IMMUNOGENIC COMPOSITION COMPRISING ELEMENTS OF C. DIFFICILE CDTB AND/OR CDTA PROTEINS

Background

C.difficile is the most important cause of nosocomial intestinal infections and is the major cause of pseudomembranous colitis in humans (Bartlett *et al Am. J. Clin. Nutr.* 11 suppl:2521-6 (1980)). The overall associated mortality rate for individuals infected with *C.difficile* was calculated to be 5.99% within 3 months of diagnosis, with higher mortality associated with advanced age, being 13.5% in patients over 80 years (Karas *et al Journal of Infection* 561:1-9 (2010)).The current treatment for *C.difficile* infection is the administration of antibiotics (metronidazole and vancomycin), however there has been evidence of strains which are resistant to these antibiotics (Shah *et al*., Expert Rev. Anti Infect. Ther. 8(5), 555–564 (2010)). Accordingly there is a need for immunogenic compositions capable of inducing antibodies to, and /or a protective immune response to, *C.difficile*.

The enterotoxicity of *C.difficile* is primarily due to the action of two toxins, toxin A and toxin B. These are both potent cytotoxins (Lyerly *et al* Current Microbiology 21:29-32 (1990).

It has been demonstrated that fragments of toxin A, in particular fragments of the C-terminal domain, can lead to a protective immune response in hamsters (Lyerly *et al* Current Microbiology 21:29-32 (1990)), WO96/12802 and WO00/61762. However the present inventors have demonstrated that antibodies against toxin A and toxin B alone are not sufficient in order to prevent disease caused by certain strains, in particular serogroup 078 and 027 strains. For this reason vaccines which are capable of protecting against these strains are still required.

Some strains, but not all, also express the binary toxin (CDT). Similar to many other binary toxins, CDT is composed of two components – an enzymatically active component (CDTa) and a catalytically inert transport component (CDTb). The catalytically inert component facilitates translocation of the CDTa into the target cell.

CDTa has an ADP-ribosylating activity, which transfers the ADP-ribose moiety of NAD/NADPH to the monomeric actin (G-actin) in the target cell and thus preventing its polymerization to F-actin and resulting in disruption of the cytoskeleton and eventual cell death (Sundriyal *et al*, Protein expression and Purification 74 (2010) 42-48).

WO2013/112867 (Merck) describes vaccines against *Clostridium difficile* comprising recombinant *C. difficile* Toxin A and Toxin B and binary toxin A (CDTa) proteins comprising specifically defined mutations relative to the native toxin sequence that are described as substantially reducing or eliminating toxicity, in combination with binary toxin B (CDTb).

The present inventors have found, that binary toxin can be used to provide an improved vaccine against *C.difficile* particularly providing protection against several of the most concerning *C.difficile* strains (such as the 027 and 078 strains). Furthermore the present inventors have demonstrated, for the first time, that only CDTa or CDTb (not both) is required in order to generate antibodies which are capable of neutralizing strains expressing binary toxin. In addition the inventors have demonstrated, for the first time, that CDTa proteins comprising mutations which reduce the ADP-ribosylating activity of CDTa, are still capable of raising an immune response. In addition, the inventors have demonstrated that truncated CDTa proteins are capable of raising an immune response. Similarly the inventors have demonstrated that truncated CDTb proteins are capable of raising an immune response, that CDTb can raise an immune response when it is in its monomeric or polymeric form and that fusion proteins comprising CDTa and CDTb or CDTb fused to isolated toxin A and/or isolated toxin B are capable of raising an immune response. Finally, the inventors have demonstrated that an immunogenic composition comprising binary toxin can be improved by adding an adjuvant, in particular an adjuvant comprising an immunologically active saponin presented in the form of a liposome or an oil in water emulsion.

Summary of Invention

In a first aspect of the invention there is provided an immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein wherein the composition does not further comprise an isolated protein having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% similarity to SEQ ID NO: 1, SEQ ID NO: 31 or SEQ ID NO: 32.

In a second aspect of the invention there is provided an immunogenic composition comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein comprising the receptor binding domain.

In a third aspect of the invention there is provided an immunogenic composition comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a mutated CDTb protein incapable of binding to CDTa.

In a fourth aspect of the invention there is provided an immunogenic composition comprising isolated *Clostridium difficile* CDTa protein wherein the isolated *Clostridium difficile* CDTa protein is a truncated CDTa protein which does not comprise the C-terminal domain.

In a fifth aspect the present invention provides an immunogenic composition comprising a fusion protein comprising a CDTa protein and a CDTb protein.

In a sixth aspect the present invention provides an immunogenic composition comprising a fusion protein between an isolated *Clostridium difficile* toxin A protein and/or an isolated *Clostridium difficile* toxin B protein fused to a CDTb protein.

In a seventh aspect the present invention provides a vaccine comprising the immunogenic composition of any one of the first five aspects and a pharmaceutically acceptable excipient.
In an eighth aspect the present invention provides the immunogenic composition of any one of the first five aspects or the vaccine of the sixth aspect, for use in the treatment or prevention of disease e.g. *C.difficile* disease.
In a ninth aspect the present invention provides the use of an immunogenic composition of any one of the first five aspects or the vaccine of the sixth aspect in the preparation of a medicament for the prevention or treatment of disease e.g. *C.difficile* disease.
In a tenth aspect the present invention provides a method of preventing or treating *C.difficile* disease comprising administering an immunogenic composition of any one of the first six aspects or the vaccine of the seventh aspect to a mammalian subject.
In a further aspect of the invention there is provided an immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein.

In a further aspect of the invention there is provided an immunogenic composition comprising either an isolated *Clostridium difficile* CDTb protein or an isolated CDTa protein but does not comprise both an isolated CDTb protein and an isolated CDTa protein.

Novel polypeptides and nucleotides as defined herein also form further aspects of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 (comprising Figures 1a-1h) - Graphs describing the size distribution of the different CdtA, CdtB and CdtA-CdtB fusion constructions as determined by sedimentation velocity analytical ultracentrifugation:

Figure 1a: AUC of C67 (CdtA (aa44-463) mut E428Q-E430Q

Figure 1b: AUC of C69 CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F-E428Q-E430Q

Figure 1c: AUC of C50 (CdtA N-term without linker (aa44-260)

Figure 1d: AUC of C61 (fusion CdtA N term with linker-CdtBshort)

Figure 1e: AUC of C62 (fusion CdtA N term without linker-CdtBlong)

Figure 1f: AUC of C52 (CdtB long)

Figure 1g: AUC of C53 (CdtB short)

Figure 1h: AUC of C55 CdtB Δ prodomain (aa. 212-876)

FIG. 2 (comprising Figures 2a-2c) - SDS PAGE profiles of CdtA, CdtB and CdtA-CdtB fusion constructions after purification:

Figure 2a: SDS PAGE of purified CdtA-CdtB fusion constructions. Lane 1: Molecular weight marker Novex sharp prestained. Lane 2: 5µg of C61 CdtA N-term link (aa. 44-268)-CdtB RBD short (aa. 636-876). Lane 3: 5µg of C62 CdtA N-term (aa. 44-260)-CdtB RBD long (aa. 621-876).

Figure 2b: SDS PAGE of purified CdtA constructions. Lane 1: Molecular weight marker Novex sharp prestained. Lane 2: 5µg of C50 CdtA WO linker (44-260). Lane 3: 5µg of C67 CdtA full length (aa44-463) mut. E428Q-E430Q. Lane 4: 5µg of C69 CdtA full length (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F-E428Q-E430Q.

Figure 2c: SDS PAGE of purified CdtB constructions. Lane 1: Molecular weight marker Novex sharp prestained. Lane 2: 5µg of C37 CdtB' Δsignal sequence (aa43-876) + GST N-term after removal of the N-term GST and activation by prodomain cleavage with chymotrypsin. Lane 3: 5µg of C55 CdtB Δ prodomain (aa. 212-876). Lane 4: 5µg of C52 CdtB receptor binding domain long (aa. 621-876). Lane 5: Molecular weight marker. Lane 6: 5µg of C38 CdtB' Δsignal sequence (aa43-876).

FIG. 3 – Graph showing anti-CDTb immunogenicity in mice immunised with *C. difficile* Binary Toxin component A or *C. difficile* Binary Toxin component B, in both cases formulated with adjuvant

FIG. 4 - Graph showing anti-CDTa immunogenicity in mice immunised with *C. difficile* Binary Toxin component A or *C. difficile* Binary Toxin component B, in both cases formulated with adjuvant

FIG. 5 – Cytotoxicity inhibition titres in HCT116 cells from mice immunised with *C. difficile* Binary Toxin component A or *C. difficile* Binary Toxin component B, in both cases formulated with adjuvant

FIG. 6 - Cytotoxicity inhibition titres in HT29 cells from mice immunised with *C. difficile* Binary Toxin component A or *C. difficile* Binary Toxin component B, in both cases formulated with adjuvant

FIG. 7 - Graph showing anti-CDTb immunogenicity in mice immunised with *C. difficile* Cdtb (activated or non activated, with and without F2 fusion comprising fragments from Toxin A and Toxin B) formulated with adjuvant

FIG. 8 - Graph showing anti-Tox A immunogenicity in mice immunised with *C. difficile* Cdtb (activated or non activated, with and without F2 fusion comprising fragments from Toxin A and Toxin B) formulated with adjuvant

FIG. 9 - Graph showing anti-Tox B immunogenicity in mice immunised with *C. difficile* Cdtb (activated or non activated, with and without F2 fusion comprising fragments from Toxin A and Toxin B) formulated with adjuvant

FIG. 10 – Tox A cytotoxicity inhibition titres in HT29 cells from mice immunised with *C. difficile* Cdtb (activated or non activated, with and without F2 fusion comprising fragments from Toxin A and Toxin B) formulated with adjuvant

FIG. 11 – Tox B cytotoxicity inhibition titres in HCT116 cells from mice immunised with *C. difficile* Cdtb (activated or non activated, with and without F2 fusion comprising fragments from Toxin A and Toxin B) formulated with adjuvant

FIG. 12 – Binary Toxin cytotoxicity inhibition titres in HT29 cells from mice immunised with *C. difficile* Binary Toxin component A or *C. difficile* Binary Toxin component B, in both cases formulated with adjuvant

FIG. 13 - Graph showing anti-CDTb immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 14 - Graph showing anti-CDTa immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 15 - Graph showing anti-Tox B immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 16 - Graph showing anti-Tox A immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 17 - Binary Toxin cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 18 - Binary Toxin cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 19 – Tox A cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 20 – Tox B cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an adjuvant formulation

FIG. 21 - Graph showing anti-CDTb immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 22 - Graph showing anti-CDTa immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 23- Graph showing anti-Tox B immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 24 - Graph showing anti-Tox A immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 25 - Binary Toxin cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 26 - Binary Toxin cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 27 – Tox A cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 28 – Tox B cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an adjuvant formulation

FIG. 29 - Graph showing anti-CDTb immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 30 - Graph showing anti-CDTa immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 31- Graph showing anti-Tox B immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 32 - Graph showing anti-Tox A immunogenicity in mice immunized with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 33 - Binary Toxin cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 34 - Binary Toxin cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 35 – Tox A cytotoxicity inhibition titres in HT29 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

FIG. 36 – Tox B cytotoxicity inhibition titres in HCT116 cells from mice immunised with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

Detailed Description

Binary toxin

The *Clostridium difficile* binary toxin comprises two different proteins, CDTa and CDTb. During infection CDTb is activated by proteolytic cleavage by a chymotrypsin-like protease to produce a CDTb protein lacking the prodomain (also referred to as CDTb''). Note that CDTb'' also lacks the CDTb signal sequence, a CDTb protein lacking the signal sequence but not lacking the prodomain is referred to as CDTb'. After proteolytic activation the CDTb oligomerises and binds to CDTa to form the complete 'binary toxin'. The binding of the binary toxin to the cell receptors leads to receptor-mediated endocytosis. As the endosome acidifies the CDTb binding domain undergoes conformational changes that allow the CDTb oligomer to form a pore, the pore formation triggers translocation of the ADP-ribosyltransferase domain (CDTa) into the target cell.

CDTb

The present invention provides an immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein. The present invention also provides an immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein as the sole C. difficile antigen. As used herein the term "as the sole C. difficile antigen" means that the immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein as the sole C. difficile antigen does not also comprise another antigen from C. difficile e.g. the immunogenic composition does not also comprise a toxin A, toxin B or CDTa protein.

The present invention provides an immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein wherein the composition does not further comprise an

isolated protein having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% similarity to SEQ ID NO: 1, SEQ ID NO: 31 or SEQ ID NO: 32. According to the invention as herein described the term 'CDTb protein' encompasses SEQ ID NO:3 or fragments or variants of SEQ ID NO:3.

In one embodiment of this first aspect of the invention, the composition does not comprise an isolated *Clostridium difficile* CDTa protein.

In one embodiment of this aspect the isolated *Clostridium difficile* CDTb protein is or comprises
(i) SEQ ID NO: 3; or
(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:3; or
(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO:3.

In one such aspect there is provided an immunogenic composition wherein the isolated *Clostridium difficile* CDTb protein is a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:3.

In another aspect there is provided an immunogenic composition wherein the isolated *Clostridium difficile* CDTb protein is a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800 or 850 contiguous amino acids of SEQ ID NO: 3.

CDTb comprises multiple domains, in particular CDTb comprises a signal peptide and a prodomain both of which are cleaved as explained in the section entitled "Binary Toxin" above.

In one embodiment the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein with the signal peptide removed. The term 'truncated CDTb protein with the signal peptide removed' refers to a fragment or variant of SEQ ID NO: 3 with substantially all of the signal peptide removed (therefore which does not comprise amino acids corresponding to substantially all of the signal peptide), there may be a few amino acids of the signal peptide remaining, for example 2, 5, 10, 15 or 20 amino acids of the signal peptide may remain. The signal peptide corresponds to amino acids 1-48 (encompassing amino acids 1-42) of SEQ ID NO: 3 or their equivalents in a binary toxin protein isolated from a different strain of

C.difficile, for example amino acids 1-42 of the amino acid sequence of CDTb from strain CD196 (Perelle, M. *et al* Infect. Immun., 65 (1997), pp. 1402–1407).
Suitably in this embodiment the isolated *Clostridium difficile* CDTb protein is or comprises

(i) SEQ ID NO: 7 or SEQ ID NO: 16; or

(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 7 or SEQ ID NO:16; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750 or 800 contiguous amino acids of SEQ ID NO: 7 or SEQ ID NO:16. In one embodiment the truncated CDTb protein with the signal peptide removed is or comprises a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 7 or SEQ ID NO:16. In a further embodiment the isolated truncated CDTb protein with the signal peptide removed is or comprises a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 800 contiguous amino acids of SEQ ID NO: 7 or SEQ ID NO:16.

In one embodiment the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein with the prodomain removed. The term 'truncated CDTb protein with the prodomain removed' refers to a fragment or variant of SEQ ID NO: 3 with substantially all of the prodomain removed (therefore which does not comprise amino acids corresponding to substantially all of the prodomain), there may be a few amino acids of the prodomain remaining, for example 2, 5, 10, 15 or 20 amino acids of the prodomain may remain. The prodomain corresponds to amino acids 48-211 (encompassing amino acids 48-166) of SEQ ID NO:3 or their equivalents in a binary toxin protein isolated from a different strain of *C.difficile*. Optionally the truncated CDTb protein with the prodomain removed also lacks the CDTb signal sequence, the CDTb signal sequence corresponds to amino acids 1-48 (encompassing amino acids 1-42) of SEQ ID NO:3 or their equivalents in a different strain. The term 'truncated CDTb protein with the prodomain removed' may also refer to a fragment or variant of SEQ ID NO: 3 which is capable of oligomerising and binding to CDTa. In this embodiment of the invention the isolated *Clostridium difficile* CDTb protein suitably is or comprises

(i) SEQ ID NO: 9 or SEQ ID NO: 51; or

(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:9 or SEQ ID NO: 51; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 650 contiguous amino acids of SEQ ID NO:9 or SEQ ID NO: 51.

In one embodiment the truncated CDTb protein with the prodomain removed is or comprises a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:9. In a further embodiment the isolated truncated CDTb

protein with the prodomain removed is or comprises a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 800 contiguous amino acids of SEQ ID NO:9.

CDTb also comprises a receptor binding domain. In one embodiment the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein comprising the receptor binding domain. The term 'truncated CDTb protein comprising the receptor binding domain' refers to a fragment or variant of SEQ ID NO:3 with substantially all but the receptor binding domain removed (therefore which does not comprise amino acids corresponding to substantially all of the protein except for the receptor binding domain), there may be a few amino acids in addition to the receptor binding domain remaining, for example 2, 5, 10, 15 or 20 amino acids except for/in addition to the receptor binding domain. In one version, the receptor binding domain corresponds to amino acids 620-876 of SEQ ID NO:3, or their equivalents in a binary toxin protein isolated from a different strain of *C.difficile*. In another version, the receptor binding domain corresponds to amino acids 636-876 of SEQ ID NO:3 or their equivalents in a binary toxin protein isolated from a different strain of *C.difficile*.
In this embodiment the isolated *Clostridium difficile* CDTb protein suitably is or comprises
(i) SEQ ID NO: 34 or SEQ ID NO: 36; or
(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 34 or SEQ ID NO: 36; or
(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150 or 200 contiguous amino acids of SEQ ID NO: 34 or SEQ ID NO: 36.
In another embodiment of this aspect of the invention, the isolated *Clostridium difficile* CDTb protein is a mutated CDTb protein incapable of binding to CDTa.
In this embodiment the isolated *Clostridium difficile* CDTb protein suitably is or comprises
(i) SEQ ID NO: 50; or
(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 50; or
(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO: 50.
The CDTb protein varies in amino acid sequence between different strains, for this reason the amino acid numbering may differ between strains. For this reason the term 'equivalents in a different strain' refers to amino acids which correspond to those of a reference strain (e.g., *C. difficile* R20291 from which SEQ ID NO:1 and SEQ ID NO:3 are derived), but which are found in a toxin from a different strain and which may thus be numbered differently. A region of 'equivalent' amino acids may be determined by aligning the sequences of the toxins from the different strains. Example binary toxin producing strains of C. difficile include

CD196, CCUG 20309, R8637, IS81, IS93, IS51, IS58, R6786, R7605, R10456 and R5989. The amino acids numbers provided throughout refer to those of reference strain R20291.

