(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 April 2007 (19.04.2007)

PCT

(10) International Publication Number
WO 2007/044382 A2

(51) International Patent Classification:
*A61K 39/08* (2006.01) *C12P 21/06* (2006.01)

(21) International Application Number:
PCT/US2006/038757

(22) International Filing Date: 5 October 2006 (05.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/724,274 7 October 2005 (07.10.2005) US
60/742,900 7 December 2005 (07.12.2005) US

(71) Applicants and
(72) Inventors: **SHONE, Clifford, C.** [GB/GB]; Health Protection Agency, Porton Down, Salisbury Wiltshire SP4 0JG (GB). **CRAWFORD, James, A.** [US/US]; 5445a Briardale Lane, Dublin, Ohio 43016 (US).

(74) Agents: **O'CONNOR, Steven, P.** et al.; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 901 New York Avenue, NW, Washington, District Of Columbia 20001-4413 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**
— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 2007/044382 A2

(54) Title: PROTEINS WITH IMPROVED SOLUBILITY AND METHODS FOR PRODUCING AND USING SAME

(57) Abstract: A method is provided for improving the solubility of proteins, for example, bacterial toxins. In one embodiment, solubility is improved by introducing point mutations that replace cysteine residues capable of forming intermolecular disulfide bonds with other amino acid residues that do not form such bonds. By abrogating the ability of the cysteine residues to form inter-molecular disulfide bonds, aggregation of the protein is reduced, thereby improving the solubility of the protein. In another embodiment, solubility of the protein is improved by producing truncated forms of the protein that express the LHN domain and a fragment of the Hc domain. Proteins made according to the method of the invention are useful, for example, as immunodiagnostic agents and vaccine components.

# PROTEINS WITH IMPROVED SOLUBILITY AND METHODS FOR PRODUCING AND USING SAME

## CROSS REFERENCE TO RELATED APPLICATIONS

[001] The present application claims the benefit of United States provisional application nos. 60/724,274, filed October 7, 2005, and 60/742,900, filed December 7, 2005, the entire disclosures of which are incorporated by reference.

## DESCRIPTION OF THE INVENTION

### Field of the Invention

[002] The present invention relates to methods for producing recombinant proteins that exhibit improved utility and process characteristics, such as solubility, compared to the corresponding native proteins. The invention also relates to proteins made according to the present methods, nucleic acids encoding the proteins, and the use of the proteins for prophylactic and therapeutic applications.

### Background of the Invention

[003] Proteins produced by organisms, and in particular microorganisms such as bacteria, are of interest because of their potential to serve as immunodiagnostic reagents, therapeutic agents, and vaccine components. Toxins are one group of proteins that have been extensively investigated for those purposes. It is often desirable, if not necessary, to purify proteins to remove contaminating materials that render them unsuitable for those uses. However, some proteins will form aggregates during

purification. Aggregates tend to exhibit low solubility and other characteristics that are undesirable, such as low immunogenicity. In some cases, aggregates arise when cysteine residues in the protein of interest form aberrant inter-molecular and/or intra-molecular disulfide bonds.

[004] One family of bacterial toxins of interest as immunodiagnostic reagents, therapeutic agents, and vaccine components are the clostridial neurotoxins, such as those from *Clostridium botulinum* and *Clostridium butyricum.* Botulinum neurotoxin (BoNT) is one of the most potent toxins known to man. Its ingestion or inhalation inhibits neurotransmitter release from synaptic vesicles, resulting in neuroparalysis and death. The use of *Clostridium botulinum* neurotoxins as vaccine components is disclosed in U.S. Patent No. 5,919,665 to J. A. Williams, which is incorporated by reference into this application. In addition, U.S. Patent Nos. 6,051,239 to Simpson *et al.*, 6,287,566 to M. T. Dertzbaugh, and 6,461,617 to Shone *et al*, each of which is incorporated by reference into this application, disclose the use of fragments of clostridial neurotoxin as vaccine components.

[005] Seven serologically distinct forms of clostridial neurotoxin exist: types A, B, C, D, E, F, and G. Full length neurotoxin type E, for example, is designated BoNT/E. Each neurotoxin type shares a common architecture in which a catalytic L-chain (LC, ~ 50 kDa) is disulfide linked to a receptor binding and translocating H-chain (HC, ~ 100 kDa). The HC polypeptide comprises all or part of two distinct functional domains. The carboxy-terminal half of the HC (~ 50 kDa), termed the $H_C$ domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the

target neuron. The amino-terminal half, termed the $H_N$ domain (~ 50 kDa), mediates the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. Although the heavy chain is required for BoNT to bind and enter the target cell, it is not toxic by itself.

[006] One particular fragment of interest is the $LH_N$ fragment, such as the $LH_N$ fragment of neurotoxin E ($LH_N$/E). $LH_N$/E corresponds to the first 845 N-terminal amino acid residues of the full length botulinum (or butyricum) neurotoxin E. It includes the LC and $H_N$ domains. During *in vivo* expression, as well as during purification, both recombinant $LH_N$/E ($rLH_N$/E) and native forms of $LH_N$/E form aggregates having a molecular mass ranging from about 120kD to several million kD. Often $LH_N$/E aggregates having masses of about 200kD, 300kD, 400kD, 500kD, 600kD, 700kD, and 800kD are observed. Although aggregated $rLH_N$/E can be recovered from insoluble lysate material by detergent extraction/reductant treatment and further purified (~90%) by anion exchange (Q Sepharose) and gel filtration (Superdex 200) chromatography, the recovered aggregate has undesirable properties. For example, purified aggregated $rLH_N$/E is recognized in a conformation sensitive ELISA to a much lesser degree (~5-10-fold) compared to the native BoNT/E control, indicating that conformational epitopes are absent and/or buried within the aggregate. Further, animal efficacy data indicate that immunization with aggregated $LH_N$/E does not protect animals against BoNT/E toxin challenge. Because conformational epitopes are known to play a key role in

eliciting protective antibody responses, these results were not totally unexpected.

[007] Different fermentation conditions, for example, slow initial growth, less potent inducers, and/or reduced induction temperatures, as well as different detergent extraction/reductant treatments and denaturation (e.g. urea)/refolding methodologies have been being explored in an effort to produce soluble, non-aggregated or less aggregated $LH_N/E$, but with limited success. Hence, there is a need for protein toxins and toxin subfragments, such as $LH_N/E$, that exhibit little or no aggregation and retain conformational epitopes that permit use of the toxins as immunodiagnostic reagents and vaccine components.

## SUMMARY OF THE INVENTION

[008] An object of the invention is to provide a method for reducing or preventing aggregate formation during purification and/or formulation of proteins, such as toxins and toxin fragments.

[009] Another object of the invention is to provide greater batch-to-batch consistency within protein products when characterized by standard methods of protein analysis.

[010] Still another object of the invention is to provide proteins with improved solubility and process characteristics.

[011] Yet another object of the invention is to provide toxin proteins and toxin fragments with improved solubility and process characteristics for use as immunodiagnostic reagents, therapeutic agents, and vaccine components.

[012] Additional objects and advantages of the invention will be set forth in part in the description that follows, and in part will be obvious from the description, or may be learned by practice of the invention. The objects and advantages of the invention will be realized and attained by the compositions and methods particularly pointed out in the appended claims.

[013] To achieve the objects and in accordance with the purpose of the invention, as embodied and broadly described herein, the disclosure describes in one embodiment a recombinant protein comprising at least one point mutation that substitutes a cysteine residue with another amino acid residue, wherein said substitution improves the solubility of the recombinant protein. In some embodiments, the protein is a toxin, such as a bacterial toxin, or a fragment thereof. The toxin may be a neurotoxin or neurotoxin fragment from, for example, *Clostridium botulinum* or *Clostridium butyricum.* In certain embodiments, the toxin is a clostridial neurotoxin, such as neurotoxin E, or a fragment thereof, such as an $LH_N/E$ fragment.

[014] The disclosure also provides toxin fragments that are more soluble than certain other fragments. For example, in some embodiments, the toxin is a clostridial neurotoxin and the fragment is an $LH_N$ fragment that further comprises amino acid sequences from the $H_C$ fragment, wherein the resulting $LH_N+H_C$ fragment is more soluble than the $LH_N$ fragment. In certain embodiments, these fragments are recombinant neurotoxin E fragments.

[015] The invention also comprises nucleic acids encoding the recombinant proteins set forth in the disclosure, vectors comprising those

nucleic acid sequences, and methods of expressing the encoded proteins in host cells.

[016] In yet other embodiments, the invention encompasses methods used in improving the solubility and process characteristics of the toxin and toxin fragments described in the specification.

[017] The invention further comprises methods of using the disclosed toxin and toxin fragments as therapeutic agents and vaccine components.

[018] Thus, the invention provides the following embodiments:

[019] In one embodiment, then invention provides a recombinant protein comprising at least one point mutation that substitutes a cysteine residue with another amino acid residue, wherein said substitution improves the solubility of the recombinant protein.

[020] In another embodiment, the protein is a toxin or non-toxic derivative of a toxin.

[021] Still other embodiments of the invention encompass a toxin or non-toxic derivative of a toxin that is of bacterial origin.

[022] In other embodiments, the bacterial toxin or toxin derivative is from either *Clostridium botulinum* or *Clostridium butyricum.*

[023] In yet other embodiments, the toxin is a neurotoxin or neurotoxin derivative.

[024] In still other embodiments, the neurotoxin is neurotoxin A, B, C, D, E, F, or G, or a non-toxic derivative thereof.

[025] In yet other embodiments, the neurotoxin or non-toxic derivative is a fragment of neurotoxin E.

[026] In other embodiments the fragment is the $LH_N/E$ fragment of neurotoxin E.

[027] In still other embodiments, the neurotoxin E fragment comprises at least one cysteine to serine amino acid substitution.

[028] In yet other embodiments, the substitution of serine for cysteine occurs at amino acid residue 26, amino acid residue 347, or both amino acid residue 26 and amino acid residue 347 compared to the $LH_N$ fragment of SEQ ID NO: 1 or SEQ ID NO: 2.

[029] The invention includes embodiments in which a protein of the invention has active endopeptidase activity.

[030] The invention also includes embodiments in which a protein of the invention has attenuated endopeptidase activity.

[031] Among other embodiments, the invention also includes nucleic acids encoding a recombinant protein of the invention.

[032] Similarly, other embodiments of the invention include a method for improving the solubility of a protein having at least one cysteine residue that forms an intermolecular disulfide bond, comprising:

(a) providing a nucleic acid sequence encoding a recombinant protein comprising at least one cysteine residue;

(b) introducing at least one point mutation into the nucleic acid sequence that substitutes at least one cysteine residue with another amino acid residue;

(c) transforming a host cell with the mutated nucleic acid sequence; and

(d) expressing the nucleic acid sequence to produce the protein.

[033] In some embodiments, the protein in the method is a toxin or non-toxic derivative thereof.

[034] In other embodiments, the protein in the method is a toxin or non-toxic derivative of bacterial origin.

[035] In yet other embodiments, the protein in the method is a bacterial toxin or toxin derivative from either *Clostridium botulinum* or *Clostridium butyricum.*

[036] In still other embodiments, the protein in the method is a toxin a neurotoxin or neurotoxin derivative.

[037] In some embodiments, the neurotoxin in the method is neurotoxin A, B, C, D, E, F, or G, or a non-toxic derivative thereof.

[038] In other embodiments, the non-toxic derivative in the method is a is a fragment of neurotoxin E.

[039] In yet other embodiments, the fragment in the method is the $LH_N$/E fragment of neurotoxin E.

[040] In some embodiments of the method, the amino acid introduced by the at least one point mutation is a serine.

[041] In other embodiments of the method, the protein is a $LH_N$ fragment of clostridial neurotoxin E and the at least one point mutation substitutes a serine for a cysteine at amino acid residue 26, amino acid residue 347, or both amino acid residue 26 and amino acid residue 347 compared to the $LH_N$ fragment of SEQ ID NO: 1 or SEQ ID NO: 2.

[042] In still other embodiments of the method, the point mutation is introduced by site-directed mutagenesis.

[043] Various embodiments of the method further comprise isolating the protein.

[044] In certain embodiments of the method, the host cell is a mammalian, plant, insect, fungal, or bacterial cell.

[045] The methods of the invention include embodiments in which a protein of the invention has active endopeptidase activity.

[046] The methods of the invention also include embodiments in which a protein of the invention has attenuated endopeptidase activity.

[047] In some embodiments, the invention provides for the use of a protein of the invention for the manufacture of a medicament for the treatment or prevention of botulism.

[048] Other embodiments of the invention include compositions comprising a protein of the invention and a pharmaceutically acceptable carrier.

[049] In still other embodiments, the invention provides methods of protecting an individual from botulism, comprising administering to the individual a composition of the invention.

[050] Yet other embodiments of the invention provide a method of producing antibodies that neutralize a clostridial neurotoxin, comprising administering a composition of the invention to an animal, allowing the animal to develop neutralizing antibodies to the clostridial neurotoxin, and isolating an antiserum that neutralizes the clostridial neurotoxin from the animal.

[051] Other embodiments encompass an antiserum produced by a method of the invention.

[052] In still other embodiments, the invention provides methods of treating exposure to a clostridial neurotoxin, comprising administering to a patient that has been exposed to the clostridial neurotoxin an antiserum of the invention.

[053] In other embodiments, the invention provides a recombinant protein comprising a truncated botulinum serotype E toxin, wherein the truncation improves the solubility of the recombinant protein.

[054] In some embodiments, the protein comprises a truncation in the Hc domain.

[055] In other embodiments, the truncated protein comprises the $LH_N$/E domain and the amino terminal 103 amino acids of the Hc domain.

[056] Still other embodiments of the invention encompass a truncated protein comprising the amino terminal 948 amino acids of the serotype E toxin.

[057] In yet other embodiments, the truncated protein comprises the $LH_N$/E domain and the amino terminal 202 amino acids of the Hc domain.

[058] Still other embodiments of the invention encompass a truncated protein comprising the amino terminal 1047 amino acids of the serotype E toxin.

[059] In other embodiments, the truncated protein comprises the $LH_N$/E domain and the amino terminal 304 amino acids of the Hc domain.

[060] Still other embodiments of the invention encompass a truncated protein comprising the amino terminal 1149 amino acids of the serotype E toxin.

[061] Additional embodiments of the invention include nucleic acids encoding a truncated botulinum serotype E toxin.

[062] In still other embodiments, the invention provides methods for improving the solubility of a clostridial neurotoxin, comprising:

(a) providing a nucleic acid sequence encoding a clostridial neurotoxin;

(b) modifying the nucleic acid sequence so that it encodes the $LH_N$ fragment and a portion of the $H_C$ fragment of the neurotoxin;

(c) transforming the modified nucleic acid sequence into a host cell capable of expressing the modified nucleic acid sequence; and

(d) expressing the modified nucleic acid sequence to produce the protein.

