Recombinant Clostridium botulinum neurotoxins

This patent application claims priority to GB 1219602.8 filed on 31 October 2012, which is hereby incorporated by reference in its entirety.

The present invention relates to nucleic acid sequences encoding *Clostridium botulinum* (*C. botulinum*) neurotoxins of serotype E (BoNT/E), and to methods for producing recombinant BoNT/E. The present invention also relates to corresponding medical uses of a recombinant BoNT/E.

Botulinum neurotoxin 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 seven different classes of botulinum neurotoxin, namely: botulinum neurotoxin serotypes A, B, C_1, D, E, F and G, 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 the most potent toxins known, with median lethal dose (LD50) values for mice ranging from 0.5 to 5 ng/kg depending on the serotype. BoNTs are adsorbed 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/C, BoNT/A and BoNT/E cleave the synaptosomal-associated protein of 25 kDa (SNAP-25); and BoNT/C cleaves syntaxin.

In nature, clostridial neurotoxins are synthesised as a single-chain polypeptide that is modified post-translationally by a proteolytic cleavage event to form two polypeptide chains joined together by a disulphide bond. Cleavage occurs at a specific cleavage site, often referred to as the activation site, that is located between the cysteine residues that provide the inter-chain disulphide bond. It is this di-chain form that is the

active form of the toxin. The two chains are termed 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 (H_C domain) and an N-terminal translocation component (H_N domain). The cleavage site is located between the L-chain and the translocation components. Following binding of the H_C domain to its target neuron and internalisation of the bound toxin into the cell via an endosome, the H_N domain translocates the L-chain across the endosomal membrane and into the cytosol, and the L-chain provides a protease function (also known as a non-cytotoxic protease).

Non-cytotoxic proteases act by proteolytically-cleaving intracellular transport proteins known as SNARE proteins (e.g. SNAP-25, VAMP, or Syntaxin) – see Gerald K (2002) "Cell and Molecular Biology" (4th edition) *John Wiley & Sons, Inc.* The acronym SNARE derives from the term Soluble NSF Attachment Receptor, where NSF means N-ethylmaleimide-Sensitive Factor. SNARE proteins are integral to intracellular vesicle fusion, and thus to secretion of molecules via vesicle transport from a cell. The protease function is a zinc-dependent endopeptidase activity and exhibits a high substrate specificity for SNARE proteins. Accordingly, once delivered to a desired target cell, the non-cytotoxic protease is capable of inhibiting cellular secretion from the target cell. The L-chain proteases of Clostridial neurotoxins are non-cytotoxic proteases that cleave SNARE proteins.

Botulinum neurotoxins are well known for their ability to cause a flaccid muscle paralysis. Said muscle-relaxant properties have led to botulinum neurotoxins (such as BoNT/A) being employed in a variety of medical and cosmetic procedures, including treatment of glabellar lines or hyperkinetic facial lines, headache, hemifacial spasm, hyperactivity of the bladder, hyperhidrosis, nasal labial lines, cervical dystonia, blepharospasm, and spasticity.

Traditionally, production of BoNT is carried out by culture of *C. botulinum* bacteria, followed by isolation and purification of the botulinum neurotoxin complex. However, production of BoNT in this way is inefficient and provides low protein yields. In addition, *C. botulinum* are spore-forming bacteria and therefore require specialist culture equipment and facilities, which are not required for the culture of

bacteria such as *Escherichia coli* (*E. coli*). The increasing use of BoNTs has therefore led to a need for alternative and/or improved methods for producing and purifying BoNT.

US 20080103098 describes a method for producing recombinant BoNT proteins in a di-chain form comprising expression of a recombinant nucleic acid construct in an *E. coli* host cell. However, said method requires the insertion of a specific, non-native (i.e. non-clostridial) pentapeptide sequence into a loop domain of the neurotoxin. The inserted pentapeptide sequence forms an activation cleavage site that is cleaved by an endogenous *E. coli* protease upon cell lysis. The method of US 20080103098 therefore teaches that in order to achieve optimal BoNT expression, the BoNT sequence must be modified by the insertion of a non-native cleavage site.

US 7132259 describes recombinant nucleic acid molecules encoding BoNT proteins. However, the nucleic acid molecules of US 7132259 are modified to replace the native cleavage site with a non-native cleavage site. Thus, the method of US 7132259 also teaches that insertion of a non-native cleavage site is required for optimal BoNT expression.

US 6495143 describes recombinant nucleic acid molecules encoding fragments of the heavy chain (H_C) of a BoNT, for use in inducing immune responses (such as in vaccination). However, the nucleic acid molecules do not encode full length BoNT sequences. Expression in *E. coli* and purification of individual H and L chains of tetanus toxin and BoNT is achievable; these isolated chains are, by themselves, non-toxic. Following the separate production of these peptide chains and under strictly controlled conditions the H and L subunits can be combined by oxidative disulphide linkage to form active di-chains. Unfortunately, this strategy has several drawbacks. Firstly, it is not practical to express and isolate large amounts of the individual chains; in particular, in the absence of the H-chain the isolated L-chain is quite insoluble in aqueous solution and is highly susceptible to proteolytic degradation. Secondly, the *in vitro* oxidation of the individually expressed and purified H and L chains to produce the active di-chain is very inefficient, and leads to low yields of active toxin and the production of many inactive incorrectly folded or oxidized forms. The purification of the correctly folded and oxidized H and L chain-containing toxin is difficult, as is its

separation from these inactive forms and the unreacted separate H and L chains. Thus, the method of US 6495143 is associated with considerable disadvantages.

There is therefore a need in the art for improved methods for producing recombinant BoNTs, in particular activated di-chain BoNTs recombinant BoNT/E.

The present invention solves one or more of the above-mentioned problems, by providing nucleic acid sequences and methods as specified in the claims.

In one aspect, the present invention provides a nucleic acid sequence comprising a sequence of contiguous nucleotides, wherein said sequence of contiguous nucleotides has at least 80% (for example, at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9, or 100%) sequence identity to the nucleic acid sequence of SEQ ID NO: 1, and wherein said sequence of contiguous nucleotides encodes a single-chain BoNT/E1 protein.

The BoNT/E serotype is divided into eight subtypes, BoNT/E1 to BoNT/E8, which share at least 90% amino acid sequence identity; BoNT/E proteins within a given subtype share a higher amino acid percentage sequence identity (for example, at least 95% or higher). As described above, the nucleic acid sequences of the invention encode a BoNT/E1 protein. An example of a BoNT/E1 protein is the protein encoded by UniParc amino acid sequence UPI000016EA7F. Another example of a BoNT/E1 protein is the protein encoded by the amino acid sequence of SEQ ID NO: 2.

The nucleic acid sequences of the present invention have been designed to advantageously provide high levels of expression in *E. coli* cells.

A number of factors influence expression levels of a given protein. One such factor is the rate at which the mRNA sequence encoding that protein is translated. This factor itself is affected by which particular codons the mRNA uses to specify each amino acid of the protein. Some codons are translated more quickly than others. The choice of codon for each amino acid can vary because mRNA codons are degenerate in nature. Several different codons can all specify the same amino acid; therefore several different mRNA sequences can encode the same protein. Different codons that specify

the same amino acid are called synonymous codons. The precise mixture of synonymous codons in a particular mRNA affects the translation rate of the encoded protein.

There are a number of different reasons that account for why some codons are translated more quickly than others. Each codon specifies an amino acid by recruiting a tRNA molecule attached to that amino acid. The speed of translation is affected by the relative abundance of the various different tRNA molecules, by the affinity with which each particular tRNA molecule binds to the codon that recruits it and also by other factors such as how well the codon-tRNA molecule pair interacts with other elements of the translation machinery. Approximate codon translation rates can be estimated by determining the frequency at which different codons are found in highly-expressed genes. However, not all frequently occurring codons result in optimal expression.

Without wishing to be bound by any particular theory, the present inventors believe that optimal expression of BoNT/E1 nucleic acid sequences is achieved by reducing the frequency (i.e. the number of occurrences in a sequence) of certain codons, hereinafter deemed "slow codons" and set out below. In this regard, the present inventors believe that said slow codons are associated with reduced translation rates.

Amino acid	**Slow codon (RNA)**	**Slow codon (DNA equivalent)**
Phenylalanine	UUU	TTT
Tyrosine	UAU	TAT
Cysteine	UGU	TGT
Histidine	CAU	CAT
Glutamine	CAA	CAA
Proline	CCA and/or CCG	CCA and/or CCG
Serine	UCA and/or UCG	TCA and/or TCG
Arginine	CGG	CGG
Leucine	UUA and/or CUA	TTA and/or CTA

The present inventors have employed a process of rational sequence design to produce the nucleic acid sequences of the invention. One way in which the nucleic acid sequences of the invention provide high expression levels of the encoded BoNT/E1 proteins is by having an optimised number of slow codons (for example, a reduction in the frequency at which slow codons appear in the sequence).

