatgactaaa	2100
attaatacac	aatttaataa	aagaaaagaa	caaattgtatc	aagctttaca	aatcaagta	2160
aatgcactta	aagcaataat	agaatctaag	tataatagtt	atactttaga	agaaaaaat	2220
gagcttacaa	ataaatatga	tattgagcaa	atagaaaatg	aacttaatca	aaaggtttct	2280
atagcaatga	ataatataga	caggttctta	actgaaagtt	ctatatctta	tttaatgaaa	2340
ttaataaatg	aagtaaaaaat	taataaatta	agagaatatg	atgaaaatgt	taaaacgtat	2400
ttattagatt	atattataaa	acatggatca	atcttgggag	agagtcagca	agaactaaat	2460
tctatggtaa	ttgataccct	aaataatagt	attcctttta	agctttcttc	ttatacagat	2520
gataaaat	taatttcata	ttttaataag	ttctttaaga	gaattaaaag	tagttctggt	2580
ttaaatatga	gatataaaaa	tgataaatac	gtagataactt	caggatatga	ttcaaatata	2640
aatattaatg	gagatgtata	taaatatcca	actaataaaa	atcaatttgg	aatatataat	2700
gataaactta	gtgaagttaa	tatatctcaa	aatgattaca	ttatatatga	taataaatat	2760
aaaaatttta	gtattagttt	ttgggtaaga	attcctaact	atgataataa	gatagtaaat	2820
gttaataatg	aatacactat	aataaattgt	atgagggata	ataattcagg	atggaaagta	2880
tctcttaatc	ataatgaaat	aatttggaca	ttgcaagata	attcaggaat	taatcaaaaa	2940
ttagcat	actatggtaa	cgcaaattgg	atctctgatt	atataaataa	gtggattttt	3000
gtaactataa	ctaattgata	attaggagat	tctaaacttt	atattaatgg	aaatttaata	3060
gataaaaaat	caatttttaa	tttaggtaat	attcatgtta	gtgacaatat	attattttaa	3120
atagttaatt	gtagttatac	aagatatatt	ggtatttagat	attttaatat	ttttgataaa	3180
gaattagatg	aaacagaaat	tcaaacttta	tataacaatg	agcctaattg	aaatatttta	3240
aaggattttt	ggggaaatta	tttgctttat	gacaaagaat	attattttatt	aaatgtgtta	3300
aaaccaaata	actttattaa	taggagaaca	gattctactt	taagcattaa	taatataaga	3360
agcactattc	ttttagctaa	tagattatat	agtggataaa	aagttaaaat	acaaagagtt	3420
aataatagta	gtactaacga	taatcttggt	agaaagaatg	atcaggtata	tattaatttt	3480
gtagccagca	aaactcactt	acttccatta	tatgctgata	cagctaccac	aaataaagag	3540
aaaacaataa	aaatatcatc	atctggcaat	agattttaatc	aagtagtagt	tatgaattca	3600
gtaggaaata	attgtacaat	gaatttttaa	aataataatg	gaaataatat	tggttggtta	3660
ggtttcaagg	cagatactgt	agttgctagt	acttggtatt	atacacatat	gagagataat	3720
acaaacagca	atggattttt	ttggaacttt	atctctgaag	aacatggatg	gcaagaaaaa	3780
acgcgttctc	accaccacca	ccaccactaa				3810

<210> 16

<211> 1269

<212> PRT

5 <213> *Clostridium botulinum*

<400> 16

ES 2 380 677 T3

Met	Ala	Thr	Ile	Asn	Ser	Phe	Asn	Tyr	Asn	Asp	Pro	Val	Asn	Asn	Arg
1				5					10					15	
Thr	Ile	Leu	Tyr	Ile	Lys	Pro	Gly	Gly	Cys	Gln	Gln	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	Ile	Pro	Gln	Asp	Phe	Leu	Pro	Pro	Thr	Ser	Leu	Lys	Asn	Gly
	50					55					60				
Asp	Ser	Ser	Tyr	Tyr	Asp	Pro	Asn	Tyr	Leu	Gln	Ser	Asp	Gln	Glu	Lys
65					70					75					80

ES 2 380 677 T3

Asp Lys Phe Leu Lys Ile Val Thr Lys Ile Phe Asn Arg Ile Asn Asp
 85 90 95
 Asn Leu Ser Gly Arg Ile Leu Leu Glu Glu Leu Ser Lys Ala Asn Pro
 100 105 110
 Tyr Leu Gly Asn Asp Asn Thr Pro Asp Gly Asp Phe Ile Ile Asn Asp
 115 120 125
 Ala Ser Ala Val Pro Ile Gln Phe Ser Asn Gly Ser Gln Ser 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 Lys 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
 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 Ser Gly Ile
 405 410 415

ES 2 380 677 T3

Ile Thr Ser Lys Thr Lys Ser Leu Val Pro Arg Gly Ser Lys Ala Ser
 420 425 430
 Ile Cys Ile Glu Ile Asn Asn Gly Glu Leu Phe Phe Val Ala Ser Glu
 435 440 445
 Asn Ser Tyr Asn Asp Asp Asn Ile Asn Thr Pro Lys Glu Ile Asp Asp
 450 455 460
 Thr Val Thr Ser Asn Asn Asn Tyr Glu Asn Asp Leu Asp Gln Val Ile
 465 470 475 480
 Leu Asn Phe Asn Ser Glu Ser Ala Pro Gly Leu Ser Asp Glu Lys Leu
 485 490 495
 Asn Leu Thr Ile Gln Asn Asp Ala Tyr Ile Pro Lys Tyr Asp Ser Asn
 500 505 510
 Gly Thr Ser Asp Ile Glu Gln His Asp Val Asn Glu Leu Asn Val Phe
 515 520 525
 Phe Tyr Leu Asp Ala Gln Lys Val Pro Glu Gly Glu Asn Asn Val Asn
 530 535 540
 Leu Thr Ser Ser Ile Asp Thr Ala Leu Leu Glu Gln Pro Lys Ile Tyr
 545 550 555 560
 Thr Phe Phe Ser Ser Glu Phe Ile Asn Asn Val Asn Lys Pro Val Gln
 565 570 575
 Ala Ala Leu Phe Val Gly Trp Ile Gln Gln Val Leu Val Asp Phe Thr
 580 585 590
 Thr Glu Ala Asn Gln Lys Ser Thr Val Asp Lys Ile Ala Asp Ile Ser
 595 600 605
 Ile Val Val Pro Tyr Ile Gly Leu Ala Leu Asn Ile Gly Asn Glu Ala
 610 615 620
 Gln Lys Gly Asn Phe Lys Asp Ala Leu Glu Leu Leu Gly Ala Gly Ile
 625 630 635 640
 Leu Leu Glu Phe Glu Pro Glu Leu Leu Ile Pro Thr Ile Leu Val Phe
 645 650 655
 Thr Ile Lys Ser Phe Leu Gly Ser Ser Asp Asn Lys Asn Lys Val Ile
 660 665 670
 Lys Ala Ile Asn Asn Ala Leu Lys Glu Arg Asp Glu Lys Trp Lys Glu
 675 680 685
 Val Tyr Ser Phe Ile Val Ser Asn Trp Met Thr Lys Ile Asn Thr Gln
 690 695 700
 Phe Asn Lys Arg Lys Glu Gln Met Tyr Gln Ala Leu Gln Asn Gln Val
 705 710 715 720
 Asn Ala Leu Lys Ala Ile Ile Glu Ser Lys Tyr Asn Ser Tyr Thr Leu
 725 730 735
 Glu Glu Lys Asn Glu Leu Thr Asn Lys Tyr Asp Ile Glu Gln Ile Glu
 740 745 750