In one embodiment the isolated *Clostridium difficile* CDTb protein is a monomer of CDTb. In a further embodiment the isolated *Clostridium difficile* CDTb protein is a multimer of CDTb. In a further embodiment the isolated *Clostridium difficile* CDTb protein is a heptamer of CDTb.

In a second aspect the present invention provides immunogenic compositions comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein comprising the receptor binding domain. In one embodiment of this aspect the isolated *Clostridium difficile* CDTb protein suitably is or comprises

(i) SEQ ID NO: 34 or SEQ ID NO: 36; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 34 or SEQ ID NO: 36; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150 or 200 contiguous amino acids of SEQ ID NO: 34 or SEQ ID NO: 36.

In a third aspect the invention provides immunogenic compositions comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a mutated CDTb protein incapable of binding to CDTa. In one embodiment of this aspect, the isolated *Clostridium difficile* CDTb protein suitably is or comprises

(i) SEQ ID NO: 50; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 50; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO: 50.

In one embodiment of the second and third aspects of the invention, the immunogenic composition comprises/further comprises an isolated *Clostridium difficile* CDTa protein comprising

(i) SEQ ID NO: 1; or

(ii) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:1; or

(iii) a fragment of CDTa having at least at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID NO:1.

CDTa

The present invention also provides immunogenic compositions comprising an isolated *Clostridium difficile* CDTa protein. The present invention also provides immunogenic compositions comprising an isolated *Clostridium difficile* CDTa protein as the sole C. difficile antigen. As used herein the term "as the sole C. difficile antigen" means that the immunogenic composition comprising an isolated *Clostridium difficile* CDTa protein as the sole C. difficile antigen does not also comprise another antigen from C. difficile e.g. the immunogenic composition does not also comprise a toxin A, toxin B or CDTb protein. According to the invention as herein described the term 'CDTa protein' encompasses SEQ ID NO:1 or fragments or variants of SEQ ID NO:1. In one embodiment the isolated *Clostridium difficile* CDTa protein comprises a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:1. In a further embodiment the isolated *Clostridium difficile* CDTa protein comprises a fragment of CDTa having at least at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID NO:1.

CDTa comprises two domains, the C-terminal domain is responsible for the ADP ribosyltransferase activity whilst the N-terminal domain is responsible for interacting with CDTb.

In one embodiment of the first three aspects of the invention, the immunogenic composition comprises/further comprises an isolated *Clostridium difficile* CDTa protein. Suitably the isolated *Clostridium difficile* CDTa protein is a truncated CDTa protein. "A truncated CDTa protein" as used herein means a CDTa protein that does not achieve its full length or its proper form, and thus is missing some of the amino acid residues that are present in full length CDTa of SEQ ID NO: 1, and which cannot perform the function for which it was intended because its structure is incapable of doing so, e.g. ADP ribosyltransferase activity and/or interacting with CDTb.

Suitably the isolated *Clostridium difficile* CDTa protein is a truncated CDTa protein which does not comprise the C-terminal domain. The term 'truncated CDTa protein which does not comprise the C-terminal domain' refers to a fragment or variant of SEQ ID NO:1 which does not comprise a substantial portion of the C-terminal domain, there may be a few amino acids of the C-terminal domain remaining, for example, 2, 5, 10, 15, 20, 25, 30, 35 or 50 amino acids of the C-terminal domain may remain. The C-terminal domain corresponds to amino acids 267-463 of SEQ ID NO:1 or their equivalents in a CDTa protein isolated from a different strain of *C.difficile*. In this embodiment the truncated *Clostridium difficile* CDTa protein suitably is or comprises

(i) SEQ ID NO: 14 or SEQ ID NO: 15

(i) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 14 or SEQ ID NO: 15; or
(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, or 190 contiguous amino acids of SEQ ID NO: 14 or SEQ ID NO: 15.
In one embodiment the truncated CDTa protein which does not comprise the C-terminal domain is a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:13. In a further embodiment the truncated CDTa protein which does not comprise the C-terminal domain is a variant of CDTa having at least 30, 50, 80, 100, 120, 150, or 190 contiguous amino acids of SEQ ID NO:13.

In a fourth aspect the invention provides an immunogenic composition comprising isolated *Clostridium difficile* CDTa protein wherein the isolated *Clostridium difficile* CDTa protein is a truncated CDTa protein which does not comprise the C-terminal domain. In one embodiment of this aspect, the isolated *Clostridium difficile* CDTa protein suitably is or comprises
(i) SEQ ID NO: 14 or SEQ ID NO: 15; or
(ii)a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:14 or SEQ ID NO: 15; or
(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, or 190 contiguous amino acids of SEQ ID NO:14 or SEQ ID NO: 15.

In a further embodiment of any of the aspects of the invention, the isolated *Clostridium difficile* CDTa protein suitably contains a mutation which reduces its ADP-ribosyltransferase activity. For example the isolated *Clostridium difficile* CDTa protein has a mutation from glutamate to another amino acid at position 428. The term 'has a mutation at position 428' refers to CDTa proteins which have a mutation at this exact location but also to a CDTa protein which is isolated from a different strain and which has a mutation at an equivalent position. The CDTa protein varies in amino acid sequence between different strains, for this reason the amino acid numbering may differ between strains, thus a CDTa protein from a different strain may have a corresponding glutamate which is not number 428 in sequence. In one embodiment the isolated *Clostridium difficile* CDTa protein has a mutation from glutamate to glutamine at position 428.

In a further embodiment of any of the aspects of the invention, the isolated *Clostridium difficile* CDTa protein suitably has a mutation from glutamate to a different amino acid at position 430, the term 'has a mutation at position 430' refers to proteins which have this exact location but also to a CDTa protein which is isolated from a different strain and which

has a mutation at an equivalent position. In one embodiment the isolated *Clostridium difficile* CDTa protein has a mutation from glutamate to glutamine at position 430.

In a further embodiment of any of the aspects of the invention, the isolated *Clostridium difficile* CDTa protein suitably is or comprises

(i) SEQ ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54; or

(ii) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54; or

(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54.

In a further embodiment of any of the aspects of the invention, the isolated *Clostridium difficile* CDTa protein suitably is or comprises

(i) SEQ ID NO: 48; or

(ii) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 48; or

(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID ID NO: 48.

Immunogenic Compositions with CDTa and/or CDTb

In a further embodiment there is provided an immunogenic composition which comprises a CDTb protein but does not comprise a CDTa protein, for example the immunogenic composition does not comprise a variant of CDTa having at least 95%, 98%, 99%, or 100% sequence identity to SEQ ID NO:1 or a fragment of CDTa having at least at least 250, 400 or 450. contiguous amino acids of SEQ ID NO:1.

In a further embodiment there is provided an immunogenic composition which comprises a CDTa protein but does not comprise a CDTb protein, for example the immunogenic composition does not comprise a variant of CDTb having at least 95%, 98%, 99%, or 100% sequence identity to SEQ ID NO:3 or a fragment of CDTb having at least 700, 750, or 800 contiguous amino acids of CDTb.

In a further embodiment there is provided an immunogenic composition which comprises either an isolated *Clostridium difficile* CDTb protein or an isolated CDTa protein but does not comprise both an isolated CDTb protein and an isolated CDTa protein.

In a further embodiment there is provided a fusion protein comprising a CDTa protein and a CDTb protein. In another embodiment there is provided immunogenic compositions comprising a fusion protein comprising a CDTa protein and a CDTb protein.

Fusion proteins comprising a CDTa protein and a CDTb protein

In a fifth aspect, the invention provides immunogenic compositions comprising a fusion protein comprising a CDTa protein and a CDTb protein. In one embodiment of this aspect, the CDTa protein suitably is truncated. For example, the CDTa protein suitably does not comprise the C-terminal domain. In this aspect, the CDTb protein suitably is truncated. In this embodiment, the CDTb protein suitably comprises the receptor binding domain.

In one embodiment of this aspect of the invention, the fusion protein suitably is or comprises

(i) SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43; or

(ii) a variant having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43; or

(iii) a fragment having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43.

“Fusion polypeptide” or “fusion protein” refers to a protein having at least two heterologous polypeptides (e.g. at least two Mycobacterium sp. polypeptides) covalently linked, either directly or via an amino acid linker. It may also refer to a protein having at least two heterologous polypeptides linked non-covalently. The polypeptides forming the fusion protein are typically linked C-terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N-terminus, or N-terminus to C-terminus. The polypeptides of the fusion protein can be in any order. This term also refers to conservatively modified variants, polymorphic variants, alleles, mutants, immunogenic fragments, and interspecies homologs of the antigens that make up the fusion protein.

The term “fused” refers to the linkage e.g. covalent linkage between two polypeptides in a fusion protein. The polypeptides are typically joined via a peptide bond, either directly to each other or via an amino acid linker. Optionally, the peptides can be joined via non-peptide covalent linkages known to those of skill in the art.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary

and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., Gene 40:39-46 (1985); Murphy et al., Proc. Natl. Acad. Sci. USA 83:8258-8262 (1986); U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length for example 1, 5, 10, 15, 20, 25, 30, 35 or 40 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

In one embodiment of any of the aspects of the invention, the immunogenic composition elicits antibodies that neutralize CDTa or CDTb or both. In a further embodiment the composition elicits antibodies that neutralize binary toxin. Whether a composition elicits antibodies against a toxin can be measured by immunising mice with the immunogenic composition, collecting sera and analysing the anti-toxin titres of the sera using by ELISA. The sera should be compared to a reference sample obtained from mice which have not been immunised. The composition of the invention elicits antibodies that neutralise CDTa if the sera against the polypeptide gives an ELISA readout more than 10%, 20%, 30%, 50%, 70%, 80%, 90%, or 100% higher than the reference sample.

In a further embodiment the immunogenic compositions of the invention elicits a protective immune response in a mammalian host against strains of *C.difficile*. In one embodiment the mammalian host is selected from the group consisting of mouse, rabbit, guinea pig, non-human primate, monkey and human. In one embodiment the mammalian host is a mouse. In a further embodiment the mammalian host is a human.

Whether an immunogenic composition elicits a protective immune response in a mammalian host against strains of *C.difficile* can be determined using a challenge assay. In such an assay the mammalian host is vaccinated with the immunogenic composition and challenged by exposure to *C.difficile*, the time which the mammal survives after challenge is compared with the time which a reference mammal that has not been immunised with the immunogenic

composition survives. An immunogenic composition elicits a protective immune response if a mammal immunised with the immunogenic composition survives at least 10%, 20%, 30%, 50%, 80%, 80%, 90%, or 100% longer than a reference mammal which has not been immunised after challenge with *C.difficile*.

Toxin A and Toxin B

In one embodiment of any of the aspects of the invention, the immunogenic compositions of the invention further comprise an isolated *Clostridium difficile* toxin A protein and/or an isolated *C.difficile* toxin B protein.

The term 'isolated *Clostridium difficile* toxin A protein' refers to a fragment or variant of SEQ ID NO: 31. In one embodiment the isolated *Clostridium difficile* toxin A protein is a fragment comprising 50, 100, 150, 200, 250, 300, 500, 750, 1000, 1250, 1500, 1750, 2000 or 2500 contiguous amino acids of SEQ ID NO:31 . In one embodiment the isolated *Clostridium difficile* toxin A protein is a variant comprising 80%, 85%, 90%, 92%, 95%, 98%, 99% or 100% identity to SEQ ID NO:31.

The term 'isolated *Clostridium difficile* toxin B protein' refers to a fragment or variant of SEQ ID NO: 32. In one embodiment the isolated *Clostridium difficile* toxin B protein is a fragment comprising 50, 100, 150, 200, 250, 300, 500, 750, 1000, 1250, 1500, 1750 or 2000 SEQ ID NO:32 . In one embodiment the isolated *Clostridium difficile* toxin B protein is a variant comprising 80%, 85%, 90%, 92%, 95%, 98%, 99% or 100% identity to SEQ ID NO:32.

In one embodiment the isolated *Clostridium difficile* toxin A protein comprises a repeating domain fragment. The term 'toxin A repeating domain' refers to the C-terminal domain of the toxin A protein from *C.difficile*, comprising repeated sequences. The toxin A repeating domain refers to amino acids 1832-2710 of toxin A from strain VPI10463 (ATCC43255) and their equivalents in a different strain, the sequence of amino acids 1832-2710 from strain VPI10463 (ATCC43255) corresponds to amino acids 1832-2710 of SEQ ID NO: 31. In a further embodiment the isolated *Clostridium difficile* toxin A protein comprises a fragment of the toxin A N-terminal domain. The toxin A N-terminal domain refers to amino acids 1-1831 of toxin A from strain VBI10463 (ATCC43255) and their equivalents in a different strain, the sequence of amino acids 1-1831 of SEQ ID NO: 31.
In one embodiment the isolated *Clostridium difficile* toxin B protein comprises a toxin B repeating domain fragment. The term 'toxin B repeating domain' refers to the C-terminal domain of the toxin B protein from *C.difficile*. This domain refers to amino acids 1834-2366

from strain VPI10463 (ATCC43255) and their equivalents in a different strain, the sequence of amino acids 1834-2366 from strain VPI10463 (ATCC43255) corresponds to amino acids 1834-2366 of SEQ ID NO: 32. In a further embodiment the isolated *Clostridium difficile* toxin B protein comprises a fragment of the toxin B N-terminal domain. The toxin B N-terminal domain refers to amino acids 1-1833 of toxin B from strain VBI10463 (ATCC43255) and their equivalents in a different strain, the sequence of amino acids 1-1833 of SEQ ID NO: 32.

The *C.difficile* toxins A and B are conserved proteins, however the sequence differs a small amount between strains, moreover the amino acid sequence for toxins A and B in different strains may differ in number of amino acids.

For these reasons the terms toxin A repeating domain and/or toxin B repeating domain to refer to a sequence which is a variant with 90%, 95%, 98%, 99% or 100% sequence identity to amino acids 1832-2710 of SEQ ID NO: 31 or a variant with 90%, 95%, 98%, 99% or 100% sequence identity to amino acids 1834-2366 of SEQ ID NO:32. Similarly the terms toxin a N-terminal domain and/or toxin B N terminal domain refer to a sequence which is avariant with 90%, 95%, 98%, 99% or 100% sequence identity to amino acids 1-1831 of SEQ ID NO:31 or a variant with 90%, 95%, 98%, 99% or 100% sequence identity to amino acids 1-1833 of SEQ ID NO:32.

Furthermore the amino acid numbering may differ between the C-terminal domains of toxin A (or toxin B) from one strain and toxin A (or toxin B) from another strain. For this reason the term 'equivalents in a different strain' refers to amino acids which correspond to those of a reference strain (e.g., *C. difficile* VPI10463), but which are found in a toxin from a different strain and which may thus be numbered differently. A region of 'equivalent' amino acids may be determined by aligning the sequences of the toxins from the different strains. The amino acids numbers provided throughout refer to those of strain VPI10463.

In a further embodiment of any of the aspects of the invention, the isolated *C.difficile* toxin A protein and the isolated *C.difficile* toxin B protein form a fusion protein. In one embodiment the fusion protein is 80%, 85%, 90%, 95%, 98%, 99% or 100% identical to a sequence selected from the group consisting of SEQ ID NO: 18, 19, 20, 21, 22, 24, 26, 28 and 30. In a further embodiment the fusion protein is a fragment of at least 800, 850, 900 or 950 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NO: 18, 19, 20, 21, 22, 24, 26, 28 and 30.

In a further embodiment of any of the aspects of the invention the immunogenic composition comprises/further comprises a fusion protein between an isolated *Clostridium difficile* toxin A protein and/or an isolated *Clostridium difficile* toxin B protein fused to a CDTb protein or to a truncated CDTb protein. In one embodiment there is provided a fusion protein comprising a fragment of toxin A, a fragment of toxin B and a CDTb protein, for example the fusion protein may comprise a fragment or variant of SEQ ID NO:18, 19, 20, 21, 22, 24, 26, 28 or 30 fused

to a CDTb protein. For example the fusion protein may comprise a fragment or variant of SEQ ID NO:18, 19, 20, 21, 22, 24, 26, 28 or 30 fused to a truncated CDTb protein.
In one embodiment the fusion protein suitably is or comprises
(i)SEQ ID NO: 44 or SEQ ID NO: 45; or
(ii) a variant having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to SEQ ID NO: 44 or SEQ ID NO: 45; or
(iii) a fragment of at least 800, 850, 900 or 950 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NO: 44 or SEQ ID NO: 45.

Fragments

The term "fragment" as defined herein may refer to a fragment comprising a T cell epitope. T cell epitopes are short contiguous stretches of amino acids which are recognised by T cells (e.g. CD4+ or CD8+ T cells). Identification of T cell epitopes may be achieved through epitope mapping experiments which are well known to the person skilled in the art (see, for example, Paul, *Fundamental Immunology*, 3rd ed., 243-247 (1993); Beißbarth et al *Bioinformatics* 2005 21(Suppl. 1):i29-i37).

