CLOSTRIDIAL NEUROTOXINS COMPRISING AN EXOGENOUS ACTIVATION LOOP

The present invention relates to clostridial neurotoxins and methods for activating and using the same.

Bacteria in the genus *Clostridia* produce highly potent and specific protein toxins, which can poison neurons and other cells to which they are delivered. Examples of such clostridial neurotoxins include the neurotoxins produced by *C. tetani* (TeNT) and by *C. botulinum* (BoNT) serotypes A-G, and X (*see* WO 2018/009903 A2), as well as those produced by *C. baratii* and *C. butyricum*.

Among the clostridial neurotoxins are some of the most potent toxins known. By way of example, botulinum neurotoxins have median lethal dose (LD_{50}) values for mice ranging from 0.5 to 5 ng/kg, depending on the serotype. Both tetanus and botulinum toxins act by inhibiting the function of affected neurons, specifically the release of neurotransmitters. While botulinum toxin acts at the neuromuscular junction and inhibits cholinergic transmission in the peripheral nervous system, tetanus toxin acts in the central nervous system.

Clostridial neurotoxins are expressed as single-chain polypeptides in *Clostridium*. Each clostridial neurotoxin has a catalytic light chain separated from the heavy chain (encompassing the N-terminal translocation domain and the C-terminal receptor binding domain) by an exposed region called the activation loop. During protein maturation proteolytic cleavage of the activation loop separates the light and heavy chain of the clostridial neurotoxin, which are held together by a disulphide bridge, to create fully active di-chain toxin. This process must be reproduced during recombinant toxin production.

Exogenous proteases such as trypsin or Lys-C are used for proteolytically activating single-chain clostridial neurotoxins. However, for some clostridial neurotoxins, incubation with Lys-C or trypsin results in partial or improper cleavage of the single-chain polypeptide resulting in the production of contaminating single-chain and/or inactive cleavage/degradation products (e.g. in the case of BoNT/E), necessitating purification of the full-length di-chain polypeptide. Thus, at present there is no universal exogenous protease for activation of clostridial neurotoxins. This is particularly problematic upon identification of a new clostridial neurotoxin or production of a modified (e.g. chimeric or hybrid) neurotoxin, which requires screening of multiple proteases to determine correct activation.

Botulinum neurotoxin serotype X (BoNT/X) was recently identified (WO 2018/009903 A2). It has been found that BoNT/X is particularly problematic to activate, and cleavage with trypsin or Lys-C completely degrades the polypeptide.

The present invention overcomes one or more of the above-mentioned problems.

The protease enterokinase exhibits a much higher substrate specificity compared to conventionally used trypsin and Lys-C. This protease recognises and cleaves immediately C-terminal to a DDDDK peptide sequence (SEQ ID NO: 72). Notably, this sequence is absent from all of the clostridial neurotoxin activation loops (*see* Figure 1), thus enterokinase has previously been ruled out as a protease for use in activating clostridial neurotoxins.

The present inventors have surprisingly found that enterokinase recognises and cleaves immediately C-terminal to an IDGR sequence present in the BoNT/C1 activation loop (*see* Figure 1). Advantageously, this sequence can also be recognised and cleaved by factor Xa, another protease exhibiting high substrate specificity (e.g. compared to trypsin and Lys-C). Moreover, the BoNT/C1 activation loop also has lysine and arginine residues, allowing cleavage by either lysine or trypsin. Thus, the present inventors have surprisingly found that the BoNT/C1 loop constitutes a universal activation loop for clostridial neurotoxins, thus providing the flexibility to use four different proteases.

In one aspect the invention provides a method for proteolytically processing a single-chain clostridial neurotoxin (e.g. of an engineered clostridial neurotoxin described herein) into a corresponding di-chain clostridial neurotoxin, the method comprising:

a. providing a single-chain clostridial neurotoxin; and

b. contacting the single-chain clostridial neurotoxin with enterokinase;

wherein the single-chain clostridial neurotoxin has an activation loop comprising the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1); and

wherein enterokinase hydrolyses a peptide bond of the activation loop thereby producing a di-chain clostridial neurotoxin (e.g. an engineered di-chain clostridial neurotoxin described herein).

In a related aspect the invention provides a method for proteolytically processing a single-chain clostridial neurotoxin (e.g. of an engineered clostridial neurotoxin described herein) into a corresponding di-chain clostridial neurotoxin, the method comprising:

a. providing a single-chain clostridial neurotoxin; and

b. contacting the single-chain clostridial neurotoxin with factor Xa ;

wherein the single-chain clostridial neurotoxin has an activation loop comprising the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1); and

wherein factor Xa hydrolyses a peptide bond of the activation loop thereby producing a di-chain clostridial neurotoxin (e.g. an engineered di-chain clostridial neurotoxin described herein).

The single-chain clostridial neurotoxin is preferably an engineered single-chain clostridial neurotoxin of the present invention, wherein the activation loop is an exogenous activation loop. Advantageously, the present inventors have found that replacing an endogenous clostridial neurotoxin activation loop with an exogenous activation loop shown as SEQ ID NO: 1 (which contains a protease cleavage site in its natural context) overcomes problems associated with modifying an endogenous activation loop to insert a protease cleavage site (e.g. a factor Xa cleavage site, such as Ile-Asp-Gly-Arg [SEQ ID NO: 18] or Ile-Glu-Gly-Arg [SEQ ID NO: 19]). In particular, modifying an endogenous activation loop to insert a protease cleavage site can lead to conformational changes, which, in turn, can have a negative effect on cleavage efficiency (*see* Example 7 herein).

In a particularly preferred embodiment the methods of the present invention comprise the use of enterokinase.

In one aspect the present invention is directed to use of enterokinase for hydrolysing a peptide bond of a polypeptide (e.g. a clostridial neurotoxin) comprising a sequence shown as SEQ ID NO: 18 or SEQ ID NO: 19 (preferably SEQ ID NO: 18). Preferably enterokinase hydrolyses a peptide bond immediately C-terminal to SEQ ID NO: 18 or SEQ ID NO: 19 (more preferably SEQ ID NO: 18) comprised within the polypeptide sequence. In one embodiment a polypeptide comprises a polypeptide sequence shown as SEQ ID NO: 1 or a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

The present invention also provides a method for manufacturing an engineered clostridial neurotoxin, the method comprising:

a. identifying an endogenous activation loop of a clostridial neurotoxin, wherein the clostridial neurotoxin is characterised in that a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C; and

b. replacing the endogenous activation loop with an exogenous activation loop thereby providing an engineered clostridial neurotoxin, wherein the exogenous activation loop comprises the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

The present invention also provides a method for manufacturing an engineered clostridial neurotoxin, the method comprising:

a. identifying an endogenous activation loop of a clostridial neurotoxin, wherein the clostridial neurotoxin is characterised in that the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C; and

b. replacing the endogenous activation loop with an exogenous activation loop thereby providing an engineered clostridial neurotoxin, wherein the exogenous activation loop comprises the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

In one embodiment the clostridial neurotoxin is characterised in that a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C and the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C.

In embodiments where the endogenous activation loop is inefficiently proteolytically processed by trypsin (and preferably a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is not hydrolysed by trypsin), the method may further comprise contacting an engineered clostridial neurotoxin with trypsin, which is capable of hydrolysing a peptide bond in the exogenous activation loop of the engineered clostridial neurotoxin. Similarly, in embodiments where the endogenous activation loop is inefficiently proteolytically processed by Lys-C (and preferably a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is not hydrolysed by Lys-C), the method may further comprise contacting an engineered clostridial neurotoxin with Lys-C, which is capable of hydrolysing a peptide bond in the exogenous activation loop of the engineered clostridial neurotoxin.

In one embodiment a method comprises a step of screening a clostridial neurotoxin for its suitability for use in a method of the invention. The screening step may comprise determining whether a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C. Alternatively or additionally, the

screening step may comprise determining whether the endogenous activation loop of a clostridial neurotoxin is inefficiently proteolytically processed by trypsin or Lys-C.

In contrast to the clostridial neurotoxin (pre-engineering) in one embodiment an engineered clostridial neurotoxin of the invention is not inefficiently proteolytically processed by enterokinase or factor Xa and/or a peptide bond outside of the exogenous activation loop of the engineered clostridial neurotoxin is not hydrolysed by enterokinase or factor Xa. Thus, the clostridial neurotoxin (pre-engineering) is preferably resistant to proteolytic processing by enterokinase and/or factor Xa.

A clostridial neurotoxin may be identified as being suitable for engineering in a method of the invention by an assay comprising contacting 1 mg of clostridial neurotoxin with at least 0.25 µg of trypsin at ≥3350 units/mg or Lys-C at ≥200 units/mg in a 50 mM Tris-HCl pH 8.0, 50 mM NaCl reaction buffer for at least 5 hours at at least 4 °C.

In one embodiment the assay comprises contacting 1 mg of clostridial neurotoxin with 0.25 µg of trypsin at ≥3350 units/mg (~1:611 molar ratio of clostridial neurotoxin to trypsin) or Lys-C at ≥200 units/mg (~1:734 molar ratio of clostridial neurotoxin to Lys-C) in a 50 mM Tris-HCl pH 8.0, 50 mM NaCl reaction buffer for 18 hours at 4 °C.

In another embodiment the assay comprises contacting 1 mg of clostridial neurotoxin with 0.40 µg of trypsin at ≥3350 units/mg (~1:978 molar ratio of clostridial neurotoxin to trypsin) or Lys-C at ≥200 units/mg (~1:1174 molar ratio of clostridial neurotoxin to Lys-C) in a 50 mM Tris-HCl pH 8.0, 50 mM NaCl reaction buffer for 5 hours at 20 °C.

The trypsin used is preferably commercially available TrypZean (Sigma# T3568). Trypsin may have a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 47. In one embodiment trypsin may have a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 47. Preferably trypsin may have a polypeptide sequence shown as SEQ ID NO: 47. One unit of said trypsin (Trypzean) is defined as the amount of enzyme that will produce a change of absorbance at 253 nm of 0.003 per min at pH 7.6 at 25 °C using 0.23 mM Na-Benzoyl-L-Arginine Ethyl Ester Solution (BAEE) as substrate in a reaction volume of 3.2 mL.

The Lys-C used is preferably commercially available Lys-C (Sigma# 000000011047825001). Lys-C may have a polypeptide sequence having at least 70% sequence identity to SEQ ID

NO: 48. In one embodiment Lys-C may have a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 48. Preferably Lys-C may have a polypeptide sequence shown as SEQ ID NO: 48. One unit of said Lys-C is defined as the amount of enzyme that will hydrolyze 1.0 µmol Tos-Gly-Pro-Lys-pNA per min at 25°C, pH 7.

If one or more cleavage products additional to those of the H-chain and L-chain of the clostridial neurotoxin are observed by SDS-PAGE (preferably when stained with Coomassie or a dye of equivalent sensitivity), then it is confirmed that a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C. Preferably if at least 3, 4, 5, 6, 7, 8, 9 or 10 cleavage products additional to those of the chain and L-chain of the clostridial neurotoxin are observed by SDS-PAGE after carrying out an assay described above, then it is confirmed that a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C.

Additionally or alternatively, if less than 70% of endogenous activation loop is proteolytically processed by trypsin or Lys-C to yield a di-chain clostridial neurotoxin (assessed by way of SDS-PAGE, preferably when stained with Coomassie or a dye of equivalent sensitivity, after carrying out an assay described above), then it is confirmed that the endogenous activation loop is inefficiently proteolytically cleaved by trypsin or Lys-C. Preferably if less than 60%, 50%, 40%, 30%, 10% or 5% of endogenous activation loop is proteolytically processed by trypsin or Lys-C (assessed by way of SDS-PAGE after carrying out an assay described above) then the clostridial neurotoxin can be characterised in that the endogenous activation loop is inefficiently proteolytically cleaved by trypsin or Lys-C. More preferably if less than 30% of endogenous activation loop is proteolytically processed by trypsin or Lys-C (assessed by way of SDS-PAGE after carrying out an assay described above) then the clostridial neurotoxin can be characterised in that the endogenous activation loop is inefficiently proteolytically cleaved by trypsin or Lys-C.

A clostridial neurotoxin (pre-engineering) is preferably one in which a peptide bond (either within or outside of the activation loop) is not, or is not substantially, hydrolysed by enterokinase or factor Xa. The term “not substantially hydrolysed” means that less than 10%, 5%, 4%, 3%, 2% or 1% of the clostridial neurotoxin present in a reaction contains a peptide bond that has been hydrolysed by enterokinase or factor Xa in a method of the invention.

In one embodiment a method of the invention further comprises contacting an engineered clostridial neurotoxin with enterokinase or factor Xa (more preferably enterokinase) thereby producing a corresponding di-chain engineered clostridial neurotoxin.

In one aspect the invention provides an engineered clostridial neurotoxin (e.g. obtainable by a method of the invention), wherein an endogenous activation loop of a clostridial neurotoxin has been replaced by an exogenous activation loop thereby providing an engineered clostridial neurotoxin,

wherein the exogenous activation loop comprises the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

In one embodiment the clostridial neurotoxin (pre-engineering) is characterised in that: a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C. In one embodiment the clostridial neurotoxin (pre-engineering) is characterised in that: the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C. In another embodiment the clostridial neurotoxin (pre-engineering) is characterised in that: a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C; and the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C. Confirmation of these features of the clostridial neurotoxin (pre-engineering) is preferably by way of an aforementioned assay.

The invention may comprise replacing an endogenous activation loop of any clostridial neurotoxin with an exogenous activation loop described herein. Preferably the clostridial neurotoxin is not BoNT/C1. The clostridial neurotoxin may be a botulinum neurotoxin or a tetanus neurotoxin. Preferably the clostridial neurotoxin is a botulinum neurotoxin (BoNT), such as BoNT/A, BoNT/B, BoNT/D, BoNT/E, BoNT/F, BoNT/G or BoNT/X.

In one embodiment a clostridial neurotoxin for use in the present invention is BoNT/X, BoNT/E or a BoNT/A1C1 hybrid. Preferably the clostridial neurotoxin is BoNT/X or BoNT/E, both of which have been characterised in that trypsin and/or Lys-C hydrolyses a peptide bond outside of the endogenous activation loop thereof, and/or in that both clostridial neurotoxins contain an endogenous activation loop that is inefficiently proteolytically processed by trypsin and/or Lys-C. Most preferably a clostridial neurotoxin for use in the present invention is BoNT/X.

The term "endogenous activation loop" as used herein means an activation loop present in a subject clostridial neurotoxin, e.g. a subject clostridial neurotoxin of the indicated serotype. For example, BoNT/A1 includes a BoNT/A1 heavy chain and light chain, thus the endogenous activation loop of BoNT/A1 is an A1 activation loop. For clostridial neurotoxin chimeras or hybrids, the person skilled in the art can identify the "endogenous activation loop", for example by determining the serotype(s) from which the L-chain and H_N domain are derived. In some embodiments, a chimera or hybrid clostridial neurotoxin may have an endogenous activation loop that is a fusion of an activation loop from two different serotypes. By way of example, a chimeric clostridial neurotoxin such as BoNT/A1C1 has a BoNT/A1 light chain and translocation domain, thus the endogenous BoNT/A1C1 activation loop is an A1 activation loop. Examples of activation loops are provided in Figure 1.

Preferably an "endogenous activation loop" is any activation loop that is not SEQ ID NO: 1. In one embodiment an "endogenous activation loop" is any activation loop that is not SEQ ID NO: 2 and/or SEQ ID NO: 3.

By contrast, an "exogenous activation loop" as used herein means an activation loop that is different to the endogenous activation loop present in a subject clostridial neurotoxin, e.g. a subject clostridial neurotoxin of the indicated serotype. For example, a BoNT/C1 activation loop has a different polypeptide sequence to a wild-type BoNT/A1 activation loop, therefore the BoNT/C1 activation loop is exogenous to BoNT/A1. For clostridial neurotoxin chimeras or hybrids, the person skilled in the art can determine whether an activation loop is an "exogenous activation loop", for example by determining the serotype(s) from which the L-chain and H_N domain are derived. Where the L-chain is a BoNT/B L-chain and the H_N domain is from BoNT/D, the endogenous activation loop may have a portion of a BoNT/B sequence and a portion of a BoNT/D sequence, and if an activation loop (e.g. a C1 activation loop) is different thereto, it is considered an "exogenous activation loop".

Determination of whether an activation loop is an "exogenous activation loop" may be made by aligning the sequence of a subject clostridial neurotoxin with the activation loop, and seeing if the activation loop is present in the subject clostridial neurotoxin sequence. If it is absent, then the activation loop can be identified as an exogenous activation loop.

Preferably, the entire endogenous activation loop is replaced by an exogenous activation loop described herein. However, in some embodiments a portion of the endogenous

activation loop is replaced, such as at least 5, 10, 15, 20, 25, 30, 35 or 40 amino acid residues of the endogenous activation are replaced.

Replacement of an endogenous activation loop may be achieved by any method known in the art. For example, replacement might be achieved by way of an amino acid modification. In one embodiment an endogenous activation loop may be replaced by deleting one or more amino acid residues of the endogenous activation loop. An endogenous activation loop may be replaced by substituting one or more amino acid residues of the endogenous activation loop with amino acid residues of an exogenous activation loop. In some embodiments an endogenous activation loop (or a portion thereof) may be deleted, and an exogenous activation loop inserted, preferably at the position formally occupied by the endogenous activation loop. Alternatively, the endogenous activation loop may be retained in an engineered clostridial neurotoxin of the invention, and preferably inactivated (e.g. by way of mutation). It is preferred that the endogenous activation loop (or a portion thereof, more preferably the entire endogenous activation loop) is not present in the engineered clostridial neurotoxin of the invention. It is preferred that the exogenous activation loop occupies the position in the clostridial neurotoxin formally occupied by the endogenous activation loop.

Methods for modifying proteins by substitution, insertion or deletion of amino acid residues are known in the art and may be employed in the practice of the present invention. By way of example, amino acid modifications may be introduced by modification of a DNA sequence encoding a clostridial neurotoxin. This can be achieved using standard molecular cloning techniques, for example by site-directed mutagenesis where short strands of DNA (oligonucleotides) coding for the desired amino acid(s) are used to replace the original coding sequence using a polymerase enzyme, or by inserting/deleting parts of the gene with various enzymes (e.g., ligases and restriction endonucleases). Alternatively a modified gene sequence can be chemically synthesised.

In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70 or SEQ ID NO: 71. In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID

NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70 or SEQ ID NO: 71. Preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70 or SEQ ID NO: 71.

In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 24, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 68 or SEQ ID NO: 69. In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 24, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 68 or SEQ ID NO: 69. Preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 24, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 68 or SEQ ID NO: 69.

In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 20, SEQ ID NO: 21 or SEQ ID NO: 24. In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 20, SEQ ID NO: 21 or SEQ ID NO: 24. Preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 20, SEQ ID NO: 21 or SEQ ID NO: 24.

Preferably an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 20. In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 20. More preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 20.

Preferably an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 21. In one embodiment an

endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 21. More preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 21.

Preferably an endogenous activation loop comprises a polypeptide sequence having at least 70% (e.g. at least 80% or 90%) sequence identity to SEQ ID NO: 24. In one embodiment an endogenous activation loop comprises a polypeptide sequence having at least 95% sequence identity to SEQ ID NO: 24. More preferably, an endogenous activation loop comprises a polypeptide sequence shown as SEQ ID NO: 24.

The present invention encompasses methods and clostridial neurotoxins in which an endogenous activation loop has been replaced by an exogenous activation loop, such as an exogenous activation loop comprising a polypeptide shown as Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1). Xaa or Yaa can be any amino acid. The number of amino acids at position Xaa and Yaa are indicated by the letters 'a' and 'b', respectively. In one embodiment 'a' and 'b' can be any integer that allows for proteolytic cleavage of the activation loop and yields an active di-chain clostridial neurotoxin. In one embodiment 'a' is at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10. In one embodiment 'b' is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. In one embodiment 'a' is ≤12, ≤11, ≤10, ≤9, ≤8, ≤7, ≤6, ≤5 or ≤4. In one embodiment 'b' is ≤20, ≤19, ≤18, ≤17, ≤16, ≤15, ≤14, ≤13, ≤12, ≤11, ≤10 or ≤9.

In one embodiment 'a' is 1-12, for example 1-10. Preferably 'a' is 1-7, such as 2-4. More preferably 'a' is 3. In one embodiment 'b' is 1-20, for example 4-15. Preferably 'b' is 6-10. More preferably 'b' is 8.

It is not intended that Xaa or Yaa be limited to only one type of amino acid. Thus, one or more residues present at position Xaa may be independently selected from the standard amino acids: aspartic acid, glutamic acid, arginine, lysine, histidine, asparagine, glutamine, serine, threonine, tyrosine, methionine, tryptophan, cysteine, alanine, glycine, valine, leucine, isoleucine, proline, and phenylalanine. One or more residues present at position Yaa may be independently selected from the standard amino acids: aspartic acid, glutamic acid, arginine, lysine, histidine, asparagine, glutamine, serine, threonine, tyrosine, methionine, tryptophan, cysteine, alanine, glycine, valine, leucine, isoleucine, proline, and phenylalanine. Preferably an amino acid at position Yaa (more preferably immediately C-terminal to the Arg residue of SEQ ID NO: 1) is not proline.

Alternatively/additionally, one or more residues present at position Xaa or Yaa may be independently selected from a non-standard amino acid (an amino acid that is not part of the standard set of 20 described above). By way of example, non-standard amino acids may include 4-hydroxyproline, 6-N-methyl lysine, 2-aminoisobutyric acid, isovaline, α -methyl serine, trans-3-methylproline, 2,4-methano-proline, cis-4-hydroxyproline, trans-4-hydroxy-proline, N-methylglycine, allo-threonine, methyl-threonine, hydroxy-ethylcysteine, hydroxyethylhomo-cysteine, nitro-glutamine, homoglutamine, pipecolic acid, tert-leucine, norvaline, 2-azaphenylalanine, 3-azaphenyl-alanine, 4-azaphenyl-alanine, L-Ornithine, L-2-amino-3-guanidinopropionic acid, or D-isomers of Lysine, Arginine and/or Ornithine, and 4-fluorophenylalanine. Methods for introducing non-standard amino acids into proteins are known in the art, and include recombinant protein synthesis using *E. coli* auxotrophic expression hosts.

Properties of the standard amino acids are indicated in the table below:

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	Charged (acidic)
Glutamic acid	Glu	E	Charged (acidic)
Arginine	Arg	R	Charged (basic)
Lysine	Lys	K	Charged (basic)
Histidine	His	H	Uncharged (polar)
Asparagine	Asn	N	Uncharged (polar)
Glutamine	Gln	Q	Uncharged (polar)
Serine	Ser	S	Uncharged (polar)
Threonine	Thr	T	Uncharged (polar)
Tyrosine	Tyr	Y	Uncharged (polar)
Methionine	Met	M	Uncharged (polar)
Tryptophan	Trp	W	Uncharged (polar)
Cysteine	Cys	C	Uncharged (polar)
Alanine	Ala	A	Uncharged (hydrophobic)
Glycine	Gly	G	Uncharged (hydrophobic)
Valine	Val	V	Uncharged (hydrophobic)
Leucine	Leu	L	Uncharged (hydrophobic)
Isoleucine	Ile	I	Uncharged (hydrophobic)
Proline	Pro	P	Uncharged (hydrophobic)
Phenylalanine	Phe	F	Uncharged (hydrophobic)

The following amino acids are considered charged amino acids: aspartic acid (negative), glutamic acid (negative), arginine (positive), and lysine (positive).

The sequence Ile-Asp/Glu-Gly-Arg comprised in SEQ ID NO: 1 refers to the site surprisingly found by the present inventors to be recognised by enterokinase (as well as factor Xa). Preferably the sequence is Ile-Asp-Gly-Arg, e.g. Cys-$(Xaa)_a$-Ile-Asp-Gly-Arg-$(Yaa)_b$-Cys. It is believed that enterokinase and factor Xa hydrolyse a peptide bond immediately C-terminal to Arg of SEQ ID NO: 1 (i.e. the peptide bond between Arg and Yaa).

In one embodiment an amino acid residue at Xaa immediately N-terminal to Ile of SEQ ID NO: 1 is an uncharged hydrophobic amino acid, preferably alanine. In some embodiments 'a' is at least 2, and Xaa comprises at least a C-terminal uncharged polar amino acid and a charged basic amino acid immediately N-terminal thereto. The charged basic amino acid is preferably lysine. Thus in embodiments where 'a' is at least 2, Xaa may comprise at least Lys-Ala, wherein Ala is immediately N-terminal to Ile of SEQ ID NO: 1.

In one embodiment Xaa comprises or consists of the sequence HKA.

In one embodiment an amino acid residue at Yaa immediately C-terminal to Arg of SEQ ID NO: 1 is an uncharged polar amino acid, preferably serine. In some embodiments 'b' is at least 2, and Yaa comprises at least an N-terminal uncharged polar amino acid and an uncharged hydrophobic amino acid immediately C-terminal thereto. The uncharged hydrophobic amino acid is preferably leucine. Thus in embodiments where 'b' is at least 2, Yaa may comprise at least Ser-Leu, wherein Ser is immediately C-terminal to Arg of SEQ ID NO: 1.

In one embodiment Yaa comprises or consists of the sequence SLYNKTLDC.

In some embodiments an exogenous activation loop has at least 70% sequence identity to SEQ ID NO: 2. In one embodiment an exogenous activation loop has at least 80%, 85% or 90% sequence identity to SEQ ID NO: 2. Preferably an exogenous activation loop has at least 95% sequence identity to SEQ ID NO: 2. More preferably, an exogenous activation loop has at least 99% sequence identity to SEQ ID NO: 2.

In a particularly preferred embodiment an exogenous loop comprises SEQ ID NO: 2. More preferably an exogenous loop consists of SEQ ID NO: 2.

The exogenous loop may also be a variant of SEQ ID NO: 2, such as SEQ ID NO:3 or a sequence having at least 70% sequence identity thereto. SEQ ID NO: 3 is a variant of SEQ ID NO: 2 in which the enterokinase recognition site IDGR has been mutated to IEGR. In one embodiment an exogenous activation loop has at least 80%, 85% or 90% sequence identity to SEQ ID NO: 3. Preferably an exogenous activation loop has at least 95% sequence identity to SEQ ID NO: 3. More preferably, an exogenous activation loop has at least 99% sequence identity to SEQ ID NO: 3.

In a particularly preferred embodiment an exogenous loop comprises SEQ ID NO: 3. More preferably an exogenous loop consists of SEQ ID NO: 3.

A clostridial neurotoxin of the present invention (e.g. engineered clostridial neurotoxin) may be encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12. In one embodiment a clostridial neurotoxin of the present invention may be encoded by a nucleotide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12. Preferably, a clostridial neurotoxin of the present invention may be encoded by a nucleotide sequence comprising (more preferably consisting of) SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12.

A clostridial neurotoxin of the present invention (e.g. engineered clostridial neurotoxin) may comprise a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13. In one embodiment a clostridial neurotoxin of the present invention may comprise a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13. Preferably, a clostridial neurotoxin of the present invention may comprise (more preferably consist of) a polypeptide sequence shown as SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13.

The clostridial neurotoxin of the present invention (e.g. engineered clostridial neurotoxin) is preferably BoNT/X, wherein the clostridial neurotoxin is encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 4. In one embodiment the clostridial neurotoxin is encoded by a nucleotide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 4. Preferably the clostridial neurotoxin is encoded by a nucleotide sequence comprising (or consisting of) SEQ ID NO: 4. The clostridial neurotoxin of the

present invention is preferably BoNT/X, wherein the clostridial neurotoxin comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 5. In one embodiment the clostridial neurotoxin comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 5. Preferably the clostridial neurotoxin comprises (or consists of) a polypeptide sequence shown as SEQ ID NO: 5.

The clostridial neurotoxin of the present invention (e.g. engineered clostridial neurotoxin) is preferably BoNT/E, wherein the clostridial neurotoxin is encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 10. In one embodiment the clostridial neurotoxin is encoded by a nucleotide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 10. Preferably the clostridial neurotoxin is encoded by a nucleotide sequence comprising (or consisting of) SEQ ID NO: 10. The clostridial neurotoxin of the present invention is preferably BoNT/E, wherein the clostridial neurotoxin comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 11. In one embodiment the clostridial neurotoxin comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 11. Preferably the clostridial neurotoxin comprises (or consists of) a polypeptide sequence shown as SEQ ID NO: 11.

In some embodiments, the polypeptide sequences of the invention (or the nucleotide sequences encoding the same) may include a purification tag, such as a His-tag. It is intended that the present invention also encompasses polypeptide sequences (and nucleotide sequences encoding the same) where the purification tag is removed.

The present invention encompasses contacting a single-chain clostridial neurotoxin (e.g. an engineered clostridial neurotoxin of the invention) with a protease capable of hydrolysing a peptide bond in an activation loop of the single-chain clostridial neurotoxin thereby producing a di-chain clostridial neurotoxin. The protease may be an endopeptidase. The protease may be enterokinase, factor Xa, Lys-C or trypsin. Preferably the protease is enterokinase or factor Xa, more preferably enterokinase.

The term “enterokinase” or “EK” encompasses enterokinases described herein, as well as any protease having structural and/or functional similarity (preferably structural and functional similarity) that is capable of hydrolysing a peptide bond of SEQ ID NO: 1. A suitable enterokinase is enterokinase light chain, which is commercially available from NEB (#P8070). One unit may be defined as the amount of enzyme required to cleave 25 µg of a

MBP-EK-paramyosin-ΔSal substrate to 95% completion in 16 hours at 25°C in a total reaction volume of 25 µl (20 mM Tris-HCl, 50 mM NaCl, 2 mM $CaCl_2$ (pH 8.0 @ 25°C)).

In one embodiment an enterokinase comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 49. In some embodiments an enterokinase comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 49. Preferably an enterokinase comprises (more preferably consists of) SEQ ID NO: 49.

In some embodiments, enterokinase may further comprise a heavy chain, wherein the heavy and light-chains are connected by a disulphide bridge. Such enterokinases are commercially available (e.g. from R&D Systems).

The term "factor Xa" encompasses factor Xa described herein, as well as any protease having structural and/or functional similarity (preferably structural and functional similarity) that is capable of hydrolysing a peptide bond of SEQ ID NO: 1. A suitable factor Xa is commercially available from NEB (#P8010). One unit may be defined as the amount of factor Xa required to cleave 50 µg of an MBP fusion protein test substrate, MBP-ΔSal (substrate MBP-ΔSal is maltose-binding protein fused to a truncated form of paramyosin, with the amino acids Ile-Glu-Gly-Arg at the fusion joint) to 95% completion in 6 hours or less at 23°C in a reaction volume of 50 µl (20 mM Tris-HCl, 100 mM NaCl, 2 mM $CaCl_2$ (pH 8.0)).

