(12) STANDARD PATENT (11) Application No. **AU 2009340404 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Modified non-cytotoxic proteases

(51) International Patent Classification(s)
A61K 38/48 (2006.01)

(21) Application No: **2009340404** (22) Date of Filing: **2009.12.16**

(87) WIPO No: **WO10/094905**

(30) Priority Data

(31) Number	(32) Date	(33) Country
0903006.5	**2009.02.23**	**GB**

(43) Publication Date: **2010.08.26**
(44) Accepted Journal Date: **2016.01.14**

(71) Applicant(s)
Ipsen Bioinnovation Limited

(72) Inventor(s)
Chaddock, John Andrew;Foster, Keith Alan

(74) Agent / Attorney
Freehills Patent Attorneys, Level 43 101 Collins Street, Melbourne, VIC, 3000

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 August 2010 (26.08.2010)

PCT

(10) International Publication Number
WO 2010/094905 A1

(51) International Patent Classification:
A61K 38/48 (2006.01)

(21) International Application Number:
PCT/GB2009/002892

(22) International Filing Date:
16 December 2009 (16.12.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0903006.5 23 February 2009 (23.02.2009) GB

(71) Applicant *(for all designated States except US)*: **SYNTAXIN LIMITED** [GB/GB]; Units 4-10, The Quadrant, Barton Lane, Abingdon, Oxfordshire OX14 3YS (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CHADDOCK, John, Andrew** [GB/GB]; Syntaxin Limited, Units 4-10, The Quadrant, Barton Lane, Abingdon Oxfordshire OX14 3YS (GB). **FOSTER, Keith, Alan** [GB/GB]; Syntaxin Limited, Units 4-10, The Quadrant, Barton Lane, Abingdon Oxfordshire OX14 3YS (GB).

(74) Agents: **MACLEAN, Martin, Robert** et al.; Mathys & Squire LLP, 120 Holborn, London EC1N 2SQ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

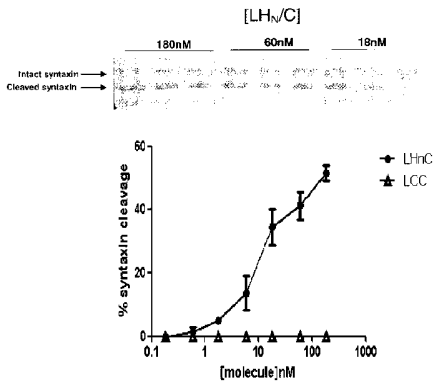
Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title**: MODIFIED NON-CYTOTOXIC PROTEASES

Figure 1



(57) **Abstract**: The present invention relates to a modified polypeptide comprising a non-cytotoxic protease, a translocation domain, a destructive protease cleavage site and a Targeting Moiety that binds to a Binding Site on a nerve cell, wherein after cleavage of the destructive cleavage site the polypeptide has reduced potency. The destructive cleavage site is recognised and cleaved by a protease present at or in an off-site target cell, and, in one embodiment, the polypeptide is a modified clostridial neurotoxin. The present invention also relates to the use of said polypeptides for treating a range of conditions, and to nucleic acids encoding said polypeptides.

WO 2010/094905 A1

Modified Non-Cytotoxic Proteases

The present invention relates to non-cytotoxic proteases having improved efficacy, and to the construction thereof.

Non-cytotoxic proteases are a well-recognised group of proteases, which act on target cells by incapacitating cellular function. Importantly, non-cytotoxic proteases do not kill the target cells upon which they act. Some of the best known examples of non-cytotoxic proteases include clostridial neurotoxins (e.g. botulinum neurotoxin; also known as BOTOX™) and IgA proteases.

Non-cytotoxic proteases act by proteolytically-cleaving intracellular transport proteins known as SNARE proteins (e.g. SNAP-25, VAMP, or Syntaxin) – see Gerald K (2002) "Cell and Molecular Biology" (4th edition) *John Wiley & Sons, Inc.* The acronym SNARE derives from the term **S**oluble **N**SF **A**ttachment **Re**ceptor, where NSF means **N**-ethylmaleimide-**S**ensitive **F**actor. SNARE proteins are integral to intracellular vesicle formation, and thus to secretion of molecules via vesicle transport from a cell. Accordingly, once delivered to a desired target cell, the non-cytotoxic protease is capable of inhibiting cellular secretion from the target cell.

Non-cytotoxic proteases may be employed in their native or substantially native forms (i.e. as holotoxins, such as BOTOX™), in which case targeting of the proteases to specific cell-types is reliant on (i) localised administration of the protease and/ or (ii) the inherent binding ability of the native protease. Alternatively, non-cytotoxic proteases may be employed in a re-targeted form in which the native protease is modified to include an exogenous ligand known as a Targeting Moiety (TM). The TM is selected to provide binding specificity for a desired target cell, and, as part of the re-targeting process, the native binding portion of the non-cytotoxic protease may be removed. Re-targeting technology is described, for example, in: EP-B-0689459; EP-B-0939818; US 6,461,617; US 7,192,596; EP-B-0826051; US 5,989,545; US 6,395,513; US

6,962,703; EP-B-0996468; US 7,052,702; EP-B-1107794; and US 6,632,440; all of which are herein incorporated by reference thereto.

In view of the ubiquitous nature of SNARE proteins, non-cytotoxic proteases have been successfully employed in a wide range of therapies.

By way of example, we refer to William J. Lipham, Cosmetic and Clinical Applications of Botulinum Toxin (Slack, Inc., 2004), which describes the use of clostridial toxins, such as botulinum neurotoxins (BoNTs), BoNT/A, BoNT/B, BoNT/C1, BoNT/D, BoNT/E, BoNT/F and BoNT/G, and tetanus neurotoxin (TeNT), to inhibit neuronal transmission in a wide variety of therapeutic and cosmetic applications - as an example, BOTOX™ is currently approved as a therapeutic for the following indications: achalasia, adult spasticity, anal fissure, back pain, blepharospasm, bruxism, cervical dystonia, essential tremor, glabellar lines or hyperkinetic facial lines, headache, hemifacial spasm, hyperactivity of bladder, hyperhidrosis, juvenile cerebral palsy, multiple sclerosis, myoclonic disorders, nasal labial lines, spasmodic dysphonia, strabismus and VII nerve disorder. In addition, clostridial toxin therapies are described for treating neuromuscular disorders (see US 6,872,397; for treating uterine disorders (see US2004/0175399); for treating ulcers and gastroesophageal reflux disease (see US2004/0086531); for treating dystonia (see US 6,319,505); for treating eye disorders (see US2004/0234532); for treating blepharospasm (see US2004/0151740); for treating strabismus (see US2004/0126396); for treating pain (see US 6,869,610, US 6,641,820, US 6,464,986, US 6,113,915); for treating fibromyalgia (see US 6,623,742, US2004/0062776); for treating lower back pain (see US2004/0037852); for treating muscle injuries (see US 6,423,319); for treating sinus headache (see US 6,838,434); for treating tension headache (see US 6,776,992); for treating headache (see US 6,458,365); for reduction of migraine headache pain (see US 5,714,469); for treating cardiovascular diseases (see US 6,767,544); for treating neurological disorders such as Parkinson's disease (see US 6,620,415, US 6,306,403); for treating neuropsychiatric disorders (see

US2004/0180061, US2003/0211121); for treating endocrine disorders (see US 6,827,931); for treating thyroid disorders (see US 6,740,321); for treating a cholinergic influenced sweat Gland (see US 6,683,049); for treating diabetes (see US 6,337,075, US 6,416,765); for treating a pancreatic disorder (see US 6,261,572, US 6,143,306); for treating cancers such as bone tumors (see US 6,565,870, US 6,368,605, US 6,139,845, US2005/0031648); for treating otic disorders (see US 6,358,926, US 6,265,379); for treating autonomic disorders such as gastrointestinal muscle disorders and other smooth muscle dysfunction (see US 5,437,291); for treatment of skin lesions associated with cutaneous cell-proliferative disorders (see US 5,670,484); for management of neurogenic inflammatory disorders (see US 6,063,768); for reducing hair loss and stimulating hair growth (see US 6,299,893); for treating downturned mouth (see US 6,358,917); for reducing appetite (see US2004/40253274); for dental therapies and procedures (see US2004/0115139); for treating neuromuscular disorders and conditions (see US2002/0010138); for treating various disorders and conditions and associated pain (see US2004/0013692) for treating pain (see WO96/33274); for treating conditions resulting from mucus hypersecretion such as asthma and COPD (see WO00/10598); for treating non-neuronal conditions such as inflammation, endocrine conditions, exocrine conditions, immunological conditions, cardiovascular conditions, bone conditions (see WO01/21213). All of the above publications are herein incorporated by reference thereto.

The use of non-cytotoxic proteases such as clostridial neurotoxins (e.g. BoNTs and TeNT) in therapeutic and cosmetic treatments of humans and other mammals is anticipated to expand to an ever-widening range of diseases and ailments that can benefit from the properties of these toxins.

Administration of a non-cytotoxic protease (including native clostridial neurotoxin clinical products) can be challenging because of a need for larger doses required to achieve a beneficial effect. Larger doses can increase the likelihood that the protease may move, for example, through the interstitial

fluids and the circulatory systems (such as the cardiovascular system and the lymphatic system) of the body, resulting in undesirable dispersal of the protease to areas not targeted for treatment. Said dispersal can lead to undesirable side effects, such as inhibition of cellular secretion in cells not targeted for treatment (e.g. inhibition of neurotransmitter release in neurons not targeted for treatment, or paralysis of a muscle not targeted for treatment). By way of specific example, a patient administered a therapeutically effective amount of a BoNT into the neck muscles for torticollis may develop dysphagia because of dispersal of the protease into the oropharynx. Similarly, a patient administered a non-cytotoxic protease to treat a neuromuscular disorder may suffer from undesirable muscle tissue inactivation due to dispersal of the protease into the muscle.

In common with any other drug substances, a therapeutic dosing range exists which identifies the lower and upper limits of safe, effective therapy. Often, the upper limit is determined by the increasing significance of off-target effects that lead to undesirable (e.g. potentially harmful) side-effects of drug treatment. In the case of non-cytotoxic proteases (notably BoNT), this could lead to the paralysis of cellular secretion in off-target cells, which, in turn, could be fatal.

The growing clinical, therapeutic and cosmetic use of non-cytotoxic proteases in therapies requiring larger doses places an ever-increasing requirement on the part of the pharmaceutical industry to develop means for minimising off-target effects, whilst maintaining the potency of the protease, such that the therapeutic dose range can be increased and the patients thus provided with increased doses which will, in turn, lead to increased efficacy of treatment.

There is therefore a need in the art for new therapies and/ or new therapeutics capable of specifically addressing undesirable, off-site targeting effects. This need is addressed by the present invention, which solves one or more of the above-mentioned problems.

In WO02/044199, Lin, Wei-Jen, *et al.*, seek to solve this problem by provision of clostridial neurotoxins modified to contain a blood protease cleavage site (ie. a site cleavable by a protease present in blood) in the binding domain of the neurotoxin, such that contact with a blood protease selectively inactivates the neurotoxin. Said binding domain (also referred to as the H_C domain) is illustrated in Figure 1B of Lin, Wei-Jen, *et al.* as the region starting at amino acid residue 873. The above-mentioned solution provided by Lin, *et al.*, however, has a number of problems, and does not adequately solve the problem of off-site targeting effects. In this regard, the present inventors have identified that clostridial neurotoxins in which the binding (H_C) domain has been removed (or otherwise inactivated) are still toxic and can still effect inhibition at their target neurons – this is confirmed by Figure 1 (see Example 39) of the present application, which illustrates SNARE protein cleavage by a clostridial neurotoxin molecule (LHN) lacking the binding (H_C). A further deficiency associated with WO02/044199 (Lin, Wei-Jen, *et al.)* is that the described technology is limited to clostridial neurotoxin molecules possessing a H_C binding domain (ie. clostridial holotxin molecules). As already discussed, however, non-cytotoxic proteases may be employed in a re-targeted form in which the native protease is modified to include an exogenous ligand known as a Targeting Moiety (TM), which provides binding specificity for a desired target cell. Thus, in the context of re-targeted non-cytotoxic proteases, the disclosure of Lin, *et al.* fails to address the problem of off-site targeting effects.

The present invention addresses the deficiencies of Lin, *et al.* and provides non-cytotoxic proteases that reduce or prevent unwanted side-effects associated with dispersal into non-targeted areas. These and related advantages are useful for various clinical, therapeutic and cosmetic applications, such as the treatment of neuromuscular disorders, neuropathic disorders, eye disorders, pain, muscle injuries, headache, cardiovascular diseases, neuropsychiatric disorders, endocrine disorders, exocrine disorders, mucus secretion-related disorders such as asthma and COPD, cancers, otic disorders and hyperkinetic facial lines, as well as, other disorders where non-

cytotoxic protease administration to a mammal can produce a beneficial effect (e.g. all of the therapies described on pages 2-3 of this specification).

In more detail, a first aspect of the present invention provides a polypeptide, comprising:

a. a non-cytotoxic protease that is capable of cleaving a SNARE protein;
b. a translocation domain that is capable of translocating the non-cytotoxic protease from within an endosome of a mammalian cell, across the endosomal membrane thereof and into the cytosol of the mammalian cell;
c. a first destructive cleavage site that is cleavable by a second protease and not by the non-cytotoxic protease, and wherein after cleavage thereof by the second protease the polypeptide has reduced potency measurable by a reduced ability to cleave said SNARE protein and/ or a reduced ability to translocate said non-cytotoxic protease across an endosomal membrane;
d. a Targeting Moiety (TM) that binds to a Binding Site present on a mammalian neuronal cell, which Binding Site is capable of undergoing endocytosis to be incorporated into an endosome within the mammalian neuronal cell; and
e. with the proviso that said first destructive cleavage site is not located within said TM.

Thus, the present invention provides a polypeptide that can be controllably inactivated and/ or destroyed at an off-site location.

In one embodiment, the destructive cleavage site is recognised and cleaved by a second protease (i.e. a destructive protease) selected from a circulating protease (e.g. an extracellular protease, such as a serum protease or a protease of the

blood clotting cascade), a tissue-associated protease (e.g. a matrix metalloprotease (MMP), such as a MMP of muscle), and an intracellular protease (preferably a protease that is absent from the target cell)).

Thus, in use, should a polypeptide of the present invention become dispersed away from its intended target cell and/ or be taken up by a non-target cell, the polypeptide will become inactivated by cleavage of the destructive cleavage site (by the second protease).

In one embodiment, the destructive cleavage site is recognised and cleaved by a second protease that is present within an off-site cell-type. In this embodiment, the off-site cell and the target cell are preferably different cell types. Alternatively (or in addition), the destructive cleavage site is recognised and cleaved by a second protease that is present at an off-site location (e.g. distal to the target cell). Accordingly, when destructive cleavage occurs extracellularly, the target cell and the off-site cell may be either the same or different cell-types. In this regard, the target cell and the off-site cell may each possess a receptor to which the same polypeptide of the invention binds).

By way of example, when treating neuromuscular disorders, a polypeptide of the present invention is targeted to the desired target cell (e.g. to a motor neuron), and includes a destructive protease cleavage site that is cleaved by a second protease present within and/ or in close proximity to muscle tissue. Accordingly, the polypeptide demonstrates minimal adverse effects on muscle tissue, and can be used at greater doses than currently tolerable by a patient, thereby leading to enhanced clinical efficacy.

The destructive cleavage site of the present invention provides for inactivation/ destruction of the polypeptide when the polypeptide is in or at an off-site location. In this regard, cleavage at the destructive cleavage site minimises the potency of the polypeptide by reducing the inherent ability of the polypeptide (when compared with an identical polypeptide lacking the same destructive cleavage

site, or possessing the same destructive site but in an uncleaved form) to translocate the non-cytotoxic component (across the endosomal membrane of a mammalian cell in the direction of the cytosol), and/ or to effect SNARE protein cleavage.

In one embodiment, the polypeptide of the invention may include a second (or subsequent) inactivation/ destruction site. Said (or subsequent) second site may be located anywhere within the polypeptide (including within the TM component). Said second (or subsequent) site may be cleaved by the same or by a different protease. Said second (or subsequent) site may have a different amino acid recognition sequence that the first inactivation/ destruction site, and may be cleaved by the same or by a different protease.

The above-mentioned reduced SNARE cleavage and/ or reduced translocation capacity can be readily measured by direct comparison of a polypeptide of the invention with an identical polypeptide (though lacking the same destructive cleavage site, or possessing the same destructive site but in an uncleaved form). In more detail, the polypeptide of the invention and the corresponding uncleaved counterpart may be assayed in parallel in any one of a variety of conventional whole cell or cell free assays. By way of example, reference is made to Examples 1-4. During said assays, the polypeptide of the invention becomes inactivated (via cleavage at the destructive cleavage site), whereas the counterpart polypeptide substantially retains full potency. Thus, in the context of the present invention, when cleaved at the destructive cleavage site, a polypeptide of the invention possesses less than 50% or less than 25%, less than 10% or less than 5%, less than 1% or less than 0.5%, less than 0.1% or less than 0.01%, or less than 0.001% or less than 0.0001% of the SNARE protein cleavage ability and/ or reduced translocation ability when compared with the uncleaved counterpart polypeptide.

In the context of whole cell assays, reduced SNARE cleavage and/ or reduced translocation ability may be determined by measuring relative SNARE protein

cleavage in a mammalian cell. This is reflective of the overall ability of the polypeptide to translocate into and subsequently cleave a SNARE protein within the cytosol of a mammalian cell. There are a variety of ways for measuring SNARE protein cleavage such as, for example, SDS-PAGE and Western Blotting followed by densitometer analysis of the cleaved products.

In the context of cell-free assays, potency can be measured in terms of relative SNARE protein cleavage, or in terms of relative translocation function (e.g. release of K^+ or NAD from liposomes, or membrane conductance measurements).

Preferred off-site targets (and thus preferred mammalian cells for the above assays) include: epithelial cells, especially lung epithelial cells; neuronal cells that are not motor neuron cells; and muscle cells.

Referring to Example 39, a modified clostridial neurotoxin (LH_N/C) was provided. This neurotoxin mimics the modified neurotoxin of Lin, *et al.* (ie. as discussed in the background part of this specification) as it lacks a functional H_C binding domain. Said modified neurotoxin was incubated in the presence of a mammalian cell (e.g. an embryonic spinal cord neuron (eSCN)) to assess it's ability to demonstrate residual clostridial neurotoxin activity in the form of SNARE protein cleavage. In parallel, a control neurotoxin consisting solely of the endopeptidase domain of botulinum neurotoxin type C (LC/C) was incubated in the same manner – the control neurotoxin therefore lacked a function H_N translocation domain. Each of the two polypeptides was then assessed for cleavage of a SNARE protein in the test cell. Surprisingly, the LH_N/C modified clostridial neurotoxin demonstrated SNARE cleavage (see Figure 1), and thus confirmed that inactivation of the H_C binding domain of botulinum neurotoxin is not adequate to reduce off-site activity. In contrast, the control neurotoxin (lacking a functional translocation component) demonstrated a lack of SNARE cleavage.

As mentioned above, the polypeptide of the present invention may include one or more (e.g. two, three, four, five or more) destructive protease cleavage sites. Where more than one destructive cleavage site is included, each cleavage site may be the same or different. In this regard, use of more than one destructive cleavage sites provides improved off-site inactivation. Similarly, use of two or more different destructive cleavage sites provides additional design flexibility. For example, when minimising off-site target effects in muscle tissue, the polypeptide of the present invention may include two different destructive sites, which are recognised and cleaved by two different muscle tissues associated proteases.

The first destructive cleavage site(s) may be engineered into the non-cytotoxic protease component or the translocation component. The second (or subsequent) site(s) may be engineered anywhere into the polypeptide. In this regard, the destructive cleavage site(s) are chosen to ensure minimal adverse effect on the potency of the polypeptide (for example by having minimal effect on the translocation domain, and/ or on the non-cytotoxic protease domain) whilst ensuring that the polypeptide is labile away from its target site/ target cell.

Preferred destructive cleavage sites (plus the corresponding second proteases) are listed in the Table immediately below. The listed cleavage sites are purely illustrative and are not intended to be limiting to the present invention.

Second protease	**Destructive cleavage site recognition sequence**	**Tolerated recognition sequence variance P4-P3-P2-P1-▼-P1'-P2'-P3'**						
		P4	**P3**	**P2**	**P1**	**P1'**	**P2'**	**P3'**
Thrombin	LVPR▼GS	A,F,G,I,L ,T,V or M	A,F,G,I, L,T,V, W or A	P	R	Not D or E	Not D or E	---
Thrombin	GR▼G			G	R	G		
Factor Xa	IEGR▼	A,F,G,I,L ,T,V or M	D or E	G	R	---	---	---

ADAM17	PLAQA▼VRSS S							
Human airway trypsin-like protease (HAT)	SKGR▼SLIGR V							
ACE (peptidyl-dipeptidase A)		---	---	---	---	Not P	Not D or E	N/A
Elastase (leukocyte)	MEA▼VTY	M, R	E	A, H	V, T	V, T, H	Y	---
Furin	RXR/KR▼	R	X	R or K	R			
Granzyme	IEPD▼	I	E	P	D	---	---	---
Caspase 1		F,W,Y,L	---	H, A,T	D	Not P,E. D.Q. K or R	---	---
Caspase 2	DVAD▼	D	V	A	D	Not P,E. D.Q. K or R	---	---
Caspase 3	DMQD▼	D	M	Q	D	Not P,E. D.Q. K or R	---	---
Caspase 4	LEVD▼	L	E	V	D	Not P,E. D.Q. K or R	---	---
Caspase 5		L or W	E	H	D	---	---	---
Caspase 6		V	E	H or I	D	Not P,E. D.Q. K or R	---	---
Caspase 7	DEVD▼	D	E	V	D	Not P,E.	---	---

						D.Q. K or R		
Caspase 8		I or L	E	T	D	Not P,E. D.Q. K or R	---	---
Caspase 9	LEHD▼	L	E	H	D	---	---	---
Caspase 10	IEHD▼	I	E	H	D	---	---	---

The present invention may employ destructive cleavage sites that are cleavable by a mammalian blood protease, such as Thrombin, Coagulation Factor VIIa, Coagulation Factor IXa, Coagulation Factor Xa, Coagulation Factor XIa, Coagulation Factor XIIa, Kallikrein, Protein C, and MBP-associated serine protease.

Lin, *et al.* describe the use of thrombin or Factor Xa cleavage sites to inactivate the H_C binding domain of a clostridial holotoxin. As discussed above, however, H_C inactivation is inadequate to achieve desirable off-site inactivation, Moreover, due to the pausity of cleavage sites disclosed, the method described by Lin, *et al.* has limited utility, for example in off-site environments where thrombin and Factor Xa are absent (or only present at low concentrations).

Matrix metalloproteases (MMPs) are a preferred group of destructive proteases in the context of the present invention. Within this group, ADAM17 (EC 3.4.24.86), also known as TACE, is preferred and cleaves a variety of membrane-anchored, cell-surface proteins to "shed" the extracellular domains. Additional, preferred MMPs include adamalysins, serralysins, and astacins.

In one embodiment of the present invention, said destructive cleavage site(s) comprises a recognition sequence having at least 3 or 4, preferably 5 or 6, more preferably 6 or 7, and particularly preferably at least 8 contiguous amino acid

residues. In this regard, the longer (in terms of contiguous amino acid residues) the recognition sequence, the less likely non-specific cleavage of the destructive site will occur via an unintended second protease.

The polypeptide of the present invention optionally includes a Targeting Moiety (TM) that binds to a Binding Site on a neuronal (eg. nerve) cell, thereby providing selectivity of the polypeptide to this species of target cell over other cells. In one embodiment, the neuronal cell is a cell of the neuromuscular junction or presynaptic cholinergic peripheral nerve terminal.

The first (and subsequent) destructive cleavage site(s) of the present invention is preferably introduced into the protease component and/ or into the translocation component. Of these two components, the protease component is preferred. Accordingly, the polypeptide may be rapidly inactivated by direct destruction of the non-cytotoxic protease and/ or translocation components. These insertion positions are preferable over a TM insertion position because, even in the case of total TM inactivation, it has been shown that the resulting polypeptide may not demonstrate adequately reduced potency on off-site cells [Chaddock, JA., et al. Protein Expression Purification 2002, 25, 219-228 and Sutton, JM, et al. Protein Expression & Purification 2005, 40(1), 31-41].

Thus, the polypeptide of the present invention does not comprise a destructive cleavage site(s) solely within the Targeting Moiety component of the polypeptide. Without wishing to be bound by any theory, it is believed that use of a destruction site within the TM component alone does not address non-specific uptake by off-site target cells. Example 39 (see also Figure 1) demonstrates that a fragment of botulinum neurotoxin type C lacking the binding domain H_C is still able to enter eSCN and cleave its substrate SNARE protein (syntaxin). A further possibility is that cleavage within the TM component might lead to a TM having increased binding affinity for off-site cells, for example, via exposure of a higher affinity binding region within the TM. In summary, the use of a destructive cleavage site(s) within the TM

component alone is considered unsatisfactory. First, off-site targeting is not adequately addressed, and, secondly, once delivered to an off-site cell, the polypeptides are still capable of (wild-type/ natural) translocation activity and/ or SNARE protein cleavage activity.

It is preferred that the TM has no destructive cleavage site. In this regard, it has been shown that the TM component may be particularly susceptible to adverse conformational changes (upon insertion of a destructive cleavage site), which adversely affect desired binding of the polypeptide. This has been shown to be a particular problem when the TM is the native targeting moiety of a clostridial neurotoxin (i.e. H_C).

Suitable TMs for use in the polypeptides of the present invention include cytokines, growth factors, neuropeptides, lectins, protein binding scaffolds, and antibodies – this term includes monoclonal antibodies, and antibody fragments such as Fab, $F(ab)'_2$, Fv, ScFv, etc.

The TM is a ligand (preferably a peptide) that binds to a neuronal cell, preferably to a neuronal cell of the neuromuscular junction. In this regard, in one embodiment the TM comprises the binding domain (H_{CC}, or H_C) of a clostridial neurotoxin (e.g. BoNT, TeNT, or from other *Clostridium spp.*), or a fragment thereof that possesses native neurotoxin binding ability. The clostridial H_C domain has evolved to bind in a highly effective manner to receptors present on neuronal cells. In accordance with this latter embodiment, the present invention provides use and corresponding methods for modifying BOTOX™ to improve its clinical utility. By way of example, suitable TM clostridial H_{CC} reference sequences include:

Botulinum type A neurotoxin - amino acid residues (Y1111-L1296)
Botulinum type B neurotoxin - amino acid residues (Y1098-E1291)
Botulinum type C neurotoxin - amino acid residues (Y1112-E1291)
Botulinum type D neurotoxin - amino acid residues (Y1099-E1276)

Botulinum type E neurotoxin - amino acid residues (Y1086-K1252)
Botulinum type F neurotoxin - amino acid residues (Y1106-E1274)
Botulinum type G neurotoxin - amino acid residues (Y1106-E1297)
Tetanus neurotoxin - amino acid residues (Y1128-D1315).

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

Similarly, by way of example, suitable TM clostridial H_C domains of reference sequences include: BoNT/A - N872-L1296; BoNT/B - E859-E1291; BoNT/C1 - N867-E1291; BoNT/D - S863-E1276; BoNT/E - R846-K1252; BoNT/F - K865-E1274; BoNT/G - N864-E1297; and TeNT - I880-D1315.

In another embodiment, the TM is selected to provide desirable binding to the neuromuscular junction. Suitable TMs are listed in WO 2006/099590, which is herein incorporated by reference thereto, and include: glucagon like hormone, a neurohormone, a neuroregulatory cytokine, a neurotrophin, a growth factor, an axon guidance signaling molecule, a sugar binding protein, a ligand that selectively binds a neurexin, a ligand for neurexin-2α, a ligand for neurexin-2β, a ligand for neurexin-3α, a ligand for neurexin-3β, a WNT, Ng-CAM(LI), NCAM, N-cadherin, a PACAP peptide such as a VIP peptide, Agrin-MUSK, a basement membrane polypeptide, and a variant of any of the foregoing polypeptides, a neuroregulatory cytokine such as ciliary neurotrophic factor (CNTF), glycophorin-A (GPA), leukemia inhibitory factor (LIF), an interleukin (IL), onostatin M, cardiotrophin-1 (CT-1), cardiotrophin-like cytokine (CLC), a neuroleukin, VEGF, an insulin-like growth factors (IGF), an epidermal growth factor (EGF), and a variant of any of the foregoing neuroregulatory cytokines. These and other TMs are selected for use in the present invention because they mimic the binding ability of clostridial neurotoxins.

As mentioned above, the destructive cleavage site(s) are introduced with minimum adverse effect on the biological properties of the polypeptide

(notably, endopeptidase activity, and/ or membrane translocation activity). In this regard, it is preferred that any potential decrease in potency of the polypeptide (compared with the same polypeptide lacking said destructive cleavage site(s)) is less than 25%, preferably less than 15%, more preferably less than 5% of the original unmodified protein. Potency here may be measured by a comparative assay such as illustrated in Examples 1-4.

When selecting destructive cleavage site(s) in the context of the present invention, it is preferred that the destructive cleavage site(s) are not substrates for any proteases that may be separately used for post-translational modification of the polypeptide of the present invention as part of its manufacturing process. In this regard, the non-cytotoxic proteases of the present invention typically employ a protease activation event (via a separate 'activation' protease cleavage site, which is structurally distinct from the destructive cleavage site of the present invention). The purpose of the activation cleavage site is to cleave a peptide bond between the non-cytotoxic protease and translocation or TM components of the polypeptide of the present invention, thereby providing an 'activated' di-chain polypeptide wherein said two components are linked together via a di-sulfide bond.

In natural clostridial holotoxin, the di-chain loop protease cleavage site occurs at K448-A449 for BoNT/A, at K441-A442 for BoNT/B, at K449-T450 for BoNT/C1, at R445-D446 for BoNT/D, at R422-K423 for BoNT/E, at K439-A440 for BoNT/F, at K446-S447 for BoNT/G, and at A457-S458 for TeNT. Thus, to help ensure that the destructive cleavage site of the polypeptides of the present invention does not adversely affect the 'activation' cleavage site and subsequent di-sulfide bond formation, the former is preferably introduced into polypeptide of the present invention at a position of at least 20, at least 30, at least 40, at least 50, and more preferably at least 60, at least 70, at least 80 (contiguous) amino acid residues away from the 'activation' cleavage site. In this regard, the activation site of a polypeptide of the invention may be readily aligned (via simple, primary sequence alignment) with the activation site

positions (listed above) for clostridial holotoxin.

The destructive cleavage site(s) are preferably exogenous (i.e. engineered/ artificial) with regard to the native components of the polypeptide. In other words, said cleavages sites are preferably not inherent to the corresponding native components of the polypeptide. By way of example, a protease or translocation component based on BoNT/A L-chain or H-chain (respectively) may be engineered according to the present invention to include a cleavage site(s). Said cleavage site(S) would not, however, be present in the corresponding BoNT native L-chain or H-chain.

In a preferred embodiment of the present invention, the destructive cleavage site(s) and the 'activation' cleavage site are not cleaved by the same protease. In one embodiment, the two cleavage sites differ from one another in that at least one, more preferably at least two, particularly preferably at least three, and most preferably at least four of the tolerated amino acids within the respective recognition sequences is/ are different.