In one embodiment, the nucleic acid sequence has a maximum of 160 slow codons (for example, a maximum of 160, 150, 140, 130, 120, 110, 100, 90, 95, 94, 93, 92, 91, 90, 89, 88, or 87 slow codons).

Thus, in one embodiment, the nucleic acid sequence has between 0 and 160 slow codons (for example 0-160, 0-150, 0-140, 0-130, 0-120, 0-110, 0-100, 0-90, 0-95, 0-94, 0-93, 0-92, 0-91, 0-90, 0-89, 0-88, or 0-87 slow codons).

In one embodiment, the nucleic acid sequence has 60-160 slow codons (for example, 60-160, 60-150, 60-140, 70-150, 70-140, 70-130, 70-120, 70-110, 70-100, 70-90, 80-130, 80-120, 80-110, 80-100, or 80-90 slow codons).

In one embodiment, optionally in combination with any one or more of the above embodiments, there are fewer slow codons in the first 50% of the nucleic acid sequence than in the second 50% of the nucleic acid sequence. The first 50% of the nucleic acid sequence is defined with reference to nucleotide position number 1 as the starting point, and therefore comprises the site of translation initiation; the second 50% of the nucleic acid sequence comprises the site of translation termination. By way of example, referring to SEQ ID NO: 1 (which has a total length of 3759 nucleotides), the first half of said sequence may be represented by nucleotide positions 1-1881 (comprising 627 nucleotide triplets), and the second half of said sequence may be represented by positions 1882-3759 (comprising 626 nucleotide triplets); alternatively, the first half of said sequence may be represented by nucleotide positions 1-1878 (comprising 626 nucleotide triplets), and the second half of said sequence may be represented by positions 1879-3759 (comprising 627 nucleotide triplets).

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises a maximum of 30 (for example, 30, 25, 20, 15, or 10) phenylalanine slow codons (RNA=UUU; DNA=TTT). In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises a maximum of 10 phenylalanine slow codons (RNA=UUU; DNA=TTT).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 30 (for example, 30, 25, 20, 19, or 18) tyrosine slow codons (RNA=UAU; DNA=TAT). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 18 tyrosine slow codons (RNA=UAU; DNA=TAT).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine and/or tyrosine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 19 (for example, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6, or 5) leucine slow codons (RNA=UUA and/or CUA; DNA=TTA and/or CTA). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine and/or tyrosine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 5 leucine slow codons (RNA=UUA and/or CUA; DNA=TTA and/or CTA).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, and/or leucine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 14 (for example, 14, 12, 10, 8, 6, 5, 4, or 3) glutamine slow codons (RNA=CAA; DNA=CAA). In one embodiment, optionally in combination with any one or more of the above embodiments (including the

previously-described embodiments relating to phenylalanine, tyrosine, and/or leucine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 3 glutamine slow codons (RNA=CAA; DNA=CAA).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, and/or glutamine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 20 (for example, 20, 19, 18, 17, or 16) serine slow codons (RNA=UCA and/or UCG; DNA=TCA and/or TCG). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, and/or glutamine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 16 serine slow codons (RNA=UCA and/or UCG; DNA=TCA and/or TCG).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, and/or serine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 23 (for example, 23, 22, 21, 20, or 19) proline slow codons (RNA=CCA and/or CCG; DNA=CCA and/or CCG). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, and/or serine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 19 proline slow codons (RNA=CCA and/or CCG; DNA=CCA and/or CCG).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, and/or proline slow codons), the nucleic acid sequence (as described above) comprises a maximum of 3 (for example 3, or 2) cysteine slow codons (RNA=UGU; DNA=TGT. In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, and/or proline slow codons), the nucleic acid sequence (as

described above) comprises a maximum of 2 cysteine slow codons (RNA=UGU; DNA=TGT).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, proline, and/or cysteine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 5 (for example 5, or 4) histidine slow codons (RNA=CAU; DNA=CAT). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, proline, and/or cysteine slow codons), the nucleic acid sequence (as described above) comprises a maximum of 4 histidine slow codons (RNA=CAU; DNA=CAT).

In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, proline, cysteine, and/or histidine slow codons), the nucleic acid sequence (as described above) comprises from 5 to 10 (for example, 5, 6, 7, 8, 9, or 10) arginine slow codons (RNA=CGG; DNA=CGG). In one embodiment, optionally in combination with any one or more of the above embodiments (including the previously-described embodiments relating to phenylalanine, tyrosine, leucine, glutamine, serine, proline, cysteine, and/or histidine slow codons), the nucleic acid sequence (as described above) comprises 10 arginine slow codons (RNA=CGG; DNA=CGG).

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises a maximum of 30 (for example, 30, 25, 20, 15, or 10; preferably 10) phenylalanine slow codons (RNA=UUU; DNA=TTT), and a maximum of 30 (for example, 30, 25, 20, 19, or 18; preferably 18) tyrosine slow codons (RNA=UAU; DNA=TAT).

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises a maximum of 30 (for example, 30, 25, 20, 15, or 10; preferably 10) phenylalanine slow codons

(RNA=UUU; DNA=TTT), a maximum of 30 (for example, 30, 25, 20, 19, or 18; preferably 18) tyrosine slow codons (RNA=UAU; DNA=TAT), and a maximum of 19 (for example, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6, or 5; preferably 5) leucine slow codons (RNA=UUA and/or CUA; DNA=TTA and/or CTA).

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises a maximum of 30 (for example, 30, 25, 20, 15, or 10; preferably 10) phenylalanine slow codons (RNA=UUU; DNA=TTT), a maximum of 30 (for example, 30, 25, 20, 19, or 18; preferably 18) tyrosine slow codons (RNA=UAU; DNA=TAT), a maximum of 19 (for example, 19, 18, 17, 16, 15, 12, 10, 9, 8, 7, 6, or 5; preferably 5) leucine slow codons (RNA=UUA and/or CUA; DNA=TTA and/or CTA), and a maximum of 14 (for example, 14, 12, 10, 8, 6, 5, 4, or 3; preferably 3) glutamine slow codons (RNA=CAA; DNA=CAA).

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises:

a maximum of 10 phenylalanine slow codons;

a maximum of 18 tyrosine slow codons;

a maximum of 2 cysteine slow codons;

a maximum of 4 histidine slow codons;

a maximum of 3 glutamine slow codons;

a maximum of 19 proline slow codons;

a maximum of 16 serine slow codons; and

a maximum of 5 leucine slow codons.

In one embodiment, optionally in combination with any one or more of the above embodiments, the nucleic acid sequence (as described above) comprises:

a maximum of 10 phenylalanine slow codons;

a maximum of 18 tyrosine slow codons;

a maximum of 2 cysteine slow codons;

a maximum of 4 histidine slow codons;

a maximum of 3 glutamine slow codons;

a maximum of 19 proline slow codons;

a maximum of 16 serine slow codons;
a maximum of 5 leucine slow codons; and
a maximum of 10 arginine slow codons.

In one embodiment, wherein the nucleic acid sequence is a nucleic acid sequence as described above, said single-chain BoNT/E1 protein comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% (for example, at least 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9, or 100%) sequence identity to the amino acid sequence of SEQ ID NO: 2.

In one embodiment, wherein the nucleic acid sequence is a nucleic acid sequence as described above, said single chain BoNT/E1 protein comprises a native activation site that is provided by an amino acid sequence selected from: KGIRK, VKGIRKS, SVKGIRKSI, VSVKGIRKSI, IVSVKGIRKSI, NIVSVKGIRKSI, KNIVSVKGIRKSI, CKNIVSVKGIRKSI, KNIVSVKGIRKSIC, and CKNIVSVKGIRKSIC.

In one embodiment, the nucleic acid sequence is a nucleic acid sequence as described above, with the proviso that the single-chain BoNT/E1 as described above or the sequence of contiguous amino acids as described above includes one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, seven or more, or eight) of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein):
glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; and asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) glycine at position 177; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; and asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) glycine at position 177, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) serine at position 198, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; and asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) alanine at position 340, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) leucine at position 773, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; alanine at position 340; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) leucine at position 963, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) glutamine at position 964, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; alanine at position 967; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) alanine at position 967, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; asparagine at position 1195.

In one embodiment, said one or more amino acids comprise (or consist of) asparagine at position 1195, and one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of: glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967.

In one embodiment, the presence of said one or more amino acids, as described above, provides a BoNT/E1 protein having improved solubility as compared to a BoNT/E1 protein lacking said amino acids. Said improved solubility increases the yield of the protein in a heterologous (*E. coli*) expression system.