Suitably the fragments of the invention are immunogenic fragments. "Immunogenic fragments" according to the present invention will typically comprise at least 9 contiguous amino acids from the full length polypeptide sequence (e.g. at least 10), such as at least 12 contiguous amino acids (e.g. at least 15 or at least 20 contiguous amino acids), in particular at least 50 contiguous amino acids, such as at least 100 contiguous amino acids (for example at least 200 contiguous amino acids). Suitably the immunogenic fragments will be at least 20%, such as at least 50%, at least 70% or at least 80% of the length of the full length polypeptide sequence.

It will be understood that in a diverse out-bred population, such as humans, different HLA types mean that specific epitopes may not be recognised by all members of the population. Consequently, to maximise the level of recognition and scale of immune response to a polypeptide, it is generally desirable that an immunogenic fragment contains a plurality of the epitopes from the full length sequence (suitably all epitopes).

Variants

"Variants" or "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified

variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences.

In respect of variants of a protein sequence, the skilled person will recognise that individual substitutions, deletions or additions to polypeptide, which alters, adds or deletes a single amino acid or a small percentage of amino acids is a "conservatively modified variant" where the alteration(s) results in the substitution of an amino acid with a functionally similar amino acid or the substitution/deletion/addition of residues which do not substantially impact the biological function of the variant.

Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

A polypeptide of the invention (such as a CDTa protein or a CDTb protein) may contain a number of conservative substitutions (for example, 1-50, such as 1-25, in particular 1-10, and especially 1 amino acid residue(s) may be altered) when compared to the reference sequence. In general, such conservative substitutions will fall within one of the amino-acid groupings specified below, though in some circumstances other substitutions may be possible without substantially affecting the immunogenic properties of the antigen. The following eight groups each contain amino acids that are typically conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(see, e.g., Creighton, *Proteins* 1984).

Suitably such substitutions do not occur in the region of an epitope, and do not therefore have a significant impact on the immunogenic properties of the antigen.

Polypeptide variants may also include those wherein additional amino acids are inserted compared to the reference sequence, for example, such insertions may occur at 1-10 locations (such as 1-5 locations, suitably 1 or 2 locations, in particular 1 location) and may, for example, involve the addition of 50 or fewer amino acids at each location (such as 20 or fewer, in particular 10 or fewer, especially 5 or fewer). Suitably such insertions do not occur in the region of an epitope, and do not therefore have a significant impact on the immunogenic properties of the antigen. One example of insertions includes a short stretch of histidine residues (e.g. 2-6 residues) to aid expression and/or purification of the antigen in question.

Polypeptide variants include those wherein amino acids have been deleted compared to the reference sequence, for example, such deletions may occur at 1-10 locations (such as 1-5 locations, suitably 1 or 2 locations, in particular 1 location) and may, for example, involve the deletion of 50 or fewer amino acids at each location (such as 20 or fewer, in particular 10 or fewer, especially 5 or fewer). Suitably such deletions do not occur in the region of an epitope, and do not therefore have a significant impact on the immunogenic properties of the antigen.

The skilled person will recognise that a particular polypeptide variant may comprise substitutions, deletions and additions (or any combination thereof).

Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity (such as at least about 95%, at least about 98% or at least about 99%) to the associated reference sequence.

The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or sub-sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 70% identity, optionally 75%, 80%, 85%, 90%, 95%, 98% or 99% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using, for example, one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be “substantially identical.” This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 25 to about 50 amino acids or nucleotides in length, or optionally over a region that is 75-100 amino acids or nucleotides in length. Suitably, the comparison is performed over a window corresponding to the entire length of the reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, references to a segment in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g.*, *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to

other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, e.g., version 7.0 (Devereaux *et al., Nuc. Acids Res.* 12:387-395 (1984).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (website at www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of

the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

POLYNUCLEOTIDE IDENTIFICATION AND CHARACTERISATION

Polynucleotides encoding the *Clostridium difficile* CDTa, CDTb, Toxin A and Toxin B proteins of the invention may be identified, prepared and/or manipulated using any of a variety of well established techniques. For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs. Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena *et al., Proc. Natl. Acad. Sci. USA* 93:10614-10619 (1996) and Heller *et al., Proc. Natl. Acad. Sci. USA* 94:2150-2155 (1997)). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as *M. tuberculosis* cells. Such polynucleotides may be amplified via polymerase chain reaction (PCR). For this approach, sequence-specific primers may be designed based on the sequences provided herein, and may be purchased or synthesised.

An amplified portion of a polynucleotide may be used to isolate a full length gene from a suitable library (e.g., a *M. tuberculosis* cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridisation techniques, a partial sequence may be labeled (e.g., by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridising filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (see Sambrook *et al., Molecular Cloning: A Laboratory Manual* (2000)). Hybridising colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single

contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, there are numerous amplification techniques for obtaining a full length coding sequence from a partial cDNA sequence. Within such techniques, amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced as described above, and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia *et al., Nucl. Acids Res.* 16:8186 (1988)), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularised by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridises to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom *et al., PCR Methods Applic.* 1:111-19 (1991)) and walking PCR (Parker *et al., Nucl. Acids. Res.* 19:3055-60 (1991)). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

POLYNUCLEOTIDE EXPRESSION IN HOST CELLS

Polynucleotide sequences or fragments thereof which encode the *Clostridium difficile* CDTa, CDTb, Toxin A and Toxin B proteins, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

Natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognised by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesised, in whole or in part, using chemical methods well known in the art (*see* Caruthers, *M.* H. *et al., Nucl. Acids Res. Symp. Ser.* pp. 215-223 (1980), Horn *et al., Nucl. Acids Res. Symp. Ser.* pp. 225-232 (1980)). Alternatively, the protein itself may be produced using chemical methods to synthesize the

amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge *et al.*, *Science* 269:202-204 (1995)) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesised peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, *Proteins, Structures and Molecular Principles* (1983)) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook *et al., Molecular Cloning, A Laboratory Manual* (2000), and Ausube*l et al., Current Protocols in Molecular Biology* (updated annually).

A variety of expression vector/host systems may be utilised to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The “control elements” or “regulatory sequences” present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5’ and 3’ untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including

constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke &Schuster, *J. Biol. Chem.* 264:5503-5509 (1989)); and the like. pGEX Vectors (Promega, Madison, Wis; GE Healthcare.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae or Pichia such as Pichia pastoris for example*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. Other vectors containing constitutive or inducible promoters include GAP, PGK, GAL and ADH. For reviews, see Ausubel *et al. (supra*) and Grant *et al., Methods Enzymol.* 153:516-544 (1987) and Romas et al. *Yeast* 8 423-88 (1992).

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, *EMBO J.* 6:307-311 (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi et al., *EMBO J.* 3:1671-1680 (1984); Broglie *et al., Science* 224:838-843

(1984); and Winter *et al., Results Probl. Cell Differ.* 17:85-105 (1991)). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (*see, e.g.*, Hobbs *in McGraw Hill Yearbook of Science and Technology* pp. 191-196 (1992)).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia larvae*. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia larvae* in which the polypeptide of interest may be expressed (Engelhard *et al., Proc. Natl. Acad. Sci. U.S.A.* 91 :3224-3227 (1994)).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan & Shenk, *Proc. Natl. Acad. Sci. U.S.A.* 81:3655-3659 (1984)). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. Methods and protocols for working with adenovirus vectors are reviewed in Wold, *Adenovirus Methods and Protocols*, 1998. Additional references regarding use of adenovirus vectors can be found in *Adenovirus: A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References*, 2004.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the

initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf. *et al., Results Probl. Cell Differ.* 20:125-162 (1994)).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al., Cell* 11:223-32 (1977)) and adenine phosphoribosyltransferase (Lowy *et al., Cell* 22:817-23 (1990)) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler *et al., Proc. Natl. Acad. Sci. U.S.A.* 77:3567-70 (1980)); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al., J. Mol. Biol.* 150:1-14 (1981)); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows

cells to utilise indole in place of tryptophan, or hisD, which allows cells to utilise histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-51 (1988)). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes *et al., Methods Mol. Biol.* 55:121-131 (1995)).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridisations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilising monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton *et al., Serological Methods, a Laboratory Manual* (1990) and Maddox *et al., J. Exp. Med.* 158:1211-1216 (1983).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridisation or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labelling or PCR amplification using a labelled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the

production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen. San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilised metal ion affinity chromatography) as described in Porath *et al., Prot. Exp. Purif.* 3:263-281 (1992) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll *et al., DNA Cell Biol.* 12:441-453 (1993)).

POLYPEPTIDE COMPOSITIONS

Generally, a polypeptide of use in the invention (for example the *Clostridium difficile* CDTa, CDTb, Toxin A and Toxin B proteins) will be an isolated polypeptide (i.e. separated from those components with which it may usually be found in nature).

For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

Polypeptides may be prepared using any of a variety of well known techniques. Recombinant polypeptides encoded by DNA sequences as described above may be readily prepared from the DNA sequences using any of a variety of expression vectors known to those of ordinary skill in the art. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast, and higher eukaryotic cells, such as mammalian cells and plant cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. Supernatants from suitable host/vector systems which secrete recombinant protein or polypeptide into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant polypeptide.

Polypeptides for use in the invention, immunogenic fragments thereof, and other variants having less than about 100 amino acids, and generally less than about 50 amino acids, may also be generated by synthetic means, using techniques well known to those of ordinary skill in the art. For example, such polypeptides may be synthesised using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc.* 85:2149-2146 (1963). Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

Within certain specific embodiments, a polypeptide may be a fusion protein that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, e.g.*, Stoute *et al., New Engl. J. Med*. 336:86-91 (1997)). A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognised by humans, or may

assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein, allowing the production of increased levels, relative to a non-fused protein, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea *et al.*, *Gene* 40:39-46 (1985); Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262 (1986); U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

Adjuvants

In a further embodiment of any of the aspects of the invention, the immunogenic composition further comprises an adjuvant. In one embodiment the adjuvant comprises aluminium hydroxide or aluminium phosphate. Alternatively the immunogenic composition of the invention may comprise an aluminium-free adjuvant, the immunogenic composition is formulated with an adjuvant that is free of aluminum or aluminum salts, that is, an aluminum-free adjuvant or adjuvant system.

In certain embodiments, the immunogenic composition is formulated with an adjuvant comprising an immunologically active saponin fraction presented in the form of a liposome. The adjuvant may further comprise a lipopolysaccharide. The adjuvant may include QS21. For example, in one embodiment, the adjuvant contains QS21 in a liposomal formulation. In one embodiment, the adjuvant system includes 3D-MPL and QS21. For example, in one embodiment, the adjuvant contains 3D-MPL and QS21 in a liposomal formulation. Optionally, the adjuvant system also contains cholesterol. In one specific embodiment, the adjuvant includes QS21 and cholesterol. Optionally, the adjuvant system contains 1, 2-Dioleoyl-sn-Glycero-3-phosphocholine (DOPC). For example, in one specific adjuvant system contains cholesterol, DOPC, 3D-MPL and QS21.
In one specific example, the immunogenic composition includes an adjuvant formulated in a dose that includes: from about 0.1 to about 0.5 mg cholesterol; from about 0.25 to about 2 mg DOPC; from about 10 µg to about 100 µg 3D-MPL; and from about 10 µg to about 100 µg QS21. In another specific example, the immunogenic composition includes an adjuvant formulated in a dose that includes: from about 0.1 to about 0.5 mg cholesterol; from about 0.25 to about 2 mg DOPC; from about 10 µg to about 70 µg 3D-MPL; and from about 10 µg to about 70 µg QS21. In one specific formulation, the adjuvant is formulated in a single dose that contains: about 0.25 mg cholesterol; about 1.0 mg DOPC; about 50 µg 3D-MPL; and about 50 µg QS21. In other embodiments, the immunogenic composition is formulated with a fractional dose (that is a dose, which is a fraction of the preceding single dose formulations, such as one half of the preceding quantity of components (cholesterol, DOPC, 3D-MPL and QS21), ¼ of the preceding quantity of components, or another fractional dose (e.g., 1/3, 1/6, etc.) of the preceding quantity of components.
In one embodiment, the immunogenic compositions according to the invention include an adjuvant containing combinations of lipopolysaccharide and Quillaja saponins that have been disclosed previously, for example in EP0671948. This patent demonstrated a strong synergy when a lipopolysaccharide (3D-MPL) was combined with a Quillaja saponin (QS21).

The adjuvant may further comprise immunostimulatory oligonucleotides (for example, CpG) or a carrier.

A particularly suitable saponin for use in the present invention is Quil A and its derivatives. Quil A is a saponin preparation isolated from the South American tree *Quillaja Saponaria Molina* and was first described by Dalsgaard *et al*. in 1974 ("Saponin adjuvants", Archiv. für die gesamte Virusforschung, Vol. 44, Springer Verlag, Berlin, p243-254) to have adjuvant activity. Purified fragments of Quil A have been isolated by HPLC which retain adjuvant activity without the toxicity associated with Quil A (EP 0 362 278), for example QS7 and QS21 (also known as QA7 and QA21). QS21 is a natural saponin derived from the bark of *Quillaja saponaria* Molina, which induces CD8+ cytotoxic T cells (CTLs), Th1 cells and a predominant IgG2a antibody response and is a preferred saponin in the context of the present invention.

When the adjuvant comprises an immunologically active saponin fraction presented in the form of a liposome, the adjuvant may further comprise a sterol. Suitably the sterol is provided at a ratio of saponin:sterol of from 1:1 to 1:100 w/w, such as from 1:1 to 1:10w/w; or 1:1 to 1:5 w/w.

In a specific embodiment, QS21 is provided in its less reactogenic composition where it is quenched with an exogenous sterol, such as cholesterol for example. Several particular forms of less reactogenic compositions wherein QS21 is quenched with an exogenous cholesterol exist. In a specific embodiment, the saponin /sterol is in the form of a liposome structure (WO 96/33739, Example 1). In this embodiment the liposomes suitably contain a neutral lipid, for example phosphatidylcholine, which is suitably non-crystalline at room temperature, for example eggyolk phosphatidylcholine, dioleoyl phosphatidylcholine (DOPC) or dilauryl phosphatidylcholine. The liposomes may also contain a charged lipid which increases the stability of the lipsome-QS21 structure for liposomes composed of saturated lipids. In these cases the amount of charged lipid is suitably 1-20% w/w, preferably 5-10%. The ratio of sterol to phospholipid is 1-50% (mol/mol), suitably 20-25%.

Suitable sterols include β-sitosterol, stigmasterol, ergosterol, ergocalciferol and cholesterol. In one particular embodiment, the adjuvant composition comprises cholesterol as sterol. These sterols are well known in the art, for example cholesterol is disclosed in the Merck Index, 11th Edn., page 341, as a naturally occurring sterol found in animal fat.

Where the active saponin fraction is QS21, the ratio of QS21 : sterol will typically be in the order of 1:100 to 1:1 (w/w), suitably between 1:10 to 1:1 (w/w), and preferably 1:5 to 1:1

(w/w). Suitably excess sterol is present, the ratio of QS21:sterol being at least 1:2 (w/w). In one embodiment, the ratio of QS21:sterol is 1:5 (w/w). The sterol is suitably cholesterol.

In one embodiment, the invention provides a dose of an immunogenic composition comprising immunologically active saponin, preferably QS21, at a level of about 1 – about 70 µg per dose, for example at an amount of about 50 µg.

In one embodiment, the invention provides a dose of an immunogenic composition comprising immunologically active saponin, preferably QS21, at a level of 60µg or less, for example between 1 and 60µg. In one embodiment, the dose of the immunogenic composition comprises QS21 at a level of approximately around 50 µg, for example between 45 and 55 µg, suitably between 46 – 54 µg or between 47 and 53 µg or between 48 and 52 µg or between 49 and 51 µg, or 50 µg.

In another embodiment the dose of the immunogenic composition comprises QS21 at a level of around 25 µg, for example between 20 – 30 µg, suitably between 21 – 29 µg or between 22 and 28 µg or between 23 and 27 µg or between 24 and 26 µg, or 25 µg.

In another embodiment, the dose of the immunogenic composition comprises QS21 at a level of around 10 µg per, for example between 5 and 15 µg, suitably between 6 and 14 µg, for example between 7 and 13 µg or between 8 and 12 µg or between 9 and 11 µg, or 10µg.

Specifically, a 0.5 ml vaccine dose volume contains 25 µg or 50 µg of QS21 per dose. Specifically, a 0.5 ml vaccine dose volume contains 50 µg of QS21 per dose.

In compositions comprising a lipopolysaccharide, the lipopolysaccharide may be present at an amount of about 1 – about 70 µg per dose, for example at an amount of about 50 µg.

The lipopolysaccharide may be a non-toxic derivative of lipid A, particularly monophosphoryl lipid A or more particularly 3-Deacylated monophoshoryl lipid A (3D – MPL).

3D-MPL is sold under the name MPL by GlaxoSmithKline Biologicals S.A. and is referred throughout the document as MPL or 3D-MPL. See, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094. 3D-MPL primarily promotes CD4+ T cell responses with an IFN-γ (Th1) phenotype. 3D-MPL can be produced according to the methods disclosed in GB 2 220 211 A. Chemically it is a mixture of 3-deacylated monophosphoryl lipid A with 3, 4, 5 or 6 acylated chains. Preferably in the compositions of the present invention small particle 3D-MPL is used. Small particle 3D-MPL has a particle size such that

it may be sterile-filtered through a 0.22μm filter. Such preparations are described in WO 94/21292.

The invention therefore provides a dose of an immunogenic composition comprising lipopolysaccharide, preferably 3D-MPL, at a level of 75μg or less, for example between 1 and 60μg.