ES 2 380 677 T3

Asn Glu Leu Asn Gln Lys Val Ser Ile Ala Met Asn Asn Ile Asp Arg
 755 760 765
 Phe Leu Thr Glu Ser Ser Ile Ser Tyr Leu Met Lys Leu Ile Asn Glu
 770 775 780
 Val Lys Ile Asn Lys Leu Arg Glu Tyr Asp Glu Asn Val Lys Thr Tyr
 785 790 795 800
 Leu Leu Asp Tyr Ile Ile Lys His Gly Ser Ile Leu Gly Glu Ser Gln
 805 810 815
 Gln Glu Leu Asn Ser Met Val Ile Asp Thr Leu Asn Asn Ser Ile Pro
 820 825 830
 Phe Lys Leu Ser Ser Tyr Thr Asp Asp Lys Ile Leu Ile Ser Tyr Phe
 835 840 845
 Asn Lys Phe Phe Lys Arg Ile Lys Ser Ser Ser Val Leu Asn Met Arg
 850 855 860
 Tyr Lys Asn Asp Lys Tyr Val Asp Thr Ser Gly Tyr Asp Ser Asn Ile
 865 870 875 880
 Asn Ile Asn Gly Asp Val Tyr Lys Tyr Pro Thr Asn Lys Asn Gln Phe
 885 890 895
 Gly Ile Tyr Asn Asp Lys Leu Ser Glu Val Asn Ile Ser Gln Asn Asp
 900 905 910
 Tyr Ile Ile Tyr Asp Asn Lys Tyr Lys Asn Phe Ser Ile Ser Phe Trp
 915 920 925
 Val Arg Ile Pro Asn Tyr Asp Asn Lys Ile Val Asn Val Asn Asn Glu
 930 935 940
 Tyr Thr Ile Ile Asn Cys Met Arg Asp Asn Asn Ser Gly Trp Lys Val
 945 950 955 960
 Ser Leu Asn His Asn Glu Ile Ile Trp Thr Leu Gln Asp Asn Ser Gly
 965 970 975
 Ile Asn Gln Lys Leu Ala Phe Asn Tyr Gly Asn Ala Asn Gly Ile Ser
 980 985 990
 Asp Tyr Ile Asn Lys Trp Ile Phe Val Thr Ile Thr Asn Asp Arg Leu
 995 1000 1005
 Gly Asp Ser Lys Leu Tyr Ile Asn Gly Asn Leu Ile Asp Lys Lys
 1010 1015 1020
 Ser Ile Leu Asn Leu Gly Asn Ile His Val Ser Asp Asn Ile Leu
 1025 1030 1035
 Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg Tyr Ile Gly Ile Arg
 1040 1045 1050
 Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu Thr Glu Ile Gln
 1055 1060 1065
 Thr Leu Tyr Asn Asn Glu Pro Asn Ala Asn Ile Leu Lys Asp Phe
 1070 1075 1080

ES 2 380 677 T3

Trp	Gly	Asn	Tyr	Leu	Leu	Tyr	Asp	Lys	Glu	Tyr	Tyr	Leu	Leu	Asn
	1085					1090					1095			
Val	Leu	Lys	Pro	Asn	Asn	Phe	Ile	Asn	Arg	Arg	Thr	Asp	Ser	Thr
	1100					1105					1110			
Leu	Ser	Ile	Asn	Asn	Ile	Arg	Ser	Thr	Ile	Leu	Leu	Ala	Asn	Arg
	1115					1120					1125			
Leu	Tyr	Ser	Gly	Ile	Lys	Val	Lys	Ile	Gln	Arg	Val	Asn	Asn	Ser
	1130					1135					1140			
Ser	Thr	Asn	Asp	Asn	Leu	Val	Arg	Lys	Asn	Asp	Gln	Val	Tyr	Ile
	1145					1150					1155			
Asn	Phe	Val	Ala	Ser	Lys	Thr	His	Leu	Leu	Pro	Leu	Tyr	Ala	Asp
	1160					1165					1170			
Thr	Ala	Thr	Thr	Asn	Lys	Glu	Lys	Thr	Ile	Lys	Ile	Ser	Ser	Ser
	1175					1180					1185			
Gly	Asn	Arg	Phe	Asn	Gln	Val	Val	Val	Met	Asn	Ser	Val	Gly	Asn
	1190					1195					1200			
Asn	Cys	Thr	Met	Asn	Phe	Lys	Asn	Asn	Asn	Gly	Asn	Asn	Ile	Gly
	1205					1210					1215			
Leu	Leu	Gly	Phe	Lys	Ala	Asp	Thr	Val	Val	Ala	Ser	Thr	Trp	Tyr
	1220					1225					1230			
Tyr	Thr	His	Met	Arg	Asp	Asn	Thr	Asn	Ser	Asn	Gly	Phe	Phe	Trp
	1235					1240					1245			
Asn	Phe	Ile	Ser	Glu	Glu	His	Gly	Trp	Gln	Glu	Lys	Thr	Arg	Ser
	1250					1255					1260			
His	His	His	His	His	His									
	1265													

<210> 17

<211> 1116

<212> ADN

5 <213> *Pseudomonas aeruginosa*

<400> 17

ES 2 380 677 T3

atgggagcc	tggccgcgct	gaccgcgcac	caggcttgcc	acagagggat	aataacttct	60
aaaactaaat	cattagttcc	gcgtggatcc	aaggccctgg	agcagtgcgg	gtacccggtg	120
cagcggctgg	tcgccctcta	cctggcgggc	cggctgtcgt	ggaaccaggt	cgaccaggtg	180
atccgcaacg	ccctggccag	ccccggcage	ggcgggcgaac	tgggcgaagc	gatccgcgag	240
cagccggagc	aggcccgctct	ggccctgacc	ctggccgcgc	ccgagagcga	gcgcttcgtc	300
cggcagggca	ccggcaacga	cgaggccggc	gcgcccaacg	ccgacgtggt	gagcctgacc	360
tgcccggtcg	ccgccggtga	atgcgcgggc	ccggcggaca	gcggcgacgc	cctgctggag	420
cgcaactatc	ccactggcgc	ggagttcctc	ggcgacggcg	gcgacgtcag	cttcagcacc	480
cgcggcacgc	agaactggac	ggtggagcgg	ctgctccagg	cgcaccgcca	actggaggag	540
cgcggctatg	tgttcgtcgg	ctaccacggc	accttcctcg	aagcggcgca	aagcatcgtc	600
ttcggcgggg	tgcgcgcgcg	cagccaggac	ctcgacgcga	tctggcgcg	tttctatatc	660
gccggcgatc	cggcgctggc	ctacggctac	gcccaggacc	aggaaccgga	cgcacgcggc	720
cggatccgca	acggtgccct	gctgcgggtc	tatgtgccgc	gctcgagcct	gccgggcttc	780
taccgcacca	gcctgaccct	ggccgcgcgc	gaggcgggcg	gcgaggtcga	acggctgac	840
ggccatccgc	tgccgctgcg	cctggacgcc	atcaccggcc	ccgaggagga	aggcgggcgc	900
ctggagacca	ttctcggtcg	gccgctggcc	gagcgcaccg	tggtgattcc	ctcggcgatc	960

cccaccgacc	cgcgcaacgt	ggcgggcgac	ctcgaccctg	ccagcatccc	cgacaaggaa	1020
caggcgatca	gcgccctgcc	ggactacgcc	agccagcccg	gcaaaccgcc	gcgcgaggac	1080
ctgacgcgtc	tcgagcacca	ccaccaccac	cactaa			1116