In one embodiment, wherein the nucleic acid sequence is a nucleic acid sequence as described above, the sequence of contiguous nucleotides has at least 770 (for example, at least 770, 775, 780, 785, 790, 795, 800, 810, 820, 830, 840, 850, 860, 870, or 880) synonymous codons when compared to the nucleic acid sequence of wild-type BoNT/E1 (SEQ ID NO: 3). Thus, in one embodiment, the nucleic acid sequence comprises at least 770 codons that differ from, but encode the same amino acid as, the corresponding codon in the nucleic acid sequence of wild-type BoNT/E1 (SEQ ID NO: 3).

In one embodiment, the nucleic acid sequence (as described above) has a G-C content of at least 41% (for example, at least 41 or 42%). In one embodiment, the nucleic acid

sequence (as described above) has a G-C content of 42%. The concept of nucleic acid G-C content (also known as GC content or G+C content) relates to the proportion of nucleotides of a given nucleic acid sequence that are either G (guanine) or C (cytosine). Thus, in one embodiment, the G-C content of a nucleic acid sequence of the invention is altered (for example by substitution of synonymous codons) to more closely match the G-C content of nucleic acids preferentially expressed in *E.* coli host cells, thus improving expression of the sequence and providing increased protein yields.

In one aspect, the invention provides an expression vector encoding a nucleic acid sequence as described above. In one embodiment, the expression vector is a pET-26b(+) vector.

In one aspect, the invention provides a host cell comprising a nucleic acid sequence as described above, or an expression vector as described above. In one embodiment, the host cell is an *E. coli* cell. In one embodiment, the *E. coli* host cell is an *E. coli* BLR (DE3) cell.

In one aspect, the invention provides a method for producing soluble single-chain BoNT/E1 protein in an *E. coli* host cell, said method comprising: expressing a nucleic acid sequence (as described above) in an *E. coli* expression system.

Methods and techniques used to express heterologous proteins in *E. coli* (*Escherichia coli*) expression systems are well known in the art.

In one embodiment, said soluble single-chain BoNT/E1 protein is expressed in the cytoplasm of said *E. coli* host cell.

In one embodiment, said soluble single-chain BoNT/E1 protein is expressed at a level of at least 3 mg/L (for example, at least 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 25, 40, 45, or 50 mg/L).

In one embodiment, the method for producing soluble single-chain BoNT/E1 protein, as described above, comprises lysis of the *E. coli* host cell to provide an *E. coli* host

cell homogenate containing said soluble single-chain BoNT/E1 protein. Methods and techniques used to lyse host cells, such as *E. coli* host cells, are known in the art. Examples include sonication and the use of a French press.

In one aspect, the invention provides a method for producing soluble di-chain BoNT/E1 protein, said method comprising: providing a soluble single-chain BoNT/E1 protein comprising a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% (for example, at least 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9, or 100%) sequence identity to the amino acid sequence of SEQ ID NO: 2, and contacting said BoNT/E1 protein with trypsin in solution.

When the single-chain BoNT/E1 protein of the invention is contacted with trypsin, the proteolytic action of trypsin cleaves the single-chain protein at a site between the L-chain protease component and the translocation component to produce a di-chain protein, where the two chains are linked by a disulphide bridge (in more detail, the two chains formed following cleavage of single-chain BoNT/E1 at the activation site are a first chain of amino acid residues 1-419 and a second chain of amino acid residues 423-1252, with residues 420, 421 and 422 removed by the cleavage event). Thus, trypsin can be used to activate the single-chain polypeptide by converting it to the active di-chain form. Thus, advantageously, the use of trypsin means that it is not necessary to engineer an exogenous (non-native) cleavage site into a BoNT/E1 of the invention.

In one embodiment, reference to trypsin embraces trypsin-like enzymes that cleave at the same protease cleavage site as trypsin.

Trypsin cleaves protein sequences in which particular amino acids lie at certain positions on either side of the cleaved peptide bond. Such sequences can be represented by the nomenclature P4-P3-P2-P1-cleaved bond-P'1-P'2-P'3-P'4; in which P1 to P4 designate amino acids positioned 1 to 4 positions to the N-terminal side of the cleaved peptide bond respectively and P'1 to P'4 designate 1, to 4 positions C-terminal of the cleaved peptide bond respectively.

Most importantly Trypsin cleaves protein sequences where either Arg or Lys amino acids occupy the P1 position. When Lys is in the P1 position there are three major types of sequence that are not sensitive to trypsin:

(1) Pro in the P'1 position usually reduces susceptibility to cleavage by trypsin (but not when Trp is in position P2).

(2) Either Cys or Asp in the P2 position together with Asp in the P'1 position reduces susceptibility to cleavage by trypsin.

(3) Cys in the P2 position together with either His or Try in the P'1 position reduces susceptibility to cleavage by trypsin

When Arg is in the P1 position there are also three major types of sequence that are not sensitive to trypsin:

(1) Pro in the P'1 position usually reduces susceptibility to cleavage by trypsin (but not when either Met, or possibly Glu, is in position P2).

(2) Cys in the P2 position together with Lys in the P'1 position reduces susceptibility to cleavage by trypsin.

(3) Arg in the P2 position together with either His or Arg in the P'1 position reduces susceptibility to cleavage by trypsin.

In one embodiment, the invention provides a method (as described above) for producing soluble di-chain BoNT/E1 protein, with the proviso that said sequence of contiguous amino acids includes one or more (for example, one or more, two or more, three or more, four or more, five or more, six or more, or seven) of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein): glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In one embodiment, the presence of said one or more amino acids, as described above (and with reference to the multiple permutations of said one or more amino acids as described above), provides a BoNT/E1 protein having improved solubility as compared to a BoNT/E1 protein lacking said amino acids. Said improved solubility can increase the yield of the protein in a heterologous expression system.

In one embodiment, wherein the invention provides a method (as described above) for producing soluble di-chain BoNT/E1 protein, the soluble single-chain BoNT/E1 protein is provided by a method as described above for producing soluble single-chain BoNT/E1 protein in an *E. coli* host cell.

In one embodiment, wherein the invention provides a method (as described above) for producing soluble di-chain BoNT/E1 protein, the method comprises separating the soluble BoNT/E1 protein from the trypsin by contacting the solution containing soluble BoNT/E1 protein and trypsin with a hydrophobic surface, wherein the soluble BoNT/E1 protein preferentially binds to the hydrophobic surface.

The present inventors have found that high yields of activated di-chain BoNT/E1 protein can be obtained by using a process of hydrophobic purification to separate the activated di-chain polypeptide from trypsin. Surprisingly, this process provides superior purification to standard purification using ion exchange chromatography, which the present inventors have found to be ineffective for separating the activated di-chain polypeptide from trypsin. In addition, the process advantageously provides an activated di-chain BoNT/E1 protein that is free from the activating protease, as part of a general purification process.

The production of active recombinant BoNT/E1 requires a proteolytic step that cleaves the molecule into the active di-chain form. This cleavage can be achieved by an *in vitro* activation step using the protease, trypsin. After the activation step, it is important to remove the protease from the final product, which also prevents any further non-specific cleavage of BoNT/E1.

The isoelectric points (pI) of trypsin and BoNT/E1 are 9.32 and 6.2 respectively, which indicates that separation of the two proteins should be achieved by Ion Exchange (IEX) chromatography, exploiting the charge difference between the two molecules. A protein's net charge is affected by the pH of its surrounding environment and will become more positively or negatively charged depending on whether it gains or loses protons. The pI is the pH value at which a molecule carries no electrical charge and will therefore not interact with a charged IEX medium. This

means that if a protein is at a pH above its pI then it will carry a net negative charge and will bind to a positively charged medium such as an anion exchanger. Similarly, if the buffer pH is below the pI then the protein will carry a net positive charge and will not bind to an anion exchanger.

Based on this principle at pH 8, it would be expected that BoNT/E (which has a pI of 6.2) would bind to an anion exchange column, whilst trypsin with a pI of 9.32 would not, allowing the two proteins to be separated. IEX is a simple and inexpensive chromatography method, as it does not require the protein loaded onto the column to be in a high salt buffer, which can lead to protein losses by precipitation.

The present inventors have tested a variety of anion exchange columns, using both strong and weak functional groups attached to cross-linked agarose beads, at pH 8. In each case a large proportion of trypsin was found not to bind to the column as predicted and was present in the flow-through. However when the columns were eluted with a linear gradient of increasing ionic strength, trypsin was eluted from the column indicating that a proportion of the trypsin was able to bind to the columns. When compared to the elution of BoNT/E1 it was found that, unexpectedly, trypsin eluted at a similar ionic strength (Table 1; Figure 1) indicating that trypsin was not separated as predicted and would be present in the final purified BoNT/E1 product with the additional possibility of further BoNT/E1 degradation.

Table 1: Elution fractions from anion exchange columns on which the separation of trypsin from BoNT/E1 was assessed. Peaks are stated in number of column volumes (CV) F/T: Flow-through from the column, FF: Fast flow resin.