In one embodiment, the dose of the immunogenic composition comprises 3D-MPL at a level of around 50 μg, for example between 45 – 55 μg, suitably between 46 – 54 μg or between 47 and 53 μg or between 48 and 52 μg or between 49 and 51 μg, or 50 μg.

In one embodiment, the dose of the immunogenic composition comprises 3D-MPL at a level of around 25 μg, for example between 20 – 30 μg, suitably between 21 – 29 μg or between 22 and 28 μg or between 23 and 27 μg or between 24 and 26 μg, or 25 μg.

In another embodiment, the dose of the immunogenic composition comprises 3D-MPL at a level of around 10 μg, for example between 5 and 15 μg, suitably between 6 and 14 μg, for example between 7 and 13 μg or between 8 and 12 μg or between 9 and 11 μg, or 10μg.

In one embodiment, the volume of the dose is 0.5 ml. In a further embodiment, the immunogenic composition is in a volume suitable for a dose which volume is higher than 0.5 ml, for example 0.6, 0.7, 0.8, 0.9 or 1 ml. In a further embodiment, the human dose is between 1 ml and 1.5 ml.

Specifically, a 0.5 ml vaccine dose volume contains 25 μg or 50 μg of 3D-MPL per dose. Specifically, a 0.5 ml vaccine dose volume contains 50 μg of 3D-MPL per dose.

The dose of the immunogenic composition according to any aspect of the invention suitably refers to human dose. By the term “human dose” is meant a dose which is in a volume suitable for human use. Generally this is between 0.3 and 1.5 ml. In one embodiment, a human dose is 0.5 ml. In a further embodiment, a human dose is higher than 0.5 ml, for example 0.6, 0.7, 0.8, 0.9 or 1 ml. In a further embodiment, a human dose is between 1 ml and 1.5 ml.

Suitable compositions of the invention are those wherein liposomes are initially prepared without MPL (as described in WO 96/33739), and MPL is then added, suitably as small particles of below 100 nm particles or particles that are susceptible to sterile filtration through

a 0.22 µm membrane. The MPL is therefore not contained within the vesicle membrane (known as MPL out). Compositions where the MPL is contained within the vesicle membrane (known as MPL in) also form an aspect of the invention. The polypeptide comprising a *C. difficile* toxin A fragment and / or a *C. difficile* toxin B fragment can be contained within the vesicle membrane or contained outside the vesicle membrane.

In a specific embodiment, QS21 and 3D-MPL are present in the same final concentration per dose of the immunogenic composition i.e. the ratio of QS21:3D-MPL is 1:1. In one aspect of this embodiment, a dose of immunogenic composition comprises a final level of 25 µg of 3D-MPL and 25 µg of QS21 or 50 µg of 3D-MPL and 50 µg of QS21.

In one embodiment, the adjuvant includes an oil-in-water emulsion. In one embodiment the adjuvant comprises an oil in water emulsion, wherein the oil in water emulsion comprises a metabolisable oil, a tocol and an emulsifier. For example, the oil-in-water emulsion can include an oil phase that incorporates a metabolisable oil, and an additional oil phase component, such as a tocol. The oil-in-water emulsion may also contain an aqueous component, such as a buffered saline solution (e.g., phosphate buffered saline). In addition, the oil-in-water emulsion typically contains an emulsifier. In one embodiment, the metabolizable oil is squalene. In one embodiment, the tocol is alpha-tocopherol. In one embodiment, the emulsifier is a nonionic surfactant emulsifier (such as polyoxyethethylene sorbitan monooleate, Polysorbate® 80 , TWEEN80™). In exemplary embodiments, the oil-in-water emulsion contains squalene and alpha tocopherol in a ratio which is equal or less than 1 (w/w).

The metabolisable oil in the oil-in-water emulsion may be present in an amount of 0.5-10mg. The tocol in the oil-in-water emulsion may be present in an amount of 0.5 – 11 mg. The emulsifying agent may be present in an amount of 0.4 – 4 mg,

In order for any oil in water composition to be suitable for human administration, the oil phase of the emulsion system has to comprise a metabolisable oil. The meaning of the term metabolisable oil is well known in the art. Metabolisable can be defined as 'being capable of being transformed by metabolism' (Dorland's Illustrated Medical Dictionary, W.B. Sanders Company, 25th edition (1974)). The oil may be any vegetable oil, fish oil, animal oil or synthetic oil, which is not toxic to the recipient and is capable of being transformed by metabolism. Nuts, seeds, and grains are common sources of vegetable oils. Synthetic oils are also part of this invention and can include commercially available oils such as NEOBEE® (caprylic/capric triglycerides made using glycerol from vegetable oil sources and

medium-chain fatty acids (MCTs) from coconut or palm kernel oils) and others. A particularly suitable metabolisable oil is squalene. Squalene (2,6,10,15,19,23-Hexamethyl-2,6,10,14,18,22-tetracosahexaene) is an unsaturated oil which is found in large quantities in shark-liver oil, and in lower quantities in olive oil, wheat germ oil, rice bran oil, and yeast, and is a particularly preferred oil for use in this invention. Squalene is a metabolisable oil by virtue of the fact that it is an intermediate in the biosynthesis of cholesterol (Merck index, 10th Edition, entry no.8619).

Suitably the metabolisable oil is present in the adjuvant composition in an amount of 0.5-10 mg, preferably 1-10, 2-10, 3-9, 4-8, 5-7, or 5-6 mg (e.g. 2-3, 5-6, or 9-10mg), specifically about 5.35 mg or about 2.14 mg per dose.

Tocols are well known in the art and are described in EP0382271. Suitably the tocol is alpha-tocopherol or a derivative thereof such as alpha-tocopherol succinate (also known as vitamin E succinate). Said tocol is suitably present in in an amount of 0.5-11 mg, preferably 1-11, 2-10, 3-9, 4-8, 5-7, 5-6 mg (e.g. 10-11, 5-6, 2.5-3.5 or 1-3 mg). In a specific embodiment the tocol is present in an amount of about 5.94 mg or about 2.38 mg per dose.

The oil in water emulsion further comprises an emulsifying agent. The emulsifying agent may suitably be polyoxyethylene sorbitan monooleate. In a particular embodiment the emulsifying agent may be Polysorbate® 80 (Polyoxyethylene (20) sorbitan monooleate) or Tween® 80.

Said emulsifying agent is suitably present in the adjuvant composition in an amount of 0.1-5, 0.2-5, 0.3-4, 0.4-3 or 2-3 mg (e.g. 0.4-1.2, 2-3 or 4-5 mg) emulsifying agent. In a specific embodiment the emulsifying agent is present in an amount of about 0.97 mg or about 2.425 mg.

In one embodiment, the amounts of specific components present in the composition are the amounts present in a 0.5 ml human dose. In a further embodiment, the immunogenic composition is in a volume suitable for a human dose which volume is higher than 0.5 ml, for example 0.6, 0.7, 0.8, 0.9 or 1 ml. In a further embodiment, the human dose is between 1 ml and 1.5 ml.

Where the adjuvant is in a liquid form and is to be combined with a liquid form of a polypeptide composition, the adjuvant composition in a human dose will be a fraction of the intended final volume of the human dose, for example approximately half of the intended final volume of the human dose, for example a 350 µl volume for an intended human dose of

0.7ml, or a 250 µl volume for an intended human dose of 0.5 ml. The adjuvant composition is diluted when combined with the polypeptide antigen composition to provide the final human dose of vaccine. The final volume of such dose will of course vary dependent on the initial volume of the adjuvant composition and the volume of polypeptide antigen composition added to the adjuvant composition. In an alternative embodiment, a liquid adjuvant is used to reconstitute a lyophilised polypeptide composition. In this embodiment, the human dose of the adjuvant composition is approximately equal to the final volume of the human dose. The liquid adjuvant composition is added to the vial containing the lyophilised polypeptide composition. The final human dose can vary between 0.5 and 1.5 ml.

The method of producing oil-in-water emulsions is well known to the person skilled in the art. Commonly, the method comprises mixing the tocol-containing oil phase with a surfactant such as a PBS/ polyoxyethylene sorbitan monooleate solution, followed by homogenisation using a homogenizer. It would be clear to a man skilled in the art that a method comprising passing the mixture twice through a syringe needle would be suitable for homogenising small volumes of liquid. Equally, the emulsification process in microfluidiser (M110S Microfluidics machine, maximum of 50 passes, for a period of 2 minutes at maximum pressure input of 6 bar (output pressure of about 850 bar)) could be adapted by the man skilled in the art to produce smaller or larger volumes of emulsion. The adaptation could be achieved by routine experimentation comprising the measurement of the resultant emulsion until a preparation was achieved with oil droplets of the required diameter.

In an oil in water emulsion, the oil and emulsifier should be in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline.

Preferably the oil-in-water emulsion systems of the present invention have a small oil droplet size in the sub-micron range. Suitably the droplet sizes will be in the range 120 to 750 nm, more preferably sizes from 120 to 600 nm in diameter. Most preferably the oil-in water emulsion contains oil droplets of which at least 70% by intensity are less than 500 nm in diameter, more preferably at least 80% by intensity are less than 300 nm in diameter, more preferably at least 90% by intensity are in the range of 120 to 200 nm in diameter.

In one embodiment, the immunogenic composition is not 3µg or 10 µg of any of SEQ ID Nos. 1 to 7 combined with an adjuvant comprising an oil in water **emulsion having** 0.125 mL SB62 emulsion (Total volume), 5.35 mg squalene, 5.94 mg DL-α-tocopherol and 2.425 mg polysorbate 80 per 0.5 ml dose. In one embodiment, the immunogenic composition is not 3µg or 10 µg of any of SEQ ID Nos. 1 to 7 combined with an adjuvant comprising an oil in

water **emulsion** 5.35 mg squalene, 5.94 mg DL-α-tocopherol and 2.425 mg polysorbate 80 per 0.5 ml dose. In one embodiment, the immunogenic composition does not contain an adjuvant comprising a oil in water **emulsion having** squalene, DL-α-tocopherol and polysorbate 80.

Immunogenic compositions and vaccines of the invention

In one embodiment the immunogenic composition has a volume of 0.5 to 1.5 ml.
In one embodiment the immunogenic composition further comprises additional antigens. In one embodiment the additional antigens are antigens derived from a bacterium selected from the group consisting of *S.pneumoniae*, *H.influenzae*, *N.meningitidis*, *E.coli*, *M.catarrhalis*, *Clostridium tetani* (tetanus), *Corynebacterium diphtheria* (diphtheria), *Bordetella pertussis* (pertussis), *S.epidermidis*, enterococci, *S.aureus,* and *Pseudomonas aeruginosa.* In a further embodiment the immunogenic composition of the invention may comprise further antigens from *C.difficile* for example the S-layer proteins (WO01/73030). Optionally the immunogenic composition further comprises a saccharide from *C.difficile*.
There is further provided a vaccine comprising an immunogenic composition of the invention and a pharmaceutically acceptable excipient.
The vaccine preparations containing immunogenic compositions of the present invention may be used to protect a mammal susceptible to *C.difficile* infection or treat a mammal with a *C.difficile* infection, by means of administering said vaccine via systemic or mucosal route. These administrations may include injection via the intramuscular, intraperitoneal, intradermal or subcutaneous routes; or via mucosal administration to the oral/alimentary, respiratory, genitourinary tracts. Although the vaccine of the invention may be administered as a single dose, components thereof may also be co-administered together at the same time or at different times (for instance pneumococcal saccharide conjugates could be administered separately, at the same time or 1-2 weeks after the administration of the any bacterial protein component of the vaccine for coordination of the immune responses with respect to each other). In addition to a single route of administration, 2 different routes of administration may be used. For example, saccharides or saccharide conjugates may be administered intramuscularly (IM) or intradermally (ID) and bacterial proteins may be administered intranasally (IN) or intradermally (ID). In addition, the vaccines of the invention may be administered IM for priming doses and IN for booster doses.
The content of toxins in the vaccine will typically be in the range 1-250μg, preferably 5-50μg, most typically in the range 5 - 25μg. Following an initial vaccination, subjects may receive one or several booster immunizations adequately spaced. Vaccine preparation is generally described in Vaccine Design ("The subunit and adjuvant approach" (eds Powell M.F. &

Newman M.J.) (1995) Plenum Press New York). Encapsulation within liposomes is described by Fullerton, US Patent 4,235,877.

In one aspect of the invention is provided a vaccine kit, comprising a vial containing an immunogenic composition of the invention, optionally in lyophilised form, and further comprising a vial containing an adjuvant as described herein. It is envisioned that in this aspect of the invention, the adjuvant will be used to reconstitute the lyophilised immunogenic composition.

A further aspect of the invention is a method of preventing or treating *C.difficile* infection comprising administering to the host an immunoprotective dose of the immunogenic composition or vaccine or kit of the invention. In one embodiment there is provided a method of preventing or treating primary and/or recurrence episodes of *c.difficile* infection comprising administering to the host an immunoprotective dose of the immunogenic composition or vaccine or kit of the invention.

In one embodiment of the invention there is provided an immunogenic composition or vaccine of the invention for use in the treatment or prevention of *C.difficile* disease. In a further embodiment of the invention there is provided an immunogenic composition or vaccine of the invention for use in the treatment or prevention of disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571. Preferably the strain is strain 078.

In a further aspect of the invention there is provided a use of an immunogenic composition or vaccine of the invention in the preparation of a medicament for the prevention or treatment of *C.difficile* disease. In a further embodiment the disease is a disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571. Preferably the strain is strain 078.

In a further aspect of the invention there is provided a method of preventing or treating *C.difficile* disease comprising administering the immunogenic composition of the invention or the vaccine of the invention to a mammalian subject such as a human subject. In a further embodiment the disease is a disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571. Preferably the strain is strain 078.

General

Around" or "approximately" are defined as within 10% more or less of the given figure for the purposes of the invention.

The terms “comprising”, “comprise” and “comprises” herein are intended by the inventors to be optionally substitutable with the terms “consisting of”, “consist of” and “consists of”, respectively, in every instance. The term “comprises” means “includes.” Thus, unless the context requires otherwise, the word “comprises,” and variations such as “comprise” and “comprising” will be understood to imply the inclusion of a stated compound or composition (*e.g.*, nucleic acid, polypeptide, antigen) or step, or group of compounds or steps, but not to the exclusion of any other compounds, composition, steps, or groups thereof. The abbreviation, “*e.g.*” is derived from the Latin exempli gratia, and is used herein to indicate a non-limiting example. Thus, the abbreviation “*e.g.*” is synonymous with the term “for example.”

The amino acid numbering used herein is derived from the sequences for CDTa, CDTb, Toxin A and Toxin B presented herein as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 31 and SEQ ID NO: 32 which are to be considered as reference sequences for these proteins.

Embodiments herein relating to “vaccine compositions” of the invention are also applicable to embodiments relating to “immunogenic compositions” of the invention, and vice versa.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Definitions of common terms in molecular biology can be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. The term “plurality” refers to two or more. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Additionally, numerical limitations given with respect to concentrations or levels of a substance, such as an antigen, may be approximate.

All references or patent applications cited within this patent specification are incorporated by reference herein in their entirety.

In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only, and are not to be construed as limiting the scope of the invention in any manner.

EXAMPLES

The AS01B adjuvant referred to is an adjuvant having 50 µg QS21 presented in the form of a liposome, 50 µg 3D-MPL, 0.25 mg cholesterol and 1.0 mg DOPC per 0.5ml dose. A dose of 50µl suitable for immunizing mice contains 5µg QS21, 5µg 3D-MPL, 0.025mg cholesterol and 0.1mg DOPC.

Example 1: Design of binary toxin antigens

The Binary Toxin (other name: ADP-ribosyltransferase toxin) is composed by two components : the enzymatic component, named CDTa and the transport and binding component, named CDTb.
Based on literature data and the known 3D structure of CDTa (*J. Biol. Chem.* 2009, vol. 284: 28713-28719), this protein could be divided into two domains. The N-terminal domain binds to CDTb and the C-terminal domain contains the enzymatic activity. Both domains are linked by a flexible peptide.
Based on literature data and information available for other B components of other bacterial binary toxins, CDTb could be divided into five domains. The first one is the prodomain, its cleavage by an enzyme having a chymotrypsin activity allows the heptamerization of the mature protein. The second domain allows the binding to CDTa. The third and fourth ones are involved in the oligomerisation and membrane insertion. Finally, the last domain is the host cell receptor binding domain.

Example 1a: Design of CDTa antigens

In order to be allowed to work with CDTa and CDTb together, CDTa must be inactivated. Two possibilities of inactivation were evaluated. The first one is the design of CDTa mutants that abolish the enzymatic activity and the second one is the use of the N-terminal domain of CDTa alone. This latter domain allows the binding to CDTb and does not contain residue involved in the enzymatic activity.

The first set of mutants were designed based on literature information (*Infection & Immunity*, 2001, vol. 69 : 6004-6011). Authors demonstrated that CDTa mutant proteins E428Q, E430Q, S388A and R345K have a significant reduced activity. Based on data shown in the publication, two mutations were preferred amongst the four : CDTa mutant E428Q and E430Q. In the publication, these mutants abolish completely the CDTa enzymatic activity. In order to rank these mutants, some structural analyses were performed for these residues :

surface-accessibility of the residues glutamate 428 (E428) and glutamate 430 (E430), effect of their mutations on the surrounding 3D structure. Based on these analyses, the CDTa mutant E428Q was chosen as preferred mutation and the CDTa mutant E430Q was selected as second choice. A double mutant E428Q, E430Q was also generated in order to be sure that the enzymatic activity was abolished.

A second set of mutants was designed because the first cytotoxicity results, obtained with the first set of mutants, were not convincing.