<210> 18

<211> 371

<212> PRT

5 <213> *Pseudomonas aeruginosa*

<400> 18

ES 2 380 677 T3

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

ES 2 380 677 T3

Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu
 275 280 285

Asp Ala Ile Thr Gly Pro Glu Glu Gly Gly Arg Leu Glu Thr Ile
 290 295 300

Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile
 305 310 315 320

Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile
 325 330 335

Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln
 340 345 350

Pro Gly Lys Pro Pro Arg Glu Asp Leu Thr Arg Leu Glu His His His
 355 360 365

His His His
 370

<210> 19

<211> 1215

<212> ADN

5 <213> *Corynebacterium diphtheriae*

<400> 19

atggctgatg	atgttggtga	ttcttctaaa	tcttttgtga	tggaaaactt	ttcttcgtac	60
cacgggacta	aacctgggta	tgtagattcc	attcaaaaag	gtatacaaaa	gccaaaatct	120
ggtacacaag	gaaattatga	cgatgattgg	aaagggtttt	atagtaccga	caataaatac	180
gacgctgcg	gatactctgt	agataacgaa	aaccgcctct	ctggaaaagc	tggaggcgtg	240
gtcaaagtga	cgtatccagg	actgacgaag	gtcctcgcac	taaaagtggg	taatgccgaa	300
actattaaga	aagagttagg	tttaagtctc	actgaaccgt	tgatggagca	agtcggaaca	360
gaagagttta	tcaaaagggt	cggggatggt	gcttcgcgtg	tagtactcag	ccttcccttt	420
gccgagggga	gttctagcgt	cgaatatatt	aataactggg	aacaggcgaa	agcgtaagc	480
gtagaacttg	agattaattt	tgaaccctgt	ggaaaacgtg	gccaagatgc	gatgtatgag	540
tatatggctc	aagcctgtgc	aagaggggata	ataacttcta	aaactaaatc	attggttcca	600
cgtggatcca	aggccttgct	atgcataaat	cttgattggg	atgtcataag	ggataaaaact	660
aagacaaaga	tagagtcttt	gaaagagcat	ggccctatca	aaaataaaat	gagcgaaagt	720
cccaataaaa	cagtatctga	ggaaaaagct	aaacaatacc	tagaagaatt	tcatcaaacc	780
gcattagagc	atcctgaatt	gtcagaactt	aaaaccgtta	ctgggaccaa	tcctcttttc	840
gctggggcta	actatgcggc	gtgggcagta	aacgttgccg	aagttatcga	tagcgaaaca	900
gctgataatt	tggaaaagac	aactgctgct	ctttcgatac	ttcctgggat	cggtagcgta	960
atgggcattg	cagacggtgc	cgttcaccac	aatacagaag	agatagtggc	acaatcaata	1020
gctttatcgt	ctttaatggg	tgctcaagct	attccattgg	taggagagct	agttgatatt	1080
ggtttcgctg	catataattt	tgtagagagt	attatcaatt	tgtttcaagt	agttcataat	1140
tcgtataatc	gtcccgcgta	ttctccgggg	cataaaacgc	aaccatttct	cgagcaccac	1200
caccaccacc	actaa					1215

<210> 20

<211> 404

10 <212> PRT

<213> *Corynebacterium diphtheriae*

<400> 20

ES 2 380 677 T3

Met Ala Asp Asp Val Val Asp Ser Ser Lys Ser Phe Val Met Glu Asn
1 5 10 15
Phe Ser Ser Tyr His Gly Thr Lys Pro Gly Tyr Val Asp Ser Ile Gln

ES 2 380 677 T3

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

ES 2 380 677 T3

355				360				365							
Glu	Ser	Ile	Ile	Asn	Leu	Phe	Gln	Val	Val	His	Asn	Ser	Tyr	Asn	Arg
370				375				380							
Pro	Ala	Tyr	Ser	Pro	Gly	His	Lys	Thr	Gln	Pro	Phe	Leu	Glu	His	His
385				390				395				400			
His	His	His	His												

<210> 21
<211> 1650
<212> ADN
<213> *Ricinus communis*

<400> 21

atggtcccca	aacaataccc	aattataaac	tttaccacag	cgggtgccac	tgtgcaaagc	60
tacacaaact	ttatcagagc	tgttcgcggt	cgtttaacaa	ctggagctga	tgtgagacat	120
gatataccag	tgttgccaaa	cagagttggt	ttgcctataa	accaacggtt	tatttttagtt	180
gaactctcaa	atcatgcaga	gctttctgtt	acattagccc	tggatgtcac	caatgcatat	240
gtggtcggct	accgtgctgg	aaatagcgca	tatttctttc	atcctgacaa	tcaggaagat	300
gcagaagcaa	tactcatct	tttactgat	gttcaaaatc	gatatacatt	cgcctttggt	360
ggtaattatg	atagacttga	acaacttgct	ggtaatctga	gagaaaatat	cgagttggga	420
aatggtccac	tagaggaggc	tatctcagcg	ctttattatt	acagtactgg	tggcactcag	480
cttccaactc	tggctcgttc	ctttataaatt	tgcattccaa	tgatttcaga	agcagcaaga	540
ttccaatata	ttgagggaga	aatgcgcacg	agaattaggt	acaaccggag	atctgcacca	600
gatcctagcg	taattacact	tgagaatagt	tggggggagac	tttccactgc	aattcaagag	660
tctaaccaag	gagcctttgc	tagtccaatt	caactgcaa	gacgtaatgg	ttccaaattc	720
agtgtgtacg	atgtgagtat	attaatccct	atcatagctc	tcattggtgta	tagatgcgca	780
cctcctaggg	ggataataac	ttctaaaact	aaatcattgg	ttccacgtgg	atccaaggct	840
gatgttttga	tggatcctga	gcccatagtg	cgtatcgtag	gtcgaaatgg	tctatgtgtt	900
gatgttaggg	atggaagatt	ccacaacgga	aacgcaatac	agtttgtggc	atgcaagtct	960
aatacagatg	caaatacagt	ctggactttg	aaaagagaca	atactattcg	atctaattgga	1020
aagtgtttaa	ctacttaagg	gtacagtcgg	ggagtctatg	tgatgatcta	tgattgcaat	1080
actgctgcaa	ctgatgccac	ccgctggcaa	atatgggata	atggaaccat	cataaatccc	1140
agatctagtc	tagtttttagc	agcgacatca	gggaacagtg	gtaccacact	tacggtgcaa	1200
accaacattt	atgcgcgttag	tcaaggttgg	cttcctacta	ataatacaca	accttttggt	1260
acaaccattg	ttgggctata	tggctctgtc	ttgcaagcaa	atagtggaca	agtatggata	1320
gaggactgta	gcagtgaaaa	ggctgaacaa	cagtgggctc	tttatgcaga	tggttcaata	1380
cgctctcagc	aaaaccgaga	taattgcctt	acaagtgatt	ctaatacacg	ggaacagtt	1440
gttaagatcc	tctcttgtgg	ccctgcatcc	tctggccaac	gatggatgtt	caagaatgat	1500
ggaaccattt	taaattttgta	tagtggattg	gtgttagatg	tgaggcgatc	ggatccgagc	1560
cttaaacaaa	tcattcttta	ccctctccat	ggtgacccaa	accaaatatg	gttaccatta	1620
tttctcgagc	accaccacca	ccaccactaa				1650