By way of example, in the case of a polypeptide chimaera containing a Factor Xa 'activation' site between clostridial L-chain and H_N components, it is preferred to employ a destructive cleavage site(s) that is a site other than a Factor Xa site, which may be inserted elsewhere in the L-chain and/ or H_N component(s). In this scenario, the polypeptide may be modified to accommodate an alternative 'activation' site between the L-chain and H_N components (for example, an enterokinase cleavage site), in which case a separate Factor Xa cleavage site(s) may be incorporated elsewhere into the polypeptide as the destructive cleavage site. Alternatively, the existing Factor Xa 'activation' site between the L-chain and H_N components may be retained, and an alternative cleavage site such as a thrombin cleavage site incorporated as the destructive cleavage site(s).

When identifying suitable sites within the primary sequence of any of the

components of the present invention for inclusion of cleavage site(s), it is preferable to select a primary sequence that closely matches with the proposed cleavage site(s) that are to be inserted. By doing so, minimal structural changes are introduced into the polypeptide. By way of example, cleavage sites typically comprise at least 3 contiguous amino acid residues. Thus, in a preferred embodiment, a cleavage site is selected that already possesses (in the correct position(s)) at least one, preferably at least two of the amino acid residues that are required in order to introduce the new cleavage site. By way of example, when the Caspase 3 cleavage site (DMQD) is to be introduced, a preferred insertion position may be identified that already includes a primary sequence selected from, for example, Dxxx, xMxx, xxQx, xxxD, DMxx, DxQx, DxxD, xMQx, xMxD, xxQD, DMQx, xMQD, DxQD, and DMxD.

By analysis of the tertiary structure of clostridial neurotoxin, the present inventors have identified a range of suitable exposed regions (in particular exposed loop regions) for insertion of the destructive site sequence(s). This analysis has been based principally on Chaddock & Marks (2006) in Cell & Molecular Life Sciences, 63, 540-551; and with additional reference to http://pathema.tigr.org/pathema/BoNT_structures.shtml); Lacy and Stevens, 1999, J. Mol Biol., 291, 1091-1104; and the following Table.

BoNT Serotype	**PDB ID**	**PDB Description**
A	1E1H	Crystal structure of recombinant botulinum neurotoxin type A light chain, self-inhibiting Zn endopeptidase
A	1XTF	Neurotoxin BoNT/A E224Q Y366F mutant
A	1XTG	Crystal structure of neurotoxin BONT/A complexed with synaptosomal-associated protein 25
A	3BTA	Crystal structure of botulinum neurotoxin serotype A
B	1EPW	Crystal Structure of Clostridium neurotoxin type B

B	1F31	Crystal structure of *Clostridium botulinum* neurotoxin B complexed with a trisaccharide
B	1F82	Botulinum neurotoxin type B catalytic domain
B	1F83	Botulinum neurotoxin type B catalytic domain with synaptobrevin-II bound
B	1G9A	Crystal structure of *Clostridium botulinum* B complexed with an inhibitor (Experiment 3)
B	1G9B	Crystal structure of *Clostridium botulinum* neurotoxin B complexed with an inhibitor (Experiment 1)
B	1G9C	Crystal structure of *Clostridium botulinum* neurotoxin B complexed with an inhibitor (Experiment 4)
B	1G9D	Crystal structure of *Clostridium botulinum* neurotoxin B complexed with an inhibitor (Experiment 2)
B	1I1E	Crystal structure of *Clostridium botulinum* neurotoxin B complexed with doxorubicin
B	1S0B	Crystal structure of botulinum neurotoxin type B at pH 4.0
B	1S0C	Crystal structure of botulinum neurotoxin type B at pH 5.0
B	1S0D	Crystal structure of botulinum neurotoxin type B at pH 5.5
B	1S0E	Crystal structure of botulinum neurotoxin type B at pH 6.0
B	1S0F	Crystal structure of botulinum neurotoxin type B at pH 7.0
B	1S0G	Crystal structure of botulinum neurotoxin type B apo form
B	1Z0H	N-terminal helix reorients in recombinant C-fragment of *Clostridium botulinum* type B
B	2ETF	Crystal structure of full length botulinum neurotoxin (type B) light chain
D	2FPQ	Crystal structure of botulinum neurotoxin type D light chain
E	1T3A	Crystal structure of *Clostridium botulinum* neurotoxin type E catalytic domain

E	1T3C	Clostridium botulinum type E catalytic domain E212Q mutant
E	1ZKW	Crystal structure of Arg347Ala mutant of botulinum neurotoxin E catalytic domain
E	1ZKX	Crystal structure of Glu158Ala/Thr159Ala/Asn160Ala- a triple mutant of *Clostridium botulinum* neurotoxin E catalytic domain
E	1ZL5	Crystal structure of Glu335Gln mutant of *Clostridium botulinum* neurotoxin E catalytic domain
E	1ZL6	Crystal structure of Tyr350Ala mutant of *Clostridium botulinum* neurotoxin E catalytic domain
E	1ZN3	Crystal structure of Glu335Ala mutant of *Clostridium botulinum* neurotoxin type E
F	2A8A	Crystal structure of *Clostridium botulinum* neurotoxin serotype F light chain
F	2A97	Crystal structure of catalytic domain of *Clostridium botulinum* neurotoxin serotype F
G	1ZB7	Crystal Structure of botulinum neurotoxin type G light chain

The above PDB identification refers to the 4 character code used by the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) to identify a specific entry in the structural database.

Additional techniques employed include use of peptide/ antibody mapping information, for example, antibody mapping of sites on the surface of HC/A (Dolimbek, BZ, 2007, Mol Immunol., 44(5):1029-41), HN/A (Atassi MZ, 2004, Protein J. 23(1):39-52), H_C/A (Oshima M., 1998, Immunol Lett., 60(1):7-12; Bavari, S 1998, Vaccine, 16(19):1850-6), HC/E (Kubota T, 1997, Appl Environ Microbiol. 63(4):1214-8) - a list of epitopes within the BoNT serotypes is publicly available and maintained at

http://pathema.tigr.org/pathema/BoNT_epitopes.shtml; and use of structural prediction software to predict the solvent accessibility of a specific peptide region – examples of publically available software include:
Swiss Model (http://swissmodel.expasy.org);
ESyPred3D (http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred); and Geno3D (http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html).

In one embodiment of the present invention, the destructive cleavage site(s) are introduced at one or more of the following position(s), which are based (for convenience purposes) on the primary amino acid sequence of BoNT/A. Whilst the insertion positions are identified by reference to BoNT/A, the primary amino acid sequences of corresponding protease domains and/ or translocation domains for BoNT/B-G etc may be readily aligned with said BoNT/A positions – by way of example, we refer to the serotype alignment illustrated in Figure 2.

For the protease component, one or more of the following positions is preferred: 27-31, 56-63, 73-75, 78-81, 99-105, 120-124, 137-144, 161-165, 169-173, 187-194, 202-214, 237-241, 243-250, 300-304, 323-335, 375-382, 391-400, and 413-423. The above numbering preferably starts from the N-terminus of the protease component of the present invention. Of these positions, the 99-105 and/ or 202-214 are most preferred. In this regard, referring to Figure 2, positions 99-105 correspond to the sequence "YSTDLGR" for serotype A, which equates to the region "KSKPLGE" for serotype B, "NSREIGE" for serotype C_1, "NERDIGK" for serotype D, "NNNLSGG" for serotype E, "NSNPAGQ" for serotype F, and "NSKPSGQ" for serotype G. Similarly, referring to Figure 2, positions 202-214 correspond to the sequence "VDTNPLLGAGKFA" for serotype A, which equates to the region "NKGASIFNRRGYF" for serotype B, "DVGEGRFSKSEFC" for serotype C_1, "NQSSAVLGKSIFC" for serotype D, "DNC----MN--EFI" for serotype E, "DN-----TD--LFI" for serotype F, and "ENKDTSIFSRRAYF" for serotype G. and "P" (202) using the numbering at the top of Figure 2 as and "P", respectively.

In a preferred embodiment, the destructive cleavage site(s) are located at a position greater than 8 amino acid residues, preferably greater than 10 amino acid residues, more preferably greater than 25 amino acid residues, particularly preferably greater than 50 amino acid residues from the N-terminus of the protease component. Similarly, in a preferred embodiment, the destructive cleavage site(s) are located at a position greater than 20 amino acid residues, preferably greater than 30 amino acid residues, more preferably greater than 40 amino acid residues, particularly preferably greater than 50 amino acid residues from the C-terminus of the protease component.

For the translocation component, one or more of the following positions is preferred: 474-479, 483-495, 507-543, 557-567, 576-580, 618-631, 643-650, 669-677, 751-767, 823-834, 845-859. The above numbering preferably acknowledges a starting position of 449 for the N-terminus of the translocation domain component of the present invention, and a starting position of 871 for the C-terminus of the H_N component. Of these positions, the 557-567 and/ or 751-767 are most preferred. In this regard, referring to Figure 2, positions 557-567 correspond to the sequence "QEFEHGKSRIA" for serotype A, which equates to the region "QTFPLDIRDIS" for serotype B, "QKLSDNVEDFT" for serotype C_1, "QKLSNNVENIT" for serotype D, "QKVPEGENNVN" for serotype E, "QKAPEGESAIS" for serotype F, and "QTFPSNIENLQ" for serotype G. Similarly, referring to Figure 2, positions 751-767 correspond to the sequence "YNQYTEEEKNNINNID" for serotype A, which equates to the region "YNIYSEKEKSNIN--IDFN" for serotype B, "YKKYSGSDKENIKS--QVE" for serotype C_1, "YKKYSGSDKENIKS--QVE" for serotype D, "YNSYTLEEKNELTNKYDIK" for serotype E, "YNNYTLDEKNRLRAEYNIY" for serotype F, and "YNRYSEEDKMNIN--IDFN" for serotype G.

In a preferred embodiment, the destructive cleavage site(s) are located at a position greater than 10 amino acid residues, preferably greater than 25 amino acid residues, more preferably greater than 40 amino acid residues, particularly

preferably greater than 50 amino acid residues from the N-terminus of the translocation component. Similarly, in a preferred embodiment, the destructive cleavage site(s) are located at a position greater than 10 amino acid residues, preferably greater than 25 amino acid residues, more preferably greater than 40 amino acid residues, particularly preferably greater than 50 amino acid residues from the C-terminus of the translocation component.

According to a second aspect of the present invention, there is provided use of a non-cytotoxic polypeptide for treating a range of diverse medical conditions and diseases. Said conditions and diseases have established therapies (see the background part of the present specification) based on very closely related (though unmodified as per the present invention) non-cytotoxic polypeptides. Accordingly, the present invention provides improvements to said therapies by use of a modified non-cytotoxic polypeptide that has a destructive cleavage site and thus reduced off-site effects.

In particular, the present invention provides use and corresponding methods for the treatment of strabismus, blepharospasm, squint, spasmodic and oromandibular dystonia, torticollis, and other beauty therapy (cosmetic) applications benefiting from cell/ muscle incapacitation (via SNARE down-regulation or inactivation).

Additional, related therapies are provided for treating a neuromuscular disorder or condition of ocular motility, e.g. concomitant and vertical strabismus, lateral rectus palsy, nystagmus, dysthyroid myopathy, etc.; dystonia, e.g. focal dystonias such as spasmodic torticollis, writer's cramp, blepharospasm, oromandibular dystonia and the symptoms thereof, e.g. bruxism, Wilson's disease, tardive dystonia, laryngeal dystonia etc.; other dystonias, e.g. tremor, tics, segmental myoclonus; spasms, such as spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, e.g. in patients with spinal cord injury, animus, back spasm, charley horse etc.; tension headaches; levator pelvic syndrome; spina bifida, tardive dyskinesia; Parkinson's and limb

(focal) dystonia and stuttering, etc.

In use, a polypeptide of the invention binds to a surface structure (the Binding Site), which is present on and preferably characteristic of a target cell. Following binding, the polypeptide (at least the protease component thereof) becomes endocytosed into a vesicle, and the translocation component then directs transport of the protease component across the endosomal membrane and into the cytosol of the target cell. Once inside the target cell, the non-cytotoxic protease inhibits the cellular exocytic fusion process, and thereby inhibits release/ secretion from the target cell.

The biologically active component of the polypeptides of the present invention is a non-cytotoxic protease. Non-cytotoxic proteases are a discrete class of molecules that do not kill cells; instead, they act by inhibiting cellular processes other than protein synthesis. Non-cytotoxic proteases are produced as part of a larger toxin molecule by a variety of plants, and by a variety of microorganisms such as Clostridium sp. and Neisseria sp.

Clostridial neurotoxins represent a major group of non-cytotoxic toxin molecules, and comprise two polypeptide chains joined together by a disulphide bond. The two chains are termed the heavy chain (H-chain), which has a molecular mass of approximately 100 kDa, and the light chain (L-chain), which has a molecular mass of approximately 50 kDa. It is the L-chain, which possesses a protease function and exhibits a high substrate specificity for vesicle and/or plasma membrane associated (SNARE) proteins involved in the exocytic process (eg. synaptobrevin, syntaxin or SNAP-25). These substrates are important components of the neurosecretory machinery.

Neisseria sp., most importantly from the species *N. gonorrhoeae*, produce functionally similar non-cytotoxic toxin molecules. An example of such a non-cytotoxic protease is IgA protease (see WO99/58571).

The choice of TM determines the specificity of the polypeptide. By way of example, the same (or similar) receptor may be present on several different cells such that one TM will bind to different cell types. In this scenario, it might be desirable only to target a single cell type. Thus, by employing a second protease ('destruction') cleavage site in a polypeptide of the present invention which is cleaved by a protease specific to one or more of the undesired cells (and/ or to their environment), it is possible to minimise off-target side effects in the undesired cells.

In another embodiment, polypeptides of the present invention may comprise two or more different TMs capable of binding to different target cell types. Alternatively (or in addition), combinations of polypeptides may be employed having different TMs so as to provide a coordinated targeting of different target cell types.

Polypeptide preparation

The polypeptides of the present invention comprise 4 principal components: a TM; a non-cytotoxic protease; a translocation domain; and a destructive protease cleavage site. Said polypeptides embrace non-cytotoxic holotoxins such as clostridial neurotoxins, and, when an exogenous TM is present, re-targeted chimaeras (often referred to as re-targeted proteases). Preparation of these molecules is conventional - by way of exemplification, we refer to: WO94/21300; WO96/33273; WO98/07864; WO00/10598; WO01/21213; WO06/059093; WO00/62814; WO00/04926; WO93/15766; WO00/61192; and WO99/58571. All of these publications are herein incorporated by reference thereto.

In more detail, the TM component of the present invention may be fused to either the protease component or the translocation component of the present invention. Said fusion is preferably by way of a covalent bond, for example either a direct covalent bond or via a spacer/ linker molecule. The protease component and the translocation component are preferably linked together via a covalent bond, for example either a direct covalent bond or via a spacer/ linker molecule. Suitable

spacer/ linked molecules are well known in the art, and typically comprise an amino acid-based sequence of between 5 and 40, preferably between 10 and 30 amino acid residues in length.

In use, the polypeptides have a di-chain conformation, wherein the protease component and the translocation component are linked together, preferably via a disulphide bond.

The polypeptides of the present invention may be prepared by conventional chemical conjugation techniques, which are well known to a skilled person. By way of example, reference is made to Hermanson, G.T. (1996), Bioconjugate techniques, Academic Press, and to Wong, S.S. (1991), Chemistry of protein conjugation and cross-linking, CRC Press.

Alternatively, the polypeptides may be prepared by recombinant preparation of a single polypeptide fusion protein (see, for example, WO98/07864). This technique is based on the *in vivo* bacterial mechanism by which native clostridial neurotoxin (ie. holotoxin) is prepared, and results in a fusion protein having the following 'simplified' structural arrangement:

NH_2 - [protease component] – [translocation component] – [TM] - COOH

According to WO98/07864, the TM is placed towards the C-terminal end of the fusion protein. The fusion protein is then 'activated' by treatment with a protease, which cleaves at a site between the protease component and the translocation component. A di-chain protein is thus produced, comprising the protease component as a single polypeptide chain covalently attached (via a disulphide bridge) to another single polypeptide chain containing the translocation component plus TM.

The WO98/07864 system is particularly suited to the preparation of fusion proteins having a TM that requires a C-terminal domain that is 'free' for

interaction with a Binding Site on a target cell.

For fusion proteins having a TM that requires an N-terminal domain that is 'free' for interaction with a Binding Site on a target cell, a modified system may be employed as described in WO06/059113.

In the modified system, the TM component of the fusion protein is located towards the middle of the linear fusion protein sequence, between the protease cleavage site and the translocation component. This ensures that the TM is attached to the translocation domain (ie. as occurs with native clostridial holotoxin), though in this case the two components are reversed in order *vis-à-vis* native holotoxin. Subsequent cleavage at the protease cleavage site exposes the N-terminal portion of the TM, and provides the di-chain polypeptide fusion protein.

The above-mentioned protease cleavage sequence(s) may be introduced (and/ or any inherent cleavage sequence removed) at the DNA level by conventional means, such as by site-directed mutagenesis. Screening to confirm the presence of cleavage sequences may be performed manually or with the assistance of computer software (e.g. the MapDraw program by DNASTAR, Inc.). Whilst any protease cleavage site may be employed (ie. clostridial, or non-clostridial), the following are preferred (either as the 'destructive' cleavage site, or as the 'activation' cleavage site):

Enterokinase	(DDDDK↓)
Factor Xa	(IEGR↓ / IDGR↓)
TEV(Tobacco Etch virus)	(ENLYFQ↓G)
Thrombin	(LVPR↓GS)
PreScission	(LEVLFQ↓GP).

Also embraced by the term protease cleavage site is an intein, which is a self-cleaving sequence. The self-splicing reaction is controllable, for example by varying the concentration of reducing agent present.

In a preferred embodiment, the fusion protein of the present invention may comprise one or more N-terminal and/ or C-terminal located purification tags. Whilst any purification tag may be employed, the following are preferred:

His-tag (e.g. 6 × histidine), preferably as a C-terminal and/ or N-terminal tag
MBP-tag (maltose binding protein), preferably as an N-terminal tag
GST-tag (glutathione-S-transferase), preferably as an N-terminal tag
His-MBP-tag, preferably as an N-terminal tag
GST-MBP-tag, preferably as an N-terminal tag
Thioredoxin-tag, preferably as an N-terminal tag
CBD-tag (Chitin Binding Domain), preferably as an N-terminal tag.

One or more peptide spacer/ linker molecules may be included in the fusion protein. For example, a peptide spacer may be employed between a purification tag and the rest of the fusion protein molecule.

Thus, a third aspect of the present invention provides a nucleic acid (e.g. DNA) sequence encoding a polypeptide as described above.

Said nucleic acid may be included in the form of a vector, such as a plasmid, which may optionally include one or more of an origin of replication, a nucleic acid integration site, a promoter, a terminator, and a ribosome binding site.

The present invention also includes a method for expressing the above-described nucleic acid sequence (i.e. the third aspect of the present invention) in a host cell, in particular in *E. coli* or via a baculovirus expression system.

The present invention also includes a method for activating a polypeptide of the

present invention, said method comprising contacting the polypeptide with a protease that cleaves the polypeptide at a recognition site (cleavage site) located between the non-cytotoxic protease component and the translocation component, thereby converting the polypeptide into a di-chain polypeptide wherein the non-cytotoxic protease and translocation components are joined together by a disulphide bond. In a preferred embodiment, the recognition site is not native to a naturally-occurring clostridial neurotoxin and/ or to a naturally-occurring IgA protease.

Polypeptide delivery

In use, the present invention employs a pharmaceutical composition, comprising a polypeptide, together with at least one component selected from a pharmaceutically acceptable carrier, excipient, adjuvant, propellant and/ or salt.

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

In the case of a polypeptide that is to be delivered locally, the polypeptide may be formulated as a cream (eg. for topical application), or for sub-dermal injection.

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

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

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

In the case of formulations for injection, it is optional to include a pharmaceutically active substance to assist retention at or reduce removal of the polypeptide from the site of administration. One example of such a pharmaceutically active substance is a vasoconstrictor such as adrenaline. Such a formulation confers the advantage of increasing the residence time of polypeptide following administration and thus increasing and/or enhancing its effect.

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

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

A particularly preferred dosing regimen is based on 0.25 ng of polypeptide as the 1X dose. In this regard, preferred dosages are in the range 1X–100X (i.e. 0.25-25 ng).

Fluid dosage forms are typically prepared utilising the polypeptide and a pyrogen-free sterile vehicle. The polypeptide, depending on the vehicle and

concentration used, can be either dissolved or suspended in the vehicle. In preparing solutions the polypeptide can be dissolved in the vehicle, the solution being made isotonic if necessary by addition of sodium chloride and sterilised by filtration through a sterile filter using aseptic techniques before filling into suitable sterile vials or ampoules and sealing. Alternatively, if solution stability is adequate, the solution in its sealed containers may be sterilised by autoclaving. Advantageously additives such as buffering, solubilising, stabilising, preservative or bactericidal, suspending or emulsifying agents and or local anaesthetic agents may be dissolved in the vehicle.

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

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

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

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

Definitions Section

Targeting Moiety (TM) means any chemical structure that functionally interacts with a Binding Site to cause a physical association between the polypeptide of the invention and the surface of a target cell. The term TM embraces any molecule (ie. a naturally occurring molecule, or a chemically/physically modified variant thereof) that is capable of binding to a Binding Site on the target cell, which Binding Site is capable of internalisation (eg. endosome formation) - also referred to as receptor-mediated endocytosis. The TM may possess an endosomal membrane translocation function, in which case separate TM and Translocation Domain components need not be present in an agent of the present invention. Throughout the preceding description, specific TMs have been described. Reference to said TMs is merely exemplary, and the present invention embraces all variants and derivatives thereof, which retain the basic binding (i.e. targeting) ability of the exemplified TMs.

As mentioned previously, preferred TMs include antibodies (eg. antibody fragments) and binding scaffolds; especially commercially available antibodies/ fragments and scaffolds designed for the purpose of binding (eg. specifically) to nerve cells.

Protein scaffolds represent a new generation of universal binding frameworks to complement the expanding repertoire of therapeutic monoclonal antibodies and derivatives such as scFvs, Fab molecules, dAbs (single-domain antibodies), diabodies and minibodies, each of which may be employed as a TM of the present invention. Scaffold systems create or modify known protein recognition domains either through creation of novel scaffolds or modification of known protein binding domains. Such scaffolds include but are not limited to:
(i) protein A based scaffolds - affibodies (Nord, K. et al 1997 "Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain". Nat Biotechnol 15, 772–777);
(ii) lipocalin based scaffolds – anticalins (Skerra 2008 "Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities". FEBS J. 275:2677-83);

(iii) fibronectin based scaffolds – adnectin (Dineen et al 2008 "The Adnectin CT-322 is a novel VEGF receptor 2 inhibitor that decreases tumor burden in an orthotopic mouse model of pancreatic cancer". BMC Cancer 8:352);

(iv) avimers (Silverman et al 2005 "Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains". Nat Biotechnol 23:1556-61);

(v) ankyrin based scaffolds – darpins (Zahnd et al 2006 "Selection and characterization of Her2 binding-designed ankyrin repeat proteins". J Biol Chem. 281:35167-75); and

(vi) centyrin scaffolds – based on a protein fold that has significant structural homology to Ig domains with loops that are analogous to CDRs. Ig domains are a common module in human proteins and have been widely applied as alternative scaffold proteins. Each of the above 'scaffold' publications is hereby incorporated (in its entirety) by reference thereto.

Binding scaffolds can be used to target particular cell types via interaction with specific cell surface proteins, receptors or other cell surface epitopes such as sugar groups. Such modified scaffolds can be engineered onto recombinant non-cytotoxic protease based polypeptides of the present invention to target specific nerve cell types of interest.

The TM of the present invention binds (preferably specifically binds) to the target cell in question. The term "specifically binds" preferably means that a given TM binds to the target cell with a binding affinity (Ka) of 10^6 M^{-1} or greater, preferably 10^7 M^{-1} or greater, more preferably 10^8 M^{-1} or greater, and most preferably, 10^9 M^{-1} or greater.

Reference to TM in the present specification embraces fragments and variants thereof, which retain the ability to bind to the target cell in question. By way of example, a variant may have at least 80%, preferably at least 90%, more preferably at least 95%, and most preferably at least 97 or at least 99% amino acid sequence homology with the reference TM. Thus, a variant may include one or more analogues of an amino acid (e.g. an unnatural amino acid), or a

substituted linkage. Also, by way of example, the term fragment, when used in relation to a TM, means a peptide having at least ten, preferably at least twenty, more preferably at least thirty, and most preferably at least forty amino acid residues of the reference TM. The term fragment also relates to the above-mentioned variants. Thus, by way of example, a fragment of the present invention may comprise a peptide sequence having at least 10, 20, 30 or 40 amino acids, wherein the peptide sequence has at least 80% sequence homology over a corresponding peptide sequence (of contiguous) amino acids of the reference peptide.

By way of example, ErbB peptide TMs (eg. EGF) may be modified to generate mutein ErbB ligands with improved properties such as increased stability. By way of example, ErbB TM muteins include ErbB peptides having amino acid modifications such as a valine residue at position 46 or 47 (EGFVal46 or 47), which confers stability to cellular proteases. ErbB TMs may also have amino acids deleted or additional amino acids inserted. This includes but is not limited to EGF having a deletion of the two C-terminal amino acids and a neutral amino acid substitution at position 51 (particularly EGF51Gln51; see US20020098178A1), and EGF with amino acids deleted (e.g. rEGF2-48; rEGF3-48 and rEGF4-48). Fragments of ErbB TMs may include fragments of TGFα which contain predicted β-turn regions (e.g. a peptide of the sequence Ac-C-H-S-G-Y-V-G-A-R-C-O-OMe), fragments of EGF such as [Ala20]EGF(14-31), and the peptide YHWYGYTPQNVI or GE11. All of the above patent specifications are incorporated herein by reference thereto.

It is routine to confirm that a TM binds to the selected target cell. For example, a simple radioactive displacement experiment may be employed in which tissue or cells representative of the target cell are exposed to labelled (eg. tritiated) TM in the presence of an excess of unlabelled TM. In such an experiment, the relative proportions of non-specific and specific binding may be assessed, thereby allowing confirmation that the TM binds to the target cell. Optionally, the assay may include one or more binding antagonists, and the

assay may further comprise observing a loss of TM binding. Examples of this type of experiment can be found in Hulme, E.C. (1990), Receptor-binding studies, a brief outline, pp. 303-311, In Receptor biochemistry, A Practical Approach, Ed. E.C. Hulme, Oxford University Press.

In the context of the present invention, reference to a peptide TM embraces peptide analogues thereof, so long as the analogue binds to the same receptor as the corresponding 'reference' TM. Said analogues may include synthetic residues such as:

ß-Nal = ß-naphthylalanine
ß-Pal = ß -pyridylalanine
hArg(Bu) = N-guanidino-(butyl)-homoarginine
$hArg(Et)_2$ = N, N'-guanidino-(dimethyl)-homoarginine
$hArg(CH_2CF_3)_2$ = N, N'-guanidino-bis-(2,2,2,-trifluoroethyl)-homoarginine
$hArg(CH_3$, hexyl) = N, N'-guanidino-(methyl, hexyl)- homoarginine
Lys(Me) = N^e-methyllysine
Lys(iPr) = N^e-isopropyllysine
AmPhe = aminomethylphenylalanine
AChxAla = aminocyclohexylalanine
Abu = α-aminobutyric acid
Tpo = 4-thiaproline
MeLeu = N-methylleucine
Orn = ornithine
Nle - norleucine
Nva = norvaline
Trp(Br) = 5-bromo-tryptophan
Trp(F) = 5-fluoro-tryptophan
$Trp(N0_2)$ = 5-nitro-tryptophan
Gaba = γ-aminobutyric acid
Bmp = J-mercaptopropionyl
Ac = acetyl

Pen - pencillamine

The polypeptides of the present invention may lack a functional H_C (or H_{CC}) domain of a clostridial neurotoxin, in which case a non-clostridial TM is typically present to bind the polypeptide to a Binding Site on the nerve cell. The H_C peptide of a native clostridial neurotoxin comprises approximately 400-440 amino acid residues, and consists of two functionally distinct domains of approximately 25kDa each, namely the N-terminal region (commonly referred to as the H_{CN} peptide or domain) and the C-terminal region (commonly referred to as the H_{CC} peptide or domain). It has been well documented that the C-terminal region (H_{CC}), which constitutes the C-terminal 160-200 amino acid residues, is responsible for binding of a clostridial neurotoxin to its natural cell receptors, namely to nerve terminals at the neuromuscular junction - this fact is also confirmed by the above publications. Thus, reference throughout this specification to a clostridial heavy-chain lacking a functional heavy chain H_C peptide (or domain) means that the clostridial heavy-chain simply lacks a functional H_{CC} peptide. In other words, the H_{CC} peptide region is either partially or wholly deleted, or otherwise modified (e.g. through conventional chemical or proteolytic treatment) to inactivate its native binding ability for nerve cells.

Alternatively, a polypeptide of the present invention may contain a functional H_C (or H_{CC}) domain of a clostridial neurotoxin as a TM. A variety of clostridial neurotoxin Hcc or Hc regions comprising a binding domain can be useful in aspects of the present invention with the proviso that these active fragments provide the binding activity and binding specificity of the natural neurotoxin. The H_C regions from the heavy chains of clostridial toxins are approximately 400-440 amino acids in length and comprise a binding domain. Research has shown that the entire length of a H_C region from a clostridial toxin heavy chain is not necessary for the binding activity of the binding domain. Thus, aspects of this embodiment can include clostridial toxin H_C regions comprising a binding domain having a length of, for example, at least 350 amino acids, at least 375 amino acids, at least 400 amino acids and at least 425 amino acids. Other

aspects of this embodiment can include clostridial toxin H_C regions comprising a binding domain having a length of, for example, at most 350 amino acids, at most 375 amino acids, at most 400 amino acids and at most 425 amino acids.

The protease of the present invention embraces all non-cytotoxic proteases that are capable of cleaving one or more proteins of the exocytic fusion apparatus in eukaryotic cells.

The protease of the present invention is preferably a bacterial protease (or fragment thereof). More preferably the bacterial protease is selected from the genera *Clostridium* or *Neisseria/ Streptococcus* (e.g. a clostridial L-chain, or a neisserial IgA protease preferably from *N. gonorrhoeae or S. pneumoniae*).

The present invention also embraces variant non-cytotoxic proteases (ie. variants of naturally-occurring protease molecules), so long as the variant proteases still demonstrate the requisite protease activity. By way of example, a variant may have at least 70%, preferably at least 80%, more preferably at least 90%, and most preferably at least 95 or at least 98% amino acid sequence homology with a reference protease sequence. Thus, the term variant includes non-cytotoxic proteases having enhanced (or decreased) endopeptidase activity – particular mention here is made to the increased K_{cat}/K_m of BoNT/A mutants Q161A, E54A, and K165L see Ahmed, S.A. (2008) Protein J. DOI 10.1007/s10930-007-9118-8, which is incorporated by reference thereto. The term fragment, when used in relation to a protease, typically means a peptide having at least 150, preferably at least 200, more preferably at least 250, and most preferably at least 300 amino acid residues of the reference protease. As with the TM 'fragment' component (discussed above), protease 'fragments' of the present invention embrace fragments of variant proteases based on a reference sequence.

The protease of the present invention preferably demonstrates a serine or metalloprotease activity (e.g. endopeptidase activity). The protease is

preferably specific for a SNARE protein (e.g. SNAP-25, synaptobrevin/VAMP, or syntaxin).

Particular mention is made to the protease domains of neurotoxins, for example the protease domains of bacterial neurotoxins. Thus, the present invention embraces the use of neurotoxin domains, which occur in nature, as well as recombinantly prepared versions of said naturally-occurring neurotoxins.