Column	**Trypsin**		**BoNT/E1**	
	Major Peak	**Minor Peak**	**Major Peak**	**Minor Peak**
ANX	F/T	8.8,11.3,12.3	10.7	17.3
QHP	F/T	9.0 ,10.6	-	-
DEAE	F/T	10.5	9.8	13.2
Q FF	F/T	9.2,10.9	16.1	10.7

The present inventors have solved the above problem. In more detail, the inventors have surprisingly identified that optimal trypsin-BoNT/E1 separation is achieved by use of a hydrophobic separation surface (for example, by hydrophobic interaction chromatography (HIC), which separates proteins according to differences in their surface hydrophobicity by utilising a reversible interaction between these proteins and the hydrophobic surface of a HIC medium).

In one embodiment, the hydrophobic surface is an inert matrix to which a ligand consisting of aryl or alkyl groups is attached.

The term “aryl” refers to aromatic groups, for example phenyl, naphthyl, thienyl, and indolyl.

The term “alkyl” refers to aliphatic groups including straight-chain, branched-chain, cyclic groups, and combinations thereof. An alkyl group may have 1 to 12 carbon atoms. Examples of alkyl groups include, but are not limited to groups such as methyl, ethyl, propyl (e.g. n-propyl, isopropyl), butyl (e.g. n-butyl, isobutyl, sec-butyl, t-butyl), pentyl, hexyl, heptyl, and octyl.

In one embodiment, the hydrophobic surface is selected from the group consisting of: butyl, phenyl or octyl ligands.

In one embodiment, the hydrophobic surface comprises butyl ligands. In one embodiment, the hydrophobic surface comprises phenyl ligands. In one embodiment, the hydrophobic surface comprises octyl ligands.

The present inventors have discovered that particularly preferable results for separating trypsin from BoNT/E are obtained with HIC using chromatography resins containing alkyl or aryl groups, for example butyl, phenyl, and octyl ligands, coupled to an inert matrix, such as cross-linked agarose or polystyrene beads (Table 2; Figure 2).

Table 2: Elution fractions from commercial hydrophobic interaction columns on which the separation of trypsin from BoNT/E was assessed. Peaks are stated in number of column volumes (CV) F/T: Flow-through from the column, FF: Fast flow resin, HP: High performance resin, (LS): Low substitution of hydrophobic groups, (HS): High substitution of hydrophobic groups.

Column	**Trypsin**		**BoNT/E**	
	Major Peak	**Minor Peak**	**Major Peak**	**Minor Peak**
Phenyl (HS) FF	23.3	-	32.2	-
Phenyl (LS) FF	F/T	-	21.4	-
Phenyl HP	F/T	16.1	24.4	-
Butyl FF	F/T	16.8	23.7	-
Butyl HP	18	Wash	27.6	-
Octyl FF	F/T	-	27.1	-

In one embodiment, the process of hydrophobic purification to separate the activated di-chain BoNT/E1 protein from trypsin reduces the concentration of trypsin at least 100-fold, at least 150-fold, at least 200-fold, at least 250-fold, at least 300-fold, at least 350-fold, at least 400-fold, at least 450-fold, or at least 500-fold. In a preferred embodiment, the process of hydrophobic purification to separate the activated di-chain BoNT/E1 protein from trypsin reduces the concentration of trypsin at least 350-fold.

In another aspect, the invention provides an active di-chain BoNT/E1 protein, wherein the first chain comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% (for example, at least 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9, or 100%) sequence identity to the amino acid sequence of positions 1-419 of SEQ ID NO: 2; wherein the second chain comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% (for example, at least 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9, or 100%) sequence identity to the amino acid sequence of positions 423-1252 of SEQ ID NO: 2; wherein the first and second chains are joined together by a disulphide bond between cysteine 412 on the first chain and cysteine 426 on the second chain; with the proviso that said sequence of contiguous amino acids includes one or more (for example, two or more, three or more, four or more, five or more, six or more, seven or more, or eight) of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein): glycine at position 177; serine at position 198; alanine at position 340; leucine at position 773; leucine at position 963; glutamine at position 964; alanine at position 967; asparagine at position 1195.

In a related aspect, the invention provides an active di-chain BoNT/E1 protein obtainable by a method (as described above) for producing soluble di-chain BoNT/E1 protein.

In one aspect, the invention provides a composition comprising an active di-chain BoNT/E1 protein (as described above), wherein said composition is substantially free from trypsin.

Thus, the composition is, advantageously, substantially free from trypsin protease (used to activate the single-chain polypeptide by converting it to the active di-chain form), thus preventing unwanted non-specific cleavage of BoNT/E1 protein.

In one embodiment, wherein the composition (as described above) is substantially free from trypsin, the composition contains less than 100 picograms (pg) trypsin per 100 nanograms (ng) of BoNT/E1 protein; for example, less than 50, 20, 10, 9, 8, 7, 6

or 5 pg trypsin per 100 ng of BoNT/E1 protein. In one embodiment, the composition (as described above) contains less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein, or less than 5 pg trypsin per 100 ng of BoNT/E1 protein. In a preferred embodiment, the composition (as described above) contains less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein.

Thus, in one embodiment, the phrase "substantially free from trypsin" means less than 100 pg trypsin per 100 ng of BoNT/E1 protein; for example, less than 50, 20, 10, 9, 8, 7, 6 or 5 pg trypsin per 100 ng of BoNT/E1 protein, preferably less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein.

Methods for determining the concentration of trypsin in a composition are known in the art. By way of example, the concentration of trypsin in a composition of the invention may be determined using a sandwich ELISA (Enzyme-Linked Immunosorbent Assay).

In a further aspect, the present invention provides a solid or liquid pharmaceutical composition comprising:

(a) an active di-chain BoNT/E1 protein as described above, and

(b) a stabilising agent.

In one embodiment, the composition (as described above) is substantially free from trypsin. In one embodiment, the composition contains less than 100 pg trypsin per 100 ng of BoNT/E1 protein, for example, less than 50, 20, 10, 9, 8, 7, 6 or 5 pg trypsin per 100 ng of BoNT/E1 protein. In one embodiment, the composition contains less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein.

Stabilising agents which can be used in compositions according to the invention include protein stabilisers, such as albumin, in particular human serum albumin (HSA), and non-protein stabilisers.

Non-protein stabilising agents which can be used in the composition according to the invention include surfactants, in particular non-ionic surfactants. Examples of non-ionic surfactants include polysorbates, such as polysorbate 20 or polysorbate 80, and block copolymers such as poloxamers (i.e. copolymers of polyethylene and propylene glycol).

In a particular embodiment, the composition does not comprise a protein as a stabilising agent.

According to a particular embodiment of the invention, the pharmaceutical composition is a liquid pharmaceutical composition comprising:

(a) an active di-chain BoNT/E1 protein, as described above;

(b) a non-protein stabilising agent that is a surfactant; and

(c) water;

wherein said liquid pharmaceutical composition does not comprise a protein stabilising agent; and

wherein said liquid pharmaceutical composition is substantially free from trypsin (e.g. said liquid pharmaceutical composition contains less than 100 pg trypsin per 100 ng of BoNT/E1 protein, or less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein, or less than 5 pg trypsin per 100 ng of BoNT/E1 protein; preferably wherein said liquid pharmaceutical composition contains less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein).

In one embodiment, the active di-chain BoNT/E1 protein is present in the composition (as described above) at a concentration of 1-100 ng/ml. In one embodiment, the active di-chain BoNT/E1 protein is present in the composition (as described above) at a concentration of 5-50 ng/ml, e.g. about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 ng/ml. In a preferred embodiment, the active di-chain BoNT/E1 protein is present at a concentration of about 20 ng/ml.

In one embodiment, the surfactant (as described above) is a polysorbate, such as a polysorbate having a mean polymerisation degree ranging from 20 to 100 monomer units, and may for example be polysorbate 80. In a preferred embodiment, the

polysorbate is vegetable-derived. The concentration of the surfactant is preferably lower than 1% v/v, for example from about 0.005% to 0.02% v/v in the case of polysorbate 80.

The pharmaceutical composition according to the invention can also comprise a crystalline agent.

By crystalline agent is meant an agent which, inter alia, maintains a mechanically strong cake structure to a lyophilised botulinum neurotoxin complex (type A, B, C, D, E, F or G) or a high purity botulinum neurotoxin (type A, B, C, D, E, F or G) . When included in solid formulations, crystalline agents also have a bulking effect. Crystalline agents notably include sodium chloride. The concentration of crystalline agent can be for example from 0.1 to 0.5 M, preferably from 0.1 to 0.4 M, notably about 0.15 to 0.3 M

The pharmaceutical composition according to the invention can also comprise a buffer to maintain pH a level comprised between 5.5 and 7.5, or between 6.0 and 7.0. The buffer can be any buffer able to maintain the adequate pH. For example, the buffer for compositions according to the invention can be chosen from the group consisting of succinate, disodium phosphate / citric acid, and an amino acid such as histidine. The concentration of the buffer can be for example from 1 to 50 mM, preferably from 5 to 20 mM, preferably about 10 mM.