<210> 22
<211> 549
<212> PRT
<213> *Ricinus communis*

<400> 22

ES 2 380 677 T3

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

ES 2 380 677 T3

35					40					45					
Val	Gly	Leu	Pro	Ile	Asn	Gln	Arg	Phe	Ile	Leu	Val	Glu	Leu	Ser	Asn
50						55					60				
His	Ala	Glu	Leu	Ser	Val	Thr	Leu	Ala	Leu	Asp	Val	Thr	Asn	Ala	Tyr
65					70					75					80
Val	Val	Gly	Tyr	Arg	Ala	Gly	Asn	Ser	Ala	Tyr	Phe	Phe	His	Pro	Asp
				85					90					95	
Asn	Gln	Glu	Asp	Ala	Glu	Ala	Ile	Thr	His	Leu	Phe	Thr	Asp	Val	Gln
			100					105					110		
Asn	Arg	Tyr	Thr	Phe	Ala	Phe	Gly	Gly	Asn	Tyr	Asp	Arg	Leu	Glu	Gln
		115					120					125			
Leu	Ala	Gly	Asn	Leu	Arg	Glu	Asn	Ile	Glu	Leu	Gly	Asn	Gly	Pro	Leu
	130					135					140				
Glu	Glu	Ala	Ile	Ser	Ala	Leu	Tyr	Tyr	Tyr	Ser	Thr	Gly	Gly	Thr	Gln
145					150					155					160
Leu	Pro	Thr	Leu	Ala	Arg	Ser	Phe	Ile	Ile	Cys	Ile	Gln	Met	Ile	Ser
				165					170					175	
Glu	Ala	Ala	Arg	Phe	Gln	Tyr	Ile	Glu	Gly	Glu	Met	Arg	Thr	Arg	Ile
			180					185					190		
Arg	Tyr	Asn	Arg	Arg	Ser	Ala	Pro	Asp	Pro	Ser	Val	Ile	Thr	Leu	Glu
		195					200					205			
Asn	Ser	Trp	Gly	Arg	Leu	Ser	Thr	Ala	Ile	Gln	Glu	Ser	Asn	Gln	Gly
		210				215					220				
Ala	Phe	Ala	Ser	Pro	Ile	Gln	Leu	Gln	Arg	Arg	Asn	Gly	Ser	Lys	Phe
225					230					235					240
Ser	Val	Tyr	Asp	Val	Ser	Ile	Leu	Ile	Pro	Ile	Ile	Ala	Leu	Met	Val
				245					250					255	
Tyr	Arg	Cys	Ala	Pro	Pro	Arg	Gly	Ile	Ile	Thr	Ser	Lys	Thr	Lys	Ser
			260					265					270		
Leu	Val	Pro	Arg	Gly	Ser	Lys	Ala	Asp	Val	Cys	Met	Asp	Pro	Glu	Pro
		275					280					285			
Ile	Val	Arg	Ile	Val	Gly	Arg	Asn	Gly	Leu	Cys	Val	Asp	Val	Arg	Asp
		290				295					300				
Gly	Arg	Phe	His	Asn	Gly	Asn	Ala	Ile	Gln	Leu	Trp	Pro	Cys	Lys	Ser
305					310					315					320
Asn	Thr	Asp	Ala	Asn	Gln	Leu	Trp	Thr	Leu	Lys	Arg	Asp	Asn	Thr	Ile
				325					330					335	
Arg	Ser	Asn	Gly	Lys	Cys	Leu	Thr	Thr	Tyr	Gly	Tyr	Ser	Pro	Gly	Val
			340					345					350		
Tyr	Val	Met	Ile	Tyr	Asp	Cys	Asn	Thr	Ala	Ala	Thr	Asp	Ala	Thr	Arg
		355					360					365			
Trp	Gln	Ile	Trp	Asp	Asn	Gly	Thr	Ile	Ile	Asn	Pro	Arg	Ser	Ser	Leu

ES 2 380 677 T3

370	375	380
Val Leu Ala Ala Thr Ser Gly Asn Ser Gly Thr Thr Leu Thr Val Gln 385 390 395 400		
Thr Asn Ile Tyr Ala Val Ser Gln Gly Trp Leu Pro Thr Asn Asn Thr 405 410 415		
Gln Pro Phe Val Thr Thr Ile Val Gly Leu Tyr Gly Leu Cys Leu Gln 420 425 430		
Ala Asn Ser Gly Gln Val Trp Ile Glu Asp Cys Ser Ser Glu Lys Ala 435 440 445		
Glu Gln Gln Trp Ala Leu Tyr Ala Asp Gly Ser Ile Arg Pro Gln Gln 450 455 460		
Asn Arg Asp Asn Cys Leu Thr Ser Asp Ser Asn Ile Arg Glu Thr Val 465 470 475 480		
Val Lys Ile Leu Ser Cys Gly Pro Ala Ser Ser Gly Gln Arg Trp Met 485 490 495		
Phe Lys Asn Asp Gly Thr Ile Leu Asn Leu Tyr Ser Gly Leu Val Leu 500 505 510		
Asp Val Arg Arg Ser Asp Pro Ser Leu Lys Gln Ile Ile Leu Tyr Pro 515 520 525		
Leu His Gly Asp Pro Asn Gln Ile Trp Leu Pro Leu Phe Leu Glu His 530 535 540		
His His His His His 545		

REIVINDICACIONES

1. Procedimiento para la producción de polipéptidos o proteínas en forma bicatenaria en las que las dos cadenas están unidas por puente disulfuro por medio de expresión recombinante en células huésped de *E. coli*, en el que

(i) el polipéptido o la proteína ejerce su actividad biológica como polipéptido o proteína bicatenaria con puente disulfuro.