Exemplary neurotoxins are produced by clostridia, and the term clostridial neurotoxin embraces neurotoxins produced by *C. tetani* (TeNT), and by *C. botulinum* (BoNT) serotypes A-G, as well as the closely related BoNT-like neurotoxins produced by *C. baratii* and *C. butyricum*. The above-mentioned abbreviations are used throughout the present specification. For example, the nomenclature BoNT/A denotes the source of neurotoxin as BoNT (serotype A). Corresponding nomenclature applies to other BoNT serotypes.

BoNTs are the most potent toxins known, with median lethal dose (LD50) values for mice ranging from 0.5 to 5 ng/kg depending on the serotype. BoNTs are adsorbed in the gastrointestinal tract, and, after entering the general circulation, bind to the presynaptic membrane of cholinergic nerve terminals and prevent the release of their neurotransmitter acetylcholine. BoNT/B, BoNT/D, BoNT/F and BoNT/G cleave synaptobrevin/vesicle-associated membrane protein (VAMP); BoNT/C, BoNT/A and BoNT/E cleave the synaptosomal-associated protein of 25 kDa (SNAP-25); and BoNT/C cleaves syntaxin.

BoNTs share a common structure, being di-chain proteins of ~150 kDa, consisting of a heavy chain (H-chain) of ~100 kDa covalently joined by a single disulphide bond to a light chain (L-chain) of ~50 kDa. The H-chain consists of two domains, each of ~50 kDa. The C-terminal domain (H_C) is required for the high-affinity neuronal binding, whereas the N-terminal domain (H_N) is proposed

to be involved in membrane translocation. The L-chain is a zinc-dependent metalloprotease responsible for the cleavage of the substrate SNARE protein.

The term L-chain fragment means a component of the L-chain of a neurotoxin, which fragment demonstrates a metalloprotease activity and is capable of proteolytically cleaving a vesicle and/or plasma membrane associated protein involved in cellular exocytosis.

Examples of suitable protease (reference) sequences include:

Botulinum type A neurotoxin	- amino acid residues (1-448)
Botulinum type B neurotoxin	- amino acid residues (1-440)
Botulinum type C neurotoxin	- amino acid residues (1-441)
Botulinum type D neurotoxin	- amino acid residues (1-445)
Botulinum type E neurotoxin	- amino acid residues (1-422)
Botulinum type F neurotoxin	- amino acid residues (1-439)
Botulinum type G neurotoxin	- amino acid residues (1-441)
Tetanus neurotoxin	- amino acid residues (1-457)
IgA protease	- amino acid residues (1-959)*

* Pohlner, J. et al. (1987). Nature 325, pp. 458-462, which is hereby incorporated by reference thereto.

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

Botulinum type A neurotoxin	- amino acid residues (M1-K448)
Botulinum type B neurotoxin	- amino acid residues (M1-K441)
Botulinum type C neurotoxin	- amino acid residues (M1-K449)
Botulinum type D neurotoxin	- amino acid residues (M1-R445)

Botulinum type E neurotoxin	- amino acid residues (M1-R422)
Botulinum type F neurotoxin	- amino acid residues (M1-K439)
Botulinum type G neurotoxin	- amino acid residues (M1-K446)
Tetanus neurotoxin	- amino acid residues (M1-A457)

A variety of clostridial toxin fragments comprising the light chain can be useful in aspects of the present invention with the proviso that these light chain fragments can specifically target the core components of the neurotransmitter release apparatus and thus participate in executing the overall cellular mechanism whereby a clostridial toxin proteolytically cleaves a substrate. The light chains of clostridial toxins are approximately 420-460 amino acids in length and comprise an enzymatic domain. Research has shown that the entire length of a clostridial toxin light chain is not necessary for the enzymatic activity of the enzymatic domain. As a non-limiting example, the first eight amino acids of the BoNT/A light chain are not required for enzymatic activity. As another non-limiting example, the first eight amino acids of the TeNT light chain are not required for enzymatic activity. Likewise, the carboxyl-terminus of the light chain is not necessary for activity. As a non-limiting example, the last 32 amino acids of the BoNT/A light chain (residues 417-448) are not required for enzymatic activity. As another non-limiting example, the last 31 amino acids of the TeNT light chain (residues 427-457) are not required for enzymatic activity. Thus, aspects of this embodiment can include clostridial toxin light chains comprising an enzymatic domain having a length of, for example, at least 350 amino acids, at least 375 amino acids, at least 400 amino acids, at least 425 amino acids and at least 450 amino acids. Other aspects of this embodiment can include clostridial toxin light chains comprising an enzymatic domain having a length of, for example, at most 350 amino acids, at most 375 amino acids, at most 400 amino acids, at most 425 amino acids and at most 450 amino acids.

The polypeptides of the present invention, especially the protease component thereof, may be PEGylated – this may help to increase stability, for example

duration of action of the protease component. PEGylation is particularly preferred when the protease comprises a BoNT/A, B or C_1 protease. PEGylation preferably includes the addition of PEG to the N-terminus of the protease component. By way of example, the N-terminus of a protease may be extended with one or more amino acid (e.g. cysteine) residues, which may be the same or different. One or more of said amino acid residues may have its own PEG molecule attached (e.g. covalently attached) thereto. An example of this technology is described in WO2007/104567, which is incorporated in its entirety by reference thereto.

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

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

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

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

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

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

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

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

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

Examples of suitable (reference) Translocation Domains include:

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

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

Botulinum type A neurotoxin	- amino acid residues (A449-K871)
Botulinum type B neurotoxin	- amino acid residues (A442-S858)
Botulinum type C neurotoxin	- amino acid residues (T450-N866)
Botulinum type D neurotoxin	- amino acid residues (D446-N862)

Botulinum type E neurotoxin	- amino acid residues (K423-K845)
Botulinum type F neurotoxin	- amino acid residues (A440-K864)
Botulinum type G neurotoxin	- amino acid residues (S447-S863)
Tetanus neurotoxin	- amino acid residues (S458-V879)

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

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

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

Alternatively, the Translocation Domain may be of a non-clostridial origin. Examples of non-clostridial (reference) Translocation Domain origins include, but not be restricted to, the translocation domain of diphtheria toxin [O=Keefe *et al.,* Proc. Natl. Acad. Sci. USA (1992) 89, 6202-6206; Silverman *et al.,* J. Biol. Chem. (1993) 269, 22524-22532; and London, E. (1992) *Biochem. Biophys. Acta., 1112, pp.25-51*], the translocation domain of *Pseudomonas* exotoxin type A [Prior *et al.* Biochemistry (1992) 31, 3555-3559], the translocation domains of anthrax toxin [Blanke *et al.* Proc. Natl. Acad. Sci. USA (1996) 93, 8437-8442], a variety of fusogenic or hydrophobic peptides of translocating function [Plank *et al.* J. Biol. Chem. (1994) 269, 12918-12924; and Wagner *et al* (1992) *PNAS, 89, pp.7934-7938*], and amphiphilic peptides [Murata *et al* (1992) *Biochem., 31, pp.1986-1992*]. The Translocation Domain may mirror the Translocation Domain present in a naturally-occurring protein, or may include amino acid variations so long as the variations do not destroy the translocating ability of the Translocation Domain.

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

Use of the (reference) Translocation Domains listed in Table (below) includes

use of sequence variants thereof. A variant may comprise one or more conservative nucleic acid substitutions and/ or nucleic acid deletions or insertions, with the proviso that the variant possesses the requisite translocating function. A variant may also comprise one or more amino acid substitutions and/ or amino acid deletions or insertions, so long as the variant possesses the requisite translocating function.

Translocation Domain source	**Amino acid residues**	**References**
Diphtheria toxin	194-380	**Silverman *et al.***, 1994, J. Biol. Chem. 269, 22524-22532 **London E.**, 1992, Biochem. Biophys. Acta., 1113, 25-51
Domain II of pseudomonas exotoxin	405-613	**Prior *et al.***, 1992, Biochemistry 31, 3555-3559 **Kihara & Pastan**, 1994, Bioconj Chem. 5, 532-538
Influenza virus haemagglutinin	GLFGAIAGFIENGWE GMIDGWYG, and Variants thereof	**Plank *et al.***, 1994, J. Biol. Chem. 269, 12918-12924 **Wagner *et al.***, 1992, PNAS, 89, 7934-7938 **Murata *et al.***, 1992, Biochemistry 31, 1986-1992
Semliki Forest virus fusogenic protein	**Translocation domain**	**Kielian *et al.***, 1996, J Cell Biol. 134(4), 863-872
Vesicular Stomatitis virus glycoprotein G	118-139	**Yao *et al.***, 2003, Virology 310(2), 319-332
SER virus F protein	Translocation domain	**Seth *et al.***, 2003, J Virol 77(11) 6520-6527
Foamy virus envelope glycoprotein	Translocation domain	**Picard-Maureau *et al.***, 2003, J Virol. 77(8), 4722-4730

The polypeptides of the present invention may further comprise a translocation facilitating domain. Said domain facilitates delivery of the non-cytotoxic protease into the cytosol of the target cell and are described, for example, in WO 08/008803 and WO 08/008805, each of which is herein incorporated by reference thereto.

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

By way of further example, a translocation facilitating domain may comprise a Clostridial toxin H_{CN} domain or a fragment or variant thereof. In more detail, a Clostridial toxin H_{CN} translocation facilitating domain may have a length of at least 200 amino acids, at least 225 amino acids, at least 250 amino acids, at least 275 amino acids. In this regard, a Clostridial toxin H_{CN} translocation facilitating domain preferably has a length of at most 200 amino acids, at most 225 amino acids, at most 250 amino acids, or at most 275 amino acids.

Specific (reference) examples include:

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

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

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

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

Botulinum type A neurotoxin	- amino acid residues (449-1110)

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

Sequence homology:

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

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

Alignment scores for determining sequence identity

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

The percent identity is then calculated as:

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

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

Conservative amino acid substitutions

Basic:	arginine
	lysine
	histidine
Acidic:	glutamic acid
	aspartic acid
Polar:	glutamine
	asparagine
Hydrophobic:	leucine
	isoleucine
	valine
Aromatic:	phenylalanine
	tryptophan

	tyrosine
Small:	glycine
	alanine
	serine
	threonine
	methionine

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

Non-naturally occurring amino acids include, without limitation, trans-3-methylproline, 2,4-methano-proline, cis-4-hydroxyproline, trans-4-hydroxy-proline, N-methylglycine, allo-threonine, methyl-threonine, hydroxy-ethylcysteine, hydroxyethylhomo-cysteine, nitro-glutamine, homoglutamine, pipecolic acid, tert-leucine, norvaline, 2-azaphenylalanine, 3-azaphenyl-alanine, 4-azaphenyl-alanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an *in vitro* system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., J. Am. Chem. Soc. 113:2722, 1991; Ellman et al., Methods Enzymol. 202:301, 1991; Chung et al., Science 259:806-9, 1993; and Chung et al., Proc. Natl. Acad. Sci.

USA 90:10145-9, 1993). In a second method, translation is carried out in Xenopus oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., J. Biol. Chem. 271:19991-8, 1996). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the polypeptide in place of its natural counterpart. See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by *in vitro* chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

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

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244: 1081-5, 1989). Sites of biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., Science 255:306-12, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of essential amino acids can also be inferred from analysis of homologies with related components (e.g. the translocation or protease components) of the polypeptides of the present invention.

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

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

There now follows a brief description of the Figures, which illustrate aspects and/ or embodiments of the present invention.

Fig. 1 - illustrates the surprising neurotoxin activity retained by a modified clostridial neurotoxin (LH_N). Said modified neurotoxin lacks a functional H_C binding domain and is therefore equivalent to the modified clostridial neurotoxins described by LIN, *et al.* (WO02/044199). In contrast, no neurotoxin activity was detected for a modified clostridial neurotoxin (LC/C), which lacks a

function H_N translocation domain.

Fig. 2 - illustrates a simply amino acid sequence homology alignment for the various BoNT serotypes. From this alignment, amino acid residues or regions from one serotype (eg. from serotype A) may be compared with corresponding residues/ regions across the serotypes by way of simple vertical alignment.

Fig. 3 - SDS-PAGE analysis of the purification of a L(#FXa)HC-EGF chimaeric protein. Lane 1 illustrates the clarified cell lysate; Lane 2 illustrates the column flow through; Lane 3 illustrates the fraction eluted following washing the column; Lanes 4, 6-12 are fractions eluted on addition of 250mM imidazole. Lane 5 is molecular mass markers (Benchmark)

Fig. 4 - SDS-PAGE analysis of the proteolysis of a L(#FXa)HC-EGF chimaeric protein by FXa. Lanes 1 & 2 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 3&4 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lane 5 is molecular mass markers (benchmark); Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Factor Xa is clearly seen in Lanes 6 and 8.

Fig. 5 - SDS-PAGE analysis of the purification of a LH(#FXa)C-EGF chimaeric protein (as prepared in Example 20). Lane 1 illustrates the clarified cell lysate; Lane 2 is molecular mass markers (Benchmark); Lane 3 illustrates the column flow through; Lane 4 illustrates the fraction eluted following washing the column; Lanes 5-12 are fractions eluted on addition of 250mM imidazole.

Fig. 6 - SDS-PAGE analysis of the purification of a LH(#FXa)C-EGF chimaeric

protein (as prepared in Example 21). Lane 1 illustrates the clarified cell lysate; Lane 2 illustrates the column flow through; Lane 3 is molecular mass markers (Benchmark); Lane 4 illustrates the fraction eluted following washing the column; Lanes 5-12 are fractions eluted on addition of 250mM imidazole.

Fig. 7 - SDS-PAGE analysis of the proteolysis of a LH(#FXa)C-EGF chimaeric protein (as prepared in Example 20) by FXa. Lanes 1 & 3 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 4 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Factor Xa is clearly seen in Lanes 7 and 9. Lane 2 is molecular mass markers (Benchmark).

Fig. 8 - Western blot analysis of the proteolysis of a LH(#FXa)C-EGF chimaeric protein (as prepared in Example 20) by FXa. Lanes 1 & 3 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 4 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Factor Xa is clearly seen in Lanes 6, 7, 8 & 9 by the visualisation of a Histidine immunoreactive band at the anticipated size. Lane 2 is molecular mass markers suitable for detection by staining (Benchmark). Lane 10 is molecular mass markers suitable for Western blot visualisation (Magic Markers).

Fig. 9 - SDS-PAGE analysis of the proteolysis of a LH(#FXa)C-EGF chimaeric protein (as prepared in Example 21) by FXa. Lanes 1 & 2 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 4 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Factor Xa is clearly seen in Lanes 7 and 9. Lane 3 is molecular mass markers (Benchmark).

Fig. 10 - SDS-PAGE analysis of the purification of a L(#Thr)HC-EGF chimaeric protein (as prepared in Example 19). Lane 1 is molecular mass markers (Benchmark); Lane 2 illustrates the clarified cell lysate; Lane 3 illustrates the column flow through; Lane 4 illustrates the fraction eluted following washing the column; Lanes 5-12 are fractions eluted on addition of 250mM imidazole.

Fig. 11 - SDS-PAGE analysis of the proteolysis of a L(#Thr)HC-EGF chimaeric protein (as prepared in Example 19) by Thrombin. Lane 1 is molecular mass markers (Benchmark). Lanes 2 & 3 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 4 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Thrombin is

clearly seen in Lanes 6, 7, 8 and 9.

Fig. 12 - Western blot analysis of the proteolysis of a L(#Thr)HC-EGF chimaeric protein (as prepared in Example 19) by Thrombin. Lane 1 is molecular mass markers (Benchmark), which are poorly visible by Western blotting. Lanes 2 & 3 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 4 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Factor Xa, in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of FXa treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Thrombin to release an ~85kDa fragment that retains the EGF domain but lacks ~20Kda of the N-terminus of the LC is clearly seen in Lanes 8 and 9.

Fig. 13 - SDS-PAGE analysis of the purification of a L(#Thr)HA-EGF chimaeric protein (as prepared in Example 24). Lane 1 illustrates the clarified cell lysate; Lane 2 illustrates the column flow through; Lane 3 illustrates the fraction eluted following washing the column; Lane 4 is molecular mass markers (Benchmark); Lanes 5-11 are fractions eluted on addition of 250mM imidazole

Fig. 14 - SDS-PAGE analysis of the proteolysis of a L(#Thr)HA-EGF chimaeric protein (as prepared in Example 24) by Thrombin. Lane 4 is molecular mass markers (Benchmark). Lanes 1 & 2 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 3 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Thrombin in the absence and presence of DTT respectively; Lanes

8 & 9 illustrate the result of Thrombin treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Thrombin is clearly seen in Lanes 6, 7, 8 and 9

Fig. 15 - Western blot analysis of the proteolysis of a L(#Thr)HA-EGF chimaeric protein (as prepared in Example 24) by Thrombin. Lane 4 is molecular mass markers (Benchmark). Lanes 1 & 2 illustrate the purified protein prior to enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 3 & 5 illustrate the protein post enterokinase activation at the LC-H_N junction, in the absence and presence of DTT respectively; Lanes 6 & 7 illustrate the enterokinase activated protein treated with Thrombin in the absence and presence of DTT respectively; Lanes 8 & 9 illustrate the result of Thrombin treatment of the protein that has not been activated with enterokinase, in the absence and presence of DTT respectively. Fragmentation of the protein following treatment with Thrombin is clearly seen in Lanes 7 and 9.

Figure 16 - illustrates the result obtained following exposure of SCN with FXa-treated L(FXa)HC-EGF compared to untreated L(FXa)HC-EGF. The protein that had been treated with Factor Xa is clearly less effective at cleaving Syntaxin than the protein that was not treated with FXa. The invention has therefore enabled a reduction in the efficacy of the modified protein.

Fig. 17 - illustrates the result obtained following exposure of SCN with 10 nM Thrombin-treated L(Thr)HA-EGF compared to 10 nM untreated L(Thr)HA-EGF. The protein that had been treated with thrombin is clearly less effective at cleaving SNAP-25 than the protein that was not treated with thrombin. The invention has therefore enabled a reduction in the efficacy of the modified protein.

As used herein, except where the context requires otherwise the term 'comprise' and variations of the term, such as 'comprising', 'comprises' and 'comprised', are not intended to exclude other additives, components, integers or steps.

Reference to any prior art in the specification is not an acknowledgment or suggestion that this prior art forms part of the common general knowledge in any jurisdiction or that this prior art could reasonably be expected to be understood, regarded as relevant, and/or combined with other pieces of prior art by a person skilled in the art.

There now follows description of specific embodiments of the invention, illustrated by the Examples.

Example 1 – Assessment of polypeptides of the invention when exposed to a mammalian cell (muscle).

Example 2 – Assessment of polypeptides of the invention when exposed to a mammalian cell having first exposed the polypeptide to circulatory proteases.

Example 3 – Assessment of the catalytic activity of polypeptides of the invention.

Example 4 – Assessment of the translocation ability of polypeptides of the invention.

Example 5 – Creation of an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC.

Example 6 – Purification of an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC.

Example 7 – Demonstration of enhanced protease sensitivity in an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC.

Example 8 – Creation of an LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC.

Example 9 – Creation of an LHA-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC.

Example 10 – Creation of an LHC-EGF chimaeric protein that incorporates a furin recognition site into the LC.

Example 11 – Creation of an LHA-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain.

Example 12 – Creation of a LHA-EGF chimaeric protein that incorporates an ADAM17 recognition site into the LC domain.

Example 13 – Creation of a recombinant BoNT/A protein that incorporates an ADAM17 recognition site into the LC

Example 14 – Creation of a recombinant BoNT/A protein that incorporates a

furin recognition site into the H_N.

Example 15 - Treatment of a patient suffering from dystonia (Spasmodic Torticollis).

Example 16 - Treatment of a patient suffering from blepharospasm.

Example 17 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC at position 210

Example 18 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC at position 195

Example 19 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC at position 210

Example 20 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 742 of the H_N

Example 21 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 750 of the H_N

Example 22 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the H_N domain at position 750 of the H_N

Example 23 – Creation of a LHD-VIPr chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 798 of the H_N

Example 24 – Creation of an LHA-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC domain

Example 25 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Factor Xa recognition site into the LC.

Example 26 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Factor Xa recognition site into the H_N.

Example 27 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Thrombin recognition site into the LC

Example 28 – Demonstration of specific cleavage of a purified LHA-EGF chimaeric protein that is engineered to incorporate a Thrombin recognition site into the LC

Example 29 – Demonstration of reduced *in vitro* cellular activity of a protein engineered to incorporate a Factor Xa protease cleavage site into the LC domain of L(FXa)HC-EGF

Example 30 – Demonstration of reduced *in vitro* cellular activity of a protein engineered to incorporate a Thrombin protease cleavage site into the LC domain of L(Thr)HA-EGF

Example 31 – Creation of a recombinant BoNT/A protein that incorporates a Thrombin recognition site into the LC

Example 32 – Creation of a recombinant BoNT/A protein that incorporates a Factor Xa recognition site into the LC.

Example 33 – Creation of a recombinant BoNT/A protein that incorporates a Factor Xa recognition site into the H_N

Example 34 – Creation of a recombinant BoNT/E protein that incorporates a Thrombin recognition site into the LC

Example 35 – Creation of a recombinant BoNT/E protein that incorporates a Factor Xa recognition site into the H_N.

Example 36 – Creation of an LHE-VIPr chimaeric protein that incorporates a Thrombin recognition site into the LC.

Example 37 – Creation of an LHE-VIPr chimaeric protein that incorporates a Factor Xa recognition site into the H_N.

Example 38 – Creation of an LHE-VIPr chimaeric protein that incorporates a Factor Xa recognition site into the LC.

Example 39 – Cleavage of SNARE protein by a modified clostridial neurotoxin (LH_N) having the properties described by LIN, *et al.* (WO02/044199)

Summary of SEQ ID NOs

SEQ ID 1	DNA sequence of LHC-EGF
SEQ ID 2	Protein sequence of LHC-EGF
SEQ ID 3	DNA sequence of L(#FXa)HC-EGF
SEQ ID 4	Protein sequence of L(#FXa)HC-EGF
SEQ ID 5	DNA sequence of L(#Thr)HC-EGF
SEQ ID 6	Protein sequence of L(#Thr)HC-EGF

SEQ ID 7	DNA sequence of LHA-EGF	
SEQ ID 8	Protein sequence of LHA-EGF	
SEQ ID 9	DNA sequence of L(#Thr)HA-EGF	
SEQ ID 10	Protein sequence of L(#Thr)HA-EGF	
SEQ ID 11	Protein sequence of L(#furin)HC-EGF	
SEQ ID 12	DNA sequence of LH(#FXa)A-EGF	
SEQ ID 13	Protein sequence of LH(#FXa)A-EGF	
SEQ ID 14	DNA sequence of L(#ADAM17)HA-EGF	
SEQ ID 15	Protein sequence of L(#ADAM17)HA-EGF	
SEQ ID 16	DNA sequence of LHA-H_c/A	
SEQ ID 17	Protein sequence of LHA-H_c/A	
SEQ ID 18	DNA sequence of L(#ADAM17)HA-H_c/A	
SEQ ID 19	Protein sequence of L(#ADAM17)HA-H_c/A	
SEQ ID 20	DNA sequence of L(#furin)HA-H_c/A	
SEQ ID 21	Protein sequence of L(#furin)HA-H_c/A	
SEQ ID 22	DNA sequence of L(#FXa)HC-EGF	(SXN1975)
SEQ ID 23	Protein sequence of L(#FXa)HC-EGF	(SXN1975)
SEQ ID 24	Protein sequence of L(#Thr)HC-EGF	(SXN1931)
SEQ ID 25	Protein sequence of L(#Thr)HC-EGF	(SXN1932)
SEQ ID 26	Protein sequence of LH(#FXa)C-EGF	(SXN1937)
SEQ ID 27	Protein sequence of LH(#FXa)C-EGF	(SXN1938)
SEQ ID 28	Protein sequence of LH(#Thr)C-EGF	(SXN1939)
SEQ ID 29	Protein sequence of LH(#FXa)D-VIPr	(SXN1930)
SEQ ID 30	Protein sequence of L(#Thr)HA-EGF	(SXN1974)
SEQ ID 31	Protein sequence of L(#Thr)HA-EGF	(SXN1974)
SEQ ID 32	Protein sequence of L(#Thr)HA-H_c/A	
SEQ ID 33	Protein sequence of L(#FXa)HA-H_c/A	
SEQ ID 34	Protein sequence of LH(FXa)A-H_c/A	
SEQ ID 35	Protein sequence of L(#Thr)HE-H_c/E	
SEQ ID 36	Protein sequence of LH(#FXa)E-H_c/E	
SEQ ID 37	Protein sequence of L(#Thr)HE-VIPr	
SEQ ID 38	Protein sequence of LH(#FXa)E-VIPr	

SEQ ID 39 Protein sequence of L(#FXa)HE-VIPr (mutation at K228D)

Examples

Example 1 – Assessment of polypeptides of the invention when exposed to a mammalian muscle cell

A purified protein created according to Example 13 is incubated in the presence of a mammalian muscle cell (coronary smooth muscle primary culture or HSkMC (150-05f) cell (available from ECACC)). In parallel, a second polypeptide (identical to the first polypeptide other than for the fact that it lacks the same destructive cleavage site) is incubated under identical conditions in the presence of the same test cell-type.

Each of the two polypeptides is then assessed for cleavage by ADAM17 (inherent to the coronary smooth muscle primary culture/ HSkMC cell) by SDS-PAGE and subsequent Western blot analysis. In this regard, a greater observed cleavage for the first polypeptide versus that observed for the second polypeptide confirms controllable inactivation of the present invention.

Example 2 – Assessment of polypeptides of the invention when exposed to a mammalian cell having first exposed the polypeptide to a circulatory protease

A first polypeptide (SEQ ID 4); prepared according to Example 5 of the present invention) is taken and incubated in the presence of a target cell having first exposed the polypeptide to circulatory proteases (for example, Factor Xa, Thrombin) *in vitro*. In parallel, a second polypeptide (SEQ ID2; identical to the first polypeptide other than for the fact that it lacks the protease cleavage site) is incubated in the same manner as for the first polypeptide.

Each of the two polypeptides is then assessed for cleavage of syntaxin in an embryonic spinal cord neuron (eSCN). In this regard, a lesser observed cleavage for the first polypeptide versus that observed for the second polypeptide confirms controllable inactivation of the present invention.

Example 3 – Assessment of the catalytic activity of polypeptides of the invention

A first polypeptide (SEQ ID 10; prepared according to Example 9 of the present invention) is incubated *in vitro* in the presence of a protease (thrombin) that cleaves the polypeptide at a destructive cleavage site introduced into the protease domain of the polypeptide. In parallel, a second polypeptide (SEQ ID 8: identical to the first polypeptide other than for the fact that it lacks the protease cleavage site) is incubated in an identical manner in the presence of the same protease.

Each of the two polypeptides is then challenged in an *in vitro* cell-free system (as described by Hallis *et al* 1996, J. Clin. Microbiol. 34 1934–1938) containing immobilised SNAP-25, and cleavage of SNAP-25 protein is measured by using specific antisera raised to the cleavage product. In this regard, a lesser observed SNARE protein cleavage for the first polypeptide versus that observed for the second polypeptide confirms controllable inactivation of the present invention.

Example 4 – Assessment of the translocation ability of polypeptides of the invention

A first polypeptide (according to the present invention) is incubated in the presence of a protease that cleaves the polypeptide at a destructive cleavage site introduced into the translocation (e.g. H_N) domain. In parallel, a second polypeptide (identical to the first polypeptide other than for the fact that it lacks the protease cleavage site) is incubated in an identical manner in the presence of the same protease.

Each of the two polypeptides is then challenged in an *in vitro* system containing a lipid bilayer membrane, and transport across the membrane is measured. For example, Shone C. (1987) describes an in vitro assay employing liposomes, which are challenged with a test molecule. Presence of the

requisite translocation function is confirmed by release from the liposomes of K+ and/or labelled NAD, which may be readily monitored [see Shone C. (1987) Eur. J. Biochem; vol. 167(1): pp. 175-180]. A further example is provided by Blaustein R. (1987), which describes a simple in vitro assay employing planar phospholipid bilayer membranes. The membranes are challenged with a test molecule and the requisite translocation function is confirmed by an increase in conductance across said membranes [see Blaustein (1987) FEBS Letts; vol. 226, no. 1: pp. 115-120].

This method is applied to study the protease inactivation of the H_N domain of serotype D BoNT. The protein of Example 23 is expressed and purified and is exposed to Factor Xa to result in cleavage of the protein within the H_N domain. The cleaved protein is assessed in the *in vitro* system described above and compared to the protein that has not been treated with Factor Xa. The experiment determines that the transport across the membrane for the Factor Xa-treated polypeptide is significantly less than that of the untreated polypeptide.

Example 5 – Creation of an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC.

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR). Simple text character analysis of the primary sequence identified the sequence $_{210}$GEGR$_{213}$ within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located on the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codon for G210 (GGC) to one that encodes Ile (ATC) was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage was assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA was incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA was checked by sequencing. The final ORF incorporating the Factor Xa site is illustrated as SEQ ID 3 and the amino acid sequence of the expression product is illustrated in SEQ ID 4.

Example 6 – Purification of an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC

The ORF created in Example 17 was cloned into an *E. coli* expression vector (a pET (Novagen) vector that has been modified to ensure mobilisation deficiency) and transformed into an *E. coli* host strain, most commonly BL21. The vector was modified to include expression of a Histidine tag at the N-terminus of the LHC-EGF ORF.

Expression of the LHC-EGF fusion protein is achieved using the following protocol. Inoculate 100 ml of modified TB containing 0.2% glucose and 100 μg/ml ampicillin in a 250 ml flask with a single colony from the LHC-EGF expression strain. Grow the culture at 37°C, 225 rpm for 16 hours. Inoculate 1L of modified TB containing 0.2% glucose and 100 μg/ml ampicillin in a 2L flask with 10ml of overnight culture. Grow cultures at 37°C until an approximate OD600nm of 0.5 is reached at which point reduce the temperature to 16°C. After 1 hour induce the cultures with 1 mM IPTG and grow at 16°C for a further

16 hours.

Purification of the LHC-EGF fusion is achieved by affinity chromatography. In detail, a falcon tube containing 25 ml 50 mM HEPES pH 7.2 200 mM NaCl and approximately 10 g of E. coli BL21 cell paste is defrosted. Sonicate the cell paste on ice 30 seconds on, 30 seconds off for 10 cycles at a power of 22 microns ensuring the sample remains cool. Spin the lysed cells at 18 000 rpm, 4°C for 30 minutes. Load the supernatant onto a 0.1 M NiSO4 charged Chelating column (20-30 ml column is sufficient) equilibrated with 50 mM HEPES pH 7.2 200 mM NaCl. Using a step gradient of 10 and 40 mM imidazole, wash away the non-specific bound protein and elute the fusion protein with 100 mM imidazole. Dialyse the eluted fusion protein against 5L of 50 mM HEPES pH 7.2 200 mM NaCl at 4°C overnight and measure the OD of the dialysed fusion protein. Add 6.4ng enterokinase/mg fusion protein and incubate at 25°C static overnight. Load onto a 0.1 M NiSO4 charged Chelating column (20-30 ml column is sufficient) equilibrated with 50 mM HEPES pH 7.2 200 mM NaCl. Wash column to baseline with 50 mM HEPES pH 7.2 200 mM NaCl. Using a step gradient of 10 and 40 mM imidazole, wash away the non-specific bound protein and elute the fusion protein with 100 mM imidazole. Dialyse the eluted fusion protein against 5L of 50 mM HEPES pH 7.2 200 mM NaCl at 4°C overnight and concentrate the fusion to about 2 mg/ml, aliquot sample and freeze at -20°C. Test purified protein using OD, BCA and purity analysis.