The pharmaceutical composition according to the invention can also comprise a disaccharide.

The disaccharide used in compositions according to the invention can be chosen from the group consisting of sucrose, trehalose, mannitol and lactose. In a specific embodiment, the disaccharide is sucrose. The concentration of the disaccharide can be for example from 5 to 50 mM, preferably from 5 to 25 mM, more preferably from 10 to 20 mM, and most preferably about 11.7 mM.

In a particular embodiment, the pharmaceutical composition is a liquid pharmaceutical composition comprising:

(a) an active di-chain BoNT/E1 protein, as described above;

(b) a non-protein stabilising agent that is a surfactant;

(c) sodium chloride,

(c) a buffer to maintain pH between 5.5 and 7.5

(e) a disaccharide, and

(f) sterile water.

wherein said liquid pharmaceutical composition does not comprise a protein stabilising agent; and

wherein said liquid pharmaceutical composition is substantially free from trypsin (e.g. said liquid pharmaceutical composition contains less than 100 pg trypsin per 100 ng of BoNT/E1 protein, or less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein, or less than 5 pg trypsin per 100 ng of BoNT/E1 protein; preferably wherein said liquid pharmaceutical composition contains less than 10 pg trypsin per 100 ng of BoNT/E1 protein, or less than 7 pg trypsin per 100 ng of BoNT/E1 protein).

According to a specific embodiment, the pharmaceutical composition according to the invention in liquid form is sealed in a vial or in a ready-to-use device, such as a syringe, with no liquid/gaseous interface, and is stable for at least three months or at least six months at 23 to 27 °C and for at least twelve months at 2-8 °C.

In one aspect, the invention provides an active di-chain BoNT/E1 protein as described above, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein as described above, or a composition as described above, or a liquid pharmaceutical composition as described above, for use in therapy.

The present inventors have identified that the active di-chain BoNT/E1 proteins of the invention, and the compositions and liquid pharmaceutical compositions thereof, can be used in therapy. Suitable therapies may include cosmetic treatments and methods of medical treatment.

Key to SEQ ID NOs

SEQ ID NO: 1 Optimised BoNT/E1 nucleic acid sequence

SEQ ID NO: 2 BoNT/E1 amino acid sequence

SEQ ID NO: 3 Wildtype BoNT/E1 nucleic acid sequence

SEQ ID NO: 1 Optimised BoNT/E1 nucleic acid sequence

ATGCCGAAAATCAACTCTTTCAACTACAACGACCCGGTTAACGACCGTACCATCCTGTAT
ATCAAACCGGGTGGTTGCCAGGAGTTCTACAAATCTTTCAACATCATGAAAAACATCTGG
ATCATCCCGGAACGTAACGTTATCGGTACCACCCCGCAGGACTTCCACCCGCCGACCTCT
CTGAAAAACGGTGACTCTTCTTACTACGACCCGAACTACCTCCAGTCTGACGAAGAAAAA
GACCGTTTCCTGAAAATCGTTACCAAAATCTTCAACCGTATCAACAACAACCTGTCTGGT
GGTATCCTGCTGGAAGAACTGTCTAAAGCTAACCCGTACCTGGGTAACGACAACACCCCG
GACAACCAGTTCCACATCGGTGACGCTTCTGCTGTTGAAATCAAATTCTCTAACGGTTCT
CAGGACATCCTGCTGCCGAACGTTATCATCATGGGTGCTGAACCGGACCTGTTCGAAACC
AACTCTTCTAACATCTCTCTGCGTAACAACTACATGCCGTCTAACCACGGTTTCGGTTCT
ATCGCTATCGTTACCTTCTCTCCGGAATACTCTTTCCGTTTCAACGACAACAGCATGAAC
GAGTTCATCCAGGACCCGGCTCTGACCCTGATGCACGAACTGATCCACTCTCTGCACGGT
CTGTACGGTGCTAAAGGTATCACCACCAAATACACCATCACCCAGAAACAGAACCCGCTG
ATCACCAACATCCGTGGTACCAACATCGAAGAGTTCCTGACCTTCGGTGGTACCGACCTG
AACATCATCACCTCTGCTCAGTCTAACGACATCTACACCAACCTGCTGGCTGACTACAAA
AAAATCGCTTCTAAACTGTCTAAAGTTCAGGTTTCTAACCCGCTGCTGAACCCGTACAAA
GACGTTTTCGAAGCTAAATACGGTCTGGACAAAGACGCTTCTGGTATCTACTCTGTTAAC
ATCAACAAATTCAACGACATCTTCAAAAAACTGTACTCTTTCACCGAGTTCGACCTGGCG
ACCAAATTCCAGGTTAAATGCCGTCAGACCTACATCGGTCAGTACAAATACTTCAAACTG
TCTAACCTGCTGAACGACTCTATCTACAACATCTCTGAAGGTTACAACATCAACAACCTG
AAAGTTAACTTCCGTGGTCAGAACGCTAACCTGAACCCGCGTATCATCACCCCGATCACC
GGTCGTGGTCTGGTTAAAAAAATCATCCGTTTCTGCAAGAATATTGTAAGCGTTAAAGGA
ATAAGAAAAAGTATCTGCATCGAAATCAACAACGGTGAACTGTTCTTCGTTGCTTCTGAA
AACTCTTACAACGACGACAACATCAACACCCCGAAAGAAATCGACGACACCGTTACCTCT
AACAACAACTACGAAAACGACCTGGACCAGGTTATCCTGAACTTCAACTCTGAATCTGCT
CCGGGTCTGTCTGACGAAAAACTGAACCTGACCATCCAGAACGACGCTTACATCCCGAAA
TACGACTCTAACGGTACCTCTGACATCGAACAGCACGACGTTAACGAACTGAACGTTTTC
TTCTACCTGGACGCTCAGAAAGTTCCGGAAGGTGAAAACAACGTTAACCTGACCTCTTCT
ATCGACACCGCTCTGCTGGAACAGCCGAAAATCTACACCTTCTTCTCTTCTGAGTTCATC
AACAACGTTAACAAACCGGTTCAGGCTGCTCTGTTCGTTTCTTGGATTCAGCAGGTTCTG
GTTGACTTCACCACCGAAGCTAACCAGAAATCTACCGTTGACAAAATCGCTGACATCTCT
ATCGTTGTTCCGTACATCGGTCTGGCTCTGAACATCGGTAACGAAGCTCAGAAAGGTAAC
TTCAAAGACGCTCTGGAACTGCTGGGTGCTGGTATCCTGCTGGAGTTCGAACCGGAACTG
CTGATCCCGACCATCCTGGTTTTCACCATCAAATCTTTCCTGGGTTCTTCTGACAACAAA
AACAAAGTTATCAAAGCTATCAACAACGCTCTGAAAGAACGTGACGAAAAATGGAAAGAA
GTTTACTCTTTCATCGTTTCTAACTGGATGACCAAAATCAACACCCAGTTCAACAAACGT
AAAGAACAGATGTACCAGGCTCTCCAGAACCAGGTTAACGCTATCAAAACCATCATCGAA
TCTAAATACAACTCTTACACCCTGGAAGAAAAAAACGAACTGACCAACAAATACGACATC