In one embodiment a factor Xa comprises a polypeptide sequence having a heavy chain with at least 70% sequence identity to SEQ ID NO: 50 and a light chain with at least 70% sequence identity to SEQ ID NO: 51, wherein the heavy and light-chain are connected by a disulphide bridge. In some embodiments a factor Xa comprises a polypeptide sequence having a heavy chain with at least 80% or 90% sequence identity to SEQ ID NO: 50 and a light chain with at least 80% or 90% sequence identity to SEQ ID NO: 51, wherein the heavy and light-chain are connected by a disulphide bridge. Preferably a factor Xa comprises (more preferably consists of) SEQ ID NO: 50 and SEQ ID NO: 51, wherein the heavy and light-chain are connected by a disulphide bridge.

The contacting can occur under any suitable conditions that result in the production of greater than 30%, 40%, 50% or 60% (preferably greater than 70%) of single-chain clostridial neurotoxin being proteolytically processed into the corresponding di-chain clostridial neurotoxin without, or without substantial, hydrolysis of a peptide bond outside of the activation loop of said clostridial neurotoxin. "Without substantial hydrolysis" may mean less

than 5%, 4%, 3%, 2% or 1% of the clostridial neurotoxins contacted contain a peptide bond outside of the activation loop that has been hydrolysed by a protease in a method of the invention.

The skilled person can select appropriate reaction times, temperatures, buffers, and molar ratios of protease to single-chain clostridial neurotoxin to achieve the above. Optimisation of such conditions can be determined empirically using routine techniques, such as SDS-PAGE (e.g. stained with Coomassie or a dye of similar sensitivity) visual analysis of the reaction products following said contacting or spectrometric techniques (e.g. mass spectrometry).

When assessed by SDS-PAGE (e.g. stained with Coomassie or a dye of similar sensitivity), a method of the invention preferably results in the production of a clostridial neurotoxin L-chain and H-chain only.

In one embodiment the proteolytic processing by a protease in a method of the invention results in the production of less than 5 degradation products of a clostridial neurotoxin L-chain or H-chain, more preferably less than 4, 3, 2 or 1 degradation products. Preferably, the L-chain and H-chain produced by a method of the invention are full-length L-chain and H-chain.

Therefore, in a particularly preferred embodiment a protease used in a method of the invention (e.g. enterokinase or factor Xa) hydrolyses only a peptide bond of SEQ ID NO: 1, more preferably only the peptide bond between Arg and Yaa of SEQ ID NO: 1.

In one embodiment the contacting occurs for at least 1 hour, e.g. at least 2, 4, 6, 8, 10, 12, 14, 16, 18 or 20 hours.

In one embodiment the contacting occurs at a temperature of at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 °C.

In one embodiment contacting occurs at a temperature of 1 to 10 °C (preferably about 4 °C). Preferably contacting occurs at a temperature of 1 to 10 °C (more preferably about 4 °C) for 10-25 hours (preferably 15-20 hours).

In one embodiment contacting occurs at a temperature of 15-25 °C (preferably about 20 °C). Preferably contacting occurs at a temperature of 15-25 °C (more preferably about 20 °C) for 10-25 hours (preferably 15-20 hours).

In one embodiment contacting occurs at a temperature of 20-30 °C (preferably about 25 °C). Preferably contacting occurs at a temperature of 20-30 °C (more preferably about 25 °C) for 10-25 hours (preferably 15-20 hours).

A contacting step of a method of the invention may comprise the use of at least 1 µg of protease per mg of clostridial neurotoxin. In one embodiment a contacting step of a method of the invention comprises the use of at least 0.1, 0.2, 0.4, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 µg of protease per mg of clostridial neurotoxin. Preferably a contacting step of a method of the invention comprises the use of at least 2 µg (more preferably at least 4 µg) of protease per mg of clostridial neurotoxin.

In one embodiment a contacting step of a method of the invention comprises the use of ≤20 µg of protease per mg of clostridial neurotoxin. In one embodiment a contacting step of a method of the invention comprises the use of ≤15 µg of protease per mg of clostridial neurotoxin. Preferably a contacting step of a method of the invention comprises the use of ≤10 µg of protease per mg of clostridial neurotoxin. More preferably, a contacting step of a method of the invention comprises the use of ≤7 µg of protease per mg of clostridial neurotoxin

A contacting step of a method of the invention may comprise the use of 0.1-20 µg of protease per mg of clostridial neurotoxin. In one embodiment a method of the invention may comprise the use of 1-10 µg of protease per mg of clostridial neurotoxin, preferably 4-7 µg of protease per mg of clostridial neurotoxin.

A contacting step of a method of the invention may comprise the use of at least 10, 20, 30, 40, 50, 60 or 70 units of enterokinase per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of at least 60 units of enterokinase (more preferably at least 70 units) per mg of clostridial neurotoxin. In some embodiments, a contacting step of a method of the invention may comprise the use of ≤150, ≤140, ≤130, ≤120, ≤110, ≤100, ≤90 units of enterokinase per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of ≤100 units of enterokinase (more preferably ≤90 units) per mg of clostridial neurotoxin. A

contacting step of a method of the invention may comprise the use of 50-110 units of enterokinase per mg of clostridial neurotoxin. In one embodiment a method of the invention may comprise the use of 70-90 units of enterokinase per mg of clostridial neurotoxin, e.g. about 80 units of enterokinase per mg of clostridial neurotoxin.

A contacting step of a method of the invention may comprise the use of at least 0.5, 1, 2, 3, 4, or 5 units of factor Xa per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of at least 3 units of factor Xa (more preferably at least 4 units) per mg of clostridial neurotoxin. In some embodiments, a contacting step of a method of the invention may comprise the use of ≤15, ≤14, ≤13, ≤12, ≤11, ≤10, ≤9, ≤8 or ≤7 units of factor Xa per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of ≤8 units of factor Xa (more preferably ≤7 units) per mg of clostridial neurotoxin. A contacting step of a method of the invention may comprise the use of 0.5-15 units of factor Xa per mg of clostridial neurotoxin. In one embodiment a method of the invention may comprise the use of 1-10 units (preferably 4-7 units) of factor Xa per mg of clostridial neurotoxin, e.g. about 5 or 6 units of factor Xa per mg of clostridial neurotoxin.

A contacting step of a method of the invention may comprise the use of at least 0.02, 0.04, 0.06 or 0.08 units of Lys-C per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of at least 0.04 units of Lys-C per mg of clostridial neurotoxin. In some embodiments, a contacting step of a method of the invention may comprise the use of ≤0.5, ≤0.4 or ≤0.2 units of Lys-C per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of ≤0.2 units of Lys-C per mg of clostridial neurotoxin. A contacting step of a method of the invention may comprise the use of 0.02-0.5 units of Lys-C per mg of clostridial neurotoxin. Preferably, a method of the invention may comprise the use of 0.04-0.2 units of Lys-C per mg of clostridial neurotoxin.

A contacting step of a method of the invention may comprise the use of at least 0.1, 0.2, 0.3 or 0.4 units of trypsin per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of at least 0.4 units of trypsin per mg of clostridial neurotoxin. In some embodiments, a contacting step of a method of the invention may comprise the use of ≤2.5, ≤2.3, ≤2.1, ≤1.9 units of trypsin per mg of clostridial neurotoxin. Preferably, a contacting step of a method of the invention may comprise the use of ≤1.8 units of trypsin per mg of clostridial neurotoxin. A contacting step of a method of the

invention may comprise the use of 0.1-2.5 units of trypsin per mg of clostridial neurotoxin. Preferably, a method of the invention may comprise the use of 0.3-2 units (more preferably 0.4-1.8 units) of trypsin per mg of clostridial neurotoxin.

In one embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/X. A reference BoNT/X sequence is shown as SEQ ID NO: 33. A histidine-tagged version of BoNT/X is presented as SEQ ID NO: 34. A reference nucleotide sequence encoding BoNT/X is shown as SEQ ID NO: 32.

In one embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/A. A reference BoNT/A sequence is shown as SEQ ID NO: 35.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/B. A reference BoNT/B sequence is shown as SEQ ID NO: 36.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/C. A reference BoNT/C_1 sequence is shown as SEQ ID NO: 37.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/D. A reference BoNT/D sequence is shown as SEQ ID NO: 38.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/E. A reference BoNT/E sequence is shown as SEQ ID NO: 39.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/F. A reference BoNT/F sequence is shown as SEQ ID NO: 40.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be BoNT/G. A reference BoNT/G sequence is shown as SEQ ID NO: 41.

In another embodiment the clostridial neurotoxin (e.g. pre-engineering) may be TeNT. A reference TeNT sequence is shown as SEQ ID NO: 42.

As discussed above, clostridial neurotoxins are formed from two polypeptide chains, the heavy chain (H-chain), which has a molecular mass of approximately 100 kDa, and the light chain (L-chain), which has a molecular mass of approximately 50 kDa. The H-chain

comprises a C-terminal targeting component (receptor binding domain or H_C domain) and an N-terminal translocation component (H_N domain).

Examples of light chain reference sequences include:

Botulinum type A neurotoxin: amino acid residues 1-448
Botulinum type B neurotoxin: amino acid residues 1-440
Botulinum type C1 neurotoxin: amino acid residues 1-441
Botulinum type D neurotoxin: amino acid residues 1-445
Botulinum type E neurotoxin: amino acid residues 1-422
Botulinum type F neurotoxin: amino acid residues 1-439
Botulinum type G neurotoxin: amino acid residues 1-441
Tetanus neurotoxin: amino acid residues 1-457

For recently-identified BoNT/X, the L-chain has been reported as corresponding to amino acids 1-439 thereof, with the L-chain boundary potentially varying by approximately 25 amino acids (e.g. 1-414 or 1-464).

The above-identified reference sequences should be considered a guide, as slight variations may occur according to sub-serotypes. By way of example, US 2007/0166332 (hereby incorporated by reference in its entirety) cites slightly different clostridial sequences:

Botulinum type A neurotoxin: amino acid residues M1-K448
Botulinum type B neurotoxin: amino acid residues M1-K441
Botulinum type C1 neurotoxin: amino acid residues M1-K449
Botulinum type D neurotoxin: amino acid residues M1-R445
Botulinum type E neurotoxin: amino acid residues M1-R422
Botulinum type F neurotoxin: amino acid residues M1-K439
Botulinum type G neurotoxin: amino acid residues M1-K446
Tetanus neurotoxin: amino acid residues M1-A457

A Translocation Domain is a molecule that enables translocation of a protease into a target cell such that a functional expression of protease activity occurs within the cytosol of the target cell. Whether any molecule (e.g. a protein or peptide) possesses the requisite translocation function of the present invention may be confirmed by any one of a number of conventional assays.

For example, Shone C. (1987) describes an *in vitro* assay employing liposomes, which are challenged with a test molecule. Presence of the requisite translocation function is confirmed by release from the liposomes of K^+ and/ or labelled NAD, which may be readily monitored [see Shone C. (1987) Eur. J. Biochem; vol. 167(1): pp. 175-180].

A further example is provided by Blaustein R. (1987), which describes a simple *in vitro* assay employing planar phospholipid bilayer membranes. The membranes are challenged with a test molecule and the requisite translocation function is confirmed by an increase in conductance across said membranes [see Blaustein (1987) FEBS Letts; vol. 226, no. 1: pp. 115-120].

Additional methodology to enable assessment of membrane fusion and thus identification of Translocation Domains suitable for use in the present invention are provided by Methods in Enzymology Vol 220 and 221, Membrane Fusion Techniques, Parts A and B, Academic Press 1993.

The present invention also embraces variant translocation domains, so long as the variant domains still demonstrate the requisite translocation activity. By way of example, a variant may have at least 70%, preferably at least 80%, more preferably at least 90%, and most preferably at least 95% or at least 98% amino acid sequence homology with a reference translocation domain. The term fragment, when used in relation to a translocation domain, means a peptide having at least 20, preferably at least 40, more preferably at least 80, and most preferably at least 100 amino acid residues of the reference translocation domain. In the case of a clostridial translocation domain, the fragment preferably has at least 100, preferably at least 150, more preferably at least 200, and most preferably at least 250 amino acid residues of the reference translocation domain (eg. H_N domain). Translocation 'fragments' of the present invention embrace fragments of variant translocation domains based on the reference sequences.

The Translocation Domain is preferably capable of formation of ion-permeable pores in lipid membranes under conditions of low pH. Preferably it has been found to use only those portions of the protein molecule capable of pore-formation within the endosomal membrane.

The Translocation Domain may be obtained from a microbial protein source, in particular from a bacterial or viral protein source. Hence, in one embodiment, the Translocation Domain is a translocating domain of an enzyme, such as a bacterial toxin or viral protein.

It is well documented that certain domains of bacterial toxin molecules are capable of forming such pores. It is also known that certain translocation domains of virally expressed membrane fusion proteins are capable of forming such pores. Such domains may be employed in the present invention.

The Translocation Domain may be of a clostridial origin, such as the H_N domain (or a functional component thereof). H_N means a portion or fragment of the H-chain of a clostridial neurotoxin approximately equivalent to the amino-terminal half of the H-chain, or the domain corresponding to that fragment in the intact H-chain. In one embodiment the H_C function of the H-chain may be removed by deletion of the H_C amino acid sequence (either at the DNA synthesis level, or at the post-synthesis level by nuclease or protease treatment). Alternatively, the H_C function may be inactivated by chemical or biological treatment. Thus, in some embodiments the H-chain may be incapable of binding to the Binding Site on a target cell to which native clostridial neurotoxin (i.e. holotoxin) binds.

Examples of suitable (reference) Translocation Domains include:

Botulinum type A neurotoxin - amino acid residues (449-871)
Botulinum type B neurotoxin - amino acid residues (441-858)
Botulinum type C neurotoxin - amino acid residues (442-866)
Botulinum type D neurotoxin - amino acid residues (446-862)
Botulinum type E neurotoxin - amino acid residues (423-845)
Botulinum type F neurotoxin - amino acid residues (440-864)
Botulinum type G neurotoxin - amino acid residues (442-863)
Tetanus neurotoxin - amino acid residues (458-879)

The above-identified reference sequence should be considered a guide as slight variations may occur according to sub-serotypes. By way of example, US 2007/0166332 (hereby incorporated by reference thereto) cites slightly different clostridial sequences:

Botulinum type A neurotoxin - amino acid residues (A449-K871)
Botulinum type B neurotoxin - amino acid residues (A442-S858)

Botulinum type C neurotoxin - amino acid residues (T450-N866)
Botulinum type D neurotoxin - amino acid residues (D446-N862)
Botulinum type E neurotoxin - amino acid residues (K423-K845)
Botulinum type F neurotoxin - amino acid residues (A440-K864)
Botulinum type G neurotoxin - amino acid residues (S447-S863)
Tetanus neurotoxin - amino acid residues (S458-V879)

In the context of the present invention, a variety of clostridial neurotoxin H_N regions comprising a translocation domain can be useful in aspects of the present invention with the proviso that these active fragments can facilitate the release of a non-cytotoxic protease (e.g. a clostridial L-chain) from intracellular vesicles into the cytoplasm of the target cell and thus participate in executing the overall cellular mechanism whereby a clostridial neurotoxin proteolytically cleaves a substrate. The H_N regions from the heavy chains of clostridial neurotoxins are approximately 410-430 amino acids in length and comprise a translocation domain. Research has shown that the entire length of a H_N region from a clostridial neurotoxin heavy chain is not necessary for the translocating activity of the translocation domain. Thus, aspects of this embodiment can include clostridial neurotoxin H_N regions comprising a translocation domain having a length of, for example, at least 350 amino acids, at least 375 amino acids, at least 400 amino acids and at least 425 amino acids. Other aspects of this embodiment can include clostridial neurotoxin H_N regions comprising translocation domain having a length of, for example, at most 350 amino acids, at most 375 amino acids, at most 400 amino acids and at most 425 amino acids.

For further details on the genetic basis of toxin production in *Clostridium botulinum* and *C. tetani*, we refer to Henderson *et al* (1997) in *The Clostridia: Molecular Biology and Pathogenesis, Academic press*.

The term H_N embraces naturally-occurring neurotoxin H_N portions, and modified H_N portions having amino acid sequences that do not occur in nature and/ or synthetic amino acid residues, so long as the modified H_N portions still demonstrate the above-mentioned translocation function.

Alternatively, the Translocation Domain may be of a non-clostridial origin. Examples of non-clostridial (reference) Translocation Domain origins include, but not be restricted to, the translocation domain of diphtheria toxin [O'Keefe *et al.,* Proc. Natl. Acad. Sci. USA (1992) 89, 6202-6206; Silverman *et al.,* J. Biol. Chem. (1993) 269, 22524-22532; and London, E. (1992)

Biochem. Biophys. Acta., 1112, pp.25-51], the translocation domain of *Pseudomonas* exotoxin type A [Prior *et al.* Biochemistry (1992) 31, 3555-3559], the translocation domains of anthrax toxin [Blanke *et al.* Proc. Natl. Acad. Sci. USA (1996) 93, 8437-8442], a variety of fusogenic or hydrophobic peptides of translocating function [Plank *et al.* J. Biol. Chem. (1994) 269, 12918-12924; and Wagner *et al* (1992) *PNAS, 89, pp.7934-7938*], and amphiphilic peptides [Murata *et al* (1992) *Biochem., 31, pp.1986-1992*]. The Translocation Domain may mirror the Translocation Domain present in a naturally-occurring protein, or may include amino acid variations so long as the variations do not destroy the translocating ability of the Translocation Domain.

Particular examples of viral (reference) Translocation Domains suitable for use in the present invention include certain translocating domains of virally expressed membrane fusion proteins. For example, Wagner *et al.* (1992) and Murata *et al.* (1992) describe the translocation (i.e. membrane fusion and vesiculation) function of a number of fusogenic and amphiphilic peptides derived from the N-terminal region of influenza virus haemagglutinin. Other virally expressed membrane fusion proteins known to have the desired translocating activity are a translocating domain of a fusogenic peptide of Semliki Forest Virus (SFV), a translocating domain of vesicular stomatitis virus (VSV) glycoprotein G, a translocating domain of SER virus F protein and a translocating domain of Foamy virus envelope glycoprotein. Virally encoded Aspike proteins have particular application in the context of the present invention, for example, the E1 protein of SFV and the G protein of the G protein of VSV.

Use of the (reference) Translocation Domains listed in Table (below) includes use of sequence variants thereof. A variant may comprise one or more conservative nucleic acid substitutions and/ or nucleic acid deletions or insertions, with the proviso that the variant possesses the requisite translocating function. A variant may also comprise one or more amino acid substitutions and/ or amino acid deletions or insertions, so long as the variant possesses the requisite translocating function.

Translocation Domain source	**Amino acid residues**	**References**
Diphtheria toxin	194-380	**Silverman *et al.***, 1994, J. Biol. Chem. 269, 22524-22532 **London E.**, 1992, Biochem. Biophys. Acta., 1113, 25-51

Translocation Domain source	Amino acid residues	References
Domain II of pseudomonas exotoxin	405-613	**Prior *et al.***, 1992, Biochemistry 31, 3555-3559 **Kihara & Pastan**, 1994, Bioconj Chem. 5, 532-538
Influenza virus haemagglutinin	GLFGAIAGFIENGWE GMIDGWYG (SEQ ID NO: 73), and Variants thereof	**Plank *et al.***, 1994, J. Biol. Chem. 269, 12918-12924 **Wagner *et al.***, 1992, PNAS, 89, 7934-7938 **Murata *et al.***, 1992, Biochemistry 31, 1986-1992
Semliki Forest virus fusogenic protein	Translocation domain	**Kielian *et al.***, 1996, J Cell Biol. 134(4), 863-872
Vesicular Stomatitis virus glycoprotein G	118-139	**Yao *et al.***, 2003, Virology 310(2), 319-332
SER virus F protein	Translocation domain	**Seth *et al.***, 2003, J Virol 77(11) 6520-6527
Foamy virus envelope glycoprotein	Translocation domain	**Picard-Maureau *et al.***, 2003, J Virol. 77(8), 4722-4730

Examples of clostridial neurotoxin H_C domain reference sequences include:

BoNT/A - N872-L1296
BoNT/B - E859-E1291
BoNT/C1 - N867-E1291
BoNT/D - S863-E1276
BoNT/E - R846-K1252
BoNT/F - K865-E1274
BoNT/G - N864-E1297
TeNT - I880-D1315

For recently-identified BoNT/X, the H_C domain has been reported as corresponding to amino acids 893-1306 thereof, with the domain boundary potentially varying by approximately 25 amino acids (e.g. 868-1306 or 918-1306).

The clostridial neurotoxins described herein may further comprise a translocation facilitating domain. Said domain facilitates delivery of the non-cytotoxic protease into the cytosol of the

target cell and are described, for example, in WO 08/008803 and WO 08/008805, each of which is herein incorporated by reference thereto.

By way of example, suitable translocation facilitating domains include an enveloped virus fusogenic peptide domain, for example, suitable fusogenic peptide domains include influenzavirus fusogenic peptide domain (eg. influenza A virus fusogenic peptide domain of 23 amino acids), alphavirus fusogenic peptide domain (eg. Semliki Forest virus fusogenic peptide domain of 26 amino acids), vesiculovirus fusogenic peptide domain (eg. vesicular stomatitis virus fusogenic peptide domain of 21 amino acids), respirovirus fusogenic peptide domain (eg. Sendai virus fusogenic peptide domain of 25 amino acids), morbiliivirus fusogenic peptide domain (eg. Canine distemper virus fusogenic peptide domain of 25 amino acids), avulavirus fusogenic peptide domain (eg. Newcastle disease virus fusogenic peptide domain of 25 amino acids), henipavirus fusogenic peptide domain (eg. Hendra virus fusogenic peptide domain of 25 amino acids), metapneumovirus fusogenic peptide domain (eg. Human metapneumovirus fusogenic peptide domain of 25 amino acids) or spumavirus fusogenic peptide domain such as simian foamy virus fusogenic peptide domain; or fragments or variants thereof.

By way of further example, a translocation facilitating domain may comprise a clostridial neurotoxin H_{CN} domain or a fragment or variant thereof. In more detail, a clostridial neurotoxin H_{CN} translocation facilitating domain may have a length of at least 200 amino acids, at least 225 amino acids, at least 250 amino acids, at least 275 amino acids. In this regard, a clostridial neurotoxin H_{CN} translocation facilitating domain preferably has a length of at most 200 amino acids, at most 225 amino acids, at most 250 amino acids, or at most 275 amino acids. Specific (reference) examples include:

Botulinum type A neurotoxin - amino acid residues (872-1110)
Botulinum type B neurotoxin - amino acid residues (859-1097)
Botulinum type C neurotoxin - amino acid residues (867-1111)
Botulinum type D neurotoxin - amino acid residues (863-1098)
Botulinum type E neurotoxin - amino acid residues (846-1085)
Botulinum type F neurotoxin - amino acid residues (865-1105)
Botulinum type G neurotoxin - amino acid residues (864-1105)
Tetanus neurotoxin - amino acid residues (880-1127)

The above sequence positions may vary a little according to serotype/ sub-type, and further examples of suitable (reference) clostridial neurotoxin H_{CN} domains include:

Botulinum type A neurotoxin	- amino acid residues (874-1110)
Botulinum type B neurotoxin	- amino acid residues (861-1097)
Botulinum type C neurotoxin	- amino acid residues (869-1111)
Botulinum type D neurotoxin	- amino acid residues (865-1098)
Botulinum type E neurotoxin	- amino acid residues (848-1085)
Botulinum type F neurotoxin	- amino acid residues (867-1105)
Botulinum type G neurotoxin	- amino acid residues (866-1105)
Tetanus neurotoxin	- amino acid residues (882-1127)

Any of the above-described facilitating domains may be combined with any of the previously described translocation domain peptides that are suitable for use in the present invention. Thus, by way of example, a non-clostridial facilitating domain may be combined with non-clostridial translocation domain peptide or with clostridial translocation domain peptide. Alternatively, a clostridial neurotoxin H_{CN} translocation facilitating domain may be combined with a non-clostridial translocation domain peptide. Alternatively, a clostridial neurotoxin H_{CN} facilitating domain may be combined or with a clostridial translocation domain peptide, examples of which include:

Botulinum type A neurotoxin	- amino acid residues (449-1110)
Botulinum type B neurotoxin	- amino acid residues (442-1097)
Botulinum type C neurotoxin	- amino acid residues (450-1111)
Botulinum type D neurotoxin	- amino acid residues (446-1098)
Botulinum type E neurotoxin	- amino acid residues (423-1085)
Botulinum type F neurotoxin	- amino acid residues (440-1105)
Botulinum type G neurotoxin	- amino acid residues (447-1105)
Tetanus neurotoxin	- amino acid residues (458-1127)

In some embodiments the clostridial neurotoxins of the present invention may lack a functional H_C domain of a clostridial neurotoxin. Accordingly, said clostridial neurotoxins are not able to bind rat synaptosomal membranes (via a clostridial H_C component) in binding assays as described in Shone et al. (1985) Eur. J. Biochem. 151, 75-82. In one embodiment, the clostridial neurotoxins preferably lack the last 50 C-terminal amino acids of a clostridial neurotoxin holotoxin. In another embodiment, the clostridial neurotoxins

preferably lack the last 100, preferably the last 150, more preferably the last 200, particularly preferably the last 250, and most preferably the last 300 C-terminal amino acid residues of a clostridial neurotoxin holotoxin. Alternatively, the H_C binding activity may be negated/ reduced by mutagenesis – by way of example, referring to BoNT/A for convenience, modification of one or two amino acid residue mutations (W1266 to L and Y1267 to F) in the ganglioside binding pocket causes the H_C region to lose its receptor binding function. Analogous mutations may be made to non-serotype A clostridial peptide components, e.g. a construct based on botulinum B with mutations (W1262 to L and Y1263 to F) or botulinum E (W1224 to L and Y1225 to F). Other mutations to the active site achieve the same ablation of H_C receptor binding activity, e.g. Y1267S in botulinum type A toxin and the corresponding highly conserved residue in the other clostridial neurotoxins. Details of this and other mutations are described in Rummel et al (2004) (Molecular Microbiol. 51:631-634), which is hereby incorporated by reference thereto.

The H_C peptide of a native clostridial neurotoxin comprises approximately 400-440 amino acid residues, and consists of two functionally distinct domains of approximately 25kDa each, namely the N-terminal region (commonly referred to as the H_{CN} peptide or domain) and the C-terminal region (commonly referred to as the H_{CC} peptide or domain). This fact is confirmed by the following publications, each of which is herein incorporated in its entirety by reference thereto: Umland TC (1997) Nat. Struct. Biol. 4: 788-792; Herreros J (2000) Biochem. J. 347: 199-204; Halpern J (1993) J. Biol. Chem. 268: 15, pp. 11188-11192; Rummel A (2007) PNAS 104: 359-364; Lacey DB (1998) Nat. Struct. Biol. 5: 898-902; Knapp (1998) Am. Cryst. Assoc. Abstract Papers 25: 90; Swaminathan and Eswaramoorthy (2000) Nat. Struct. Biol. 7: 1751-1759; and Rummel A (2004) Mol. Microbiol. 51(3), 631-643. Moreover, it has been well documented that the C-terminal region (H_{CC}), which constitutes the C-terminal 160-200 amino acid residues, is responsible for binding of a clostridial neurotoxin to its natural cell receptors, namely to nerve terminals at the neuromuscular junction - this fact is also confirmed by the above publications. Thus, reference throughout this specification to a clostridial heavy-chain lacking a functional heavy chain H_C peptide (or domain) such that the heavy-chain is incapable of binding to cell surface receptors to which a native clostridial neurotoxin binds means that the clostridial heavy-chain simply lacks a functional H_{CC} peptide. In other words, the H_{CC} peptide region may be either partially or wholly deleted, or otherwise modified (e.g. through conventional chemical or proteolytic treatment) to inactivate its native binding ability for nerve terminals at the neuromuscular junction.

Thus, in one embodiment, a clostridial neurotoxin H_N peptide of the present invention lacks part of a C-terminal peptide portion (H_{CC}) of a clostridial neurotoxin and thus lacks the H_C binding function of native clostridial neurotoxin. By way of example, in one embodiment, the C-terminally extended clostridial H_N peptide lacks the C-terminal 40 amino acid residues, or the C-terminal 60 amino acid residues, or the C-terminal 80 amino acid residues, or the C-terminal 100 amino acid residues, or the C-terminal 120 amino acid residues, or the C-terminal 140 amino acid residues, or the C-terminal 150 amino acid residues, or the C-terminal 160 amino acid residues of a clostridial neurotoxin heavy-chain. In another embodiment, the clostridial H_N peptide of the present invention lacks the entire C-terminal peptide portion (H_{CC}) of a clostridial neurotoxin and thus lacks the H_C binding function of native clostridial neurotoxin. By way of example, in one embodiment, the clostridial H_N peptide lacks the C-terminal 165 amino acid residues, or the C-terminal 170 amino acid residues, or the C-terminal 175 amino acid residues, or the C-terminal 180 amino acid residues, or the C-terminal 185 amino acid residues, or the C-terminal 190 amino acid residues, or the C-terminal 195 amino acid residues of a clostridial neurotoxin heavy-chain. By way of further example, the clostridial H_N peptide of the present invention lacks a clostridial H_{CC} reference sequence selected from the group consisting of:

Botulinum type A neurotoxin - amino acid residues (Y1111-L1296)
Botulinum type B neurotoxin - amino acid residues (Y1098-E1291)
Botulinum type C neurotoxin - amino acid residues (Y1112-E1291)
Botulinum type D neurotoxin - amino acid residues (Y1099-E1276)
Botulinum type E neurotoxin - amino acid residues (Y1086-K1252)
Botulinum type F neurotoxin - amino acid residues (Y1106-E1274)
Botulinum type G neurotoxin - amino acid residues (Y1106-E1297)
Tetanus neurotoxin - amino acid residues (Y1128-D1315).

The above-identified reference sequences should be considered a guide as slight variations may occur according to sub-serotypes.

The present invention is suitable for application to many different varieties of clostridial neurotoxin. Thus, in the context of the present invention, the term “clostridial neurotoxin” embraces toxins produced by *C. botulinum* (botulinum neurotoxin serotypes A, B, C1, D, E, F, G, H, and X), *C. tetani* (tetanus neurotoxin), *C. butyricum* (botulinum neurotoxin serotype E), and *C. baratii* (botulinum neurotoxin serotype F), as well as modified clostridial neurotoxins or derivatives derived from any of the foregoing. The term “clostridial neurotoxin”

also embraces botulinum neurotoxin serotype H. Preferably the clostridial neurotoxin is not BoNT/C1.

Botulinum neurotoxin (BoNT) is produced by *C. botulinum* in the form of a large protein complex, consisting of BoNT itself complexed to a number of accessory proteins. There are at present nine different classes of botulinum neurotoxin, namely: botulinum neurotoxin serotypes A, B, C1, D, E, F, G, H, and X all of which share similar structures and modes of action. Different BoNT serotypes can be distinguished based on inactivation by specific neutralising anti-sera, with such classification by serotype correlating with percentage sequence identity at the amino acid level. BoNT proteins of a given serotype are further divided into different subtypes on the basis of amino acid percentage sequence identity.