Example 7 – Demonstration of enhanced protease sensitivity in an LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC

The purified chimaeric protein of Example 6 is assessed for its stability in the presence of protease using the methodology outlines in Example 2 and 3. In summary, the LHC-EGF chimaeric protein is exposed to a range of concentrations of Factor Xa protease (obtained, for example, from New England Biolabs #P8010L) *in vitro* over a period of 1-120 minutes. The

proteolysis is terminated by addition of a specific inhibitor of Factor Xa (for example Dansyl-glu-gly-arg-chloromethyl ketone (CALBIOCHEM, #251700)). A control protein chimaera of LHC-EGF that does not include the additional Factor Xa site is used to compare the effect of the protease on LC activity (using Example 3), and functionality of the chimaera when exposed to a target cell (using Example 2 and measuring syntaxin cleavage in an embryonic spinal cord neuron (eSCN)).

Example 8 – Creation of an LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin (LVPRGS). Simple text character analysis of the primary sequence identified the sequence $_{194}$ISPRFM$_{199}$ within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located near the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons for S_{195} to Val (TCT to GTT) and M_{195} to Ser (ATG to TCC) changes the region $_{194}$ISPRFM$_{199}$ to IVPRFS to make it a substrate for Thrombin cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference

to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is illustrated as SEQ ID 5 and the amino acid sequence of the expression product is illustrated in SEQ ID 6.

Example 9 – Creation of an LHA-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/A and the EGF sequence (SEQ ID 8) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin (GRG). Simple text character analysis of the primary sequence identified the sequence $_{103}GRM_{105}$ within the LC domain. The location of the peptide in the tertiary structure of the LC/A is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 7 (encoding the ORF of SEQ ID 8) using a primer designed to switch the codon for Met105 (ATG) to one that encodes Gly (GGT) was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example

pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is illustrated as SEQ ID 9 and the amino acid sequence of the expression product is illustrated in SEQ ID 10.

Example 10 – Creation of an LHC-EGF chimaeric protein that incorporates a furin recognition site into the LC

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for furin (RXR▼K/R). Simple text character analysis of the primary sequence identified the sequence $_{210}GEGR_{213}$ within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located on the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the peptide region from GEGR to RSRR was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the furin site is illustrated as SEQ ID 11 and the amino acid

sequence of the expression product is illustrated in SEQ ID 12.

Example 11 – Creation of an LHA-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/A and the human epidermal growth factor sequence (SEQ ID 8) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR). Simple text character analysis of the primary sequence identified the sequence $_{562}$GKSR$_{565}$ within the H_N domain. The location of the peptide in the tertiary structure of the H_N/A is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 7 (encoding SEQ ID 8) using a primer designed to switch the peptide region from GKSR to IEGR was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is illustrated as SEQ ID 13 and the amino acid sequence of the expression product is illustrated in SEQ ID 14.

Example 12 – Creation of a LHA-EGF chimaeric protein that incorporates an ADAM17 recognition site into the LC domain

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/A and the human epidermal growth factor sequence (SEQ ID 8) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for ADAM17 (PLAQAVRSSS). Simple text character analysis of the primary sequence identifies a region of the structure ($_{206}$PLLGAGKFAT$_{215}$ within the LC domain) that is amenable to protein engineering. The location of the peptide in the tertiary structure of the LC is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 7 (which encodes SEQ ID 8) was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). The mutagenesis of the LC was performed to modify the coding region from $_{206}$PLLGAGKFAT$_{215}$ to PLAQAVRSSS.

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the additional ADAM17 sites is illustrated as SEQ ID 15 and the amino acid sequence of the expression product is illustrated in SEQ ID 16.

Example 13 – Creation of a recombinant BoNT/A protein that incorporates an ADAM17 recognition site into the LC domain

The primary sequence of a recombinant endopeptidase active BoNT/A containing an engineered activation protease site specific for enterokinase (SEQ ID 18) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for ADAM17 (PLAQAVRSSS). Simple text character analysis of the primary sequence identifies a region of the BoNT structure ($_{206}$PLLGAGKFAT$_{215}$ within the LC domain) that is amenable to protein engineering. The location of the peptide in the tertiary structure of the LC is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 17 (which encodes SEQ ID 18) was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). The mutagenesis of the LC was performed to modify the coding region from $_{206}$PLLGAGKFAT$_{215}$ to PLAQAVRSSS.

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the additional ADAM17 sites is illustrated as SEQ ID 19 and the amino acid sequence of the expression product is illustrated in SEQ ID 20.

Example 14 – Creation of a recombinant BoNT/A protein that incorporates a furin recognition site into the H_N

The primary sequence of a recombinant endopeptidase active BoNT/A containing an engineered activation protease site specific for enterokinase (SEQ ID 18) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for furin (RXR▼K/R). Simple text character analysis of the primary sequence identified the sequence $_{563}$KSR$_{565}$ within the H_N domain that is amenable to protein engineering. . The location of the peptide in the tertiary structure of the H_N domain is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 17 (which encodes SEQ ID 18) using a primer designed to switch the codon for K_{563} (AAA) to Arg (CGT) and to insert an Arg (CGC) after the existing R_{565} changes the sequence $_{563}$KSR$_{565}$ to RSRR which is a substrate for cleavage by furin. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the additional ADAM17 sites is illustrated as SEQ ID 21 and the amino acid sequence of the expression product is illustrated in SEQ ID 22.

Example 15 - Treatment of a patient suffering from dystonia (Spasmodic

Torticollis)

A male, suffering from spasmodic torticollis, as manifested by spasmodic or tonic contractions of the neck musculature, producing stereotyped abnormal deviations of the head, the chin being rotated to one side, and the shoulder being elevated toward the side at which the head is rotated, had previously been treated with a therapeutically effective amount of BoNT/A into the neck muscles for torticollis, but developed dysphagia because of dispersal of the protease into the oropharynx. The patient is subsequently treated by injection with up to about 300 units, or more, of polypeptide of the present invention (eg. a botulinum toxin type A neurotoxin modified to include a Factor Xa protease sensitive site), in the dystonic neck muscles. After 3-7 days the symptoms are substantially alleviated and the patient is able to hold his head and shoulder in a normal position for at least 3 months. Following the treatment with the modified neurotoxin the patient does not experience any dysphagia. By utilising the modified botulinum toxin type A, the physician is able to inject more product into the area requiring therapy without fear of an increase in side effects. Enhanced dose leads to enhanced duration of action and therefore improved therapy.

Example 16 - Treatment of a patient suffering from blepharospasm

A 58 year old female with blepharospasm is treated by injecting between about 1 to about 5 units of a polypeptide of the present invention (eg. a botulinum toxin type A neurotoxin modified to include a ADAM17 protease sensitive site, as described in Example 13) into the lateral pre-tarsal orbicularis oculi muscle of the upper lid and the lateral pre-tarsal orbicularis oculi of the lower lid, the amount injected varying based upon both the size of the muscle to be injected and the extent of muscle paralysis desired. Alleviation of the blepharospasm occurs in about 1 to about 7 days. By utilising the modified botulinum toxin type A, the physician is able to inject more product into the area requiring therapy without fear of an increase in side effects. Enhanced dose leads to enhanced duration of action and therefore improved therapy.

Example 17 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the LC at position 210 [SXN101975]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR). A site for insertion of a Factor Xa site is identified in the primary sequence $_{210}$GEGR within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located near the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) was achieved using a primer designed to switch the codons for $_{210}$G to I to make it a substrate for Factor Xa cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is illustrated as SEQ ID 23 and the amino acid sequence of the expression product is illustrated in SEQ ID 24.

Example 18 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC at position 195 [SXN101931]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin. A site for insertion of a Thrombin site is identified in the primary sequence $_{194}$ISPRFM$_{199}$ within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located near the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons for S_{195} to Val (TCT to GTT) and M_{195} to Ser (ATG to TCC) changes the region $_{194}$ISPRFM$_{199}$ to IVPRFS to make it a substrate for Thrombin cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 25.

Example 19 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC at position 210 [SXN101932]

The primary sequence of a chimaeric protein constructed by a genetic fusion of

the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin. A site for insertion of a Thrombin site is identified in the primary sequence $_{210}$GEGRFS within the LC domain. The location of the peptide in the tertiary structure of the LC/C is predicted from examination of the location of the homologous peptide sequence in the LC/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC/A equivalent peptide sequence is located near the surface of the LC. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons $_{211}$EGR to TPR to create a sequence GTPRFS which is a substrate for Thrombin cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 26.

Example 20 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 742 of the H_N [SXN101937]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence

(SEQ ID 2) is reviewed and a site for insertion of a Factor Xa site is identified in the primary sequence $_{742}$IDLE$_{755}$ within the H_N domain. The location of the peptide in the tertiary structure of the H_N/C is predicted from examination of the location of the homologous peptide sequence in the H_N/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N/A equivalent peptide sequence is located near the surface of the H_N. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons for $_{742}$LE to GR to make it a substrate for Factor Xa cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 27.

Example 21 – Creation of a LHC-EGF chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 750 of the H_N [SXN101938]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed and a site for insertion of a Factor Xa site is identified in the primary sequence $_{750}$SGSD$_{753}$ within the H_N domain. The location of the

peptide in the tertiary structure of the H_N/C is predicted from examination of the location of the homologous peptide sequence in the H_N/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N/A equivalent peptide sequence is located near the surface of the H_N. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons for $_{750}$SGSD to IDGR make it a substrate for Factor Xa cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 28.

Example 22 – Creation of a LHC-EGF chimaeric protein that incorporates a Thrombin recognition site into the H_N domain at position 750 of the H_N [SXN101939]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/C and the human epidermal growth factor sequence (SEQ ID 2) is reviewed and a site for insertion of a Thrombin site is identified in the primary sequence $_{750}$SGSD$_{753}$ within the H_N domain. The location of the peptide in the tertiary structure of the H_N/C is predicted from examination of the location of the homologous peptide sequence in the H_N/A using the X-ray

crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N/A equivalent peptide sequence is located near the surface of the H_N. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 1 (encoding SEQ ID 2) using a primer designed to switch the codons for SGSD to GVPR to make it a substrate for Thrombin cleavage. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 29.

Example 23 – Creation of a LHD-VIPr chimaeric protein that incorporates a Factor Xa recognition site into the H_N domain at position 798 of the H_N [SXN101930]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/D and an analogue of the human vasoactive intestinal peptide (VIPr) is reviewed and a site for insertion of a Factor Xa site is identified in the primary sequence $_{798}$SGSD within the H_N domain. The location of the peptide in the tertiary structure of the H_N/D is predicted from examination of the location of the homologous peptide sequence in the H_N/A using the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol

(http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N/A equivalent peptide sequence is located near the surface of the H_N. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the gene using a primer designed to switch the codons for $_{798}$SGSD to IDGR to make it a substrate for Factor Xa cleavage is performed. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 30.

Example 24 – Creation of an LHA-EGF chimaeric protein that incorporates a Thrombin recognition site into the LC domain [SXN1974]

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/A and the human epidermal growth factor sequence (SEQ ID 8) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin (GRG). Simple text character analysis of the primary sequence identified the sequence $_{103}$GRM$_{105}$ within the LC domain. The location of the peptide in the tertiary structure of the LC is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 7 (encoding SEQ ID 8) using a primer designed to switch the peptide region from GRM to GRG was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 31.

Example 25 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Factor Xa recognition site into the LC [SXN1975]

A novel molecule incorporating a Factor Xa recognition site into the LC of LHC-EGF is constructed according to Example 17. Using methodology similar to that described in Example 6, the protein of Example 17 is expressed and purified. The methodology was adapted for use on an AKTA Xpress purification system. Essentially, the clarified *E.coli* lysates were applied to a 5ml HisTrap FF Crude column on the Xpress system. The program was set to wash the columns with 10 column volumes of binding buffer (50mM Tris pH8.0, 200mM NaCl) and 10 col. vols. of 40mM imidazole in binding buffer (collected together with the flow through). Elution was with 5 col. vols. of 250mM imidazole in binding buffer. The protein was collected in a loop and held until system was ready to desalt (in 50mM Tris pH8.0, 150mM NaCl). The desalted protein was collected in a 2ml 96 well plate. Figure 3 illustrates purification of LHC-EGF from *E. coli.*

Using methodology described in Example 7, the protein is treated with Factor Xa protease and samples analysed by SDS-PAGE. Figure 4 illustrates the cleavage of the protein in the presence of Factor Xa. Cleavage products are observed in the non-reduced and reduced samples. The estimated mass of the cleavage products is in agreement with the anticipated cleavage point of the engineered protein

Example 26 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Factor Xa recognition site into the H_N [SXN1937 & SXN1938]

A novel molecule incorporating a Factor Xa recognition site into the H_N of LHC-EGF is constructed according to Example 20, and a second novel jmolecule incorporating a Factor Xa recognition site into a different location within the H_N of LHC-EGF is constructed according to Example 21. Using methodology similar to that described in Example 24, the proteins of Example 20 and 21 are expressed and purified. Figure 5 illustrates purification of LHC-EGF from Example 20 from *E. coli*, and Figure 6 illustrates purification of LHC-EGF from Example 21 from *E. coli*.

Using methodology described in Example 7, the protein of Example 20 is treated with Factor Xa protease and samples analysed by SDS-PAGE. Figure 7 illustrates the cleavage of the protein in the presence of Factor Xa, as assessed by staining of SDS-PAGE gels. Figure 8 illustrates the profile of the samples when assessed by Western blotting using anti-His tag antibodies to probe for the presence of the His tag. The estimated mass of the cleavage products is in agreement with the anticipated cleavage point of the engineered protein.

Using methodology described in Example 7, the protein of Example 21 is treated with Factor Xa protease and samples analysed by SDS-PAGE. Figure 9 illustrates the cleavage of the protein in the presence of Factor Xa. The estimated mass of the cleavage products is in agreement with the anticipated

cleavage point of the engineered protein.

Example 27 – Demonstration of specific cleavage of a purified LHC-EGF chimaeric protein that is engineered to incorporate a Thrombin recognition site into the LC [SXN1932]

A novel molecule incorporating a Thrombin recognition site into the LC of LHC-EGF is constructed according to Example 19. Using methodology similar to that described in Example 25, the protein of Example 19 is expressed and purified. Figure 10 illustrates purification of LHC-EGF from *E. coli.*

Using methodology described in Example 7, the protein is treated with Thrombin protease and samples analysed by SDS-PAGE. Figure 11 illustrates the cleavage of the protein in the presence of Thrombin, as assessed by SDS-PAGE. Figure 12 illustrates the cleavage of the protein in the presence of Thrombin, as assessed by Western blotting using anti-EGF antibodies. The estimated mass of the cleavage products is in agreement with the anticipated cleavage point of the engineered protein

Example 28 – Demonstration of specific cleavage of a purified LHA-EGF chimaeric protein that is engineered to incorporate a Thrombin recognition site into the LC [SXN1974]

A novel molecule incorporating a Factor Xa recognition site into the LC of LHA-EGF is constructed according to Example 24. Using methodology similar to that described in Example 25, the protein of Example 24 is expressed and purified. Figure 13 illustrates purification of LHA-EGF from *E. coli.*

Using methodology described in Example 7, the protein is treated with Thrombin protease and samples analysed by SDS-PAGE. Figure 14 illustrates the cleavage of the protein in the presence of Thrombin. Figure 15 illustrates the Western blot profile of the same PAGE, using anti-EGF as primary antibody. The estimated mass of the cleavage products is in agreement with the anticipated cleavage point of the engineered protein

Example 29 – Demonstration of reduced *in vitro* cellular activity of a protein engineered to incorporate a FXa protease cleavage site into the LC domain of LHC-EGF [SXN1975]

The protein product of Example 25 is expressed and purified. The purified protein is exposed to FXa protease for prior to assessment in an *in vitro* spinal cord neuron (SCN) assay. The preparation of SCN is a well established technique and is described in the literature [B.R. Ransom, E. Neale, M. Henkart, P.N. Bullock, P.G. Nelson, Mouse spinal cord in cell culture. I. Morphology and intrinsic neuronal electrophysiologic properties, J. Neurophysiol. 40 (1977) 1132–1150; S.C. Fitzgerald, A Dissection and Tissue Culture Manual of the Nervous System, Alan R. Liss Inc, New York, 1989]. Test protein is prepared at a variety of concentrations by dilution into culture media. SCNs are exposed to the test proteins for 24hr prior to removal of media and preparation of the cellular material for analysis by SDS-PAGE and Western blotting. Following separation of cellular proteins on Novex 4-20% Tris-glycine polyacrylamide gels, the proteins are transferred to nitrocellulose and subsequently probed for the presence of the appropriate SNARE protein using antibodies obtained from commercial sources. In this case, the antibodies were specific for the SNARE syntaxin.

Referring to Figure 16, the protein that has been treated with Factor Xa is clearly less effective at cleaving Syntaxin than the protein that was not treated with FXa. The invention has therefore enabled a reduction in the efficacy of the modified protein.

Example 30 - Demonstration of reduced *in vitro* cellular activity of a protein engineered to incorporate a Thrombin protease cleavage site into the LC domain of LHA-EGF [SXN1974]

The protein product of Example 24 is expressed and purified. The purified protein is exposed to Thrombin protease for prior to assessment in an *in vitro* spinal cord neuron (SCN) assay. The preparation of SCN is a well established

technique and is described in the literature [B.R. Ransom, E. Neale, M. Henkart, P.N. Bullock, P.G. Nelson, Mouse spinal cord in cell culture. I. Morphology and intrinsic neuronal electrophysiologic properties, J. Neurophysiol. 40 (1977) 1132–1150; S.C. Fitzgerald, A Dissection and Tissue Culture Manual of the Nervous System, Alan R. Liss Inc, New York, 1989]. Test protein is prepared at a variety of concentrations by dilution into culture media. SCNs are exposed to the test proteins for 24hr prior to removal of media and preparation of the cellular material for analysis by SDS-PAGE and Western blotting. Following separation of cellular proteins on Novex 4-20% Tris-glycine polyacrylamide gels, the proteins are transferred to nitrocellulose and subsequently probed for the presence of the appropriate SNARE protein using antibodies obtained from commercial sources. In this case, the antibodies were specific for the SNARE SNAP-25. Figure 17 demonstrates SNAP-25-cleavage by thrombin-treated L(Thr)HA-EGF compared to untreated L(Thr)HA-EGF.

Example 31 – Creation of a recombinant BoNT/A protein that incorporates a Thrombin recognition site into the LC

The primary sequence of a recombinant endopeptidase active BoNT/A containing an engineered activation protease site specific for enterokinase (SEQ ID 18) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for thrombin (GRG). Simple text character analysis of the primary sequence identified the sequence $_{103}$GRM$_{105}$ within the LC domain that is amenable to protein engineering. The location of the peptide in the tertiary structure of the H_N domain is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the H_N peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 17 (which encodes SEQ ID 18) using

a primer designed to switch the codons for M_{105} to G changes the sequence $_{103}GRM_{105}$ to GRG which is a substrate for cleavage by thrombin. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final amino acid sequence of the expression product is illustrated in SEQ ID 32.

Example 32 – Creation of a recombinant BoNT/A protein that incorporates a Factor Xa recognition site into the LC

The primary sequence of a recombinant endopeptidase active BoNT/A containing an engineered activation protease site specific for enterokinase (SEQ ID 18) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR). Simple text character analysis of the primary sequence identified the sequence IDSL within the LC domain that is amenable to protein engineering. The location of the peptide in the tertiary structure of the LC domain is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 17 (which encodes SEQ ID 18) using a primer designed to switch the codons for $_{276}SL$ to GR changes the sequence

IDSL to IDGR which is a substrate for cleavage by Factor Xa. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final amino acid sequence of the expression product is illustrated in SEQ ID 33.

Example 33 – Creation of a recombinant BoNT/A protein that incorporates a Factor Xa recognition site into the H_N

The primary sequence of a recombinant endopeptidase active BoNT/A containing an engineered activation protease site specific for enterokinase (SEQ ID 18) is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR). Simple text character analysis of the primary sequence identified the sequence $_{562}$GKSR$_{565}$ within the H_N domain that is amenable to protein engineering. The location of the peptide in the tertiary structure of the H_N domain is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis of the SEQ ID 17 (which encodes SEQ ID 18) using a primer designed to switch the peptide region from GKSR to IEGR which is a substrate for cleavage by Factor Xa. Mutagenesis was achieved utilising

standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final amino acid sequence of the expression product is illustrated in SEQ ID 34.

Example 34 – Creation of a recombinant BoNT/E protein that incorporates a Thrombin recognition site into the LC

The primary sequence of a recombinant endopeptidase active BoNT/E [nucleotide accession AM695755; Uniprot number A8Y867] is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin (LVPRGS). Simple text character analysis of the primary sequence identified the sequence $_{186}$FSPEYS$_{191}$ within the LC domain that is amenable to protein engineering. The location of the peptide in the tertiary structure of the H_N domain is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore good for accessibility by proteases.

Site directed mutagenesis is achieved using a primer designed to switch the peptide region from FSPEYS to IVPRFS which is a substrate for cleavage by Thrombin. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final amino acid sequence of the expression product is illustrated in SEQ ID 35.

Example 35 – Creation of a recombinant BoNT/E protein that incorporates a Factor Xa recognition site into the H_N

The primary sequence of BoNT/E [nucleotide accession AM695755; Uniprot number A8Y867] is reviewed for a potential insertion site for a Factor Xa recognition peptide (IEGR). Comparison of the primary sequence of BoNT/E with that of BoNT/A and the corresponding location of the peptide in the tertiary structure of the H_N domain predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA), concludes that the region $_{727}$TLEE is suitable for protein engineering to IEGR.

Site directed mutagenesis is achieved using a primer designed to switch the peptide region from TLEE to IEGR which is a substrate for cleavage by Factor Xa. Mutagenesis was achieved utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology).

E. coli codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA

is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final amino acid sequence of the expression product is illustrated in SEQ ID 36.

Example 36 – Creation of an LHE-VIPr chimaeric protein that incorporates a Thrombin recognition site into the LC

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/E and an analogue of the human vasoactive intestinal peptide (VIPr) sequence is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Thrombin (GRG). Simple text character analysis of the primary sequence identified the sequence $_{103}GGI_{105}$ within the LC domain of the chimaera. The location of the peptide in the tertiary structure of the LC/E is predicted from the X-ray crystal structure of LC/E (pdb: 1T3A) as the guide. Freely available software (such as Jmol (http://www.rcsb.org/pdb/explore/jmol.do?structureId=1T3A&bionumber=1) are used to identify that the LC peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis is achieved using a primer designed to switch the peptide region from GGI to GRG utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Thrombin site is used to encode the amino acid sequence of the expression product is illustrated in

SEQ ID 37.

Example 37 – Creation of an LHE-VIPr chimaeric protein that incorporates a Factor Xa recognition site into the H_N

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/E and an analogue of vasoactive intestinal peptide (VIPr) sequence is reviewed for the presence of amino acid strings that bear resemblance to the prototypical recognition site for Factor Xa (IEGR).

Simple text character analysis of the primary sequence identified the sequence $_{585}$GENN within the H_N domain. The location of the peptide in the tertiary structure of the H_N/E is predicted from the X-ray crystal structure of BoNT/A (pdb: 3BTA) as the guide. Freely available software (such as FirstGlance in Jmol (http://molvis.sdsc.edu/fgij/fg.htm?mol=3bta)) are used to identify that the LC peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis is achieved using a primer designed to switch the peptide region from GENN to IEGR utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 38.

Example 38 – Creation of an LHE-VIPr chimaeric protein that incorporates

a Factor Xa recognition site into the LC

The primary sequence of a chimaeric protein constructed by a genetic fusion of the LH_N fragment of BoNT/E (incorporating a mutated substrate recognition domain (K228D) and an analogues of the human vasoactive intestinal peptide (VIPr) is reviewed for the presence of amino acid strings that are exposed on the surface of the protein and can be engineered to resemble the prototypical recognition site for Factor Xa (IEGR).

Analysis of the primary sequence identified the sequence $_{268}$VAQY within the LC domain. The location of the peptide in the tertiary structure of the LC/E is predicted from the X-ray crystal structure of BoNT/E (pdb: 1T3A) as the guide. Freely available software (such as Jmol (http://www.rcsb.org/pdb/explore/jmol.do?structureId=1T3A&bionumber=1) are used to identify that the LC peptide sequence is located on the surface. The location is therefore a good region for accessibility by proteases.

Site directed mutagenesis is achieved using a primer designed to switch the peptide region from VAQY to IEGR utilising standard molecular tools for performing mutagenesis (for example, the Stratagene Quickchange mutagenesis methodology). *E. coli* codon usage is assessed by reference to software programs such as Graphical Codon Usage Analyser (Geneart), and the %GC content and codon usage ratio assessed by reference to published codon usage tables (for example GenBank Release 143, September 13 2004) to ensure that the mutagenesis does not result in poor codon utilisation. The mutagenised DNA is incorporated into a standard cloning vector, for example pCR4, prior to transformation into *E. coli* host. The integrity of the ORF DNA is checked by sequencing. The final ORF incorporating the Factor Xa site is used to encode the amino acid sequence of the expression product is illustrated in SEQ ID 39.

Example 39 – Cleavage of SNARE protein by a modified clostridial neurotoxin (LH_N) having the properties described by LIN, *et al.*

(WO02/044199)

Embryonic spinal cord neurons were prepared by dissection from E15 Sprague Dawley rats and dissociated before plating onto Matrigel-coated 96 well plates at 125,000 cells per well in medium (MEM buffered with sodium bicarbonate, 5% inactivated horse serum, 0.6% D-glucose, 2% N1 medium supplement, 40ng/ml corticosterone, 20ng/ml tri-iodothryronine).

After three weeks the cells were incubated with fresh medium containing either recombinant light chain of serotype C (LC/C) or a modified clostridial neurotoxin consisting of the translocation and light chains of serotype C (LHn/C) at half log concentrations between 180 nM and 0.18 nM) for 24hrs at 37°C in a humidified, 5% CO_2 atmosphere.

Cells were lysed with SDS PAGE loading buffer containing DTT. Proteins were separated by SDS PAGE (12 % Tris-Bis), transferred to nitrocellulose membrane and syntaxin detected using rabbit anti-syntaxin 2 antibody (Synaptic Systems, cat#110022). Bound antibody was detected with anti-rabbit IgG-peroxidase conjugate, followed by Westdura for fluorescent signal. Images were scanned and quantitated using Syngene software (GeneTools). The Figure 1 shows cleaved syntaxin as a percentage of total syntaxin, and confirms a neurotoxin activity for the modified clostridial neurotoxin lacking a functional H_C binding domain (LHn/C), but no detectable neurotoxin activity for the modified clostridial neurotoxin lacking a functional H_N translocation domain (LC/C).