[063] In yet other embodiments, the invention provides for use of a truncated botulinum serotype E toxin for the manufacture of a medicament for the treatment or prevention of botulism.

[064] In other embodiments, the invention provides a composition comprising a truncated botulinum serotype E toxin and a pharmaceutically acceptable carrier.

[065] In still other embodiments, the invention provides a method of protecting an individual from botulism, comprising administering to the

individual a composition comprising a truncated botulinum serotype E toxin and a pharmaceutically acceptable carrier.

[066] Other embodiments of the invention include a method of producing antibodies that neutralize a clostridial neurotoxin, comprising administering a composition comprising a truncated botulinum serotype E toxin and a pharmaceutically acceptable carrier to an animal, allowing the animal to develop neutralizing antibodies to the clostridial neurotoxin, and isolating an antiserum that neutralizes the clostridial neurotoxin from the animal.

[067] In still other embodiments, the invention provides an antiserum produced by the method of the preceding paragraph.

[068] In yet other embodiment, the invention provides methods of treating exposure to a clostridial neurotoxin, comprising administering to a patient that has been exposed to the clostridial neurotoxin an antiserum of the invention.

[069] In yet another embodiment, the invention provides a mutated botulinum serotype E toxin comprising either or both of a leucine residue substituted for the tryptophan residue at position 1223 and a phenylalanine residue for the tyrosine residue at position 1224 of SEQ ID NO: 1 or SEQ ID NO: 2.

[070] The invention also provides, in additional embodiments, for the use of a protein of the preceding paragraph for the manufacture of a medicament for the treatment or prevention of botulism.

[071] In still other embodiments, the invention provides an *in vitro* method for improving the solubility of a protein having at least one cysteine residue that forms an intermolecular disulfide bond, comprising:

(a) providing a nucleic acid sequence encoding a recombinant protein comprising at least one cysteine residue;

(b) introducing at least one point mutation into the nucleic acid sequence that substitutes at least one cysteine residue with another amino acid residue;

(c) transforming a host cell with the mutated nucleic acid sequence; and

(d) expressing the nucleic acid sequence to produce the protein.

[072] Yet other embodiments of the invention encompass an *in vitro* method for improving the solubility of a clostridial neurotoxin, comprising:

(a) providing a nucleic acid sequence encoding a clostridial neurotoxin;

(b) modifying the nucleic acid sequence so that it encodes the $LH_N$ fragment and a portion of the $H_C$ fragment of the neurotoxin;

(c) transforming the modified nucleic acid sequence into a host cell capable of expressing the modified nucleic acid sequence; and

(d) expressing the modified nucleic acid sequence to produce the protein.

[073] In still other embodiments, the invention provides a method of producing antibodies that neutralize a clostridial neurotoxin, comprising

isolating antibodies elicited by an inoculated polypeptide, wherein said polypeptide is a protein according to the invention.

[074] Yet other embodiments of the invention provided for the use of an antiserum of the invention for the manufacture of a medicament for treating exposure to clostridial neurotoxin.

[075] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed. The accompanying drawings illustrate embodiments of the invention and together with the description, serve to explain the principles of the invention.

## BRIEF DESCRIPTION OF THE DRAWINGS

[076] Figs. 1A and 1B are a polyacrylamide gel and chromatogram, respectively, showing production of recombinant $LH_N$/E in *Escherichia coli* as a high molecular weight aggregate.

[077] Figs. 2A and 2B demonstrate high levels of mutated recombinant $LH_N$/E in *Escherichia coli* as non-aggregated proteins.

[078] Figs. 3A and 3B are Coomassie Blue stained gels demonstrating that the solubility of the $LH_N$/E-Hc protein increases as the length of the Hc sequence included in the protein increases.

[079] Figs. 4A and 4B are western blots demonstrating that the solubility of the $LH_N$/E-Hc protein increases as the length of the Hc sequence included in the protein increases.

[080] Figs. 5A and 5B are a Coomassie Blue stained gel and western blot analysis, respectively, of the solubility of the $LH_N/E$ and $LH_N/E$- $H_C406$ proteins in the presence of the reducing agent DTT.

[081] Fig. 6 is a western blot comparing protein levels in total lysate, the soluble fraction, and the insoluble fraction for $LH_N$ proteins comprising Cys to Ser replacements at positions 26 and 347, and having or lacking an Hc fragment.

## DESCRIPTION OF THE EMBODIMENTS

[082] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. All documents, or portions of documents, cited herein, including but not limited to patents, patent applications, articles, books, and treatises, are hereby expressly incorporated by reference in their entirety for any purpose. In the event that one or more of the incorporated documents or portions of documents defines a term that contradicts that term's definition in the application, this application controls.

[083] The use of the singular includes the plural unless specifically stated otherwise. The word "a" or "an" means "at least one" unless specifically stated otherwise. The use of "or" means "and/or" unless stated otherwise. The meaning of the phrase "at least one" is equivalent to the meaning of the phrase "one or more." Furthermore, the use of the term "including," as well as other forms, such as "includes" and "included," is not limiting. Also, terms such as "element" or "component" encompass both

elements or components comprising one unit and elements or components comprising more than one unit unless specifically stated otherwise.

[084] Neurotoxic proteins and fragments of these proteins are important immunodiagnostic reagents, therapeutic agents, and vaccine components. Functional neurotoxins are hazardous to work with, however, so investigators prefer to use recombinant proteins that have been genetically modified to reduce or eliminate their neurotoxicity. Unfortunately, it can be difficult to purify some of the recombinant, non-toxic, proteins because they often form aggregates, which have reduced solubility and are less effective reagents for use in immunodiagnostic, therapeutic, and vaccine applications. For example, although aggregated $rLH_N/E$ can be purified, it is recognized in a conformation-sensitive ELISA to a much lesser degree (~5-10-fold) than is the native BoNT/E toxin, indicating that conformational epitopes are absent and/or buried within the aggregate. Also, immunization with aggregated $LH_N/E$ does not protect animals against BoNT/E toxin challenge.

[085] Accordingly, the disclosure provides recombinant proteins with improved solubility. For example, in one embodiment, the disclosure describes a recombinant protein comprising at least one point mutation that substitutes a cysteine residue with another amino acid residue, wherein said substitution improves the solubility of the recombinant protein. In some embodiments, the protein is a toxin, such as a bacterial toxin, for example, a neurotoxin or neurotoxin fragment from *Clostridium botulinum* or *Clostridium butyricum*. In some embodiments, the protein is a fragment of a neurotoxin, such as an $LH_N$ fragment, for example, an $LH_N/E$ fragment. In certain

embodiments, the first and third cysteine residues, counting from the amino terminus of a naturally-occurring neurotoxin amino acid sequence, have been replaced with non-cysteine amino acids, such as serine, in the recombinant neurotoxin protein or fragment thereof. Although mutations in clostridial neurotoxin E are exemplified, clostridial neurotoxins A, B, C, D, F, and G can be modified in the same manner.

[086] Nucleic acid sequences encoding various neurotoxins have been cloned and those nucleic acid sequences are known in the art. For example, a nucleic acid sequence of a full length neurotoxin E from *C. botulinum* is provided in GenBank accession no. AB082519. A nucleic acid sequence of a full length neurotoxin E from *C. butyricum* is provided in GenBank accession no. AB088207.

[087] An example of an amino acid sequence of *C. botulinum* BoNT/E neurotoxin is:

MPKINSFNYN DPVNDRTILY IKPGGCQEFY KSFNIMKNIW IIPERNVIGT
TPQDFHPPTS LKNGDSSYYD PNYLQSDEEK DRFLKIVTKI FNRINNNLSG
GILLEELSKA NPYLGNDNTP DNQFHIGDAS AVEIKFSNGS QDILLPNVII
MGAEPDLFET NSSNISLRNN YMPSNHRFGS IAIVTFSPEY SFRFNDNCMN
EFIQDPALTL MHELIHSLHG LYGAKGITTK YTITQKQNPL ITNIRGTNIE
EFLTFGGTDL NIITSAQSND IYTNLLADYK KIASKLSKVQ VSNPLLNPYK
DVFEAKYGLD KDASGIYSVN INKFNDIFKK LYSFTEFDLR TKFQVKCRQT
YIGQYKYFKL SNLLNDSIYN ISEGYNINNL KVNFRGQNAN LNPRIITPIT
GRGLVKKIIR FCKNIVSVKG IRKSICIEIN NGELFFVASE NSYNDDNINT
PKEIDDTVTS NNNYENDLDQ VILNFNSESA PGLSDEKLNL TIQNDAYIPK
YDSNGTSDIE QHDVNELNVF FYLDAQKVPE GENNVNLTSS IDTALLEQPK
IYTFFSSEFI NNVNKPVQAA LFVSWIQQVL VDFTTEANQK STVDKIADIS
IVVPYIGLAL NIGNEAQKGN FKDALELLGA GILLEFEPEL LIPTILVFTI
KSFLGSSDNK NKVIKAINNA LKERDEKWKE VYSFIVSNWM TKINTQFNKR
KEQMYQALQN QVNAIKTIIE SKYNSYTLEE KNELTNKYDI KQIENELNQK
VSIAMNNIDR FLTESSISYL MKIINEVKIN KLREYDENVK TYLLNYIIQH
GSILGESQQE LNSMVTDTLN NSIPFKLSSY TDDKILISYF NKFFKRIKSS
SVLNMRYKND KYVDTSGYDS NININGDVYK YPTNKNQFGI YNDKLSEVNI
SQNDYIIYDN KYKNFSISFW VRIPNYDNKI VNVNNEYTII NCMRDNNSGW
KVSLNHNEII WTFEDNRGIN QKLAFNYGNA NGISDYINKW IFVTITNDRL

GDSKLYINGN LIDQKSILNL GNIHVSDNIL FKIVNCSYTR YIGIRYFNIF
DKELDETEIQ TLYSNEPNTN ILKDFWGNYL LYDKEYYLLN VLKPNNFIDR
RKDSTLSINN IRSTILLANR LYSGIKVKIQ RVNNSSTNDN LVRKNDQVYI
NFVASKTHLF PLYADTATTN KEKTIKISSS GNRFNQVVVM NSVGNCTMNF
KNNNGNNIGL LGFKADTVVA STWYYTHMRD HTNSNGCFWN FISEEHGWQE
K

(SEQ ID NO: 1). This sequence includes the Met at residue 1. SEQ ID NO: 1 is the reference sequence for all numbering regarding *C. botulinum* BoNT/E and fragments thereof used in this specification, irrespective of whether those sequences have or do not have the first Met. The $LH_N$/E fragment extends from the amino terminus to amino acid residue 845 (Lys) in SEQ ID NO: 1.

[088] An example of an amino acid sequence of *C. butryicum* BoNT/E neurotoxin is:

MPTINSFNYN DPVNNRTILY IKPGGCQQFY KSFNIMKNIW IIPERNVIGT
IPQDFLPPTS LKNGDSSYYD PNYLQSDQEK DKFLKIVTKI FNRINDNLSG
RILLEELSKA NPYLGNDNTP DGDFIINDAS AVPIQFSNGS QSILLPNVII
MGAEPDLFET NSSNISLRNN YMPSNHGFGS IAIVTFSPEY SFRFKDNSMN
EFIQDPALTL MHELIHSLHG LYGAKGITTK YTITQKQNPL ITNIRGTNIE
EFLTFGGTDL NIITSAQSND IYTNLLADYK KIASKLSKVQ VSNPLLNPYK
DVFEAKYGLD KDASGIYSVN INKFNDIFKK LYSFTEFDLA TKFQVKCRQT
YIGQYKYFKL SNLLNDSIYN ISEGYNINNL KVNFRGQNAN LNPRIITPIT
GRGLVKKIIR FCKNIVSVKG IRKSICIEIN NGELFFVASE NSYNDDNINT
PKEIDDTVTS NNNYENDLDQ VILNFNSESA PGLSDEKLNL TIQNDAYIPK
YDSNGTSDIE QHDVNELNVF FYLDAQKVPE GENNVNLTSS IDTALLEQPK
IYTFFSSEFI NNVNKPVQAA LFVGWIQQVL VDFTTEANQK STVDKIADIS
IVVPYIGLAL NIGNEAQKGN FKDALELLGA GILLEFEPEL LIPTILVFTI
KSFLGSSDNK NKVIKAINNA LKERDEKWKE VYSFIVSNWM TKINTQFNKR
KEQMYQALQN QVNALKAIIE SKYNSYTLEE KNELTNKYDI EQIENELNQK
VSIAMNNIDR FLTESSISYL MKLINEVKIN KLREYDENVK TYLLDYIIKH
GSILGESQQE LNSMVIDTLN NSIPFKLSSY TDDKILISYF NKFFKRIKSS
SVLNMRYKND KYVDTSGYDS NININGDVYK YPTNKNQFGI YNDKLSEVNI
SQNDYIIYDN KYKNFSISFW VRIPNYDNKI VNVNNEYTII NCMRDNNSGW
KVSLNHNEII WTLQDNSGIN QKLAFNYGNA NGISDYINKW IFVTITNDRL
GDSKLYINGN LIDKKSILNL GNIHVSDNIL FKIVNCSYTR YIGIRYFNIF
DKELDETEIQ TLYNNEPNAN ILKDFWGNYL LYDKEYYLLN VLKPNNFINR
RTDSTLSINN IRSTILLANR LYSGIKVKIQ RVNNSSTNDN LVRKNDQVYI
NFVASKTHLL PLYADTATTN KEKTIKISSS GNRFNQVVVM NSVGNCTMNF
KNNNGNNIGL LGFKADTVVA STWYYTHMRD NTNSNGFFWN FISEEHGWQE
K

(SEQ ID NO: 2). This sequence includes the Met at residue 1. SEQ ID NO: 2 is the reference sequence for all numbering regarding *C. butyricum* BoNT/E and fragments thereof used in this specification, irrespective of whether those sequences have or do not have the first Met. The $LH_N$/E fragment extends from the amino terminus to amino acid residue 845 (Lys) in SEQ ID NO: 2.

[089] In the case of neurotoxin proteins from *C. botulinum* and *C. butyricum*, any amino acid sequences disclosed in which the initial methionine is not included are also considered BoNT/E or $LH_N$/E fragments, as appropriate, even though the numbering of the amino acid residues in that particular fragment assumes the Met at position 1.

[090] BoNT/E and fragments thereof that have endopeptidase activity have a glutamate (E) at position 213 and a histidine (H) at position 216. The endopeptidase activity can be abolished by mutating these sequences. For example, the specification describes BoNT/E and fragments thereof in which the glutamate is replaced with glutamine (Q) (i.e., E213Q) and the histidine is replaced with a tyrosine (Y) (i.e., H216Y). These proteins lack endopeptidase activity.