BoNTs are absorbed in the gastrointestinal tract, and, after entering the general circulation, bind to the presynaptic membrane of cholinergic nerve terminals and prevent the release of their neurotransmitter acetylcholine. BoNT/B, BoNT/D, BoNT/F and BoNT/G cleave synaptobrevin/vesicle-associated membrane protein (VAMP); BoNT/C1, BoNT/A and BoNT/E cleave the synaptosomal-associated protein of 25 kDa (SNAP-25); and BoNT/C1 cleaves syntaxin. BoNT/X has been found to cleave SNAP-25, VAMP1, VAMP2, VAMP3, VAMP4, VAMP5, Ykt6, and syntaxin 1.

Tetanus toxin is produced in a single serotype by *C. tetani*. *C. butyricum* produces BoNT/E, while *C. baratii* produces BoNT/F.

The term “clostridial neurotoxin” is also intended to embrace modified clostridial neurotoxins and derivatives thereof, including but not limited to those described below. A modified clostridial neurotoxin or derivative may contain one or more amino acids that has been modified as compared to the native (unmodified) form of the clostridial neurotoxin, or may contain one or more inserted amino acids that are not present in the native (unmodified) form of the clostridial neurotoxin. By way of example, a modified clostridial neurotoxin may have modified amino acid sequences in one or more domains relative to the native (unmodified) clostridial neurotoxin sequence. Such modifications may modify functional aspects of the toxin, for example biological activity or persistence. Thus, in one embodiment, the engineered clostridial neurotoxin of the invention is an engineered modified clostridial neurotoxin, or an engineered modified clostridial neurotoxin derivative, or an engineered clostridial neurotoxin derivative.

A modified clostridial neurotoxin may have one or more modifications in the amino acid sequence of the heavy chain (such as a modified H_C domain), wherein said modified heavy chain binds to target nerve cells with a higher or lower affinity than the native (unmodified) clostridial neurotoxin. Such modifications in the H_C domain can include modifying residues in the ganglioside binding site of the H_C domain or in the protein (SV2 or synaptotagmin) binding site that alter binding to the ganglioside receptor and/or the protein receptor of the target nerve cell. Examples of such modified clostridial neurotoxins are described in WO 2006/027207 and WO 2006/114308, both of which are hereby incorporated by reference in their entirety.

A modified clostridial neurotoxin may have one or more modifications in the amino acid sequence of the light chain, for example modifications in the substrate binding or catalytic domain which may alter or modify the SNARE protein specificity of the modified L-chain. Examples of such modified clostridial neurotoxins are described in WO 2010/120766 and US 2011/0318385, both of which are hereby incorporated by reference in their entirety.

A modified clostridial neurotoxin may comprise one or more modifications that increases or decreases the biological activity and/or the biological persistence of the modified clostridial neurotoxin. For example, a modified clostridial neurotoxin may comprise a leucine- or tyrosine-based motif, wherein said motif increases or decreases the biological activity and/or the biological persistence of the modified clostridial neurotoxin. Suitable leucine-based motifs include xDxxxLL (SEQ ID NO: 74), xExxxLL (SEQ ID NO: 75), xExxxIL (SEQ ID NO: 76), and xExxxLM (SEQ ID NO: 77) (wherein x is any amino acid). Suitable tyrosine-based motifs include Y-x-x-Hy (SEQ ID NO: 78) (wherein Hy is a hydrophobic amino acid). Examples of modified clostridial neurotoxins comprising leucine- and tyrosine-based motifs are described in WO 2002/08268, which is hereby incorporated by reference in its entirety.

The term “clostridial neurotoxin” is intended to embrace hybrid and chimeric clostridial neurotoxins. A hybrid clostridial neurotoxin comprises at least a portion of a light chain from one clostridial neurotoxin or subtype thereof, and at least a portion of a heavy chain from another clostridial neurotoxin or clostridial neurotoxin subtype. In one embodiment the hybrid clostridial neurotoxin may contain the entire light chain from one clostridial neurotoxin subtype and the heavy chain from another clostridial neurotoxin subtype. In another embodiment, a chimeric clostridial neurotoxin may contain a portion (e.g. the binding domain) of the heavy chain of one clostridial neurotoxin subtype, with another portion of the heavy chain being from another clostridial neurotoxin subtype. Similarly or alternatively, the

therapeutic element may comprise light chain portions from different clostridial neurotoxins. Such hybrid or chimeric clostridial neurotoxins are useful, for example, as a means of delivering the therapeutic benefits of such clostridial neurotoxins to patients who are immunologically resistant to a given clostridial neurotoxin subtype, to patients who may have a lower than average concentration of receptors to a given clostridial neurotoxin heavy chain binding domain, or to patients who may have a protease-resistant variant of the membrane or vesicle toxin substrate (e.g., SNAP-25, VAMP and syntaxin). Hybrid and chimeric clostridial neurotoxins are described in US 8,071,110, which publication is hereby incorporated by reference in its entirety. Thus, in one embodiment, the engineered clostridial neurotoxin of the invention is an engineered hybrid clostridial neurotoxin, or an engineered chimeric clostridial neurotoxin.

In a particularly preferred embodiment, a clostridial neurotoxin is BoNT/X comprising at least one domain from a non-BoNT/X clostridial neurotoxin (e.g. a BoNT/X hybrid or chimera).

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/X L-chain and a non-BoNT/X H_N and H_C domain;
ii. A BoNT/X H_N domain and a non-BoNT/X L-chain and H_C domain
iii. A BoNT/X H_C domain and a non-BoNT/X L-chain and H_N domain;
iv. A BoNT/X L-chain and H_N domain and a non-BoNT/X H_C domain
v. A BoNT/X L-chain and H_C domain and a non-BoNT/X H_N domain; or
vi. A BoNT/X H_N domain and H_C domain and a non-BoNT/X L-chain.

In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising at least 70% sequence identity to SEQ ID NO: 6. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising at least 80% or 90% sequence identity to SEQ ID NO: 6. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising (more preferably consisting of) SEQ ID NO: 6. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises a polypeptide sequence having at least 70%

sequence identity to SEQ ID NO: 7. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 7. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises (more preferably consists of) a polypeptide sequence shown as SEQ ID NO: 7.

In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/B H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising at least 70% sequence identity to SEQ ID NO: 8. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising at least 80% or 90% sequence identity to SEQ ID NO: 8. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain is encoded by a nucleotide sequence comprising (more preferably consisting of) SEQ ID NO: 8. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 9. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 9. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/X L-chain and H_N domain and a BoNT/A H_C domain comprises (more preferably consists of) a polypeptide sequence shown as SEQ ID NO: 9.

In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/C H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/D H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/E H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/F H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a BoNT/G H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/X L-chain and H_N domain and a TeNT H_C domain.

In one embodiment, a clostridial neurotoxin is BoNT/A comprising at least one domain from a non-BoNT/A clostridial neurotoxin.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/A L-chain and a non-BoNT/A H_N and H_C domain;
ii. A BoNT/A H_N domain and a non-BoNT/A L-chain and H_C domain
iii. A BoNT/A H_C domain and a non-BoNT/A L-chain and H_N domain;
iv. A BoNT/A L-chain and H_N domain and a non-BoNT/A H_C domain
v. A BoNT/A L-chain and H_C domain and a non-BoNT/A H_N domain; or
vi. A BoNT/A H_N domain and H_C domain and a non-BoNT/A L-chain.

In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain is encoded by a nucleotide sequence comprising at least 70% sequence identity to SEQ ID NO: 12. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain is encoded by a nucleotide sequence comprising at least 80% or 90% sequence identity to SEQ ID NO: 12. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain is encoded by a nucleotide sequence comprising (more preferably consisting of) SEQ ID NO: 12. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 13. In one embodiment an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 13. Preferably an engineered clostridial neurotoxin of the invention comprising a BoNT/A L-chain and H_N domain and a BoNT/C1 H_C domain comprises (more preferably consists of) a polypeptide sequence shown as SEQ ID NO: 13.

In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/B H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/D H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention

comprises a BoNT/A L-chain and H_N domain and a BoNT/E H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/F H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/G H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a BoNT/X H_C domain. In one embodiment an engineered clostridial neurotoxin of the invention comprises a BoNT/A L-chain and H_N domain and a TeNT H_C domain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/B L-chain and a non-BoNT/B H_N and H_C domain;
ii. A BoNT/B H_N domain and a non-BoNT/B L-chain and H_C domain
iii. A BoNT/B H_C domain and a non-BoNT/B L-chain and H_N domain;
iv. A BoNT/B L-chain and H_N domain and a non-BoNT/B H_C domain
v. A BoNT/B L-chain and H_C domain and a non-BoNT/B H_N domain; or
vi. A BoNT/B H_N domain and H_C domain and a non-BoNT/B L-chain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/D L-chain and a non-BoNT/XD H_N and H_C domain;
ii. A BoNT/D H_N domain and a non-BoNT/D L-chain and H_C domain
iii. A BoNT/D H_C domain and a non-BoNT/D L-chain and H_N domain;
iv. A BoNT/D L-chain and H_N domain and a non-BoNT/D H_C domain
v. A BoNT/D L-chain and H_C domain and a non-BoNT/D H_N domain; or
vi. A BoNT/D H_N domain and H_C domain and a non-BoNT/D L-chain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/E L-chain and a non-BoNT/E H_N and H_C domain;
ii. A BoNT/E H_N domain and a non-BoNT/E L-chain and H_C domain
iii. A BoNT/E H_C domain and a non-BoNT/E L-chain and H_N domain;
iv. A BoNT/E L-chain and H_N domain and a non-BoNT/E H_C domain
v. A BoNT/E L-chain and H_C domain and a non-BoNT/E H_N domain; or
vi. A BoNT/E H_N domain and H_C domain and a non-BoNT/E L-chain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/F L-chain and a non-BoNT/F H_N and H_C domain;
ii. A BoNT/F H_N domain and a non-BoNT/F L-chain and H_C domain
iii. A BoNT/F H_C domain and a non-BoNT/F L-chain and H_N domain;
iv. A BoNT/F L-chain and H_N domain and a non-BoNT/F H_C domain
v. A BoNT/F L-chain and H_C domain and a non-BoNT/F H_N domain; or
vi. A BoNT/F H_N domain and H_C domain and a non-BoNT/F L-chain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A BoNT/G L-chain and a non-BoNT/G H_N and H_C domain;
ii. A BoNT/G H_N domain and a non-BoNT/G L-chain and H_C domain
iii. A BoNT/G H_C domain and a non-BoNT/G L-chain and H_N domain;
iv. A BoNT/G L-chain and H_N domain and a non-BoNT/G H_C domain
v. A BoNT/G L-chain and H_C domain and a non-BoNT/G H_N domain; or
vi. A BoNT/G H_N domain and H_C domain and a non-BoNT/G L-chain.

For example, in one embodiment a clostridial neurotoxin of the invention (comprising an exogenous activation loop) may comprise:

i. A TeNT L-chain and a non-TeNT H_N and H_C domain;
ii. A TeNT H_N domain and a non-TeNT L-chain and H_C domain
iii. A TeNT H_C domain and a non-TeNT L-chain and H_N domain;
iv. A TeNT L-chain and H_N domain and a non-TeNT H_C domain
v. A TeNT L-chain and H_C domain and a non-TeNT H_N domain; or
vi. A TeNT H_N domain and H_C domain and a non-TeNT L-chain.

The term “clostridial neurotoxin” may also embrace newly discovered botulinum neurotoxin protein family members expressed by non-clostridial microorganisms, such as the Enterococcus encoded toxin which has closest sequence identity to BoNT/X, the *Weissella oryzae* encoded toxin called BoNT/Wo (NCBI Ref Seq: WP_027699549.1), which cleaves VAMP2 at W89-W90, the *Enterococcus faecium* encoded toxin (GenBank: OTO22244.1), which cleaves VAMP2 and SNAP25, and the *Chryseobacterium pipero* encoded toxin (NCBI Ref.Seq: WP_034687872.1).

The term "clostridial neurotoxin" is intended to embrace re-targeted clostridial neurotoxins. In a re-targeted clostridial neurotoxin, the clostridial neurotoxin is modified to include an exogenous ligand known as a Targeting Moiety (TM). The TM is selected to provide binding specificity for a desired target cell, and as part of the re-targeting process the native binding portion of the clostridial neurotoxin (e.g. the H_C domain, or the H_{CC} domain) may be removed. Re-targeting technology is described, for example, in: EP-B-0689459; WO 1994/021300; EP-B-0939818; US 6,461,617; US 7,192,596; WO 1998/007864; EP-B-0826051; US 5,989,545; US 6,395,513; US 6,962,703; WO 1996/033273; EP-B-0996468; US 7,052,702; WO 1999/017806; EP-B-1107794; US 6,632,440; WO 2000/010598; WO 2001/21213; WO 2006/059093; WO 2000/62814; WO 2000/04926; WO 1993/15766; WO 2000/61192; and WO 1999/58571; all of which are hereby incorporated by reference in their entirety. Thus, in one embodiment, the engineered clostridial neurotoxin of the invention is an engineered re-targeted clostridial neurotoxin. The engineered clostridial neurotoxin of the present invention may lack a functional H_C domain of a clostridial neurotoxin and also lack any functionally equivalent TM. Accordingly, said polypeptides lack the natural binding function of a clostridial neurotoxin and are not able to bind rat synaptosomal membranes (via a clostridial H_C component, or via any functionally equivalent TM) in binding assays as described in Shone *et al.* (1985) Eur. J. Biochem. 151, 75-82. In one embodiment, the TM is preferably not a Wheat Germ Agglutinin (WGA) peptide.

In one embodiment an engineered clostridial neurotoxin of the invention may comprise an engineered LH_N polypeptide described herein.

In one embodiment an engineered clostridial neurotoxin may comprise an engineered LH_N polypeptide described herein and a targeting moiety (TM).

Reference engineered LH_N polypeptide sequences are presented herein as SEQ ID NOs: 53-60, however the engineered LH_N polypeptide sequence may have at least 70% sequence identity to any of SEQ ID NOs: 53-60. In one embodiment the engineered LH_N polypeptide sequence may have at least 80% or 90% sequence identity to any of SEQ ID NOs: 53-60. Preferably the engineered LH_N polypeptide sequence comprises (more preferably consists of) any of SEQ ID NOs: 53-60.

In one embodiment a TM may comprise anthrax toxin protective antigen (PA) or a fragment thereof. A reference sequence for PA is shown as SEQ ID NO: 52. In some embodiments PA comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO:

52 or a fragment thereof. In some embodiments PA comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 52 or a fragment thereof. In other embodiments PA comprises (or consists of) a polypeptide sequence shown as SEQ ID NO: 52 or a fragment thereof.

Thus, in one embodiment an engineered clostridial neurotoxin of the present invention may comprise: a clostridial neurotoxin non-cytotoxic protease domain, a clostridial neurotoxin translocation domain (e.g. LH_N of a clostridial neurotoxin), and a TM comprising PA or a fragment thereof. Said engineered clostridial neurotoxin comprises an exogenous activation loop comprising polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

Thus, in one embodiment, an engineered clostridial neurotoxin comprises PA or a fragment thereof and LH_N/A, LH_N/B, LH_N/D, LH_N/E, LH_N/F, LH_N/G, LH_N/X or LH_N/TeNT, wherein the endogenous clostridial neurotoxin activation loop has been replaced with an exogenous activation loop comprising polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

Lys-C is not suitable for use with conventional clostridial neurotoxins comprising LH_N and a PA TM, as Lys-C hydrolses one or more peptide bonds outside of the endogenous activation loop of said clostridial neurotoxin, for example it has been found that Lys-C hydrolyses one or more peptide bonds in the PA TM.

Thus, in some embodiments, an engineered clostridial neurotoxin comprises PA (or a fragment thereof) and:

i. Amino acid residues 1-871 of SEQ ID NO: 35
ii. Amino acid residues 1-858 of SEQ ID NO: 36
iii. Amino acid residues 1-862 of SEQ ID NO: 38
iv. Amino acid residues 1-845 of SEQ ID NO: 39
v. Amino acid residues 1-864 of SEQ ID NO: 40
vi. Amino acid residues 1-863 of SEQ ID NO: 41
vii. Amino acid residues 1-879 of SEQ ID NO: 42 or
viii. Amino acid residues 1-924 of SEQ ID NO: 33;

wherein the endogenous clostridial neurotoxin activation loop has been replaced with an exogenous activation loop comprising polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1).

In one embodiment an engineered clostridial neurotoxin has a polypeptide sequence with at least 70% sequence identity to a polypeptide comprising:

a. SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

In one embodiment an engineered clostridial neurotoxin has a polypeptide sequence with at least 80% or 90% sequence identity to a polypeptide comprising:

a. SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

Preferably an engineered clostridial neurotoxin has a polypeptide sequence comprising:

a. SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

In one embodiment an engineered clostridial neurotoxin has a polypeptide sequence with at least 70% sequence identity to a polypeptide comprising:

a. a fragment of SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

In one embodiment an engineered clostridial neurotoxin has a polypeptide sequence with at least 80% or 90% sequence identity to a polypeptide comprising:

a. a fragment of SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

Preferably an engineered clostridial neurotoxin has a polypeptide sequence comprising:

a. a fragment of SEQ ID NO: 52; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

In one embodiment a PA fragment may be PAd1, which is located at residues 1-258 of SEQ ID NO: 52. In one embodiment a PA fragment may be PAd2, which is located at residues 259-487 of SEQ ID NO: 52. In one embodiment a PA fragment may be PAd3, which is located at residues 488-594 of SEQ ID NO: 52. In one embodiment a PA fragment may be PAd4, which is located at residues 595-735 of SEQ ID NO: 52. In other embodiments a PA fragment may contain any combination of PAd1, PAd2, PAd3 or PAd4.

Full-length 83 kDa PA (PA83) may be proteolytically processed by a furin or other furin-like protease thus removing an N-terminal fragment (PA20). The 63 kDa processed form, known as PA63 is an oligomerisable form of PA.

In one embodiment the PA fragment may comprise (or consist of) one or more of PA63, PAd3-d4, PAd2-d4, and PAd4.

In one embodiment the PA fragment may be a C-terminal receptor-binding domain of PA or a PA fragment (or variant) thereof that retains binding activity to ANTXR2 or a nociceptor neuron binding protein.

The present invention also embraces clostridial neurotoxins that have a non-native protease cleavage site. In such clostridial neurotoxins, the native protease cleavage site (also known as the activation site, as described above) is modified or replaced with a protease cleavage site that is not native to that clostridial neurotoxin (i.e. an exogenous cleavage site). Such a site will require an exogenous protease for cleavage, which allows for improved control over the timing and location of cleavage events. Non-native protease cleavage sites that may be employed in clostridial neurotoxins include:

TEV(Tobacco Etch virus)	(ENLYFQ↓G) (SEQ ID NO: 79)
Thrombin	(LVPR↓GS) (SEQ ID NO: 80)
PreScission	(LEVLFQ↓GP) (SEQ ID NO: 81).

Additional protease cleavage sites include recognition sequences that are cleaved by a non-cytotoxic protease, for example by the light chain of a clostridial neurotoxin. These include the SNARE (e.g. SNAP-25, syntaxin, VAMP) protein recognition sequences that are cleaved by non-cytotoxic proteases such as the light chain of a clostridial neurotoxin. Clostridial neurotoxins comprising non-native protease cleavage sites are described in US 7,132,259, EP 1206554-B2 and US 2007/0166332, all of which are hereby incorporated by reference in their entirety. Also embraced by the term protease cleavage site is an intein, which is a self-

cleaving sequence. The self-splicing reaction is controllable, for example by varying the concentration of reducing agent present.

The present invention also embraces clostridial neurotoxins comprising a "destructive cleavage site". In said clostridial neurotoxins, a non-native protease cleavage site is incorporated into the clostridial neurotoxin, at a location chosen such that cleavage at said site will decrease the activity of, or inactivate, the clostridial neurotoxin. The destructive protease cleavage site can be susceptible to cleavage by a local protease, in the event that the clostridial neurotoxin, following administration, migrates to a non-target location. Suitable non-native protease cleavage sites include those described above. Clostridial neurotoxins comprising a destructive cleavage site are described in WO 2010/094905 and WO 2002/044199, both of which are hereby incorporated by reference in their entirety.

The engineered clostridial neurotoxins of the present invention, especially the light chain component thereof, may be PEGylated – this may help to increase stability, for example duration of action of the light chain component. PEGylation is particularly preferred when the light chain comprises a BoNT/A, B or C1 protease. PEGylation preferably includes the addition of PEG to the N-terminus of the light chain component. By way of example, the N-terminus of a light chain may be extended with one or more amino acid (e.g. cysteine) residues, which may be the same or different. One or more of said amino acid residues may have its own PEG molecule attached (e.g. covalently attached) thereto. An example of this technology is described in WO2007/104567, which is hereby incorporated by reference in its entirety.

The engineered clostridial neurotoxins of the present invention may be free from the complexing proteins that are present in a naturally occurring clostridial neurotoxin complex.

The engineered clostridial neurotoxins of the present invention can be produced using recombinant nucleic acid technologies. Thus, in one embodiment, an engineered clostridial neurotoxin (as described above) is a recombinant engineered clostridial neurotoxin.

In another aspect, the present invention provides a nucleic acid (for example, a DNA) comprising a nucleic acid sequence encoding an engineered clostridial neurotoxin as described above. In one embodiment, the nucleic acid sequence is prepared as part of a DNA vector comprising a promoter and a terminator.

In a preferred embodiment, the vector has a promoter selected from:

Promoter	Induction Agent	Typical Induction Condition
Tac (hybrid)	IPTG	0.2 mM (0.05-2.0mM)
AraBAD	L-arabinose	0.2% (0.002-0.4%)
T7-lac operator	IPTG	0.2 mM (0.05-2.0mM)

In another preferred embodiment, the vector has a promoter selected from:

Promoter	Induction Agent	Typical Induction Condition
Tac (hybrid)	IPTG	0.2 mM (0.05-2.0mM)
AraBAD	L-arabinose	0.2% (0.002-0.4%)
T7-lac operator	IPTG	0.2 mM (0.05-2.0mM)
T5-lac operator	IPTG	0.2 mM (0.05-2.0mM)

The nucleic acid molecules of the invention may be made using any suitable process known in the art. Thus, the nucleic acid molecules may be made using chemical synthesis techniques. Alternatively, the nucleic acid molecules of the invention may be made using molecular biology techniques.

The DNA construct of the present invention is preferably designed *in silico*, and then synthesised by conventional DNA synthesis techniques.

The above-mentioned nucleic acid sequence information is optionally modified for codon-biasing according to the ultimate host cell (e.g. *E. coli*) expression system that is to be employed.

In one aspect the present invention provides a nucleotide sequence encoding an engineered clostridial neurotoxin of the present invention. The nucleotide sequence comprises a sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12. In one embodiment the nucleotide sequence comprises a sequence having at least 80% or 90% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12. Preferably the nucleotide sequence comprises (more preferably consists of) SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12.

The nucleotide sequence of the invention encodes a polypeptide comprising SEQ ID NO: 1.

The terms "nucleotide sequence" and "nucleic acid" are used synonymously herein. Preferably the nucleotide sequence is a DNA sequence.

The invention provides a method of producing a single-chain (engineered) clostridial neurotoxin protein having a light chain and a heavy chain, the method comprising expressing a nucleic acid described herein in a suitable host cell, lysing the host cell to provide a host cell homogenate containing the single-chain (engineered) clostridial neurotoxin protein, and isolating the single-chain (engineered) clostridial neurotoxin protein. In one aspect, the present invention provides a method for proteolytically processing an (engineered) clostridial neurotoxin of the present invention into a corresponding di-chain clostridial neurotoxin, the method comprising contacting the (engineered) clostridial neurotoxin with a protease (preferably an endopeptidase, such as enterokinase or factor Xa) thereby producing a di-chain clostridial neurotoxin (e.g. wherein the light chain and heavy chain are joined together by a disulphide bond).

The present invention therefore provides a di-chain clostridial neurotoxin obtainable by a method of the invention.

The term "obtainable" as used herein also encompasses the term "obtained". In one embodiment the term "obtainable" means obtained.

A clostridial neurotoxin of the present invention suitably finds utility in medicine or in cosmetics. In use, the clostridial neurotoxin is preferably in a di-chain form.

The (engineered) clostridial neurotoxins of the invention may be used to prevent or treat certain medical or cosmetic diseases and conditions. Thus, in a further aspect, the present invention provides an (engineered) clostridial neurotoxin as described above, for use in medicine.

In a related aspect, the present invention provides an (engineered) clostridial neurotoxin as described above, for use in the prevention or treatment of a disease or condition selected from: a condition associated with unwanted immune secretion, strabismus, blepharospasm, squint, dystonia (e.g. spasmodic dystonia, oromandibular dystonia, focal dystonia, tardive dystonia, laryngeal dystonia, limb dystonia, cervical dystonia), torticollis (e.g. spasmodic

torticollis), beauty therapy (cosmetic) applications benefiting from cell/muscle incapacitation (via SNARE down-regulation or inactivation), neuromuscular disorder or condition of ocular motility (e.g. concomitant strabismus, vertical strabismus, lateral rectus palsy, nystagmus, dysthyroid myopathy), writer's cramp, blepharospasm, bruxism, Wilson's disease, tremor, tics, segmental myoclonus, spasms, spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, animus, back spasm, charley horse, tension headaches, levator pelvic syndrome, spina bifida, tardive dyskinesia, Parkinson's disease, stuttering, hemifacial spasm, eyelid disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure, achalasia, dysphagia, lacrimation, hyperhydrosis, excessive salivation, excessive gastrointestinal secretions, muscle pain (e.g. pain from muscle spasms), headache pain (e.g. tension headache), brow furrows, skin wrinkles, cancer, uterine disorders, uro-genital disorders, urogenital-neurological disorders, chronic neurogenic inflammation, and a smooth muscle disorder.

Where an (engineered) clostridial neurotoxin of the invention comprises a BoNT/X sequence (or portion thereof) said clostridial neurotoxin may be able to target other types of secretory cells other than neurons, due to its ability to cleave VAMP4, VAMP5 and/or Ykt6. In some embodiments, the secretory cell targeted is a secretory immune cell. A “secretory immune cell” as used herein, refers to immune cells that secrets cytokines, chemokines, or antibodies. Such secretory immune cells may be innate immune cells including, without limitation, natural killer cells, mast cells, eosinophils, basophils, macrophages, neutrophils, and dendritic cells. Secretory immune cells that secret antibodies (e.g. white blood cells) may also be targeted by the clostridial neurotoxins of the present disclosure. Non-limiting examples of antibody secreting cells include, without limitation, plasma B cells, plasmocytes, plasmacytes, and effector B cells. In some embodiments, the clostridial neurotoxin may modulate an immune response. Thus, further contemplated herein are therapeutic use of a clostridial neurotoxin of the invention to treat a condition associated with unwanted secretion, preferably unwanted immune secretion. Conditions associated with unwanted immune secretion include, without limitation: inflammation, psoriasis, allergy, haemophagocytic lymphohistiocytosis, and alcoholic pancreatic disease.

In one aspect, the invention provides a pharmaceutical composition comprising an (engineered) clostridial neurotoxin or a di-chain clostridial neurotoxin of the invention and a pharmaceutically acceptable carrier, excipient, adjuvant, propellant and/or salt.

The (engineered) clostridial neurotoxins of the present invention may be formulated for oral, parenteral, continuous infusion, inhalation or topical application. Compositions suitable for injection may be in the form of solutions, suspensions or emulsions, or dry powders which are dissolved or suspended in a suitable vehicle prior to use.

In the case of an (engineered) clostridial neurotoxin that is to be delivered locally, the (engineered) clostridial neurotoxin may be formulated as a cream (e.g. for topical application), or for sub-dermal injection.

Local delivery means may include an aerosol, or other spray (e.g. a nebuliser). In this regard, an aerosol formulation of an (engineered) clostridial neurotoxin enables delivery to the lungs and/or other nasal and/or bronchial or airway passages.

(Engineered) clostridial neurotoxins of the invention may be administered to a patient by intrathecal or epidural injection in the spinal column at the level of the spinal segment involved in the innervation of an affected organ.

A preferred route of administration is via laproscopic and/ or localised, particularly intramuscular, injection.

The dosage ranges for administration of the (engineered) clostridial neurotoxins of the present invention are those to produce the desired therapeutic effect. It will be appreciated that the dosage range required depends on the precise nature of the (engineered) clostridial neurotoxin or composition, the route of administration, the nature of the formulation, the age of the patient, the nature, extent or severity of the patient's condition, contraindications, if any, and the judgement of the attending physician. Variations in these dosage levels can be adjusted using standard empirical routines for optimisation.

Suitable daily dosages (per kg weight of patient) are in the range 0.0001-1 ng/kg, preferably 0.0001-0.5 ng/kg, more preferably 0.002-0.5 ng/kg, and particularly preferably 0.004-0.5 ng/kg. The unit dosage can vary from less than 1 picogram to 30ng, but typically will be in the region of 0.01 to 1 ng per dose, which may be administered daily or preferably less frequently, such as weekly or six monthly.

A particularly preferred dosing regimen is based on 0.05 ng of (engineered) clostridial neurotoxin as the 1X dose. In this regard, preferred dosages are in the range 1X–100X (i.e. 0.05-5 ng).

Fluid dosage forms are typically prepared utilising the (engineered) clostridial neurotoxin and a pyrogen-free sterile vehicle. The (engineered) clostridial neurotoxin, depending on the vehicle and concentration used, can be either dissolved or suspended in the vehicle. In preparing solutions the (engineered) clostridial neurotoxin can be dissolved in the vehicle, the solution being made isotonic if necessary by addition of sodium chloride and sterilised by filtration through a sterile filter using aseptic techniques before filling into suitable sterile vials or ampoules and sealing. Alternatively, if solution stability is adequate, the solution in its sealed containers may be sterilised by autoclaving. Advantageously additives such as buffering, solubilising, stabilising, preservative or bactericidal, suspending or emulsifying agents and or local anaesthetic agents may be dissolved in the vehicle.

Dry powders, which are dissolved or suspended in a suitable vehicle prior to use, may be prepared by filling pre-sterilised ingredients into a sterile container using aseptic technique in a sterile area. Alternatively the ingredients may be dissolved into suitable containers using aseptic technique in a sterile area. The product is then freeze dried and the containers are sealed aseptically.

Parenteral suspensions, suitable for intramuscular, subcutaneous or intradermal injection, are prepared in substantially the same manner, except that the sterile components are suspended in the sterile vehicle, instead of being dissolved and sterilisation cannot be accomplished by filtration. The components may be isolated in a sterile state or alternatively it may be sterilised after isolation, e.g. by gamma irradiation.

Advantageously, a suspending agent for example polyvinylpyrrolidone is included in the composition(s) to facilitate uniform distribution of the components.

Administration in accordance with the present invention may take advantage of a variety of delivery technologies including microparticle encapsulation, viral delivery systems or high-pressure aerosol impingement.

Embodiments related to the various methods of the invention are intended to be applied equally to other methods, the clostridial neurotoxins, e.g. engineered clostridial neurotoxins

(whether in single-chain or di-chain forms), uses or pharmaceutical compositions, and *vice versa*.