(ii) el resto de aminoácido C terminal de la primera cadena es un resto de aminoácido básico;

(iii) la segunda cadena de la proteína/del polipéptido presenta de manera N terminal en la dirección N-C de 1 a 20 restos de aminoácido y una secuencia de pentapéptido VPXGS denominada como PRS, en la que X significa cualquier aminoácido de origen natural, en la que V significa Val, Leu, Ile, Ala, Phe, Pro o Gly, en la que P significa Pro, Leu, Ile, Ala, Phe, Val o Gly, en la que G significa Gly, Leu, Ile, Ala, Pro, Phe o Val y en la que S significa Ser, Tyr, Trp o Thr; y

(iv) el procedimiento comprende las siguientes etapas:

(a) modificar el polipéptido o la proteína, al nivel de ácido nucleico, de modo que el polipéptido o la proteína en su forma modificada en su región de bucle presente la secuencia VPXGS y en dirección N terminal de esta secuencia PRS a una distancia de 1 a 20 restos de aminoácido un resto de aminoácido básico, siendo X, V, P, G y S tal como se definen anteriormente, y estando definida la región de bucle como la secuencia de aminoácidos que se encuentra entre los dos restos de cisteína que participan en el puente disulfuro;

(b) introducir el constructo modificado al nivel de ácido nucleico en células de *E. coli*;

(c) cultivar y a continuación lisar las células huésped; y

(d) aislar el polipéptido o la proteína bicatenaria con puente disulfuro.

2. El procedimiento según la reivindicación 1, en el que el resto de aminoácido básico en (ii) es un resto de Arg o Lys.

3. El procedimiento según la reivindicación 1 ó 2, en el que la primera cadena del polipéptido/de la proteína es la cadena ligera del polipéptido/de la proteína y la segunda cadena es la cadena pesada del polipéptido/de la proteína, especialmente en el que la proteína es una proteína híbrida.

4. El procedimiento según una de las reivindicaciones 1 a 3, en el que el polipéptido o la proteína es una neurotoxina botulínica, especialmente la neurotoxina botulínica del serotipo A (BoNT(A)), y especialmente el fragmento LH_N de BoNT(A).

5. El procedimiento según la reivindicación 4, en el que la secuencia PRS VPXGS se inserta entre los aminoácidos Leu₄₄₂ y Lys₄₄₈ de BoNT(A) con la delección de los aminoácidos 443-447, insertándose especialmente la secuencia PRS VPRGS, VPYGS, VPHGS o VPQGS.

6. El procedimiento según una de las reivindicaciones 1 a 4, en el que la secuencia PRS VPXGS se inserta en el octapéptido Lys₄₃₈ - Ile₄₄₅ de BoNT(B), en el 15-mero His₄₃₈ - Asp₄₅₂ de BoNT(C1) o en el 13-mero Lys₄₁₃ - Ile₄₂₅ de BoNT(E) con la delección de al menos un aminoácido, insertándose especialmente la secuencia PRS VPXGS en forma del 17-mero GIITSKTKSLVPRGSKA o del 18-mero RGIITSKTKSLVPRGSKA.

7. El procedimiento según la reivindicación 3, en el que la proteína híbrida presenta los siguientes componentes A, B y C:

- un dominio efector que, mediante su actividad enzimática, puede inhibir la secreción en células diana o destruir las mismas o un dominio de toxina (componente A);

- una secuencia de bucle que presenta la secuencia VPXGS (componente B), así como

- un dominio de unión celular que confiere a la proteína de fusión/híbrida una especificidad celular (componente C); y el componente D opcional:

- un dominio de translocación.

8. El procedimiento según la reivindicación 7, en el que el dominio de toxina (A) es el dominio de la toxina diftérica, de la exotoxina de *Pseudomonas* o de ricina, especialmente en el que el dominio de toxina (A) es el fragmento PE40 (dominio III, dominio II y dominio Ib) o el fragmento PE38 (dominio III y dominio II) de la exotoxina de *Pseudomonas* o la cadena A de ricina.

9. El procedimiento según la reivindicación 7 u 8, en el que el dominio de unión celular (C) es un anticuerpo monoclonal, una afilina, una proteína de repetición de anquirina, una anticalina, un factor de crecimiento tal como TGF-alfa, FGF, VEGF o IGF-1 o una citocina tal como IL2, IL4 o IL6.

10. El procedimiento según la reivindicación 3, en el que la proteína híbrida presenta los siguientes componentes A, B y C:

- una proteína o un oligopéptido que confiere a la proteína de fusión una mejor solubilidad, provoca una mayor tasa de expresión y/o posibilita una purificación por afinidad (componente A),

- una secuencia de bucle que presenta la secuencia VPXGS (componente B), así como

- un polipéptido cualquiera (componente C).

11. El procedimiento según la reivindicación 10, en el que el componente A es la glutatión-S-transferasa (GST), la proteína de unión a maltosa (MBP), una marca de His, una marca Strep o una marca FLAG.
12. El procedimiento según una de las reivindicaciones anteriores, en el que las células de *E. coli* son células K12 de *E. coli*, especialmente células K12 de *E. coli* de las cepas M15[pREP4], XL1-BLUE o UT5600.
- 5 13. Polipéptido o proteína, encontrándose el polipéptido/la proteína como polipéptido/proteína bicatenaria con puente disulfuro y siendo biológicamente activo, caracterizado porque el extremo C terminal de la primera cadena del polipéptido/de la proteína es un resto de aminoácido básico y la segunda cadena del polipéptido/de la proteína presenta de manera N terminal en la dirección N-C de 1 a 20 restos de aminoácido y una secuencia de pentapéptido VPXGS denominada como PRS, en la que X significa cualquier aminoácido de origen natural, en la que V significa Val, Leu, Ile, Ala, Phe, Pro o Gly, en la que P significa Pro, Leu, Ile, Ala, Phe, Val o Gly, en la que G significa Gly, Leu, Ile, Ala, Pro, Phe o Val y en la que S significa Ser, Tyr, Trp o Thr.
- 10 14. El polipéptido o la proteína según la reivindicación 13, en el que el resto de aminoácido básico es un resto de Arg o Lys.
- 15 15. El polipéptido o la proteína según la reivindicación 13 ó 14, en el que la primera cadena del polipéptido/de la proteína es la cadena ligera del polipéptido/de la proteína y la segunda cadena es la cadena pesada del polipéptido/de la proteína, especialmente en el que el extremo C terminal de la primera cadena es un resto de Lys.
16. El polipéptido o la proteína según una de las reivindicaciones 13 a 15, en el que la segunda cadena presenta de manera N terminal la secuencia de pentapéptido VPXGS, la secuencia de hexapéptido XVPXGS o la secuencia de heptapéptido XXVPXGS.
- 20 17. El polipéptido o la proteína según una de las reivindicaciones 13, 14 y 16, siendo el polipéptido/la proteína una neurotoxina botulínica, un derivado o un fragmento de la neurotoxina botulínica, especialmente el fragmento LH_N, o presenta la actividad biológica de una neurotoxina botulínica, especialmente siendo el polipéptido/la proteína la neurotoxina botulínica del serotipo A (BoNT(A)) o presentando la actividad biológica de BoNT(A), especialmente, siendo el polipéptido o la proteína el fragmento LH_N de BoNT(A) o presentando la actividad biológica del fragmento LH_N de BoNT(A).
- 25 18. El polipéptido o la proteína según una de las reivindicaciones 13 a 17, en el que la segunda cadena de manera N terminal presenta la secuencia de heptapéptido SLVPXGS, especialmente en la que X es R, Y, H o Q.
19. El polipéptido o la proteína según una de las reivindicaciones 13 a 16 ó 18, siendo la proteína una proteína híbrida, especialmente presentando la proteína híbrida los siguientes componentes A, B y C:
 - 30 - un dominio efector que, mediante su actividad enzimática, puede inhibir la secreción en células diana o destruir las mismas o un dominio de toxina (componente A);
 - una secuencia de bucle que presenta la secuencia VPXGS (componente B), así como un dominio de unión celular que confiere a la proteína de fusión/híbrida una especificidad celular (componente C); y el componente D opcional:
- 35 un dominio de translocación
20. El polipéptido o la proteína según la reivindicación 19, en el que el dominio de toxina (A) es el dominio de la toxina diftérica, de la exotoxina de *Pseudomonas* o de ricina, especialmente en el que el dominio de toxina (A) es el fragmento PE40 (dominio III, dominio II y dominio Ib) o el fragmento PE38 (dominio III y dominio II) de la exotoxina de *Pseudomonas* o la cadena A de ricina.
- 40 21. El polipéptido o la proteína según la reivindicación 19 ó 20, en el que el dominio de unión celular (C) es un anticuerpo monoclonal, una afilina, una proteína de repetición de anquirina, una anticalina, un factor de crecimiento tal como TGF-alfa, FGF, VEGF o IGF-1 o una citocina tal como IL2, IL4 o IL6.
22. El polipéptido o la proteína según la reivindicación 19, presentando la proteína híbrida los siguientes componentes A, B y C:
 - 45 - una proteína o un oligopéptido que confiere a la proteína de fusión una mejor solubilidad, provoca una mayor tasa de expresión y/o posibilita una purificación por afinidad (componente A),
 - una secuencia de bucle que presenta la secuencia VPXGS (componente B), así como
 - un polipéptido cualquiera (componente C), especialmente en el que el componente A es la glutatión-S-transferasa (GST), la proteína de unión a maltosa (MBP), una marca de His, una marca Strep o una marca FLAG.
- 50 23. Preparación farmacéutica que comprende el polipéptido o la proteína según una de las reivindicaciones 13 a 22.