[091] Examples of recombinant proteins comprising at least one point mutation that substitutes a cysteine residue with another amino acid residue, wherein said substitution improves the solubility of the recombinant protein include, but are not limited to:

a) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 1, wherein the cysteine at position 26 of SEQ ID NO: 1 is replaced with a serine;

b) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 1, wherein the cysteine at position 347 of SEQ ID NO: 1 is replaced with a serine;

c) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 1, wherein the cysteine at position 26 and the cysteine at position 347 of SEQ ID NO: 1 are each replaced with a serine;

d) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 2, wherein the cysteine at position 26 of SEQ ID NO: 2 is replaced with a serine;

e) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 2, wherein the cysteine at position 347 of SEQ ID NO: 2 is replaced with a serine; and

f) a protein comprising residues 2 to 845 of the amino acid sequence set forth in SEQ ID NO: 2, wherein the cysteine at position 26 and the cysteine at position 347 of SEQ ID NO: 2 are each replaced with a serine.

[092] In alternative embodiments, each of the proteins described in parts (a)-(f) of the preceding paragraph may consist of, rather than comprise, the respective amino acid sequences. Optionally, each of the proteins described in parts (a)-(f) of the preceding paragraph may further comprise a methionine at their respective amino termini. The solubility, immunogenicity, or both solubility and immunogenicity of the proteins described in parts (a)-(f) of the preceding paragraph is improved compared to proteins comprising or

consisting of the corresponding amino acid sequence lacking the mentioned cysteine to serine replacement(s). In some embodiments, the proteins contain the E213Q and H216Y point mutations that abolish endopeptidase activity.

[093] Additional examples of BoNT/E proteins or fragments thereof comprising Cys to Ser replacements are set forth in SEQ ID NOS: 7-14.

[094] The disclosure also provides recombinant neurotoxins fragments that are more soluble than certain other fragments. For example, in some embodiments, the fragment is an $LH_N$ fragment that further comprises amino acid sequences from the $H_C$ fragment, wherein the resulting $LH_N$+$H_C$ fragment is more soluble than the $LH_N$ fragment. In certain embodiments, the $LH_N$ fragment is an $LH_N$/E fragment and the $H_C$ fragment is an $H_C$/E fragment. The $LH_N$ fragment may optionally comprise at least one point mutation that substitutes a cysteine residue with another amino acid residue.

[095] Examples of recombinant neurotoxins fragments that are more soluble than certain other fragments include, but are not limited to:

a) a protein comprising residues 2 to 948 of the amino acid sequence set forth in SEQ ID NO: 1;

b) a protein comprising residues 2 to 1047 of the amino acid sequence set forth in SEQ ID NO: 1;

c) a protein comprising residues 2 to 1149 of the amino acid sequence set forth in SEQ ID NO: 1;

d) a protein comprising residues 2 to 948 of the amino acid sequence set forth in SEQ ID NO: 2;

e) a protein comprising residues 2 to 1047 of the amino acid sequence set forth in SEQ ID NO: 2; and

f) a protein comprising residues 2 to 1149 of the amino acid sequence set forth in SEQ ID NO: 2.

[096] In alternative embodiments, each of the proteins described in parts (a)-(f) of the preceding paragraph may consist of, rather than comprise, the respective amino acid sequences. Optionally, each of the proteins described in parts (a)-(f) of the preceding paragraph may further comprise a methionine at their respective amino terminus. The proteins described in parts (a)-(f) of the preceding paragraph may also optionally comprise a Cys to Ser substitution at amino acid residue 26, 347, or 26 and 347 of SEQ ID NO: 1, or SEQ ID NO: 2, as appropriate. The solubility, immunogenicity, or both solubility and immunogenicity of the proteins described in parts (a)-(f) of the preceding paragraph is improved compared to proteins consisting of amino acids 2-845 of the corresponding amino acid sequence. That is, the proteins described in parts (a)-(f) of the preceding paragraph have improved solubility, improved immunogenicity, or improved solubility and improved immunogenicity compared to the $LH_N$ fragment of SEQ ID NO: 1 or SEQ ID NO: 2. In some embodiments, the proteins contain the E213Q and H216Y point mutations that abolish endopeptidase activity.

[097] Additional examples of proteins comprising an extended $LH_N$ fragment are set forth in SEQ ID NO: 10 and SEQ ID NO: 14. Expression constructs $LH_N$/E-Hc103; $LH_N$/E-Hc202; $LH_N$/E-Hc304; and $LH_N$/E-Hc406 also

contain nucleic acid sequences encoding proteins comprising an extended $LH_N$ fragment.

[098] Proteins that comprise amino acid residues 1223 or 1224 (relative to SEQ ID NO: 1 or SEQ ID NO: 2) of neurotoxin E may further comprise an amino acid substitution at either residue 1223, residue 1224, or both residue 1223 and 1224. For example, the protein may comprise a tryptophan (W) to leucine (L) mutation at postions 1223 (i.e., W1223L), a tyrosine (Y) to phenylalanine (F) mutation at position 1224 (i.e., Y1224F), or a W1223L and a Y1224F double mutation. Examples of protein comprising these mutations are set forth in SEQ ID NO: 9 and SEQ ID NO: 13.

[099] The invention also comprises nucleic acids encoding the various recombinant proteins described in the specification, vectors comprising those nucleic acid sequences, and methods of expressing the encoded protein in a host cell. Thus, in some embodiments, the nucleic acids encode modified *C. botulinum* or *C. butyricum* neurotoxins, such as neurotoxin E, or fragments thereof, such as an $LH_N$ fragment, that have improved solubility, immunogenicity, or both improved solubility and immunogenicity compared to the corresponding unmodified neurotoxin or neurotoxin fragment. Methods of measuring solubility and immunogenicity are known in the art, and include, but are not limited to, the methods described in the Examples.

[0100] An improvement in solubility can be accomplished by changing codons in the nucleic acid sequence that code for the amino acid cysteine to another amino acid that does not form a disulfide bond. Alternatively, or in addition, solubility can be improved by extending the sequence of a fragment,

such as an $LH_N$ fragment of a clostridial neurotoxin, by providing additional sequences from an adjoining segment, such as an $H_C$ fragment of a clostridial neurotoxin.

[0101] Methods of manipulating nucleic acids and of expressing the encoded proteins are known in the art, and include those described in Molecular Cloning, A Laboratory Manual (2nd Ed., Sambrook, Fritsch and Maniatis, Cold Spring Harbor) and Current Protocols in Molecular Biology (Eds. Aufubel, Brent. Kingston, More, Feidman, Smith and Stuhl. Greene Publ. Assoc., Wiley-Interscience, NY, N.Y., 1992). Thus, it is possible to modify a nucleic acid sequence by replacing the codon for cysteine with a codon for another amino acid. In general, a cysteine is replaced with a serine, but other amino acid substitutions are also possible, such as replacement of cysteine with alanine, glycine, valine, leucine, isoleucine, or modified forms of these amino acids, so long as the replacement amino acid does not readily form disulfide bonds. Alternatively, the cysteine residue may simply be deleted from the sequence. Obviously, a deletion must remove the codon for the cysteine from the nucleic acid sequence without introducing a frameshift. Techniques for making substitution and deletion mutations at predetermined sites in a nucleic acid having a known sequence are well known and include, but are not limited to, primer mutagenesis and other forms of site-directed mutagenesis.

[0102] Similarly, methods of joining two sequence fragments, such as an $LH_N$ and an $H_C$ fragment of a clostridial neurotoxin, and of truncating a sequence, are known in the art. These include, but are not limited to, PCR-

based techniques and techniques for ligating together two or more nucleic acid sequences.

[0103] Certain methods of expressing proteins are described in the Examples. Other methods can also be used, however. Generally, in order to express a protein, such as a bacterial toxin or fragment thereof, a suitable cell line is transformed with a DNA sequence encoding that protein under the control of known regulatory sequences. The transformed host cells are cultured and the protein recovered and isolated from the culture medium. The isolated expressed proteins are substantially free from other proteins with which they are co-produced as well as from other contaminants. Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO), the monkey kidney COS-1 cell line, or mammalian CV-1 cells. The selection of suitable mammalian host cells and methods for transformation, culturing, amplification, screening, product production and purification are known in the art. (*See, e.g.*, Gething and Sambrook, *Nature*, 293:620-625 (1981); Kaufman *et al.*, *Mol Cell Biol.*, 5(7):1750-1759 (1985); Howley *et al.*, U.S. Patent 4,419,446.))

[0104] Bacterial cells may also be used as suitable hosts for expression of a bacterial toxin or fragment thereof. For example, various strains of *E. coli* (e.g., HB101, MC1061) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis, Pseudomonas*, other bacilli and the like may also be used. For expression of a protein in bacterial cells, DNA encoding the propeptide is generally not necessary.

[0105] In some embodiments, the bacterial toxin or fragment thereof is expressed using a vector that contains a DNA sequence encoding the protein and appropriate expression control sequences. Expression control sequences for such vectors are known to those skilled in the art and may be selected depending upon the host cells. In other embodiments, the bacterial toxin or fragment thereof is expressed as a fusion protein comprising the protein sequence of the bacterial toxin or fragment thereof and, for example, a tag to stabilize the resulting fusion protein or to simplify purification of the bacterial toxin or fragment thereof. Such tags are known in the art. Representative examples include sequences which encode a series of histidine residues, the epitope tag FLAG, the Herpes simplex glycoprotein D, beta-galactosidase, maltose binding protein, streptavidin tag or glutathione S-transferase.

[0106] The invention also encompasses the methods used for improving the solubility and process characteristics of a protein. For example, in some embodiments, the disclosure provides methods for improving the solubility and process characteristics of a protein having at least one cysteine residue that forms an intermolecular disulfide bond, comprising:

(a) providing a nucleic acid sequence encoding a recombinant protein comprising at least one cysteine residue;

(b) introducing at least one point mutation into the nucleic acid sequence that substitutes at least one cysteine residue with another amino acid residue;

(c) transforming a host cell with the mutated nucleic acid sequence; and

(d) expressing the nucleic acid sequence to produce the protein.

[0107] In other embodiments, the method comprises improving the solubility of a botulinum neurotoxin, comprising:

(a) providing a nucleic acid sequence encoding a botulinum neurotoxin;

(b) modifying the nucleic acid sequence so that it encodes the $LH_N$ fragment and a portion of the $H_C$ fragment of the neurotoxin;

(c) transforming the modified nucleic acid sequence into a host cell capable of expressing the modified nucleic acid sequence; and

(d) expressing the modified nucleic acid sequence to produce the protein.

[0108] The methods for improving the solubility of a protein, such as a botulinum neurotoxin, can be entirely *in vitro* methods. In other embodiments, as discussed herein, the methods can include an *in vivo* aspect, such expressing the nucleic acid *in vivo*.

[0109] Unless otherwise stated, a "soluble" recombinant protein is one that is exists in solution in the cytoplasm of the host cell. If the protein contains a signal sequence the soluble protein is exported to the periplasmic space in bacteria hosts and is secreted into the culture medium in eukaryotic cells capable of secretion or by bacterial host possessing the appropriate genes. In contrast, an insoluble protein is one which exists in denatured form inside cytoplasmic granules (called an inclusion bodies) in the host cell. A

soluble protein is a protein which is not found in an inclusion body inside the host cell or is found both in the cytoplasm and in inclusion bodies and in this case the protein may be present at high or low levels in the cytoplasm.

[0110] A soluble protein is distinct from a "solubilized" protein. An insoluble recombinant protein found inside an inclusion body may be solubilized (i.e., rendered into a soluble form) by treating purified inclusion bodies with denaturants such as guanidine hydrochloride, urea or sodium dodecyl sulfate (SDS). These denaturants must then be removed from the solubilized protein preparation to allow the recovered protein to renature (refold). A distinction is also made between proteins that are soluble (i.e., dissolved) in a solution devoid of significant amounts of ionic detergents (e.g., SDS) or denaturants (e.g., urea, guanidine hydrochloride) and proteins that exist as a suspension of insoluble protein molecules dispersed within the solution. A soluble protein will not be removed from a solution containing the protein by centrifugation using conditions sufficient to remove bacteria present in a liquid medium (e.g.,., centrifugation at 5,000g for 4-5 minutes). A method of testing whether a protein is soluble or insoluble is described in U.S. Patent No. 5,919,665, which is incorporated by reference.

[0111] The invention further encompasses methods of using the disclosed toxin and toxin fragments as therapeutic agents and vaccine components. Optionally, the disclosed toxin and toxin fragments are tested to ensure that they are free or substantially free of endotoxin activity. Methods of testing for endotoxin activity are known in the art.

[0112] Toxins and toxin fragments useful in vaccine compositions are those that can stimulate an antibody response that neutralizes a wild-type toxin of the same type. For example, when the toxin or toxin fragment is derived from clostridial type E neurotoxin, then the toxin or toxin fragment composition stimulates antibodies that neutralize the toxin activity of wild-type BoNT/E. By way of example only, one method for selecting clostridial neurotoxin toxins or neurotoxin fragments that can stimulate an antibody response that neutralizes wild-type BoNT activity is to determine whether the clostridial neurotoxin or neurotoxin fragment is immunoreactive with polyclonal neutralizing antibodies to wild-type BoNT of same type, such as BoNT/E. Methods of determining whether clostridial neurotoxin or neurotoxin fragment immunoreact with antibodies to wild-type BoNT include ELISA, western blot, double immunodiffusion assay, RIA, and the like. Another exemplary method comprises using the clostridial neurotoxin or neurotoxin fragments as an immunogen in mice, then determining whether the mice are protected from challenge with wild-type BoNT, such as wild-type BoNT/E.

[0113] A toxin or toxin fragment can be combined with a pharmaceutically acceptable carrier. Physiologically acceptable diluents include physiological saline solutions, and buffered saline solutions at neutral pH such as phosphate buffered saline. Other types of physiological carriers include liposomes and polymers. Optionally, the toxin or toxin fragments can be combined with an adjuvant. In some embodiments, the adjuvant is IC31™, produced by Intercell AG, Vienna, Austria. (See EP 1 326 634B and EP 1 296 713B.) In other embodiments, the adjuvant is a Toll-like receptor (TLR)

agonist, such as a TLR 4 agonist, a TLR7 agonist, or a TLR9 agonist. TLR9 agonists include, for example, immunostimulatory CpG nucleic acid sequences. Other examples of adjuvants that can be used include, but are not limited to, Freund's incomplete adjuvant, Freund's complete adjuvant, alum, monophosphoryl lipid A, alum phosphate or hydroxide, and QS-21.

[0114] For vaccine formulations, the toxins or toxin fragments can also be combined with immunomodulators, such as interleukins and interferons, for example IL-1, IL-12, and IFN-$\gamma$.

[0115] When the toxin or toxin fragment is a clostridial neurotoxin or neurotoxin fragment, multiple types of clostridial neurotoxin or neurotoxin fragments can be used together in a formulation, or a single type can be used alone. Thus, vaccine formulations and compositions include, but are not limited to, BoNT/E or a fragment of BoNT/E, such as $LH_N/E$, including $LH_N/E$ that is mutated and $LH_N/E$ that is extended by the inclusion of amino acid sequences from the $H_C$ fragments, either alone or in combination with wild-type, mutant, or fragments of one or more of clostridial neurotoxins type A, B, C, D, F, or G. Many vaccine formulations are known to those of skill in the art.