AAACAGATCGAAAACGAACTGAACCAGAAAGTTTCTATCGCTATGAACAACATCGACCGT
TTCCTGACCGAATCTTCTATCTCTTACCTGATGAAACTCATCAACGAAGTTAAAATCAAC
AAACTGCGTGAATACGACGAAAACGTTAAAACCTACCTGCTGAACTACATCATCCAGCAC
GGTTCTATCCTGGGTGAATCTCAGCAGGAACTGAACTCTATGGTTACCGACACCCTGAAC
AACTCTATCCCGTTCAAACTGTCTTCTTACACCGACGACAAAATCCTGATCTCTTACTTC
AACAAATTCTTTAAACGCATTAAGAGTTCATCGGTTCTGAATATGCGGTACAAAAATGAT
AAATATGTCGATACTTCTGGATATGATAGCAATATCAACATTAACGGCGACGTGTATAAA
TATCCGACAAATAAAAACCAGTTTGGGATATATAACGACAAGCTGTCGGAGGTCAATATT
TCTCAAAACGACTATATCATTTACGATAATAAATATAAAAACTTTAGCATTAGTTTTTGG
GTTCGTATACCTAATTATGACAATAAAATTGTAAATGTGAATAACGAGTATACCATTATA
AACTGTATGCGCGACAATAACAGTGGTTGGAAGGTATCGCTGAACCATAATGAGATTATC
TGGACCCTGCAGGATAATGCAGGTATAAACCAGAAACTGGCTTTTAACTATGGAAACGCA
AATGGGATCTCAGATTACATTAATAAATGGATTTTTGTTACCATTACGAACGATCGCTTA
GGCGACTCAAAACTTTATATTAATGGCAATCTGATAGATCAGAAATCAATCTTAAATTTG
GGCAATATTCATGTCTCTGATAACATCTTGTTCAAGATCGTTAATTGCAGTTACACTCGT
TATATTGGCATTCGTTACTTTAATATCTTCGATAAAGAACTGGACGAGACGGAAATCCAG
ACTCTGTATTCAAACGAGCCCAATACTAATATATTGAAAGATTTTTGGGGTAACTATCTT
TTATATGATAAAGAATACTATCTCCTGAATGTATTGAAGCCAAACAATTTCATAGATAGA
CGCAAGGATAGCACATTAAGTATCAACAATATCAGATCTACTATACTGTTAGCAAATCGC
CTCTACTCCGGTATTAAAGTGAAGATTCAGCGGGTTAATAACTCCAGTACCAATGATAAT
CTGGTCCGTAAGAACGATCAGGTATACATCAATTTCGTCGCGAGCAAAACTCATCTCTTC
CCGCTTTACGCCGATACAGCTACGACAAACAAGGAAAAAACCATAAAAATTTCCAGCTCC
GGAAACAGATTCAATCAAGTAGTTGTAATGAACTCTGTGGGTAATAATTGTACGATGAAC
TTTAAGAATAACAATGGGAACAATATTGGACTTTTGGGCTTCAAAGCCGACACAGTGGTG
GCGTCCACCTGGTATTACACGCACATGCGGGACCATACGAATTCGAACGGTTGCTTCTGG
AACTTTATCTCGGAAGAACACGGGTGGCAAGAAAAATAA

SEQ ID NO: 2 BoNT/E1 amino acid sequence

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTS
LKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTP
DNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGS
IAIVTFSPEYSFRFNDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITTKYTITQKQNPL
ITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLSKVQVSNPLLNPYK
DVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKFQVKCRQTYIGQYKYFKL
SNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPITGRGLVKKIIRFCKNIVSVKG
IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESA
PGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSS
IDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADIS
IVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNK
NKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIE
SKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKIN
KLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF
NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNI
SQNDYIIYDNKYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEII
WTLQDNAGINQKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNL
GNIHVSDNILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYL
LYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDN
LVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNNCTMN
FKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK

SEQ ID NO: 3 Wildtype BoNT/E1 nucleic acid sequence

ATGCCAAAAATTAATAGTTTTAATTATAATGATCCTGTTAATGATAGAACAATTTTATAT
ATTAAACCAGGCGGTTGTCAAGAATTTTATAAATCATTTAATATTATGAAAAATATTTGG
ATAATTCCAGAGAGAAATGTAATTGGTACAACCCCCCAAGATTTTCATCCGCCTACTTCA
TTAAAAAATGGAGATAGTAGTTATTATGACCCTAATTATTTACAAAGTGATGAAGAAAAG
GATAGATTTTTAAAAATAGTCACAAAAATATTTAATAGAATAAATAATAATCTTTCAGGA
GGGATTTTATTAGAAGAACTGTCAAAAGCTAATCCATATTTAGGGAATGATAATACTCCA
GATAATCAATTCCATATTGGTGATGCATCAGCAGTTGAGATTAAATTCTCAAATGGTAGC
CAAGACATACTATTACCTAATGTTATTATAATGGGAGCAGAGCCTGATTTATTTGAAACT
AACAGTTCCAATATTTCTCTAAGAAATAATTATATGCCAAGCAATCACGGTTTTGGATCA
ATAGCTATAGTAACATTCTCACCTGAATATTCTTTTAGATTTAATGATAATAGTATGAAT
GAATTTATTCAAGATCCTGCTCTTACATTAATGCATGAATTAATACATTCATTACATGGA
CTATATGGGGCTAAAGGGATTACTACAAAGTATACTATAACACAAAAACAAAATCCCCTA
ATAACAAATATAAGAGGTACAAATATTGAAGAATTCTTAACTTTTGGAGGTACTGATTTA
AACATTATTACTAGTGCTCAGTCCAATGATATCTATACTAATCTTCTAGCTGATTATAAA
AAAATAGCGTCTAAACTTAGCAAAGTACAAGTATCTAATCCACTACTTAATCCTTATAAA
GATGTTTTTGAAGCAAAGTATGGATTAGATAAAGATGCTAGCGGAATTTATTCGGTAAAT
ATAAACAAATTTAATGATATTTTTAAAAAATTATACAGCTTTACGGAATTTGATTTAGCA
ACTAAATTTCAAGTTAAATGTAGGCAAACTTATATTGGACAGTATAAATACTTCAAACTT
TCAAACTTGTTAAATGATTCTATTTATAATATATCAGAAGGCTATAATATAAATAATTTA
AAGGTAAATTTTAGAGGACAGAATGCAAATTTAAATCCTAGAATTATTACACCAATTACA
GGTAGAGGACTAGTAAAAAAAATCATTAGATTTTGTAAAAATATTGTTTCTGTAAAAGGC
ATAAGGAAATCAATATGTATCGAAATAAATAATGGTGAGTTATTTTTTGTGGCTTCCGAG
AATAGTTATAATGATGATAATATAAATACTCCTAAAGAAATTGACGATACAGTAACTTCA
AATAATAATTATGAAAATGATTTAGATCAGGTTATTTTAAATTTTAATAGTGAATCAGCA
CCTGGACTTTCAGATGAAAAATTAAATTTAACTATCCAAAATGATGCTTATATACCAAAA
TATGATTCTAATGGAACAAGTGATATAGAACAACATGATGTTAATGAACTTAATGTATTT
TTCTATTTAGATGCACAGAAAGTGCCCGAAGGTGAAAATAATGTCAATCTCACCTCTTCA
ATTGATACAGCATTATTAGAACAACCTAAAATATATACATTTTTTTCATCAGAATTTATT
AATAATGTCAATAAACCTGTGCAAGCAGCATTATTTGTAAGCTGGATACAACAAGTGTTA
GTAGATTTTACTACTGAAGCTAACCAAAAAAGTACTGTTGATAAAATTGCAGATATTTCT
ATAGTTGTTCCATATATAGGTCTTGCTTTAAATATAGGAAATGAAGCACAAAAAGGAAAT
TTTAAAGATGCACTTGAATTATTAGGAGCAGGTATTTTATTAGAATTTGAACCCGAGCTT
TTAATTCCTACAATTTTAGTATTCACGATAAAATCTTTTTTAGGTTCATCTGATAATAAA
AATAAAGTTATTAAAGCAATAAATAATGCATTGAAAGAAAGAGATGAAAAATGGAAAGAA
GTATATAGTTTTATAGTATCGAATTGGATGACTAAAATTAATACACAATTTAATAAAAGA
AAAGAACAAATGTATCAAGCTTTACAAAATCAAGTAAATGCAATTAAAACAATAATAGAA
TCTAAGTATAATAGTTATACTTTAGAGGAAAAAAATGAGCTTACAAATAAATATGATATT
AAGCAAATAGAAAATGAACTTAATCAAAAGGTTTCTATAGCAATGAATAATATAGACAGG
TTCTTAACTGAAAGTTCTATATCCTATTTAATGAAATTAATAAATGAAGTAAAAATTAAT
AAATTAAGAGAATATGATGAGAATGTCAAAACGTATTTATTGAATTATATTATACAACAT
GGATCAATCTTGGGAGAGAGTCAGCAAGAACTAAATTCTATGGTAACTGATACCCTAAAT
AATAGTATTCCTTTTAAGCTTTCTTCTTATACAGATGATAAAATTTTAATTTCATATTTT
AATAAATTCTTTAAGAGAATTAAAAGTAGTTCAGTTTTAAATATGAGATATAAAAATGAT
AAATACGTAGATACTTCAGGATATGATTCAAATATAAATATTAATGGAGATGTATATAAA
TATCCAACTAATAAAAATCAATTTGGAATATATAATGATAAACTTAGTGAAGTTAATATA
TCTCAAAATGATTACATTATATATGATAATAAATATAAAAATTTTAGTATTAGTTTTTGG
GTAAGAATTCCTAACTATGATAATAAGATAGTAAATGTTAATAATGAATACACTATAATA
AATTGTATGAGAGATAATAATTCAGGATGGAAAGTATCTCTTAATCATAATGAAATAATT
TGGACATTGCAAGATAATGCAGGAATTAATCAAAAATTAGCATTTAACTATGGTAACGCA
AATGGTATTTCTGATTATATAAATAAGTGGATTTTTGTAACTATAACTAATGATAGATTA