Fig. 1

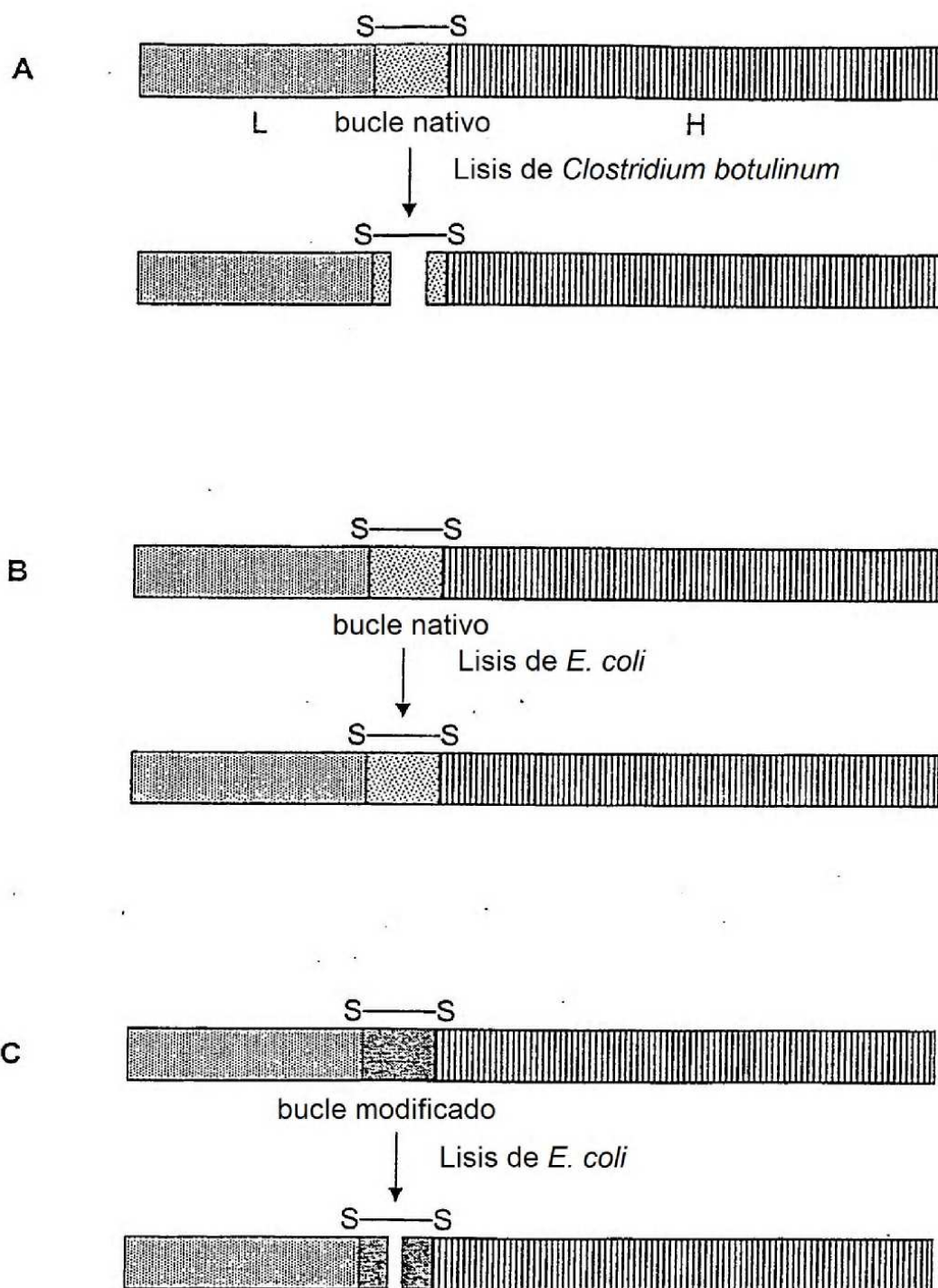


Fig. 2

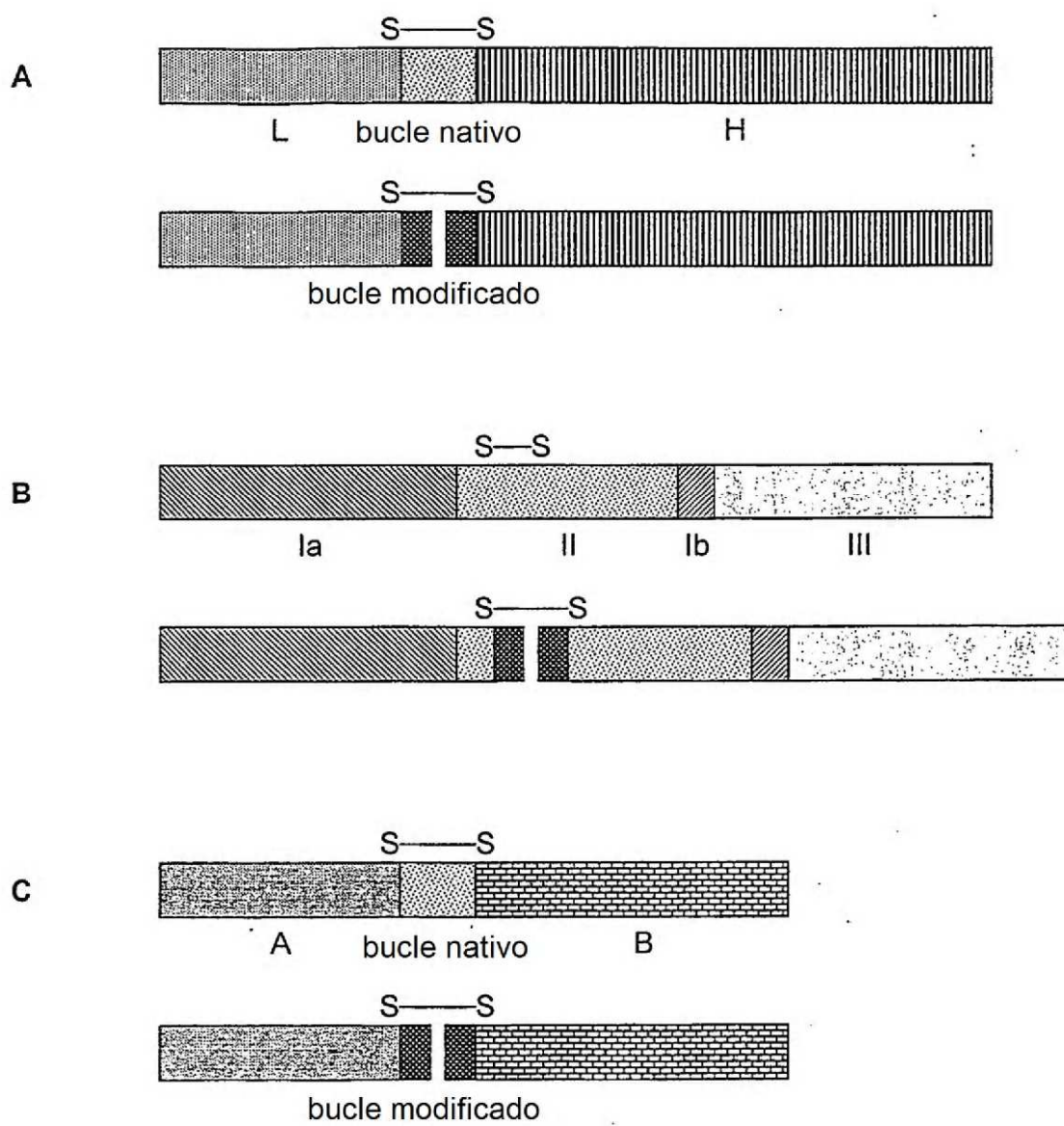


Fig. 3

1) BoNT(A)-WT	TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG GCC TTA AAT GAT TTA TGT C V R G I I I T S K T K K S L D K G Y N K A L N D L C 430 443 447 454
2) BoNT(A)-Mod1	TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG CGT GGA TCC AAG GCC TTA AAT GAT TTA TGT C V R G I I I T S K T K K S L V P R G S K A L N D L C ↓
3) BoNT(A)-Mod2	TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG TAC GGA TCC AAG GCC TTA AAT GAT TTA TGT C V R G I I I T S K T K K S L V P Y G S K A L N D L C ↓
4) BoNT(A)-Mod3	TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG CAT GGA TCC AAG GCC TTA AAT GAT TTA TGT C V R G I I I T S K T K K S L V P H G S K A L N D L C ↓
5) BoNT(A)-Mod4	TGT GTA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG CAA GGA TCC AAG GCC TTA AAT GAT TTA TGT C V R G I I I T S K T K K S L V P Q G S K A L N D L C ↓

Fig. 4

1) BoNT(B)-WT	TGT AAA AGT GTT AAA GCT CCA GGA ATA TGT C K S V K A P G I C 437 446
2) BoNT(B)-Mod1	TGT AAA AGT GGG ATA ATA ACT TCT AAA ACT AAA TCA TTG GTT CCA CGT GGA TCC AAG GCC TCA GGA ATA TGT C K S G I I T S K T K K S L V P R G S K A S G I C 437 460
3) BoNT(C1)-WT	TGT CAT AAA GCA ATA GAT GGT AGA TCA TTA TAT AAT AAA ACA TTA GAT TGT C H K A I D G R S L Y N K T L D C 437 453