[0116] The toxin or toxin fragment is added to a vaccine formulation in an amount effective to stimulate a protective immune response in an animal challenged with wild-type toxin. Thus, in preparing a vaccine formulation, the toxin or toxin fragment is used for the manufacture of a medicament for the treatment or prevention of botulism. Generation of protective antibodies that neutralize the wild-type toxin can be measured by testing the ability of the vaccine to protect an animal, such as a mouse, from challenge with a lethal

dose of wild-type toxin. The amounts of the toxin or toxin fragment in the vaccine composition that can form a protective immune response are generally about 0.1 μg to 100 mg per kg of body weight. In some cases, about 1 μg to about 1 mg/kg body weight is used. Often, about 1 μg to about 100 μg toxin or toxin fragment per kg of body weight will be sufficient to stimulate a protective immune response, such as protective antibodies. An amount of toxin or toxin fragment that stimulates a protective immune response is considered to be an "effective amount."

[0117] Depending upon the circumstances, such as the animal to be vaccinated, either a single or multiple doses of the vaccine composition are administered to provide protective immunity against the wild-type toxin. The vaccine composition can be administered to an animal in a variety of ways, including subcutaneously, intramuscularly, intravenously, intradermally, orally, intranasally, ocularly, and intraperitoneally.

[0118] Any animal that is susceptible to the wild-type toxin can be vaccinated with the toxin or toxin fragment in an immunostimulatory composition. Examples of animals susceptible to clostridial neurotoxins include, but are not limited to, rabbits, rodents, birds, horses, cattle, and humans. Accordingly, a vaccine composition comprising a clostridial neurotoxin or neurotoxin fragment, such as the clostridial neurotoxins and neurotoxin fragments described herein, can be used to protect rabbits, rodents, birds, horses, cattle, and humans, including infant humans, from botulism, or from one or more of the symptoms of botulism, such as diarrheal disease, paralysis (either mild or severe), or death.

[0119] Toxin and toxin fragments can also be used to prepare compositions comprising neutralizing antibodies that immunoreact with the wild-type toxin. The resulting antisera can be used for the manufacture of a medicament for treating exposure to clostridial neurotoxin. Thus, antibody compositions, such as the isolated antisera or antibodies purified therefrom, can be used as a passive immune serum to prevent or treat botulism in patients exposed to the wild-type toxin. In such cases, the patient is a human, including an infant, suspected of having come in contact with the toxin, or is a human, including an infant, who has had known contact with the toxin, but is not yet showing symptoms of exposure. The antibody composition can also be used in a method of treating to ameliorate symptoms in patients that are suffering from the presence of toxin in their body. When the toxin is a clostridial neurotoxin, the symptoms include diarrhea and paralysis.

[0120] Methods of preparing passive immune sera are known in the art. For example, a vaccine composition can be administered to an animal such as a horse or a human until a neutralizing antibody response to wild type toxin is generated. Neutralizing antibodies can then be harvested, purified, and administered to patients exposed to, or exhibiting symptoms of contact with, the toxin to thereby treat or prevent botulism. In some cases, the antibodies are not purified after harvesting. When the neutralizing antibodies are from humans, the antibody preparation will generally be free of viruses, such as HIV and hepatitis. Methods of preparing human antisera are known in the art, and include the methods used to prepare IVIg. The neutralizing antibodies can be administered intravenously, intramuscularly, intradermally, or

subcutaneously. Antibiotic therapy can be used in conjunction. Dosages for neutralizing antibodies generally vary from about 1 mg to 1000 mg/kg. Often, they are administered at a dosage of about 50-200 mg/kg of body weight.

[0121] The invention will be further clarified by the following examples, which are intended to be purely exemplary of the invention and in no way limiting.

[0122] **Examples**

[0123] **Example 1. Site-Directed Mutagenesis of $LH_N$/E to Remove One or More Cysteine Residues**

[0124] Endopeptidase-ablating mutations (E213Q and H216Y relative to SEQ ID NO: 1 and SEQ ID NO: 2) were introduced into the $LH_N$/E coding sequence and the resulting cassettes cloned into plasmid vector pET26b. Various *E. coli* host strains were transformed and assessed for the ability to direct expression of a r$LH_N$/E fragment. While high levels of target protein could be produced, recombinant $LH_N$/E was expressed in all hosts as high molecular mass aggregate (Figs. 1A and 1B). Those SDS-PAGE and gel filtration studies conducted under reducing and non-reducing conditions showed that $LH_N$/E aggregation results, at least in part, from intermolecular disulfide bond formation.

[0125] It was hypothesized that aggregation could be due to the formation of cysteine-linked disulfide bonds between multiple $LH_N$/E polypeptides. Molecular biology approaches were pursued to increase expression of soluble, non-aggregated r$LH_N$/E protein. Using computational analyses of the BoNT/E catalytic domain crystal structure two surface-proximal cysteine residues (Cys26 and Cys347 in SEQ ID NO: 1 and SEQ ID

NO: 2) have been identified that most likely participate in intermolecular disulfide-bond-mediated bridging. Those residues were targeted for mutagenesis, and the aggregation properties of the mutated proteins were assessed.

[0126] $LH_N/E$ was cloned into the expression vector pET26b (Novagen) and this plasmid clone was used for the mutagenesis procedure. Specifically, site directed mutagenesis (QuikChange II XL site-directed mutagenesis kit, Stratagene) was used to introduce two point mutations into the $LH_N/E$ gene that change codons 26 and 347 (relative to a nucleic acid sequence encoding SEQ ID NO: 1 or SEQ ID NO: 2) from cysteine to structurally similar serine. While it is preferable to substitute the cysteine residue with a structurally similar amino acid, any amino acid may be substituted as long as that amino acid is incapable of forming an intermolecular disulfide bond.

[0127] The primer names and sequences used for mutagenesis are shown, with the nucleotides responsible for the cysteine to serine changes underlined. For both mutations, a TGC codon (cysteine) was changed to an AGC codon (serine). These two primers were used for the C26S mutation: C26S-LhnEfor: 5`-GTATATTAAACCGGGCGGCAGCCAGGAGTTTTATAAA AGC-3` (SEQ ID NO: 3) and C26S-LhnErev: 5`-GCTTTTATAAAACTCCTGG CTGCCGCCCGGTTTAATATAC-3` (SEQ ID NO: 4). These two primers were used for the C347S mutation: C347S-LhnEfor: 5`-GTACCAAATTTCAGGTGA AGAGCCGCCAAACCTACATCG-3` (SEQ ID NO: 5) and C347S-LhnErev: 5`-CGATGTAGGTTTGGCGGCTCTTCACCTGAAATTTGGTAC-3` (SEQ ID NO: 6).

[0128] Clones of pET26b/$LH_N$/E containing each single point mutation and both point mutations were made. The C26S single mutant, C347S single mutant, and C26S/C347S double mutant clones were expressed and analyzed for disulfide bond-mediated aggregation. As shown in Figs 2A and 2B, analysis of reduced and non-reduced samples revealed that intermolecular disulfide bonding was abolished in each of the $LH_N$/E clones (C26S single mutant, C347S single mutant, and C26S, C347S double mutant), whereas intermolecular disulfide bond formation was observable for the parental $LH_N$/E clone lacking the serine substitutions.

[0129] **Example 2. Extending the $LH_N$/E Fragment with $H_C$ Sequence Improves Solubility**

[0130] The solubility of clostridial neurotoxin proteins can also be enhanced by creating proteins in which an $LH_N$ domain, or a fragment of an $LH_N$ domain, is expressed along with amino acid sequence from an Hc domain.

[0131] Recombinant truncated forms of the botulinum serotype E toxin, such as the $LH_N$ fragments, have proven difficult to produce (express) and purify due to low solubility. Even at low concentrations, insoluble forms are often expressed in a non-native multimeric and aggregated state which renders them poor immunogens and unable to elicit protective levels of toxin neutralizing antibodies. To address this issue and enable the production of soluble (and possibly monomeric) protein, recombinant derivatives of the $LH_N$/E protein have been produced that carry various lengths of the adjoining Hc domain.

[0132] The following expression constructs were tested for protein solubility in the *E. coli* strain ER2566: $LH_N$/E; $LH_N$/E-Hc103; $LH_N$/E-Hc202; $LH_N$/E-Hc304; and $LH_N$/E-Hc406. Fifteen milliliters of LB media containing kanamycin (30 μg/mL) were inoculated with each strain containing the construct listed above. Inoculations were made directly from frozen glycerol stocks. These cultures grew at 37°C overnight with shaking. The next morning, the overnight cultures were diluted into 1L of 2xYT containing kanamycin (30 μg/mL) (15 mL into 1L of 2xYT) in Fernbach flasks. The cultures shook at 37°C for three hours. $OD_{600}$ ranged from 0.7-0.9 for all of the cultures. The Fernbach flasks were moved into 20°C incubators, IPTG was added to 1 mM final concentration, and the flasks shook at 20°C for 4 hours.

[0133] The 1L cultures were collected by centrifugation and cell pellets were resuspended in 200 ml of 50 mM Tris, pH 8.0, 25 mM EDTA, pH 8.0. The cell suspensions were disrupted using the microfluidizer. Thirty-five milliliters of each cell lysate was centrifuged at 10,000xg for 30 minutes. The supernatants containing soluble protein were transferred into clean tubes and the insoluble pellets were resuspended in 35 mL of 50 mM Tris, pH 8.0, 25 mM EDTA, pH 8.0.

[0134] Equal volumes of total cell lysate, and soluble and insoluble fractions, were prepared for SDS-PAGE by the addition of sample buffer and boiling for 5 minutes. Equal amounts of each sample were subjected to SDS-PAGE. One set of gels was stained using MicroWave Blue reagent. Another set was transferred to PVDF membrane and subjected to western blotting

using antisera specific for BoNT/E obtained from the Health Protection Agency.

[0135] The $LH_N$/E solubility is enhanced with the addition of amino acid residues from the Hc domain. This can be seen by comparing the amount of $LH_N$/E-Hc in soluble versus insoluble fractions in the Coomassie stained gels of Figure 3 and the western blots of Figure 4. $LH_N$/E, which is devoid of any Hc sequence, is detected predominately in the insoluble fraction. This is also observed for $LH_N$/E-Hc103 and $LH_N$/E-Hc202. However, the solubility of recombinant proteins containing longer segments of the Hc domains is greatly enhanced. This can be seen for $LH_N$/E-Hc406, which fractionates predominately with the soluble fraction, and also with $LH_N$/E-Hc304, which also displays enhanced solubility. These conclusions were confirmed by densitometry scans of the Coomassie stained gel.

[0136] We also compared the effect of treating $LH_N$/E and $LH_N$/E-Hc406 with the reducing agent DTT. Fig. 5A shows a Coomassie stained gel and Fig. 5B a western blot of the DTT treated (+DTT) and untreated (-DTT) samples. In both the presence and absence of DTT, more $LH_N$/E-Hc406 was found in the soluble (S) than in the insoluble (I) fraction.

[0137] The addition of $H_C$ sequence to the $LH_N$/E fragment improves its solubility, as does the replacement of Cys 26 and Cys 347 with serine residues. We have also prepared expression constructs in which we combined these approaches. Following induction of protein expression, the pellet was harvested, then lysed and microfluidized. The lysate was separated into soluble and insoluble fractions by centrifugation. The levels of

protein in the total lysate, soluble fraction, and insoluble fraction compared by running on a gel and western blotting with the anti-BoNT/E antisera. As shown in Fig. 6, the $LH_N$/E-Hc406 (C26S, C347S) mutant, like the $LH_N$/E (C26S, C347S) mutant, partitions predominantly to the soluble fraction.

[0138] **Example 3. Immunogenicity of $LH_N$/E Cys to Ser fragments and $LH_N$/E-Hc fragments**

[0139] Abolishing the ability of $LH_N$/E to form aberrant intermolecular disulfide bonds by replacing cysteine residues with amino acids that do not form disulfide bonds and extending the $LH_N$/E fragment with $H_C$ sequence are two techniques that improve the yield of monomeric or less aggregated protein. These modifications will enhance the immunogenicity and protective efficacy of the $LH_N$/E fragment and the enzymatic activity of non-endopeptidase ablated toxins and toxin subfragments.

[0140] Immunogenicity of the recombinant proteins is tested in mice. Mice are immunized either with 10 μg of a $LH_N$/E protein in which one or more cysteine residues has been replaced with another amino acid that does not form disulfide bonds, with 10 μg of a $LH_N$/E protein that has been extended by inclusion of $H_C$ sequence, with 10 μg of inactivated BoNT/E, or with other proteins described in the Examples. The proteins are suspended in an adjuvant emulsion. Control mice are immunized with saline emulsified in adjuvant for use as negative controls. The mice are immunized i.p. four times at 2-week intervals. One week after the last immunization, the mice are bled and the serum is analyzed by immunoblot for the presence of specific antibody. ELISA is used to determine the titer of the antisera. Two weeks

after the last immunization, each mouse is challenged i.p. with 2 lethal doses of BoNT/E. Four days after challenge, the mice are scored for survivors.

[0141] **Example 4. Amino Acid Sequences Encoded by Certain Constructs**

[0142] The following amino acid sequences are encoded by constructs in which the type E neurotoxin is from *Clostridium botulinum*. In the sequences, the mutated cysteine residues, which have been substituted with serine residues, are indicated in bold and underline.

[0143] 1. $LH_N$/E (endopeptidase active):

PKINSFNYNDPVNDRTILYIKPGG**S**QEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP
TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG
NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHRFGSIAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHELIHSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKF
QVK**S**RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ
IENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQ
HGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK
(SEQ ID NO: 7).

[0144] 2. $LH_N$/E (endopeptidase attenuated):

PKINSFNYNDPVNDRTILYIKPGG**S**QEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP
TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG
NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHRFGSIAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKF
QVK**S**RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI

VSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ
IENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQ
HGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK
(SEQ ID NO: 8).

[0145] 3. BoNT/E neurotoxin (endopeptidase attenuated):

PKINSFNYNDPVNDRTILYIKPGGSQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP
TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG
NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHRFGSIAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKF
QVKSRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ
IENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQ
HGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMR
YKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYK
NFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTFEDNRGIN
QKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD
NILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNEPNTNILKDFWGNYLLYDK
EYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL
VRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTM
NFKNNNGNNIGLLGFKADTVVASTLFYTHMRDHTNSNGCFWNFISEEHGWQEK
(SEQ ID NO: 9).

[0146] 4. Extended $LH_N$/E neurotoxin (endopeptidase attenuated):

PKINSFNYNDPVNDRTILYIKPGGSQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP
TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG
NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHRFGSIAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKF
QVKSRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ
IENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQ
HGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMR

YKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYK
NFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTFEDNRGIN
QKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD
NILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE
(SEQ ID NO: 10).