In this second set, a CDTa mutant containing 7 mutations (including the two mutations already described) was designed. All these mutations were designed based on literature information (available for CDTa or its Clostridium perfringens homologs Ia) and 3D structure analysis. All mutated residues are located around the catalytic site of CDTa. These residues have been modified in order to avoid ligand or water molecule binding. This CDTa "super mutant" contains the mutations R345A, Q350A, N385A, R402A, S388F, E428Q and E430Q.

Based on this "super mutated" CDTa, two other CDTa mutated variants were evaluated in order to eliminate E428Q and E430Q mutations (construct C108 contains the E430Q mutation but not the mutation E428Q, construct C107 does not contain both mutations).

CDTa Nterminal domain alone

It was described in the literature (*Infection & Immunity*, 2001, vol. 69 : 6004-6011) that the $CDTa^{1-240}$ is the minimal CDTa fragment that still allows a binding to Ib (B component of binary toxin of *Clostridium perfringens*). This fragment will be tested in the lab but based on known 3D structure, it was suggested that this domain will probably not be optimal in term of correct folding of this domain of CDTa.

Antigen design was performed based on the known 3D structure (Protein Data Bank accession number : 2WN4, *J. Biol. Chem.*, 2009, vol. 284 : 28713-28719) to improve the expression and folding of an isolated CDTa N-terminal domain. On the 3D structure, a linker peptide of eight amino acids allows the separation between the N- and C-terminal domains of CDTa. Two isolated CDTa N-terminal domains were designed, the first one contains this flexible peptide and the second one not.

CDTa : sequences summary

A summary of all CDTa sequences is presented in table 1.

Table 1

Name	Length (aa)*	Location	Comments
CDTa	463	1-463	Full length of CDTa coming from strain R20291
CDTa'	421	44-463	CDTa without signal peptide **(C34)**
CDTa_E428Q	421	44-463	CDTa' with mutation of Glu^{428} into Gln **(C44)**

CDTa_E430Q	421	44-463	CDTa' with mutation of Glu^{430} into Gln **(C54)**
CDTa_E428 430Q	421	44-463	CDTa' with the two mutations Glu^{428} into Gln and Glu^{430} into Gln **(C67)**
CDTa_7mutations	421	44-463	CDTa' containing 7 mutated amino acids **(C69)**
CDTa_N_litt	198	44-240	Minimum CDTa N-terminal domain that still allows binding to Ib **(C51)**
CDTa_NADlink	226	44-268	CDTa N-terminal domain based on antigen design work **(C49)**
CDTa_NAD	218	44-260	CDTa N-terminal domain based on antigen design work **(C50)**

*Length contains additional N-terminal Methionine but not the His-tag

Example 1b: Design of CDTb antigens

CDTb mature

In order to avoid the chymotrypsin activation step in the CDTb process, it was tried to express only the mature CDTb protein (without its signal peptide and prodomain).

In the literature (*Protein Expression and Purification*, 2010, vol. 74 : 42-48), the mature CDTb was described as starting at Leucine 210. This mature CDTb was named CDTb".

After in house experimental data, it seems that the activated CDTb starts at Serine 212. This result was supported by analysis of a 3D modelised structure of CDTb. This model was built using SwissModel (*Bioinformatics*, 2006, vol. 22 : 195-201). The template used for the homology modeling was the B component of *Bacillus anthracis*, named Protective Antigen or PA (Protein Data Bank accession number : 3TEW).

CDTb receptor-binding domain alone

Given the fact that a fusion containing only the receptor-binding domains of Toxin A and B is sufficient to induce neutralizing antibodies, it was decided to produce and evaluate the CDTb receptor-binding domain alone.

The 3D structure model obtained for CDTb is accurate for the four first domains of CDTb but not for the receptor-binding domain (these domains of CDTb and PA are too different). To design constructs expressing this domain alone, the C-terminal part of the fourth domain was analysed on the 3D structure model in order to decide where the last domain will start.

Two versions of the CDTb-receptor binding domain were designed. In the first one, this domain starts just after the modelised 3D structure of the fourth domain. In this version, the CDTb-receptor-binding domain will probably have a long flexible peptide in its N-terminal

part. The second version starts where the 2D predicted structures performed on the C-terminal part of CDTb (predictions done using Psipred program, *Bioinformatics*, 2000, vol. 16 : 404-405) become more compacts after a lack of predicted secondary structures. This could indicate the beginning of a new structural domain. In this second version, no flexible peptide is present at the N-terminal part of the isolated CDTb receptor-binding domain.

CDTb Ca^{2+} binding motif mutation

Following literature information, mutations in the Ca^{2+} binding domain of the B component of Iota toxin of *Clostridium perfringens* (Ib) abolish the binding with the A component of this binary toxin (Ia). These mutations could be very interesting in the case of a vaccine composition containing a mixture of mature CDTb protein and a wild type CDTa protein.

Using multiple protein sequence alignment, these mutations were located on the CDTb sequence and mutated. It concerns residues Asp^{220}, Asp^{222} and Asp^{224}. They were mutated into Ala residues.

CDTb prodomain

In order to try to decrease the degradation issues observed with C55 in gel, some co-expression tests were evaluated. The working hypothesis of doing that is to improve the folding of the mature CDTb.

Two limits of prodomain were proposed. The first one starts at residue 43 of CDTb (after the signal peptide cleavage) and finishes at residue Met^{211} (given that the experimentally determined first residue of the mature CDTb is Ser^{212}). The second prodomain was designed based on the predicted 3D structure of CDTb. The linker existing between the prodomain and the first structural domain of the mature CDTb protein is removed in this construct.

CDTb : sequence summary

A summary of all CDTb sequences is presented in table 2.

Table 2

Name	**Length (aa)***	**Location**	**Comments**
CDTb	876	1-876	Full length of CDTb coming from strain R20291
CDTb'	835	43-876	CDTb without signal peptide **(C38)**
CDTb"	668	210-876	CDTb without signal peptide and prodomain as defined in the literature
CDTb"_xp data	666	212-876	CDTb without signal peptide and prodomain, as demonstrated by in house experimental results **(C55)**

CDTbClg	258	620-876	CDTb receptor-binding domain containing natural flexible peptide in its N-terminal part, based on antigen design work **(C52)**
CDTbCsh	242	636-876	CDTb receptor-binding domain, based on antigen design work **(C53)**
CDTb Ca2+ mutated	666	212-876	Mature CDTb (without signal peptide and without prodomain) containing 3 mutations D220A, D222A and D224A **(C97)**
CDTbprodomainLg	170	43-211	CDTb prodomain **(C58)**
CDTbprodomainSh	145	43-186	CDTb prodomain without the linker existing between the prodomain and the first structural domain of mature CDTb **(C59)**

*Length contains additional N-terminal Methionine but not the His-tag

Example 1c: Design of CDTa – CDTb fusions

Background information

The aim of these constructs is to obtain both components of the Binary Toxin into one process.

A lot of different kinds of fusions could be designed but, as proof of concept, the first fusion evaluated is the combination of CDTa N-terminal domain (named CDTaNADlink and CDTaNAD) with CDTb receptor-binding domain (named CDTbCsh and CDTbClg).

Fusion CDTaNterm - CDTb receptor-binding domain alone

Without additional experimental data on each partner of the fusion, all possible combinations were initiated but always with the CDTa domain as first partner of the fusion.

In these fusions, the CDTaNADlink and the CDTaNAD domains have two and one residues less than the designed isolated domains, respectively. These CDTa additional amino acids were kept in the isolated designs in order to avoid potential issues during the expression process.

A summary of all CDTa-CDTb fusion sequences is presented in table 3.

Table 3

Name	**Length (aa)***	**Location**	
		CDTa	CDTb
CDTaNADlink-CDTbCsh **(C61)**	465	44-266	636-876

CDTaNAD-CDTbCsh **(C63)**	458	44-259	636-876
CDTaNADlink-CDTbClg **(C60)**	481	44-266	620-876
CDTaNAD-CDTbClg **(C62)**	474	44-259	620-876

*Length contains additional N-terminal Methionine but not the His-tag

Example 1d: design of ToxA-ToxB- CDTb receptor-binding domain fusion

The objective of this fusion is the combination of receptor-binding domains of the three major toxins of *Clostridium difficile* into one construct.

Given the fact that F2 and CDTb receptor-binding domains are not supposed to adopt the same fold, a linker / spacer must be used between the two partners of the fusion in order to allow a correct independent folding of them. Two fusions were designed.

In the first one (named F2_CDTbClg), the long designed version of the receptor-binding domain is fused at the C-terminal part of F2. In this version, the long flexible N-terminal peptide of CDTb receptor-binding domain will function as spacer.

In the second fusion (named F2_GG_NVCDTbCsh), the short designed version of the receptor-binding domain is fused at the C-terminal part of F2. In order to allow a correct folding of the two partners, the length of the linker created in this fusion must be increased. In order to do that, the CDTb-receptor binding domain was extended with two natural residues, moreover two exogenous Glycines were added between F2 and the longer version of CDTbCsh.

A summary of all F2-CDTb fusion sequences is presented in table 4.

Table 4

Name	**Length (aa)***	**Location**		**Comments**
		F2	CDTb	
F2_CDTbClg	1223	ToxA : 2121-2686 ToxB : 1968-2366	620-876	
F2_GG_NVCDTbCsh	1211	ToxA : 2121-2686 ToxB : 1968-2366	634-876	Two exogenous Gly residues added as spacer between the two partners of the fusion

*Length contains additional N-terminal Methionine but not the His-tag

Example 2: Cloning, expression and purification of CdtA protein

Expression plasmid and recombinant strain: CdtA full length

Genes encoding the protein of full length without signal peptide of CdtA with and without mutations (see tables below) and a His tag in C-term were cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. Final constructs were generated by the transformation of *E. coli* strain HMS174 (DE3) or BLR (DE3) pLysS (C34) with each recombinant expression vector separately according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

CdtA

C number	
C34	CdtA (aa44-463)
C44	CdtA (aa44-463) mut. E428Q
C49	CdtA linker (44-268)
C50	CdtA WO linker (44-260)
C54	CdtA (aa44-463) mut. E430Q
C67	CdtA (aa44-463) mut. E428Q-E430Q
C68	CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-E428Q-E430Q
C69	CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F-E428Q-E430Q
C107	CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F
C108	CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F-E430Q
C110	CdtA (aa44-463) mut. R345A-Q350A-N385A-R402A-S388F-E428Q

Host strain:

HMS 174 (DE3). HMS174 strains provide the *recA* mutation in a K-12 background. Strains having the designation (DE3) are lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase. λ DE3 lysogens are designed for protein expression from pET vectors **Genotype:** F^- *recA1 hsdR*(r_{K12}^- m_{K12}^+) (Rif^R).

BLR(DE3) pLysS. BLR is a recA derivative of BL21. Strains having the designation (DE3) are lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases, pLysS strains express T7 lysozyme wich further supress basal expression of the T7 RNA polymerase prior to induction.

Genotype : *E.coli* BLR::DE3 strain, F^- *ompT* *hsdS*$_B$(r_B^- m_B^-) *gal dcm* (*DE3*) Δ(srl-recA)306::Tn*10* pLysS (Cam^R, Tet^R).

Expression of the recombinant proteins:

E.coli transformants were stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) to obtain O.D.600nm between 0.1 -0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

Each overnight culture were diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until O.D.620 reached 0.5/0.6.

At O.D.600nm around 0.6, the cultures were cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After overnight induction (around 16 hours), $O.D._{600nm}$ were evaluated after induction and culture were centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Expression plasmid and recombinant strain: CdtA- N-term

Genes encoding protein of N-terminal, without signal peptide of CdtA (see tables below) and a His tag in C-term were cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. Final constructs were generated by the transformation of *E. coli* strain HMS174 (DE3) with each recombinant expression vectors separately according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid

transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

CdtA N-term

C number	
C49	CdtA linker (44-268)
C50	CdtA WO linker (44-260)

Host strain:

HMS 174 (DE3). HMS174 strains provide the *recA* mutation in a K-12 background. Strains having the designation (DE3) are lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase. λ DE3 lysogens are designed for protein expression from pET vectors

Genotype: F^- *recA1 hsdR*(r_{K12}^- m_{K12}^+) (Rif^R).

Expression of the recombinant proteins:

E.coli transformants were stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) to obtain O.D.600nm between 0.1 -0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

This overnight culture was diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until O.D.620 reached 0.5/0.6.

At O.D.600nm around 0.6, the culture was cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After overnight induction (around 16 hours), $O.D._{600nm}$ was evaluated after induction and culture was centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Purification

The following procedure was used to purify constructs C34, C44, C49, C50, C54, C67, C69, C107 and C110.

The bacterial pellets were re-suspended in 20mM or 50mM bicine buffers (pH 7.5 or pH 8.0), containing 500mM NaCl , 0mM or 5mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl)phosphine hydrochloride) and a mixture of protease inhibitors (Complete, Roche, without EDTA). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-proteins were purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with the same buffer used to bacterial re-suspension. After loading on the column, the column was washed with a 20mM or 50mM bicine buffer (pH7.5 or pH8.0), containing 500mM NaCl, 10mM imidazole, 5mM TCEP. Elution was performed using a 50mM bicine buffer pH7.6, 500mM NaCl, 1mM TCEP and imidazole (250mM or 500mM).

After desalting (BIORAD Bio-Gel P6 Desalting) and concentration (Amicon Ultra 10kDa) steps, the product was loaded on SEC chromatography (SUPERDEX ^{TM} 75 or 200) in 20mM or 50mM bicine buffer(pH7.5 or pH8.0), 150mM NaCl, 1mM TCEP, for further purification step.

Fractions containing Cdta antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

Example 4 – Cloning, expression and purification of *C.difficile* CdtB protein

Expression plasmid and recombinant strain: CdtB Full length.

Genes encoding the truncated protein of CdtB without signal peptide (Pro-CdtB`) and a His tag in C-term were cloned into the pGEX-6p1 expression vector (GE Healthcare) using the BamHI/XhoI restriction sites using standard procedures. This vector included GST (Gluthathione-S-transferase) as fusion partner in N-terminal of either CdtB`(GST-Pro-Cdtb'). The final construct was generated by the transformation of *E. coli* strain BL21 (DE3) with the recombinant expression vector according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

Genes encoding the truncated protein of CdtB without signal peptide (Pro-CdtB`: C38) and without signal peptide and prodomain (CdtB``: C40 or C55) and a His tag in C-term were

cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. Final constructs were generated by the transformation of *E. coli* B834 (DE3) modified strain for C55 and BL21(DE3) for C38 and C40 with the appropriate recombinant expression vectors according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

CdtB

C number	
C37	CdtB' Δsignal sequence (aa43-876) + GST N-term
C38	CdtB' Δsignal sequence (aa43-876)
C40	CdtB''Δ signal sequence and prodomain (aa210-876)
C55	CdtB Δ prodomain (aa. 212-876)

Host Strain

BL21(DE3). BL21(DE3) is a non-methionine auxotroph derivative of B834. Strains having the designation (DE3) are lysogenic for a λ prophage that contains an IPTG inducible T7 RNA polymerase. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Genotype : *E.coli* BL21::DE3 strain, F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (*DE3*).

B834 is the parental strain for BL21. These protease-deficient hosts are methionine auxotrophs. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Modification: Including PGL gene to avoid phosphogluconoylation in the biotin locus (Strain is auxotroph for biotin).

Genotype : B834 ::DE3 strain, F– *ompT hsdS*B(rB – mB–) *gal dcm met* (DE3)

Modification : Δ(bioA-bioD)::PGL

Expression of the recombinant proteins:

E.coli transformants were stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose +/- kanamycin (50 µg/ml) or ampicillin (100 µg/ml) to obtain $O.D._{600nm}$ between 0.1 -0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

Overnight culture were diluted to 1:20 in 500 ml of LBT medium containing +/- kanamycin (50 µg/ml) or ampicillin (100 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until $O.D._{620}$ reached 0.5/0.6.

At an O.D.at 600nm of around 0.6, cultures were cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After the overnight inductions (around 16 hours), O.D. at 600nm were evaluated after induction and cultures were centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Expression plasmid and recombinant strain.

Genes encoding the truncated protein of CdtB without prodomain CdtB mature, Knock-out of Ca++ bindinding site (inhibit binding of CdtA to CdtB) and a His tag in C-term was cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. The final construct was generated by the transformation of *E. coli* B834 (DE3) modified strain with the recombinant expression vector according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

CdtB Ca++ binding domain knock-out

C number	
C97	CdtB mature mut Asp 220-222-224 Ala (aa. 212-876)

Host Strain

B834 is the parental strain for BL21. These protease-deficient hosts are methionine auxotrophs. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Modification: Including PGL gene to avoid phosphogluconoylation in the biotin locus (Strain is auxotroph for bioti n).

Genotype : B834 ::DE3 strain, F– *ompT hsdS*B(rB – mB–) *gal dcm met* (DE3)
Modification : Δ(bioA-bioD)::PGL

Expression of the recombinant proteins:

A *E.coli* transformant was stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) to obtain $O.D._{600nm}$ between 0.1 -0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

This overnight culture was diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until $O.D._{620}$ reached 0.5/0.6.

At an O.D.at 600nm of around 0.6, the culture was cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After the overnight induction (around 16 hours), O.D. at 600nm was evaluated after induction and culture was centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Purification

C37

The bacterial pellet was re-suspended in 50mM bicine buffer (pH 8.0) containing 500mM NaCl , 5mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl)phosphine hydrochloride) and a mixture of protease inhibitor (Complete, Roche). Bacteria were lysed using a French Press system
3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with the same buffer used to bacterial re-suspension. After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 150mM NaCl, 25mM imidazole, 1mM TCEP. Elution was performed using a 50mM bicine buffer pH8.0 containing 150mM NaCl, 250mM imidazole, 1mM TCEP.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing
150mM NaCl and 1mM TCEP, the product was treated (overnight at 4°C) with PreScission protease (GE-Healthcare) in order to cleave the GST tag. After overnight treatment, 0.2% Tween 20 was added to the digestion mixture.
Then the protein was passed through a GST affinity column (GE GSTrap FF) pre-equilibrated with buffer 50mM bicine buffer pH8.0 containing 150mM NaCl , 1mM TCEP, 0.2% tween20 and 20mM reduced glutation, in order to remove the cleaved tag, un-cleaved fusion protein and the PreScission protease.