SEQUENCE HOMOLOGY

Any of a variety of sequence alignment methods can be used to determine percent identity, including, without limitation, global methods, local methods and hybrid methods, such as, e.g., segment approach methods. Protocols to determine percent identity are routine procedures within the scope of one skilled in the art. Global methods align sequences from the beginning to the end of the molecule and determine the best alignment by adding up scores of individual residue pairs and by imposing gap penalties. Non-limiting methods include, e.g., CLUSTAL W, see, e.g., Julie D. Thompson et al., CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position- Specific Gap Penalties and Weight Matrix Choice, 22(22) Nucleic Acids Research 4673-4680 (1994); and iterative refinement, see, e.g., Osamu Gotoh, Significant Improvement in Accuracy of Multiple Protein. Sequence Alignments by Iterative Refinement as Assessed by Reference to Structural Alignments, 264(4) J. Mol. Biol. 823-838 (1996). Local methods align sequences by identifying one or more conserved motifs shared by all of the input sequences. Non-limiting methods include, e.g., Match-box, see, e.g., Eric Depiereux and Ernest Feytmans, Match-Box: A Fundamentally New Algorithm for the Simultaneous Alignment of Several Protein Sequences, 8(5) CABIOS 501 -509 (1992); Gibbs sampling, see, e.g., C. E. Lawrence et al., Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment, 262(5131) Science 208-214 (1993); Align-M, see, e.g., Ivo Van Walle et al., Align-M - A New Algorithm for Multiple Alignment of Highly Divergent Sequences, 20(9) Bioinformatics:1428-1435 (2004).

Thus, percent sequence identity is determined by conventional methods. See, for example, Altschul et al., Bull. Math. Bio. 48: 603-16, 1986 and Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-19, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix of Henikoff and Henikoff (ibid.) as shown below (amino acids are indicated by the standard one-letter codes).

The "percent sequence identity" between two or more nucleic acid or amino acid sequences is a function of the number of identical positions shared by the sequences. Thus, % identity may be calculated as the number of identical nucleotides / amino acids divided by the total number of nucleotides / amino acids, multiplied by 100. Calculations of % sequence identity

may also take into account the number of gaps, and the length of each gap that needs to be introduced to optimize alignment of two or more sequences. Sequence comparisons and the determination of percent identity between two or more sequences can be carried out using specific mathematical algorithms, such as BLAST, which will be familiar to a skilled person.

5

ALIGNMENT SCORES FOR DETERMINING SEQUENCE IDENTITY

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

The percent identity is then calculated as:

30

$$\frac{\text{Total number of identical matches}}{[\text{length of the longer sequence plus the number of gaps introduced into the longer sequence in order to align the two sequences}]} \times 100$$

35

Substantially homologous polypeptides are characterized as having one or more amino acid substitutions, deletions or additions. These changes are preferably of a minor nature, that is conservative amino acid substitutions (see below) and other substitutions that do not significantly affect the folding or activity of the polypeptide; small deletions, typically of one to about 30 amino acids; and small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues, or an affinity tag.

CONSERVATIVE AMINO ACID SUBSTITUTIONS

Basic:	arginine
	lysine
	histidine
Acidic:	glutamic acid
	aspartic acid
Polar:	glutamine
	asparagine
Hydrophobic:	leucine
	isoleucine
	valine
Aromatic:	phenylalanine
	tryptophan
	tyrosine
Small:	glycine
	alanine
	serine
	threonine
	methionine

In addition to the 20 standard amino acids, non-standard amino acids (such as 4-hydroxyproline, 6-N-methyl lysine, 2-aminoisobutyric acid, isovaline and α -methyl serine) may be substituted for amino acid residues of the polypeptides of the present invention. A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural amino acids may be substituted for polypeptide amino acid residues. The polypeptides of the present invention can also comprise non-naturally occurring amino acid residues.

Non-naturally occurring amino acids include, without limitation, trans-3-methylproline, 2,4-methano-proline, cis-4-hydroxyproline, trans-4-hydroxy-proline, N-methylglycine, allo-threonine, methyl-threonine, hydroxy-ethylcysteine, hydroxyethylhomo-cysteine, nitro-glutamine, homoglutamine, pipecolic acid, tert-leucine, norvaline, 2-azaphenylalanine, 3-azaphenyl-alanine, 4-azaphenyl-alanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an in vitro system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell free system comprising an E. coli S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., J. Am. Chem. Soc. 113:2722, 1991; Ellman et al., Methods Enzymol. 202:301, 1991; Chung et al., Science 259:806-9, 1993; and Chung et al., Proc. Natl. Acad. Sci. USA 90:10145-9, 1993). In a second method, translation is carried out in Xenopus oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., J. Biol. Chem. 271:19991-8, 1996). Within a third method, E. coli cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the polypeptide in place of its natural counterpart. See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by in vitro chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for amino acid residues of polypeptides of the present invention.

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244: 1081-5, 1989). Sites of biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See,

for example, de Vos et al., Science 255:306-12, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of essential amino acids can also be inferred from analysis of homologies with related components (e.g. the translocation or protease components) of the polypeptides of the present invention.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-7, 1988) or Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-6, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30:10832-7, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 20 ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide the skilled person with a general dictionary of many of the terms used in this disclosure.

This disclosure is not limited by the exemplary methods and materials disclosed herein, and any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of this disclosure. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, any nucleic acid sequences are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

The headings provided herein are not limitations of the various aspects or embodiments of this disclosure.

Amino acids are referred to herein using the name of the amino acid, the three letter abbreviation or the single letter abbreviation. The term “protein", as used herein, includes proteins, polypeptides, and peptides. As used herein, the term “amino acid sequence” is

synonymous with the term “polypeptide” and/or the term “protein”. In some instances, the term “amino acid sequence” is synonymous with the term “peptide”. In some instances, the term “amino acid sequence” is synonymous with the term “enzyme”. The terms "protein" and "polypeptide" are used interchangeably herein. In the present disclosure and claims, the conventional one-letter and three-letter codes for amino acid residues may be used. The 3-letter code for amino acids as defined in conformity with the IUPACIUB Joint Commission on Biochemical Nomenclature (JCBN). It is also understood that a polypeptide may be coded for by more than one nucleotide sequence due to the degeneracy of the genetic code.

Other definitions of terms may appear throughout the specification. Before the exemplary embodiments are described in more detail, it is to be understood that this disclosure is not limited to particular embodiments described, and as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be defined only by the appended claims.

Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within this disclosure. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also encompassed within this disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in this disclosure.

It must be noted that as used herein and in the appended claims, the singular forms “a”, “an”, and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a clostridial neurotoxin” includes a plurality of such candidate agents and reference to “the clostridial neurotoxin” includes reference to one or more clostridial neurotoxins and equivalents thereof known to those skilled in the art, and so forth.

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will now be described, by way of example only, with reference to the following Figures and Examples.

Figure 1 shows a comparison of the protein sequence of the activation loop for all BoNT serotypes and a tetanus toxin with two flanking cysteines forming a disulphide bridge connecting the light and heavy chain of a toxin molecule. Factor Xa cleavage site (IDGR) in BoNT/C1 and BoNT/CD underlined.

Figure 2 shows that all four proteases tested: trypsin (TrypZean), Lys-C, factor Xa (FXa) and enterokinase (EK) have the ability to cleave BoNT/C1 activation loop and create a di-chain molecule as compared to C - protease untreated control. (**A**,**B**) BoNT/C1(0) (SEQ ID NO: 15) treated with Factor Xa (FXa), enterokinase (EK) and trypsin (1-16 hour time course) as indicated tested by SDS-PAGE in non-reducing (**A**) and reducing conditions (**B**). Similarly, proteolytic digest by Lys-C creates di-chain molecules of BoNT/C1 (**C**). Both trypsin and Lys-C show off-target cleavage within the heavy chain of BoNT/C1. 1-Benchmark (5ul); 2-Control sample (-LysC) – DTT; 3-Control sample (-LysC) +DTT; 4-Activated – DTT; 5-Activated +DTT.

Figure 3 (**A**) shows BoNT/X cleavage by Lys-C. Samples were tested in non-reduced and reduced (+DTT) conditions. 1-Benchmark (5ul); 2-No protease control; 3-LysC 0.125 µg/ml; 4-LysC 0.25 µg/ml; 5-LysC 0.5 µg/ml; 6-LysC 1 µg/ml; 7-LysC 2 µg/ml; 8-LysC 4 µg/ml; 9-No protease control +DTT; 10-LysC 0.125 µg/ml +DTT; 11-LysC 0.25 µg/ml +DTT; 12-LysC 0.5 µg/ml +DTT; 13-LysC 1 µg/ml +DTT; 14-LysC 2 µg/ml +DTT; 15-LysC 4 µg/ml +DTT; and 16-Benchmark (5ul). (**B**) shows BoNT/X cleavage by Trypsin (TrypZean), Factor Xa and Enterokinase. Samples were tested in non-reduced and reduced (+DTT) conditions. 1-Benchmark (5ul); 2-No protease control; 3-TrypZean 0.125 µg/ml; 4-TrypZean 0.25 µg/ml; 5-No protease control; 6-TrypZean 1 µg/ml; 7-TrypZean 2 µg/ml; 8-TrypZean 4 µg/ml; 9-Factor Xa 5 µg/ml; 10-Enterokinase 0.01 µg/ml; 11-No protease control +DTT; 12-TrypZean 0.125 µg/ml +DTT; 13-TrypZean 0.25 µg/ml +DTT; 14-No protease control +DTT; 15-TrypZean 1 µg/ml +DTT; 16-TrypZean 2 µg/ml +DTT; 17-TrypZean 4 µg/ml +DTT; 18-Factor Xa 5 µg/ml; 19-Enterokinase 0.01 µg/ml; and 20-Benchmark (5ul).

Figure 4 shows engineered BoNT/X (SEQ ID NO: 5) treated with indicated proteases and successful BoNT di-chain formation evident by comparison between non-reduced (-DTT) and

reduced (+DTT) conditions. Lanes 4-7 show in-process samples. Lanes 8-11 show final samples after last polishing step. 1-capture HisHP load; 2-control -DTT –protease; 3-control +DTT – protease; 4-activated -DTT +EK; 5-activated +DTT +EK; 6-activated -DTT +FXa; 7-activated +DTT +FXa; 8-final -DTT EK active.; 9-final +DTT EK activ.; 10-final -DTT FXa activ.; 11-final +DTT FXa activ.

Figure 5 (**A**) shows BoNT/E tested with 10 µg/ml endoproteinase Lys-C from *Lysobacter enzymogenes* ("Lys-C") and *Pseudomonas aeruginosa* ("rLys-C") for 2 h @ 37 °C. Samples were tested in non-reduced and reduced (+TCEP) conditions. (**B**) shows BoNT/E treated with indicated amounts of Trypsin for 7 h at 20°C. (C) – sample stored at -20°C. (T) – no protease control at 20°C. Samples were tested in reduced (+DTT) conditions.

Figure 6 shows engineered BoNT/E (SEQ ID NO: 11) treated with indicated proteases and successful BoNT di-chain formation evident by comparison between non-reduced (-DTT) and reduced (+DTT) conditions. 1-Benchmark (5ul); 2-Control sample (– EK) – DTT; 3-Control sample (– EK) +DTT; 4-Activated (+ EK) – DTT; 5-Activated (+ EK) +DTT; 6-Benchmark (5ul); 7-Control sample (– FXa) – DTT; 8-Control sample (– FXa) +DTT; 9-Activated (+ FXa) – DTT; 10-Activated (+ FXa) +DTT.

Figure 7 (**A**) shows engineered BoNT/A1C1 (SEQ ID NO: 13) treated with Factor Xa protease and successful BoNT di-chain formation evident by comparison between non-reduced (-DTT) and reduced (+DTT) conditions. 1-BenchMark ladder; 2-Control (-FXa – DTT); 3-Control (-FXa +DTT); 4-Activated (+FXa -DTT); 5-Activated (+FXa +DTT). (**B**) shows cleavage of BoNT/A1 with FXa or EK after 2 hours compared to a positive control (di-chain BoNT/A1).

Figure 8 shows BoNT/A1C1 and BoNT/C1 dose dependent inhibition of glutamate release from primary rat neurons.

Figure 9 shows intact mass analysis of non-reduced engineered BoNT/E (SEQ ID NO: 11) activated by enterokinase with an indicated mass of 143853 Da.

Figure 10 shows intact mass analysis of reduced engineered BoNT/E (SEQ ID NO: 11) activated by enterokinase with an indicated mass of 47518Da and 96338Da.

Figure 11 shows intact mass analysis of non-reduced engineered BoNT/E (SEQ ID NO: 11) activated by factor Xa with an indicated mass of 143850 Da.

Figure 12 shows intact mass analysis of reduced engineered BoNT/E (SEQ ID NO: 11) activated by factor Xa with an indicated mass of 47518Da and 96335Da.

Figure 13 shows LH_N/A1 containing an EK cleavage site inserted into the activation loop (SEQ ID NO: 44) and treated with EK compared to a native A1 loop (SEQ ID NO: 46) treated with Lys-C. 1-Benchmark (5ul); 2-empty; 3-SEQ ID NO: 44 +EK –DTT; 4-SEQ ID NO: 44 +EK +DTT; 5-SEQ ID NO: 44 -EK –DTT; 6-SEQ ID NO: 44 -EK +DTT; 7-Benchmark; 8-SEQ ID NO: 46 -LysC –DTT; 9-SEQ ID NO: 46 -LysC –DTT; 10-SEQ ID NO: 46 +LysC –DTT; 11-SEQ ID NO: 46 -LysC +DTT; 12-SEQ ID NO: 46 -LysC +DTT; and 13-SEQ ID NO: 46 +LysC +DTT.

Figure 14 shows activation of engineered BoNT/XA (SEQ ID NO: 7) with FXa.

Figure 15 shows activation of engineered BoNT/XB (SEQ ID NO: 9) with FXa.

SEQUENCE LISTING

Where an initial Met amino acid residue or a corresponding initial codon is indicated in any of the following SEQ ID NOs, said residue/codon is optional.

SEQ ID NO:1 (C1 Activation Loop Consensus Sequence)
Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys

SEQ ID NO: 2 (C1 Activation Loop)
CHKAIDGRSLYNKTLDC

SEQ ID NO: 3 (C1 Activation Loop Variant)
CHKAIEGRSLYNKTLDC

SEQ ID NO: 4 (Nucleotide Sequence of BoNT/X with a C1 Activation Loop)
ATGAAACTGGAAATCAACAAATTCAACTACAACGATCCGATCGATGGCATTAATGTTATTACCATGCG
TCCGCCTCGTCATAGCGATAAAATCAATAAAGGTAAAGGTCCGTTCAAAGCCTTTCAGGTGATTAAAA
ACATTTGGATTGTGCCGGAACGCTACAACTTTACCAATAATACCAACGATCTGAACATTCCGAGCGAA
CCGATTATGGAAGCAGATGCCATTTATAACCCGAACTATCTGAATACCCCGAGCGAAAAAGATGAATT
TCTGCAGGGTGTTATCAAAGTGCTGGAACGCATTAAAAGCAAACCGGAAGGTGAAAAACTGCTGGAAC
TGATTAGCAGCAGCATTCCGCTGCCGCTGGTTAGCAATGGTGCACTGACCCTGAGCGATAATGAAACC
ATTGCATATCAAGAGAACAACAACATTGTGAGCAATCTGCAGGCAAACCTGGTTATTTATGGTCCGGG
TCCTGATATTGCAAATAATGCAACCTATGGTCTGTATAGCACCCCGATTAGTAATGGTGAAGGTACAC
TGAGCGAAGTTAGCTTTAGCCCGTTTTATCTGAAACCGTTTGATGAAAGCTATGGCAATTATCGTAGC
CTGGTGAATATCGTGAACAAATTCGTGAAACGTGAATTTGCACCTGATCCGGCAAGCACCCTGATGCA
TGAACTGGTTCATGTTACCCATAATCTGTATGGTATTAGCAACCGCAACTTCTACTATAACTTTGACA
CCGGCAAAATTGAAACCAGCCGTCAGCAGAATAGCCTGATTTTTGAAGAACTGCTGACCTTTGGTGGC

ATTGATAGCAAAGCAATTAGCAGCCTGATCATCAAGAAAATTATCGAAACCGCCAAGAACAACTATAC
CACGCTGATTAGCGAACGCCTGAATACCGTTACCGTTGAAAATGATCTGCTGAAATATATCAAAAACA
AAATCCCGGTTCAGGGTCGTCTGGGTAACTTTAAACTGGATACCGCAGAATTCGAGAAAAAGCTGAAT
ACCATTCTGTTTGTGCTGAACGAAAGCAATCTGGCACAGCGTTTTAGCATTCTGGTTCGTAAACATTA
CCTGAAAGAACGTCCGATTGATCCGATTTATGTGAACATTCTGGATGACAATAGCTACAGCACCCTGG
AAGGTTTTAACATTAGCAGTCAGGGTAGCAATGATTTCCAAGGTCAGCTGCTGGAAAGCAGCTATTTT
GAAAAAATTGAAAGCAATGCCCTGCGTGCCTTTATCAAAATTTGTCATAAAGCCATTGATGGTCGCAG
CCTGTATAACAAAACCCTGGATTGTATTGAGGTGGAAAACAAAGACCTGTTTCTGATTAGCAACAAAG
ATAGCCTGAACGATATTAACCTGAGCGAAGAAAAAATCAAACCGGAAACCACCGTGTTCTTCAAAGAT
AAACTGCCTCCGCAGGATATTACGCTGAGCAATTATGATTTTACCGAAGCCAATAGCATTCCGAGCAT
TAGCCAGCAGAACATTCTGGAACGTAATGAAGAACTGTATGAACCGATTCGCAATAGCCTGTTTGAAA
TCAAAACCATCTATGTGGATAAGCTGACCACCTTTCATTTTCTGGAAGCCCAGAATATTGATGAGAGC
ATTGATAGCAGCAAAATTCGTGTTGAACTGACCGATAGCGTTGATGAAGCACTGAGCAATCCGAATAA
AGTTTATAGCCCGTTCAAGAACATGAGCAACACCATTAATAGCATTGAAACCGGTATTACCAGCACCT
ACATCTTTTATCAGTGGCTGCGTAGCATCGTGAAAGATTTTAGTGATGAAACCGGCAAAATCGACGTG
ATTGATAAAAGCAGCGATACCCTGGCAATTGTTCCGTATATTGGTCCGCTGCTGAATATTGGTAATGA
TATTCGTCATGGCGATTTTGTGGGTGCAATTGAACTGGCAGGCATTACCGCACTGCTGGAATATGTTC
CGGAATTTACCATTCCGATTCTGGTTGGTCTGGAAGTTATTGGTGGCGAACTGGCACGTGAACAGGTT
GAAGCAATTGTTAATAATGCCCTGGATAAACGCGATCAGAAATGGGCAGAAGTTTACAATATTACCAA
AGCACAGTGGTGGGGCACCATTCATTTACAGATTAATACCCGTCTGGCCCATACCTATAAAGCCCTGA
GCCGTCAGGCAAATGCCATTAAAATGAATATGGAATTTCAGCTGGCCAACTACAAAGGCAACATTGAT
GATAAAGCCAAGATCAAAAACGCCATCAGCGAAACCGAAATTCTGCTGAACAAAAGCGTTGAACAGGC
CATGAAAAACACCGAGAAATTCATGATTAAACTGAGCAACAGCTACCTGACCAAAGAAATGATTCCGA
AAGTTCAGGACAACCTGAAAAACTTTGATCTGGAAACCAAAAAGACCCTGGACAAGTTCATCAAAGAG
AAAGAAGATATCCTGGGCACCAATCTGAGCAGCAGCCTGCGTCGTAAAGTTAGCATTCGTCTGAATAA
AAACATTGCCTTCGACATCAACGATATCCCGTTTAGCGAATTTGATGATCTGATCAACCAGTACAAAA
ACGAGATCGAAGATTATGAAGTGCTGAATCTGGGTGCAGAAGATGGCAAAATCAAAGATCTGAGCGGT
ACAACCAGCGATATCAATATTGGTTCAGATATCGAACTGGCCGATGGTCGTGAAAATAAAGCGATTAA
GATTAAAGGCAGCGAGAACAGCACCATCAAAATTGCAATGAACAAATATCTGCGTTTTAGCGCGACCG
ATAACTTTAGCATTAGCTTTTGGATCAAACATCCGAAACCGACCAATCTGCTTAATAACGGTATTGAA
TATACCCTGGTCGAGAACTTTAATCAGCGTGGTTGGAAAATTAGCATCCAGGATAGCAAACTGATTTG
GTATCTGCGCGATCACAATAACAGCATCAAAATCGTTACACCGGATTATATTGCGTTTAATGGCTGGA
ACCTGATTACCATTACAAACAATCGTAGCAAAGGCAGCATCGTGTATGTTAACGGTAGCAAAATTGAA
GAGAAGGACATTAGCAGCATTTGGAATACCGAAGTGGATGATCCGATTATCTTCCGCCTGAAAAATAA
CCGTGATACCCAGGCATTTACCCTGCTGGATCAGTTTAGCATTTATCGCAAAGAACTGAACCAGAACG
AAGTGGTGAAACTGTATAACTACTACTTCAACAGCAACTACATTCGCGATATTTGGGGTAATCCGCTG
CAGTACAACAAAAAATACTATCTGCAGACCCAGGACAAACCTGGTAAAGGTCTGATCCGCGAATATTG
GAGCAGCTTTGGTTATGATTATGTGATTCTGAGCGATAGCAAGACCATTACCTTTCCGAATAATATCC
GTTATGGTGCCCTGTATAATGGTAGCAAAGTGCTGATCAAGAACAGCAAAAAACTGGATGGTCTGGTG
CGCAATAAAGATTTCATTCAGCTGGAAATCGATGGCTATAATATGGGTATTAGCGCAGATCGCTTTAA
CGAGGATACCAACTATATTGGCACCACCTATGGTACAACCCATGATCTGACCACCGATTTTGAAATTA
TTCAGCGCCAAGAGAAATACCGCAATTATTGTCAGCTGAAAACCCCGTATAACATCTTTCATAAAAGC
GGTCTGATGAGCACCGAAACCAGCAAACCGACCTTCCATGATTATCGCGATTGGGTTTATAGCAGCGC
ATGGTATTTTCAGAACTATGAAAATCTGAACCTGCGCAAACATACCAAAACCAACTGGTATTTTATCC
CGAAAGATGAAGGTTGGGATGAAGATCTGGAAGTGCTGTTTCAGGGTCCGCATCATCACCACCATCAC
CATCATCATCACTGA

SEQ ID NO: 5 (Polypeptide Sequence of BoNT/X with a C1 Activation Loop)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNTNDLNIPSE
PIMEADAIYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIPLPLVSNGALTLSDNET
IAYQENNNIVSNLQANLVIYGPGPDIANNATYGLYSTPISNGEGTLSEVSFSPFYLKPFDESYGNYRS
LVNIVNKFVKREFAPDPASTLMHELVHVTHNLYGISNRNFYYNFDTGKIETSRQQNSLIFEELLTFGG
IDSKAISSLIIKKIIETAKNNYTTLISERLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLN
TILFVLNESNLAQRFSILVRKHYLKERPIDPIYVNILDDNSYSTLEGFNISSQGSNDFQGQLLESSYF
EKIESNALRAFIKICHKAIDGRSLYNKTLDCIEVENKDLFLISNKDSLNDINLSEEKIKPETTVFFKD

KLPPQDITLSNYDFTEANSIPSISQQNILERNEELYEPIRNSLFEIKTIYVDKLTTFHFLEAQNIDES
IDSSKIRVELTDSVDEALSNPNKVYSPFKNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDV
IDKSSDTLAIVPYIGPLLNIGNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQV
EAIVNNALDKRDQKWAEVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNID
DKAKIKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDKFIKE
KEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNLGAEDGKIKDLSG
TTSDINIGSDIELADGRENKAIKIKGSENSTIKIAMNKYLRFSATDNFSISFWIKHPKPTNLLNNGIE
YTLVENFNQRGWKISIQDSKLIWYLRDHNNSIKIVTPDYIAFNGWNLITITNNRSKGSIVYVNGSKIE
EKDISSIWNTEVDDPIIFRLKNNRDTQAFTLLDQFSIYRKELNQNEVVKLYNYYFNSNYIRDIWGNPL
QYNKKYYLQTQDKPGKGLIREYWSSFGYDYVILSDSKTITFPNNIRYGALYNGSKVLIKNSKKLDGLV
RNKDFIQLEIDGYNMGISADRFNEDTNYIGTTYGTTHDLTTDFEIIQRQEKYRNYCQLKTPYNIFHKS
GLMSTETSKPTFHDYRDWVYSSAWYFQNYENLNLRKHTKTNWYFIPKDEGWDEDLEVLFQGPHHHHHH
HHHH

SEQ ID NO: 6 (Nucleotide Sequence of BoNT/XA [LH_NX-H_CA] with a C1 Activation Loop)

ATGAAACTGGAAATCAACAAATTCAACTACAACGATCCGATCGATGGCATTAATGTTATTACCATGCG
TCCGCCTCGTCATAGCGATAAAATCAATAAAGGTAAAGGTCCGTTCAAAGCCTTTCAGGTGATTAAAA
ACATTTGGATTGTGCCGGAACGCTACAACTTTACCAATAATACCAACGATCTGAACATTCCGAGCGAA
CCGATTATGGAAGCAGATGCCATTTATAACCCGAACTATCTGAATACCCCGAGCGAAAAAGATGAATT
TCTGCAGGGTGTTATCAAAGTGCTGGAACGCATTAAAAGCAAACCGGAAGGTGAAAAACTGCTGGAAC
TGATTAGCAGCAGCATTCCGCTGCCGCTGGTTAGCAATGGTGCACTGACCCTGAGCGATAATGAAACC
ATTGCATATCAAGAGAACAACAACATTGTGAGCAATCTGCAGGCAAACCTGGTTATTTATGGTCCGGG
TCCTGATATTGCAAATAATGCAACCTATGGTCTGTATAGCACCCCGATTAGTAATGGTGAAGGTACAC
TGAGCGAAGTTAGCTTTAGCCCGTTTTATCTGAAACCGTTTGATGAAAGCTATGGCAATTATCGTAGC
CTGGTGAATATCGTGAACAAATTCGTGAAACGTGAATTTGCACCTGATCCGGCAAGCACCCTGATGCA
TGAACTGGTTCATGTTACCCATAATCTGTATGGTATTAGCAACCGCAACTTCTACTATAACTTTGACA
CCGGCAAAATTGAAACCAGCCGTCAGCAGAATAGCCTGATTTTTGAAGAACTGCTGACCTTTGGTGGC
ATTGATAGCAAAGCAATTAGCAGCCTGATCATCAAGAAAATTATCGAAACCGCCAAGAACAACTATAC
CACGCTGATTAGCGAACGCCTGAATACCGTTACCGTTGAAAATGATCTGCTGAAATATATCAAAAACA
AAATCCCGGTTCAGGGTCGTCTGGGTAACTTTAAACTGGATACCGCAGAATTCGAGAAAAAGCTGAAT
ACCATTCTGTTTGTGCTGAACGAAAGCAATCTGGCACAGCGTTTTAGCATTCTGGTTCGTAAACATTA
CCTGAAAGAACGTCCGATTGATCCGATTTATGTGAACATTCTGGATGACAATAGCTACAGCACCCTGG
AAGGTTTTAACATTAGCAGTCAGGGTAGCAATGATTTTCAGGGCCAGCTGCTGGAAAGCAGCTATTTT
GAAAAAATTGAATCCAATGCGCTGCGTGCCTTTATCAAAATTTGTCATAAAGCCATTGATGGTCGCAG
CCTGTATAACAAAACCCTGGATTGTATTGAAGTGGAAAACAAAGACCTGTTCCTGATTAGCAATAAAG
ATAGCCTGAACGATATCAACCTGAGCGAAGAAAAAATCAAACCGGAAACCACCGTGTTCTTCAAAGAT
AAACTGCCTCCGCAGGATATTACCCTGAGCAATTATGATTTTACCGAAGCCAATAGCATTCCGAGCAT
TAGCCAGCAGAACATTCTGGAACGTAATGAAGAACTGTATGAACCGATTCGCAATAGCCTGTTTGAAA
TCAAAACCATCTATGTGGATAAGCTGACCACCTTTCATTTTCTGGAAGCCCAGAATATTGATGAGAGC
ATTGATAGCAGCAAAATTCGTGTTGAACTGACCGATAGCGTTGATGAAGCACTGAGCAATCCGAATAA
AGTTTATAGCCCGTTCAAGAACATGAGCAACACCATTAATAGCATTGAAACCGGTATTACCAGCACCT
ACATCTTTTATCAGTGGCTGCGTAGCATCGTGAAAGATTTTAGTGATGAAACCGGCAAAATCGACGTG
ATTGATAAAAGCAGCGATACCCTGGCCATTGTTCCGTATATTGGTCCGCTGCTGAATATTGGTAATGA
TATTCGTCATGGCGATTTTGTGGGTGCAATTGAACTGGCAGGCATTACCGCACTGCTGGAATATGTTC
CGGAATTTACCATTCCGATTCTGGTTGGTCTGGAAGTTATTGGTGGCGAACTGGCACGTGAACAGGTT
GAAGCAATTGTTAATAATGCCCTGGATAAACGCGATCAGAAATGGGCAGAAGTTTACAATATTACCAA
AGCACAGTGGTGGGGCACCATTCATTTACAGATTAATACCCGTCTGGCCCATACCTATAAAGCCCTGA
GCCGTCAGGCAAATGCCATTAAAATGAATATGGAATTTCAGCTGGCCAACTACAAAGGCAACATTGAT
GATAAAGCCAAGATCAAAAACGCCATCAGCGAAACCGAAATTCTGCTGAACAAAAGCGTTGAACAGGC
CATGAAAAACACCGAGAAATTCATGATTAAACTGAGCAACAGCTACCTGACCAAAGAAATGATTCCGA
AAGTTCAGGACAACCTGAAAAACTTTGATCTGGAAACCAAAAAGACCCTGGACAAGTTCATCAAAGAG
AAAGAAGATATCCTGGGCACCAATCTGAGCAGCAGCCTGCGTCGTAAAGTTAGCATTCGTCTGAATAA
AAACATTGCCTTCGACATCAACGATATCCCGTTTAGCGAATTTGATGATCTGATCAACCAGTACAAAA
ACGAGATCGAAGATTATGAAGTGCTGAATCTGGGTGCAGAAGATGGCAAAATCAAAGATCTGAGCGGT
ACAACCAGCGATATTAACATTGGTAGCGATATCGAAATCATCAACACCAGCATTCTGAATCTGCGCTA