SEQUENCE LISTING

SEQ ID NO: 1

ATGATTTCCGAATTTGGCTCGGAGTTCATGCCAATTACGATTAACAATTTTAACTATAGTGATCCGGTGG
ATAATAAAAACATTTTATACCTGGATACCCACTTGAATACTCTTGCCAATGAGCCTGAAAAAGCCTTTCG
CATAACGGGTAACATTTGGGTCATTCCGGACCGTTTTAGCCGGAACTCTAACCCTAATCTGAATAAACCT
CCGCGTGTCACGTCTCCGAAAAGTGGGTATTACGATCCAAATTATCTGAGTACCGATTCAGACAAGGATA
CGTTTCTGAAAGAAATCATAAAACTTTTCAAAAGAATCAACTCCCGTGAAATCGGTGAAGAGCTGATCTA
CCGTCTGTCGACGGACATTCCTTTTCCGGGAAACAATAACACTCCCATTAATACCTTCGACTTTGATGTC
GATTTCAACTCAGTCGATGTGAAAACCCGCCAGGGTAATAACTGGGTTAAAACTGGATCCATTAACCCGT

CCGTTATTATCACAGGTCCTCGTGAAAATATTATAGATCCTGAGACCTCCACGTTCAAGCTGACGAATAA
CACTTTTGCGGCACAGGAAGGGTTTGGTGCCCTTTCAATTATCTCTATCTCTCCGCGCTTCATGTTAACG
TATTCTAACGCAACCAACGATGTTGGCGAGGGCCGCTTCAGCAAAAGTGAATTCTGTATGGATCCCATTC
TGATCTTGATGCATGAGCTTAACCACGCTATGCATAATCTTTATGGTATTGCAATCCCAAACGATCAGAC
GATCTCCAGCGTTACATCTAACATATTCTACAGCCAATATAATGTGAAGCTCGAATATGCAGAGATTTAC
GCCTTCGGTGGGCCGACCATTGACCTCATTCCAAAGTCTGCCCGTAAGTACTTTGAGGAAAAAGCGTTGG
ATTACTATCGTAGCATCGCGAAACGCCTGAATTCAATTACAACTGCAAACCCATCTAGCTTCAACAAATA
CATCGGAGAATATAAACAAAAGCTGATACGCAAATATCGCTTTGTGGTCGAATCGTCCGGGGAAGTGACA
GTTAATCGAAATAAATTTGTTGAACTCTATAATGAATTAACGCAGATCTTCACAGAATTTAATTATGCTA
AAATCTATAATGTACAGAACCGGAAAATTTATCTCAGTAATGTATACACACCGGTGACTGCTAACATTCT
GGACGATAACGTCTACGATATTCAAAATGGCTTTAATATCCCGAAGAGCAACTTGAATGTCCTCTTCATG
GGGCAGAACTTGTCACGTAACCCAGCGCTGCGAAAAGTTAACCCAGAAAATATGTTGTACCTCTTTACAA
AATTCTGTGTAGACGCCGACGATGACGATAAACTGTACAACAAAACCCTGCAATGCCGTGAACTTCTGGT
TAAGAACACCGACCTGCCGTTCATTGGGGACATCAGTGATGTCAAAACGGATATTTTTCTTCGGAAGGAT
ATTAATGAGGAAACCGAAGTGATACCTGACAATGTGTCGGTAGATCAGGTAATCCTGAGTAAGAACACCA
GCGAGCATGGGCAGCTGGATCTGTTGTATCCGAGCATTGACAGCGAGTCGGAAATACTGCCCGGCGAAAA
TCAAGTTTTTTATGACAATCGGACCCAGAATGTTGATTATCTGAATAGTTACTATTACTTGGAGAGCCAA
AAATTATCAGATAATGTGGAAGACTTTACCTTTACCCGGTCTATCGAAGAGGCGCTGGATAACAGCGCGA
AAGTTTACACTTATTTTCCCACGCTCGCAAACAAAGTTAATGCTGGCGTACAGGGTGGATTATTTCTTAT
GTGGGCGAATGATGTGGTAGAGGACTTTACAACCAACATCCTGCGCAAAGACACTTTAGACAAAATTTCT
GACGTCTCGGCCATTATCCCGTATATAGGTCCGGCCTTAAACATAAGCAATTCGGTTCGCCGTGGCAACT
TCACAGAAGCCTTCGCTGTGACTGGTGTGACCATTCTGTTGGAAGCATTTCCTGAGTTTACGATCCCGGC
TCTGGGCGCATTTGTAATTTACTCTAAAGTTCAGGAACGAAATGAAATTATAAAAACTATCGATAATTGC
CTGGAACAGCGTATCAAGAGATGGAAGGATTCCTATGAGTGGATGATGGGGACCTGGCTGTCAAGAATTA
TCACACAGTTTAATAACATATCCTATCAAATGTATGATAGCTTAAACTATCAAGCAGGAGCGATAAAGGC
GAAAATTGACCTGGAATACAAGAAATATTCTGGTTCGGATAAAGAGAATATTAAAAGTCAGGTGGAAAAT
CTGAAAAATAGTTTAGATGTCAAAATTTCTGAGGCGATGAATAACATTAACAAATTCATCCGCGAGTGCA
GTGTAACTTATTTGTTTAAGAATATGTTACCCAAAGTTATCGACGAACTGAATGAATTTGATCGTAATAC
CAAAGCCAAATTGATCAACCTCATCGACTCTCATAACATCATTCTGGTGGGAGAAGTCGACAAACTGAAA
GCTAAGGTGAATAACAGCTTCCAGAATACAATTCCGTTTAATATTTTCTCATACACCAATAACTCGCTGC
TTAAAGATATTATCAACGAATATTTTAATCTGGAGGGTGGCGGTGGCAGTGGCGGTGGCGGATCCGGCGG
TGGCGGTAGCGCACTGGATAATTCAGATTCCGAATGTCCACTGTCACACGATGGTTATTGTCTTCATGAT
GGCGTGTGCATGTATATAGAAGCGTTAGATAAATACGCTTGCAACTGCGTGGTTGGCTATATCGGCGAAC
GTTGTCAGTATCGTGATTTAAAGTGGTGGGAATTACGCTAATGA

SEQ ID NO: 2

ISEFGSEFMPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPP
RVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVD
FNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTY
SNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYA
FGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTV
NRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG
QNLSRNPALRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDI
NEETEVIPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQK

LSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD
VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCL
EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENL
KNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKA
KVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDG
VCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID NO: 3

ATGATTTCCGAATTTGGCTCGGAGTTCATGCCAATTACGATTAACAATTTTAACTATAGTGATCCGGTGG
ATAATAAAAACATTTTATACCTGGATACCCACTTGAATACTCTTGCCAATGAGCCTGAAAAAGCCTTTCG
CATAACGGGTAACATTTGGGTCATTCCGGACCGTTTTAGCCGGAACTCTAACCCTAATCTGAATAAACCT
CCGCGTGTCACGTCTCCGAAAAGTGGGTATTACGATCCAAATTATCTGAGTACCGATTCAGACAAGGATA
CGTTTCTGAAAGAAATCATAAAACTTTTCAAAAGAATCAACTCCCGTGAAATCGGTGAAGAGCTGATCTA
CCGTCTGTCGACGGACATTCCTTTTCCGGGAAACAATAACACTCCCATTAATACCTTCGACTTTGATGTC
GATTTCAACTCAGTCGATGTGAAAACCCGCCAGGGTAATAACTGGGTTAAAACTGGATCCATTAACCCGT
CCGTTATTATCACAGGTCCTCGTGAAAATATTATAGATCCTGAGACCTCCACGTTCAAGCTGACGAATAA
CACTTTTGCGGCACAGGAAGGGTTTGGTGCCCTTTCAATTATCTCTATCTCTCCGCGCTTCATGTTAACG
TATTCTAACGCAACCAACGATGTTATCGAGGGCCGCTTCAGCAAAAGTGAATTCTGTATGGATCCCATTC
TGATCTTGATGCATGAGCTTAACCACGCTATGCATAATCTTTATGGTATTGCAATCCCAAACGATCAGAC
GATCTCCAGCGTTACATCTAACATATTCTACAGCCAATATAATGTGAAGCTCGAATATGCAGAGATTTAC
GCCTTCGGTGGGCCGACCATTGACCTCATTCCAAAGTCTGCCCGTAAGTACTTTGAGGAAAAAGCGTTGG
ATTACTATCGTAGCATCGCGAAACGCCTGAATTCAATTACAACTGCAAACCCATCTAGCTTCAACAAATA
CATCGGAGAATATAAACAAAAGCTGATACGCAAATATCGCTTTGTGGTCGAATCGTCCGGGGAAGTGACA
GTTAATCGAAATAAATTTGTTGAACTCTATAATGAATTAACGCAGATCTTCACAGAATTTAATTATGCTA
AAATCTATAATGTACAGAACCGGAAAATTTATCTCAGTAATGTATACACACCGGTGACTGCTAACATTCT
GGACGATAACGTCTACGATATTCAAAATGGCTTTAATATCCCGAAGAGCAACTTGAATGTCCTCTTCATG
GGGCAGAACTTGTCACGTAACCCAGCGCTGCGAAAAGTTAACCCAGAAAATATGTTGTACCTCTTTACAA
AATTCTGTGTAGACGCCGACGATGACGATAAACTGTACAACAAAACCCTGCAATGCCGTGAACTTCTGGT
TAAGAACACCGACCTGCCGTTCATTGGGGACATCAGTGATGTCAAAACGGATATTTTTCTTCGGAAGGAT
ATTAATGAGGAAACCGAAGTGATACCTGACAATGTGTCGGTAGATCAGGTAATCCTGAGTAAGAACACCA
GCGAGCATGGGCAGCTGGATCTGTTGTATCCGAGCATTGACAGCGAGTCGGAAATACTGCCCGGCGAAAA
TCAAGTTTTTTATGACAATCGGACCCAGAATGTTGATTATCTGAATAGTTACTATTACTTGGAGAGCCAA
AAATTATCAGATAATGTGGAAGACTTTACCTTTACCCGGTCTATCGAAGAGGCGCTGGATAACAGCGCGA
AAGTTTACACTTATTTTCCCACGCTCGCAAACAAAGTTAATGCTGGCGTACAGGGTGGATTATTTCTTAT
GTGGGCGAATGATGTGGTAGAGGACTTTACAACCAACATCCTGCGCAAAGACACTTTAGACAAAATTTCT
GACGTCTCGGCCATTATCCCGTATATAGGTCCGGCCTTAAACATAAGCAATTCGGTTCGCCGTGGCAACT
TCACAGAAGCCTTCGCTGTGACTGGTGTGACCATTCTGTTGGAAGCATTTCCTGAGTTTACGATCCCGGC
TCTGGGCGCATTTGTAATTTACTCTAAAGTTCAGGAACGAAATGAAATTATAAAAACTATCGATAATTGC
CTGGAACAGCGTATCAAGAGATGGAAGGATTCCTATGAGTGGATGATGGGGACCTGGCTGTCAAGAATTA
TCACACAGTTTAATAACATATCCTATCAAATGTATGATAGCTTAAACTATCAAGCAGGAGCGATAAAGGC
GAAAATTGACCTGGAATACAAGAAATATTCTGGTTCGGATAAAGAGAATATTAAAAGTCAGGTGGAAAAT
CTGAAAAATAGTTTAGATGTCAAAATTTCTGAGGCGATGAATAACATTAACAAATTCATCCGCGAGTGCA
GTGTAACTTATTTGTTTAAGAATATGTTACCCAAAGTTATCGACGAACTGAATGAATTTGATCGTAATAC
CAAAGCCAAATTGATCAACCTCATCGACTCTCATAACATCATTCTGGTGGGAGAAGTCGACAAACTGAAA

GCTAAGGTGAATAACAGCTTCCAGAATACAATTCCGTTTAATATTTTCTCATACACCAATAACTCGCTGC
TTAAAGATATTATCAACGAATATTTTAATCTGGAGGGTGGCGGTGGCAGTGGCGGTGGCGGATCCGGCGG
TGGCGGTAGCGCACTGGATAATTCAGATTCCGAATGTCCACTGTCACACGATGGTTATTGTCTTCATGAT
GGCGTGTGCATGTATATAGAAGCGTTAGATAAATACGCTTGCAACTGCGTGGTTGGCTATATCGGCGAAC
GTTGTCAGTATCGTGATTTAAAGTGGTGGGAATTACGCTAATGA

SEQ ID NO: 4

MISEFGSEFMPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP
PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDV
DFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLT
YSNATNDVIEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIY
AFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVT
VNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFM
GQNLSRNPALRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKD
INEETEVIPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQ
KLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKIS
DVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNC
LEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVEN
LKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLK
AKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHD
GVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID NO: 5

ATGATTTCCGAATTTGGCTCGGAGTTCATGCCAATTACGATTAACAATTTTAACTATAGTGATCCGGTGG
ATAATAAAAACATTTTATACCTGGATACCCACTTGAATACTCTTGCCAATGAGCCTGAAAAAGCCTTTCG
CATAACGGGTAACATTTGGGTCATTCCGGACCGTTTTAGCCGGAACTCTAACCCTAATCTGAATAAACCT
CCGCGTGTCACGTCTCCGAAAAGTGGGTATTACGATCCAAATTATCTGAGTACCGATTCAGACAAGGATA
CGTTTCTGAAAGAAATCATAAAACTTTTCAAAAGAATCAACTCCCGTGAAATCGGTGAAGAGCTGATCTA
CCGTCTGTCGACGGACATTCCTTTTCCGGGAAACAATAACACTCCCATTAATACCTTCGACTTTGATGTC
GATTTCAACTCAGTCGATGTGAAAACCCGCCAGGGTAATAACTGGGTTAAAACTGGATCCATTAACCCGT
CCGTTATTATCACAGGTCCTCGTGAAAATATTATAGATCCTGAGACCTCCACGTTCAAGCTGACGAATAA
CACTTTTGCGGCACAGGAAGGGTTTGGTGCCCTTTCAATTATCTCTATCGTTCCGCGCTTCTCCTTAACG
TATTCTAACGCAACCAACGATGTTGGCGAGGGCCGCTTCAGCAAAAGTGAATTCTGTATGGATCCCATTC
TGATCTTGATGCATGAGCTTAACCACGCTATGCATAATCTTTATGGTATTGCAATCCCAAACGATCAGAC
GATCTCCAGCGTTACATCTAACATATTCTACAGCCAATATAATGTGAAGCTCGAATATGCAGAGATTTAC
GCCTTCGGTGGGCCGACCATTGACCTCATTCCAAAGTCTGCCCGTAAGTACTTTGAGGAAAAAGCGTTGG
ATTACTATCGTAGCATCGCGAAACGCCTGAATTCAATTACAACTGCAAACCCATCTAGCTTCAACAAATA
CATCGGAGAATATAAACAAAAGCTGATACGCAAATATCGCTTTGTGGTCGAATCGTCCGGGGAAGTGACA
GTTAATCGAAATAAATTTGTTGAACTCTATAATGAATTAACGCAGATCTTCACAGAATTTAATTATGCTA
AAATCTATAATGTACAGAACCGGAAAATTTATCTCAGTAATGTATACACACCGGTGACTGCTAACATTCT
GGACGATAACGTCTACGATATTCAAAATGGCTTTAATATCCCGAAGAGCAACTTGAATGTCCTCTTCATG
GGGCAGAACTTGTCACGTAACCCAGCGCTGCGAAAAGTTAACCCAGAAAATATGTTGTACCTCTTTACAA
AATTCTGTGTAGACGCCGACGATGACGATAAACTGTACAACAAAACCCTGCAATGCCGTGAACTTCTGGT
TAAGAACACCGACCTGCCGTTCATTGGGGACATCAGTGATGTCAAAACGGATATTTTTCTTCGGAAGGAT

ATTAATGAGGAAACCGAAGTGATACCTGACAATGTGTCGGTAGATCAGGTAATCCTGAGTAAGAACACCA
GCGAGCATGGGCAGCTGGATCTGTTGTATCCGAGCATTGACAGCGAGTCGGAAATACTGCCCGGCGAAAA
TCAAGTTTTTTATGACAATCGGACCCAGAATGTTGATTATCTGAATAGTTACTATTACTTGGAGAGCCAA
AAATTATCAGATAATGTGGAAGACTTTACCTTTACCCGGTCTATCGAAGAGGCGCTGGATAACAGCGCGA
AAGTTTACACTTATTTTCCCACGCTCGCAAACAAAGTTAATGCTGGCGTACAGGGTGGATTATTTCTTAT
GTGGGCGAATGATGTGGTAGAGGACTTTACAACCAACATCCTGCGCAAAGACACTTTAGACAAAATTTCT
GACGTCTCGGCCATTATCCCGTATATAGGTCCGGCCTTAAACATAAGCAATTCGGTTCGCCGTGGCAACT
TCACAGAAGCCTTCGCTGTGACTGGTGTGACCATTCTGTTGGAAGCATTTCCTGAGTTTACGATCCCGGC
TCTGGGCGCATTTGTAATTTACTCTAAAGTTCAGGAACGAAATGAAATTATAAAAACTATCGATAATTGC
CTGGAACAGCGTATCAAGAGATGGAAGGATTCCTATGAGTGGATGATGGGGACCTGGCTGTCAAGAATTA
TCACACAGTTTAATAACATATCCTATCAAATGTATGATAGCTTAAACTATCAAGCAGGAGCGATAAAGGC
GAAAATTGACCTGGAATACAAGAAATATTCTGGTTCGGATAAAGAGAATATTAAAAGTCAGGTGGAAAAT
CTGAAAAATAGTTTAGATGTCAAAATTTCTGAGGCGATGAATAACATTAACAAATTCATCCGCGAGTGCA
GTGTAACTTATTTGTTTAAGAATATGTTACCCAAAGTTATCGACGAACTGAATGAATTTGATCGTAATAC
CAAAGCCAAATTGATCAACCTCATCGACTCTCATAACATCATTCTGGTGGGAGAAGTCGACAAACTGAAA
GCTAAGGTGAATAACAGCTTCCAGAATACAATTCCGTTTAATATTTTCTCATACACCAATAACTCGCTGC
TTAAAGATATTATCAACGAATATTTTAATCTGGAGGGTGGCGGTGGCAGTGGCGGTGGCGGATCCGGCGG
TGGCGGTAGCGCACTGGATAATTCAGATTCCGAATGTCCACTGTCACACGATGGTTATTGTCTTCATGAT
GGCGTGTGCATGTATATAGAAGCGTTAGATAAATACGCTTGCAACTGCGTGGTTGGCTATATCGGCGAAC
GTTGTCAGTATCGTGATTTAAAGTGGTGGGAATTACGCTAATGA

SEQ ID NO: 6

MISEFGSEFMPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKP
PRVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDV
DFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISIVPRFSLT
YSNATNDVGEGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIY
AFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVT
VNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFM
GQNLSRNPALRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKD
INEETEVIPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQ
KLSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKIS
DVSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNC
LEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVEN
LKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLK
AKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHD
GVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID NO: 7

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt
tatcccggaacgtgatacctttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt

attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
ctgggcgctggtaaattcgcaactgatcctgcggttaccctggctcacgaactgattcatgcaggccacc
gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa
aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacggcaaatcccgtatcgcactgactaactccgttaacgaagctctgctcaacccgtccc
gtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttctt
gggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaatt
gcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaagacg
acttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccatccc
ggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaac
gcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaagg
ttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaa
ggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatcgac
gatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaaccagt
gctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcgtc
tctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctg
aaggacaaagtgaacaataccttatcgaccgacatccctttcagctcagtaaatatgtcgataaccaac
gccttttgtccactctagaaggtggcggtgggtccggtggcggtggctcaggcgggggcggtagcgcact
agacaactctgactctgaatgcccgctgtctcacgacggttactgcctgcacgacggtgtttgcatgtac
atcgaagctctggacaaatacgcttgcaactgcgttgttggttacatcggtgaacgttgccagtaccgtg
acctgaaatggtgggaactgcgtgcgctagaagcaCACCATCATCACcaccatcaccatcaccattaatg
a

SEQ ID NO: 8

MGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQ
VPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINV
IQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPL
LGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSL
QENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTE
IYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLK

NFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNI
EAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQ
EFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKI
ADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDN
ALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNID
DLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRL
KDKVNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMY
IEALDKYACNCVVGYIGERCQYRDLKWWELRALEAHHHHHHHHHH

SEQ ID NO: 9

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt
tatcccggaacgtgataccttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtggtctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt
attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
ctgggcgctggtaaattcgcaactgatcctgcggttaccctggctcacgaactgattcatgcaggccacc
gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaatttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa
aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacggcaaatcccgtatcgcactgactaactccgttaacgaagctctgctcaacccgtccc
gtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttctt
gggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaatt
gcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaagacg
acttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccatccc
ggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaac
gcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaagg
ttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaa
ggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatcgac
gatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaaccagt

gctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcgtc
tctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctg
aaggacaaagtgaacaataccttatcgaccgacatccctttcagctcagtaaatatgtcgataaccaac
gcctttgtccactctagaaggtggcggtgggtccggtggcggtggctcaggcgggggcggtagcgcact
agacaactctgactctgaatgcccgctgtctcacgacggttactgcctgcacgacggtgtttgcatgtac
atcgaagctctggacaaatacgcttgcaactgcgttgttggttacatcggtgaacgttgccagtaccgtg
acctgaaatggtgggaactgcgtgcgctagaagcaCACCATCATCACcaccatcaccatcaccattaatg
a

SEQ ID NO: 10
MGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQ
VPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRGLLTSIVRGIPFWGGSTIDTELKVIDTNCINV
IQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPL
LGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSL
QENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTE
IYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLK
NFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNI
EAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQ
EFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKI
ADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDN
ALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNID
DLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRL
KDKVNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMY
IEALDKYACNCVVGYIGERCQYRDLKWWELRALEAHHHHHHHHHH

SEQ ID 11
ISEFGSEFMPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPP
RVTSPKSGYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVD
FNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTY
SNATNDVRSRRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYA
FGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTV
NRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMG
QNLSRNPALRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDI
NEETEVIPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQK
LSDNVEDFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISD
VSAIIPYIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCL
EQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENL
KNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKA
KVNNSFQNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDG
VCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 12
atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt

tatcccggaacgtgatacctttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt
attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
ctgggcgctggtaaattcgcaactgatcctgcggttaccctggctcacgaactgattcatgcaggccacc
gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa
aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacATcGaaGGccgtatcgcactgactaactccgttaacgaagctctgctcaacccgtccc
gtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttctt
gggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaatt
gcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaagacg
acttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccatccc
ggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaac
gcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaagg
ttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaa
ggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatcgac
gatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaaccagt
gctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcgtc
tctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctg
aaggacaaagtgaacaataccttatcgaccgacatccctttcagctcagtaaatatgtcgataaccaac
gccttttgtccactctagaaggtggcggtgggtccggtggcggtggctcaggcgggggcggtagcgcact
agacaactctgactctgaatgcccgctgtctcacgacggttactgcctgcacgacggtgtttgcatgtac
atcgaagctctggacaaatacgcttgcaactgcgttgttggttacatcggtgaacgttgccagtaccgtg
acctgaaatggtgggaactgcgtgcgctagaagcaCACCATCATCACcaccatcaccatcaccattaatg
a

SEQ ID 13

GSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQV
PVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVI

QPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLL
GAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ
ENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI
YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKN
FTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIE
AAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQE
FEHIEGRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIA
DITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNA
LSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDD
LSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLK
DKVNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYI
EALDKYACNCVVGYIGERCQYRDLKWWELRALEAHHHHHHHHHH

SEQ ID 14

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt
tatcccggaacgtgataccttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt
attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
GCGCAGgctGTTCGTTCCTCTTCTgatcctgcggttaccctggctcacgaactgattcatgcaggccacc
gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa
aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacggcaaatcccgtatcgcactgactaactccgttaacgaagctctgctcaacccgtccc
gtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttctt
gggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaatt
gcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaagacg
acttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccatccc
ggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaac

gcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaagg
ttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaa
ggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatcgac
gatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaaccagt
gctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcgtc
tctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctg
aaggacaaagtgaacaataccttatcgaccgacatccctttcagctcagtaaatatgtcgataaccaac
gccttttgtccactctagaaggtggcggtgggtccggtggcggtggctcaggcgggggcggtagcgcact
agacaactctgactctgaatgcccgctgtctcacgacggttactgcctgcacgacggtgtttgcatgtac
atcgaagctctggacaaatacgcttgcaactgcgttgttggttacatcggtgaacgttgccagtaccgtg
acctgaaatggtgggaactgcgtgcgctagaagcaCACCATCATCACcaccatcaccatcaccattaatg
a

SEQ ID 15

GSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQV
PVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVI
QPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLA
QAVRSSSDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ
ENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI
YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKN
FTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIE
AAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQE
FEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIA
DITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNA
LSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDD
LSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLK
DKVNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYI
EALDKYACNCVVGYIGERCQYRDLKWWELRALEAHHHHHHHHHH

SEQ ID 16

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgcttacatcaaaatcccgaacgc
tggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggttatcccggaacgtgataccttactaacccggaag
aaggtgacctgaacccgccaccggaagcgaaacaggtgccggtatcttactatgactccacctacctgtctaccgataacgaaaag
gacaactacctgaaaggtgttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcggt
atcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgttattcagccggacggttcctat
cgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatcatccagttcgagtgtaagagctttggtcacgaagttctgaac
ctcacccgtaacggctacggttccactcagtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacga
acccactgctgggcgctggtaaattcgcaactgatcctgcggttaccctggctcacgaactgattcatgcaggccaccgcctgtacggt
atcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtccggtctggaagttagcttcgaagaactgcg
tacttttggcggtcacgacgctaaattcatcgactctctgcaagaaaacgagttccgtctgtactactataacaagttcaaagatatcgc
atccaccctgaacaaagcgaaatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctca

gcgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaaatttacaccgaagacaa
cttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgacaaggcagtattcaaaatcaacatcgtgccgaaagttaa
ctacactatctacgatggtttcaacctgcgtaacaccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaac
ttcacaaaactgaaaaacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAA
CTAAATCTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggattt
attcttcagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatcgaagcagcc
gaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacgagccggaaaacatttctatcgaaaacc
tgagctctgatatcatcggccagctggaactgatgccgaacatcgaacgtttcccaaacggtaaaaagtacgagctggacaaatata
ccatgttccactacctgcgcgcgcaggaatttgaacacggcaaatcccgtatcgcactgactaactccgttaacgaagctctgctcaa
cccgtcccgtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttcttgggttgggttga
acagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaattgcggatatcactatcatcatcccgtacatcggt
ccggctctgaacattggcaacatgctgtacaaagacgacttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatc
ccggaaatcgccatcccggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaacgc
gctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaaggttaatactcagatcgacctc
atccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaaggcaatcattaactaccagtacaaccagtacac
cgaggaagaaaaaaacaacatcaacttcaacatcgacgatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaac
atcaacaagttcctgaaccagtgctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcg
tctctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctgaaggacaaagtgaaca
ataccttatcgaccgacatccctttcagctcagtaaatatgtcgataaccaacgcctttgtccactttcaccgaatacatcaaaaacat
catcaacaccagtctagaaATCCTGAACCTGCGTTACGAATCTAACCACCTGATCGACCTGTCTCG
TTACGCTTCTAAAATCAACATCGGTTCTAAAGTTAACTTCGACCCGATCGACAAAAACCAG
ATCCAGCTGTTCAACCTGGAATCTTCTAAAATCGAAGTTATCCTGAAAAACGCTATCGTTTA
CAACTCTATGTACGAAAACTTCTCTACCTCTTTCTGGATTCGTATCCCGAAATACTTTAACT
CTATCTCTCTGAACAACGAATACACCATCATCAACTGCATGGAAAACAACTCTGGTTGGAA
AGTTTCTCTGAACTACGGTGAAATCATCTGGACCCTGCAAGACACCCAGGAAATCAAACAG
CGTGTTGTTTTCAAATACTCTCAGATGATCAACATCTCTGACTACATCAACCGTTGGATCTT
CGTTACCATCACCAACAACCGTCTGAACAACTCTAAAATCTACATCAACGGTCGTCTGATC
GACCAGAAACCGATCTCTAACCTGGGTAACATCCACGCTTCTAACAACATCATGTTCAAAC
TGGACGGTTGCCGTGACACCCACCGTTACATCTGGATCAAATACTTCAACCTGTTCGACAA
AGAACTGAACGAAAAAGAAATCAAAGACCTGTACGACAACCAGTCTAACTCTGGTgcactagtg
ATTTTGAAGGACTTTTGGGGCGACTATCTCCAGTACGACAAACCTTACTATATGCTGAATT
TGTATGATCCCAACAAATATGTGGATGTGAATAACGTTGGTATTAGGGGTTACATGTATTT
GAAGGGTCCAAGGGGGTCAGTCATGACAACCAATATCTACTTAAATTCCTCTCTTTACCGA
GGGACAAAATTCATTATCAAAAAGTATGCTAGTGGAAATAAAGATAATATAGTCAGAAACAA
TGATCGCGTTTACATTAACGTGGTAGTCAAAAATAAGGAGTATAGACTAGCTACGAATGCA
TCGCAGGCGGGAGTGGAGAAGATACTGAGCGCACTAGAAATACCTGACGTAGGAAACTTA
AGCCAGGTTGTCGTTATGAAATCAAAGAACGATCAAGGAATTACTAATAAGTGTAAGATGA

ACTTACAAGATAACAATGGCAATGATATAGGCTTCATCGGGTTTCATCAATTTAACAACATA
GCGAAACTCGTAGCCTCTAACTGGTACAACCGTCAAATCGAACGAAGTTCCCGTACTCTA
GGTTGCTCGTGGGAGTTCATCCCAGTAGACGACGGGTGGGGCGAACGGCCGCTTgcgctag
caCACCATCATCACcaccatcaccatcaccattaatga

SEQ ID 17

HMGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLN
PPPEAKQVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDT
ELKVIDTNCINVIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDF
TFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLE
VSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYL
LSEDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYD
GFNLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNL
QCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIE
NLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVY
TFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYK
DDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALSKRNEKWDEVYKYIVTN
WLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLSSKLNESINKAMI
NINKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDKVNNTLST
DIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKVNFDPIDKNQI
QLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNY
GEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGNIH
ASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDK
PYYMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIV
RNNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMN
LQDNNGNDIGFIGFHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHH
HHHHHHHH

SEQ ID 18

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt
tatcccggaacgtgataccttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt
attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
GCGCAGgctGTTCGTTCCTCTTCTgatcctgcggttaccctggctcacgaactgattcatgcaggccacc

gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa
aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacggcaaatcccgtatcgcactgactaactccgttaacgaagctctgctcaacccgtccc
gtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgttctt
gggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaaatt
gcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaagacg
acttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccatccc
ggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgacaac
gcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctggctaagg
ttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctaccaa
ggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatcgac
gatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaaccagt
gctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgcgtc
tctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgtctg
aaggacaaagtgaacaataccttatcgaccgacatccctttttcagctcagtaaatatgtcgataaccaac
gccttttgtccactttcaccgaatacatcaaaaacatcatcaacaccagtctagaaATCCTGAACCTGCG
TTACGAATCTAACCACCTGATCGACCTGTCTCGTTACGCTTCTAAAATCAACATCGGTTCTAAAGTTAAC
TTCGACCCGATCGACAAAAACCAGATCCAGCTGTTCAACCTGGAATCTTCTAAAATCGAAGTTATCCTGA
AAAACGCTATCGTTTACAACTCTATGTACGAAAACTTCTCTACCTCTTTCTGGATTCGTATCCCGAAATA
CTTTAACTCTATCTCTCTGAACAACGAATACACCATCATCAACTGCATGGAAAACAACTCTGGTTGGAAA
GTTTCTCTGAACTACGGTGAAATCATCTGGACCCTGCAAGACACCCAGGAAATCAAACAGCGTGTTGTTT
TCAAATACTCTCAGATGATCAACATCTCTGACTACATCAACCGTTGGATCTTCGTTACCATCACCAACAA
CCGTCTGAACAACTCTAAAATCTACATCAACGGTCGTCTGATCGACCAGAAACCGATCTCTAACCTGGGT
AACATCCACGCTTCTAACAACATCATGTTCAAACTGGACGGTTGCCGTGACACCCACCGTTACATCTGGA
TCAAATACTTCAACCTGTTCGACAAAGAACTGAACGAAAAAGAAATCAAAGACCTGTACGACAACCAGTC
TAACTCTGGTgcactagtgATTTTGAAGGACTTTTGGGGCGACTATCTCCAGTACGACAAACCTTACTAT
ATGCTGAATTTGTATGATCCCAACAAATATGTGGATGTGAATAACGTTGGTATTAGGGGTTACATGTATT
TGAAGGGTCCAAGGGGGTCAGTCATGACAACCAATATCTACTTAAATTCCTCTCTTTACCGAGGGACAAA
ATTCATTATCAAAAAGTATGCTAGTGGAAATAAAGATAATATAGTCAGAAACAATGATCGCGTTTACATT
AACGTGGTAGTCAAAAATAAGGAGTATAGACTAGCTACGAATGCATCGCAGGCGGGAGTGGAGAAGATAC
TGAGCGCACTAGAAATACCTGACGTAGGAAACTTAAGCCAGGTTGTCGTTATGAAATCAAAGAACGATCA
AGGAATTACTAATAAGTGTAAGATGAACTTACAAGATAACAATGGCAATGATATAGGCTTCATCGGGTTT

CATCAATTTAACAACATAGCGAAACTCGTAGCCTCTAACTGGTACAACCGTCAAATCGAACGAAGTTCCC
GTACTCTAGGTTGCTCGTGGGAGTTCATCCCAGTAGACGACGGGTGGGGCGAACGGCCGCTTgcgctagc
aCACCATCATCACcaccatcaccatcaccattaatga

SEQ ID 19

GSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQV
PVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVI
QPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLA
QAVRSSSDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ
ENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI
YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKN
FTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIE
AAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQE
FEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIA
DITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNA
LSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDD
LSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLK
DKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKVNF
DPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKV
SLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLGN
IHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDKPYYM
LNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYIN
VVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGFH
QFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHHHHHHHHH

