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.

Summary

In one aspect, there is provided 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.

In another aspect, there is provided 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.

In yet another aspect, there is provided 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, wherein said sequence of contiguous amino acids has at least 95% sequence identity to the amino acid sequence of SEQ ID NO: 2, contacting said BoNT/E1 protein with trypsin in solution, and 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.

In yet another aspect, there is provided 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.

In yet another aspect, there is provided an active di-chain BoNT/E1 protein obtainable by the method as described herein.

In yet another aspect, there is provided a composition comprising an active di-chain BoNT/E1 protein as described herein, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein as described herein; wherein said composition is substantially free from trypsin.

In yet another aspect, there is provided a liquid pharmaceutical composition comprising: an active di-chain BoNT/E1 protein as described herein, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein as described herein; 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).

In yet another aspect, there is provided an active di-chain BoNT/E1 protein as described herein, or an active di-chain BoNT/E1 protein obtainable by proteolytic cleavage of the single-chain BoNT/E1 protein as described herein, or a composition as

described herein, or a liquid pharmaceutical composition as described herein, for use in therapy.

Description

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