TGAAAGCAATCATCTGATTGATCTGAGCCGTTATGCGTCCAAAATCAATATTGGCAGCAAAGTGAATT
TCGACCCGATCGATAAAAATCAGATCCAGCTGTTTAATCTGGAAAGCTCCAAAATTGAGGTGATTCTG
AAAAACGCGATTGTGTACAATAGCATGTATGAGAATTTCTCAACCAGCTTCTGGATTCGCATTCCGAA
ATACTTTAACAGCATCAGCCTGAACAACGAGTATACCATTATCAACTGCATGGAAAACAATAGCGGTT
GGAAAGTGAGCCTGAATTATGGTGAAATTATCTGGACCCTGCAGGATACCCAAGAAATCAAACAGCGT
GTTGTGTTCAAATACAGCCAGATGATTAACATCAGCGATTACATTAACCGCTGGATCTTTGTTACCAT
TACCAACAATCGCCTGAATAACAGCAAGATCTATATTAACGGTCGTCTGATTGACCAGAAACCGATTA
GTAATCTGGGTAATATTCATGCCAGCAACAACATCATGTTCAAACTGGATGGTTGTCGTGATACCCAT
CGTTATATTTGGATCAAGTATTTTAACCTGTTTGATAAAGAACTGAACGAAAAAGAAATTAAGGATCT
GTATGATAACCAGTCCAATAGCGGCATCCTGAAGGATTTTTGGGGTGATTATCTGCAGTATGACAAAC
CGTATTATATGCTGAACCTGTACGATCCGAACAAATATGTGGATGTGAATAATGTGGGTATCCGTGGC
TATATGTATCTGAAAGGTCCGCGTGGTAGCGTTATGACCACCAACATTTATCTGAATAGCAGCCTGTA
TCGTGGCACCAAATTCATCATCAAAAAATACGCCAGCGGCAACAAAGATAATATTGTGCGTAATAATG
ACCGCGTGTATATCAATGTGGTGGTGAAGAATAAAGAATATCGTCTGGCAACCAATGCAAGCCAGGCA
GGCGTTGAAAAAATTCTGAGCGCACTGGAAATCCCGGATGTGGGTAATCTGAGCCAGGTTGTTGTTAT
GAAAAGCAAAAATGATCAGGGCATCACCAACAAGTGCAAAATGAATCTGCAGGACAATAACGGCAACG
ACATTGGTTTTATTGGCTTTCACCAGTTTAACAACATTGCCAAACTGGTTGCGAGCAATTGGTATAAT
CGTCAGATTGAACGTAGCAGTCGTACCCTGGGTTGTAGCTGGGAATTTATTCCGGTTGATGATGGTTG
GGGTGAACGTCCGCTGCATCATCACCACCATCACCATCACCACCATTAA

SEQ ID NO: 7 (Polypeptide Sequence of BoNT/XA [LH_NX-H_CA] with a C1 Activation Loop)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNTNDLNIPSE
PIMEADAIYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIPLPLVSNGALTLSDNET
IAYQENNNIVSNLQANLVIYGPGPDIANNATYGLYSTPISNGEGTLSEVSFSPFYLKPFDESYGNYRS
LVNIVNKFVKREFAPDPASTLMHELVHVTHNLYGISNRNFYYNFDTGKIETSRQQNSLIFEELLTFGG
IDSKAISSLIIKKIIETAKNNYTTLISERLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLN
TILFVLNESNLAQRFSILVRKHYLKERPIDPIYVNILDDNSYSTLEGFNISSQGSNDFQGQLLESSYF
EKIESNALRAFIKICHKAIDGRSLYNKTLDCIEVENKDLFLISNKDSLNDINLSEEKIKPETTVFFKD
KLPPQDITLSNYDFTEANSIPSISQQNILERNEELYEPIRNSLFEIKTIYVDKLTTFHFLEAQNIDES
IDSSKIRVELTDSVDEALSNPNKVYSPFKNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDV
IDKSSDTLAIVPYIGPLLNIGNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQV
EAIVNNALDKRDQKWAEVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNID
DKAKIKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDKFIKE
KEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNLGAEDGKIKDLSG
TTSDINIGSDIEIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQIQLFNLESSKIEVIL
KNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQR
VVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTH
RYIWIKYFNLFDKELNEKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRG
YMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQA
GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAKLVASNWYN
RQIERSSRTLGCSWEFIPVDDGWGERPLHHHHHHHHHH

SEQ ID NO: 8 (Nucleotide Sequence of BoNT/XB [LH_NX-H_CB] with a C1 Activation Loop)

ATGAAACTGGAAATCAACAAATTCAACTACAACGATCCGATCGATGGCATTAATGTTATTACCATGCG
TCCGCCTCGTCATAGCGATAAAATCAATAAAGGTAAAGGTCCGTTCAAAGCCTTTCAGGTGATTAAAA
ACATTTGGATTGTGCCGGAACGCTACAACTTTACCAATAATACCAACGATCTGAACATTCCGAGCGAA
CCGATTATGGAAGCAGATGCCATTTATAACCCGAACTATCTGAATACCCCGAGCGAAAAAGATGAATT
TCTGCAGGGTGTTATCAAAGTGCTGGAACGCATTAAAAGCAAACCGGAAGGTGAAAAACTGCTGGAAC
TGATTAGCAGCAGCATTCCGCTGCCGCTGGTTAGCAATGGTGCACTGACCCTGAGCGATAATGAAACC
ATTGCATATCAAGAGAACAACAACATTGTGAGCAATCTGCAGGCAAACCTGGTTATTTATGGTCCGGG
TCCTGATATTGCAAATAATGCAACCTATGGTCTGTATAGCACCCCGATTAGTAATGGTGAAGGTACAC
TGAGCGAAGTTAGCTTTAGCCCGTTTTATCTGAAACCGTTTGATGAAAGCTATGGCAATTATCGTAGC
CTGGTGAATATCGTGAACAAATTCGTGAAACGTGAATTTGCACCTGATCCGGCAAGCACCCTGATGCA
TGAACTGGTTCATGTTACCCATAATCTGTATGGTATTAGCAACCGCAACTTCTACTATAACTTTGACA
CCGGCAAAATTGAAACCAGCCGTCAGCAGAATAGCCTGATTTTTGAAGAACTGCTGACCTTTGGTGGC

ATTGATAGCAAAGCAATTAGCAGCCTGATCATCAAGAAAATTATCGAAACCGCCAAGAACAACTATAC
CACGCTGATTAGCGAACGCCTGAATACCGTTACCGTTGAAAATGATCTGCTGAAATATATCAAAAACA
AAATCCCGGTTCAGGGTCGTCTGGGTAACTTTAAACTGGATACCGCAGAATTCGAGAAAAAGCTGAAT
ACCATTCTGTTTGTGCTGAACGAAAGCAATCTGGCACAGCGTTTTAGCATTCTGGTTCGTAAACATTA
CCTGAAAGAACGTCCGATTGATCCGATTTATGTGAACATTCTGGATGACAATAGCTACAGCACCCTGG
AAGGTTTTAACATTAGCAGTCAGGGTAGCAATGATTTTCAGGGCCAGCTGCTGGAAAGCAGCTATTTT
GAAAAAATTGAATCCAATGCGCTGCGTGCCTTTATCAAAATTTGTCATAAAGCCATTGATGGTCGCAG
CCTGTATAACAAAACCCTGGATTGTATTGAAGTGGAAAACAAAGACCTGTTCCTGATTAGCAATAAAG
ATAGCCTGAACGATATCAACCTGAGCGAAGAAAAAATCAAACCGGAAACCACCGTGTTCTTCAAAGAT
AAACTGCCTCCGCAGGATATTACCCTGAGCAATTATGATTTTACCGAAGCCAATAGCATTCCGAGCAT
TAGCCAGCAGAACATTCTGGAACGTAATGAAGAACTGTATGAACCGATTCGCAATAGCCTGTTTGAAA
TCAAAACCATCTATGTGGATAAGCTGACCACCTTTCATTTTCTGGAAGCCCAGAATATTGATGAGAGC
ATTGATAGCAGCAAAATTCGTGTTGAACTGACCGATAGCGTTGATGAAGCACTGAGCAATCCGAATAA
AGTTTATAGCCCGTTCAAGAACATGAGCAACACCATTAATAGCATTGAAACCGGTATTACCAGCACCT
ACATCTTTTATCAGTGGCTGCGTAGCATCGTGAAAGATTTTAGTGATGAAACCGGCAAAATCGACGTG
ATTGATAAAAGCAGCGATACCCTGGCCATTGTTCCGTATATTGGTCCGCTGCTGAATATTGGTAATGA
TATTCGTCATGGCGATTTTGTGGGTGCAATTGAACTGGCAGGCATTACCGCACTGCTGGAATATGTTC
CGGAATTTACCATTCCGATTCTGGTTGGTCTGGAAGTTATTGGTGGCGAACTGGCACGTGAACAGGTT
GAAGCAATTGTTAATAATGCCCTGGATAAACGCGATCAGAAATGGGCAGAAGTTTACAATATTACCAA
AGCACAGTGGTGGGGCACCATTCATTTACAGATTAATACCCGTCTGGCCCATACCTATAAAGCCCTGA
GCCGTCAGGCAAATGCCATTAAAATGAATATGGAATTTCAGCTGGCCAACTACAAAGGCAACATTGAT
GATAAAGCCAAGATCAAAAACGCCATCAGCGAAACCGAAATTCTGCTGAACAAAAGCGTTGAACAGGC
CATGAAAAACACCGAGAAATTCATGATTAAACTGAGCAACAGCTACCTGACCAAAGAAATGATTCCGA
AAGTTCAGGACAACCTGAAAAACTTTGATCTGGAAACCAAAAAGACCCTGGACAAGTTCATCAAAGAG
AAAGAAGATATCCTGGGCACCAATCTGAGCAGCAGCCTGCGTCGTAAAGTTAGCATTCGTCTGAATAA
AAACATTGCCTTCGACATCAACGATATCCCGTTTAGCGAATTTGATGATCTGATCAACCAGTACAAAA
ACGAGATCGAAGATTATGAAGTGCTGAATCTGGGTGCAGAAGATGGCAAAATCAAAGATCTGAGCGGT
ACAACCAGCGATATTAACATTGGTAGCGATATCGAAATCCTGAACAACATTATTCTGAACCTGCGCTA
TAAAGATAACAACCTGATTGATCTGAGTGGCTATGGTGCAAAAGTTGAAGTTTATGATGGTGTGGAAC
TGAACGACAAAAACCAGTTCAAACTGACCAGCAGCGCAAATTCAAAAATTCGCGTTACCCAGAACCAG
AACATCATTTTTAACAGCGTGTTTCTGGATTTCAGCGTGAGCTTTTGGATTCGTATTCCGAAATATAA
GAACGACGGCATCCAGAACTATATCCACAATGAATATACCATCATCAACTGCATGAAGAATAACAGCG
GTTGGAAAATTAGCATCCGTGGCAATCGTATTATTTGGACCCTGATCGATATTAATGGCAAAACCAAG
AGCGTGTTTTTCGAGTATAACATCCGTGAAGATATCAGCGAATACATCAACCGTTGGTTTTTTGTGAC
CATTACCAACAATCTGAACAACGCCAAAATCTACATTAACGGCAAACTGGAAAGCAACACCGATATCA
AAGATATTCGTGAAGTGATTGCCAACGGCGAGATTATCTTTAAACTGGATGGTGATATTGATCGCACC
CAGTTTATTTGGATGAAATACTTCAGCATCTTCAACACCGAACTGAGCCAGAGCAATATTGAAGAACG
CTATAAAATCCAGAGCTACAGCGAGTATCTGAAAGACTTTTGGGGTAATCCGCTGATGTACAACAAAG
AATACTACATGTTTAATGCCGGTAACAAAAACAGCTATATCAAACTGAAAAAGGATAGTCCGGTGGGT
GAAATTCTGACCCGTAGCAAATATAACCAGAATAGCAAGTATATCAACTATCGCGATCTGTACATCGG
CGAGAAATTTATCATTCGTCGTAAAAGCAACTCCCAGAGCATTAACGATGATATTGTGCGCAAAGAGG
ATTACATCTACCTGGATTTTTTCAACCTGAATCAAGAGTGGCGTGTGTACACCTATAAGTACTTCAAA
AAAGAAGAAATGAAACTGTTTCTGGCACCGATCTATGATAGCGACGAATTTTACAATACCATTCAGAT
TAAAGAATATGATGAACAGCCGACCTATAGCTGTCAGCTGCTGTTTAAAAAGGATGAAGAAAGCACGG
ATGAAATTGGCCTGATTGGTATCCATCGTTTTTATGAAAGCGGCATCGTGTTCGAAGAGTACAAAGAT
TATTTCTGCATCAGCAAATGGTATCTTAAAGAGGTGAAACGCAAACCGTATAATCTGAAACTGGGTTG
CAATTGGCAGTTCATCCCGAAAGATGAAGGTTGGACCGAACATCATCACCACCATCACCATCATCATC
ACTGA

SEQ ID NO: 9 (Polypeptide Sequence of BoNT/XB [LH_NX-H_CB] with a C1 Activation Loop)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNTNDLNIPSE
PIMEADAIYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIPLPLVSNGALTLSDNET
IAYQENNNIVSNLQANLVIYGPGPDIANNATYGLYSTPISNGEGTLSEVSFSPFYLKPFDESYGNYRS
LVNIVNKFVKREFAPDPASTLMHELVHVTHNLYGISNRNFYYNFDTGKIETSRQQNSLIFEELLTFGG
IDSKAISSLIIKKIIETAKNNYTTLISERLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLN
TILFVLNESNLAQRFSILVRKHYLKERPIDPIYVNILDDNSYSTLEGFNISSQGSNDFQGQLLESSYF
EKIESNALRAFIKICHKAIDGRSLYNKTLDCIEVENKDLFLISNKDSLNDINLSEEKIKPETTVFFKD
KLPPQDITLSNYDFTEANSIPSISQQNILERNEELYEPIRNSLFEIKTIYVDKLTTFHFLEAQNIDES
IDSSKIRVELTDSVDEALSNPNKVYSPFKNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDV
IDKSSDTLAIVPYIGPLLNIGNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQV
EAIVNNALDKRDQKWAEVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNID
DKAKIKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDKFIKE
KEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNLGAEDGKIKDLSG
TTSDINIGSDIEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFKLTSSANSKIRVTQNQ
NIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWTLIDINGKTK
SVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDRT
QFIWMKYFSIFNTELSQSNIEERYKIQSYSEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVG
EILTRSKYNQNSKYINYRDLYIGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFK
KEEMKLFLAPIYDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKD
YFCISKWYLKEVKRKPYNLKLGCNWQFIPKDEGWTEHHHHHHHHHH

SEQ ID NO: 10 (Nucleotide Sequence of BoNT/E with a C1 Activation Loop)

atgccgaaaatcaactctttcaactacaacgacccggttaacgaccgtaccatcctgtatatcaaacc
gggtggttgccaggagttctacaaatctttcaacatcatgaaaaacatctggatcatcccggaacgta
acgttatcggtaccaccccgcaggacttccacccgccgacctctctgaaaaacggtgactcttcttac
tacgacccgaactacctccagtctgacgaagaaaaagaccgtttcctgaaaatcgttaccaaaatctt
caaccgtatcaacaacaacctgtctggtggtatcctgctggaagaactgtctaaagctaacccgtacc
tgggtaacgacaacaccccggacaaccagttccacatcggtgacgcttctgctgttgaaatcaaattc
tctaacggttctcaggacatcctgctgccgaacgttatcatcatgggtgctgaaccggacctgttcga
aaccaactcttctaacatctctctgcgtaacaactacatgccgtctaaccacggtttcggttctatcg
ctatcgttaccttctctccggaatactctttccgtttcaacgacaacagcatgaacgagttcatccag
gacccggctctgaccctgatgcaccaactgatctattctctgcacggtctgtacggtgctaaaggtat
caccaccaaatacaccatcacccagaaacagaacccgctgatcaccaacatccgtggtaccaacatcg
aagagttcctgaccttcggtggtaccgacctgaacatcatcacctctgctcagtctaacgacatctac
accaacctgctggctgactacaaaaaaatcgcttctaaactgtctaaagttcaggtttctaacccgct
gctgaacccgtacaaagacgttttcgaagctaaatacggtctggacaaagacgcttctggtatctact
ctgttaacatcaacaaattcaacgacatcttcaaaaaactgtactctttcaccgagttcgacctggcg
accaaattccaggttaaatgccgtcagacctacatcggtcagtacaaatacttcaaactgtctaacct
gctgaacgactctatctacaacatctctgaaggttacaacatcaacaacctgaaagttaacttccgtg
gtcagaacgctaacctgaacccgcgtatcatcaccccgatcaccggtcgtggtctggttaaaaaaatc
atccgtttcTGCCACAAAGCGATTGATGGCCGCTCTCTCTATAACAAAACGCTGGATTGCatcgaaat
caacaacggtgaactgttcttcgttgcttctgaaaactcttacaacgacgacaacatcaacaccccga
aagaaatcgacgacaccgttacctctaacaacaactacgaaaacgacctggaccaggttatcctgaac
ttcaactctgaatctgctccgggtctgtctgacgaaaaactgaacctgaccatccagaacgacgctta
catcccgaaatacgactctaacggtacctctgacatcgaacagcacgacgttaacgaactgaacgttt
tcttctacctggacgctcagaaagttccggaaggtgaaaacaacgttaacctgacctcttctatcgac
accgctctgctggaacagccgaaaatctacaccttcttctcttctgagttcatcaacaacgttaacaa
accggttcaggctgctctgttcgtttcttggattcagcaggttctggttgacttcaccaccgaagcta
accagaaatctaccgttgacaaaatcgctgacatctctatcgttgttccgtacatcggtctggctctg
aacatcggtaacgaagctcagaaaggtaacttcaaagacgctctggaactgctgggtgctggtatcct
gctggagttcgaaccggaactgctgatcccgaccatcctggttttcaccatcaaatctttcctgggtt
cttctgacaacaaaaacaaagttatcaaagctatcaacaacgctctgaaagaacgtgacgaaaaatgg
aaagaagtttactctttcatcgtttctaactggatgaccaaaatcaacacccagttcaacaaacgtaa
agaacagatgtaccaggctctccagaaccaggttaacgctatcaaaaccatcatcgaatctaaataca

actcttacaccctggaagaaaaaaacgaactgaccaacaaatacgacatcaaacagatcgaaaacgaa
ctgaaccagaaagtttctatcgctatgaacaacatcgaccgtttcctgaccgaatcttctatctctta
cctgatgaaactcatcaacgaagttaaaatcaacaaactgcgtgaatacgacgaaaacgttaaaacct
acctgctgaactacatcatccagcacggttctatcctgggtgaatctcagcaggaactgaactctatg
gttaccgacaccctgaacaactctatcccgttcaaactgtcttcttacaccgacgacaaaatcctGAT
CTCTTACTTCAACAAATTCTTTAAAcgcATTAAGAGTTCATCGGTTctgaatATGCGGTACAAAAATG
ATAAAtatGTCGATACTTCTGGATATgatAGCAATATCAACATTAACGGCGACGTGTATAAATATccg
ACAAATAAAAACCAGTTTGGGATATATAACGACAAGctgTCGGAGGTCAATattTCTCAAAACGACta
tATCattTACGATAATaaaTATAAAAACTTTAGCATTAGTtttTGGGTTcgtATACCTAATtatGACA
ATaaaattGTAAATGTGAATAACGAGTATACCATTATAAACTGTATGcgcGACAATAACAGTGGTTGG
AAGGTATCGctgAACCATAATGAGATTATCTGGACCctgcagGATAATgcaGGTATAAACCAGAAACT
GGCTTTTAACTATGGAAACGCAAATGGGATCTCAGATTACATTaataaaTGGatttttGTTaccATTA
CGAACGATcgcTTAGGCGACTCAAAACTTTATATTAATggcAATctgATAGATCAGAAATCAATCTTA
AATTTGGGCAATATTCATGTCTCTgatAACATCTTGTTCAAGATCGTTAATTGCAGTTACACTcgtTA
TATTGGCATTCGTTACTTTAATATCTTCgataaaGAActgGACGAGACGGAAATCcagACTCTGTATT
CAAACGAGCCCAATACTAATATATTGAAAGATTTTTGGGGTAACTATCTTTTATATGATAAAGAATAC
TATCTCCTGaatGTATTGAAGCCAAACAATTTCATAGATAGACGCAAGGATAGCACATTAAGTATCAA
CAATATCAGATCTACTATActgttaGCAAATCGCCTcTACTCCggtATTAAAGTGAAGATTcagCGGG
TTAATAACTCCAGTACCAATGATAATCTGGTCCGTAAGAACGATCAGGTATACATCaatTTCGTCGCG
AGCAAAACTcatCTCTTCCCGCTTTACGCCgatACAGCTACGACAAACAAGGAAAAAACCATAAAAAT
TTCCAGCTCCGGAAACAGATTCAATCAAGTAGTTGTAATGAACTCTGTGGGTaatAATTGTACGATGA
ACTTTaagAATAACAATGGGAACAATattGGACTTTTGGGCTTcAAAGCCGACACAGTGGTGGCGTCC
ACCTGGTATTACACGcacATGcggGACCATACGAATTCGAACGGTTGCTTCTGGAACTTTATCTCGGA
AgaaCACGGGTGGCAAGAAAAA

SEQ ID NO: 11 (Polypeptide Sequence of BoNT/E with a C1 Activation Loop)

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSSY
YDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKF
SNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSMNEFIQ
DPALTLMHQLIYSLHGLYGAKGITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIY
TNLLADYKKIASKLSKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLA
TKFQVKCRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKI
IRFCHKAIDGRSLYNKTLDCIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILN
FNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSID
TALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLAL
NIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKW
KEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENE
LNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSM
VTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYP
TNKNQFGIYNDKLSEVNISQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGW
KVSLNHNEIIWTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSIL
NLGNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEY
YLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVA
SKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMNFKNNNGNNIGLLGFKADTVVAS
TWYYTHMRDHTNSNGCFWNFISEEHGWQEK.

SEQ ID NO: 12 (Nucleotide Sequence of BoNT/A1C1 with a C1 Activation Loop)

ATGCCATTCGTCAACAAGCAATTCAACTACAAAGACCCAGTCAACGGCGTCGACATCGCATACATCAA
GATTCCGAACGCCGGTCAAATGCAGCCGGTTAAGGCTTTTAAGATCCACAACAAGATTTGGGTTATCC
CGGAGCGTGACACCTTCACGAACCCGGAAGAAGGCGATCTGAACCCGCCACCGGAAGCGAAGCAAGTC
CCTGTCAGCTACTACGATTCGACGTACCTGAGCACGGATAACGAAAAAGATAACTACCTGAAAGGTGT
GACCAAGCTGTTCGAACGTATCTACAGCACGGATCTGGGTCGCATGCTGCTGACTAGCATTGTTCGCG
GTATCCCGTTCTGGGGTGGTAGCACGATTGACACCGAACTGAAGGTTATCGACACTAACTGCATTAAC
GTTATTCAACCGGATGGTAGCTATCGTAGCGAAGAGCTGAATCTGGTCATCATTGGCCCGAGCGCAGA
CATTATCCAATTCGAGTGCAAGAGCTTTGGTCACGAGGTTCTGAATCTGACCCGCAATGGCTATGGTA

GCACCCAGTACATTCGTTTTTCGCCGGATTTTACCTTCGGCTTTGAAGAGAGCCTGGAGGTTGATACC
AATCCGTTGCTGGGTGCGGGCAAATTCGCTACCGATCCGGCTGTCACGCTGGCCCATgAACTGATCcA
CGCAGGCCACCGCCTGTACGGCATTGCCATCAACCCAAACCGTGTGTTCAAGGTTAATACGAATGCAT
ACTACGAGATGAGCGGCCTGGAAGTCAGCTTCGAAGAACTGCGCACCTTCGGTGGCCATGACGCTAAA
TTCATTGACAGCTTGCAAGAGAATGAGTTCCGTCTGTACTACTATAACAAATTCAAAGACATTGCAAG
CACGTTGAACAAGGCCAAAAGCATCGTTGGTACTACCGCGTCGTTGCAGTATATGAAGAATGTGTTTA
AAGAGAAGTACCTGCTGTCCGAGGATACCTCCGGCAAGTTTAGCGTTGATAAGCTGAAGTTTGACAAA
CTGTACAAGATGCTGACCGAGATTTACACCGAGGACAACTTTGTGAAATTCTTCAAAGTGTTGAATCG
TAAAACCTATCTGAATTTTGACAAAGCGGTTTTCAAGATTAACATCGTGCCGAAGGTGAACTACACCA
TCTATGACGGTTTTAACCTGCGTAACACCAACCTGGCGGCGAACTTTAACGGTCAGAATACGGAAATC
AACAACATGAATTTCACGAAGTTGAAGAACTTCACGGGTCTGTTCGAGTTCTATAAGCTGCTGTGCCA
CAAAGCGATTGATGGCCGCTCTCTCTATAACAAAACGCTGGATTGCATTAAGGTAAACAATTGGGATC
TGTTCTTTTCGCCATCCGAAGATAATTTTACCAACGACCTGAACAAGGGTGAAGAAATCACCAGCGAT
ACGAATATTGAAGCAGCGGAAGAGAATATCAGCCTGGATCTGATCCAGCAGTACTATCTGACCTTTAA
CTTCGACAATGAACCGGAGAACATTAGCATTGAGAATCTGAGCAGCGACATTATCGGTCAGCTGGAAC
TGATGCCGAATATCGAACGTTTCCCGAACGGCAAAAAGTACGAGCTGGACAAGTACACTATGTTCCAT
TACCTGCGTGCACAGGAGTTTGAACACGGTAAAAGCCGTATCGCGCTGACCAACAGCGTTAACGAGGC
CCTGCTGAACCCGAGCCGTGTCTATACCTTCTTCAGCAGCGACTATGTTAAGAAAGTGAACAAAGCCA
CTGAGGCCGCGATGTTCCTGGGCTGGGTGGAACAGCTGGTATATGACTTCACGGACGAGACGAGCGAA
GTGAGCACTACCGACAAAATTGCTGATATTACCATCATTATCCCGTATATTGGTCCGGCACTGAACAT
TGGCAACATGCTGTACAAAGACGATTTTGTGGGTGCCCTGATCTTCTCCGGTGCCGTGATTCTGCTGG
AGTTCATTCCGGAGATTGCGATCCCGGTGTTGGGTACCTTCGCGCTGGTGTCCTACATCGCGAATAAG
GTTCTGACGGTTCAGACCATCGATAACGCGCTGTCGAAACGTAATGAAAAATGGGACGAGGTTTACAA
ATACATTGTTACGAATTGGCTGGCGAAAGTCAATACCCAGATCGACCTGATCCGTAAGAAAATGAAAG
AGGCGCTGGAGAATCAGGCGGAGGCCACCAAAGCAATTATCAACTACCAATACAACCAGTACACGGAA
GAAGAGAAGAATAACATTAACTTCAATATCGATGATTTGAGCAGCAAGCTGAATGAATCTATCAACAA
AGCGATGATCAATATCAACAAGTTTTTGAATCAGTGTAGCGTTTCGTACCTGATGAATAGCATGATTC
CGTATGGCGTCAAACGTCTGGAGGACTTCGACGCCAGCCTGAAAGATGCGTTGCTGAAATACATTTAC
GACAATCGTGGTACGCTGATTGGCCAAGTTGACCGCTTGAAAGACAAAGTTAACAATACCCTGAGCAC
CGACATCCCATTTCAACTGAGCAAGTATGTTGATAATCAACGTCTGTTGAGCACTTTCACCGAGTATA
TCAAAAACATTAATGACAGCAAAATTCTGAGCCTGCAGAATCGTAAGAATACGCTGGTAGATACCAGT
GGATATAATGCGGAAGTCTCAGAAGAGGGTGATGTACAGCTGAACCCGATCTTTCCGTTCGACTTTAA
ACTGGGGTCTAGTGGTGAAGATCGCGGTAAAGTGATCGTTACCCAAAACGAGAACATTGTGTATAACA
GCATGTACGAGAGTTTCTCAATTTCTTTCTGGATTCGCATCAATAAATGGGTTTCTAATTTGCCTGGC
TATACCATCATTGATAGCGTCAAAAACAACTCGGGCTGGTCGATTGGCATTATTAGCAACTTTCTGGT
GTTTACCCTGAAACAGAATGAGGATTCGGAACAGAGCATTAACTTCTCCTACGACATCAGCAACAATG
CACCAGGGTATAACAAATGGTTCTTCGTAACGGTGACGAACAATATGATGGGCAATATGAAAATCTAC
ATTAACGGGAAACTTATCGACACCATTAAAGTGAAAGAGCTTACTGGGATCAATTTTAGTAAAACCAT
TACCTTTGAGATCAACAAAATTCCGGACACGGGTCTGATTACCTCCGATTCGGATAATATCAATATGT
GGATTCGCGACTTTTATATCTTCGCCAAAGAACTTGATGGCAAAGATATCAACATTTTGTTTAATTCC
CTGCAGTATACCAATGTCGTTAAGGACTATTGGGGCAATGATCTCCGCTACAATAAAGAATACTACAT
GGTTAACATCGACTATCTCAATCGCTACATGTATGCTAACTCGCGTCAAATTGTGTTTAACACACGTC
GTAACAACAACGATTTTAACGAAGGTTATAAAATCATTATCAAACGGATCCGCGGCAATACGAACGAT
ACTCGTGTTCGTGGCGGTGACATTCTGTATTTCGACATGACGATTAATAATAAAGCGTACAATCTGTT
CATGAAGAACGAAACCATGTACGCCGATAACCATTCCACTGAAGATATCTACGCAATCGGACTTCGCG
AACAGACCAAAGACATTAACGACAACATCATCTTTCAGATTCAACCGATGAATAATACCTACTACTAT
GCCTCCCAGATCTTCAAAAGTAATTTCAACGGCGAAAACATTTCAGGCATTTGCTCAATCGGCACTTA
TCGGTTCCGGTTAGGTGGTGATTGGTATCGTCACAACTACCTTGTTCCCACAGTGAAACAAGGCAACT
ATGCATCGCTCTTAGAAAGCACATCTACGCATTGGGGTTTTGTGCCAGTCAGTGAATAA

SEQ ID NO: 13 (Polypeptide Sequence of BoNT/A1C1 with a C1 Activation Loop)

MPFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQV
PVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCIN
VIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDT
NPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAK
FIDSLQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDK
LYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEI
NNMNFTKLKNFTGLFEFYKLLCHKAIDGRSLYNKTLDCIKVNNWDLFFSPSEDNFTNDLNKGEEITSD
TNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFH
YLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSE
VSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANK
VLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTE
EEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIY
DNRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNINDSKILSLQNRKNTLVDTS
GYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPG
YTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNMKIY
INGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNS
LQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGNTND
TRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYY
ASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE

SEQ ID NO: 14 (Nucleotide Sequence of BoNT/C1(0) (Endonegative))