SEQ ID 20

atgggatccatggagttcgttaacaaacagttcaactataaagacccagttaacggtgttgacattgctt
acatcaaaatcccgaacgctggccagatgcagccggtaaaggcattcaaaatccacaacaaaatctgggt
tatcccggaacgtgatacctttactaacccggaagaaggtgacctgaacccgccaccggaagcgaaacag
gtgccggtatcttactatgactccacctacctgtctaccgataacgaaaaggacaactacctgaaaggtg
ttactaaactgttcgagcgtatttactccaccgacctgggccgtatgctgctgactagcatcgttcgcgg
tatcccgttctggggcggttctaccatcgataccgaactgaaagtaatcgacactaactgcatcaacgtt
attcagccggacggttcctatcgttccgaagaactgaacctggtgatcatcggcccgtctgctgatatca
tccagttcgagtgtaagagctttggtcacgaagttctgaacctcacccgtaacggctacggttccactca
gtacatccgtttctctccggacttcaccttcggttttgaagaatccctggaagtagacacgaacccactg
ctgggcgctggtaaattcgcaactgatcctgcggttaccctggctcacgaactgattcatgcaggccacc
gcctgtacggtatcgccatcaatccgaaccgtgtcttcaaagttaacaccaacgcgtattacgagatgtc
cggtctggaagttagcttcgaagaactgcgtacttttggcggtcacgacgctaaattcatcgactctctg
caagaaaacgagttccgtctgtactactataacaagttcaaagatatcgcatccaccctgaacaaagcga
aatccatcgtgggtaccactgcttctctccagtacatgaagaacgtttttaaagaaaaatacctgctcag
cgaagacacctccggcaaattctctgtagacaagttgaaattcgataaactttacaaaatgctgactgaa
atttacaccgaagacaacttcgttaagttctttaaagttctgaaccgcaaaacctatctgaacttcgaca
aggcagtattcaaaatcaacatcgtgccgaaagttaactacactatctacgatggtttcaacctgcgtaa
caccaacctggctgctaattttaacggccagaacacggaaatcaacaacatgaacttcacaaaactgaaa

aacttcactggtctgttcgagttttacaagctgctgtgcGTCGACGGCATCATTACCTCCAAAACTAAAT
CTGACGATGACGATAAAAACAAAGCGCTGAACCTGCAGtgtatcaaggttaacaactgggatttattctt
cagcccgagtgaagacaacttcaccaacgacctgaacaaaggtgaagaaatcacctcagatactaacatc
gaagcagccgaagaaaacatctcgctggacctgatccagcagtactacctgacctttaatttcgacaacg
agccggaaaacatttctatcgaaaacctgagctctgatatcatcggccagctggaactgatgccgaacat
cgaacgtttcccaaacggtaaaaagtacgagctggacaaatataccatgttccactacctgcgcgcgcag
gaatttgaacacggcCGTtcccgtCGCatcgcactgactaactccgttaacgaagctctgctcaacccgt
cccgtgtatacaccttcttctctagcgactacgtgaaaaaggtcaacaaagcgactgaagctgcaatgtt
cttgggttgggttgaacagcttgtttatgattttaccgacgagacgtccgaagtatctactaccgacaaa
attgcggatatcactatcatcatcccgtacatcggtccggctctgaacattggcaacatgctgtacaaag
acgacttcgttggcgcactgatcttctccggtgcggtgatcctgctggagttcatcccggaaatcgccat
cccggtactgggcacctttgctctggtttcttacattgcaaacaaggttctgactgtacaaaccatcgac
aacgcgctgagcaaacgtaacgaaaaatgggatgaagtttacaaatatatcgtgaccaactggctgcgcta
aggttaatactcagatcgacctcatccgcaaaaaaatgaaagaagcactggaaaaccaggcggaagctac
caaggcaatcattaactaccagtacaaccagtacaccgaggaagaaaaaaacaacatcaacttcaacatc
gacgatctgtcctctaaactgaacgaatccatcaacaaagctatgatcaacatcaacaagttcctgaacc
agtgctctgtaagctatctgatgaactccatgatcccgtacggtgttaaacgtctggaggacttcgatgc
gtctctgaaagacgccctgctgaaatacatttacgacaaccgtggcactctgatcggtcaggttgatcgt
ctgaaggacaaagtgaacaataccttatcgaccgacatccctttttcagctcagtaaatatgtcgataacc
aacgcctttttgtccactttcaccgaatacatcaaaaacatcatcaacaccagtctagaaATCCTGAACCT
GCGTTACGAATCTAACCACCTGATCGACCTGTCTCGTTACGCTTCTAAAATCAACATCGGTTCTAAAGTT
AACTTCGACCCGATCGACAAAAACCAGATCCAGCTGTTCAACCTGGAATCTTCTAAAATCGAAGTTATCC
TGAAAAACGCTATCGTTTACAACTCTATGTACGAAAACTTCTCTACCTCTTTCTGGATTCGTATCCCGAA
ATACTTTAACTCTATCTCTCTGAACAACGAATACACCATCATCAACTGCATGGAAAACAACTCTGGTTGG
AAAGTTTCTCTGAACTACGGTGAAATCATCTGGACCCTGCAAGACACCCAGGAAATCAAACAGCGTGTTG
TTTTCAAATACTCTCAGATGATCAACATCTCTGACTACATCAACCGTTGGATCTTCGTTACCATCACCAA
CAACCGTCTGAACAACTCTAAAATCTACATCAACGGTCGTCTGATCGACCAGAAACCGATCTCTAACCTG
GGTAACATCCACGCTTCTAACAACATCATGTTCAAACTGGACGGTTGCCGTGACACCCACCGTTACATCT
GGATCAAATACTTCAACCTGTTCGACAAAGAACTGAACGAAAAAGAAATCAAAGACCTGTACGACAACCA
GTCTAACTCTGGTgcactagtgATTTTGAAGGACTTTTGGGGCGACTATCTCCAGTACGACAAACCTTAC
TATATGCTGAATTTGTATGATCCCAACAAATATGTGGATGTGAATAACGTTGGTATTAGGGGTTACATGT
ATTTGAAGGGTCCAAGGGGGTCAGTCATGACAACCAATATCTACTTAAATTCCTCTCTTTACCGAGGGAC
AAAATTCATTATCAAAAAGTATGCTAGTGGAAATAAAGATAATATAGTCAGAAACAATGATCGCGTTTAC
ATTAACGTGGTAGTCAAAAATAAGGAGTATAGACTAGCTACGAATGCATCGCAGGCGGGAGTGGAGAAGA
TACTGAGCGCACTAGAAATACCTGACGTAGGAAACTTAAGCCAGGTTGTCGTTATGAAATCAAAGAACGA
TCAAGGAATTACTAATAAGTGTAAGATGAACTTACAAGATAACAATGGCAATGATATAGGCTTCATCGGG
TTTCATCAATTTAACAACATAGCGAAACTCGTAGCCTCTAACTGGTACAACCGTCAAATCGAACGAAGTT
CCCGTACTCTAGGTTGCTCGTGGGAGTTCATCCCAGTAGACGACGGGTGGGGCGAACGGCCGCTTgcgct
agcaCACCATCATCACcaccatcaccatcaccattaatga

SEQ ID 21

GSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQV
PVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVI
QPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLL

GAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQ
ENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEI
YTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKN
FTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIE
AAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQE
FEHGRSRRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKI
ADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDN
ALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNID
DLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRL
KDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKVN
FDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWK
VSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNLG
NIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDKPYY
MLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVYI
NVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIGF
HQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHHHHHHHHHH

SEQ ID 22

atgccgatcaccatcaacaacttcaactacagcgatccggtggataacaaaaacatcctgtacctggata
cccatctgaataccctggcgaacgaaccggaaaaagcgtttcgtatcaccggcaacatttgggttattcc
ggatcgttttagccgtaacagcaacccgaatctgaataaaccgccgcgtgttaccagcccgaaaagcggt
tattacgatccgaactatctgagcaccgatagcgataaagataccttcctgaaagaaatcatcaaactgt
tcaaacgcatcaacagccgtgaaattggcgaagaactgatctatcgcctgagcaccgatattccgtttcc
gggcaacaacaacaccccgatcaacacctttgatttcgatgtggatttcaacagcgttgatgttaaaacc
cgccagggtaacaattgggtgaaaaccggcagcattaacccgagcgtgattattaccggtccgcgcgaaa
acattattgatccggaaaccagcacctttaaactgaccaacaacacctttgcggcgcaggaaggttttgg
cgcgctgagcattattagcattagcccgcgctttatgctgacctatagcaacgcgaccaacgatgttatt
gaaggccgtttcagcaaaagcgaattttgcatggacccgatcctgatcctgatgcatgaactgaaccatg
cgatgcataacctgtatggcatcgcgattccgaacgatcagaccattagcagcgtgaccagcaacatctt
ttacagccagtacaacgtgaaactggaatatgcggaaatctatgcgtttggcggtccgaccattgatctg
attccgaaaagcgcgcgcaaatacttcgaagaaaaagcgctggattactatcgcagcattgcgaaacgtc
tgaacagcattaccaccgcgaatccgagcagcttcaacaaatatatcggcgaatataaacagaaactgat
ccgcaaatatcgctttgtggtggaaagcagcggcgaagttaccgttaaccgcaataaattcgtggaactg
tacaacgaactgacccagatcttcaccgaatttaactatgcgaaaatctataacgtgcagaaccgtaaaa
tctacctgagcaacgtgtataccccggtgaccgcgaatattctggatgataacgtgtacgatatccagaa
cggctttaacatcccgaaaagcaacctgaacgttctgtttatgggccagaacctgagccgtaatccggcg
ctgcgtaaagtgaacccggaaaacatgctgtacctgttcaccaaattttgcGTCGAcGCGGACGATGACG
ATAAACTGTACAACAAAACCCTGCAGtgtcgtgaactgctggtgaaaaacaccgatctgccgtttattgg
cgatatcagcgatgtgaaaaccgatatcttcctgcgcaaagatatcaacgaagaaaccgaagtgatcccg
gataacgtgagcgttgatcaggtgatcctgagcaaaaacaccagcgaacatggtcagctggatctgctgt
atccgagcattgatagcgaaagcgaaattctgccgggcgaaaaccaggtgttttacgataaccgtaccca
gaacgtggattacctgaacagctattactacctggaaagccagaaactgagcgataacgtggaagatttt
acctttacccgcagcattgaagaagcgctggataacagcgcgaaagtttacacctattttccgaccctgg
cgaacaaagttaatgcgggtgttcagggcggtctgtttctgatgtgggcgaacgatgtggtggaagattt

caccaccaacatcctgcgtaaagataccctggataaaatcagcgatgttagcgcgattattccgtatatt
ggtccggcgctgaacattagcaatagcgtgcgtcgtggcaattttaccgaagcgtttgcggttaccggtg
tgaccattctgctggaagcgtttccggaatttaccattccggcgctgggtgcgtttgtgatctatagcaa
agtgcaggaacgcaacgaaatcatcaaaaccatcgataactgcctggaacagcgtattaaacgctggaaa
gatagctatgaatggatgatgggcacctggctgagccgtattatcacccagttcaacaacatcagctacc
agatgtacgatagcctgaactatcaggcgggtgcgattaaagcgaaaatcgatctggaatacaaaaaata
cagcggcagcgataaagaaaacatcaaaagccaggttgaaaacctgaaaaacagcctggatgtgaaaatt
agcgaagcgatgaataacatcaacaaattcatccgcgaatgcagcgtgacctacctgttcaaaaacatgc
tgccgaaagtgatcgatgaactgaacgaatttgatcgcaacaccaaagcgaaactgatcaacctgatcga
tagccacaacattattctggtgggcgaagtggataaactgaaagcgaaagttaacaacagcttccagaac
accatcccgtttaacatcttcagctataccaacaacagcctgctgaaagatatcatcaacgaatacttca
atctagaaggtggcggtgggtccggtggcggtggctcaggcgggggcggtagcgcactagacaactctga
ctctgaatgcccgctgtctcacgacggttactgcctgcacgacggtgtttgcatgtacatcgaagctctg
gacaaatacgcttgcaactgcgttgttggttacatcggtgaacgttgccagtaccgtgacctgaaatggt
gggaactgcgt

SEQ ID 23

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVI
EGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL
YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIP
DNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVEDF
TFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYI
GPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKRWK
DSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDVKI
SEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSFQN
TIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIEAL
DKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 24

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISIVPRFSLTYSNATNDVG
EGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL
YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIY
YPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVE
DFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIP
YIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKR

WKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDV
KISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIE
ALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 25

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVG
TPRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL
YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIY
YPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVE
DFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIP
YIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKR
WKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKSQVENLKNSLDV
KISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIE
ALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 26

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVG
EGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL
YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIY
YPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVE
DFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIP
YIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKR
WKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDGRYKKYSGSDKENIKSQVENLKNSLDV
KISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIE
ALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 27

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVG
EGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL

YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIY
YPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVE
DFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIP
YIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKR
WKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYIDGRKENIKSQVENLKNSLDV
KISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIE
ALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 28

MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPNLNKPPRVTSPKSG
YYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRLSTDIPFPGNNNTPINTFDFDVDFNSVDVKT
RQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNNTFAAQEGFGALSIISISPRFMLTYSNATNDVG
EGRFSKSEFCMDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDL
IPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFVVESSGEVTVNRNKFVEL
YNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTANILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPA
LRKVNPENMLYLFTKFCVDADDDDKLYNKTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIY
YPDNVSVDQVILSKNTSEHGQLDLLYPSIDSESEILPGENQVFYDNRTQNVDYLNSYYYLESQKLSDNVE
DFTFTRSIEEALDNSAKVYTYFPTLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIP
YIGPALNISNSVRRGNFTEAFAVTGVTILLEAFPEFTIPALGAFVIYSKVQERNEIIKTIDNCLEQRIKR
WKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYGVPRKENIKSQVENLKNSLDV
KISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVDKLKAKVNNSF
QNTIPFNIFSYTNNSLLKDIINEYFNLEGGGGSGGGGSGGGGSALDNSDSECPLSHDGYCLHDGVCMYIE
ALDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 29

MTWPVKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPSLSKPPRPTSKYQS
YYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYLVVGSPFMGDSSTPEDTFDFTRHTTNIAVEK
FENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQGQQSNPSFEGFGTLSILKVAPEFLLTFSDVTSNQS
SAVLGKSIFCMDPVIALMHELTHSLHQLYGINIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEI
IPQIERSQLREKALGHYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFDKDNTGNFVVNIDKFNSL
YSDLTNVMSEVVYSSQYNVKNRTHYFSRHYLPVFANILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPA
LQKLSSESVVDLFTKVCVDGGGGSADDDDKHSDAVFTDNYTRLRRQLAVRRYLNSILNALAGGGGSGGGG
SGGGGSALALQCIKVKNNRLPYVADKDSISQEIFENKIITDETNVQNYSDKFSLDESILDGQVPINPEIV
DPLLPNVNMEPLNLPGEEIVFYDDITKYVDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFL
PSLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAFA
TAGVAFLLEGFPEFTIPALGVFTFYSSIQEREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNH
INYQMYDSLSYQADAIKAKIDLEYKKYIDGRKENIKSQVENLKNSLDVKISEAMNNINKFIRECSVTYLF
KNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVDRLKAKVNESFENTMPFNIFSYTNNSLLKDIIN
EYFNLEA

SEQ ID 30

MEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQVPV
SYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRGLLTSIVRGIPFWGGSTIDTELKVIDTNCINVIQP
DGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGA
GKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQEN
EFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT
EDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT
GLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAA
EENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFE
HGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADI
TIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALS
KRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLS
SKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
VNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDQYCLHDGVCMYIEA
LDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 31

MEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQVPV
SYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRGLLTSIVRGIPFWGGSTIDTELKVIDTNCINVIQP
DGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGA
GKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQEN
EFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLTEIYT
EDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFT
GLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAA
EENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRAQEFE
HGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADI
TIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTIDNALS
KRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNIDDLS
SKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDRLKDK
VNNTLSTDIPFQLSKYVDNQRLLSTLEGGGGSGGGGSGGGGSALDNSDSECPLSHDQYCLHDGVCMYIEA
LDKYACNCVVGYIGERCQYRDLKWWELR

SEQ ID 32

HMGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAK
QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRGLLTSIVRGIPFWGGSTIDTELKVIDTNCIN
VIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNP
LLGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDS
LQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLT
EIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKL
KNFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTN
IEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRA
QEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDK
IADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTID
NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNI

DDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDR
LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKV
NFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGW
KVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL
GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDKPY
YMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY
INVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIG
FHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHHHHHHHHHH

SEQ ID 33

HMGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAK
QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCIN
VIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNP
LLGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDG
RQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLT
EIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKL
KNFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTN
IEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRA
QEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDK
IADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTID
NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNI
DDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDR
LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKV
NFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGW
KVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL
GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDKPY
YMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY
INVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIG
FHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHHHHHHHHHH

SEQ ID 34

HMGSMEFVNKQFNYKDPVNGVDIAYIKIPNAGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAK
QVPVSYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCIN
VIQPDGSYRSEELNLVIIGPSADIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNP
LLGAGKFATDPAVTLAHELIHAGHRLYGIAINPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDS
LQENEFRLYYYNKFKDIASTLNKAKSIVGTTASLQYMKNVFKEKYLLSEDTSGKFSVDKLKFDKLYKMLT
EIYTEDNFVKFFKVLNRKTYLNFDKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKL
KNFTGLFEFYKLLCVDGIITSKTKSDDDDKNKALNLQCIKVNNWDLFFSPSEDNFTNDLNKGEEITSDTN
IEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERFPNGKKYELDKYTMFHYLRA
QEFEHIEGRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDK
IADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVSYIANKVLTVQTID
NALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNINFNI
DDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVDR
LKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSLEILNLRYESNHLIDLSRYASKINIGSKV

NFDPIDKNQIQLFNLESSKIEVILKNAIVYNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGW
KVSLNYGEIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNNRLNNSKIYINGRLIDQKPISNL
GNIHASNNIMFKLDGCRDTHRYIWIKYFNLFDKELNEKEIKDLYDNQSNSGALVILKDFWGDYLQYDKPY
YMLNLYDPNKYVDVNNVGIRGYMYLKGPRGSVMTTNIYLNSSLYRGTKFIIKKYASGNKDNIVRNNDRVY
INVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGNDIGFIG
FHQFNNIAKLVASNWYNRQIERSSRTLGCSWEFIPVDDGWGERPLALAHHHHHHHHHH

SEQ ID 35

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSSYYD
PNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFSNGS
QHILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTIVPRFSFRFNDNSINEFIQDPALTL
MHELIHSLHGLYGAKGITTTCIITQQQNPLITNRKGINIEEFLTFGGNDLNIITVAQYNDIYTNLLNDYR
KIASKLSKVQVSNPQLNPYKDIFQEKYGLDKDASGIYSVNINKFDDILKKLYSFTEFDLATKFQVKCRET
YIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIIKPITGRGLVKKIIRFCKNIVSVKG
IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNL
TIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFI
NNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGA
GILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKR
KEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYL
MKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF
NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDN
KYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNA
NGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIF
DKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANR
LYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVM
NSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK

SEQ ID 36

MPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSSYYD
PNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFSNGS
QHILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSINEFIQDPALTL
MHELIHSLHGLYGAKGITTTCIITQQQNPLITNRKGINIEEFLTFGGNDLNIITVAQYNDIYTNLLNDYR
KIASKLSKVQVSNPQLNPYKDIFQEKYGLDKDASGIYSVNINKFDDILKKLYSFTEFDLATKFQVKCRET
YIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIIKPITGRGLVKKIIRFCKNIVSVKG
IRKSICIEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNL
TIQNDAYIPKYDSNGTSDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFI
NNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGA
GILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKR
KEQMYQALQNQVNAIKTIIESKYNSYIEGRKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYL
MKLINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYF
NKFFKRIKSSSVLNMRYKNDKYVDTSGYDSNININGDVYKYPTNKNQFGIYNDKLSEVNISQNDYIIYDN
KYKNFSISFWVRIPNYDNKIVNVNNEYTIINCMRDNNSGWKVSLNHNEIIWTLQDNAGINQKLAFNYGNA
NGISDYINKWIFVTITNDRLGDSKLYINGNLIDQKSILNLGNIHVSDNILFKIVNCSYTRYIGIRYFNIF

DKELDETEIQTLYSNEPNTNILKDFWGNYLLYDKEYYLLNVLKPNNFIDRRKDSTLSINNIRSTILLANR
LYSGIKVKIQRVNNSSTNDNLVRKNDQVYINFVASKTHLFPLYADTATTNKEKTIKISSSGNRFNQVVVM
NSVGNNCTMNFKNNNGNNIGLLGFKADTVVASTWYYTHMRDHTNSNGCFWNFISEEHGWQEK

SEQ ID 37

MGSMPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSS
YYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGRGLLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFS
NGSQHILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSINEFIQDPA
LTLMHELIHSLHGLYGAKGITTTCIITQQQNPLITNRKGINIEEFLTFGGNDLNIITVAQYNDIYTNLLN
DYRKIASKLSKVQVSNPQLNPYKDIFQEKYGLDKDASGIYSVNINKFDDILKKLYSFTEFDLATKFQVKC
RETYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIIKPITGRGLVKKIIRFCVDGGG
GSADDDDKHSDAVFTDNYTRLRRQLAVRRYLNSILNALAGGGGSGGGGSGGGGSALVLQCIEINNGELFF
VASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGT
SDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWI
QQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTIL
VFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIK
TIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYD
ENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKG

SEQ ID 38

MGSMPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSS
YYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFS
NGSQHILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSINEFIQDPA
LTLMHELIHSLHGLYGAKGITTTCIITQQQNPLITNRKGINIEEFLTFGGNDLNIITVAQYNDIYTNLLN
DYRKIASKLSKVQVSNPQLNPYKDIFQEKYGLDKDASGIYSVNINKFDDILKKLYSFTEFDLATKFQVKC
RETYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIIKPITGRGLVKKIIRFCVDGGG
GSADDDDKHSDAVFTDNYTRLRRQLAVRRYLNSILNALAGGGGSGGGGSGGGGSALVLQCIEINNGELFF
VASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGT
SDIEQHDVNELNVFFYLDAQKVPEIEGRVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWI
QQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTIL
VFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIK
TIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYD
ENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKG

SEQ ID 39

MGSMPKINSFNYNDPVNDRTILYIKPGGCQEFYKSFNIMKNIWIIPERNVIGTTPQDFHPPTSLKNGDSS
YYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEELSKANPYLGNDNTPDNQFHIGDASAVEIKFS
NGSQHILLPNVIIMGAEPDLFETNSSNISLRNNYMPSNHGFGSIAIVTFSPEYSFRFNDNSINEFIQDPA
LTLMHELIHSLHGLYGADGITTTCIITQQQNPLITNRKGINIEEFLTFGGNDLNIITIEGRNDIYTNLLN
DYRKIASKLSKVQVSNPQLNPYKDIFQEKYGLDKDASGIYSVNINKFDDILKKLYSFTEFDLATKFQVKC
RETYIGQYKYFKLSNLLNDSIYNISEGYNINNLKVNFRGQNANLNPRIIKPITGRGLVKKIIRFCVDGGG
GSADDDDKHSDAVFTDNYTRLRRQLAVRRYLNSILNALAGGGGSGGGGSGGGGSALVLQCIEINNGELFF
VASENSYNDDNINTPKEIDDTVTSNNNYENDLDQVILNFNSESAPGLSDEKLNLTIQNDAYIPKYDSNGT

SDIEQHDVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWI
QQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDALELLGAGILLEFEPELLIPTIL
VFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIK
TIIESKYNSYTLEEKNELTNKYDIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYD
ENVKTYLLNYIIQHGSILGESQQELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFKG

CLAIMS:

1. A polypeptide, comprising:

a. a non-cytotoxic protease that is capable of cleaving a SNARE protein;

b. a translocation domain that is capable of translocating the non-cytotoxic protease from within an endosome of a mammalian cell, across the endosomal membrane thereof and into the cytosol of the mammalian cell;

c. a first destructive cleavage site that is cleavable by a second protease and not by the non-cytotoxic protease, and wherein after cleavage thereof by the second protease the polypeptide has reduced potency;

d. a Targeting Moiety (TM) that binds to a Binding Site on a nerve cell of the neuromuscular junction, which Binding Site is capable of undergoing endocytosis to be incorporated into an endosome within the nerve cell; and

e. with the proviso that said first destructive cleavage site is not located within said TM.

2. A polypeptide according to Claim 1, wherein the reduced potency is as measurable by a reduced ability to cleave said SNARE protein and/ or a reduced ability to translocate said non-cytotoxic protease.

3. A polypeptide according to Claim 1 or 2, wherein the destructive cleavage site is cleaved by a protease selected from a circulating protease; a tissue-associated protease; or an intracellular protease.

4. A polypeptide according to Claim 3, wherein the circulating protease is an extracellular protease and the tissue-associated protease is a matrix metalloprotease (MMP).

5. A polypeptide according to claim 5, wherein the extracellular protease is a serum protease or a protease of the blood clotting cascade and the matrix metalloprotease (MMP) is a muscle MMP.

6. A polypeptide according to Claim 3, wherein the intracellular protease is a protease that is absent from the target cell.

7. A polypeptide according to any one of Claims 3 to 6, wherein the protease is Thrombin, Factor Xa, ADAM17, Human airway trypsin-like protease (HAT), ACE (peptidyl-dipeptidase A), Elastase, Furin, Granzyme, Caspase, a Matrix metalloprotease (MMP), a TACE, an adamalysin, a serralysin, a astacin, Coagulation Factor VIIa, Coagulation Factor IXa, Coagulation Factor XIa, Coagulation Factor XIIa, Kallikrein, Protein C, or MBP-associated serine protease.

8. A polypeptide according to Claim 7, wherein the caspase is one of Caspase 1-10.

9. A polypeptide according to any preceding claim, wherein the non- cytotoxic protease comprises a clostridial neurotoxin L-chain or a fragment thereof that is capable of cleaving a SNARE protein; or wherein the non-cytotoxic protease comprises an IgA protease or a fragment thereof that is capable of cleaving a SNARE protein.

10. A polypeptide according to any of Claims 1 to 9, wherein the polypeptide comprises at least one first destructive cleavage site present in the non-cytotoxic protease component and/ or at least one first destructive cleavage site present in the translocation domain.

11. A polypeptide according to any preceding claim, wherein the translocation domain comprises a clostridial neurotoxin translocation domain.

12. A polypeptide according to any preceding Claim, wherein the TM comprises a clostridial neurotoxin Hc or Hcc domain or a fragment thereof that is capable of binding to a neuronal cell; or wherein the TM comprises a peptide selected from a glucagon like hormone, a neurohormone, a neuroregulatory cytokine, a

neurotrophin, a growth factor, an axon guidance signaling molecule, a sugar binding protein, a ligand that selectively binds a neurexin, a ligand for neurexin-2α, a ligand for neurexin-2β, a ligand for neurexin-3α, a ligand for neurexin-3β, a WNT, Ng-CAM(LI), NCAM, N-cadherin, a PACAP peptide such as a VIP peptide, Agrin-MUSK, a basement membrane polypeptide, and a variant of any of the foregoing polypeptides; a neuroregulatory cytokine such as a ciliary neurotrophic factor (CNTF), a glycophorin-A (GPA), a leukemia inhibitory factor (LIF), an interleukin (IL), an onostatin M, a cardiotrophin-1 (CT-1), a cardiotrophin-like cytokine (CLC), a neuroleukin, VEGF, an insulin-like growth factor (IGF), an epidermal growth factor (EGF), and a variant of any of the foregoing neuroregulatory cytokines.

13. A polypeptide according to any preceding claim, wherein after cleavage at the destructive cleavage site, the polypeptide has a reduced ability to cleave a SNARE protein.

14. A polypeptide according to Claim 1, wherein said polypeptide comprises an amino acid sequence having at least 80% sequence identity with an amino acid sequence selected from: SEQ ID NO: 4, 6, 8, 10, 11, 13, 15, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31 , 32, 33, 34, 35, 36, 37, 38, or 39.

15. A polypeptide according to claim 1, wherein said polypeptide comprises an amino acid sequence having at least 90% sequence identity with an amino acid sequence selected from SEQ ID NO: 4, 6, 8, 10, 11, 13, 15, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, or 39.

16. A polypeptide according to claim 1, wherein said polypeptide comprises an amino acid sequence having at least 95% sequence identity with an amino acid sequence selected from SEQ ID NO: 4, 6, 8, 10, 11, 13, 15, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, or 39.

17. A nucleic acid encoding a polypeptide according to any preceding claim.

18. A nucleic acid according to Claim 17, wherein the nucleic acid comprises a nucleic acid sequence having at least 80% sequence identity to a nucleic acid sequence selected from SEQ ID NO: 3, 5, 7, 9, 12, 14, 18, 20, or 22.

2009340404

19. A nucleic acid according to Claim 17, wherein the nucleic acid comprises a nucleic acid sequence having at least 90% sequence identity to a nucleic acid sequence selected from SEQ ID NO: 3, 5, 7, 9, 12, 14, 18, 20, or 22.

20. A nucleic acid according to Claim 17, wherein the nucleic acid comprises a nucleic acid sequence having at least 95% sequence identity to a nucleic acid sequence selected from SEQ ID NO: 3, 5, 7, 9, 12, 14, 18, 20, or 22.

21. A polypeptide according to any of Claims 1 to 16 for use in suppressing a condition or disease selected from strabismus, blepharospasm, squint, dystonia, beauty therapy (cosmetic) applications benefiting from cell/ muscle incapacitation (via SNARE down-regulation or inactivation), neuromuscular disorder or condition of ocular motility, writer's cramp, blepharospasm, bruxism, Wilson's disease, tremor, tics, segmental myoclonus; spasms, spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, animus, back spasm, Charley horse, tension headaches, levator pelvic syndrome, spina bifida, tardive dyskinesia, Parkinson's and stuttering.

22. A method for suppressing a condition or disease selected from strabismus, blepharospasm, squint, dystonia, torticollis, beauty therapy (cosmetic) applications benefiting from cell/ muscle incapacitation (via SNARE down-regulation or inactivation), neuromuscular disorder or condition of ocular motility, writer's cramp, blepharospasm, bruxism, Wilson's disease, tremor, tics, segmental myoclonus; spasms, spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, animus, back spasm, Charley horse, tension headaches, levator pelvic syndrome, spina bifida, tardive dyskinesia, Parkinson's and stuttering;

said method comprising administering to a patient an effective amount of a polypeptide according to any of Claims 1 to 16.

23. Use of a polypeptide according to any one of claims 1 to 16 in the manufacture of a medicament for use in suppressing a condition or disease selected from strabismus, blepharospasm, squint, dystonia, beauty therapy (cosmetic) applications benefiting from cell/ muscle incapacitation (via SNARE down-

regulation or inactivation), neuromuscular disorder or condition of ocular motility, writer's cramp, blepharospasm, bruxism, Wilson's disease, tremor, tics, segmental myoclonus; spasms, spasticity due to chronic multiple sclerosis, spasticity resulting in abnormal bladder control, animus, back spasm, Charley horse, tension headaches, levator pelvic syndrome, spina bifida, tardive dyskinesia, Parkinson's and stuttering.

24. A polypeptide according to Claim 21, method according to Claim 22 or use according to claim 23, wherein the dystonia is selected from the group consisting of spasmodic dystonia, oromandibular dystonia, focal dystonia, tardive dystonia, laryngeal dystonia and limb dystonia.

25. A polypeptide according to Claim 21, method according to Claim 22 or use according to claim 23, wherein the torticollis is spasmodic torticollis.

26. A polypeptide according to Claim 21, method according to Claim 22 or use according to claim 23, wherein the neuromuscular disorder or condition of ocular motility is selected from the group consisting of concomitant strabismus, vertical strabismus, lateral rectus palsy and nystagmus dysthyroid myopathy.