[0147] The following amino acid sequences are encoded by constructs in which the type E neurotoxin is from *Clostridium butyricum*. In the sequences, the mutated cysteine residues, which have been substituted with serine residues, are indicated in bold and underline.

[0148] 5. $LH_N$/E (endopeptidase active):

PTINSFNYNDPVNNRTILYIKPGG**S**QQFYKSFNIMKNIWIIPERNVIGTIPQDFLPP
TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG
NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHGFGSIAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHELIHSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF
QVK**S**RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQ
IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIK
HGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFK
(SEQ ID NO: 11).

[0149] 6. $LH_N$/E (endopeptidase attenuated):

PTINSFNYNDPVNNRTILYIKPGG**S**QQFYKSFNIMKNIWIIPERNVIGTIPQDFLPP
TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG
NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHGFGSIAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF
QVK**S**RQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQ

IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIK
HGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFK
(SEQ ID NO: 12).

[0150] 7. BoNT/E neurotoxin (endopeptidase attenuated):

PTINSFNYNDPVNNRTILYIKPGGSQQFYKSFNIMKNIWIIPERNVIGTIPQDFLPP
TSLKNGDSSYYDPNYLQSDQEKDKFLKIVTKIFNRINDNLSGRILLEELSKANPYLG
NDNTPDGDFIINDASAVPIQFSNGSQSILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHGFGSIAIVTFSPEYSFRFKDNSMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLATKF
QVKSRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVGWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNALKAIIESKYNSYTLEEKNELTNKYDIEQ
IENELNQKVSIAMNNIDRFLTESSISYLMKLINEVKINKLREYDENVKTYLLDYIIK
HGSILGESQQELNSMVIDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMR
YKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYK
NFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNSGIN
QKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDKKSILNLGNIHVSD
NILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYNNEPNANILKDFWGNYLLYDK
EYYLLNVLKPNNFINRRTDSTLSINNIRSTILLANRLYSGIKVKIQRVNNSSTNDNL
VRKNDQVYINFVASKTHLLPLYADTATTNKEKTIKISSSGNRFNQVVVMNSVGNCTM
NFKNNNGNNIGLLGFKADTVVASTLFYTHMRDNTNSNGFFWNFISEEHGWQEK
(SEQ ID NO: 13).

[0151] 8. Extended $LH_N$/E (endopeptidase attenuated):

PKINSFNYNDPVNDRTILYIKPGGSQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPP
TSLKNGDSSYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLG
NDNTPDNQFHIGDASAVEIKFSNGSQDILLPNVIIMGAEPDLFETNSSNISLRNNYM
PSNHRFGSIAIVTFSPEYSFRFNDNCMNEFIQDPALTLMHQLIYSLHGLYGAKGITT
KYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKL
SKVQVSNPLLNPYKDVFEAKYGLDKDASGIYSVNINKFNDIFKKLYSFTEFDLRTKF
QVKSRQTYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIITPIT
GRGLVKKIIRFCKNIVSVKGIRKSICIEINNGELFFVASENSYNDDNINTPKEIDDT
VTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGTSDIEQHDV
NELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAAL
FVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELL
GAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFI
VSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQ
IENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQ
HGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKRIKSSSVLNMR
YKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDNKYK

NFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTFEDNRGIN
QKLAFNYGNANGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSD
NILFKIVNCSYTRYIGIRYFNIFDKELDETEIQTLYSNE
(SEQ ID NO: 14).

[0152] Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein, such as, for example, *C. botulinum* neurotoxins of type A, B, C, D, F, or G mutated or truncated according to the method of the invention that exhibit improved solubility. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

WHAT IS CLAIMED IS:

1. A recombinant protein comprising at least one point mutation that substitutes a cysteine residue with another amino acid residue, wherein said substitution improves the solubility of the recombinant protein.

2. The protein of claim 1, wherein the protein is a toxin or non-toxic derivative of a toxin.

3. The protein of claim 2, wherein the toxin or non-toxic derivative of a toxin is of bacterial origin.

4. The protein of claim 3, wherein the bacterial toxin or toxin derivative is from either *Clostridium botulinum* or *Clostridium butyricum*.

5. The protein of claim 4, wherein the toxin is a neurotoxin or neurotoxin derivative.

6. The protein of claim 5, wherein the neurotoxin is neurotoxin A, B, C, D, E, F, or G, or a non-toxic derivative thereof.

7. The protein of claim 6, wherein the neurotoxin or non-toxic derivative is a fragment of neurotoxin E.

8. The protein of claim 7, wherein the fragment is the $LH_N$/E fragment of neurotoxin E.

9. The protein of claim 7, wherein the neurotoxin E fragment comprises at least one cysteine to serine amino acid substitution.

10. The protein of claim 9, wherein the substitution of serine for cysteine occurs at amino acid residue 26, amino acid residue 347, or both amino acid residue 26 and amino acid residue 347 compared to the $LH_N$ fragment of SEQ ID NO: 1 or SEQ ID NO: 2.

11. The protein of claim 6, wherein the protein has active endopeptidase activity.

12. The protein of claim 6, wherein the protein has attenuated endopeptidase activity.

13. A nucleic acid encoding a recombinant protein of any one of claims 1 to 12.

14. A method for improving the solubility of a protein having at least one cysteine residue that forms an intermolecular disulfide bond, comprising:

(a) providing a nucleic acid sequence encoding a recombinant protein comprising at least one cysteine residue;

(b) introducing at least one point mutation into the nucleic acid sequence that substitutes at least one cysteine residue with another amino acid residue;

(c) transforming a host cell with the mutated nucleic acid sequence; and

(d) expressing the nucleic acid sequence to produce the protein.

15. The method of claim 14, wherein the protein is a toxin or non-toxic derivative thereof.

16. The method of claim 15, wherein the toxin or non-toxic derivative is of bacterial origin.

17. The method of claim 16, wherein the bacterial toxin or toxin derivative is from either *Clostridium botulinum* or *Clostridium butyricum*.

18. The method of claim 17, wherein the toxin is a neurotoxin or neurotoxin derivative.

19. The method of claim 18, wherein the neurotoxin is neurotoxin A, B, C, D, E, F, or G, or a non-toxic derivative thereof.

20. The method of claim 19, wherein the non-toxic derivative is a fragment of neurotoxin E.

21. The method of claim 20, wherein the fragment is the $LH_N$/E fragment of neurotoxin E.

22. The method of claim 14, wherein the amino acid introduced by the at least one point mutation is a serine.

23. The method of claim 14, wherein the protein is a $LH_N$ fragment of clostridial neurotoxin E and the at least one point mutation substitutes a serine for a cysteine at amino acid residue 26, amino acid residue 347,

or both amino acid residue 26 and amino acid residue 347 compared to the $LH_N$ fragment of SEQ ID NO: 1 or SEQ ID NO: 2.

24. The method of claim 14, wherein the point mutation is introduced by site-directed mutagenesis.

25. The method of claim 14, further comprising isolating the protein.

26. The method of claim 14, wherein the host cell is a mammalian, plant, insect, fungal, or bacterial cell.

27. The method of claim 19, wherein the protein has active endopeptidase activity.

28. The method of claim 19, wherein the protein has attenuated endopeptidase activity.

29. Use of a protein of any one of claims 1 to 10 or claim 12 for the manufacture of a medicament for the treatment or prevention of botulism.

30. A composition comprising a protein of any one of claims 1 to 10 or claim 12 and a pharmaceutically acceptable carrier.

31. A method of protecting an individual from botulism, comprising administering to the individual a composition of claim 30.

32. A method of producing antibodies that neutralize a clostridial neurotoxin, comprising administering the composition of claim 30 to an

animal, allowing the animal to develop neutralizing antibodies to the clostridial neurotoxin, and isolating an antiserum that neutralizes the clostridial neurotoxin from the animal.

33. An antiserum produced by the method of claim 32.

34. A method of treating exposure to a clostridial neurotoxin, comprising administering to a patient that has been exposed to the clostridial neurotoxin the antiserum of claim 33.

35. A recombinant protein comprising a truncated botulinum serotype E toxin, wherein the truncation improves the solubility of the recombinant protein.

36. The protein of claim 35, wherein the truncation is in the Hc domain.

37. The protein of claim 35, wherein the truncated protein comprises the $LH_N$/E domain and the amino terminal 103 amino acids of the Hc domain.

38. The protein of claim 35, wherein the truncated protein comprises the amino terminal 948 amino acids of the serotype E toxin.

39. The protein of claim 35, wherein the truncated protein comprises the $LH_N$/E domain and the amino terminal 202 amino acids of the Hc domain.

40. The protein of claim 35, wherein the truncated protein comprises the amino terminal 1047 amino acids of the serotype E toxin.

41. The protein of claim 35, wherein the truncated protein comprises the $LH_N$/E domain and the amino terminal 304 amino acids of the Hc domain.

42. The protein of claim 35, wherein the truncated protein comprises the amino terminal 1149 amino acids of the serotype E toxin.

43. A nucleic acid encoding a recombinant protein of any one of claims 35 to 42.

44. A method for improving the solubility of a clostridial neurotoxin, comprising:

(a) providing a nucleic acid sequence encoding a clostridial neurotoxin;

(b) modifying the nucleic acid sequence so that it encodes the $LH_N$ fragment and a portion of the $H_C$ fragment of the neurotoxin;

(c) transforming the modified nucleic acid sequence into a host cell capable of expressing the modified nucleic acid sequence; and

(d) expressing the modified nucleic acid sequence to produce the protein.

45. Use of a protein of any one of claims 35 to 42 for the manufacture of a medicament for the treatment or prevention of botulism.

46. A composition comprising a protein of any one of claims 35 to 42 and a pharmaceutically acceptable carrier.

47. A method of protecting an individual from botulism, comprising administering to the individual a composition of claim 46.

48. A method of producing antibodies that neutralize a clostridial neurotoxin, comprising administering the composition of claim 46 to an animal, allowing the animal to develop neutralizing antibodies to the clostridial neurotoxin, and isolating an antiserum that neutralizes the clostridial neurotoxin from the animal.

49. An antiserum produced by the method of claim 48.

50. A method of treating exposure to a clostridial neurotoxin, comprising administering to a patient that has been exposed to the clostridial neurotoxin the antiserum of claim 49.

51. A mutated botulinum serotype E toxin comprising either or both of a leucine residue substituted for the tryptophan residue at position 1223 and a phenylalanine residue for the tyrosine residue at position 1224 of SEQ ID NO: 1 or SEQ ID NO: 2.

52. Use of the protein of claim 51 for the manufacture of a medicament for the treatment or prevention of botulism.

53. An *in vitro* method for improving the solubility of a protein having at least one cysteine residue that forms an intermolecular disulfide bond, comprising:

(a) providing a nucleic acid sequence encoding a recombinant protein comprising at least one cysteine residue;

(b) introducing at least one point mutation into the nucleic acid sequence that substitutes at least one cysteine residue with another amino acid residue;

(c) transforming a host cell with the mutated nucleic acid sequence; and

(d) expressing the nucleic acid sequence to produce the protein.

54. An *in vitro* method for improving the solubility of a clostridial neurotoxin, comprising:

(a) providing a nucleic acid sequence encoding a clostridial neurotoxin;

(b) modifying the nucleic acid sequence so that it encodes the $LH_N$ fragment and a portion of the $H_C$ fragment of the neurotoxin;

(c) transforming the modified nucleic acid sequence into a host cell capable of expressing the modified nucleic acid sequence; and

(d) expressing the modified nucleic acid sequence to produce the protein.

55. A method of producing antibodies that neutralize a clostridial neurotoxin, comprising isolating antibodies elicited by an inoculated

polypeptide, wherein said polypeptide is a protein according to any one of claims 6-10 or claim 12.

56. A method of producing antibodies that neutralize a clostridial neurotoxin, comprising isolating antibodies elicited by an inoculated polypeptide, wherein said polypeptide is a protein according to any one of claims 35-42.

57. Use of an antiserum according to claim 33 for the manufacture of a medicament for treating exposure to clostridial neurotoxin.

58. Use of an antiserum according to claim 49 for the manufacture of a medicament for treating exposure to clostridial neurotoxin.

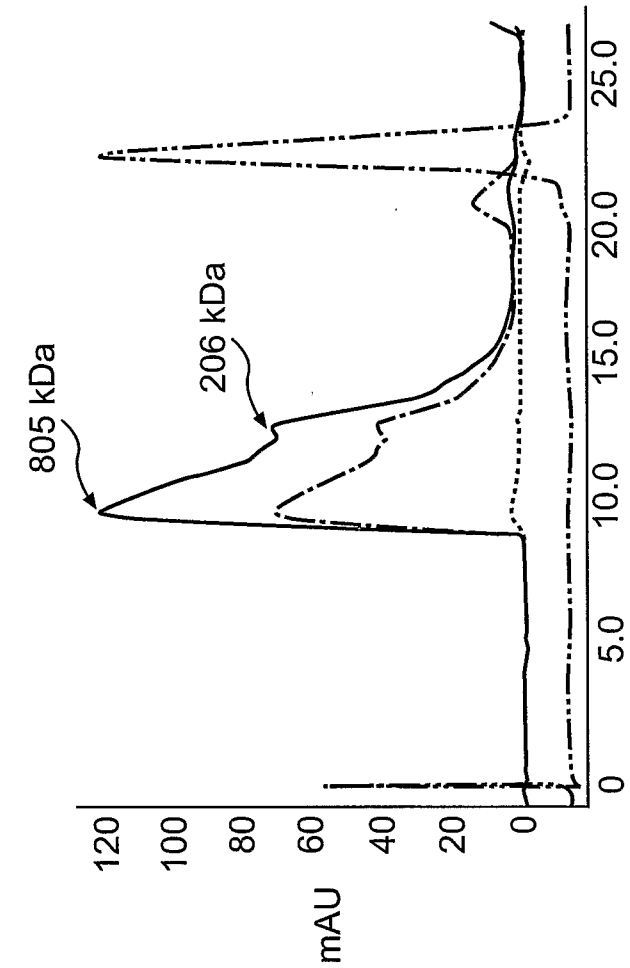


FIG. 1B

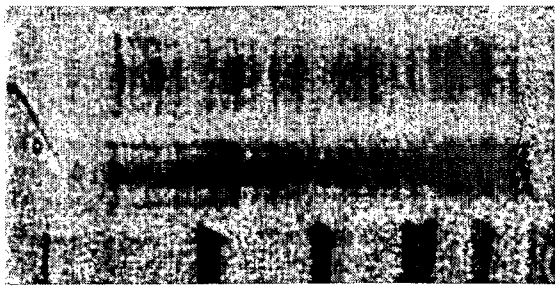

FIG. 1A

| LANE | SAMPLE | REDUCTANT |
|---|---|---|
| 1 | MW MARKER | NA |
| 2 | LHn-E C26S | + |
| 3 | LHn-E C26S | - |
| 4 | LHn-E C347S | + |
| 5 | LHn-E C347S | - |
| 6 | LHn-E C26S+C347S | + |
| 7 | LHn-E C26S+C347S | - |
| 8 | NATIVE LHn-E | + |
| 9 | NATIVE LHn-E | - |