ATGCCGATCACGATTAATAATTTCAACTATAGCGATCCGGTGGACAATAAGAATATTCTGTATCTGGA
TACTCATCTGAATACGCTGGCTAACGAACCGGAGAAAGCGTTCCGCATCACAGGCAACATCTGGGTTA
TTCCCGATCGCTTTTCACGCAACAGCAACCCTAATCTGAACAAACCTCCTCGTGTCACCAGTCCTAAA
TCCGGTTATTACGACCCAAACTATCTGAGTACGGATAGCGATAAAGATCCCTTTCTGAAAGAGATCAT
TAAGCTGTTCAAACGCATTAACTCTCGCGAAATTGGGGAAGAGCTGATCTATCGGCTTTCGACAGATA
TCCCGTTCCCAGGTAACAATAATACCCCGATTAATACTTTCGACTTTGATGTTGATTTCAATTCTGTG
GATGTGAAAACGCGTCAAGGCAATAATTGGGTGAAAACTGGTAGCATTAACCCGAGTGTAATTATCAC
AGGTCCCCGTGAGAACATCATCGACCCGGAAACCTCTACCTTCAAGCTGACGAACAACACGTTTGCTG
CACAGGAAGGGTTTGGTGCCCTGTCAATCATTTCCATCTCACCGCGTTTCATGTTAACCTACTCCAAT
GCCACAAATGATGTTGGCGAAGGACGTTTTAGCAAATCAGAATTTTGCATGGACCCAATTCTCATTCT
GATGggCacGCTGAACaATGCGATGCACAACTTGTATGGCATTGCTATTCCAAACGATCAAACCATTA
GCTCCGTTACCAGTAATATCTTCTATAGCCAGTATAATGTCAAATTGGAGTATGCCGAAATTTACGCC
TTTGGAGGCCCGACCATTGACCTGATTCCGAAATCTGCACGCAAATACTTCGAAGAAAAGGCGTTAGA
TTACTATCGCAGCATCGCGAAACGCCTGAACTCGATTACCACGGCCAATCCGTCGTCGTTCAACAAAT
ACATTGGTGAATATAAACAGAAACTGATTCGCAAATATCGGTTTGTCGTAGAAAGCTCTGGTGAAGTG
ACTGTAAACCGCAACAAATTTGTCGAACTCTACAACGAGTTGACCCAAATCTTTACCGAGTTTAACTA
CGCAAAGATCTATAACGTACAGAACCGCAAGATTTATCTTAGCAATGTATACACACCGGTTACTGCGA
ACATCTTAGACGACAATGTGTATGATATTCAGAATGGCTTTAACATCCCGAAATCAAATCTGAACGTT
CTGTTTATGGGCCAGAACCTGAGTCGTAATCCAGCACTGCGTAAAGTGAACCCGGAAAATATGCTCTA
CTTGTTTACCAAATTTTGCCACAAAGCGATTGATGGCCGCTCTCTCTATAACAAAACGCTGGATTGTC
GTGAGTTACTTGTGAAGAACACTGATTTACCGTTCATTGGGGATATCTCCGACGTGAAAACCGATATC
TTCCTGCGCAAAGACATTAATGAAGAAACGGAAGTCATCTATTACCCCGACAATGTGAGCGTTGATCA
GGTCATTTTATCGAAGAACACCTCCGAACATGGTCAGTTGGATTTGCTGTACCCTAGCATTGACTCGG
AGAGTGAAATCCTTCCGGGCGAAAATCAAGTGTTTTACGACAACCGTACCCAAAATGTTGATTATTTG
AATTCTTATTACTACCTGGAATCTCAGAAATTGAGCGACAATGTGGAAGATTTCACGTTCACACGCTC
CATTGAGGAAGCGCTGGATAATAGCGCGAAAGTGTATACGTATTTCCCTACCTTGGCGAATAAAGTAA
ATGCTGGTGTCCAGGGAGGCTTATTTCTGATGTGGGCGAATGATGTGGTAGAAGATTTTACGACCAAT
ATTTTGCGTAAGGACACCTTAGATAAAATTAGCGATGTTAGCGCCATCATCCCCTATATTGGCCCAGC
ACTGAATATCTCGAACTCTGTGCGTCGCGGAAACTTCACCGAAGCATTTGCGGTGACCGGGGTTACTA
TTCTGTTGGAAGCCTTTCCGGAGTTTACTATTCCGGCGCTGGGTGCGTTTGTGATTTATTCGAAAGTA
CAAGAACGCAATGAAATTATCAAAACCATCGATAATTGCCTGGAACAACGCATTAAACGCTGGAAGGA
TTCTTATGAATGGATGATGGGCACCTGGTTATCCCGTATTATCACACAGTTTAACAACATCTCGTATC
AGATGTACGATTCACTGAACTACCAAGCAGGGGCGATCAAAGCCAAGATCGACTTAGAATACAAGAAA

TATTCAGGTAGCGATAAAGAGAATATTAAAAGCCAGGTTGAAAACCTGAAGAACTCTCTGGATGTCAA
AATTTCAGAGGCTATGAACAACATTAACAAATTTATCCGCGAATGTAGCGTCACGTATCTGTTTAAAA
ACATGCTCCCGAAAGTGATTGATGAGCTCAACGAGTTTGATCGCAACACAAAGGCCAAACTGATTAAC
CTGATTGATAGTCACAATATTATTTTAGTCGGTGAAGTTGACAAGCTGAAGGCTAAGGTCAATAACAG
CTTTCAGAACACTATTCCGTTTAATATTTTCTCCTATACGAACAATAGTCTGCTGAAAGACATTATCA
ACGAATACTTCAACAATATTAATGACAGCAAAATTCTGAGCCTGCAGAATCGTAAGAATACGCTGGTA
GATACCAGTGGATATAATGCGGAAGTCTCAGAAGAGGGTGATGTACAGCTGAACCCGATCTTTCCGTT
CGACTTTAAACTGGGGTCTAGTGGTGAAGATCGCGGTAAAGTGATCGTTACCCAAAACGAGAACATTG
TGTATAACAGCATGTACGAGAGTTTCTCAATTTCTTTCTGGATTCGCATCAATAAATGGGTTTCTAAT
TTGCCTGGCTATACCATCATTGATAGCGTCAAAAACAACTCGGGCTGGTCGATTGGCATTATTAGCAA
CTTTCTGGTGTTTACCCTGAAACAGAATGAGGATTCGGAACAGAGCATTAACTTCTCCTACGACATCA
GCAACAATGCACCAGGGTATAACAAATGGTTCTTCGTAACGGTGACGAACAATATGATGGGCAATATG
AAAATCTACATTAACGGGAAACTTATCGACACCATTAAAGTGAAAGAGCTTACTGGGATCAATTTTAG
TAAAACCATTACCTTTGAGATCAACAAAATTCCGGACACGGGTCTGATTACCTCCGATTCGGATAATA
TCAATATGTGGATTCGCGACTTTTATATCTTCGCCAAAGAACTTGATGGCAAAGATATCAACATTTTG
TTTAATTCCCTGCAGTATACCAATGTCGTTAAGGACTATTGGGGCAATGATCTCCGCTACAATAAAGA
ATACTACATGGTTAACATCGACTATCTCAATCGCTACATGTATGCTAACTCGCGTCAAATTGTGTTTA
ACACACGTCGTAACAACAACGATTTTAACGAAGGTTATAAAATCATTATCAAACGGATCCGCGGCAAT
ACGAACGATACTCGTGTTCGTGGCGGTGACATTCTGTATTTCGACATGACGATTAATAATAAAGCGTA
CAATCTGTTCATGAAGAACGAAACCATGTACGCCGATAACCATTCCACTGAAGATATCTACGCAATCG
GACTTCGCGAACAGACCAAAGACATTAACGACAACATCATCTTTCAGATTCAACCGATGAATAATACC
TACTACTATGCCTCCCAGATCTTCAAAAGTAATTTCAACGGCGAAAACATTTCAGGCATTTGCTCAAT
CGGCACTTATCGGTTCCGGTTAGGTGGTGATTGGTATCGTCACAACTACCTTGTTCCCACAGTGAAAC
AAGGCAACTATGCATCGCTCTTAGAAAGCACATCTACGCATTGGGGTTTTGTGCCAGTCAGTGAATaa
tg

SEQ ID NO: 15 (Polypeptide Sequence of BoNT/C1(0) (Endonegative))

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPK
SGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSV
DVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSN
ATNDVGEGRFSKSEFCMDPILILMGTLNNAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYA
FGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEV
TVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNV
LFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDI
FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYL
NSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTN
ILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKV
QERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK
YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLIN
LIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLV
DTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN
LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM
KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINIL
FNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGN
TNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNT
YYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE

SEQ ID NO: 16 (Nucleotide Sequence of BoNT/C1)

ATGCCGATCACGATTAATAATTTCAACTATAGCGATCCGGTGGACAATAAGAATATTCTGTATCTGGA
TACTCATCTGAATACGCTGGCTAACGAACCGGAGAAAGCGTTCCGCATCACAGGCAACATCTGGGTTA
TTCCCGATCGCTTTTCACGCAACAGCAACCCTAATCTGAACAAACCTCCTCGTGTCACCAGTCCTAAA
TCCGGTTATTACGACCCAAACTATCTGAGTACGGATAGCGATAAAGATCCCTTTCTGAAAGAGATCAT
TAAGCTGTTCAAACGCATTAACTCTCGCGAAATTGGGGAAGAGCTGATCTATCGGCTTTCGACAGATA
TCCCGTTCCCAGGTAACAATAATACCCCGATTAATACTTTCGACTTTGATGTTGATTTCAATTCTGTG
GATGTGAAAACGCGTCAAGGCAATAATTGGGTGAAAACTGGTAGCATTAACCCGAGTGTAATTATCAC

AGGTCCCCGTGAGAACATCATCGACCCGGAAACCTCTACCTTCAAGCTGACGAACAACACGTTTGCTG
CACAGGAAGGGTTTGGTGCCCTGTCAATCATTTCCATCTCACCGCGTTTCATGTTAACCTACTCCAAT
GCCACAAATGATGTTGGCGAAGGACGTTTTAGCAAATCAGAATTTTGCATGGACCCAATTCTCATTCT
GATGCACGAGCTGAACCATGCGATGCACAACTTGTATGGCATTGCTATTCCAAACGATCAAACCATTA
GCTCCGTTACCAGTAATATCTTCTATAGCCAGTATAATGTCAAATTGGAGTATGCCGAAATTTACGCC
TTTGGAGGCCCGACCATTGACCTGATTCCGAAATCTGCACGCAAATACTTCGAAGAAAAGGCGTTAGA
TTACTATCGCAGCATCGCGAAACGCCTGAACTCGATTACCACGGCCAATCCGTCGTCGTTCAACAAAT
ACATTGGTGAATATAAACAGAAACTGATTCGCAAATATCGGTTTGTCGTAGAAAGCTCTGGTGAAGTG
ACTGTAAACCGCAACAAATTTGTCGAACTCTACAACGAGTTGACCCAAATCTTTACCGAGTTTAACTA
CGCAAAGATCTATAACGTACAGAACCGCAAGATTTATCTTAGCAATGTATACACACCGGTTACTGCGA
ACATCTTAGACGACAATGTGTATGATATTCAGAATGGCTTTAACATCCCGAAATCAAATCTGAACGTT
CTGTTTATGGGCCAGAACCTGAGTCGTAATCCAGCACTGCGTAAAGTGAACCCGGAAAATATGCTCTA
CTTGTTTACCAAATTTTGCCACAAAGCGATTGATGGCCGCTCTCTCTATAACAAAACGCTGGATTGTC
GTGAGTTACTTGTGAAGAACACTGATTTACCGTTCATTGGGGATATCTCCGACGTGAAAACCGATATC
TTCCTGCGCAAAGACATTAATGAAGAAACGGAAGTCATCTATTACCCCGACAATGTGAGCGTTGATCA
GGTCATTTTATCGAAGAACACCTCCGAACATGGTCAGTTGGATTTGCTGTACCCTAGCATTGACTCGG
AGAGTGAAATCCTTCCGGGCGAAAATCAAGTGTTTTACGACAACCGTACCCAAAATGTTGATTATTTG
AATTCTTATTACTACCTGGAATCTCAGAAATTGAGCGACAATGTGGAAGATTTCACGTTCACACGCTC
CATTGAGGAAGCGCTGGATAATAGCGCGAAAGTGTATACGTATTTCCCTACCTTGGCGAATAAAGTAA
ATGCTGGTGTCCAGGGAGGCTTATTTCTGATGTGGGCGAATGATGTGGTAGAAGATTTTACGACCAAT
ATTTTGCGTAAGGACACCTTAGATAAAATTAGCGATGTTAGCGCCATCATCCCCTATATTGGCCCAGC
ACTGAATATCTCGAACTCTGTGCGTCGCGGAAACTTCACCGAAGCATTTGCGGTGACCGGGGTTACTA
TTCTGTTGGAAGCCTTTCCGGAGTTTACTATTCCGGCGCTGGGTGCGTTTGTGATTTATTCGAAAGTA
CAAGAACGCAATGAAATTATCAAAACCATCGATAATTGCCTGGAACAACGCATTAAACGCTGGAAGGA
TTCTTATGAATGGATGATGGGCACCTGGTTATCCCGTATTATCACACAGTTTAACAACATCTCGTATC
AGATGTACGATTCACTGAACTACCAAGCAGGGGCGATCAAAGCCAAGATCGACTTAGAATACAAGAAA
TATTCAGGTAGCGATAAAGAGAATATTAAAAGCCAGGTTGAAAACCTGAAGAACTCTCTGGATGTCAA
AATTTCAGAGGCTATGAACAACATTAACAAATTTATCCGCGAATGTAGCGTCACGTATCTGTTTAAAA
ACATGCTCCCGAAAGTGATTGATGAGCTCAACGAGTTTGATCGCAACACAAAGGCCAAACTGATTAAC
CTGATTGATAGTCACAATATTATTTTAGTCGGTGAAGTTGACAAGCTGAAGGCTAAGGTCAATAACAG
CTTTCAGAACACTATTCCGTTTAATATTTTCTCCTATACGAACAATAGTCTGCTGAAAGACATTATCA
ACGAATACTTCAACAATATTAATGACAGCAAAATTCTGAGCCTGCAGAATCGTAAGAATACGCTGGTA
GATACCAGTGGATATAATGCGGAAGTCTCAGAAGAGGGTGATGTACAGCTGAACCCGATCTTTCCGTT
CGACTTTAAACTGGGGTCTAGTGGTGAAGATCGCGGTAAAGTGATCGTTACCCAAAACGAGAACATTG
TGTATAACAGCATGTACGAGAGTTTCTCAATTTCTTTCTGGATTCGCATCAATAAATGGGTTTCTAAT
TTGCCTGGCTATACCATCATTGATAGCGTCAAAAACAACTCGGGCTGGTCGATTGGCATTATTAGCAA
CTTTCTGGTGTTTACCCTGAAACAGAATGAGGATTCGGAACAGAGCATTAACTTCTCCTACGACATCA
GCAACAATGCACCAGGGTATAACAAATGGTTCTTCGTAACGGTGACGAACAATATGATGGGCAATATG
AAAATCTACATTAACGGGAAACTTATCGACACCATTAAAGTGAAAGAGCTTACTGGGATCAATTTTAG
TAAAACCATTACCTTTGAGATCAACAAAATTCCGGACACGGGTCTGATTACCTCCGATTCGGATAATA
TCAATATGTGGATTCGCGACTTTTATATCTTCGCCAAAGAACTTGATGGCAAAGATATCAACATTTTG
TTTAATTCCCTGCAGTATACCAATGTCGTTAAGGACTATTGGGGCAATGATCTCCGCTACAATAAAGA
ATACTACATGGTTAACATCGACTATCTCAATCGCTACATGTATGCTAACTCGCGTCAAATTGTGTTTA
ACACACGTCGTAACAACAACGATTTTAACGAAGGTTATAAAATCATTATCAAACGGATCCGCGGCAAT
ACGAACGATACTCGTGTTCGTGGCGGTGACATTCTGTATTTCGACATGACGATTAATAATAAAGCGTA
CAATCTGTTCATGAAGAACGAAACCATGTACGCCGATAACCATTCCACTGAAGATATCTACGCAATCG
GACTTCGCGAACAGACCAAAGACATTAACGACAACATCATCTTTCAGATTCAACCGATGAATAATACC
TACTACTATGCCTCCCAGATCTTCAAAAGTAATTTCAACGGCGAAAACATTTCAGGCATTTGCTCAAT
CGGCACTTATCGGTTCCGGTTAGGTGGTGATTGGTATCGTCACAACTACCTTGTTCCCACAGTGAAAC
AAGGCAACTATGCATCGCTCTTAGAAAGCACATCTACGCATTGGGGTTTTGTGCCAGTCAGTGAATaa

SEQ ID NO: 17 (Polypeptide Sequence of BoNT/C1)

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPK
SGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSV
DVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSN

ATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYA
FGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEV
TVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNV
LFMGQNLSRNPALRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDI
FLRKDINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYL
NSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTN
ILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKV
QERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKK
YSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLIN
LIDSHNIILVGEVDKLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLV
DTSGYNAEVSEEGDVQLNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSN
LPGYTIIDSVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM
KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINIL
FNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNNDFNEGYKIIIKRIRGN
TNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNT
YYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYRHNYLVPTVKQGNYASLLESTSTHWGFVPVSE

SEQ ID NO: 18 (Enterokinase and Factor Xa Cleavage Site)

IDGR

SEQ ID NO: 19 (Enterokinase and Factor Xa Cleavage Site Variant)

IEGR

SEQ ID NO: 20 (BoNT/X Activation Loop)

CPRNGLLYNAIYRNSKNYLNNIDLEDKKTTSKTNVSYPCSLLNGC

SEQ ID NO: 21 (BoNT/A1 & A6 Activation Loop)

CVRGIITSKTKSLDKGYNKALNDLC

SEQ ID NO: 22 (BoNT/B2, B3 & B6 Activation Loop)

CKSVRAPGIC

SEQ ID NO: 23 (BoNT/D Activation Loop)

CLRLTKNSRDDSTC

SEQ ID NO: 24 (BoNT/E1 to E5, E9 & E12 Activation Loop)

CKNIVSVKGIRKSIC

SEQ ID NO: 25 (BoNT/F1 and F6 Activation Loop)

CKSVIPRKGTKAPPRLC

SEQ ID NO: 26 (BoNT/F2 and F3 Activation Loop)

CKSIIPRKGTKQSPSLC

SEQ ID NO: 27 (BoNT/F4 Activation Loop)

CKSIIPRKGTKAPPRLC

SEQ ID NO: 28 (BoNT/F5 Activation Loop)

CLNSSFKKNTKKPLC

SEQ ID NO: 29 (BoNT/F7 Activation Loop)

CKSIVSKKGTKNSLC

SEQ ID NO: 30 (TeNT Activation Loop)

CKKIIPPTNIRENLYNRTASLTDLGGELC

SEQ ID NO: 31 (BoNT/G Activation Loop)

CKPVMYKNTGKSEQC

SEQ ID NO: 32 (Nucleotide Sequence of Wild-Type BoNT/X-10HT)

ATGAAACTGGAAATCAACAAATTCAACTACAACGATCCGATCGATGGCATTAATGTTATTACCATGCGTCCGCCT
CGTCATAGCGATAAAATCAATAAAGGTAAAGGTCCGTTCAAAGCCTTTCAGGTGATTAAAAACATTTGGATTGTG
CCGGAACGCTACAACTTTACCAATAATACCAACGATCTGAACATTCCGAGCGAACCGATTATGGAAGCAGATGCC
ATTTATAACCCGAACTATCTGAATACCCCGAGCGAAAAAGATGAATTTCTGCAGGGTGTTATCAAAGTGCTGGAA
CGCATTAAAAGCAAACCGGAAGGTGAAAAACTGCTGGAACTGATTAGCAGCAGCATTCCGCTGCCGCTGGTTAGC
AATGGTGCACTGACCCTGAGCGATAATGAAACCATTGCATATCAAGAGAACAACAACATTGTGAGCAATCTGCAG
GCAAACCTGGTTATTTATGGTCCGGGTCCTGATATTGCAAATAATGCAACCTATGGTCTGTATAGCACCCCGATT
AGTAATGGTGAAGGTACACTGAGCGAAGTTAGCTTTAGCCCGTTTTATCTGAAACCGTTTGATGAAAGCTATGGC
AATTATCGTAGCCTGGTGAATATCGTGAACAAATTCGTGAAACGTGAATTTGCACCTGATCCGGCAAGCACCCTG
ATGCATGAACTGGTTCATGTTACCCATAATCTGTATGGTATTAGCAACCGCAACTTCTACTATAACTTTGACACC
GGCAAAATTGAAACCAGCCGTCAGCAGAATAGCCTGATTTTTGAAGAACTGCTGACCTTTGGTGGCATTGATAGC
AAAGCAATTAGCAGCCTGATCATCAAGAAAATTATCGAAACCGCCAAGAACAACTATACCACGCTGATTAGCGAA
CGCCTGAATACCGTTACCGTTGAAAATGATCTGCTGAAATATATCAAAAACAAAATCCCGGTTCAGGGTCGTCTG
GGTAACTTTAAACTGGATACCGCAGAATTCGAGAAAAAGCTGAATACCATTCTGTTTGTGCTGAACGAAAGCAAT
CTGGCACAGCGTTTTAGCATTCTGGTTCGTAAACATTACCTGAAAGAACGTCCGATTGATCCGATTTATGTGAAC
ATTCTGGATGACAATAGCTACAGCACCCTGGAAGGTTTTAACATTAGCAGTCAGGGTAGCAATGATTTCCAAGGT
CAGCTGCTGGAAAGCAGCTATTTTGAAAAAATTGAAAGCAATGCCCTGCGTGCCTTCATTAAAATCTGTCCGCGT
AATGGTCTGCTGTATAATGCCATTTATCGCAACAGCAAAAACTACCTGAACAACATTGATCTGGAAGATAAAAAG
ACCACGAGCAAAACCAATGTTAGCTATCCGTGTAGCCTGCTGAATGGTTGTATTGAAGTTGAAAACAAAGACCTG
TTCCTGATTAGCAACAAAGATAGCCTGAACGATATTAACCTGAGCGAAGAAAAAATCAAACCGGAAACCACCGTG
TTCTTCAAAGATAAACTGCCTCCGCAGGATATTACGCTGAGCAATTATGATTTTACCGAAGCCAATAGCATTCCG
AGCATTAGCCAGCAGAACATTCTGGAACGTAATGAAGAACTGTATGAACCGATTCGCAATAGCCTGTTTGAAATC
AAAACCATCTATGTGGATAAGCTGACCACCTTTCATTTTCTGGAAGCCCAGAATATTGATGAGAGCATTGATAGC
AGCAAAATTCGTGTTGAACTGACCGATAGCGTTGATGAAGCACTGAGCAATCCGAATAAAGTTTATAGCCCGTTC
AAGAACATGAGCAACACCATTAATAGCATTGAAACCGGTATTACCAGCACCTACATCTTTTATCAGTGGCTGCGT
AGCATCGTGAAAGATTTTAGTGATGAAACCGGCAAAATCGACGTGATTGATAAAAGCAGCGATACCCTGGCAATT
GTTCCGTATATTGGTCCGCTGCTGAATATTGGTAATGATATTCGTCATGGCGATTTTGTGGGTGCAATTGAACTG
GCAGGCATTACCGCACTGCTGGAATATGTTCCGGAATTTACCATTCCGATTCTGGTTGGTCTGGAAGTTATTGGT
GGCGAACTGGCACGTGAACAGGTTGAAGCAATTGTTAATAATGCCCTGGATAAACGCGATCAGAAATGGGCAGAA

GTTTACAATATTACCAAAGCACAGTGGTGGGGCACCATTCATTTACAGATTAATACCCGTCTGGCCCATACCTAT
AAAGCCCTGAGCCGTCAGGCAAATGCCATTAAAATGAATATGGAATTTCAGCTGGCCAACTACAAAGGCAACATC
GATGATAAAGCCAAAATCAAAAACGCCATCAGCGAAACCGAAATCCTGCTGAACAAAAGCGTTGAACAGGCAATG
AAAAACACCGAGAAATTCATGATCAAACTGAGCAACAGCTATCTGACCAAAGAAATGATTCCGAAAGTGCAGGAT
AACCTGAAAAATTTCGATCTGGAAACCAAGAAAACCCTGGACAAATTTATCAAAGAGAAAGAGGACATTCTGGGC
ACCAATCTGAGCAGCAGCCTGCGTCGTAAAGTTAGCATTCGTCTGAATAAAAACATTGCCTTCGACATCAACGAT
ATCCCGTTTAGCGAATTTGATGATCTGATCAACCAGTACAAAAACGAGATCGAAGATTATGAAGTGCTGAATCTG
GGTGCAGAAGATGGGAAAATCAAAGATCTGAGCGGTACAACCAGCGATATCAATATTGGTTCAGATATCGAACTG
GCCGATGGTCGTGAAAATAAAGCCATTAAGATTAAAGGCAGCGAGAACAGCACCATCAAAATTGCAATGAACAAA
TATCTGCGTTTTAGCGCGACCGATAACTTTAGCATTAGCTTTTGGATCAAACATCCGAAACCGACCAATCTGCTT
AATAACGGTATTGAATATACCCTGGTCGAGAACTTTAATCAGCGTGGTTGGAAAATTAGCATCCAGGATAGCAAA
CTGATTTGGTATCTGCGCGATCACAATAACAGCATCAAAATCGTTACACCGGATTATATTGCGTTTAATGGCTGG
AACCTGATTACCATTACAAACAATCGTAGCAAAGGCAGCATTGTGTATGTGAACGGTAGCAAAATTGAAGAGAAG
GATATTAGCAGCATCTGGAATACCGAAGTGGATGATCCGATTATCTTTCGCCTGAAAAACAATCGCGATACCCAG
GCGTTTACCCTGCTGGATCAGTTTAGCATTTATCGGAAAGAACTGAACCAGAACGAAGTGGTGAAACTGTATAAC
TACTACTTCAACAGCAACTACATTCGCGATATTTGGGGTAATCCGCTGCAGTACAACAAAAAATACTATCTGCAG
ACCCAGGACAAACCTGGTAAAGGTCTGATCCGCGAATATTGGAGCAGCTTTGGTTATGATTATGTGATTCTGAGC
GATAGCAAGACGATTACCTTTCCGAATAATATCCGTTATGGTGCCCTGTATAACGGCAGCAAAGTTCTGATCAAA
AATAGCAAAAAACTGGATGGTCTGGTGCGCAATAAAGATTTCATTCAGCTGGAAATCGATGGCTATAATATGGGT
ATTAGCGCAGATCGCTTTAACGAGGATACCAACTATATTGGCACCACCTATGGTACAACCCATGATCTGACCACC
GATTTTGAAATTATTCAGCGCCAAGAGAAATACCGCAATTATTGTCAGCTGAAAACCCCGTATAACATCTTTCAT
AAAAGCGGTCTGATGAGCACCGAAACCAGCAAACCGACCTTTCATGATTATCGTGACTGGGTTTATAGCAGCGCA
TGGTATTTTCAGAACTATGAAAATCTGAACCTGCGCAAACATACCAAAACCAACTGGTATTTTATCCCGAAAGAT
GAAGGTTGGGATGAAGATCTTGAAGTTCTGTTTCAGGGTCCGCATCATCACCACCATCACCATCATCATCAC

SEQ ID NO: 33 (Polypeptide Sequence of BoNT/X)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNT
NDLNIPSEPIMEADAIYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIP
LPLVSNGALTLSDNETIAYQENNNIVSNLQANLVIYGPGPDIANNATYGLYSTPISNGEG
TLSEVSFSPFYLKPFDESYGNYRSLVNIVNKFVKREFAPDPASTLMHELVHVTHNLYGIS
NRNFYYNFDTGKIETSRQQNSLIFEELLTFGGIDSKAISSLIIKKIIETAKNNYTTLISE
RLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLNTILFVLNESNLAQRFSILVR
KHYLKERPIDPIYVNILDDNSYSTLEGFNISSQGSNDFQGQLLESSYFEKIESNALRAFI
KICPRNGLLYNAIYRNSKNYLNNIDLEDKKTTSKTNVSYPCSLLNGCIEVENKDLFLISN
KDSLNDINLSEEKIKPETTVFFKDKLPPQDITLSNYDFTEANSIPSISQQNILERNEELY
EPIRNSLFEIKTIYVDKLTTFHFLEAQNIDESIDSSKIRVELTDSVDEALSNPNKVYSPF
KNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDVIDKSSDTLAIVPYIGPLLNI
GNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQVEAIVNNALDKRD
QKWAEVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNIDDKAK
IKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDK
FIKEKEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNL
GAEDGKIKDLSGTTSDINIGSDIELADGRENKAIKIKGSENSTIKIAMNKYLRFSATDNF
SISFWIKHPKPTNLLNNGIEYTLVENFNQRGWKISIQDSKLIWYLRDHNNSIKIVTPDYI
AFNGWNLITITNNRSKGSIVYVNGSKIEEKDISSIWNTEVDDPIIFRLKNNRDTQAFTLL
DQFSIYRKELNQNEVVKLYNYYFNSNYIRDIWGNPLQYNKKYYLQTQDKPGKGLIREYWS
SFGYDYVILSDSKTITFPNNIRYGALYNGSKVLIKNSKKLDGLVRNKDFIQLEIDGYNMG
ISADRFNEDTNYIGTTYGTTHDLTTDFEIIQRQEKYRNYCQLKTPYNIFHKSGLMSTETS
KPTFHDYRDWVYSSAWYFQNYENLNLRKHTKTNWYFIPKDEGWDED

SEQ ID NO: 34 (Polypeptide Sequence of Wild-Type BoNT/X-10HT)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNTNDLNIPSEPIMEADA
IYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIPLPLVSNGALTLSDNETIAYQENNNIVSNLQ
ANLVIYGPGPDIANNATYGLYSTPISNGEGTLSEVSFSPFYLKPFDESYGNYRSLVNIVNKFVKREFAPDPASTL
MHELVHVTHNLYGISNRNFYYNFDTGKIETSRQQNSLIFEELLTFGGIDSKAISSLIIKKIIETAKNNYTTLISE
RLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLNTILFVLNESNLAQRFSILVRKHYLKERPIDPIYVN
ILDDNSYSTLEGFNISSQGSNDFQGQLLESSYFEKIESNALRAFIKICPRNGLLYNAIYRNSKNYLNNIDLEDKK
TTSKTNVSYPCSLLNGCIEVENKDLFLISNKDSLNDINLSEEKIKPETTVFFKDKLPPQDITLSNYDFTEANSIP
SISQQNILERNEELYEPIRNSLFEIKTIYVDKLTTFHFLEAQNIDESIDSSKIRVELTDSVDEALSNPNKVYSPF

KNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDVIDKSSDTLAIVPYIGPLLNIGNDIRHGDFVGAIEL
AGITALLEYVPEFTIPILVGLEVIGGELAREQVEAIVNNALDKRDQKWAEVYNITKAQWWGTIHLQINTRLAHTY
KALSRQANAIKMNMEFQLANYKGNIDDKAKIKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQD
NLKNFDLETKKTLDKFIKEKEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNL
GAEDGKIKDLSGTTSDINIGSDIELADGRENKAIKIKGSENSTIKIAMNKYLRFSATDNFSISFWIKHPKPTNLL
NNGIEYTLVENFNQRGWKISIQDSKLIWYLRDHNNSIKIVTPDYIAFNGWNLITITNNRSKGSIVYVNGSKIEEK
DISSIWNTEVDDPIIFRLKNNRDTQAFTLLDQFSIYRKELNQNEVVKLYNYYFNSNYIRDIWGNPLQYNKKYYLQ
TQDKPGKGLIREYWSSFGYDYVILSDSKTITFPNNIRYGALYNGSKVLIKNSKKLDGLVRNKDFIQLEIDGYNMG
ISADRFNEDTNYIGTTYGTTHDLTTDFEIIQRQEKYRNYCQLKTPYNIFHKSGLMSTETSKPTFHDYRDWVYSSA
WYFQNYENLNLRKHTKTNWYFIPKDEGWDEDLEVLFQGPHHHHHHHHHH