GGAGATTCTAAACTTTATATTAATGGAAATTTAATAGATCAAAAATCAATTTTAAATTTA
GGTAATATTCATGTTAGTGACAATATATTATTTAAAATAGTTAATTGTAGTTATACAAGA
TATATTGGTATTAGATATTTTAATATTTTTGATAAAGAATTAGATGAAACAGAAATTCAA
ACTTTATATAGCAATGAACCTAATACAAATATTTTGAAGGATTTTTGGGGAAATTATTTG
CTTTATGACAAAGAATACTATTTATTAAATGTGTTAAAACCAAATAACTTTATTGATAGG
AGAAAAGATTCTACTTTAAGCATTAATAATATAAGAAGCACTATTCTTTTAGCTAATAGA
TTATATAGTGGAATAAAAGTTAAAATACAAAGAGTTAATAATAGTAGTACTAACGATAAT
CTTGTTAGAAAGAATGATCAGGTATATATTAATTTTGTAGCCAGCAAAACTCACTTATTT
CCATTATATGCTGATACAGCTACCACAAATAAAGAGAAAACAATAAAAATATCATCATCT
GGCAATAGATTTAATCAAGTAGTAGTTATGAATTCAGTAGGAAATAATTGTACAATGAAT
TTTAAAAATAATAATGGAAATAATATTGGGTTGTTAGGTTTCAAGGCAGATACTGTAGTT
GCTAGTACTTGGTATTATACACATATGAGAGATCATACAAACAGCAATGGATGTTTTTGG
AACTTTATTTCTGAAGAACATGGATGGCAAGAAAAATAA

List of figures

Figure 1

Elution fractions from anion exchange columns on which the separation of trypsin from BoNT/E1 was assessed. The peak of trypsin, BoNT/E1 and the salt gradient are marked. Fig. 1A: Q-Sepharose HP; Fig. 1B: DEAE Sepharose.

Figure 2

Elution fractions from hydrophobic interaction columns on which the separation of trypsin from BoNT/E1 was assessed. The peak of trypsin, BoNT/E1 and the salt gradient are marked. Fig. 2A: Phenyl Sepharose HP; Fig. 2B: Butyl Sepharose HP; Fig. 2C:) Octyl Sepharose FF.

Figure 3

Soluble expression level of rBoNT/E1 culture determined by western blotting, compared to commercial BoNT/E1.

Figure 4

SDS-PAGE of rBoNT/E1 under non-reducing and reducing conditions confirming formation of the di-chain structure.

Examples

EXAMPLE 1

Construction of an optimised BoNT/E1 nucleic acid sequence

The DNA sequence was initially designed by back translation of the BoNT/E1 amino acid sequence (SEQ ID NO: 2). A restriction sequence (PstI) was added to the N-terminus and a stop codon and additional restriction sequences, XbaI-Stop codon-HindIII, to the C-terminus. The DNA sequence was then optimised for expression based on the number and location of slow codons (as defined above).

The sequence was optimised to select against slow codons. This was applied particularly at the start of the sequence to get good initiation and start translation. Where slow codons were included (to allow for usage according to expression host codon bias), these were towards the end of the sequence (where the beginning of the sequence is defined as where translation is initiated).

Once the sequence had been designed, the optimised DNA sequence was synthesised in two parts using a unique/native PstI site for later assembly into the full-length toxin gene. The sequence of the first gene included an NdeI site at the amino terminus and a PstI site at the carboxyl terminus. This part of the gene was 2895 bp in length, encoding the BoNT/E1 LC and amino portion of the HC. The sequence of the second gene included a PstI site at the N terminus and HindIII site at the carboxyl terminus, was 882 bp in length and encoded the carboxyl portion of the BoNT/E1 HC.

EXAMPLE 2

Construction of the expression vector BoNT/E1 nucleic acid sequence

An expression vector based on the vector pET-26b(+) (Novagen) was employed, which includes the cloning restriction sites NdeI and HindIII located at the start and end of the DNA encoding the BoNT/E1 ORF. The pET-26b(+) vector was mobilisable-deficient but could be mobilised if co-resident with other mobilisable plasmids. The pET-26b (+) vector was modified to remove mobility genes and render it non-mobilisable.

The expression vector was digested with NdeI and PstI and the purified vector backbone was ligated with the first fragment of BoNT/E1 DNA that had been digested with the same restriction enzymes to create an intermediate product. In the second cloning step, BoNT/E1 DNA from the second fragment that had been digested with PstI and HindIII was ligated into the intermediate product from step one (that had also been digested with the same restriction enzymes). This led to the creation of the final product of BoNT/E1 DNA in the expression vector.

EXAMPLE 3

Insertion of BoNT/E1 expression vector into host

This example is based on the use of *E. coli* BLR (DE3) cells, though the procedures and methods are equally applicable to any other *E. coli* expression strain. *E. coli* BLR (DE3) competent cells were stored below -70°C until required. The transformation of the cells was carried out using an adaptation of the manufacturer's protocol. The cells were defrosted on ice and sub aliquots of 10 µL were prepared. An aliquot was transformed using heat shock at 42°C for 80 seconds with 1 µL of plasmid DNA. After recovering on ice for 5 minutes, 90 µL of animal free SOC broth was added to the transformations which were then transferred to shaking incubators and incubated for 1 hour at 37°C and 250 rpm. After incubation 90 µL of each transformation was transferred and spread onto animal free LB agar plates supplemented with 50 µg/mL kanamycin. The plates were incubated at 37°C for 16 hours.

EXAMPLE 4

Culturing of host and expression of soluble rBoNT/E1 protein

A single colony of BoNT/E1 transformed in BLR(DE3) cells was used to inoculate a 250 ml conical flask containing 100 ml modified Terrific Broth (mTB) supplemented with 0.2% glucosamine and 30 µg/ml kanamycin. This method would be equally applicable when using a Microbank bead or glycerol stock (10-100 µl) to inoculate the flask.

The flask was incubated for 16 hours at 37 °C with 250 RPM shaking. 10 ml of this starter culture was used to inoculate 2 L conical flasks each containing 1 L supplemented with 0.2% glucosamine and 30 µg/ml kanamycin. Cells were grown at 37 °C for ~2 hours at 225 RPM until an OD_{600} of 0.5 was reached. At this point, the

culture temperature was dropped to 16 °C. After 1 hour, the cells were induced to express BoNT/E1 by addition of 1 mM IPTG for 20 hours. Cells were harvested by centrifugation for 20 min at 4 °C, weighed and then stored at -20 °C.

EXAMPLE 5

Extraction of BoNT/E1 protein from host and analysis of expression level

Expression cell pastes of rBoNT/E1 were thawed at room temperature and resuspended by pipetting in 3 ml of Tris-NaCL re-suspension buffer per gram of cells supplemented with 10 µl benzonase. Cells were lysed by sonication at a 4 µm amplitude – 10x 30 s on + >45 s off. The lysate was centrifuged at 4000*g* for 1 h at 4 °C to obtain the soluble rBoNT/E1 in the supernatant.

Bradford Assay to Determine Total Protein Concentration of Prepared Lysates

A sample (50 µL) of either diluted rBoNT/E1 lysate or BSA standard was added to 1 mL plastic cuvettes. 450 µL of Coomassie Bradford Assay reagent was added to each cuvette and allowed to incubate at room temperature for 10 minutes before reading OD_{600}. The values obtained for the BSA standards were used to determine the amount of protein in the lysate samples.

Preparation of Lysate Samples for Semi-Quantitative Western Blotting Analysis

A commercial sample of BoNT/E1 protein purchased from Metabiologics was used to make up SDS-PAGE standards. SDS-PAGE samples were then prepared from the lysate samples from the expressed cell cultures to a known total protein concentration.

Western Blotting

Gels were loaded and ran at 200V for 55 minutes and blotted 0.4 mA for 1 hour onto nitrocellulose membrane in methanol free blotting buffer. The nitrocellulose blots were blocked for 1 hour with 0.5% BSA in PBS-0.1% Tween 20 and then probed with an antibody to BoNT/E1 for 1 hour. The blots were detected with HRP conjugated secondary antibody developed with SuperSignal DuraWest substrate. The developed blots were imaged using a Syngene Imaging Instrument (Figure 3).

EXAMPLE 6

Initial purification and activation of target BoNT/E1 protein to di-chain form

This example was based on one combination of capture and intermediate column steps, though the combination could be altered or reversed to use the same properties in a different order. The clarified supernatant was brought to a high salt concentration and loaded onto a hydrophobic capture column (butyl sepharose). The bound rBoNT/E1 was eluted from the column using a gradient of low salt Tris buffer. The eluted protein was then further purified using an ion-exchange column such as Q-sepharose, eluting a gradient of high salt Tris buffer. Trypsin was then added to the eluted rBoNT/E1 sample to a final concentration of 2.5 μg/ml and incubated at 37 °C for 40 min. This nicked the BoNT/E1 activation loop and formed the final BoNT/E1 di-chain structure, as confirmed by reducing SDS-PAGE (Figure 4).