The GST-free protein was collected in the flow through and loaded again on a 5ml GE Histrap column (GE) pre-equilibrated with 50mM bicine buffer pH8.0 containing 150mM NaCl, 1mM TCEP, 0.2% tween20. After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 150mM NaCl, 0.2% tween20, 1mM TCEP and 10mM imidazole. Elution was performed using a 50mM bicine buffer pH8.0 containing 150mM NaCl, 0.2% tween20, 1mM TCEP and 500mM imidazole.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing 150 mM NaCl, 1mM TCEP and 0.2% tween 20, the product was treated with α-chymotrypsin (from bovine pancreas - Sigma), followed by trypsin inhibitor treatment (from egg white- Sigma). The complete activation of Cdtb by chymotrypsin was monitored by SDS-PAGE.

The fully activated product was loaded on SEC chromatography (SUPERDEX ^{TM} 75) in 50mM bicine buffer pH8.0 containing 300mM NaCl, 1mM TCEP. Fractions containing CdtB antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

C38

The bacterial pellet was re-suspended in 50mM bicine buffer (pH 8.0) containing 150 mM NaCl , 5mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl)phosphine hydrochloride), 0.4%empigen and a mixture of protease inhibitor (Complete, Roche). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with 50mM bicine buffer (pH 8.0) containing 150mM NaCl , 1mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl)phosphine hydrochloride) and 0.15% empigen . After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 150mM NaCl, 20mM imidazole, 1mM TCEP and 0.2% tween 20. Elution was performed using a 50mM bicine buffer pH8.0 containing 150mM NaCl, 500mM imidazole, 1mM TCEP and 0.2% tween 20.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing 150mM NaCl, 1mM TCEP and 0.2% tween 20, the product was treated with α-chymotrypsin (from bovine pancreas - Sigma), followed by trypsin inhibitor treatment (from egg white- Sigma). The complete activation of Cdb by chymotrypsin was monitored by SDS-PAGE.

The fully activated product was loaded on SEC chromatography (SUPERDEX ^{TM} 75) in 50mM bicine buffer pH8.0, 300mM NaCl, 1mM TCEP. Fractions containing Cdtb protein were selected on the basis of purity by SDS-PAGE and loaded again on a 5ml GE Histrap column (GE) pre-equilibrated with 50mM bicine buffer (pH 8.0) containing 300mM NaCl , 1mM TCEP. After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 300mM NaCl, 10mM imidazole, 1mM TCEP. Elution was performed using a 50mM bicine buffer pH8.0 containing 300mM NaCl, 500mM imidazole, 1mM TCEP.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing 300mM NaCl, 1mM TCEP the protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

C40

The bacterial pellet was re-suspended in 20mM bicine buffer (pH 8.0) containing 500mM NaCl , 5mM$CaCl_2$ and a mixture of protease inhibitor (Complete, Roche). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a 1ml GE Histrap column (GE) pre-equilibrated with 20mM bicine buffer (pH 8.0) containing 500mM NaCl , 5mM$CaCl_2$. After loading on the column, the column was washed with a 20mM bicine buffer pH8.0, containing 500mM NaCl, 5mM $CaCl_2$ and 5mM imidazole. Elution was performed using a 20mM bicine buffer pH8.0 containing 150mM NaCl,5mM $CaCl_2$ and 250mM imidazole.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 20mM bicine buffer pH8.0 containing 150mM NaCl, 1mM TCEP the product was loaded on SEC chromatography (SUPERDEX ^{TM} 75) in the same buffer. Fractions containing Cdtb antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

C55

The bacterial pellet was re-suspended in 50mM bicine buffer (pH 8.0) containing 150 mM NaCl,
5mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl) phosphine hydrochloride), 0.4% empigen and a mixture of protease inhibitors (Complete, Roche). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with 50mM bicine buffer (pH 8.0) containing 150mM NaCl , 0.15% empigen, 1mM TCEP. After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 150 mM NaCl, 0.2% tween 20, 20mM imidazole and 1mM TCEP. Elution was performed using a 50mM bicine buffer pH8.0 containing 150mM NaCl, 0.2% tween 20, 500mM imidazole and 1mM TCEP.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing 300mM NaCl, 1mM TCEP the product was loaded on SEC chromatography (SUPERDEX ™ 75) in the same buffer. Fractions containing Cdtb antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

Expression of the recombinant proteins: CdtB receptor binding domain:

Expression plasmid and recombinant strain.

Genes encoding the truncated protein of CdtB only recptor binding domain (C52-C53) and a His tag in C-term was cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. The final constructs were generated by the transformation of *E. coli* B834 (DE3) modified strain with the recombinant expression vector according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

Host Strain

B834 is the parental strain for BL21. These protease-deficient hosts are methionine auxotrophs. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Modification: Including PGL gene to avoid phosphogluconoylation in the biotin locus (Strain is auxotroph for bioti n).

Genotype : B834 ::DE3 strain, F– *ompT hsdS*B(rB – mB–) *gal dcm met* (DE3)

Modification : Δ(bioA-bioD)::PGL

CdtB Rec. Bdng domain

C number	
C52	CdtB receptor binding domain long (aa. 620-876)
C53	CdtB receptor binding domain short (aa. 636-876)

Expression of the recombinant proteins:

A *E.coli* transformants were stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) to obtain $O.D._{600nm}$ between 0.1 - 0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

These overnight cultures were diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until $O.D._{620}$ reached 0.5/0.6.

At an O.D.at 600nm of around 0.6, the cultures were cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After the overnight inductions (around 16 hours), O.D. at 600nm were evaluated after induction and cultures were centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Purification

C52 and C53

The bacterial pellets were re-suspended in 50mM bicine buffer pH 8.0, containing 500mM NaCl and a mixture of protease inhibitors (Complete, Roche, without EDTA). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-proteins were purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with the same buffer used to bacterial re-suspension. After loading on the column, the column was washed with a 20mM bicine buffer

pH7.5 containing 500mM NaCl, 25mM imidazole. Elution was performed using a 50mM bicine buffer pH7.5, 500mM NaCl, and 250mM imidazole.

After desalting (BIORAD Bio-Gel P6 Desalting) and concentration (Amicon Ultra 10kDa) steps, the product was loaded on SEC chromatography (SUPERDEX ™ 75) in 20mM buffer pH7.5,
150mM NaCl.

Fractions containing Cdtb antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

Example 6 – Cloning, expression and purification of *C.difficile CdtA N-term and* CdtB receptor binding domain fusion proteins

Expression plasmid and recombinant strain.

Genes encoding the fusion protein of CdtA N-term (C49 or C50) with CdtB receptor binding domain protein long or short version (C61 or C62) and a His tag in C-term were cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. The final constructs were generated by the transformation of *E. coli* B834 (DE3) modified strain with the appropriate recombinant expression vector according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

Fusion CdtA N-term/ CdtB- receptor binding domain	
C number	
C61	CdtA N-term link (aa. 44-268)-CdtB RBD short (aa. 636-876)
C62	CdtA N-term (aa. 44-260)-CdtB RBD long (aa. 621-876)

Host Strain

Genotype : *E.coli* BL21::DE3 strain, F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (*DE3*).

B834 is the parental strain for BL21. These protease-deficient hosts are methionine auxotrophs. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Modification: Including PGL gene to avoid phosphogluconoylation in the biotin locus (Strain is auxotroph for biotin).

---

[016] Genotype : B834 ::DE3 strain, F– *ompT hsdS*B(rB – mB–) *gal dcm met* (DE3)
Modification : Δ(bioA-bioD)::PGL

Expression of the recombinant proteins:

E.coli transformants were stripped from each agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) to obtain $O.D._{600nm}$ between 0.1 -0.2. Cultures were incubated overnight at 37 °C, 250 RPM.

These overnight cultures were diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until $O.D._{620}$ reached 0.5/0.6.

At an O.D.at 600nm of around 0.6, the cultures were cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After the overnight induction (around 16 hours), O.D. at 600nm was evaluated after induction and culture was centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.

Purification

C61

The bacterial pellet was re-suspended in 50mM bicine buffer (pH 8.0) containing 300mM NaCl,5mM TCEP (Thermo Scientific Pierce, (2-carboxyethyl) phosphine hydrochloride), 0.4% empigen and a mixture of protease inhibitors (Complete, Roche). Bacteria were lysed using a French Press system 3 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 20 000g for 30 min at 4°C.

The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a 5ml GE Histrap column (GE) pre-equilibrated with 50mM bicine buffer (pH 8.0) containing 300mM NaCl , 0.15% empigen, 1mM TCEP. After loading on the column, the column was washed with a 50mM bicine buffer pH8.0, containing 300

mM NaCl, 0.2% tween 20, 25mM imidazole and 1mM TCEP. Elution was performed using a 50mM bicine buffer pH8.0 containing 150mM NaCl, 0.2% tween 20, 500mM imidazole and 1mM TCEP.

After desalting step (BIORAD Bio-Gel P6 Desalting) in 50mM bicine buffer pH8.0 containing 300mM NaCl, 1mM TCEP the product was loaded on SEC chromatography (SUPERDEX ™ 200) in the same buffer. Fractions containing the recombinant antigen were selected on the basis of purity by SDS-PAGE. Protein concentration was determined using Lowry RC/DC Protein Assay of BioRad. The purified bulk was sterile-filtered on 0.22 µm and stored at -80°C.

Example 7 – Cloning and expression of *C.difficile* CdtB mature co-expressed (C55) with prodomain of CdtB C58.

Expression plasmid and recombinant strain.

Genes encoding protein pro domain of CdtB without His tag was cloned into the pET21b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures. The final construct was generated by the transformation of *E. coli* B834 (DE3) modified strain with the recombinant expression vector of prodomain CdtB and CdtB mature protein C55- information about cloning of C55 see example 3 according to standard method with CaCl2-treated cells (Hanahan D. « Plasmid transformation by Simanis. » In Glover, D. M. (Ed), DNA cloning. IRL Press London. (1985): p. 109-135.).

Pro domain of CdtB alone

C number	
C58	Pro-domaine CdtB long (aa. 43-211)

Host Strain

B834 is the parental strain for BL21. These protease-deficient hosts are methionine auxotrophs. λ DE3 lysogens are designed for protein expression from pET vectors This strain is also deficient in the *lon* and *ompT* proteases.

Modification: Including PGL gene to avoid phosphogluconoylation in the biotin locus (Strain is auxotroph for biotin).

[016] Genotype : B834 ::DE3 strain, F– *ompT hsdS*B(rB – mB–) *gal dcm met* (DE3)
Modification : Δ(bioA-bioD)::PGL

Expression of the recombinant proteins:

A *E.coli* transformant was stripped from agar plate and used to inoculate 200 ml of LBT broth ± 1% (w/v) glucose + kanamycin (50 µg/ml) and ampicillin (100µg/ ml) to obtain $O.D._{600nm}$ between 0.1 -0.2. Culture was incubated overnight at 37 °C, 250 RPM.

This overnight culture was diluted to 1:20 in 500 ml of LBT medium containing kanamycin (50 µg/ml) and ampicillin (100 µg/ml) and grown at 37°C at a stirring speed of 250 rpm until $O.D._{620}$ reached 0.5/0.6.

At an O.D.at 600nm of around 0.6, the culture was cooled down before inducing the expression of the recombinant protein by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; EMD Chemicals Inc., catalogue number: 5815) and incubated overnight at 23 °C, 250 RPM.

After the overnight induction (around 16 hours), O.D. at 600nm was evaluated after induction and culture was centrifuged at 14 000 RPM for 15 minutes and pellets were frozen at -20°C separately.
Purification
same as C55 produced alone

Example 8: Molecular weight evaluation of CdtA, CdtB and CdtA-CdtB fusion constructions

Analytical ultracentrifugation was used to determine the homogeneity and size distribution in solution of the different species within a protein sample by measuring the rate at which molecules move in response to a centrifugal force. This is based on the calculation of the coefficients of sedimentation of the different species that are obtained by sedimentation velocity experiment, which depend on their molecular shape and mass.

1. Protein samples are spun in a Beckman-Coulter ProteomeLab XL-1 analytical ultracentrifuge at 8000RPM, 25000RPM or 42000RPM depending of the target protein size, after the AN-60Ti rotor had been equilibrated to 15°C.

2. For data collection, scans were recorded at 280nm every 5 minutes.
3. Data analysis was performed using the program SEDFIT for determination of the C(S) distribution. Determination of the partial specific volume of the proteins was performed with the SEDNTERP software from their amino acid sequence. Sednterp was also used to determine the viscosity and the density of the buffer.
4. Determination of the molecular weight of the different species has been determined from the C(S) distribution plot (concentration vs sedimentation coefficient), considering that it's a better representation of the raw data than the C(M) distribution (concentration vs molecular weight) to characterize the size distribution of a mixture.

Figures 1a to 1h describe the size distribution of the different CdtA. CdtB and CdtA-CdtB fusion constructions as determined by sedimentation velocity analytical ultracentrifugation.

The calculated molecular weight of the main species for C67 and C69 mutated full length CdtA protein may correspond with a monomer, while the C50 truncated CdtA N-terminal construction is present in solution as a mix of monomer and dimer (figures 1a, 1b and 1c).

Both C61 and C62 fusions of CdtA-CdtB are mainly dimeric, with a minor proportion of monomer (figures 1d and 1e).

Constructions of the CdtB receptor binding domain C52 and C52 are mainly dimeric with presence of small amount of monomer (figures 1f and 1g).

Full length CdtB without prodomain C55 is highly aggregated after purification, presenting a heterogeneous size distribution by AUC (figure 1h).

Example 9: SDS PAGE profile of CdtA, CdtB and CdtA-CdtB fusion constructions after purification

Purified proteins from each construction were separated on denaturing and reducing SDS PAGE in order to assess the sequence integrity.
Figure 2a shows that CdtA-CdtB fusion constructions C61 and C62 are present in majority at the expected molecular weight. Same observations are made for CdtA constructions on figure 2b.

It is shown on figure 2c that chymotrypsin activation of C37 CdtB (aa. 43-876) construction results in the truncation of the prodomain for the obtention of a protein (lane 2) at a

molecular weight comparable to the mature CdtB represented at the lane 3 by C55 (aa. 212-876). SDS PAGE profile of C55 contains significant amount of secondary products that couln't be separated from the complete protein, which is coherent with the highly aggregated profile observed by AUC on figure 2h.

CdtB expressed with prodomain C38 (aa. 43-876) was purified as a an heterogeneous preparation composed at the majority of a doublet of the expected molecular weight containing a significant amount of secondary products.

Example 9: Immunisation of mice with C.difficile CdtA and CdtB sub-units proteins in a AS01B formulation.

Mice immunisation

Groups of 25 female Balb/C mice were immunized IM at days 0, 14 and 28 with 5 µg of full CdtA and CdtB binary toxin purified sub-units.These antigens were injected in an AS01B formulation.

Anti-CdtA and anti-CdtB ELISA titers were determined in individual sera collected at day 42 (Post III 14). Results are shown in Figures 3-4.

A binary toxin cytotoxicity inhibition assay was also performed on pooled Post III sera (day42). Results are shown in Figures 5-6

Anti-CdtA and anti-CdtB ELISA response: Protocol

Full CdtA (C34) or full CdtB (C37)sub-units were coated at 1µg/ml (for CdtA) or 2µg/ml (for CdtB) in phosphate buffered saline (PBS) on high-binding microtitre plates (Nunc MAXISORP™), overnight at 4° C. The plates were blocked with PBS-BSA 1% for 30 min at RT with agitation. The mice anti-sera are prediluted 1/500 in PBS-BSA0.2%-TWEEN™ 0.05%. and then, further twofold dilutions were made in microplates and incubated at RT for 30 min. After washing, bound mouse antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated Anti-Mouse (ref: 110-035-003) diluted 1:5000 in PBS-BSA0.2%-tween 0.05%. The detection antibodies were incubated for 30 min. at room temperature (RT) with agitation. The color was developed using 4 mg O-phenylenediamine (OPD) + 5 µl H_2O_2 per 10 ml pH 4.5 0.1M citrate buffer for 15 minutes in the dark at room temperature. The reaction was stopped with 50 µl HCl, and the optical density (OD) was read at 490 nm relative to 620 nm.

The level of anti-CdtA or anti-CdtB antibodies are expressed in mid-point titers. A GMT was calculated for the 25 samples in each treatment group.

Binary toxin cytotoxicity inhibition assay

Human colonic eptithelial cells (HT29 or HCT-116 cells) were cultured at 37°C with 5%CO_2 in DMEM +10% fetal bovine serum +1% glutamine +1% antibiotics (penicillin-streptomycin-amphotericin) and were seeded in 96-well black tissue culture plates (Greiner Bio-one, Ref : 655090) at a density of 4.10^4 cells/well for HT29 and 1.10^4cells/well for HCT116.
After 24h, the cell media was removed from the wells.

The mice anti-sera were prediluted 1:50 in cell media and then, further three-fold dilutions were made in microplate (NUNC, Ref : 163320). 50µl of serial dilutions of mice pooled antisera were added to the black plates. 50µl of a mix of CdtA (25ng/ml) and chemotrypsin-activated CdtB (75 ng/ml) were then added and the black plates incubated at 37°C with 5% CO_2 for 6 days.