Figure 1

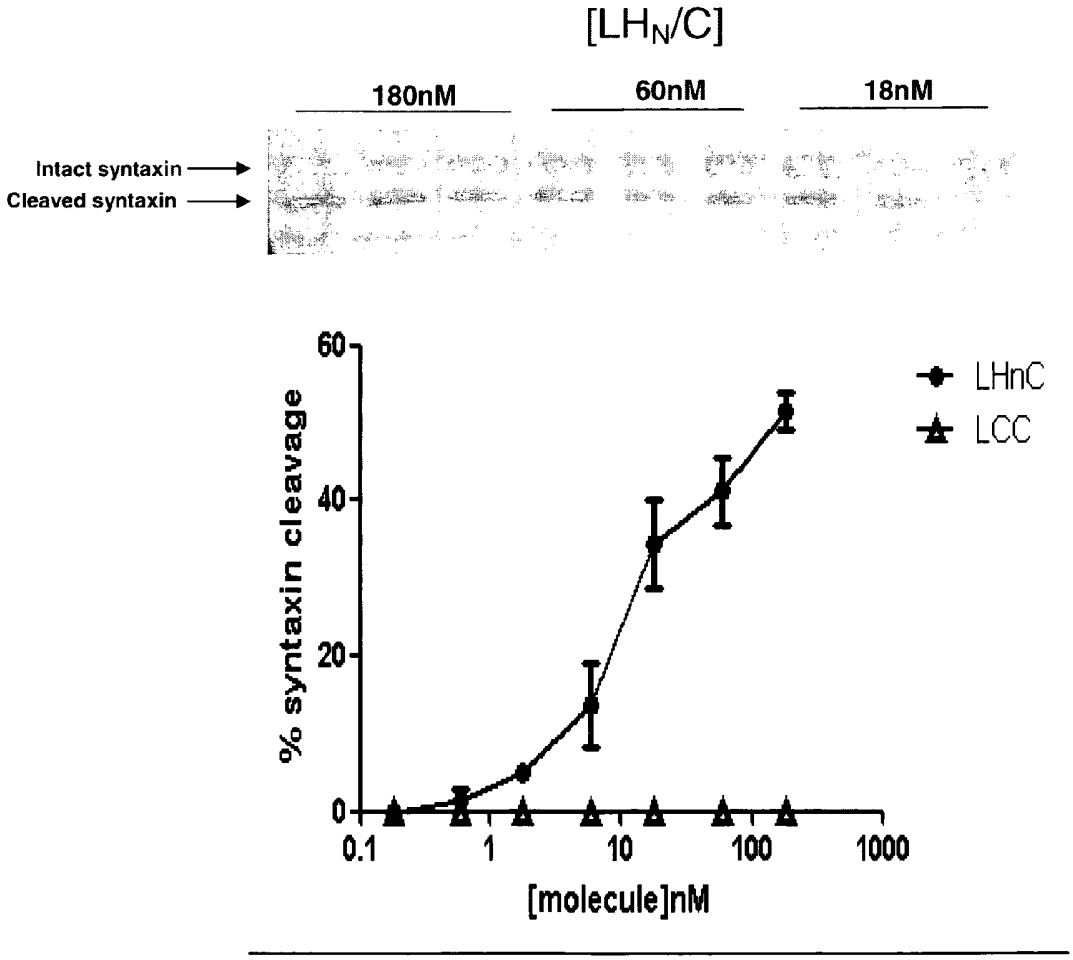

[LHN/C]
180nM
60nM
18nM
Intact syntaxin
Cleaved syntaxin
% syntaxin cleavage
60
40
20
0
0.1
1
10
100
1000
[molecule]nM
LHnC
LCC

Figure 2

WO 2010/094905 PCT/GB2009/002892

```
LHA : EDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKIN-IVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVDGII- : 435
LHB : EDSEGKYSIDVESFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAYEEIS-KEHLAVYKIQMCVDEEK- : 442
LHC : VESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTAN-ILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNP--ENMLYLFTKFCVDAID- : 442
LHD : KDNTGNFVVNIDKFNSLYSDLTNVMSEVVYSSQYNVKNRTHYFSRHYLPVFAN-ILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKLSS--ESVVDLFTKVCVDKSEE : 443
LHE : KDASGIYSVNINKFNDIFKKLYS-FTEFDLRTKFQVKCRQTYIGQYKY-FKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNPRIITPIT-GRGLVKKIIRFCVD---- : 414
LHF : QDSNGNYTVNISKFNAIYKKLFS-FTECDLAQKFQVKNRSNYLFHFKP-FRLLDLLDDNIYSISEGFNIG--SLRVNNNGQNINLNSRIVGPIP-DNGLVERFVGLCVD---- : 423
LHG : EDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEIS-LEHLVIYRIAMCKP--V- : 439
      d  G   61  kF  65  L    3E     5 6  R  Y           66    Y i  G5N6         GQN   N            6       Cvd


LHA : -TSKTKSLIEGRNKALNLQC--IKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERF--PNGKKY : 543
LHB : -LYDDDDKDRWGS---SLQC--IDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPIN---ELILDTDLISKIE-LPSENTESLTDFN-VDVPVYEKQ--PAIKKI : 542
LHC : ------GRSLYN---KTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS-KNTSEHGQLDLLYPSIDSESE---ILPGENQVFYDNRTQ : 539
LHD : KLYDDDDKDRWG---SSLQC--IKVKNNRLPYVADKDSISQEIFENKIITDETNVQNYSDKFSLD---ESILDGQVPINPEIVDPLLPNVNMEPL---NLPGEEIVFYDDITK : 545
LHE : -IEGRKGIR-------KLQC--IEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLD---QVILNFNSESAPG-LSDEKLNLTIQND-AYIPKYDSNGTSDIEQH : 512
LHF : -IEGRKGTK-------NLQC--IKVNNRDLFFVASESSYNENGINSPKEIDDTTITNNNYKKNLD---EVILDYNSDAIPN-LSSRLLNTTAQND-SYVPKYDSNGTSEIKEY : 521
LHG : -MYKNTGK--------SEQC--IIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSID---QLILDNDLSSGID-LPNENTEPFTNFDDIDIPVYIKQ--SALKKI : 535
                     lQC  6 6 N  L 5                    T          61   2 iL          6


LHA : ELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGAL : 656
LHB : FTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAF : 655
LHC : NVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFP-TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAF : 651
LHD : YVDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLP-SLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAF : 657
LHE : DVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDAL : 625
LHF : TVDKLNVFFYLYAQKAPEGESAISLTSSVNTALLDASKVYTFFSSDFINTVNKPVQAALFISWIQQVINDFTTEATQKSTIDKIADISLIVPYVGLALNIGNEVQKGNFKEAI : 634
LHG : FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAF : 648
      1    5 YL  Q          lT S   AL    46Y35f        N  v a 6F  W   66 DFt       T DK6 D63 66PY6G ALN6gN   4 1F  A


LHA : IFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNIN--F : 764
LHB : EIAGASILLEFIPELLIPVVGAFLLESYID---NKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNIN--I : 763
LHC : AVTGVTILLEAFPEFTIPALGAFVIYSKVQ---ERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKS-- : 759
LHD : ATAGVAFLLEGFPEFTIPALGVFTFYSSIQ---EREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKS-- : 765
LHE : ELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY : 738
LHF : ELLGAGILLEFVPELLIPTILVFTIKSFINSDDSKNKIIKAINNALRERELKWKEVYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDGIKKIIEYKYNNYTLDEKNRLRAEY : 747
LHG : EIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNIN--I : 756
            G  iL6E  PE  6P 6  F   S 6    4   6  I N L  R  4W   Y  66  W6  6 TQf      My  L  Q  a    I   Y  Y3   K  6
```

```
          800         *       820         *       840         *       860         *       880         *       900
LHA : NIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVD-RLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK----- : 871
LHB : DFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAE-YEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNS----- : 870
LHC : QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVD-KLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFN----- : 866
LHD : QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVD-RLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFN----- : 872
LHE : DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQ-ELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK----- : 845
LHF : NIYSIKEELNKKVSLAMQNIDRFLTESSISYLMKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSEYTNDKILIHILIRFYK----- : 855
LHG : DFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVN-ILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYIS----- : 863
          6   L1  6  A6 NI1 F6   S63YL   6    6  L  5D   k  L6  I      L       l   V       6PF 6  Yt 1  L      5

LHA : MEFVNKQFNYKDPVNGVDIAYIKIPN-AGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPPEAKQVPV-SYYDSTYLSTDNEKDNYLKGVTKLFERIYSTDLGRMLLTSI : 111
LHB : MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWIIPERYTFGYKPE-DFNKSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRIKSKPLGEKLLEMI : 112
LHC : MPITINNFNYSDPVDNKNILYLDTHLNTLANEPEKAFRITGNIWVIPDRFSRNSNPN--LNKPPRVTSPKS-GYYDPNYLSTDSDKDTFLKEIIKLFKRINSREIGEELIYRL : 110
LHD : MTWPVKDFNYSDPVNDNDILYLRIPQNKLITTPVKAFMITQNIWVIPERFSSDTNPS--LSKPPRPTSKYQ-SYYDPSYLSTDEQKDTFLKGIIKLFKRINERDIGKKLINYL : 110
LHE : MPK-INSFNYNDPVNDRTILYIKP---GGCQEFYKSFNIMKNIWIIPERNVIGTTPQ-DFHPPTSLKNGDS-SYYDPNYLQSDEEKDRFLKIVTKIFNRINNNLSGGILLEEL : 107
LHF : MPVNINNFNYNDPINNTTILYMKMPYYEDSNKYYKAFEIMDNVWIIPERNIIGKKPS-DFYPPISLDSGSS-AYYDPNYLTTDAEKDRFLKTVIKLFNRINSNPAGQVLLEEI : 111
LHG : MPVNIKNFNYNDPINNDDIIMMEPFNDPGPGTYYKAFRIIDRIWIVPERFTYGFQPD-QFNASTGVFSKDVYEYYDPTYLKTDAEKDKFLKTMIKLFNRINSKPSGQRLLDMI : 112
      MP6T6N1FNYNDP61111I6Y646P514G3N25YKAF4IT1N6W66PER535G3KP2GDFNKP3S6TS4DSCSYYDP1YL331E2K145LKT6IK6FNRINS4P6G24L6EE6

LHA : VRGIPFWGG-STIDTELKVIDTN-CINVIQPD-GSYR---SEELNLVIIGPSADIIQFECKSFGHEV----LNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFA : 214
LHB : INGIPYLGDRRVPLEEFNTNIASVTVNKLISNPGEVERKKGIFANLIIFGPGPVLNENETTDIGIQN----HFASREGFGGIMQMKFCPEYVSVFNNVQENKGASIFNRRGYF : 221
LHC : STDIPFPGNNNTPINTFDFDVDFNSVDVKTRQGNNWVKTGSINPSVIITGPRENIIDPETSTFKLTNN---TFAAQEGFGALSIISISPRFMLTYSNATNDVGEGRFSKSEFC : 220
LHD : VVGSPFMGDSSTPEDTFDFTRHTTNIAVEKFENGSWKVTNIITPSVLIFGPLPNILDYTASLTLQCQQ---SNPSFEGFGTLSILKVAPEFLLTFSDVTSNQSSAVLGKSIFC : 220
LHE : SKANPYLGNDNTPDNQFHIGDAS-AVEIKFSN-GSQD---ILLPNVIIMGAEPDLFETNSSNISLRNN---YMPSNHRFGSIAIVTFSPEYSFRFN---DNC----MN--EFI : 203
LHF : KNGKPYLGNDHTAVNEFCANNRSTSVEIKESN-GTTD---SMLLNLVILGPGPNILECSTFPVRIFPNNIAYDPSEKGFGSIQLMSFSTEYEYAFN---DN-----TD--LFI : 210
LHG : VDAIPYLGNASTPPDKFAANVANVSINKKIIQPGAEDQIKGLMTNLIIFGPGPVLSDNFTDSMIMNG----HSPISEGFGARMMIRFCPSCLNVFNNVQENKDTSIFSRRAYF : 221
      61GIP56G11STP612F1611A3T3616K2S1PG35D4TKS66PN66I6GPGP1662NE3S36462NNNIAY1P32EG5G36S664FSPE565T5N1VT11K33S6614SE5I

          *       240         *       260         *       280         *       300         *       320         *       34
LHA : TDPAVTLAHELIHAGHRLYGIAI-NPNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYYNKFKDIASTLN-K-AKSIVGTTASLQYMKNVFKEKYLLS : 324
LHB : SDPALILMHELIHVLHGLYGIK--VDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVDRLNKV-LVCISDPNININIYKNKFKDKYKFV : 331
LHC : MDPILILMHELNHAMHNLYGIAIPNDQTISSVTSNIFYSQYNVKLEYAEIYAFGGPTIDLIPKSARKYFEEKALDYYRSIAKRLNSITTANPSSFNKYIGEYKQKLIRKYRFV : 333
LHD : MDPVIALMHELTHSLHQLYGINIPSDKRIRPQVSEGFFSQDGPNVQFEELYTFGGLDVEIIPQIERSQLREKALGHYKDIAKRLNNINKTIPSSWISNIDKYKKIFSEKYNFD : 333
LHE : QDPALTLMHELIHSLHGLYGAKG-ITTKYTITQKQNPLITNIRGTNIEEFLTFGGTDLNIITSAQSNDIYTNLLADYKKIASKLS---KVQVSNPLLNP--YKDVFEAKYGLD : 310
LHF : ADPAISLAHELIHVLHGLYGAKG-VTNKKVIEVDQGALMAAEKDIKIEEFITFGGQDLNIITNSTNQKIYVILLSNYTAIASRLS---QVNRNNSALNTTYYKNFFQWKYGLD : 319
LHG : ADPALTLMHELIHVLHGLYGIK--ISNLPITPNTKEFFMQHSDPVQAEELYTFGGHDPSVISPSTDMNIYNKALQNFQDIANRLN-I-VSSAQGSGIDISLYKQIYKNKYDFV : 330
         ADPA63LMHELIHV6HGLYGIKIP63146662N32G556QS341625EE6YTFGGQD616I3SST1K26YEK6L1154DIAS4LNN6NKVS6S13N6161YYK165KEKY1FD
```

```
LHA : EDTSGKFSVDKLKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDKAVFKIN-IVPKVNYTIYDGFNLRNTNLAANFNGQNTEINNMNFTKLKNFTGLFEFYKLLCVDGII- : 435
LHB : EDSEGKYSIDVESFDKLYKSLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIEEGFNISDKDMEKEYRGQNKAINKQAYEEIS-KEHLAVYKIQMCVDEEK- : 442
LHC : VESSGEVTVNRNKFVELYNELTQIFTEFNYAKIYNVQNRKIYLSNVYTPVTAN-ILDDNVYDIQNGFNIPKSNLNVLFMGQNLSRNPALRKVNP--ENMLYLFTKFCVDAID- : 442
LHD : KDNTGNFVVNIDKFNSLYSDLTNVMSEVVYSSQYNVKNRTHYFSRHYLPVFAN-ILDDNIYTIRDGFNLTNKGFNIENSGQNIERNPALQKLSS--ESVVDLFTKVCVDKSEE : 443
LHE : KDASGIYSVNINKFNDIFKKLYS-FTEFDLRTKFQVKCRQTYIGQYKY-FKLSNLLNDSIYNISEGYNIN--NLKVNFRGQNANLNPRIITPIT-GRGLVKKIIRFCVD---- : 414
LHF : QDSNGNYTVNISKFNAIYKKLFS-FTECDLAQKFQVKNRSNYLFHFKP-FRLLDLLDDNIYSISEGFNIG--SLRVNNNGQNINLNSRIVGPIP-DNGLVERFVGLCVD---- : 423
LHG : EDPNGKYSVDKDKFDKLYKALMFGFTETNLAGEYGIKTRYSYFSEYLPPIKTEKLLDNTIYTQNEGFNIASKNLKTEFNGQNKAVNKEAYEEIS-LEHLVIYRIAMCKP--V- : 439
      2D33GK536161KF1K65KKLTS653ET16AKK5K6KNRK3Y6S25KPP646N16611N6Y3I2EG5N631K1646E5NGQN6E6NP265KE63NKEG66E55646CVDAI6E

LHA : -TSKTKSLIEGRNKALNLQC--IKVNNWDLFFSPSEDNFTNDLNKGEEITSDTNIEAAEENISLDLIQQYYLTFNFDNEPENISIENLSSDIIGQLELMPNIERF--PNGKKY : 543
LHB : -LYDDDDKDRWGS---SLQC--IDVDNEDLFFIADKNSFSDDLSKNERIEYNTQSNYIENDFPIN---ELILDTDLISKIE-LPSENTESLTDFN-VDVPVYEKQ--PAIKKI : 542
LHC : ------GRSLYN---KTLQCRELLVKNTDLPFIGDISDVKTDIFLRKDINEETEVIYYPDNVSVD---QVILS-KNTSEHGQLDLLYPSIDSESE---ILPGENQVFYDNRTQ : 539
LHD : KLYDDDDKDRWG---SSLQC--IKVKNNRLPYVADKDSISQEIFENKIITDETNVQNYSDKFSLD---ESILDGQVPINPEIVDPLLPNVNMEPL---NLPGEEIVFYDDITK : 545
LHE : -IEGRKGIR-------KLQC--IEINNGELFFVASENSYNDDNINTPKEIDDTVTSNNNYENDLD---QVILNFNSESAPG-LSDEKLNLTIQND-AYIPKYDSNGTSDIEQH : 512
LHF : -IEGRKGTK-------NLQC--IKVNNRDLFFVASESSYNENGINSPKEIDDTTITNNNYKKNLD---EVILDYNSDAIPN-LSSRLLNTTAQND-SYVPKYDSNGTSEIKEY : 521
LHG : -MYKNTGK--------SEQC--IIVNNEDLFFIANKDSFSKDLAKAETIAYNTQNNTIENNFSID---QLILDNDLSSGID-LPNENTEPFTNFDDIDIPVYIKQ--SALKKI : 535
      K6YD1KG44R5GNKAK3LQCRE6K61NEDLF56A1K1S53116FK3E4ITD1T263NYE11FS61LIQ26IL15163SEPEQ6SSENLN3132N1D6Y6PKYE4QGFS164KY

LHA : ELDKYTMFHYLRAQEFEHGKSRIALTNSVNEALLNPSRVYTFFSSDYVKKVNKATEAAMFLGWVEQLVYDFTDETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGAL : 656
LHB : FTDENTIFQYLYSQTFPLDIRDISLTSSFDDALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNVGNETAKGNFENAF : 655
LHC : NVDYLNSYYYLESQKLSDNVEDFTFTRSIEEALDNSAKVYTYFP-TLANKVNAGVQGGLFLMWANDVVEDFTTNILRKDTLDKISDVSAIIPYIGPALNISNSVRRGNFTEAF : 651
LHD : YVDYLNSYYYLESQKLSNNVENITLTTSVEEALGYSNKIYTFLP-SLAEKVNKGVQAGLFLNWANEVVEDFTTNIMKKDTLDKISDVSVIIPYIGPALNIGNSALRGNFNQAF : 657
LHE : DVNELNVFFYLDAQKVPEGENNVNLTSSIDTALLEQPKIYTFFSSEFINNVNKPVQAALFVSWIQQVLVDFTTEANQKSTVDKIADISIVVPYIGLALNIGNEAQKGNFKDAL : 625
LHF : TVDKLNVFFYLYAQKAPEGESAISLTSSVNTALLDASKVYTFFSSDFINTVNKPVQAALFISWIQQVINDFTTEATQKSTIDKIADISLIVPYVGLALNIGNEVQKGNFKEAI : 634
LHG : FVDGDSLFEYLHAQTFPSNIENLQLTNSLNDALRNNNKVYTFFSTNLVEKANTVVGASLFVNWVKGVIDDFTSESTQKSTIDKVSDVSIIIPYIGPALNVGNETAKENFKNAF : 648
      561KLN655YLEAQK6PE16E163LT3S61EALL1SN46Y35FS3156NKVNKPV2AA6F6SW622661DFT3EA32KST6DK6AD63666PY6GPALN6GNE6Q4G1FK1AF

LHA : IFSGAVILLEFIPEIAIPVLGTFALVSYIA---NKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKKMKEALENQAEATKAIINYQYNQYTEEEKNNIN--F : 764
LHB : EIAGASILLEFIPELLIPVVGAFLLESYID---NKNKIIKTIDNALTKRNEKWSDMYGLIVAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNIN--I : 763
LHC : AVTGVTILLEAFPEFTIPALGAFVIYSKVQ---ERNEIIKTIDNCLEQRIKRWKDSYEWMMGTWLSRIITQFNNISYQMYDSLNYQAGAIKAKIDLEYKKYSGSDKENIKS-- : 759
LHD : ATAGVAFLLEGFPEFTIPALGVFTFYSSIQ---EREKIIKTIENCLEQRVKRWKDSYQWMVSNWLSRITTQFNHINYQMYDSLSYQADAIKAKIDLEYKKYSGSDKENIKS-- : 765
LHE : ELLGAGILLEFEPELLIPTILVFTIKSFLGSSDNKNKVIKAINNALKERDEKWKEVYSFIVSNWMTKINTQFNKRKEQMYQALQNQVNAIKTIIESKYNSYTLEEKNELTNKY : 738
LHF : ELLGAGILLEFVPELLIPTILVFTIKSFINSDDSKNKIIKAINNALRERELKWKEVYSWIVSNWLTRINTQFNKRKEQMYQALQNQVDGIKKIIEYKYNNYTLDEKNRLRAEY : 747
LHG : EIGGAAILMEFIPELIVPIVGFFTLESYVG---NKGHIIMTISNALKKRDQKWTDMYGLIVSQWLSTVNTQFYTIKERMYNALNNQSQAIEKIIEDQYNRYSEEDKMNIN--I : 756
      E6AGAAIL6EF6PE666P66GVFT6KS561SDDN4NK66KTI1NAL42R124WKD6YS5663NW6346NTQF1KI4EQMY2ALNNQA1A6KAIIEY2YN4Y3EEEKNN6NSE5
```

```
LHA : NIDDLSSKLNESINKAMININKFLNQCSVSYLMNSMIPYGVKRLEDFDASLKDALLKYIYDNRGTLIGQVD-RLKDKVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIK----- : 871
LHB : DFNDINSKLNEGINQAIDNINNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAE-YEKSKVNKYLKTIMPFDLSIYTNDTILIEMFNKYNS----- : 870
LHC : QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNEFDRNTKAKLINLIDSHNIILVGEVD-KLKAKVNNSFQNTIPFNIFSYTNNSLLKDIINEYFN----- : 866
LHD : QVENLKNSLDVKISEAMNNINKFIRECSVTYLFKNMLPKVIDELNKFDLRTKTELINLIDSHNIILVGEVD-RLKAKVNESFENTMPFNIFSYTNNSLLKDIINEYFN----- : 872
LHE : DIKQIENELNQKVSIAMNNIDRFLTESSISYLMKIINEVKINKLREYDENVKTYLLNYIIQHGSILGESQQ-ELNSMVTDTLNNSIPFKLSSYTDDKILISYFNKFFK----- : 845
LHF : NIYSIKEELNKKVSLAMQNIDRFLTESSISYLMKLINEAKINKLSEYDKRVNQYLLNYILENSSTLGTSSVPELNNLVSNTLNNSIPFELSEYTNDKILIHILIRFYK----- : 855
LHG : DFNDIDFKLNQSINLAINNIDDFINQCSISYLMNRMIPLAVKKLKDFDDNLKRDLLEYIDTNELYLLDEVN-ILKSKVNRHLKDSIPFDLSLYTKDTILIQVFNNYIS----- : 863
      16116KNKL12K6S2A61NI14F6N2CS63YLMK466P6A614L4E5D4N6K41L6NYID2NN6IL6G2V1P4LKSKVN13LK136PF16SSYT1136L626FNE55K
```