FIG. 2A

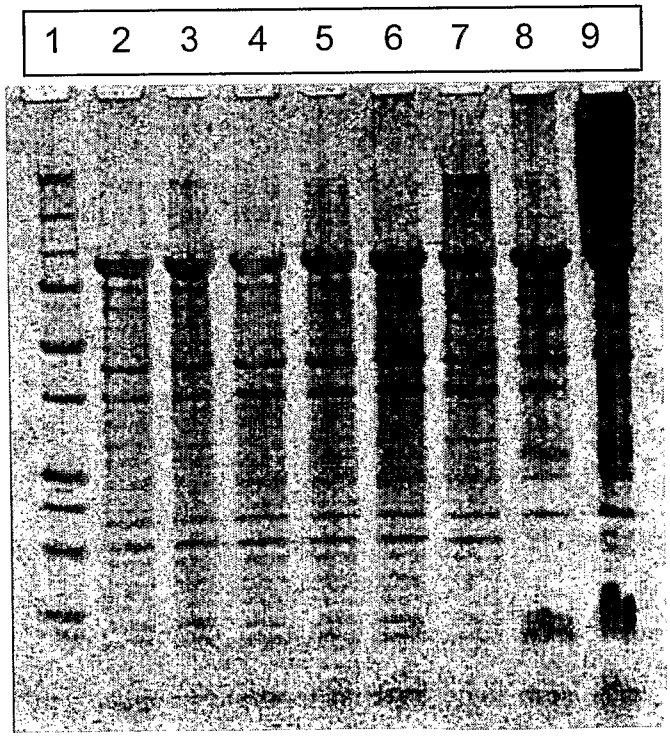


FIG. 2B

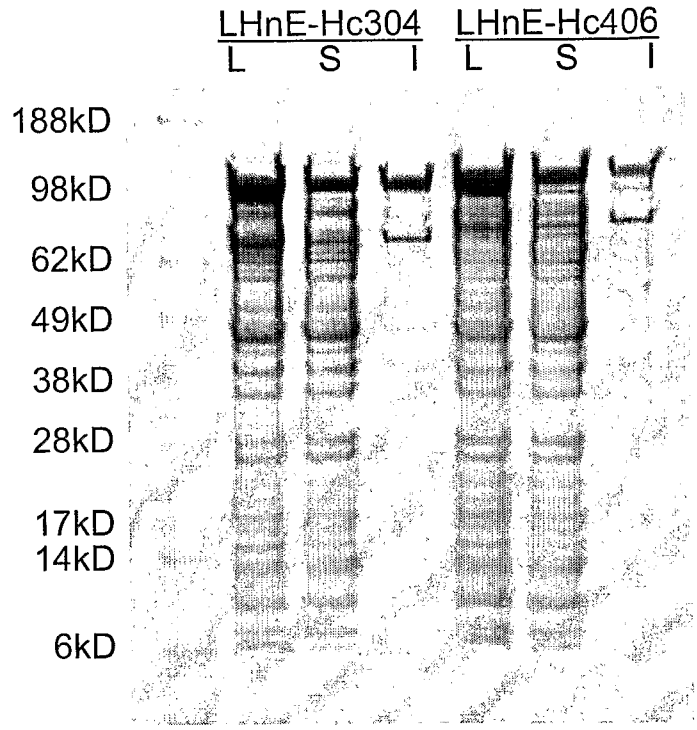
LHnE-Hc304
L S I
LHnE-Hc406
L S I
188kD
98kD
62kD
49kD
38kD
28kD
17kD
14kD
6kD

FIG. 3A

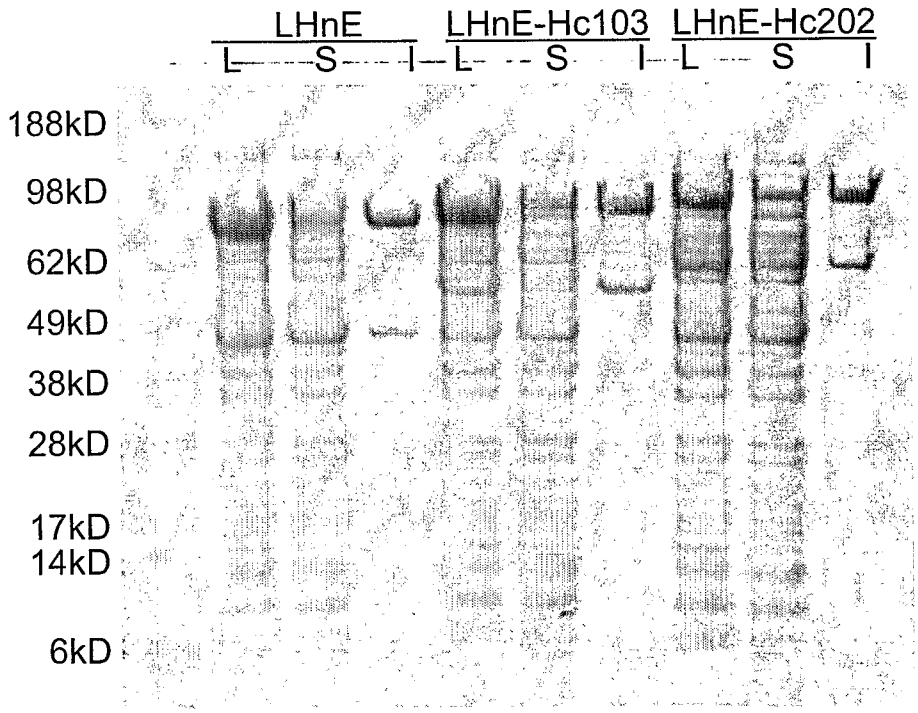
LHnE
L S I
LHnE-Hc103
L S I
LHnE-Hc202
L S I
188kD
98kD
62kD
49kD
38kD
28kD
17kD
14kD
6kD

FIG. 3B

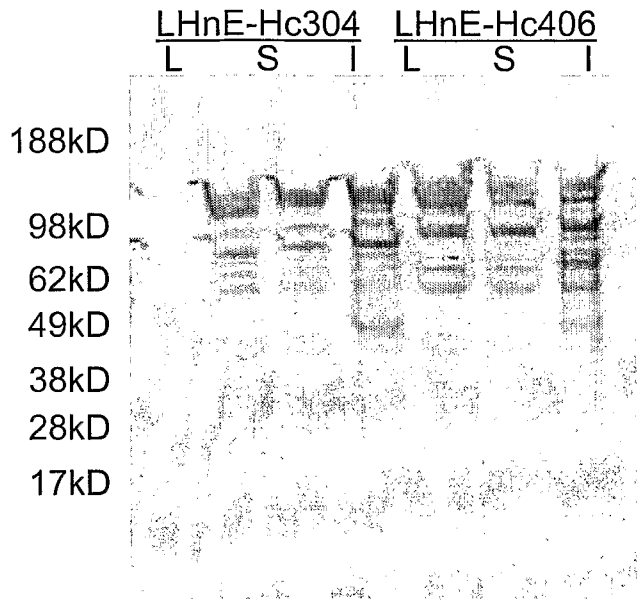
LHnE-Hc304
LHnE-Hc406
L S I L S I
188kD
98kD
62kD
49kD
38kD
28kD
17kD

FIG. 4A

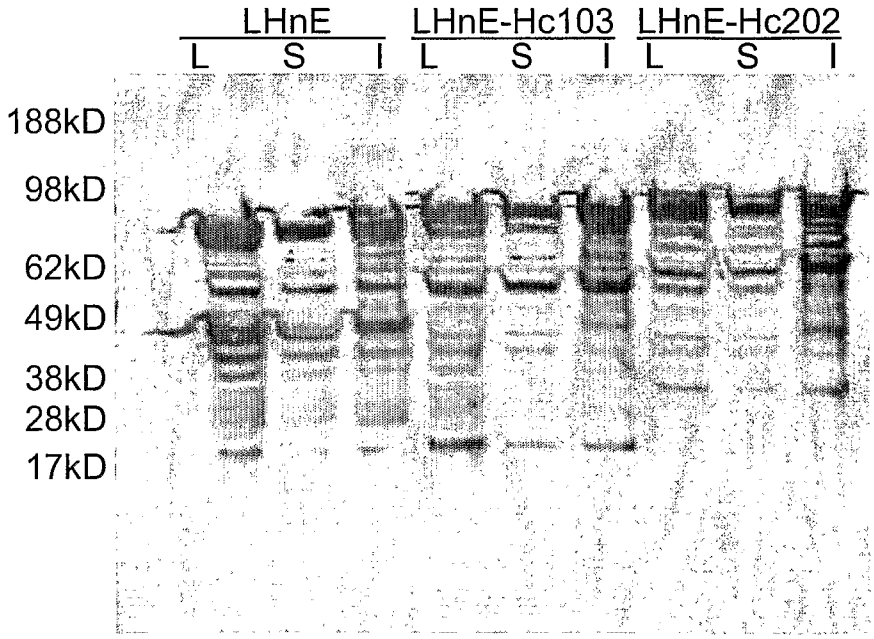
LHnE
LHnE-Hc103
LHnE-Hc202
L S I L S I L S I
188kD
98kD
62kD
49kD
38kD
28kD
17kD

FIG. 4B

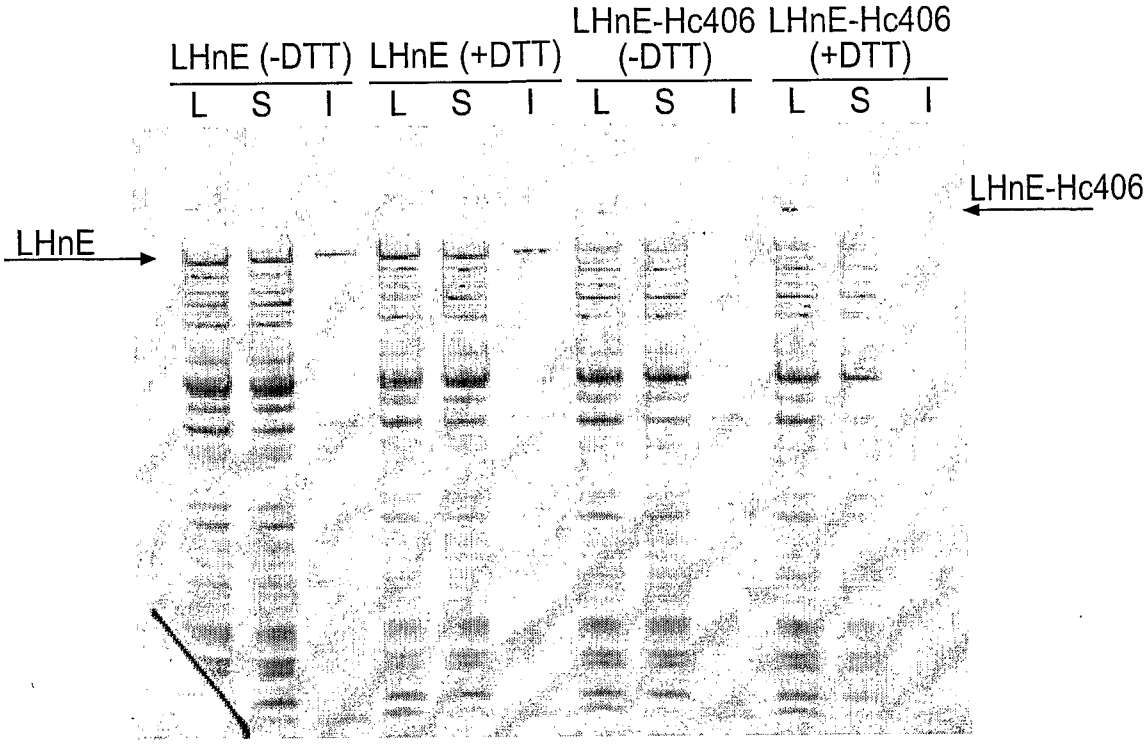
LHnE (-DTT)
LHnE (+DTT)
LHnE-Hc406 (-DTT)
LHnE-Hc406 (+DTT)
L S I L S I L S I L S I
LHnE-Hc406
LHnE

FIG. 5A

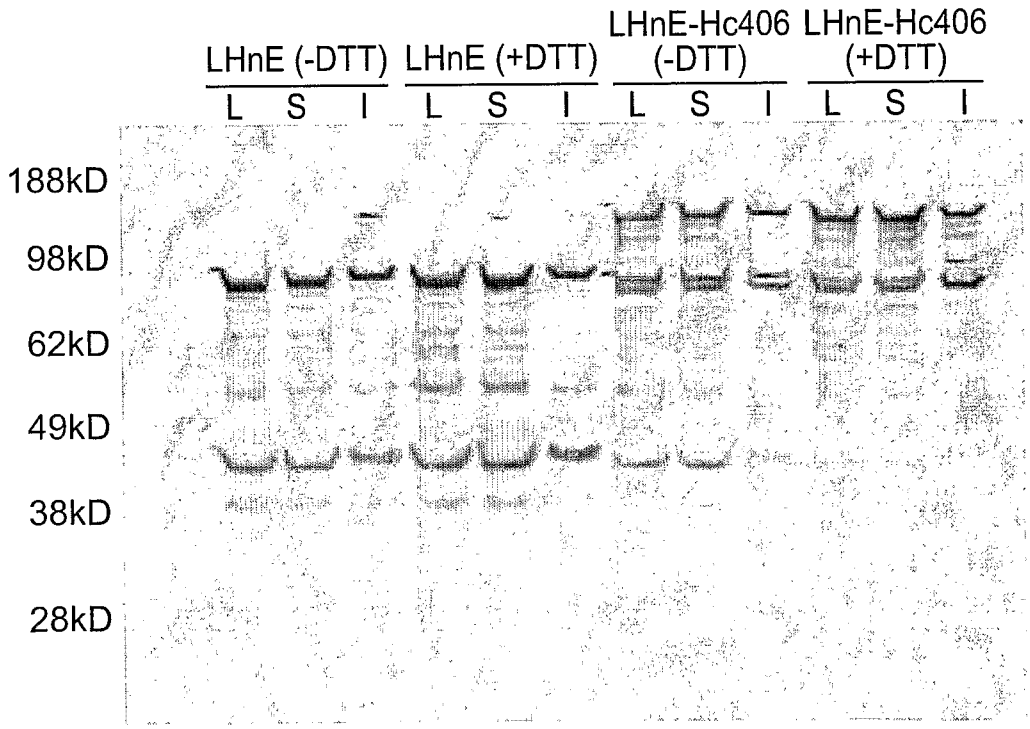
LHnE (-DTT)
LHnE (+DTT)
LHnE-Hc406 (-DTT)
LHnE-Hc406 (+DTT)
L S I L S I L S I L S I
188kD
98kD
62kD
49kD
38kD
28kD