After 6 days, the mix of antisera and toxin were removed from the wells and 100µl of Hoescht stain (BD Pharmingen, Ref : 561908) diluted 1:500 in phosphate buffer saline (PBS) was added in each well for 2 hours in the dark at room temperature.

After coloration, the Hoescht stain was removed from the wells and the cells fluorescence cells was measured using an Axiovision microscope.

The surface covered by fluorescent staining was determined in each well and cytotoxicity inhibition titers were defined as the reciprocal dilution inducing a 50% inhibition of the fluorescent signal.

Example 10: Immunisation of mice with C.difficile CdtB chemotrypsin-activated or not, mixed with F2 or not, formulated in AS01B.

Mice immunisation

Groups of 25 female Balb/C mice were immunized IM at days 0, 14 and 28 with 5µg of CdtB binary toxin purified sub-unit chemotrypsin-activated or not, mixed with 5µg of F2 or not. These antigens were injected in an AS01B formulation.

Anti-CdtB, anti-ToxA and anti-ToxB ELISA titers were determined in individual sera collected at day 42 (Post III 14). Results are shown in Figures 7-9.

A binary toxin, ToxA and ToxB cytotoxicity inhibition assay was also performed on pooled Post III sera (day42). Results are shown in Figures 10-12.

anti-CdtB, anti-ToxA and anti-ToxB ELISA response: Protocol

Full CdtB (C37)sub-unit, F2 Cter ToxA and F2 Cter ToxB were coated at 0.5µg/ml (for CdtB), 2µg/ml (for ToxA F2 Cter) and 1µg/ml (for ToxB F2 Cter) in phosphate buffered saline (PBS) on high-binding microtitre plates (Nunc MAXISORP™), overnight at 4° C. The plates were blocked with PBS-BSA 1% for 30 min at RT with agitation. The mice anti-sera are prediluted 1/500 in PBS-BSA0.2%-TWEEN™ 0.05%. and then, further twofold dilutions were made in microplates and incubated at RT for 30 min. After washing, bound mouse antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated Anti-Mouse (ref: 110-035-003) diluted 1:5000 in PBS-BSA0.2%-tween 0.05%. The detection antibodies were incubated for 30 min. at room temperature (RT) with agitation. The color was developed using 4 mg O-phenylenediamine (OPD) + 5 µl H_2O_2 per 10 ml pH 4.5 0.1M citrate buffer for 15 minutes in the dark at room temperature. The reaction was stopped with 50 µl HCl, and the optical density (OD) was read at 490 nm relative to 620 nm.

The level of anti-CdtB antibodies are expressed in mid-point titers.

The level of anti- F2Cter ToxA and F2Cter ToxB antibodies present in each individual sera is determined by comparison to a reference serum added on each plate and expressed in µg/ml.

A GMT was calculated for the 25 samples in each treatment group.

Binary toxin, ToxA and ToxB cytotoxicity inhibition assay

Human colonic eptithelial cells (HT29 or HCT-116 cells) were cultured at 37°C with 5%CO_2 in DMEM +10% fetal bovine serum +1% glutamine +1% antibiotics (penicillin-streptomycin-amphotericin) and were seeded in 96-well black tissue culture plates (Greiner Bio-one, Ref : 655090) at a density of 4.10^4 cells/well for HT29 and 1.10^4cells/well for HCT116.
After 24h, the cell media was removed from the wells.

The mice anti-sera were prediluted 1:5 for g1 (CdtB non-activated) and g2 (CdtB activated) and 1:20 for g3 (CdtB non-activated + F2) and g4 (Cdtb activated + F2) in cell media, for ToxA inhibition cytotoxicity assay, 1:10 for ToxB inhibition cytotoxicity assay and 1:50 for binary toxin inhibition assay. Then, further three-fold dilutions were made in microplate (NUNC, Ref : 163320). 50µl of serial dilutions of mice pooled antisera were added to the

black plates. 50µl of ToxA (0.01µg/ml) on HT29, ToxB (0.022µg/ml) on HCT116 and a mix of CdtA (25ng/ml) and chemotrypsin-activated CdtB (75 ng/ml) on HT29 and HCT116 were then added in the black plates and incubated at 37°C with 5% CO_2 for 6 days.

After 6 days, the mix of antisera and toxin were removed from the wells and 100µl of Hoescht stain (BD Pharmingen, Ref : 561908) diluted 1:500 in phosphate buffer saline (PBS) was added in each well for 2 hours in the dark at room temperature.

After coloration, the Hoescht stain was removed from the wells and the cells fluorescence cells was measured using an Axiovision microscope.

The surface covered by fluorescent staining was determined in each well and cytotoxicity inhibition titers were defined as the reciprocal dilution inducing a 50% inhibition of the fluorescent signal.

Example 11: Immunisation of mice with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 6µg/dose in an AS01B formulation

Mice immunisation

Groups of 20 female Balb/C mice were immunized IM at days 0, 14 and 28 with 6µg of CdtA-CdtB fusion (C61 and C62), or 3µg of CdtA (C34, C50 or C67) and/or 3µg of CdtB (C37, C52, C55 or C55/C58) mixed or not with 6µg of F2. These antigens were injected in an AS01B formulation.

Anti-CdtA, anti-CdtB, anti-ToxA and anti-ToxB ELISA titers were determined in individual sera collected at day 42 (Post III 14). Results are shown in Figures 13-16.
A binary toxin, ToxA and ToxB cytotoxicity inhibition assay was also performed on pooled Post III sera (day42). Results are shown in Figures 17-20.

anti-CdtA, anti-CdtB, anti-ToxA F2Cter and anti-ToxB F2 Cter ELISA response: Protocol

CdtA mut E428Q (C44), Full CdtB (C37)sub-units, F2 Cter ToxA and F2 Cter ToxB were coated at 1µg/ml (for CdtA), 0.5µg/ml (for CdtB), 2µg/ml (for ToxA F2 Cter) and 1µg/ml (for ToxB F2 Cter) in phosphate buffered saline (PBS) on high-binding microtitre plates (Nunc MAXISORP™), overnight at 4° C. The plates were blocked with PBS-BSA 1% for 30 min at RT with agitation. The mice anti-sera were prediluted 1:100 (for CdtA, CdtB, ToxB) or 1:200 (for ToxA) for Post II and 1:500 (for CdtA and ToxA), 1:500 or 1:2000 (for CdtB) and 1:250 (for ToxB) for Post III in PBS-BSA0.2%-TWEEN™ 0.05%. Then, further twofold dilutions

were made in microplates and incubated at RT for 30 min. After washing, bound mouse antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated Anti-Mouse (ref: 110-035-003) diluted 1:5000 in PBS-BSA0.2%-tween 0.05%. The detection antibodies were incubated for 30 min. at room temperature (RT) with agitation. The color was developed using 4 mg O-phenylenediamine (OPD) + 5 µl H_2O_2 per 10 ml pH 4.5 0.1M citrate buffer for 15 minutes in the dark at room temperature. The reaction was stopped with 50 µl HCl, and the optical density (OD) was read at 490 nm relative to 620 nm.

The level of anti-CdtA, anti-CdtB, anti- F2Cter ToxA and F2Cter ToxB antibodies present in each individual sera is determined by comparison to a reference serum added on each plate and expressed in µg/ml. A GMT was calculated for the 20 samples in each treatment group.

Binary toxin, ToxA and ToxB cytotoxicity inhibition assay

Human colonic eptithelial cells (HT29 or HCT-116 cells) were cultured at 37°C with 5%CO_2 in DMEM +10% fetal bovine serum +1% glutamine +1% antibiotics (penicillin-streptomycin-amphotericin) and were seeded in 96-well black tissue culture plates (Greiner Bio-one, Ref : 655090) at a density of 4.10^4 cells/well for HT29 and 1.10^4cells/well for HCT116.
After 24h, the cell media was removed from the wells.

The mice anti-sera were prediluted 1:50 in cell media, for ToxA inhibition cytotoxicity assay, 1:10 for ToxB inhibition cytotoxicity assay , 1:50 for binary toxin inhibition assay on HT29 and 1:30 (for Post II) and 1:30 or 1:100 (for Post III) for binary toxin inhibition assay on HCT116. Then, further three-fold dilutions were made in microplate (NUNC, Ref : 163320). 50µl of serial dilutions of mice pooled antisera were added to the black plates. 50µl of ToxA (0.025µg/ml) on HT29, ToxB (0.6µg/ml) on HCT116 and a mix of CdtA (25ng/ml) and chemotrypsin-activated CdtB (75 ng/ml) on HT29 and HCT116 were then added in the black plates and incubated at 37°C with 5% CO_2 for 6 days.

After 6 days, the mix of antisera and toxin were removed from the wells and 100µl of Hoescht stain (BD Pharmingen, Ref : 561908) diluted 1:500 in phosphate buffer saline (PBS) was added in each well for 2 hours in the dark at room temperature.

After coloration, the Hoescht stain was removed from the wells and the cells fluorescence cells was measured using an Axiovision microscope.

The surface covered by fluorescent staining was determined in each well and cytotoxicity inhibition titers were defined as the reciprocal dilution inducing a 50% inhibition of the fluorescent signal.

Example 12: Immunisation of mice with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 2µg/dose in an AS01B formulation

Mice immunisation

Groups of 20 female Balb/C mice were immunized IM at days 0, 14 and 28 with 2µg of CdtA-CdtB fusion (C61 and C62), or 1µg of CdtA (C34, C50 or C67) and/or 1µg of CdtB (C37, C52, C55 or C55/C58) mixed or not with 2µg of F2. These antigens were injected in an AS01B formulation.

Anti-CdtA, anti-CdtB, anti-ToxA and anti-ToxB ELISA titers were determined in individual sera collected at day 42 (Post III 14). Results are shown in Figures 21-24.

A binary toxin, ToxA and ToxB cytotoxicity inhibition assay was also performed on pooled Post III sera (day42). Results are shown in Figures 25-28.

anti-CdtA, anti-CdtB, anti-ToxA and anti-ToxB ELISA response: Protocol

CdtA mut E428Q (C44), Full CdtB (C37)sub-units, F2 Cter ToxA and F2 Cter ToxB were coated at 1µg/ml (for CdtA), 0.5µg/ml (for CdtB), 2µg/ml (for ToxA F2 Cter) and 1µg/ml (for ToxB F2 Cter) in phosphate buffered saline (PBS) on high-binding microtitre plates (Nunc MAXISORP™), overnight at 4° C. The plates were blocked with PBS-BSA 1% for 30 min at RT with agitation. The mice anti-sera were prediluted 1:100 (for CdtB, ToxA,ToxB) and 1:100 or 1:250 (for CdtA) for Post II and 1:500 for Post III in PBS-BSA0.2%-TWEEN™ 0.05%. Then, further twofold dilutions were made in microplates and incubated at RT for 30 min. After washing, bound mouse antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated Anti-Mouse (ref: 110-035-003) diluted 1:5000 in PBS-BSA0.2%-tween 0.05%. The detection antibodies were incubated for 30 min. at room temperature (RT) with agitation. The color was developed.

Example 13: Immunisation of mice with different binary toxin vaccine candidates (CdtA/CdtB) combined with F2 at 10µg/dose in a non-adjuvanted formulation

Mice immunisation

Groups of 20 female Balb/C mice were immunized IM at days 0, 14 and 28 with 10µg of CdtA-CdtB fusion (C61 and C62), or 5µg of CdtA (C34, C50 or C67) and/or 5µg of CdtB (C37, C52, C55 or C55/C58) mixed or not with 10µg of F2. These antigens were injected in a non-adjuvanted formulation.

Anti-CdtA, anti-CdtB, anti-ToxA and anti-ToxB ELISA titers were determined in individual sera collected at day 42 (Post III 14). Results are shown in Figures 29-32.

A binary toxin, ToxA and ToxB cytotoxicity inhibition assay was also performed on pooled Post III sera (day42). Results are shown in Figures 33-36.

anti-CdtA, anti-CdtB, anti-ToxA and anti-ToxB ELISA response: Protocol

CdtA mut E428Q (C44), Full CdtB (C37)sub-units, F2 Cter ToxA and F2 Cter ToxB were coated at 1µg/ml (for CdtA), 0.5µg/ml (for CdtB), 2µg/ml (for F2 Cter ToxA) and 1µg/ml (for F2 Cter ToxB) in phosphate buffered saline (PBS) on high-binding microtitre plates (Nunc MAXISORP^{TM}), overnight at 4° C. The plates were blocked with PBS-BSA 1% for 30 min at RT with agitation. The mice anti-sera were prediluted 1:100 (for CdtA, CdtB, ToxA,ToxB) for Post II and 1:100 (for CdtA, ToxA, ToxB), 1:100 or 1:200 (for CdtB) for Post III in PBS-BSA0.2%-TWEEN^{TM} 0.05%. Then, further twofold dilutions were made in microplates and incubated at RT for 30 min. After washing, bound mouse antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated Anti-Mouse (ref: 110-035-003) diluted 1:5000 in PBS-BSA0.2%-tween 0.05%. The detection antibodies were incubated for 30 min. at room temperature (RT) with agitation. The color was developed using 4 mg O-phenylenediamine (OPD) + 5 µl H_2O_2 per 10 ml pH 4.5 0.1M citrate buffer for 15 minutes in the dark at room temperature. The reaction was stopped with 50 µl HCl, and the optical density (OD) was read at 490 nm relative to 620 nm.

A GMT was calculated for the 20 samples in each treatment group.

Binary toxin, ToxA and ToxB cytotoxicity inhibition assay: protocol

Human colonic eptithelial cells (HT29 or HCT-116 cells) were cultured at 37°C with 5%CO_2 in DMEM +10% fetal bovine serum +1% glutamine +1% antibiotics (penicillin-streptomycin-amphotericin) and were seeded in 96-well black tissue culture plates (Greiner Bio-one, Ref : 655090) at a density of 4.10^4 cells/well for HT29 and 1.10^4cells/well for HCT116.

After 24h, the cell media was removed from the wells.

The mice anti-sera were prediluted 1:50 in cell media, for ToxA inhibition cytotoxicity assay, 1:10 for ToxB inhibition cytotoxicity assay , 1:50 for binary toxin inhibition assay on HT29 and 1:30 (for Post II) and 1:30 or 1:100 (for Post III) for binary toxin inhibition assay on HCT116. Then, further three-fold dilutions were made in microplate (NUNC, Ref : 163320).

50µl of serial dilutions of mice pooled antisera were added to the black plates. 50µl of ToxA (0.025µg/ml) on HT29, ToxB (0.6µg/ml) on HCT116 and a mix of CdtA (25ng/ml) and chemotrypsin-activated CdtB (75 ng/ml) on HT29 and HCT116 were then added in the black plates and incubated at 37°C with 5% CO_2 for 6 days.

After 6 days, the mix of antisera and toxin were removed from the wells and 100µl of Hoescht stain (BD Pharmingen, Ref : 561908) diluted 1:500 in phosphate buffer saline (PBS) was added in each well for 2 hours in the dark at room temperature.

After coloration, the Hoescht stain was removed from the wells and the cells fluorescence cells was measured using an Axiovision microscope.

The surface covered by fluorescent staining was determined in each well and cytotoxicity inhibition titers were defined as the reciprocal dilution inducing a 50% inhibition of the fluorescent signal.

Example 14: Cloning and expression of *C.difficile F2 and* CdtB receptor binding domain fusion proteins

Expression plasmid and recombinant strain.

Genes encoding the fusion protein of F2 protein with CdtB receptor binding domain protein long or short version (C64 and C65) and a His tag in C-term were cloned into the pET24b(+) expression vector (Novagen) using the NdeI/XhoI restriction sites using standard procedures.