4) BoNT(C1)-Mod1	TGT CAT AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG CGT GGA TCC AAG GCC TTA GAT TGT C H R G I I T S K T K K S L V P R G S K A L D C 437 459
5) BoNT(E)-WT	TGT AAA AAT ATT GTT TCT GTA AAA GGC ATA AGG AAA TCA ATA TGT C K N I V S V K G I R K S I C 412 426
6) BoNT(E)-Mod1	TGT AAA AGT GGG ATA ATA ACT TCT AAA ACT AAA TCA TTG GTT CCA CGT GGA TCC AAG GCC TCA ATA TGT C K S G I I T S K T K K S L V P R G S K A S I C 412 434

Fig. 5

1) PE40-WT	TGC CAC CTG CCG CTG GAG ACT TTC ACC CGT CAT CGC CAG CCG GGC TGG GAA CAA CTG GAG CAG TGC C H L P L E T F T R H R Q P R G W E Q L E Q C 13 35
2) PE40-Mod1	TGC CAC AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GTT CCG CGT GGA TCC AAG GCC CTG GAG CAG TGC C H R G I I T S K T K K S L V P R G S K A L E Q C 13 36
3) DT-WT	TGT GCA GGA AAT CGT GTC AGG CGA TCA GTA GGT TGC TCA TTG TCA TGC C A G N R V R R S V G S S L S C 186 201
4) DT-Mod1	TGT GCA AGA GGG ATA ATA ACT TCT AAA ACT AAA TCA TTG GTT CCA CGT GGA TCC AAG GCC TTG TCA TGC C A R G I I T S K T K K S L V P R G S K A L S C 186 208
5) Ricina-WT	TGC GCA CCT CCA CCA TCG TCA CAG TTT TCT TTG CTT ATA AGG CCA GTG GTA CCA AAT TTT AAT GCT GAT GTT TGT C A P P S S Q F S L L I R P V V P N F N A D V C 259 283
6) Ricina-Mod1	TGC GCA CCT CCT AGG GGG ATA ATA ACT TCT AAA ACT AAA TCA TTG GTT CCA CGT GGA TCC AAG GCT GAT GTT TGT C A P P R G I I T S K T K K S L V P R G S K A D V C 259 283