SEQ ID NO: 35 (BoNT/A - UniProt P10845)

MPFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN
PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG
STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGY
GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPN
RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKA
KSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKV
LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT
GLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGEE
ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNG
KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEA
AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSG
AVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAK
VNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKA
MININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
VNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINI
GSKVNFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNN
EYTIINCMENNSGWKVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTIT
NNRLNNSKIYINGRLIDQKPISNLGNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELN
EKEIKDLYDNQSNSGILKDFWGDYLQYDKPYYMLNLYDPNKYVDVNNVGIRGYMYLKGPR
GSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYINVVVKNKEYRLATNASQA
GVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFHQFNNIAK
LVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPL

SEQ ID NO: 36 (BoNT/B - UniProt P10844)

MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN
KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG
DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH
FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY
GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV
DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN
IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA
YEEISKEHLAVYKIQMCKSVKAPGICIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSN
YIENDFPINELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPAIKKIFTDENTIFQY
LYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVND
FVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLI
PVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMY
KALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSV
SYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDL
SIYTNDTILIEMFNKYNSEILNNIILNLRYKDNNLIDLSGYGAKVEVYDGVELNDKNQFK
LTSSANSKIRVTQNQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNS
GWKISIRGNRIIWTLIDINGKTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYING
KLESNTDIKDIREVIANGEIIFKLDGDIDRTQFIWMKYFSIFNTELSQSNIEERYKIQSY

SEYLKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGEILTRSKYNQNSKYINYRDLY
IGEKFIIRRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEEEKLFLAPISD
SDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS
KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE

SEQ ID NO: 37 (BoNT/C - UniProt P18640)

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNK
PPRVTSPKSGYYDPNYLSTDSDKDPFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNN
NTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTF
AAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYG
IAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSI
AKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVELYNELTQIFTE
FNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCHKAIDGRSLYNKTLDCRELLVKNTDLPFIGDISDVKTDIFLRK
DINEETEVIYYPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQN
VDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLM
WANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILL
EAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQF
NNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNIN
KFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVSEEGDVQ
LNPIFPFDFKLGSSGEDRGKVIVTQNENIVYNSMYESFSISFWIRINKWVSNLPGYTIID
SVKNNSGWSIGIISNFLVFTLKQNEDSEQSINFSYDISNNAPGYNKWFFVTVTNNMMGNM
KIYINGKLIDTIKVKELTGINFSKTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKEL
DGKDINILFNSLQYTNVVKDYWGNDLRYNKEYYMVNIDYLNRYMYANSRQIVFNTRRNNN
DFNEGYKIIIKRIRGNTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYA
IGLREQTKDINDNIIFQIQPMNNTYYYASQIFKSNFNGENISGICSIGTYRFRLGGDWYR
HNYLVPTVKQGNYASLLESTSTHWGFVPVSE

SEQ ID NO: 38 (BoNT/D - UniProt P19321)

MTWPVKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSK
PPRPTSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDS
STPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSN
PSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYG
INIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDI
AKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSE
VVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPA
LQKLSSESVVDLFTKVCLRLTKNSRDDSTCIKVKNNRLPYVADKDSISQEIFENKIITDE
TNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDYL
NSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWANE
VVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGFP
EFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHIN
YQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIR
ECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENTM
PFNIFSYTNNSLLKDIINEYFNSINDSKILSLQNKKNALVDTSGYNAEVRVGDNVQLNTI
YTNDFKLSSSGDKIIVNLNNNILYSAIYENSSVSFWIKISKDLTNSHNEYTIINSIEQNS
GWKLCIRNGNIEWILQDVNRKYKSLIFDYSESLSHTGYTNKWFFVTITNNIMGYMKLYIN
GELKQSQKIEDLDEVKLDKTIVFGIDENIDENQMLWIRDFNIFSKELSNEDINIVYEGQI
LRNVIKDYWGNPLKFDTEYYIINDNYIDRYIAPESNVLVLVQYPDRSKLYTGNPITIKSV
SDKNPYSRILNGDNIILHMLYNSRKYMIIRDTDTIYATQGGECSQNCVYALKLQSNLGNY
GIGIFSIKNIVSKNKYCSQIFSSFRENTMLLADIYKPWRFSFKNAYTPVAVTNYETKLLS
TSSFWKFISRDPGWVE

SEQ ID NO: 39 (BoNT/E - UniProt Q00496)

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS
LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP
DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHRFGS
IAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL
ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK
DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKL
SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG
IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA
PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS
IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS
IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK
NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE
SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKIN
KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF
NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI
SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII
WTFEDNRGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL
GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL
LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN
LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTMNF
KNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK

SEQ ID NO: 40 (BoNT/F - UniProt A7GBG3)

MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD
PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN
EHTPINEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVY
DPSNDGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGAR
GVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATR
LSRVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKF
KVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKG
LVEKIVKFCKSVIPRKGTKAPPRLCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLN
NNYRNNLDEVILDYNSETIPQISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFF
YLHAQKVPEGETNISLTSSIDTALSEESQVYTFFSSEFINTINKPVHAALFISWINQVIR
DFTTEATQKSTFDKIADISLVVPYVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELL
IPTILVFTIKSFIGSSENKNKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRK
EQMYQALQNQVDAIKTVIEYKYNNYTSDERNRLESEYNINNIREELNKKVSLAMENIERF
ITESSIFYLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNN
SIPFELSSYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIY
STNRNQFGIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDC
IRNNNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGN
SRIYINGNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETL
YSDEPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSN
TRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKL
IRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTS
SNGCFWSFISKEHGWQEN

SEQ ID NO: 41 (BoNT/G - UniProt Q60393)

MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN
ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG
NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH
SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY
GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA
NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL

AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY
EEISLEHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIANKDSFSKDLAKAETIAYN
TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD
SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK
GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI
PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI
KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI
NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS
IPFDLSLYTKDTILIQVFNNYISNISSNAILSLSYRGGRLIDSSGYGATMNVGSDVIFND
IGNGQFKLNNSENSNITAHQSKFVVYDSMFDNFSINFWVRTPKYNNNDIQTYLQNEYTII
SCIKNDSGWKVSIKGNRIIWTLIDVNAKSKSIFFEYSIKDNISDYINKWFSITITNDRLG
NANIYINGSLKKSEKILNLDRINSSNDIDFKLINCTDTTKFVWIKDFNIFGRELNATEVS
SLYWIQSSTNTLKDFWGNPLRYDTQYYLFNQGMQNIYIKYFSKASMGETAPRTNFNNAAI
NYQNLYLGLRFIIKKASNSRNINNDNIVREGDYIYLNIDNISDESYRVYVLVNSKEIQTQ
LFLAPINDDPTFYDVLQIKKYYEKTTYNCQILCEKDTKTFGLFGIGKFVKDYGYVWDTYD
NYFCISQWYLRRISENINKLRLGCNWQFIPVDEGWTE

SEQ ID NO: 42 (TeNT – UniProt P04958)

MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN
PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN
SYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN
KNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH
GLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDYK
AIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFTE
IELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVNT
NAFRNVDGSGLVSKLIGLCKKIIPPTNIRENLYNRTASLTDLGGELCIKIKNEDLTFIAE
KNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTKGIPYAP
EYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYFPSVI
SKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQGYEGN
FIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWIEVYK
LVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEINNLKN
KLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANSKFIG
ITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDVILKKSTILNLDINNDIISDIS
GFNSSVITYPDAQLVPGINGKAIHLVNNESSEVIVHKAMDIEYNDMFNNFTVSFWLRVPK
VSASHLEQYGTNEYSIISSMKKHSLSIGSGWSVSLKGNNLIWTLKDSAGEVRQITFRDLP
DKFNAYLANKWVFITITNDRLSSANLYINGVLMGSAEITGLGAIREDNNITLKLDRCNNN
NQYVSIDKFRIFCKALNPKEIEKLYTSYLSITFLRDFWGNPLRYDTEYYLIPVASSSKDV
QLKNITDYMYLTNAPSYTNGKLNIYYRRLYNGLKFIIKRYTPNNEIDSFVKSGDFIKLYV
SYNNNEHIVGYPKDGNAFNNLDRILRVGYNAPGIPLYKKMEAVKLRDLKTYSVQLKLYDD
KNASLGLVGTHNGQIGNDPNRDILIASNWYFNHLKDKILGCDWYFVPTDEGWTND

SEQ ID NO: 43 (Nucleotide Sequence of LH_N/A1 with an EK Cleavage Site)

ATGGAGTTCGTTAACAAACAGTTCAACTATAAAGACCCAGTTAACGGTGTTGACATTGCTTACATCAAAATCCCG
AACGCTGGCCAGATGCAGCCGGTAAAGGCATTCAAAATCCACAACAAAATCTGGGTTATCCCGGAACGTGATACC
TTTACTAACCCGGAAGAAGGTGACCTGAACCCGCCACCGGAAGCGAAACAGGTGCCGGTATCTTACTATGACTCC
ACCTACCTGTCTACCGATAACGAAAAGGACAACTACCTGAAAGGTGTTACTAAACTGTTCGAGCGTATTTACTCC
ACCGACCTGGGCCGTATGCTGCTGACTAGCATCGTTCGCGGTATCCCGTTCTGGGGCGGTTCTACCATCGATACC
GAACTGAAAGTAATCGACACTAACTGCATCAACGTTATTCAGCCGGACGGTTCCTATCGTTCCGAAGAACTGAAC
CTGGTGATCATCGGCCCGTCTGCTGATATCATCCAGTTCGAGTGTAAGAGCTTTGGTCACGAAGTTCTGAACCTC
ACCCGTAACGGCTACGGTTCCACTCAGTACATCCGTTTCTCTCCGGACTTCACCTTCGGTTTTGAAGAATCCCTG
GAAGTAGACACGAACCCACTGCTGGGCGCTGGTAAATTCGCAACTGATCCTGCGGTTACCCTGGCTCACGAACTG
ATTCATGCAGGCCACCGCCTGTACGGTATCGCCATCAATCCGAACCGTGTCTTCAAAGTTAACACCAACGCGTAT
TACGAGATGTCCGGTCTGGAAGTTAGCTTCGAAGAACTGCGTACTTTTGGCGGTCACGACGCTAAATTCATCGAC
TCTCTGCAAGAAAACGAGTTCCGTCTGTACTACTATAACAAGTTCAAAGATATCGCATCCACCCTGAACAAAGCG
AAATCCATCGTGGGTACCACTGCTTCTCTCCAGTACATGAAGAACGTTTTTAAAGAAAAATACCTGCTCAGCGAA
GACACCTCCGGCAAATTCTCTGTAGACAAGTTGAAATTCGATAAACTTTACAAAATGCTGACTGAAATTTACACC

GAAGACAACTTCGTTAAGTTCTTTAAAGTTCTGAACCGCAAAACCTATCTGAACTTCGACAAGGCAGTATTCAAA
ATCAACATCGTGCCGAAAGTTAACTACACTATCTACGATGGTTTCAACCTGCGTAACACCAACCTGGCTGCTAAT
TTTAACGGCCAGAACACGGAAATCAACAACATGAACTTCACAAAACTGAAAAACTTCACTGGTCTGTTCGAGTTT
TACAAGCTGCTGTGCGTCGACGGCATCATTACCTCCAAAACTAAATCTGACGATGACGATAAAAACAAAGCGCTG
AACCTGCAGTGTATCAAGGTTAACAACTGGGATTTATTCTTCAGCCCGAGTGAAGACAACTTCACCAACGACCTG
AACAAAGGTGAAGAAATCACCTCAGATACTAACATCGAAGCAGCCGAAGAAAACATCTCGCTGGACCTGATCCAG
CAGTACTACCTGACCTTTAATTTCGACAACGAGCCGGAAAACATTTCTATCGAAAACCTGAGCTCTGATATCATC
GGCCAGCTGGAACTGATGCCGAACATCGAACGTTTCCCAAACGGTAAAAAGTACGAGCTGGACAAATATACCATG
TTCCACTACCTGCGCGCGCAGGAATTTGAACACGGCAAATCCCGTATCGCACTGACTAACTCCGTTAACGAAGCT
CTGCTCAACCCGTCCCGTGTATACACCTTCTTCTCTAGCGACTACGTGAAAAAGGTCAACAAAGCGACTGAAGCT
GCAATGTTCTTGGGTTGGGTTGAACAGCTTGTTTATGATTTTACCGACGAGACGTCCGAAGTATCTACTACCGAC
AAAATTGCGGATATCACTATCATCATCCCGTACATCGGTCCGGCTCTGAACATTGGCAACATGCTGTACAAAGAC
GACTTCGTTGGCGCACTGATCTTCTCCGGTGCGGTGATCCTGCTGGAGTTCATCCCGGAAATCGCCATCCCGGTA
CTGGGCACCTTTGCTCTGGTTTCTTACATTGCAAACAAGGTTCTGACTGTACAAACCATCGACAACGCGCTGAGC
AAACGTAACGAAAAATGGGATGAAGTTTACAAATATATCGTGACCAACTGGCTGGCTAAGGTTAATACTCAGATC
GACCTCATCCGCAAAAAAATGAAAGAAGCACTGGAAAACCAGGCGGAAGCTACCAAGGCAATCATTAACTACCAG
TACAACCAGTACACCGAGGAAGAAAAAAACAACATCAACTTCAACATCGACGATCTGTCCTCTAAACTGAACGAA
TCCATCAACAAAGCTATGATCAACATCAACAAGTTCCTGAACCAGTGCTCTGTAAGCTATCTGATGAACTCCATG
ATCCCGTACGGTGTTAAACGTCTGGAGGACTTCGATGCGTCTCTGAAAGACGCCCTGCTGAAATACATTTACGAC
AACCGTGGCACTCTGATCGGTCAGGTTGATCGTCTGAAGGACAAAGTGAACAATACCTTATCGACCGACATCCCT
TTTCAGCTCAGTAAATATGTCGATAACCAACGCCTTTTGTCCACTCTAGAAGCACACCATCATCACCACCATCAC
CATCACCAT

SEQ ID NO: 44 (Polypeptide Sequence of LH_N/A1 with an EK Cleavage Site)

MEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQVPVSYYDS
TYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVIQPDGSYRSEELN
LVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHEL
IHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKA
KSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFK
INIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVDGIITSKTKSDDDDKNKAL
NLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDII
GQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEA
AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPV
LGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQ
YNQYTEEEKNNINFNIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYD
NRGTLIGQVDRLKDKVNNTLSTDIPFQLSKYVDNQRLLSTLEAHHHHHHHHHH

SEQ ID NO: 45 (Nucleotide Sequence of BoNT with a Native A1 Loop)

ATGAAACTGGAAATCAACAAATTCAACTACAACGATCCGATCGATGGCATTAATGTTATTACCATGCGTCCGCCT
CGTCATAGCGATAAAATCAATAAAGGTAAAGGTCCGTTCAAAGCCTTTCAGGTGATTAAAAACATTTGGATTGTG
CCGGAACGCTACAACTTTACCAATAATACCAACGATCTGAACATTCCGAGCGAACCGATTATGGAAGCAGATGCC
ATTTATAACCCGAACTATCTGAATACCCCGAGCGAAAAAGATGAATTTCTGCAGGGTGTTATCAAAGTGCTGGAA
CGCATTAAAAGCAAACCGGAAGGTGAAAAACTGCTGGAACTGATTAGCAGCAGCATTCCGCTGCCGCTGGTTAGC
AATGGTGCACTGACCCTGAGCGATAATGAAACCATTGCATATCAAGAGAACAACAACATTGTGAGCAATCTGCAG
GCAAACCTGGTTATTTATGGTCCGGGTCCTGATATTGCAAATAATGCAACCTATGGTCTGTATAGCACCCCGATT
AGTAATGGTGAAGGTACACTGAGCGAAGTTAGCTTTAGCCCGTTTTATCTGAAACCGTTTGATGAAAGCTATGGC
AATTATCGTAGCCTGGTGAATATCGTGAACAAATTCGTGAAACGTGAATTTGCACCTGATCCGGCAAGCACCCTG
ATGCATGAACTGGTTCATGTTACCCATAATCTGTATGGTATTAGCAACCGCAACTTCTACTATAACTTTGACACC
GGCAAAATTGAAACCAGCCGTCAGCAGAATAGCCTGATTTTTGAAGAACTGCTGACCTTTGGTGGCATTGATAGC
AAAGCAATTAGCAGCCTGATCATCAAGAAAATTATCGAAACCGCCAAGAACAACTATACCACGCTGATTAGCGAA
CGCCTGAATACCGTTACCGTTGAAAATGATCTGCTGAAATATATCAAAAACAAAATCCCGGTTCAGGGTCGTCTG
GGTAACTTTAAACTGGATACCGCAGAATTCGAGAAAAAGCTGAATACCATTCTGTTTGTGCTGAACGAAAGCAAT
CTGGCACAGCGTTTTAGCATTCTGGTTCGTAAACATTACCTGAAAGAACGTCCGATTGATCCGATTTATGTGAAC
ATTCTGGATGACAATAGCTACAGCACCCTGGAAGGTTTTAACATTAGCAGTCAGGGTAGCAATGATTTCCAAGGT
CAGCTGCTGGAAAGCAGCTATTTTGAAAAAATTGAAAGCAATGCCCTGCGTGCCTTCATTAAAATCTGTCCGCGT
AATGGTCTGCTGTATAATGCCATTTATCGCAACAGCAAAAATCTGGAAGTTCTGTTTCAGGGTCCGCATCATCAC
CACCATCACCATCATCATCACCTGGAAGTGTTATTTCAGGGACCGTATCTGAATAACATTGATCTGGAAGATAAA
AAGACCACGAGCAAAACCAATGTTAGCTATCCGTGTAGCCTGCTGAATGGTTGTATTGAAGTTGAAAACAAAGAC

CTGTTCCTGATTAGCAACAAAGATAGCCTGAACGATATTAACCTGAGCGAAGAAAAAATCAAACCGGAAACCACC
GTGTTCTTCAAAGATAAACTGCCTCCGCAGGATATTACGCTGAGCAATTATGATTTTACCGAAGCCAATAGCATT
CCGAGCATTAGCCAGCAGAACATTCTGGAACGTAATGAAGAACTGTATGAACCGATTCGCAATAGCCTGTTTGAA
ATCAAAACCATCTATGTGGATAAGCTGACCACCTTTCATTTTCTGGAAGCCCAGAATATTGATGAGAGCATTGAT
AGCAGCAAAATTCGTGTTGAACTGACCGATAGCGTTGATGAAGCACTGAGCAATCCGAATAAAGTTTATAGCCCG
TTCAAGAACATGAGCAACACCATTAATAGCATTGAAACCGGTATTACCAGCACCTACATCTTTTATCAGTGGCTG
CGTAGCATCGTGAAAGATTTTAGTGATGAAACCGGCAAAATCGACGTGATTGATAAAAGCAGCGATACCCTGGCA
ATTGTTCCGTATATTGGTCCGCTGCTGAATATTGGTAATGATATTCGTCATGGCGATTTTGTGGGTGCAATTGAA
CTGGCAGGCATTACCGCACTGCTGGAATATGTTCCGGAATTTACCATTCCGATTCTGGTTGGTCTGGAAGTGATT
GGTGGCGAACTGGCACGTGAACAGGTTGAAGCAATTGTTAATAATGCCCTGGATAAACGCGATCAGAAATGGGCA
GAAGTTTACAATATTACCAAAGCACAGTGGTGGGGCACCATTCATTTACAGATTAATACCCGTCTGGCCCATACC
TATAAAGCCCTGAGCCGTCAGGCAAATGCCATTAAAATGAATATGGAATTTCAGCTGGCCAACTACAAAGGCAAC
ATTGATGATAAAGCCAAGATCAAAAACGCCATCAGCGAAACCGAAATTCTGCTGAACAAAAGCGTTGAACAGGCC
ATGAAAAACACCGAGAAATTCATGATTAAACTGAGCAACAGCTACCTGACCAAAGAAATGATTCCGAAAGTTCAG
GACAACCTGAAAAACTTTGACCTGGAAACCAAAAAAACCCTGGACAAGTTCATCAAAGAGAAAGAAGATATCCTG
GGCACCAATCTGAGCAGCAGCCTGCGTCGTAAAGTTAGCATTCGTCTGAATAAAAACATTGCCTTCGACATCAAC
GATATCCCGTTTAGCGAATTTGATGATCTGATCAACCAGTACAAAAACGAGATCGAAGATTATGAAGTGCTGAAT
CTGGGTGCAGAAGATGGCAAAATCAAAGATCTGAGCGGTACAACCAGCGATATCAATATTGGTTCAGATATCGAA
CTGGCCGATGGTCGTGAAAATAAAGCCATTAAGATTAAAGGCAGCGAGAACAGCACCATCAAAATTGCAATGAAC
AAATATCTGCGTTTTAGCGCCACCGATAACTTTAGCATTAGCTTTTGGATCAAACATCCGAAACCGACCAATCTG
CTTAATAACGGTATTGAATATACCCTGGTCGAGAACTTTAATCAGCGTGGTTGGAAAATTAGCATCCAGGATAGC
AAACTGATTTGGTATCTGCGCGATCACAATAACAGCATCAAAATCGTTACACCGGATTATATTGCGTTTAATGGC
TGGAACCTGATCACCATTACGAATAATCGTAGCAAAGGCAGCATCGTGTATGTGAATGGTAGCAAAATTGAAGAG
AAGGACATTAGCAGCATTTGGAATACCGAAGTGGATGATCCGATTATCTTCCGCCTGAAAAATAACCGTGATACC
CAGGCATTTACCCTGCTGGATCAGTTTAGCATTTATCGGAAAGAACTGAACCAGAACGAAGTGGTGAAACTGTAT
AACTACTACTTCAACAGCAACTACATTCGCGATATTTGGGGTAATCCGCTGCAGTACAACAAAAAATACTATCTG
CAGACCCAGGACAAACCTGGTAAAGGTCTGATCCGCGAATATTGGAGCAGCTTTGGTTATGATTATGTGATTCTG
AGCGATAGCAAGACGATTACCTTTCCGAACAATATCCGTTATGGTGCCCTGTATAACGGTAGCAAAGTTCTGATC
AAGAACAGCAAGAAATTAGATGGTCTGGTGCGCAATAAAGATTTCATTCAGCTGGAAATCGATGGCTATAATATG
GGTATTAGCGCAGATCGCTTTAACGAGGATACCAACTATATTGGCACCACCTATGGTACAACCCATGATCTGACC
ACCGATTTTGAAATTATTCAGCGCCAAGAGAAATACCGCAATTATTGTCAGCTGAAAACCCCGTATAACATCTTT
CATAAAAGCGGTCTGATGAGCACCGAAACCAGCAAACCGACCTTCCATGATTATCGCGATTGGGTTTATAGCAGC
GCATGGTATTTTCAGAACTATGAAAATCTGAACCTGCGCAAACATACCAAAACCAACTGGTATTTTATCCCGAAA
GATGAAGGTTGGGACGAAGAT

SEQ ID NO: 46 (Polypeptide Sequence of BoNT with a Native A1 Loop)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNTNDLNIPSEPIMEADA
IYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIPLPLVSNGALTLSDNETIAYQENNNIVSNLQ
ANLVIYGPGPDIANNATYGLYSTPISNGEGTLSEVSFSPFYLKPFDESYGNYRSLVNIVNKFVKREFAPDPASTL
MHELVHVTHNLYGISNRNFYYNFDTGKIETSRQQNSLIFEELLTFGGIDSKAISSLIIKKIIETAKNNYTTLISE
RLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLNTILFVLNESNLAQRFSILVRKHYLKERPIDPIYVN
ILDDNSYSTLEGFNISSQGSNDFQGQLLESSYFEKIESNALRAFIKICPRNGLLYNAIYRNSKNLEVLFQGPHHH
HHHHHHLEVLFQGPYLNNIDLEDKKTTSKTNVSYPCSLLNGCIEVENKDLFLISNKDSLNDINLSEEKIKPETT
VFFKDKLPPQDITLSNYDFTEANSIPSISQQNILERNEELYEPIRNSLFEIKTIYVDKLTTFHFLEAQNIDESID
SSKIRVELTDSVDEALSNPNKVYSPFKNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDVIDKSSDTLA
IVPYIGPLLNIGNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQVEAIVNNALDKRDQKWA
EVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNIDDKAKIKNAISETEILLNKSVEQA
MKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDKFIKEKEDILGTNLSSSLRRKVSIRLNKNIAFDIN
DIPFSEFDDLINQYKNEIEDYEVLNLGAEDGKIKDLSGTTSDINIGSDIELADGRENKAIKIKGSENSTIKIAMN
KYLRFSATDNFSISFWIKHPKPTNLLNNGIEYTLVENFNQRGWKISIQDSKLIWYLRDHNNSIKIVTPDYIAFNG
WNLITITNNRSKGSIVYVNGSKIEEKDISSIWNTEVDDPIIFRLKNNRDTQAFTLLDQFSIYRKELNQNEVVKLY
NYYFNSNYIRDIWGNPLQYNKKYYLQTQDKPGKGLIREYWSSFGYDYVILSDSKTITFPNNIRYGALYNGSKVLI
KNSKKLDGLVRNKDFIQLEIDGYNMGISADRFNEDTNYIGTTYGTTHDLTTDFEIIQRQEKYRNYCQLKTPYNIF
HKSGLMSTETSKPTFHDYRDWVYSSAWYFQNYENLNLRKHTKTNWYFIPKDEGWDED

SEQ ID NO: 47 (Polypeptide Sequence of Trypsin)

MHPLLILAFVGAAVAFPSDDDDKIVGGYTCAENSVPYQVSLNAGYHFCGGSLINDQWVVS
AAHCYQYHIQVRLGEYNIDVLEGGEQFIDASKIIRHPKYSSWTLDNDILLIKLSTPAVIN

ARVSTLALPSACASGSTECLISGWGNTLSSGVNYPDLLQCLEAPLLSHADCEASYPGEIT
NNMICAGFLEGGKDSCQGDSGGPVACNGQLQGIVSWGYGCAQKGKPGVYTKVCNYVDWIQ
ETIAANS

SEQ ID NO: 48 (Polypeptide Sequence of Lys-C)

G V S G S C N I D V V C P E G N G H R D V I R S V A A Y S K Q G T M W C T G
S L V N N S A N D K K M Y F L T A N H C G M T T A A I A S S M V V Y W N Y Q
N S T C R A P G S S S S G A N G D G S L A Q S Q T G A V V R A T N A A S D F
T L L E L N T A A N P A Y N L F W A G W D R R D Q N F A G A T A I H H P N V
A E K R I S H S T V A T E I S G Y N G A T G T S H L H V F W Q A S G G V T E
P G S S G S P I Y S P E K R V L G Q L H 211 G G P S S C S A T G A D R S D Y
Y G R V F T S W T G G G T S A T R L S D W L D A A G T G A Q F I D G L D S T
G T P P V

SEQ ID NO: 49 (Polypeptide Sequence of Enterokinase Light Chain)

IVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGLH
MASNLTSPQIETRLIDQIVINRHYNKRRKNNDIAMMHLEMKVNYTDYIQPICLPEENQVF
PPGRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYDAGGV
DSCQGDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLH

SEQ ID NO: 50 (Polypeptide Sequence of Factor Xa Heavy Chain)

IVGGRDCAEGECPWQALLVNEENEGFCGGTILNEFYVLTAAHCLHQAKRFTVRVGDRNTEQEEGNEMAHEVEMTV
KHSRFVKETYDFDIAVLRLKTPIRFRRNVAPACLPEKDWAEATLMTQKTGIVSGFGRTHEKGRLSSTLKMLEVPY
VDRSTCKLSSSFTITPNMFCAGYDTQPEDACQGDSGGPHVTRFKDTYFVTGIVSWGEGCARKGKFGVYTKVSNFL
KWIDKIMKARAGAAGSRGHSEAPATWTVPPPLPL

SEQ ID NO: 51 (Polypeptide Sequence of Factor Xa Light Chain)

ANSFLEEVKQGNLERECLEEACSLEEAREVFEDAEQTDEFWSKYKDGDQCEGHPCLNQGHCKDGIGDYTCTCAEG
FEGKNCEFSTREICSLDNGGCDQFCREERSEVRCSCAHGYVLGDDSKSCVSTERFPCGKFTQGRS

SEQ ID NO: 52 (Polypeptide Sequence of Anthrax Toxin Protective Antigen - NCBI Ref Seq: NP_052806)

1 mkkrkvlipl malstilvss tgnleviqae vkqenrllne sesssqgllg yyfsdlnfqa
61 pmvvtssttg dlsipssele nipsenqyfq saiwsgfikv kksdeytfat sadnhvtmwv
121 ddqevinkas nsnkirlekg rlyqikiqyq renptekgld fklywtdsqn kkevissdnl
181 qlpelkqkss nsrkkrstsa gptvpdrdnd gipdsleveg ytvdvknkrt flspwisnih
241 ekkgltkyks spekwstasd pysdfekvtg ridknvspea rhplvaaypi vhvdmeniil
301 sknedqstqn tdsqtrtisk ntstsrthts evhgnaevha sffdiggsvs agfsnsnsst
361 vaidhslsla gertwaetmg lntadtarln aniryvntgt apiynvlptt slvlgknqtl
421 atikakenql sqilapnnyy psknlapial naqddfsstp itmnynqfle lektkqlrld
481 tdqvygniat ynfengrvrv dtgsnwsevl pqiqettari ifngkdlnlv erriaavnps
541 dplettkpdm tlkealkiaf gfnepngnlq yqgkditefd fnfdqqtsqn iknqlaelna
601 tniytvldki klnakmnili rdkrfhydrn niavgadesv vkeahrevin sstegllllni
661 dkdirkilsg yiveiedteg lkevindryd mlnisslrqd gktfidfkky ndklplyisn
721 pnykvnvyav tkentiinps engdtstngi kkilifskkg yeig

SEQ ID NO: 53 (LH_N/A with a C1 Activation Loop)

MPFVNKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN
PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGG
STIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGY
GSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPN
RVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKA
KSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKV
LNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT

GLFEFYKLLCHKA**IDGR**SLYNKTLDCIKVNNWDLFFSPSEDNFTNDLNKGEE
ITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNG
KKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEA
AMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSG
AVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAK
VNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKA
MININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
VNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK

SEQ ID NO: 54 (LH_N/B with a C1 Activation Loop)

MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPEDFN
KSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMIINGIPYLG
DRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNH
FASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYFSDPALILMHELIHVLHGLY
GIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIV
DRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYKSLMFGFTETN
IAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQA
YEEISKEHLAVYKIQMCHKA**IDGR**SLYNKTLDCIDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPI
NELILDTDLISKIELPSENTESLTDFNVDVPVYEKQPAIKKIFTDENTIFQY
LYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVND
FVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAFEIAGASILLEFIPELLI
PVVGAFLLESYIDNKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMY
KALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLNEGINQAIDNINNFINGCSV
SYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDL
SIYTNDTILIEMFNKYNS