FIG. 5B

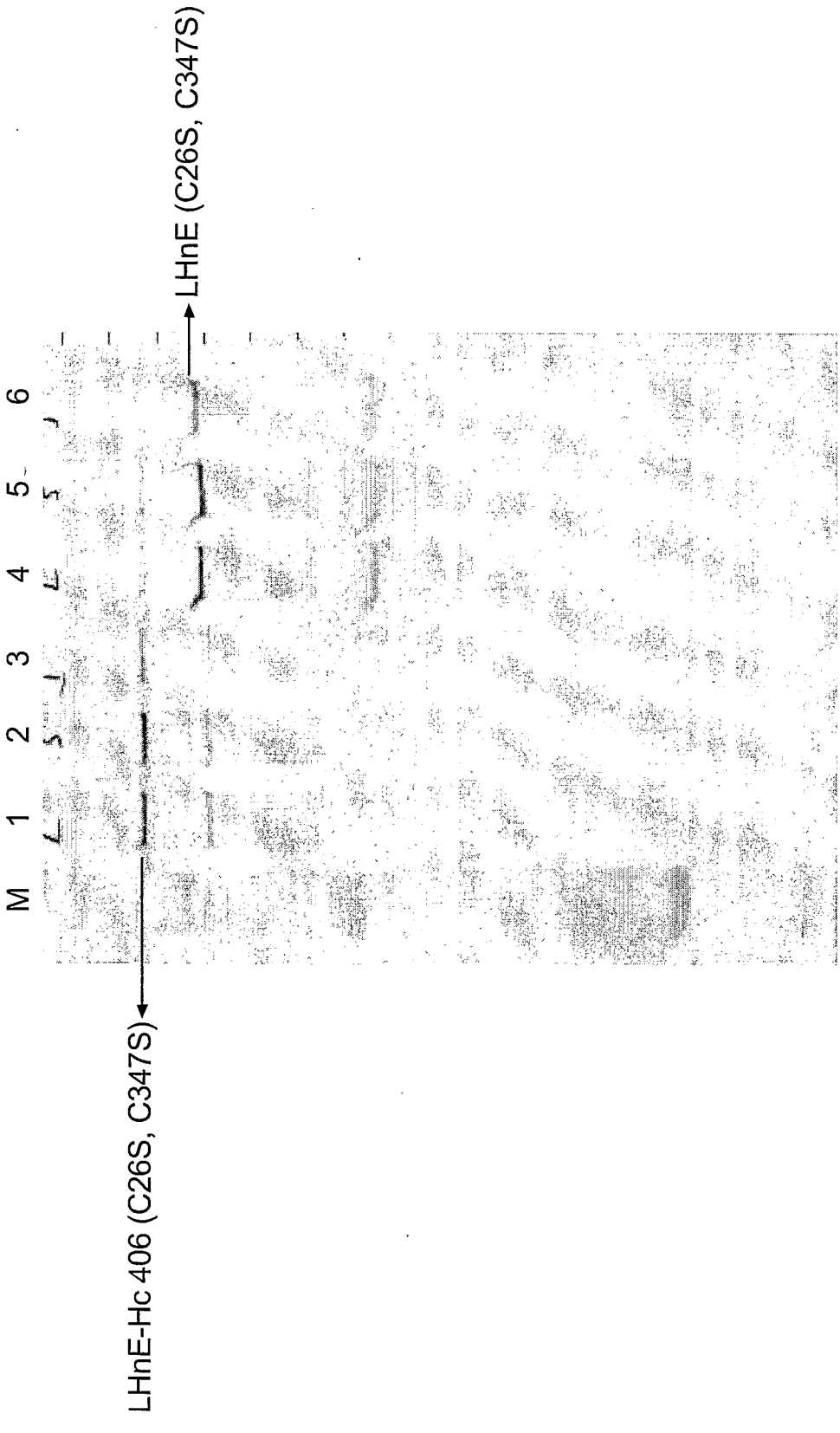


FIG. 6

SEQUENCE LISTING

<110> EMERGENT BIOSOLUTIONS, INC.

<120> PROTEINS WITH IMPROVED SOLUBILITY AND METHODS FOR PRODUCING AND USING SAME

<130> 09613.0016-00304

<140>
<141>

<150> 60/742,900
<151> 2005-10-07

<150> 60/724,274
<151> 2005-10-07

<160> 14

<170> PatentIn Ver. 3.3

<210> 1
<211> 1251
<212> PRT
<213> Clostridium botulinum

<400> 1

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

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

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

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

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

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

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

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

Ala Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu
130 135 140

Leu Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr
145 150 155 160

```
Asn Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His
                165                 170                 175

Arg Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe
            180                 185                 190

Arg Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu
        195                 200                 205

Thr Leu Met His Glu Leu Ile His Ser Leu His Gly Leu Tyr Gly Ala
    210                 215                 220

Lys Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu
225                 230                 235                 240

Ile Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly
                245                 250                 255

Gly Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr
            260                 265                 270

Thr Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys
        275                 280                 285

Val Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu
    290                 295                 300

Ala Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn
305                 310                 315                 320

Ile Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu
                325                 330                 335

Phe Asp Leu Arg Thr Lys Phe Gln Val Lys Cys Arg Gln Thr Tyr Ile
            340                 345                 350

Gly Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile
        355                 360                 365

Tyr Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe
    370                 375                 380

Arg Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr
385                 390                 395                 400

Gly Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val
                405                 410                 415

Ser Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly
            420                 425                 430

Glu Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile
        435                 440                 445

Asn Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr
    450                 455                 460
```

Glu Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala
465 470 475 480

Pro Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala
485 490 495

Tyr Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His
500 505 510

Asp Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val
515 520 525

Pro Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala
530 535 540

Leu Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile
545 550 555 560

Asn Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile
565 570 575

Gln Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr
580 585 590

Val Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu
595 600 605

Ala Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tyr Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser
885 890 895

Glu Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr
900 905 910

Lys Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn
915 920 925

Lys Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg
930 935 940

Asp Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile
945 950 955 960

Trp Thr Phe Glu Asp Asn Arg Gly Ile Asn Gln Lys Leu Ala Phe Asn
965 970 975

Tyr Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe
980 985 990

Val Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn
995 1000 1005

Gly Asn Leu Ile Asp Gln Lys Ser Ile Leu Asn Leu Gly Asn Ile His
1010 1015 1020

Val Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg
1025 1030 1035 1040

Tyr Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu
1045 1050 1055

Thr Glu Ile Gln Thr Leu Tyr Ser Asn Glu Pro Asn Thr Asn Ile Leu
1060 1065 1070

```
Lys Asp Phe Trp Gly Asn Tyr Leu Leu Tyr Asp Lys Glu Tyr Tyr Leu
       1075                1080                1085

Leu Asn Val Leu Lys Pro Asn Asn Phe Ile Asp Arg Arg Lys Asp Ser
   1090                1095                1100

Thr Leu Ser Ile Asn Asn Ile Arg Ser Thr Ile Leu Leu Ala Asn Arg
1105                1110                1115                1120

Leu Tyr Ser Gly Ile Lys Val Lys Ile Gln Arg Val Asn Asn Ser Ser
               1125                1130                1135

Thr Asn Asp Asn Leu Val Arg Lys Asn Asp Gln Val Tyr Ile Asn Phe
           1140                1145                1150

Val Ala Ser Lys Thr His Leu Phe Pro Leu Tyr Ala Asp Thr Ala Thr
       1155                1160                1165

Thr Asn Lys Glu Lys Thr Ile Lys Ile Ser Ser Ser Gly Asn Arg Phe
   1170                1175                1180

Asn Gln Val Val Val Met Asn Ser Val Gly Asn Cys Thr Met Asn Phe
1185                1190                1195                1200

Lys Asn Asn Asn Gly Asn Asn Ile Gly Leu Leu Gly Phe Lys Ala Asp
               1205                1210                1215

Thr Val Val Ala Ser Thr Trp Tyr Tyr Thr His Met Arg Asp His Thr
           1220                1225                1230

Asn Ser Asn Gly Cys Phe Trp Asn Phe Ile Ser Glu Glu His Gly Trp
       1235                1240                1245

Gln Glu Lys
   1250
```

<210> 2
<211> 1251
<212> PRT
<213> Clostridium butyricum

<400> 2

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

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

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

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

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

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

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

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

Ala Ser Ala Val Pro Ile Gln Phe Ser Asn Gly Ser Gln Ser Ile Leu
    130                 135                 140

Leu Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr
145                 150                 155                 160

Asn Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His
                165                 170                 175

Gly Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe
            180                 185                 190

Arg Phe Lys Asp Asn Ser Met Asn Glu Phe Ile Gln Asp Pro Ala Leu
        195                 200                 205

Thr Leu Met His Glu Leu Ile His Ser Leu His Gly Leu Tyr Gly Ala
    210                 215                 220

Lys Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu
225                 230                 235                 240

Ile Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly
                245                 250                 255

Gly Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr
            260                 265                 270

Thr Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys
        275                 280                 285

Val Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu
    290                 295                 300

Ala Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn
305                 310                 315                 320

Ile Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu
                325                 330                 335

Phe Asp Leu Ala Thr Lys Phe Gln Val Lys Cys Arg Gln Thr Tyr Ile
            340                 345                 350

Gly Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile
        355                 360                 365

Tyr Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe
    370                 375                 380
```

Arg Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr
385 390 395 400

Gly Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val
405 410 415

Ser Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly
420 425 430

Glu Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile
435 440 445

Asn Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr
450 455 460

Glu Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala
465 470 475 480

Pro Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala
485 490 495

Tyr Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His
500 505 510

Asp Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val
515 520 525

Pro Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala
530 535 540

Leu Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile
545 550 555 560

Asn Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Gly Trp Ile
565 570 575

Gln Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr
580 585 590

Val Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu
595 600 605

Ala Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tyr Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser
                885                 890                 895

Glu Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr
            900                 905                 910

Lys Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn
        915                 920                 925

Lys Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg
    930                 935                 940

Asp Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile
945                 950                 955                 960

Trp Thr Leu Gln Asp Asn Ser Gly Ile Asn Gln Lys Leu Ala Phe Asn
                965                 970                 975

Tyr Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe
            980                 985                 990
```

Val Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn
995 1000 1005

Gly Asn Leu Ile Asp Lys Lys Ser Ile Leu Asn Leu Gly Asn Ile His
1010 1015 1020

Val Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg
1025 1030 1035 1040

Tyr Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu
1045 1050 1055

Thr Glu Ile Gln Thr Leu Tyr Asn Asn Glu Pro Asn Ala Asn Ile Leu
1060 1065 1070

Lys Asp Phe Trp Gly Asn Tyr Leu Leu Tyr Asp Lys Glu Tyr Tyr Leu
1075 1080 1085

Leu Asn Val Leu Lys Pro Asn Asn Phe Ile Asn Arg Arg Thr Asp Ser
1090 1095 1100

Thr Leu Ser Ile Asn Asn Ile Arg Ser Thr Ile Leu Leu Ala Asn Arg
1105 1110 1115 1120

Leu Tyr Ser Gly Ile Lys Val Lys Ile Gln Arg Val Asn Asn Ser Ser
1125 1130 1135

Thr Asn Asp Asn Leu Val Arg Lys Asn Asp Gln Val Tyr Ile Asn Phe
1140 1145 1150

Val Ala Ser Lys Thr His Leu Leu Pro Leu Tyr Ala Asp Thr Ala Thr
1155 1160 1165

Thr Asn Lys Glu Lys Thr Ile Lys Ile Ser Ser Ser Gly Asn Arg Phe
1170 1175 1180

Asn Gln Val Val Val Met Asn Ser Val Gly Asn Cys Thr Met Asn Phe
1185 1190 1195 1200

Lys Asn Asn Asn Gly Asn Asn Ile Gly Leu Leu Gly Phe Lys Ala Asp
1205 1210 1215

Thr Val Val Ala Ser Thr Trp Tyr Tyr Thr His Met Arg Asp Asn Thr
1220 1225 1230

Asn Ser Asn Gly Phe Phe Trp Asn Phe Ile Ser Glu Glu His Gly Trp
1235 1240 1245

Gln Glu Lys
1250

<210> 3
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 3
gtatattaaa ccgggcggca gccaggagtt ttataaaagc 40

<210> 4
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 4
gcttttataa aactcctggc tgccgcccgg tttaatatac 40

<210> 5
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 5
gtaccaaatt tcaggtgaag agccgccaaa cctacatcg 39

<210> 6
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 6
cgatgtaggt ttggcggctc ttcacctgaa atttggtac 39

<210> 7
<211> 844
<212> PRT
<213> Clostridium botulinum

<400> 7
Pro Lys Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asp Arg Thr
1 5 10 15

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

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

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

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

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

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

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

Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu Leu
    130                 135                 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145                 150                 155                 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Arg
                165                 170                 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Glu Leu Ile His Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320
```

```
Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Arg Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
            420                 425                 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
        435                 440                 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
    450                 455                 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465                 470                 475                 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
                485                 490                 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
            500                 505                 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
        515                 520                 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
    530                 535                 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545                 550                 555                 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile Gln
                565                 570                 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
            580                 585                 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
        595                 600                 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
    610                 615                 620
```

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

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

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

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

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

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

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

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

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

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

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

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

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

Lys Ile Leu Ile Ser Tyr Phe Asn Lys Phe Phe Lys
835 840

<210> 8
<211> 844
<212> PRT
<213> Clostridium botulinum

<400> 8
Pro Lys Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asp Arg Thr
1 5 10 15

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

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

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

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

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

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

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

Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu Leu
    130                 135                 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145                 150                 155                 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Arg
                165                 170                 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Arg Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350
```

```
Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
            420                 425                 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
        435                 440                 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
    450                 455                 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465                 470                 475                 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
                485                 490                 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
            500                 505                 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
        515                 520                 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
    530                 535                 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545                 550                 555                 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile Gln
                565                 570                 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
            580                 585                 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
        595                 600                 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
    610                 615                 620

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

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

```
Asp Asn Lys Asn Lys Val Ile Lys Ala Ile Asn Asn Ala Leu Lys Glu
            660                 665                 670

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

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

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

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

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

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

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

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

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

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

Lys Ile Leu Ile Ser Tyr Phe Asn Lys Phe Phe Lys
        835                 840
```

<210> 9
<211> 1250
<212> PRT
<213> Clostridium botulinum

<400> 9

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

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

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

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

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

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

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

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

Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu Leu
    130                 135                 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145                 150                 155                 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Arg
                165                 170                 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Arg Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380
```

```
Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
            420                 425                 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
        435                 440                 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
    450                 455                 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465                 470                 475                 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
                485                 490                 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
            500                 505                 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
        515                 520                 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
    530                 535                 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545                 550                 555                 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile Gln
                565                 570                 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
            580                 585                 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
        595                 600                 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
    610                 615                 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser Glu
885 890 895

Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr Lys
900 905 910

Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn Lys
915 920 925

Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg Asp
930 935 940

Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile Trp
945 950 955 960

Thr Phe Glu Asp Asn Arg Gly Ile Asn Gln Lys Leu Ala Phe Asn Tyr
965 970 975

Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe Val
980 985 990