Fig. 6

Nº 1: BoNT(A)-L /dir.	ACCACTCCATGGCATTGTTTAATAACAATTTAAT
Nº 2: BoNT(A)-L/inv.	ACCACCGTCGACTTAGTGGTGGTGGTGGTGGTGGTCTCGAGACGCGTATTTAAGGC CTTGGATCCACGCGGAACATAATGATTTAGTTTGTAGAAGTTATTATC
Nº 3: BoNT(A)-H _N /dir.	ACCACCAGGCCTTAAATGATTTATGTATCAAAGTTAA
Nº 4: BoNT(A)-H _N /inv.	ACCACCCTCGAGAATATTCTTAATATATTTCAGTAAATG
Nº 5: BoNT(A)-H _{CN} /inv.	ACCACCCTCGAGACCTGAATTTGATTGATTATCATATA
Nº 6: BoNT(B)-L/dir.	ACCACTCCATGGCAGTTACAATAAATAATTTTAATTATA
Nº 7: BoNT(B)-L/inv.	ACCACTAGGCCTTGGATCCACGTGGAACCAATGATTTAGTTTGTAGAAGTTATTA TCCCACCTTTACACATTTGTATCTTATATAC
Nº 8: BoNT(B)-H _N /dir.	ACCACTAGGCCTCAGGAATATGTATTGATGTTGA
Nº 9: BoNT(B)-H _N /inv.	ACCACTCTCGAGAATTTGCTATTATATTATTATAACA
Nº 10: BoNT(C)-L/dir.	ACCACTCCATGGCAATAACAATTAACAACCTTTAATTATTC
Nº 11: BoNT(C)-L/inv.	GAGAAATAGGCCTTGGATCCACGCGGAACATAATGATTTAGTTTGTAGAAGTTATTA TCCCTCTATGACAAAATTTTGTAAATAAATAAAGC
Nº 12: BoNT(C)-H _N /dir.	ACCACTAGGCCTTAGATTGTAGAGAGCTTTTAG
Nº 13: BoNT(C)-H _N /inv.	ACCACTCTCGAGTATATCTTTTAATAAAGAATTATTAG
Nº 14: BoNT(C)-H _C /inv.	ACCACTCTCGAGCTCTACCCATCCTGGATCCC
Nº 15: PE-II-bucle/dir.	CATGGGCAGCCTGGCCGCGCTGACCGCGCACCAGGCTTGCCACAGAGGGATAAT AACTTCTAAAACTAAATCATTAGTTCGCGTGGATCCAAGGCCCTGGAGCAGTG CGGGTAC
Nº 16: PE-II-bucle/inv.	CCGCACTGCTCCAGGGCCTTGGATCCACGCGGAACATAATGATTTAGTTTGTAGAA GTTATTATCCCTCTGTGGCAAGCCTGGTGCGCGGTGAGCGCGGCCAGGCTGCC
Nº 17: PE-II/dir.	ACCACTCCATGGGTTTCAGGGTACCCGGTGCGAGCGGCTGGTCGCC
Nº 18: PE-III/inv.	ACCACTACGCGTCAGGTCCTCGCGCGGCGGTTTG
Nº 19: DT-A/dir.	ACCACTCCATGGGCTGATGATGTTGTTGATTCTTC
Nº 20: DT-A/inv.	ACCACTAGGCCTTGGATCCACGTGGAACCAATGATTTAGTTTGTAGAAGTTATTA CCCTCTTGACAGGCTTGAGCCATATAC
Nº 21: DT-B _N /dir.	ACCACTAGGCCTTGTCATGCATAAATCTTGATTGG
Nº 22: DT-B _N /inv.	ACCACTCTCGAGAAATGGTTGCGTTTATGCCC
Nº 23: Ricina-A/dir.	ACCACTCCATGGTCCCCAAACAATACCCAATTATAAAC
Nº 24: Ricina-A/inv.	ACCACTCTCGAGTCAGGACCTAGGAGGTGCGCATCTATACAC
Nº 25: Ricina-B/dir.	ACCACTCCTAGGGGGATAATAACTTCTAAAACTAAATCATTTGGTTCCACGTGGA TCCAAGGCTGATGTTGTATGGATCC
Nº 26: Ricina-B/inv.	ACCACTCTCGAGAAATAATGGTAACCATATTTGGTTTGG

Fig. 7

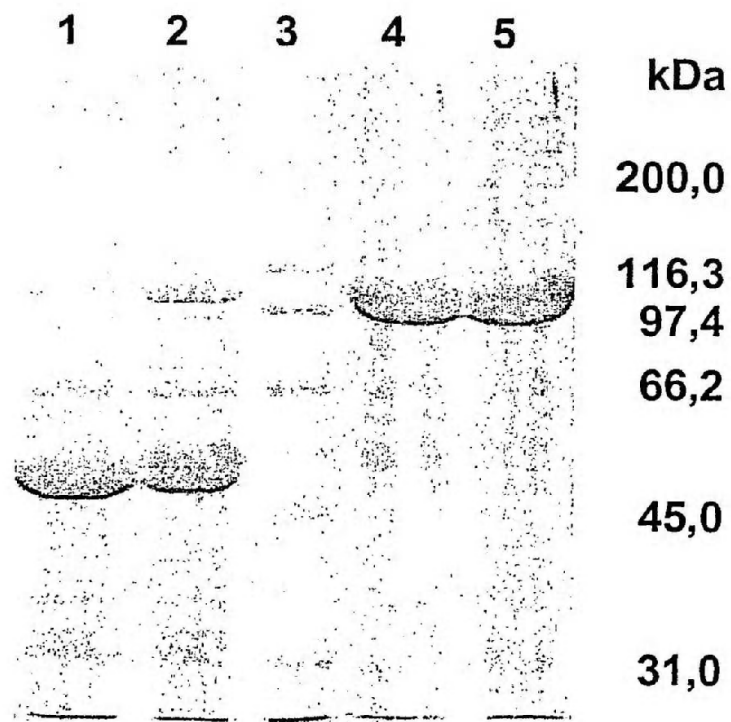


Fig. 8

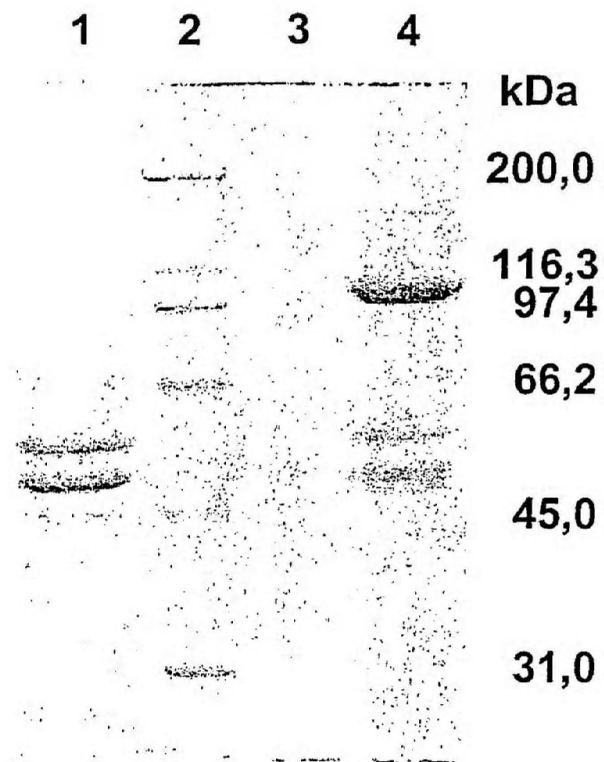


Fig. 9