EXAMPLE 7

Final purification of target BoNT/E1 protein free from activating protease

The activated rBoNT/E1 sample was loaded immediately in high salt buffer onto a hydrophobic column (butyl sepharose). The column was washed with high salt buffer to remove the weakly associated trypsin, before a gradient of low salt Tris buffer was applied to further remove the trypsin from the column and the bound rBoNT/E1 protein. The rBoNT/E1 protein was then eluted late in the gradient, away from the trypsin.

Assay to Determine Trypsin levels

A trypsin ELISA was developed to determine the levels present in the column fractions and in the final BoNT/E1 sample. An anti-trypsin capture antibody was coated to micro-titre plates for 1hour at 37°C. Trypsin standards and test samples were added onto the plate (100 μL/well) and incubated for 1h at 37°C before detection with a second anti-trypsin antibody. The amount of trypsin in each sample/column fraction was then interpolated from the standards and overlaid on the purification chromatogram to confirm the separation of the trypsin from the BoNT/E1 (Figure 2B).

EXAMPLE 8

Formulation comprising active di-chain BoNT/E1 substantially free from trypsin

The following six liquid compositions comprising active di-chain BoNT/E1 were prepared (Table 3).

	1	2	3	4	5	6
Polysorbate 80	0.10 mg/mL	0.10 mg/mL	0.10 mg/mL	0.10 mg/mL	-	-
Poloxamer	-	-	-	-	0.04 mg/mL	0.04 mg/mL
Sucrose	4.0 mg/mL	-	4.0 mg/mL	-	4.0 mg/mL	-
Mannitol	-	4.0 mg/mL	-	4.0 mg/mL	-	4.0 mg/mL
Sodium Chloride	8.76 mg/mL	8.76 mg/mL	8.76 mg/mL	8.76 mg/mL	8.76 mg/mL	8.76 mg/mL
pH	6.5	6.5	6.5	6.5	6.5	6.5
Buffer	L-Histidine / Hydrochloric acid	L-Histidine / Hydrochloric acid	Di sodium phosphate / Citric acid anhydrous	Di sodium phosphate / Citric acid anhydrous	L-Histidine / Hydrochloric acid	L-Histidine / Hydrochloric acid
Di-Chain BoNT/E1	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL	20 ng/mL
MilliQ water	q.s. to 1 mL	q.s. to 1 mL	q.s. to 1 mL	q.s. to 1 mL	q.s. to 1 mL	q.s. to 1 mL

All six compositions were stored at 25°C for 12 weeks. The stability of the di-chain BoNT/E1 protease function was assessed during that period using a cell free endopeptidase assay. The monthly degradation rates for the six formulations were below 5% per month over the 12 weeks, which shows the di-chain BoNT/E1 protease function of the six compositions remains stable at 25°C for at least 12 weeks.

CLAIMS

1. A nucleic acid sequence comprising a sequence of contiguous nucleotides, wherein said sequence of contiguous nucleotides has at least 90% sequence identity to the nucleic acid sequence of SEQ ID NO: 1,
and wherein said sequence of contiguous nucleotides encodes a single-chain BoNT/E1 protein.

2. The nucleic acid sequence according to Claim 1, wherein the nucleic acid has a maximum of 160 slow codons.

3. The nucleic acid sequence according to Claim 1 or Claim 2, wherein said single-chain BoNT/E1 protein comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% sequence identity to the amino acid sequence of SEQ ID NO: 2.

4. The nucleic acid sequence according to any one of Claims 1-3, with the proviso that the single-chain BoNT/E1 according to Claim 1 or Claim 2 or said sequence of contiguous amino acids according to Claim 3 includes one or more of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein):
glycine at position 177
serine at position 198
alanine at position 340
leucine at position 773
leucine at position 963
glutamine at position 964
alanine at position 967
asparagine at position 1195.

5. The nucleic acid sequence according to any preceding claim, wherein said sequence of contiguous nucleotides has at least 785 synonymous codons when compared to the nucleic acid sequence of wild-type BoNT/E1 (SEQ ID NO: 3).

6. A method for producing soluble single-chain BoNT/E1 protein in an *E. coli* host cell, said method comprising:
expressing a nucleic acid sequence according to any preceding claim in an *E. coli* expression system.

7. The method according to Claim 6, wherein said soluble single-chain BoNT/E1 protein is expressed in the cytoplasm of said *E. coli* host cell.

8. The method according to Claim 6 or Claim 7, wherein said soluble single-chain BoNT/E1 protein is expressed at a level of at least 5 mg/L.

9. The method according to any of Claims 6-8, comprising lysis of the *E. coli* host cell to provide an *E. coli* host cell homogenate containing said soluble single-chain BoNT/E1 protein.

10. A method for producing soluble di-chain BoNT/E1 protein, said method comprising:
providing a soluble single-chain BoNT/E1 protein comprising a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% sequence identity to the amino acid sequence of SEQ ID NO: 2,
and contacting said BoNT/E1 protein with trypsin in solution.

11. The method according to Claim 10, with the proviso that said sequence of contiguous amino acids includes one or more of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein):
glycine at position 177
serine at position 198
alanine at position 340
leucine at position 773
leucine at position 963
glutamine at position 964
alanine at position 967

asparagine at position 1195.

12. The method according to Claim 10 or Claim 11, wherein the soluble single-chain BoNT/E1 protein is provided by the method of any of Claims 6-9.

13. The method according to any of Claims 10-12, comprising separating the soluble BoNT/E1 protein from the trypsin by contacting the solution containing soluble BoNT/E1 protein and trypsin with a hydrophobic surface, wherein the soluble BoNT/E1 protein preferentially binds to the hydrophobic surface.

14. The method according to Claim 13, wherein the hydrophobic surface is an inert matrix to which a ligand consisting of aryl or alkyl groups is attached.

15. The method according to Claim 14, wherein the hydrophobic surface is selected from the group consisting of: butyl, phenyl or octyl ligands.

16. An active di-chain BoNT/E1 protein,
wherein the first chain comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% sequence identity to the amino acid sequence of positions 1-419 of SEQ ID NO: 2;
wherein the second chain comprises a sequence of contiguous amino acids, and wherein said sequence of contiguous amino acids has at least 95% sequence identity to the amino acid sequence of positions 423-1252 of SEQ ID NO: 2;
wherein the first and second chains are joined together by a disulphide bond between cysteine 412 on the first chain and cysteine 426 on the second chain;
with the proviso that said sequence of contiguous amino acids includes one or more of the following amino acids (wherein the amino acid position numbering starts with the N-terminal amino acid residue and ends with the C-terminal amino acid residue of the BoNT/E1 protein):
glycine at position 177
serine at position 198
alanine at position 340
leucine at position 773
leucine at position 963

glutamine at position 964
alanine at position 967
asparagine at position 1195.

17. An active di-chain BoNT/E1 protein obtainable by the method according to any of Claims 10-15.

18. A composition comprising an active di-chain BoNT/E1 protein according to Claim 16 or Claim 17, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein according to any of Claims 6-9; wherein said composition is substantially free from trypsin.

19. The composition according to Claim 18, where said composition contains less than 10 pg trypsin per 100 ng BoNT/E1 protein, or less than 7 pg trypsin per 100 ng BoNT/E1 protein, or less than 5 pg trypsin per 100 ng BoNT/E1 protein.

20. A liquid pharmaceutical composition comprising:

an active di-chain BoNT/E1 protein according to Claim 16 or Claim 17, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein according to any of Claims 6-9;

a non-protein stabilising agent that is a surfactant; and

water;

wherein said liquid pharmaceutical composition does not comprise a protein stabilising agent; and

wherein said liquid pharmaceutical composition is substantially free from trypsin (e.g. said liquid pharmaceutical composition contains less than 10 pg trypsin per 100 ng BoNT/E1 protein, or less than 7 pg trypsin per 100 ng BoNT/E1 protein, or less than 5 pg trypsin per 100 ng BoNT/E1 protein).

21. The liquid pharmaceutical composition according to claim 20, wherein said liquid pharmaceutical composition further comprises:

sodium chloride,

a buffer to maintain pH between 5.5 and 7.5, and

a disaccharide;

wherein the water is sterile water.

22. An active di-chain BoNT/E1 protein according to Claim 16 or Claim 17, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein according to any of Claims 6-9, or a composition according to Claim 18 or Claim 19, or a liquid pharmaceutical composition according to Claim 20 or Claim 21, for use in therapy.