```
Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn Gly
        995                1000                1005

Asn Leu Ile Asp Gln Lys Ser Ile Leu Asn Leu Gly Asn Ile His Val
   1010                1015                1020

Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg Tyr
1025                1030                1035                1040

Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu Thr
               1045                1050                1055

Glu Ile Gln Thr Leu Tyr Ser Asn Glu Pro Asn Thr Asn Ile Leu Lys
           1060                1065                1070

Asp Phe Trp Gly Asn Tyr Leu Leu Tyr Asp Lys Glu Tyr Tyr Leu Leu
       1075                1080                1085

Asn Val Leu Lys Pro Asn Asn Phe Ile Asp Arg Arg Lys Asp Ser Thr
   1090                1095                1100

Leu Ser Ile Asn Asn Ile Arg Ser Thr Ile Leu Leu Ala Asn Arg Leu
1105                1110                1115                1120

Tyr Ser Gly Ile Lys Val Lys Ile Gln Arg Val Asn Asn Ser Ser Thr
               1125                1130                1135

Asn Asp Asn Leu Val Arg Lys Asn Asp Gln Val Tyr Ile Asn Phe Val
           1140                1145                1150

Ala Ser Lys Thr His Leu Phe Pro Leu Tyr Ala Asp Thr Ala Thr Thr
       1155                1160                1165

Asn Lys Glu Lys Thr Ile Lys Ile Ser Ser Ser Gly Asn Arg Phe Asn
   1170                1175                1180

Gln Val Val Val Met Asn Ser Val Gly Asn Cys Thr Met Asn Phe Lys
1185                1190                1195                1200

Asn Asn Asn Gly Asn Asn Ile Gly Leu Leu Gly Phe Lys Ala Asp Thr
               1205                1210                1215

Val Val Ala Ser Thr Leu Phe Tyr Thr His Met Arg Asp His Thr Asn
           1220                1225                1230

Ser Asn Gly Cys Phe Trp Asn Phe Ile Ser Glu Glu His Gly Trp Gln
       1235                1240                1245

Glu Lys
   1250
```

<210> 10
<211> 1065
<212> PRT
<213> Clostridium botulinum

<400> 10

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

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

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

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

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

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

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

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

Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu Leu
    130                 135                 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145                 150                 155                 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Arg
                165                 170                 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300
```

```
Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Arg Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
            420                 425                 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
        435                 440                 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
    450                 455                 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465                 470                 475                 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
                485                 490                 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
            500                 505                 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
        515                 520                 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
    530                 535                 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545                 550                 555                 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile Gln
                565                 570                 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
            580                 585                 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
        595                 600                 605
```

```
Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
    610                 615                 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser Glu
                885                 890                 895

Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr Lys
            900                 905                 910
```

```
Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn Lys
        915                 920                 925

Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg Asp
    930                 935                 940

Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile Trp
945                 950                 955                 960

Thr Phe Glu Asp Asn Arg Gly Ile Asn Gln Lys Leu Ala Phe Asn Tyr
                965                 970                 975

Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe Val
            980                 985                 990

Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn Gly
        995                1000                1005

Asn Leu Ile Asp Gln Lys Ser Ile Leu Asn Leu Gly Asn Ile His Val
   1010                1015                1020

Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg Tyr
1025                1030                1035                1040

Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu Thr
                1045                1050                1055

Glu Ile Gln Thr Leu Tyr Ser Asn Glu
            1060                1065
```

<210> 11
<211> 844
<212> PRT
<213> Clostridium butyricum

<400> 11

```
Pro Thr Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asn Arg Thr
  1               5                  10                  15

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

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

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

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

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

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

```
Leu Gly Asn Asp Asn Thr Pro Asp Gly Asp Phe Ile Ile Asn Asp Ala
        115                 120                 125

Ser Ala Val Pro Ile Gln Phe Ser Asn Gly Ser Gln Ser Ile Leu Leu
    130                 135                 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145                 150                 155                 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Gly
                165                 170                 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Lys Asp Asn Ser Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Glu Leu Ile His Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Ala Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415
```

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
420 425 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
435 440 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
450 455 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465 470 475 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
485 490 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
500 505 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
515 520 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
530 535 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545 550 555 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Gly Trp Ile Gln
565 570 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
580 585 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
595 600 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

Lys Ile Leu Ile Ser Tyr Phe Asn Lys Phe Phe Lys
        835                 840
```

<210> 12
<211> 844
<212> PRT
<213> Clostridium butyricum

<400> 12

```
Pro Thr Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asn Arg Thr
  1               5                  10                  15

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

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

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

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

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

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

Leu Gly Asn Asp Asn Thr Pro Asp Gly Asp Phe Ile Ile Asn Asp Ala
        115                 120                 125

Ser Ala Val Pro Ile Gln Phe Ser Asn Gly Ser Gln Ser Ile Leu Leu
    130                 135                 140
```

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145 150 155 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Gly
165 170 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
180 185 190

Phe Lys Asp Asn Ser Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
195 200 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
210 215 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225 230 235 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
245 250 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
260 265 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
275 280 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
290 295 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305 310 315 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
325 330 335

Asp Leu Ala Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
340 345 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
355 360 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
370 375 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385 390 395 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
405 410 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
420 425 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
435 440 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
450 455 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465 470 475 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
485 490 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
500 505 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
515 520 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
530 535 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545 550 555 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Gly Trp Ile Gln
565 570 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
580 585 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
595 600 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

Lys Ile Leu Ile Ser Tyr Phe Asn Lys Phe Phe Lys
835 840

<210> 13
<211> 1250
<212> PRT
<213> Clostridium butyricum

<400> 13
Pro Thr Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asn Arg Thr
1 5 10 15

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

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

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

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

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

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

Leu Gly Asn Asp Asn Thr Pro Asp Gly Asp Phe Ile Ile Asn Asp Ala
115 120 125

Ser Ala Val Pro Ile Gln Phe Ser Asn Gly Ser Gln Ser Ile Leu Leu
130 135 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145 150 155 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Gly
165 170 175

```
Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
            180                 185                 190

Phe Lys Asp Asn Ser Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
        195                 200                 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
    210                 215                 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225                 230                 235                 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
                245                 250                 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
            260                 265                 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
        275                 280                 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
    290                 295                 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305                 310                 315                 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
                325                 330                 335

Asp Leu Ala Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
            340                 345                 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
        355                 360                 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
    370                 375                 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385                 390                 395                 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
                405                 410                 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
            420                 425                 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
        435                 440                 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
    450                 455                 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465                 470                 475                 480
```

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
485 490 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
500 505 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
515 520 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
530 535 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545 550 555 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Gly Trp Ile Gln
565 570 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
580 585 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
595 600 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser Glu
                885                 890                 895

Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr Lys
            900                 905                 910

Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn Lys
        915                 920                 925

Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg Asp
    930                 935                 940

Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile Trp
945                 950                 955                 960

Thr Leu Gln Asp Asn Ser Gly Ile Asn Gln Lys Leu Ala Phe Asn Tyr
                965                 970                 975

Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe Val
            980                 985                 990

Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn Gly
        995                1000                1005

Asn Leu Ile Asp Lys Lys Ser Ile Leu Asn Leu Gly Asn Ile His Val
   1010                1015                1020

Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg Tyr
1025                1030                1035                1040

Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu Thr
               1045                1050                1055

Glu Ile Gln Thr Leu Tyr Asn Asn Glu Pro Asn Ala Asn Ile Leu Lys
           1060                1065                1070

Asp Phe Trp Gly Asn Tyr Leu Leu Tyr Asp Lys Glu Tyr Tyr Leu Leu
       1075                1080                1085
```

Asn Val Leu Lys Pro Asn Asn Phe Ile Asn Arg Arg Thr Asp Ser Thr
1090 1095 1100

Leu Ser Ile Asn Asn Ile Arg Ser Thr Ile Leu Leu Ala Asn Arg Leu
1105 1110 1115 1120

Tyr Ser Gly Ile Lys Val Lys Ile Gln Arg Val Asn Asn Ser Ser Thr
1125 1130 1135

Asn Asp Asn Leu Val Arg Lys Asn Asp Gln Val Tyr Ile Asn Phe Val
1140 1145 1150

Ala Ser Lys Thr His Leu Leu Pro Leu Tyr Ala Asp Thr Ala Thr Thr
1155 1160 1165

Asn Lys Glu Lys Thr Ile Lys Ile Ser Ser Ser Gly Asn Arg Phe Asn
1170 1175 1180

Gln Val Val Val Met Asn Ser Val Gly Asn Cys Thr Met Asn Phe Lys
1185 1190 1195 1200

Asn Asn Asn Gly Asn Asn Ile Gly Leu Leu Gly Phe Lys Ala Asp Thr
1205 1210 1215

Val Val Ala Ser Thr Leu Phe Tyr Thr His Met Arg Asp Asn Thr Asn
1220 1225 1230

Ser Asn Gly Phe Phe Trp Asn Phe Ile Ser Glu Glu His Gly Trp Gln
1235 1240 1245

Glu Lys
1250

<210> 14
<211> 1065
<212> PRT
<213> Clostridium butyricum

<400> 14
Pro Lys Ile Asn Ser Phe Asn Tyr Asn Asp Pro Val Asn Asp Arg Thr
1 5 10 15

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

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

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

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

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

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

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

Ser Ala Val Glu Ile Lys Phe Ser Asn Gly Ser Gln Asp Ile Leu Leu
130 135 140

Pro Asn Val Ile Ile Met Gly Ala Glu Pro Asp Leu Phe Glu Thr Asn
145 150 155 160

Ser Ser Asn Ile Ser Leu Arg Asn Asn Tyr Met Pro Ser Asn His Arg
165 170 175

Phe Gly Ser Ile Ala Ile Val Thr Phe Ser Pro Glu Tyr Ser Phe Arg
180 185 190

Phe Asn Asp Asn Cys Met Asn Glu Phe Ile Gln Asp Pro Ala Leu Thr
195 200 205

Leu Met His Gln Leu Ile Tyr Ser Leu His Gly Leu Tyr Gly Ala Lys
210 215 220

Gly Ile Thr Thr Lys Tyr Thr Ile Thr Gln Lys Gln Asn Pro Leu Ile
225 230 235 240

Thr Asn Ile Arg Gly Thr Asn Ile Glu Glu Phe Leu Thr Phe Gly Gly
245 250 255

Thr Asp Leu Asn Ile Ile Thr Ser Ala Gln Ser Asn Asp Ile Tyr Thr
260 265 270

Asn Leu Leu Ala Asp Tyr Lys Lys Ile Ala Ser Lys Leu Ser Lys Val
275 280 285

Gln Val Ser Asn Pro Leu Leu Asn Pro Tyr Lys Asp Val Phe Glu Ala
290 295 300

Lys Tyr Gly Leu Asp Lys Asp Ala Ser Gly Ile Tyr Ser Val Asn Ile
305 310 315 320

Asn Lys Phe Asn Asp Ile Phe Lys Lys Leu Tyr Ser Phe Thr Glu Phe
325 330 335

Asp Leu Arg Thr Lys Phe Gln Val Lys Ser Arg Gln Thr Tyr Ile Gly
340 345 350

Gln Tyr Lys Tyr Phe Lys Leu Ser Asn Leu Leu Asn Asp Ser Ile Tyr
355 360 365

Asn Ile Ser Glu Gly Tyr Asn Ile Asn Asn Leu Lys Val Asn Phe Arg
370 375 380

Gly Gln Asn Ala Asn Leu Asn Pro Arg Ile Ile Thr Pro Ile Thr Gly
385 390 395 400

Arg Gly Leu Val Lys Lys Ile Ile Arg Phe Cys Lys Asn Ile Val Ser
405 410 415

Val Lys Gly Ile Arg Lys Ser Ile Cys Ile Glu Ile Asn Asn Gly Glu
420 425 430

Leu Phe Phe Val Ala Ser Glu Asn Ser Tyr Asn Asp Asp Asn Ile Asn
435 440 445

Thr Pro Lys Glu Ile Asp Asp Thr Val Thr Ser Asn Asn Asn Tyr Glu
450 455 460

Asn Asp Leu Asp Gln Val Ile Leu Asn Phe Asn Ser Glu Ser Ala Pro
465 470 475 480

Gly Leu Ser Asp Glu Lys Leu Asn Leu Thr Ile Gln Asn Asp Ala Tyr
485 490 495

Ile Pro Lys Tyr Asp Ser Asn Gly Thr Ser Asp Ile Glu Gln His Asp
500 505 510

Val Asn Glu Leu Asn Val Phe Phe Tyr Leu Asp Ala Gln Lys Val Pro
515 520 525

Glu Gly Glu Asn Asn Val Asn Leu Thr Ser Ser Ile Asp Thr Ala Leu
530 535 540

Leu Glu Gln Pro Lys Ile Tyr Thr Phe Phe Ser Ser Glu Phe Ile Asn
545 550 555 560

Asn Val Asn Lys Pro Val Gln Ala Ala Leu Phe Val Ser Trp Ile Gln
565 570 575

Gln Val Leu Val Asp Phe Thr Thr Glu Ala Asn Gln Lys Ser Thr Val
580 585 590

Asp Lys Ile Ala Asp Ile Ser Ile Val Val Pro Tyr Ile Gly Leu Ala
595 600 605

Leu Asn Ile Gly Asn Glu Ala Gln Lys Gly Asn Phe Lys Asp Ala Leu
610 615 620

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro Thr Asn Lys Asn Gln Phe Gly Ile Tyr Asn Asp Lys Leu Ser Glu
885 890 895

Val Asn Ile Ser Gln Asn Asp Tyr Ile Ile Tyr Asp Asn Lys Tyr Lys
900 905 910

Asn Phe Ser Ile Ser Phe Trp Val Arg Ile Pro Asn Tyr Asp Asn Lys
915 920 925

Ile Val Asn Val Asn Asn Glu Tyr Thr Ile Ile Asn Cys Met Arg Asp
930 935 940

Asn Asn Ser Gly Trp Lys Val Ser Leu Asn His Asn Glu Ile Ile Trp
945 950 955 960

Thr Phe Glu Asp Asn Arg Gly Ile Asn Gln Lys Leu Ala Phe Asn Tyr
965 970 975

Gly Asn Ala Asn Gly Ile Ser Asp Tyr Ile Asn Lys Trp Ile Phe Val
980 985 990

Thr Ile Thr Asn Asp Arg Leu Gly Asp Ser Lys Leu Tyr Ile Asn Gly
995 1000 1005

Asn Leu Ile Asp Gln Lys Ser Ile Leu Asn Leu Gly Asn Ile His Val
1010 1015 1020

Ser Asp Asn Ile Leu Phe Lys Ile Val Asn Cys Ser Tyr Thr Arg Tyr
1025 1030 1035 1040

Ile Gly Ile Arg Tyr Phe Asn Ile Phe Asp Lys Glu Leu Asp Glu Thr
1045 1050 1055

Glu Ile Gln Thr Leu Tyr Ser Asn Glu
1060 1065