SEQ ID NO: 55 (LH_N/D with a C1 Activation Loop)

MTWPVKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSK
PPRPTSKYQSYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDS
STPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSN
PSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFCMDPVIALMHELTHSLHQLYG
INIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDI
AKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSLYSDLTNVMSE
VVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPA
LQKLSSESVVDLFTKVCHKA**IDGR**SLYNKTLDCIKVKNNRLPYVADKDSISQEIFENKIITDE
TNVQNYSDKFSLDESILDGQVPINPEIVDPLLPNVNMEPLNLPGEEIVFYDDITKYVDYL
NSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLPSLAEKVNKGVQAGLFLNWANE
VVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFATAGVAFLLEGFP
EFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHIN
YQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKISEAMNNINKFIR
ECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENTM
PFNIFSYTNNSLLKDIINEYFN

SEQ ID NO: 56 (LH_N/E with a C1 Activation Loop)

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS
LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP
DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHRFGS
IAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL
ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK
DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKFQVKCRQTYIGQYKYFKL
SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRF
CHKA**IDGR**SLYNKTLDCIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLS
DEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS
IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS
IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK
NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE
SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKIN
KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF
NKFFK

SEQ ID NO: 57 (LH_N/F with a C1 Activation Loop)

MPVVINSFNYNDPVNDDTILYMQIPYEEKSKKYYKAFEIMRNVWIIPERNTIGTDPSDFD
PPASLENGSSAYYDPNYLTTDAEKDRYLKTTIKLFKRINSNPAGEVLLQEISYAKPYLGN
EHTPINEFHPVTRTTSVNIKSSTNVKSSIILNLLVLGAGPDIFENSSYPVRKLMDSGGVY
DPSNDGFGSINIVTFSPEYEYTFNDISGGYNSSTESFIADPAISLAHELIHALHGLYGAR
GVTYKETIKVKQAPLMIAEKPIRLEEFLTFGGQDLNIITSAMKEKIYNNLLANYEKIATR
LSRVNSAPPEYDINEYKDYFQWKYGLDKNADGSYTVNENKFNEIYKKLYSFTEIDLANKF
KVKCRNTYFIKYGFLKVPNLLDDDIYTVSEGFNIGNLAVNNRGQNIKLNPKIIDSIPDKG
LVEKIVKFCHKA**IDGR**SLYNKTLDCIRVNNRELFFVASESSYNENDINTPKEIDDTTNLN
NNYRNNLDEVILDYNSETIPQISNQTLNTLVQDDSYVPRYDSNGTSEIEEHNVVDLNVFF
YLHAQKVPEGETNISLTSSIDTALSEESQVYTFFSSEFINTINKPVHAALFISWINQVIR
DFTTEATQKSTFDKIADISLVVPYVGLALNIGNEVQKENFKEAFELLGAGILLEFVPELL
IPTILVFTIKSFIGSSENKNKIIKAINNSLMERETKWKEIYSWIVSNWLTRINTQFNKRK
EQMYQALQNQVDAIKTVIEYKYNNYTSDERNRLESEYNINNIREELNKKVSLAMENIERF
ITESSIFYLMKLINEAKVSKLREYDEGVKEYLLDYISEHRSILGNSVQELNDLVTSTLNN
SIPFELSSYTNDKILILYFNKLYK

SEQ ID NO: 58 (LH_N/G with a C1 Activation Loop)

MPVNIKXFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPDQFN
ASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMIVDAIPYLG
NASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNGH
SPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYFADPALTLMHELIHVLHGLY
GIKISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIA
NRLNIVSSAQGSGIDISLYKQIYKNKYDFVEDPNGKYSVDKDKFDKLYKALMFGFTETNL
AGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAY
EEISLEHLVIYRIAMCHKA**IDGR**SLYNKTLDCIIVNNEDLFFIANKDSFSKDLAKAETIAYN
TQNNTIENNFSIDQLILDNDLSSGIDLPNENTEPFTNFDDIDIPVYIKQSALKKIFVDGD
SLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVK
GVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAFEIGGAAILMEFI
PELIVPIVGFFTLESYVGNKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTI
KERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNINIDFNDIDFKLNQSINLAINNIDDFI
NQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVNILKSKVNRHLKDS
IPFDLSLYTKDTILIQVFNNYIS

SEQ ID NO: 59 (LH_N/TeNT with a C1 Activation Loop)

MPITINNFRYSDPVNNDTIIMMEPPYCKGLDIYYKAFKITDRIWIVPERYEFGTKPEDFN
PPSSLIEGASEYYDPNYLRTDSDKDRFLQTMVKLFNRIKNNVAGEALLDKIINAIPYLGN
SYSLLDKFDTNSNSVSFNLLEQDPSGATTKSAMLTNLIIFGPGPVLNKNEVRGIVLRVDN
KNYFPCRDGFGSIMQMAFCPEYVPTFDNVIENITSLTIGKSKYFQDPALLLMHELIHVLH
GLYGMQVSSHEIIPSKQEIYMQHTYPISAEELFTFGGQDANLISIDIKNDLYEKTLNDYK
AIANKLSQVTSCNDPNIDIDSYKQIYQQKYQFDKDSNGQYIVNEDKFQILYNSIMYGFTE
IELGKKFNIKTRLSYFSMNHDPVKIPNLLDDTIYNDTEGFNIESKDLKSEYKGQNMRVNT
NAFRNVDGSGLVSKLIGLCHKA**IDGR**SLYNKTLDCIKIKNEDLTFIAE
KNSFSEEPFQDEIVSYNTKNKPLNFNYSLDKIIVDYNLQSKITLPNDRTTPVTKGIPYAP
EYKSNAASTIEIHNIDDNTIYQYLYAQKSPTTLQRITMTNSVDDALINSTKIYSYFPSVI
SKVNQGAQGILFLQWVRDIIDDFTNESSQKTTIDKISDVSTIVPYIGPALNIVKQGYEGN
FIGALETTGVVLLLEYIPEITLPVIAALSIAESSTQKEKIIKTIDNFLEKRYEKWIEVYK
LVKAKWLGTVNTQFQKRSYQMYRSLEYQVDAIKKIIDYEYKIYSGPDKEQIADEINNLKN
KLEEKANKAMININIFMRESSRSFLVNQMINEAKKQLLEFDTQSKNILMQYIKANSKFIG
ITELKKLESKINKVFSTPIPFSYSKNLDCWVDNEEDIDV

SEQ ID NO: 60 (LH_N/X with a C1 Activation Loop)

MKLEINKFNYNDPIDGINVITMRPPRHSDKINKGKGPFKAFQVIKNIWIVPERYNFTNNT
NDLNIPSEPIMEADAIYNPNYLNTPSEKDEFLQGVIKVLERIKSKPEGEKLLELISSSIP
LPLVSNGALTLSDNETIAYQENNNIVSNLQANLVIYGPGPDIANNATYGLYSTPISNGEG
TLSEVSFSPFYLKPFDESYGNYRSLVNIVNKFVKREFAPDPASTLMHELVHVTHNLYGIS
NRNFYYNFDTGKIETSRQQNSLIFEELLTFGGIDSKAISSLIIKKIIETAKNNYTTLISE
RLNTVTVENDLLKYIKNKIPVQGRLGNFKLDTAEFEKKLNTILFVLNESNLAQRFSILVR
KHYLKERPIDPIYVNILDDNSYSTLEGFNISSQGSNDFQGQLLESSYFEKIESNALRAFI
KICHKA**IDGR**SLYNKTLDCIEVENKDLFLISN

KDSLNDINLSEEKIKPETTVFFKDKLPPQDITLSNYDFTEANSIPSISQQNILERNEELY
EPIRNSLFEIKTIYVDKLTTFHFLEAQNIDESIDSSKIRVELTDSVDEALSNPNKVYSPF
KNMSNTINSIETGITSTYIFYQWLRSIVKDFSDETGKIDVIDKSSDTLAIVPYIGPLLNI
GNDIRHGDFVGAIELAGITALLEYVPEFTIPILVGLEVIGGELAREQVEAIVNNALDKRD
QKWAEVYNITKAQWWGTIHLQINTRLAHTYKALSRQANAIKMNMEFQLANYKGNIDDKAK
IKNAISETEILLNKSVEQAMKNTEKFMIKLSNSYLTKEMIPKVQDNLKNFDLETKKTLDK
FIKEKEDILGTNLSSSLRRKVSIRLNKNIAFDINDIPFSEFDDLINQYKNEIEDYEVLNL
GAEDGKIKDLSGTTSDINIGSDIE

EXAMPLES

MATERIALS & METHODS

Materials

- 5 ml HiTrap Butyl HP (GE#: 28411005)
- 5 ml HiTrap Q HP (GE#: 17-1154-01)
- 5 ml HiTrap Phenyl HP column (GE# 17-5195-01)
- CHT Type II column (Biorad# 7324756)
- TrypZean (Sigma# T3568)
- Lys-C (Sigma# 000000011047825001)
- Enterokinase, light chain (NEB#P8070)
- Factor Xa (NEB#P8010)
- ACQUITY UPLC Protein BEH C4 Column (Waters# 186004495)

Protein purification

E. coli BL21 (DE3) or NiCo (DE3) (NEB) were utilized for protein expression. In general, bacteria were cultured at 37°C until induction, temperature was dropped to 16°C and protein expression was induced with 1mM IPTG overnight.

BoNT/AC with a C1 Loop (SEQ ID NO: 13)

Bacterial pellets were disrupted in lysis buffer (50 mM Tris-HCl pH=8) by sonication and clarified by centrifugation. Ammonium sulphate concentration was adjusted to 1.3M and target protein was captured using Butyl HP resin (GE). Fractions containing target protein were desalted and loaded onto Q HP resin (GE). Purified protein was activated overnight at 4°C with 6 µg/1mg of BoNT Factor Xa (NEB), followed by polishing using Phenyl HP resin (GE).

BoNT/E with a C1 loop (SEQ ID NO: 11)

Bacterial pellets were disrupted in lysis buffer (100mM sodium phosphate pH=7.8; 100mM NaCl) by sonication and clarified by centrifugation. Ammonium sulphate concentration was adjusted to 1.25 M and target protein was captured using Butyl HP resin (GE). Fractions containing target protein were desalted and loaded onto Q HP resin (GE). Purified protein was activated overnight at 4°C with either 5 µg/ 1mg of BoNT Factor Xa (NEB) or 80U/ml of Enterokinase (NEB), followed by polishing using CHT Type II resin (Biorad).

BoNT/X (SEQ ID NO: 5)

Bacterial pellets were disrupted in lysis buffer (50 mM Tris-HCl pH=8, 500mM NaCl) by sonication and clarified by centrifugation. Target protein was captured using HisTrap HP column (GE). Fractions containing target protein were desalted and loaded onto Q HP resin (GE). Purified protein was activated overnight at 4°C with either 5 µg/1mg of BoNT Factor Xa (NEB) or 80U/1mg of BoNT of Enterokinase (NEB), followed by polishing using 1ml HisTrap column (GE).

LC/MS

Samples were buffer exchanged into 50mM ammonium bicarbonate prior to analysis. Samples were either intact protein or reduced by incubation with 10mM DTT for 30 minutes at 37°C. Samples were tested using Waters Acquity H-Class UPLC system combined with Waters Xevo G2-XS QToF mass spectrometer.

- Mobile phase A 0.1% formic acid in water
- Mobile phase B 0.1% formic acid in acetonitrile
- Column: ACQUITY UPLC Protein BEH C4 (Waters)

EXAMPLE 1

The BoNT/C1 activation loop can be cleaved by multiple proteases

Inactive BoNT/C1 (0) (SEQ ID NO: 15) was incubated with a set of proteases: trypsin, Lys-C, enterokinase and Factor Xa. All proteases showed cleavage within the activation loop. BoNT/C1 (0) (SEQ ID NO: 15) was digested with enterokinase (EK) or factor Xa (FXa) overnight at 4°C and 25°C. In addition, BoNT/C1 (0) was trypsin digested over a 16hr timecourse at 20°C (Figure 2 A, B). All three proteases can cleave the BoNT/C1 activation loop and create a di-chain molecule when compared to protease untreated control. However, additional cleavage products were visible after tryptic and Lys-C digest.

EXAMPLE 2

Characterisation and improvement of BoNT/X proteolytic activation

Partially purified wild-type BoNT/X-10HT (SEQ ID NO: 34) was incubated overnight at 4°C with increasing amounts of trypsin (TrypZean) and Lys-C, as well as factor Xa (FXa) and enterokinase (EK).

Figure 3 shows that wild-type BoNT/X was completely degraded by both Lys-C (Figure 3A) and trypsin (Figure 3B, lanes 12-13 and 15-17). Notably, FXa and EK were unable to cleave the protein into a di-chain form (Figure 3B, lanes 18 and 19, respectively).

In an attempt to improve activation of BoNT/X, the BoNT/X activation loop was replaced by the BoNT/C1 activation loop (SEQ ID NO: 2) creating engineered BoNT protein SEQ ID NO: 5. The engineered BoNT was purified using several chromatographic steps and treated with enterokinase (EK) or factor Xa (FXa) to verify that the presence of BoNT/C loop allowed for production of a di-chain molecule. Surprisingly, Figure 4 shows that EK and FXa specifically cleave engineered BoNT/X into the di-chain form.

EXAMPLE 3

Characterisation and improvement of BoNT/E proteolytic activation

Wild-type BoNT/E was cleaved with Lys-C and trypsin (TryZean). Figure 5A shows that Lys-C inaccurately cleaves/degrades BoNT/E. Treatment with trypsin resulted in truncation of BoNT/E meaning an additional purification step was required to separate the full-length protein from the truncation product (Figure 5B).

In an attempt to improve activation of BoNT/E, the BoNT/E activation loop was replaced by the BoNT/C1 activation loop (SEQ ID NO: 2) creating engineered BoNT protein SEQ ID NO: 11. The engineered BoNT was purified using several chromatographic steps and treated with enterokinase (EK) or factor Xa (FXa) to verify that the presence of BoNT/C loop allowed for production of a di-chain molecule. Surprisingly, Figure 6 shows that EK and FXa specifically cleave engineered BoNT/E into the di-chain form.

EXAMPLE 4

Proteolytic activation of a BoNT/A1C1 chimera

The BoNT/C loop was introduced into a BoNT/A1C1 chimera (LH_N/A1 with a C1 H_C domain) to facilitate proteolytic protein activation. The BoNT/A1 activation loop of BoNT/A1C1 was replaced with a BoNT/C1 loop creating engineered BoNT protein SEQ ID NO: 13. The engineered BoNT was purified using several chromatographic steps and treated with factor Xa (FXa) to verify that the presence of BoNT/C loop allowed for production of a di-chain molecule. Figure 7A shows that FXa specifically cleaves engineered BoNT/A1C1 into the di-chain form. For comparison purposes, wild-type Met BoNT/A1 (commercially available from MetabiologicsA1080116) containing the A1 activation loop was incubated with FXa and EK. Figure 7B shows that FXa does not cleave the A1 activation loop, while EK cleaves with only minimal activity, and both FXa and EK results in formation of additional improper cleavage products (degradation products).

EXAMPLE 5

BoNTs containing the C1 loop retain SNARE cleavage activity

Rat primary cortical neurons were treated for 24h with BoNT/A1C1 (SEQ ID NO: 13) of Example 4 containing the BoNT/C1 activation loop and purified recombinant BoNT/C1 (SEQ ID NO: 17). SNARE-dependent glutamate release from the cells stimulated by potassium chloride was measured after incubation (Figure 8). These data confirm activity of clostridial neurotoxins modified to include the BoNT/C1 activation loop.

EXAMPLE 6

Factor Xa and Enterokinase cleave BoNT/C activation loop at the same site IDGR↓SL

Purified BoNT/E containing the BoNT/C1 loop (SEQ ID NO: 11) was activated by either enterokinase or factor Xa proteases and was incubated with 10mM DTT to reduce disulphide bridges and separate light and heavy chain. Liquid chromatography–mass spectrometry analysis of intact protein mass was performed on reduced and non-reduced engineered BoNT/E (SEQ ID NO: 11) samples to map the cleavage sites of both proteases. Both proteases cleaved BoNT/E to produce light and heavy chain of the same size indicating that both enterokinase and factor Xa cleave at the same site (Table 1).

Table 1. Comparison of predicted and measured masses of engineered BoNT/E (SEQ ID NO: 11)after cleavage at IDGR↓SL site within the C-loop. Heavy chain mass indicates cleavage at the predicted site by both enterokinase and factor Xa.

	predicted theoretical mass [Da]	**observed mass after EK cleavage [Da]**	**observed mass after FXa cleavage [Da]**
intact molecule	143952	143853 (Fig.9)	143850 (Fig.11)
light chain	47633	47518 (Fig.10)	47518 (Fig.12)
heavy chain	96337	96338 (Fig.10)	96335 (Fig. 12)

COMPARATIVE EXAMPLE 7

Insertion of a protease recognition site into an endogenous loop

The BoNT/C1 activation loop is the only BoNT activation loop that contains a naturally occurring cleavage site for site specific proteases FXa (and surprisingly EK) (*see* Figure 1). All other loops are cleaved by non-specific proteases such as trypsin or Lys-C. Cleavage by Lys-C and trypsin often leads to undesired protein truncation as the cleavage site is determined by protease accessibility rather than a specific recognition sequence.

Natural activation loops from different serotypes evolved to allow protease accessibility and process toxin into a di-chain form by *Clostridium*. Without wishing to be bound by theory, it is believed that mutating these loops to create a protease recognition site can lead to conformational changes, which could negatively affect cleavage efficiency.

To test this hypothesis, a polypeptide having the BoNT/A1 light chain and translocation domain (LH_N/A1) was modified to include an EK protease recognition sequence DDDDK (SEQ ID NO: 44). The efficiency of proteolytic cleavage of modified LH_N/A1 with EK was assessed and compared with cleavage of the wild-type A1 activation loop with Lys-C (note owing to an absence of EK recognition sites in the wild-type loop a direct comparison using EK is not possible). Figure 13 shows that cleavage of the modified loop is much less efficient than the wild-type loop.

EXAMPLE 8

Proteolytic activation of a BoNT/XA chimera

A BoNT/XA chimera containing the light chain and translocation domain of BoNT/X, the binding domain of BoNT/A1, and the BoNT/C1 activation loop was manufactured (SEQ ID NO: 7). Figure 14 shows that the di-chain form of the engineered BoNT/XA chimera was produced following activation with FXa.

EXAMPLE 9

Proteolytic activation of a BoNT/XB chimera

A BoNT/XB chimera containing the light chain and translocation domain of BoNT/X, the binding domain of BoNT/B, and the BoNT/C1 activation loop was manufactured (SEQ ID NO: 9). Figure 15 shows that the di-chain form of the engineered BoNT/XB chimera was produced following activation with FXa.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention. Although the present invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in biochemistry and biotechnology or related fields are intended to be within the scope of the following claims.

CLAIMS

1. A method for proteolytically processing a single-chain clostridial neurotoxin into a corresponding di-chain clostridial neurotoxin, the method comprising:

a. providing a single-chain clostridial neurotoxin; and

b. contacting the single-chain clostridial neurotoxin with enterokinase or factor Xa;

wherein the single-chain clostridial neurotoxin has an activation loop comprising the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1), wherein a = 1-10 and b = 4-15; and

wherein enterokinase or factor Xa hydrolyses a peptide bond of the activation loop thereby producing a di-chain clostridial neurotoxin.

2. The method according to claim 1, wherein a = 2-4.

3. The method according to claim 1 or 2, wherein a = 3.

4. The method according to any one of the preceding claims, wherein b = 6-10.

5. The method according to any one of the preceding claims, wherein b = 8.

6. The method according to any one of the preceding claims, wherein the activation loop comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

7. The method according to any one of the preceding claims, wherein the activation loop comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

8. The method according to any one of the preceding claims, wherein the activation loop comprises a polypeptide sequence having SEQ ID NO: 2 or SEQ ID NO: 3.

9. The method according to any one of the preceding claims, wherein the activation loop comprises SEQ ID NO: 2.

10. The method according to any one of the preceding claims, wherein the activation loop consists of SEQ ID NO: 2.

11. The method according to any one of the preceding claims, wherein the single-chain clostridial neurotoxin is contacted with enterokinase.

12. The method according to any one of the preceding claims, wherein the clostridial neurotoxin lacks a functional H_C domain of a clostridial neurotoxin.

13. The method according to any one of the preceding claims, wherein the clostridial neurotoxin is a retargeted clostridial neurotoxin comprising a non-clostridial Targeting Moiety (TM).

14. The method according to any one of the preceding claims, wherein the clostridial neurotoxin is a re-targeted clostridial neurotoxin comprising a TM, wherein the TM comprises an anthrax toxin protective antigen (PA) or a fragment thereof.

15. The method according to claim 14, wherein the clostridial neurotoxin has a polypeptide sequence with at least 70% sequence identity to a polypeptide comprising:

a. SEQ ID NO: 52 or a fragment thereof; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

16. The method according to any one of the preceding claims, wherein the clostridial neurotoxin is not BoNT/C1.

17. The method according to any one of the preceding claims, wherein the clostridial neurotoxin is selected from: BoNT/A, BoNT/B, BoNT/D, BoNT/E, BoNT/F, BoNT/G, BoNT/X, and TeNT.

18. The method according to any one of the preceding claims, wherein the clostridial neurotoxin is BoNT/X, BoNT/E, a chimeric BoNT or a hybrid BoNT.

19. The method according to any one of the preceding claims, wherein the single-chain clostridial neurotoxin:

a. is encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12; or
b. comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13.

20. A method for manufacturing an engineered clostridial neurotoxin, the method comprising:

a. identifying an endogenous activation loop of a clostridial neurotoxin, wherein the clostridial neurotoxin is characterised in that:

i. a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C; and/or

ii. the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C; and

b. replacing the endogenous activation loop with an exogenous activation loop thereby providing an engineered clostridial neurotoxin, wherein the exogenous activation loop comprises the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1), wherein a = 1-10 and b = 4-15.

21. The method according to claim 20, wherein the endogenous activation loop is one or more selected from SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70 or SEQ ID NO: 71.

22. The method according to claim 20 or 21, wherein a clostridial neurotoxin is identified as being suitable for use in the method by contacting the clostridial neurotoxin with trypsin or Lys-C and confirming hydrolysis of a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin.

23. The method according to any one of claims 20-22 further comprising contacting the engineered clostridial neurotoxin with enterokinase or factor Xa thereby producing a corresponding di-chain engineered clostridial neurotoxin.

24. The method according to any one of claims 20-23, wherein a = 2-4.

25. The method according to any one of claims 20-24, wherein a = 3.

26. The method according to any one of claims 20-25, wherein b = 6-10.

27. The method according to any one of claims 20-26, wherein b = 8.

28. The method according to any one of claims 20-27, wherein the exogenous activation loop comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

29. The method according to any one of claims 20-28, wherein the exogenous activation loop comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

30. The method according to any one of claims 20-29, wherein the exogenous activation loop comprises a polypeptide sequence having SEQ ID NO: 2 or SEQ ID NO: 3.

31. The method according to any one of claims 20-30, wherein the exogenous activation loop comprises SEQ ID NO: 2.

32. The method according to any one of claims 20-31, wherein the exogenous activation loop consists of SEQ ID NO: 2.

33. The method according to any one of claims 23-32, wherein the engineered clostridial neurotoxin is contacted with enterokinase.

34. The method according to any one of claims 20-33, wherein the clostridial neurotoxin lacks a functional H_C domain of a clostridial neurotoxin.

35. The method according to any one of claims 20-34, wherein the clostridial neurotoxin is a retargeted clostridial neurotoxin comprising a non-clostridial Targeting Moiety (TM).

36. The method according to any one of claims 20-35, wherein the clostridial neurotoxin is a re-targeted clostridial neurotoxin comprising a TM, wherein the TM comprises an anthrax toxin protective antigen (PA) or a fragment thereof.

37. The method according to claim 36, wherein the clostridial neurotoxin has a polypeptide sequence with at least 70% sequence identity to a polypeptide comprising:

a. SEQ ID NO: 52 or a fragment thereof; and
b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60.

38. The method according to any one of claims 20-37, wherein the clostridial neurotoxin is not BoNT/C1.

39. The method according to any one of claims 20-38, wherein the clostridial neurotoxin is selected from: BoNT/A, BoNT/B, BoNT/D, BoNT/E, BoNT/F, BoNT/G, BoNT/X, and TeNT.

40. The method according to any one of claims 20-39, wherein the clostridial neurotoxin is BoNT/X, BoNT/E, a chimeric BoNT or a hybrid BoNT.

41. The method according to any one of claims 20-40, wherein the engineered clostridial neurotoxin:

a. is encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12; or

b. comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13.

42. An engineered clostridial neurotoxin (e.g. obtainable by the method of any one of claims 20-41), wherein an endogenous activation loop of a clostridial neurotoxin has been replaced by an exogenous activation loop thereby providing an engineered clostridial neurotoxin,

wherein the exogenous activation loop comprises the polypeptide sequence Cys-$(Xaa)_a$-Ile-Asp/Glu-Gly-Arg-$(Yaa)_b$-Cys (SEQ ID NO: 1), wherein a = 1-10 and b = 4-15, and

wherein the clostridial neurotoxin is characterised in that: a peptide bond outside of the endogenous activation loop of the clostridial neurotoxin is hydrolysed by trypsin or Lys-C; and/or the endogenous activation loop is inefficiently proteolytically processed by trypsin or Lys-C.

43. The engineered clostridial neurotoxin according to claim 42, wherein the endogenous activation loop is one or more selected from SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 63, SEQ ID NO: 64, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 69, SEQ ID NO: 70 or SEQ ID NO: 71.

44. The engineered clostridial neurotoxin according to claim 42 or 43, wherein a = 2-4.

45. The engineered clostridial neurotoxin according to any one of claims 42-44, wherein a = 3.

46. The engineered clostridial neurotoxin according to any one of claims 42-45, wherein b = 6-10.

47. The engineered clostridial neurotoxin according to any one of claims 42-46, wherein b = 8.

48. The engineered clostridial neurotoxin according to any one of claims 42-47, wherein the exogenous activation loop comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

49. The engineered clostridial neurotoxin according to any one of claims 42-48, wherein the exogenous activation loop comprises a polypeptide sequence having at least 80% or 90% sequence identity to SEQ ID NO: 2 or SEQ ID NO: 3.

50. The engineered clostridial neurotoxin according to any one of claims 42-49, wherein the exogenous activation loop comprises a polypeptide sequence having SEQ ID NO: 2 or SEQ ID NO: 3.

51. The engineered clostridial neurotoxin according to any one of claims 42-50, wherein the exogenous activation loop comprises SEQ ID NO: 2.

52. The engineered clostridial neurotoxin according to any one of claims 42-51, wherein the exogenous activation loop consists of SEQ ID NO: 2.

53. The engineered clostridial neurotoxin according to any one of claims 42-52, wherein the clostridial neurotoxin lacks a functional H_C domain of a clostridial neurotoxin.

54. The engineered clostridial neurotoxin according to any one of claims 42-53, wherein the clostridial neurotoxin is a retargeted clostridial neurotoxin comprising a non-clostridial Targeting Moiety (TM).

55. The engineered clostridial neurotoxin according to any one of claims 42-54, wherein the clostridial neurotoxin is not BoNT/C1.

56. The engineered clostridial neurotoxin according to any one of claims 42-55, wherein the clostridial neurotoxin is selected from: BoNT/A, BoNT/B, BoNT/D, BoNT/E, BoNT/F, BoNT/G, BoNT/X, and TeNT.

57. The engineered clostridial neurotoxin according to any one of claims 42-56, wherein the clostridial neurotoxin is BoNT/X, BoNT/E, a chimeric BoNT or a hybrid BoNT.

58. The engineered clostridial neurotoxin according to any one of claims 42-57, wherein the single-chain clostridial neurotoxin:

a. is encoded by a nucleotide sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12; or

b. comprises a polypeptide sequence having at least 70% sequence identity to SEQ ID 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11 or SEQ ID NO: 13.

59. The engineered clostridial neurotoxin according to any one of claims 42-58, wherein the engineered clostridial neurotoxin is a re-targeted clostridial neurotoxin comprising a targeting moiety (TM), wherein the TM comprises an anthrax toxin protective antigen (PA) or a fragment thereof.

60. The engineered clostridial neurotoxin according to claim 59, wherein the engineered clostridial neurotoxin has a polypeptide sequence with at least 70% sequence identity to a polypeptide comprising:

a. SEQ ID NO: 52 or a fragment thereof; and

b. SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 or SEQ ID NO: 60

61. A method for proteolytically processing an engineered clostridial neurotoxin according to any one of claims 42-60 into a corresponding di-chain clostridial neurotoxin, the method comprising contacting the engineered clostridial neurotoxin with enterokinase or factor Xa thereby producing a di-chain clostridial neurotoxin.

62. A di-chain clostridial neurotoxin obtainable by the method of any one of claims 1-19, 23-41 or 61.

63. A pharmaceutical composition comprising the engineered clostridial neurotoxin according to any one of claims 42-60 or the di-chain clostridial neurotoxin according to claim 62, and a pharmaceutically acceptable carrier, excipient, adjuvant, propellant and/or salt.

64. An engineered clostridial neurotoxin according to any one of claims 42-60, the di-chain clostridial neurotoxin according to claim 62 or the pharmaceutical composition according to claim 63 for use in treating one or more of: a condition associated with unwanted immune secretion, strabismus, blepharospasm, squint, dystonia (e.g. spasmodic dystonia, oromandibular dystonia, focal dystonia, tardive dystonia, laryngeal dystonia, limb dystonia, cervical dystonia), torticollis (e.g. spasmodic torticollis), beauty therapy (cosmetic) applications benefiting from cell/muscle incapacitation (via SNARE down-regulation or inactivation), neuromuscular disorder or condition of ocular motility (e.g. concomitant strabismus, vertical strabismus, lateral rectus palsy, nystagmus, dysthyroid myopathy), writer's cramp, blepharospasm, bruxism, Wilson's disease, tremor, tics, segmental myoclonus, spasms, spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, animus, back spasm, charley horse, tension headaches, levator pelvic syndrome, spina bifida, tardive dyskinesia, Parkinson's disease, stuttering, hemifacial spasm, eyelid disorder, cerebral palsy, focal spasticity, spasmodic colitis, neurogenic bladder, anismus, limb spasticity, tics, tremors, bruxism, anal fissure, achalasia, dysphagia, lacrimation, hyperhydrosis, excessive salivation, excessive gastrointestinal secretions, muscle pain (e.g. pain from muscle spasms), headache pain (e.g. tension headache), brow furrows, skin wrinkles, cancer, uterine disorders, uro-genital disorders, urogenital-neurological disorders, chronic neurogenic inflammation, and a smooth muscle disorder.

65. Use of enterokinase for hydrolysing a peptide bond of a polypeptide comprising a polypeptide sequence shown as SEQ ID NO: 18 or SEQ ID NO: 19.

66. A nucleotide sequence encoding an engineered clostridial neurotoxin comprising a sequence having at least 70% sequence identity to SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 or SEQ ID NO: 12.