Figure 3

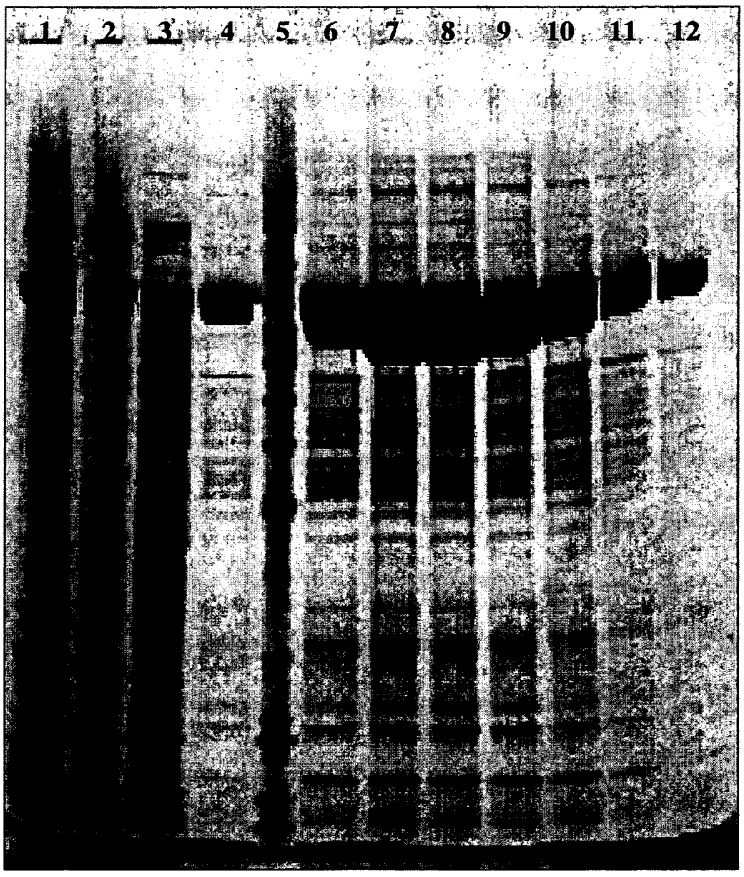


Figure 4

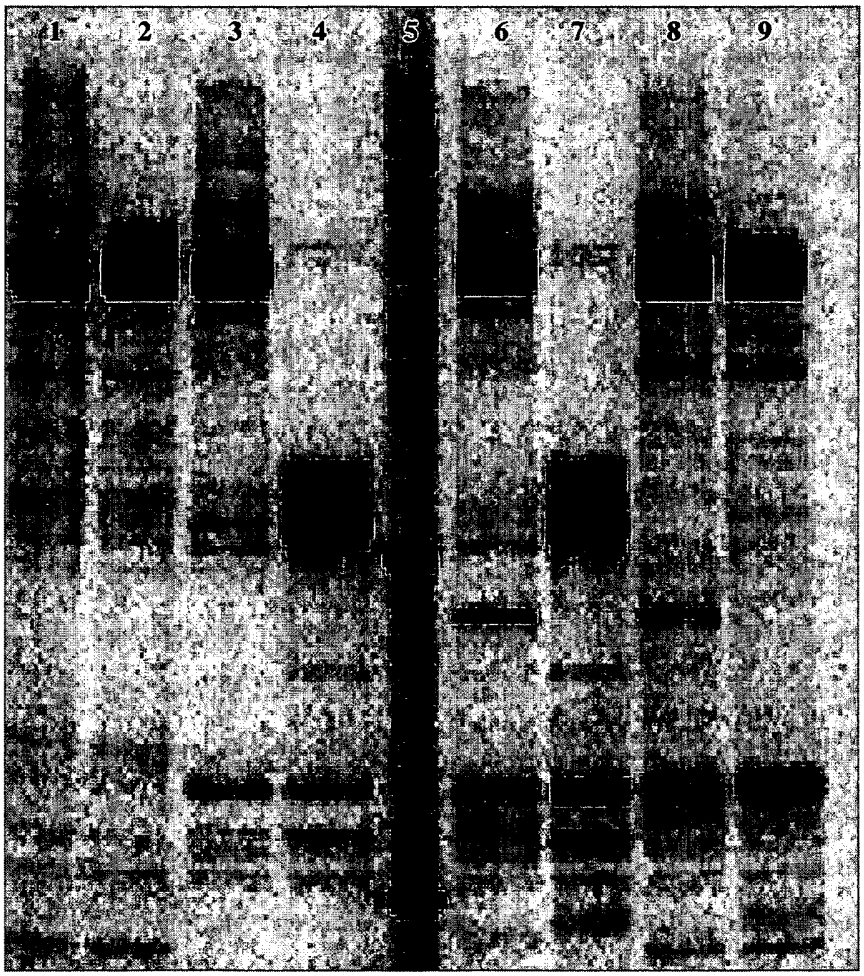


Figure 5

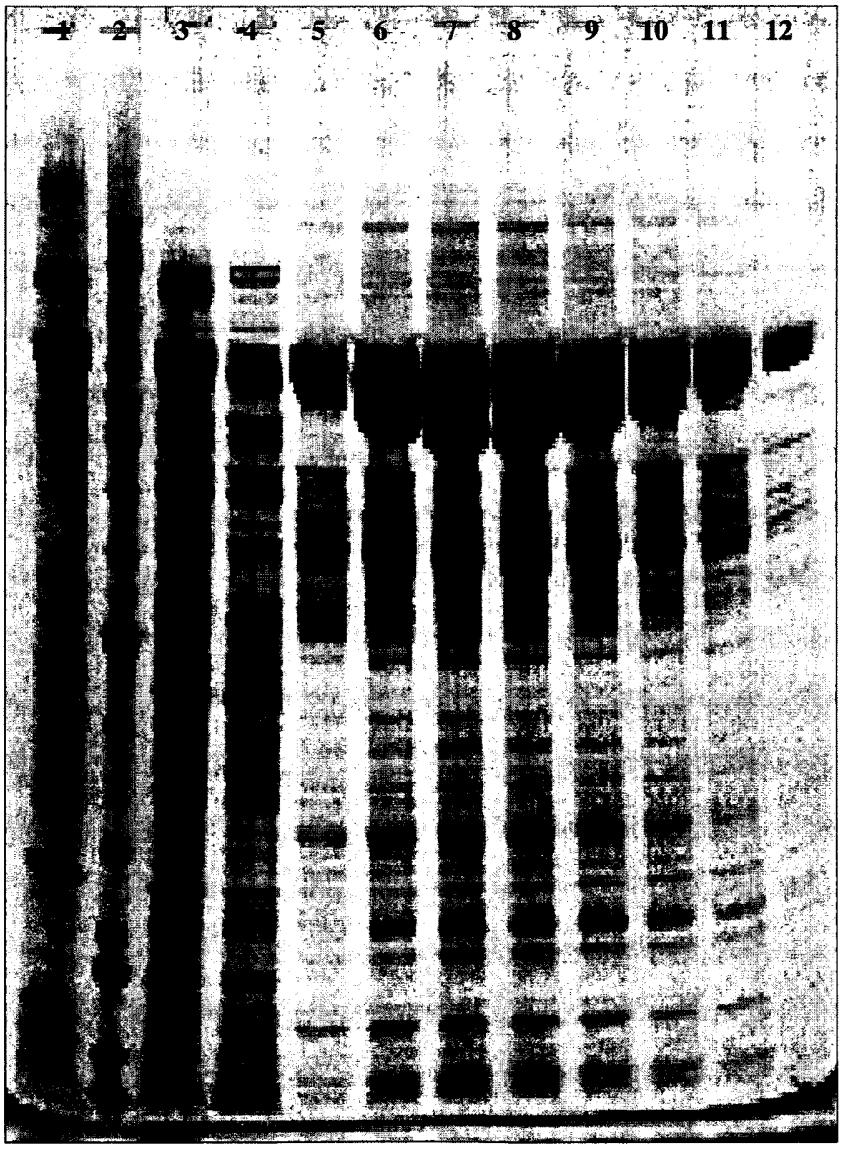


Figure 6

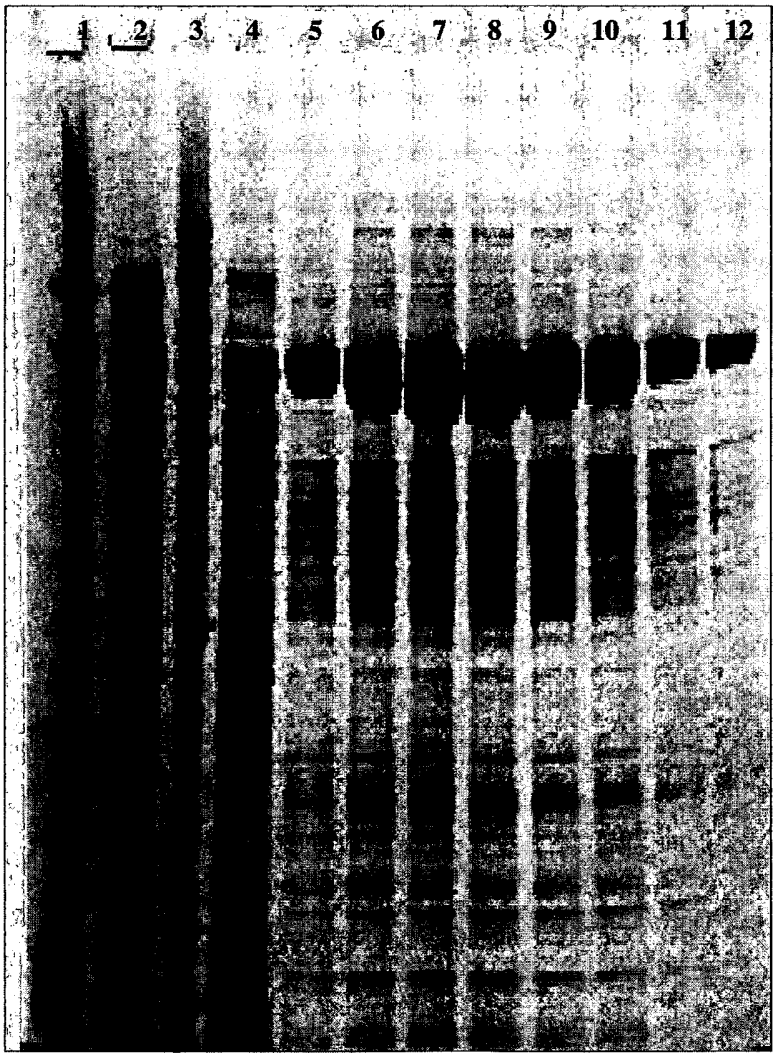


Figure 7

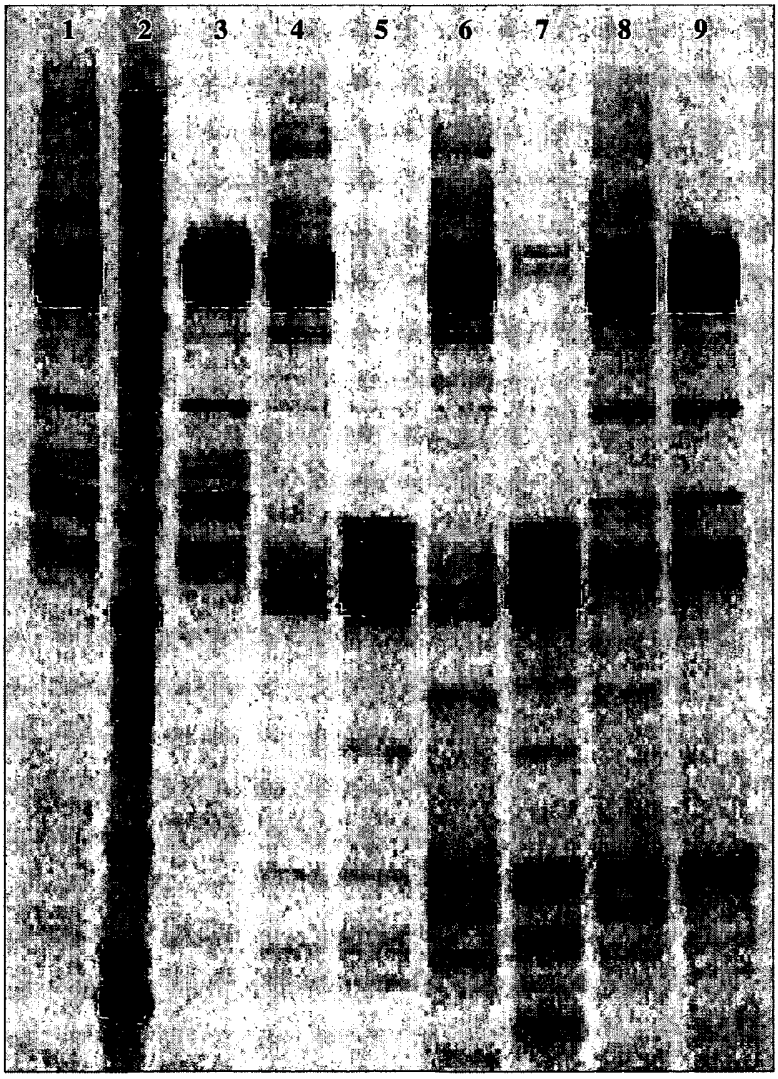


Figure 8

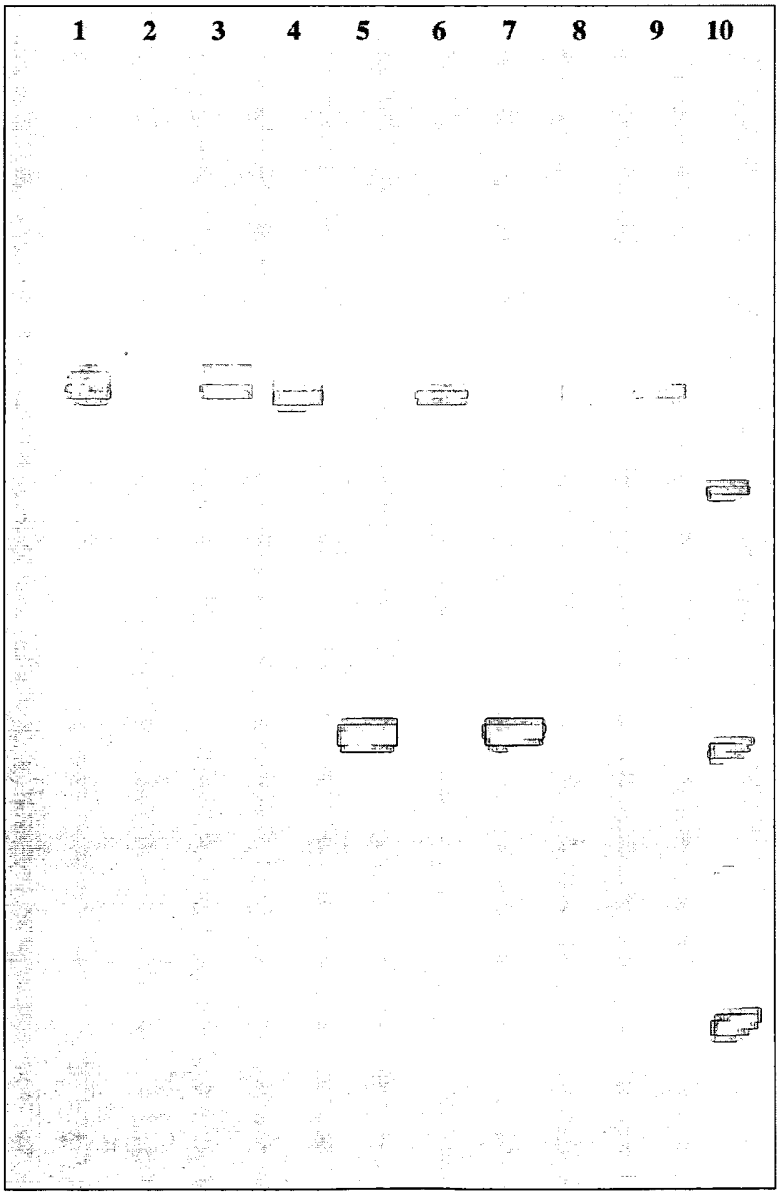


Figure 9

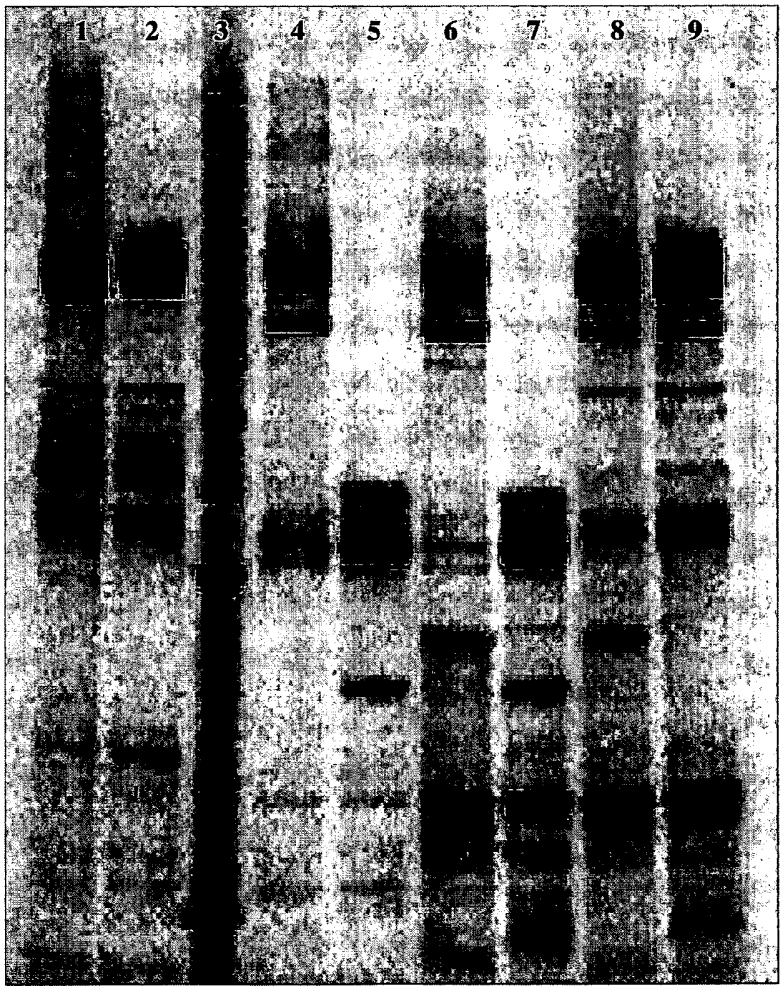


Figure 10

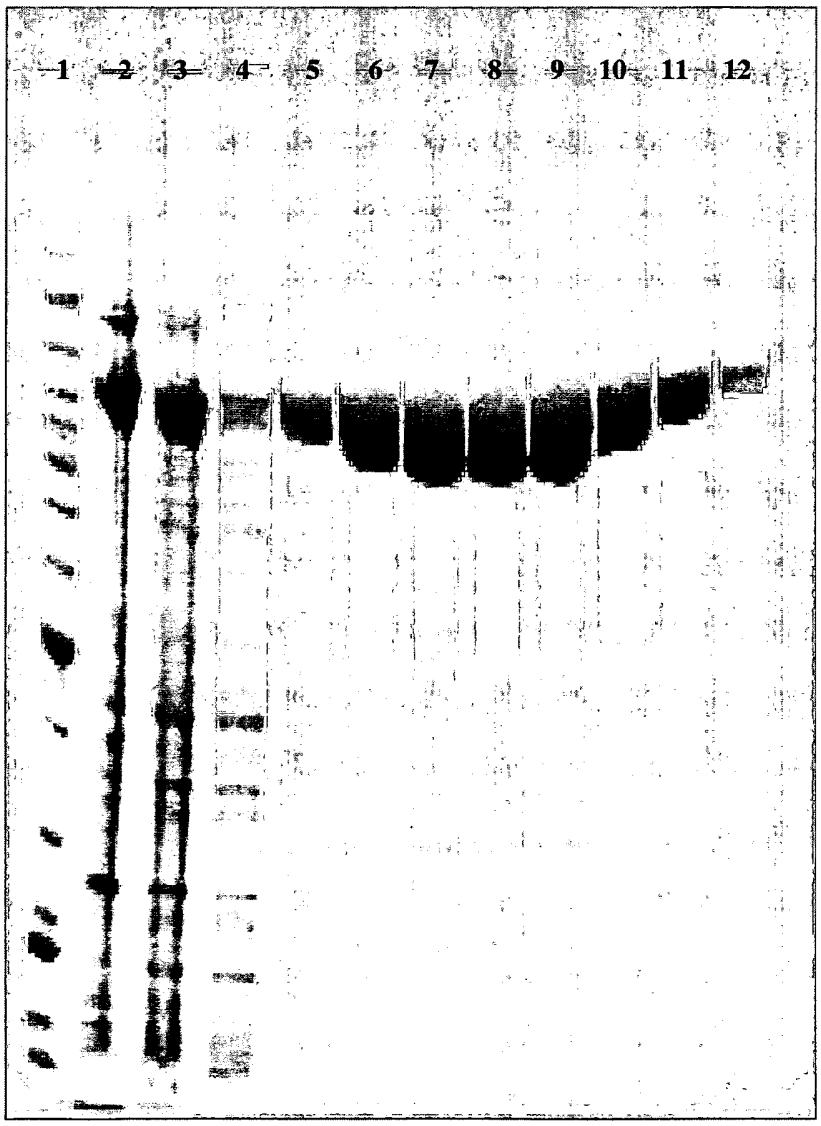


Figure 11

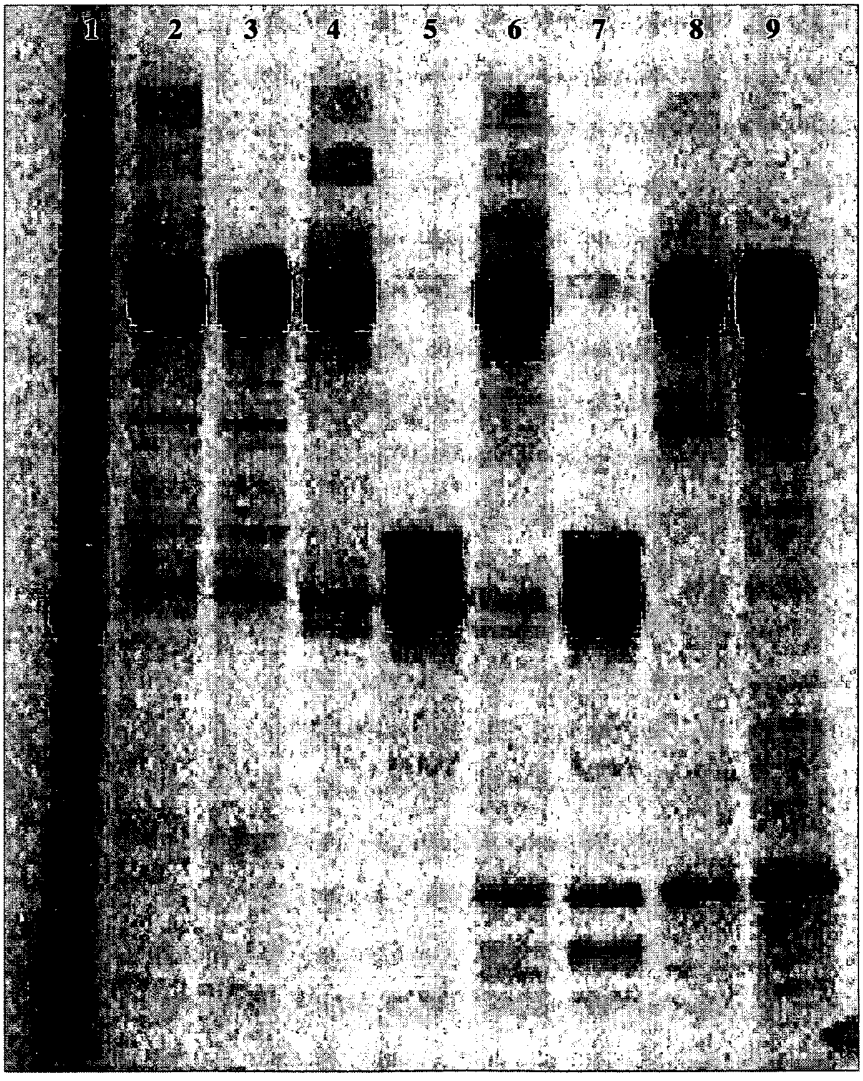


Figure 12

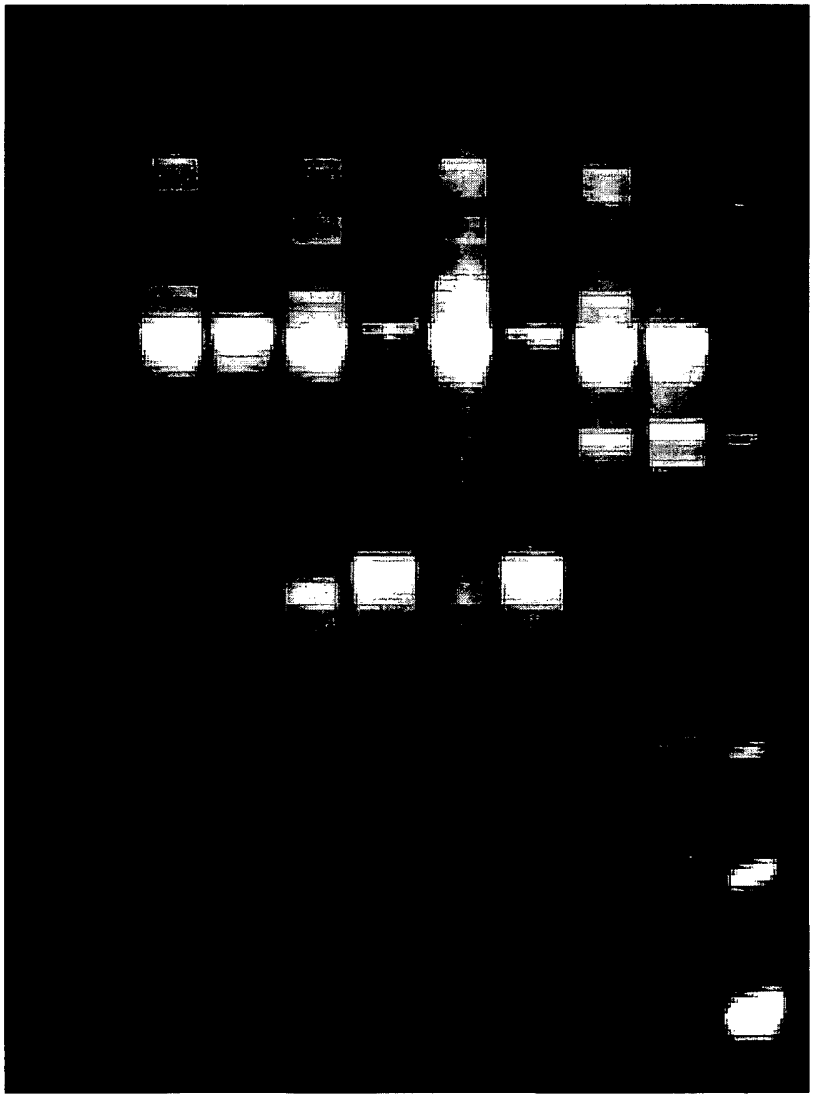


Figure 13

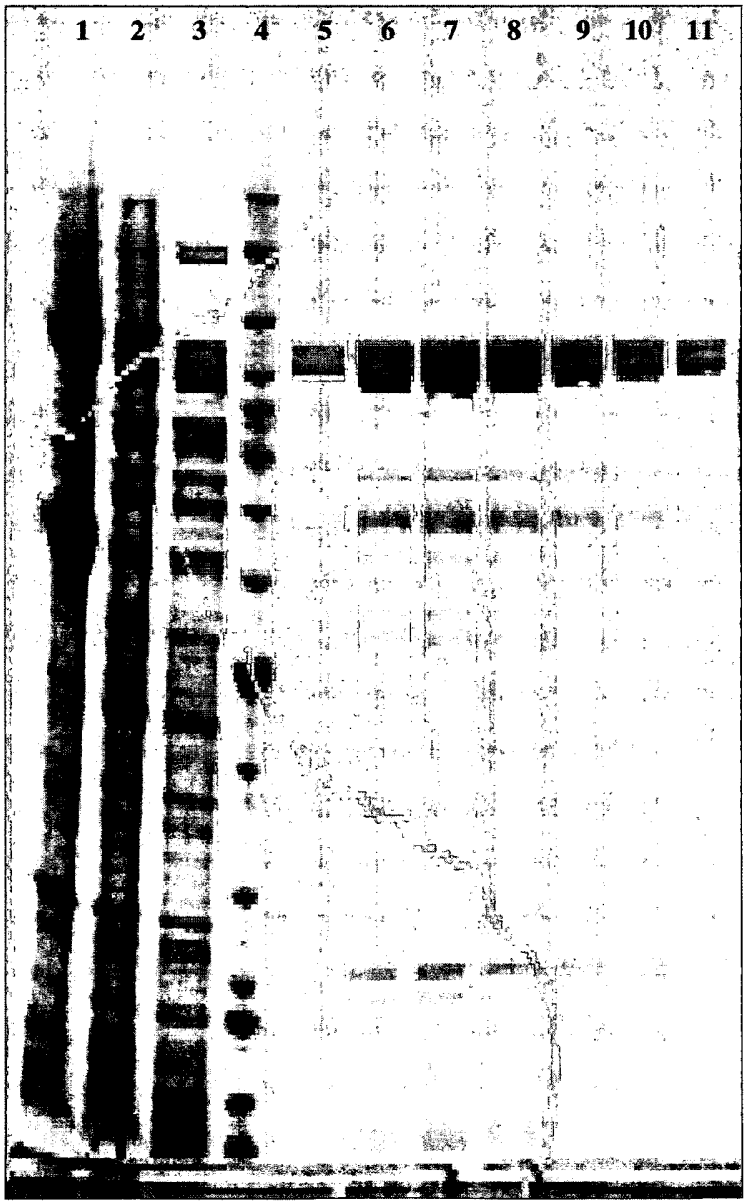


Figure 14

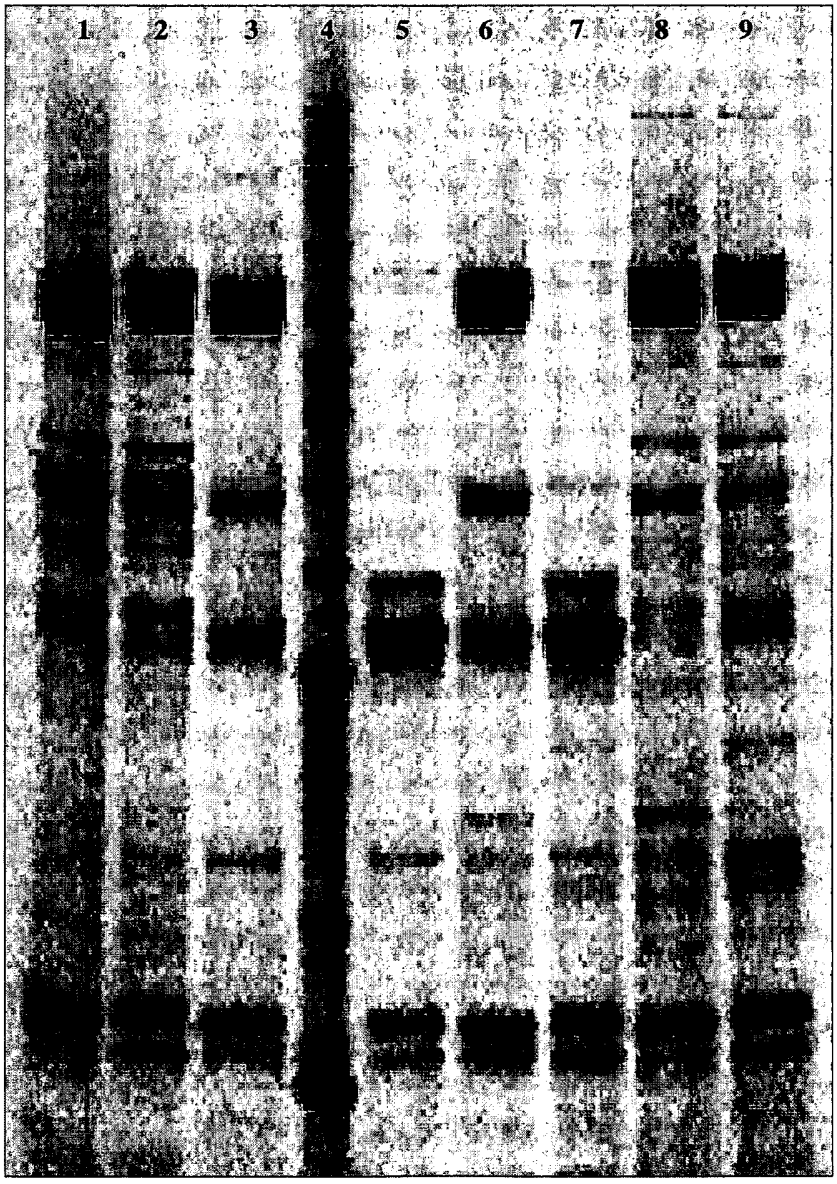


Figure 15

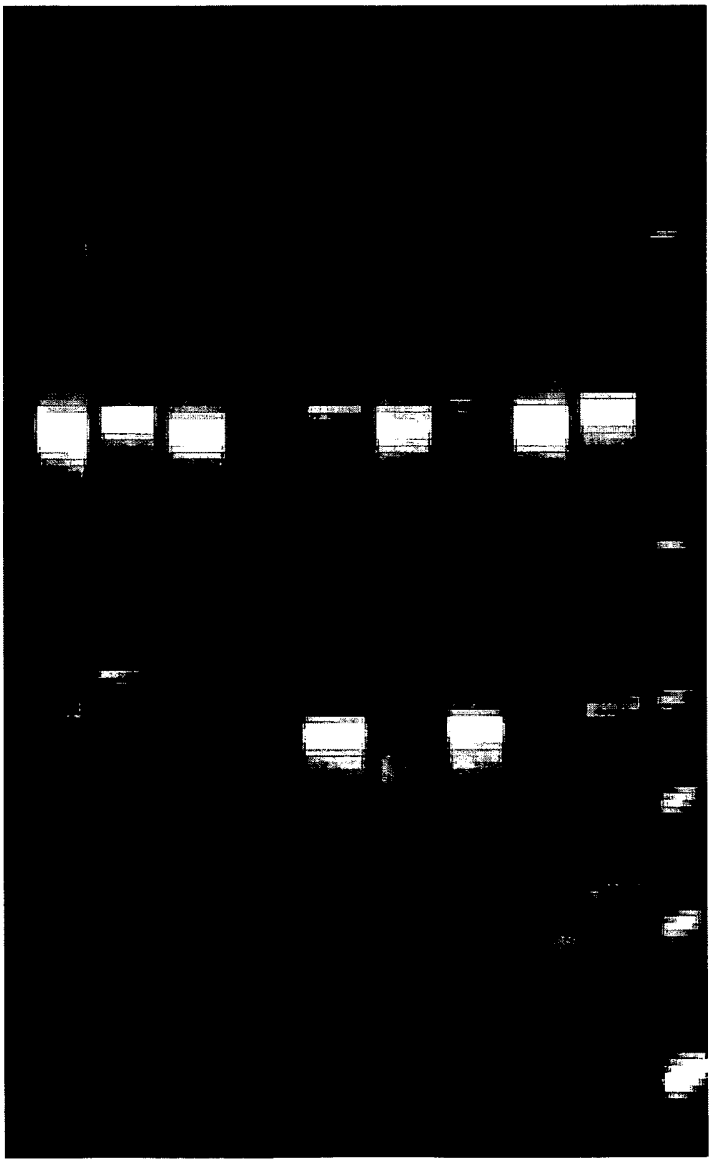


Figure 16

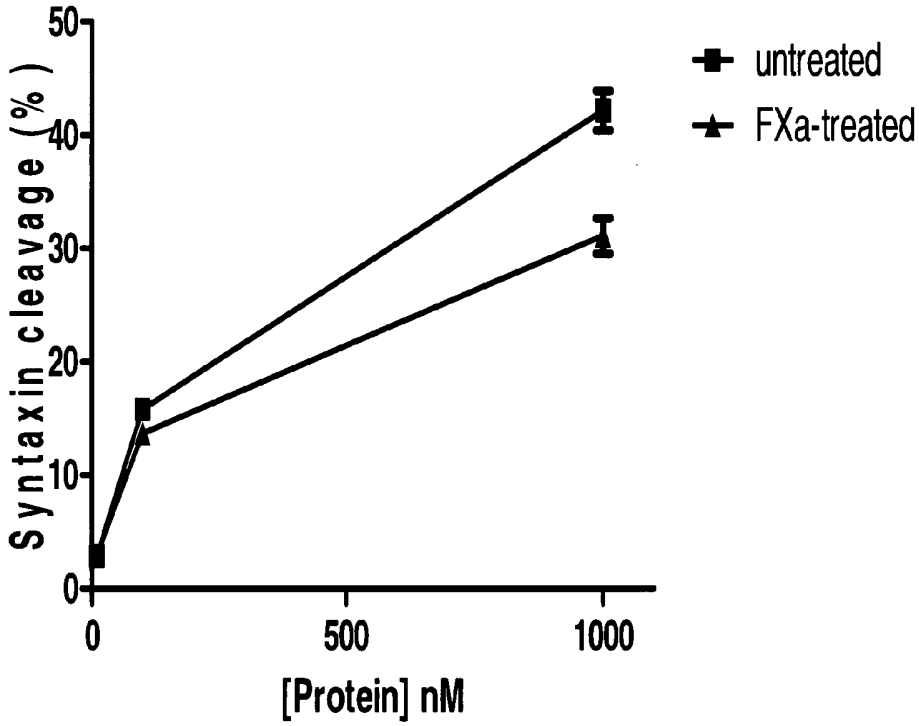
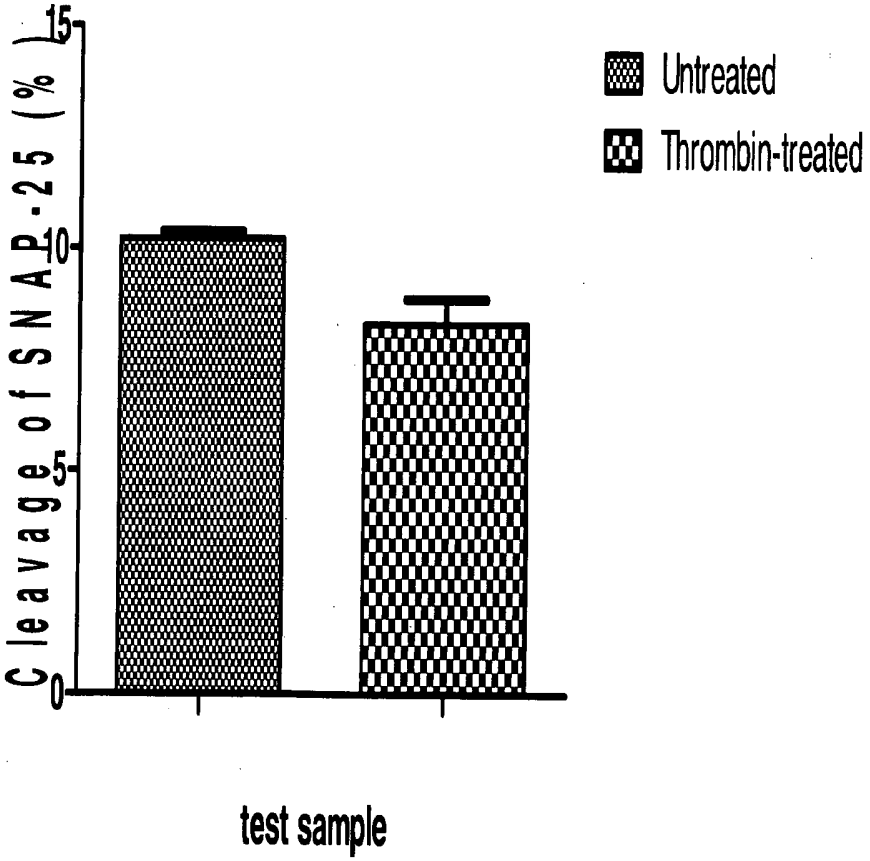

Syntaxin cleavage (%)
50
40
30
20
10
0
0
500
1000
[Protein] nM
untreated
FXa-treated

Figure 17


Cleavage of SNAP-25 (%)
15
10
5
0
Untreated
Thrombin-treated
test sample