Fusion F2 / CdtB - receptor binding domain

C number	
C64	CdtA N-term (aa. 44-260)-CdtB RBD long (aa. 621-876)
C65	F2-CdtB RBD short (aa. 636-876)

Sequence Summary (Table A)

Description	Construct reference	Amino acid sequence	Polynucleotide sequence

CDTa full length (strain R20291)	N/A	SEQ.I.D.NO:1	SEQ.I.D.NO:2
CDTb full length (strain R20291)	N/A	SEQ.I.D.NO:3	SEQ.I.D.NO:4
CDTa without signal peptide	C34	SEQ.I.D.NO:5	SEQ.I.D.NO:6
CDTb' (minus signal peptide) ligated to Glutathione-S-transferase protein.(GST underlined)	C37	SEQ.I.D.NO:7	SEQ.I.D.NO:8
CDTb'' (minus pro-domain and signal peptide)	C40	SEQ.I.D.NO:9	N/A
CDTa mutation E428Q	C44	SEQ.I.D.NO:10	SEQ.I.D.NO:11
CDTa mutation E430Q	C54	SEQ.I.D.NO:12	N/A
CDTa N terminal domain (residue 44 to residue 240)	Gülke et al 2001	SEQ.I.D.NO:13	N/A
CDTa without signal peptide, with a linker between the N-term domain and the C-term domain (containing the enzymatic activity). This construct covers the fragment from amino acid 44 to aa 268.	C49	SEQ.I.D.NO:14	N/A
CDTa without signal peptide or linker. This construct covers the fragment from aa 44 to aa 260.	C50	SEQ.I.D.NO:15	
CDTb minus signal peptide (CDTb')	C38	SEQ ID NO:16	SEQ ID NO:17
Fusion 1	F1	SEQ ID NO:18	
Fusion 2	F2	SEQ ID NO:19	
Fusion 3	F3	SEQ ID NO:20	
Fusion 4	F4	SEQ ID NO:21	
Fusion 5	F5	SEQ ID NO:22	
Fusion F54 Gly	N/A	SEQ ID NO:24	SEQ ID NO:23
Fusion F54 New	N/A	SEQ ID NO:26	SEQ ID NO:25
Fusion F5 ToxB	N/A	SEQ ID NO:28	SEQ ID NO:27
Fusion F52 New	N/A	SEQ ID NO:30	SEQ ID NO:29
Toxin A	N/A	SEQ ID NO:31	
Toxin B	N/A	SEQ ID NO:32	

CDTb'' (minus pro-domain and signal peptide) ligated to Glutathione-S-transferase protein.	C39	SEQ ID NO:33	N/A
CdtB receptor binding domain with linker in N-term of sequence, from aa 620-876	C52	SEQ ID NO:34	SEQ ID NO:35
CdtB receptor binding domain without linker in N-term of sequence, from aa 636-876	C53	SEQ ID NO:36	SEQ ID NO:37
CDTb with prodomain removed (CDTb'', aa212-876)	C55	SEQ ID NO:51	
CDTb prodomain sequence (long, aa43-211)	C58	SEQ ID NO:38	N/A
CDTb prodomain sequence (short, aa43-186)	C59	SEQ ID NO:39	N/A
Fusion CDTa N-term with linker (aa44-268) to CDTb receptor binding domain with linker in N term of sequence (aa621-876)	C60	SEQ ID NO:40	N/A
Fusion CDTa N-term with linker (aa44-268) to CDTb receptor binding domain without linker in N term of sequence (aa636-876)	C61	SEQ ID NO:41	N/A
Fusion CDTa N-term without linker (aa44-260) to CDTb receptor binding domain with linker in N term of sequence (aa621-876)	C62	SEQ ID NO:42	N/A
Fusion CDTa N-term without linker (aa44-260) to CDTb receptor binding domain without linker in N term of sequence (aa636-876)	C63	SEQ ID NO:43	N/A
Fusion F2- CDTb receptor binding domain with linker in N term of sequence (aa621-876)	C64	SEQ ID NO:44	N/A
Fusion of F2 to CDTb receptor binding domain without linker in N	C65	SEQ ID NO:45	N/A

term of sequence (aa636-876) with 2 heterogeneous Gly residues between F2 and CTDb sequences			
CDTa without signal peptide, with two mutations (E428Q, E430Q, aa 44-463).	C67	SEQ ID NO:46	SEQ ID NO:47
CDTa without signal peptide, with seven mutations (R345A, Q350A, N385A, R402A, S388F, E428Q, E430Q, aa 44-463).	C69	SEQ ID NO:48	SEQ ID NO:49
CDTb without signal sequence and prodomain (mature fragment based on MS data) with Ca2+ binding motif mutation (aa212-876, mut Asp-9-11-13 Ala)	C97	SEQ ID NO:50	N/A
CDTa without signal peptide, with five mutations (R345A, Q350A, N385A, R402A, S388F, aa 44-463).	C107	SEQ ID NO:52	SEQ ID NO:53
CDTa without signal peptide, with six mutations (R345A, Q350A, N385A, R402A, S388F, E430Q, aa 44-463).	C108	SEQ ID NO:54	SEQ ID NO:55
CdtA without signal peptide, with six mutations (R345A-Q350A-N385A-R402A-S388F-E428Q, aa 44-463).	C110	SEQ ID NO:56	N/A

SEQUENCE LISTING

SEQ ID 1 – CDTa full length polypeptide sequence

MKKFRKHKRISNCISILLILYLTLGGLLPNNIYAQDLQSYSEKVCNTTYKAPIERPEDFLKDKE
KAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQTRNYFYDYQIEANSREKEYKEL
RNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLEKFNEFKETIQNKLFKQDGFKDIS
LYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIEQGYSIKIDKIVRIVIDGKHYIKAE
ASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNELADVNDYMRGGYTAINNYLISNGPV
NNPNPELDSKITNIENALKREPIPTNLTVYRRSGPQEFGLTLTSPEYDFNKLENIDAFKSKWE
GQALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGAYLSAIPGYAGEYEVLLNHGSKFKI
NKIDSYKDGTITKLIVDATLIP

SEQ ID 2 – CDTa full length polynucleotide sequence

ATGAAAAAATTTAGGAAACATAAAAGGATTAGTAATTGTATATCTATATTGTTGATATTAT
ATCTAACTTTAGGTGGTTTGTTACCTAATAACATTTATGCACAAGACTTACAAAGCTATA
GTGAAAAAGTTTGCAATACTACTTACAAGGCTCCTATAGAAAGACCAGAAGATTTTCTTA
AAGATAAAGAAAAGGCTAAAGAATGGGAAAGAAAAGAAGCAGAAAGAATAGAGCAAAAA
CTTGAAAGATCTGAAAAAGAAGCATTAGAATCATATAAAAAAGATTCTGTAGAAATAAGT
AAATATTCTCAGACAAGAAATTATTTTTATGATTATCAAATAGAAGCAAATTCTCGAGAAA
AAGAATATAAAGAACTTCGAAATGCTATATCAAAAAATAAAATAGATAAACCTATGTATGT
CTATTATTTTGAATCTCCAGAAAAATTTGCATTTAATAAAGTAATAAGAACAGAAAATCAA
AACGAAATTTCATTAGAAAAATTTAATGAGTTTAAAGAAACTATACAAAACAAATTATTTA
AGCAAGATGGATTTAAAGATATTTCTTTATATGAACCTGGAAAAGGTGATGAAAAACCTA
CACCATTACTTATGCACTTAAAATTACCTAGAAATACTGGTATGTTACCATATACAAATAC
TAACAATGTAAGTACATTAATAGAGCAAGGATATAGTATAAAAATAGATAAAATTGTTCGT
ATAGTTATAGATGGGAAGCACTATATTAAAGCAGAAGCATCTGTTGTAAGTAGTCTTGAT
TTTAAAGATGATGTAAGTAAGGGGGATTCTTGGGGTAAAGCAAATTATAATGATTGGAG
TAATAAATTAACACCTAATGAACTTGCTGATGTAAATGATTATATGCGTGGAGGATATAC
TGCAATTAATAATTATTTAATATCAAATGGTCCAGTAAATAATCCTAACCCAGAATTAGAT
TCTAAAATCACAAACATTGAAAATGCATTAAAACGTGAACCTATTCCAACTAATTTAACTG
TATATAGAAGATCTGGTCCTCAAGAATTTGGTTTAACTCTTACTTCCCCTGAATATGATTT
TAACAAACTAGAAAATATAGATGCTTTTAAATCAAAATGGGAAGGACAAGCACTGTCTTA
TCCAAACTTTATTAGTACTAGTATTGGTAGTGTGAATATGAGTGCATTTGCTAAAAGAAA
AATAGTACTACGTATAACTATACCTAAAGGTTCTCCTGGAGCTTATCTATCAGCTATTCC
AGGTTATGCAGGTGAATATGAAGTGCTTTTAAATCATGGAAGCAAATTTAAAATCAATAA

AATTGATTCTTACAAAGATGGTACTATAACAAAATTAATTGTTGATGCAACATTGATACCT
TAA

SEQ ID 3 – CDTb full length polypeptide sequence

MKIQMRNKKVLSFLTLTAIVSQALVYPVYAQTSTSNHSNKKKEIVNEDILPNNGLMGYYFTDE
HFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKSDVKSIRWTGRIIPSKDGEYTLSTDRDDVLM
QVNTESTISNTLKVNMKKGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPEENLFLR
DYSNIEKDDPFIPNNNFFDPKLMSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVKWEDSFA
EQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKAIKTEARDPLVAAYPIVGVGMEKLIISTN
EHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTANVTTNYSHTTDNSTAVQDSNG
ESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLS
PGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSSGQI
VTEGNSWSDYISQIDSISASIILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKK
DGLLYFNDIPIDESCVELIFDDNTANKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTNFDDY
NNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKI
KAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEV
KIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFK
DIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVD

SEQ ID 4 – CDTb full length polynucleotide sequence

ATGAAAATACAAATGAGGAATAAAAAGGTATTAAGTTTTTTAACACTTACAGCTATAGTTA
GTCAAGCACTAGTATATCCTGTATATGCTCAAACTAGTACAAGTAATCATTCTAATAAGA
AAAAAGAAATTGTAAATGAAGATATACTCCCAAACAATGGATTAATGGGATATTATTTCA
CAGATGAGCACTTTAAAGATTTAAAATTAATGGCACCCATAAAAGATGGTAATTTAAAAT
TTGAAGAAAAGAAAGTAGATAAACTTCTGGATAAAGACAAATCAGATGTAAAATCTATAC
GATGGACAGGAAGAATAATTCCTTCTAAGGATGGTGAATATACATTATCAACTGATAGA
GATGATGTCTTAATGCAAGTAAATACTGAGAGTACTATATCAAATACACTTAAAGTTAATA
TGAAAAAGGGTAAAGAATATAAAGTTAGAATAGAGCTACAAGATAAAAATTTAGGTTCAA
TAGATAATTTATCATCACCTAATCTTTATTGGGAATTAGATGGTATGAAGAAAATTATACC
AGAAGAAAATTTATTCTTAAGAGATTATTCTAATATAGAAAAAGATGATCCATTTATCCCA
AATAACAATTTCTTTGACCCAAAGTTGATGTCTGATTGGGAAGACGAAGATTTGGATACA
GATAATGATAATATACCAGATTCATATGAACGAAATGGATATACTATTAAGGACTTAATT
GCAGTTAAGTGGGAAGATAGTTTTGCAGAACAAGGCTATAAGAAATATGTATCAAATTAT
TTAGAGTCAAATACTGCTGGAGATCCATATACAGATTATGAAAAAGCTTCAGGTTCTTTT
GACAAGGCTATAAAGACTGAAGCAAGAGATCCGTTAGTTGCAGCATATCCAATTGTTGG
AGTAGGTATGGAAAAATTAATTATATCTACAAATGAACATGCCTCTACTGATCAAGGTAA
AACTGTTTCCAGAGCTACTACTAACAGTAAAACTGAATCTAATACAGCTGGTGTGTCTGT

TAATGTAGGATATCAAAATGGATTCACAGCTAATGTAACTACAAATTATTCCCATACAAC
AGATAATTCAACTGCTGTTCAAGATAGTAATGGAGAATCATGGAATACTGGATTAAGTAT
AAACAAAGGAGAATCTGCATATATAAATGCAAATGTTAGATATTACAACACAGGTACTGC
ACCTATGTACAAAGTGACACCAACAACAAATTTAGTGTTAGATGGAGATACATTATCAAC
TATCAAAGCACAAGAAAATCAAATTGGCAATAATCTATCTCCTGGAGATACTTATCCCAA
AAAAGGGCTTTCACCTCTAGCTCTTAACACAATGGATCAATTTAGCTCTAGACTGATTCC
TATAAATTATGATCAATTAAAAAAATTAGATGCTGGAAAGCAAATTAAATTAGAAACAACA
CAAGTAAGTGGAAATTTTGGTACAAAAAATAGTTCTGGACAAATAGTAACAGAAGGAAAT
AGTTGGTCAGACTATATAAGTCAAATTGACAGTATTTCTGCATCTATTATATTAGATACAG
AGAATGAATCTTACGAAAGAAGAGTTACTGCTAAAAATTTACAGGATCCAGAAGATAAAA
CACCTGAACTTACAATTGGAGAAGCAATTGAAAAAGCTTTTGGCGCTACTAAAAAAGAT
GGTTTGTTATATTTTAATGATATACCAATAGATGAAAGTTGTGTTGAACTCATATTTGATG
ATAATACAGCCAATAAGATTAAAGATAGTTTAAAAACTTTGTCTGATAAAAAGATATATAA
TGTTAAACTTGAAAGAGGAATGAATATACTTATAAAAACACCAACTTACTTTACTAATTTT
GATGATTATAATAATTACCCTAGTACATGGAGTAATGTCAATACTACGAATCAAGATGGT
TTACAAGGCTCAGCAAATAAATTAAATGGTGAGACGAAGATTAAAATCCCTATGTCTGAG
CTAAAACCTTATAAACGTTATGTTTTTAGTGGATATTCAAAGGATCCTTTAACATCTAATT
CAATAATTGTAAAGATAAAAGCAAAAGAAGAGAAAACGGATTATTTGGTACCAGAACAA
GGATATACAAAATTTAGTTATGAATTTGAAACTACTGAAAAAGATTCTTCTAATATAGAGA
TAACATTAATTGGTAGTGGTACAACATACTTAGATAACTTATCTATTACAGAGCTAAATAG
TACTCCTGAAATACTTGATGAACCAGAAGTTAAAATTCCAACTGACCAAGAAATAATGGA
TGCACATAAAATATATTTTGCAGATTTAAATTTTAATCCAAGTACAGGAAATACTTATATA
AATGGTATGTATTTTGCACCAACACAAACTAATAAAGAAGCTCTCGATTATATCCAAAAA
TATAGAGTTGAAGCTACTTTACAATATTCTGGATTTAAAGATATTGGAACTAAAGATAAA
GAAATGCGTAATTATTTAGGAGATCCAAATCAGCCTAAAACTAATTATGTTAATCTTAGG
AGTTATTTTACAGGTGGAGAAAATATTATGACATACAAGAAATTAAGAATATATGCAATTA
CTCCAGACGATAGAGAGTTATTAGTTCTTAGTGTTGATTAG

SEQ ID 5 – CDTa C34 construct polypeptide sequence

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNELAD
VNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLTVYRRSGPQEFGLT
LTSPEYDFNKLENIDAFKSKWEGQALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGAYL
SAIPGYAGEYEVLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

SEQ ID 6 – CDTb C34 construct polynucleotide sequence

ATGGTTTGCAATACCACCTATAAAGCACCGATTGAACGTCCGGAAGATTTTCTGAAAGA
TAAAGAAAAAGCCAAAGAATGGGAACGCAAAGAAGCAGAACGTATTGAACAGAAACTG
GAACGTAGCGAAAAAGAAGCACTGGAAAGCTACAAAAAAGATAGCGTGGAAATTTCAAA
ATATAGCCAGACCCGCAATTATTTCTATGATTATCAGATTGAAGCCAATAGCCGTGAAAA
AGAATATAAAGAACTGCGCAATGCCATTAGCAAAAACAAAATTGATAAACCGATGTATGT
GTATTATTTCGAAAGTCCGGAAAAATTTGCCTTTAACAAAGTGATTCGCACCGAAAATCA
GAATGAAATTAGCCTGGAAAAATTCAATGAATTTAAAGAAACCATTCAGAATAAACTGTT
TAAACAGGATGGCTTTAAAGATATTTCACTGTATGAACCGGGTAAAGGTGATGAAAAAC
CGACACCGCTGCTGATGCATCTGAAACTGCCTCGTAATACCGGTATGCTGCCGTATAC
CAATACCAATAATGTTAGCACCCTGATTGAACAGGGCTATAGCATCAAAATTGATAAAAT
TGTGCGCATTGTGATTGATGGCAAACATTATATCAAAGCCGAAGCCAGCGTTGTTTCAA
GCCTGGATTTTAAAGATGATGTGAGCAAAGGCGATAGCTGGGGTAAAGCAAACTATAAT
GATTGGAGCAATAAACTGACCCCGAATGAACTGGCAGATGTGAATGATTATATGCGTGG
TGGTTATACCGCCATTAACAATTATCTGATTAGCAATGGTCCGGTGAATAATCCGAATCC
GGAACTGGATAGCAAAATTACCAATATTGAAAATGCCCTGAAACGCGAACCGATTCCGA
CCAATCTGACCGTTTATCGTCGTAGCGGTCCGCAAGAATTTGGTCTGACCCTGACCAGT
CCGGAATATGACTTTAACAAACTGGAAAATATTGATGCCTTTAAAAGCAAATGGGAAGG
TCAGGCACTGAGCTATCCGAACTTTATTAGCACCAGCATTGGTAGCGTTAATATGAGCG
CATTTGCCAAACGTAAAATTGTGCTGCGTATTACCATTCCGAAAGGTAGTCCGGGTGCA
TATCTGAGCGCAATTCCGGGTTATGCCGGTGAATATGAAGTTCTGCTGAATCATGGCAG
CAAATTCAAAATTAACAAAATTGATAGCTATAAAGATGGCACCATTACCAAACTGATTGT
TGATGCAACCCTGATTCCGTAA

SEQ ID 7 – CDTb C37 construct. CDTb' (minus signal peptide) ligated to Glutathione-S-transferase protein (GST underlined) polypeptide sequence.

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD
VKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKL
PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIP
QIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPLGSHMEIVNEDILPNNGLMG
YYFTDEHFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKSDVKSIRWTGRIIPSKDGEYTLSTDR
DDVLMQVNTESTISNTLKVNMKKGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPE
ENLFLRDYSNIEKDDPFIPNNNFFDPKLMSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVK
WEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKAIKTEARDPLVAAYPIVGVGME
KLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTANVTTNYSHTTDNSTA
VQDSNGESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQEN

QIGNNLSPGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGT
KNSSGQIVTEGNSWSDYISQIDSISASIILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEK
AFGATKKDGLLYFNDIPIDESCVELIFDDNTANKIKDSLKTLSDKKIYNVKLERGMNILIKTPTY
FTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPL
TSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTP
EILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEA
TLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDREL
LVLSVD

SEQ ID 8 – CDTb C37construct. CDTb' (minus pro-domain) ligated to Glutathione-S-transferase protein (GST underlined) polynucleotide sequence.

atgtcccctatactaggttattggaaaattaagggccttgtgcaacccactcgacttcttttggaatatcttgaagaaaaatatgaag
agcatttgtatgagcgcgatgaaggtgataaatggcgaaacaaaaagtttgaattgggtttggagtttcccaatcttccttattatatt
gatggtgatgttaaattaacacagtctatggccatcatacgttatatagctgacaagcacaacatgttgggtggttgtccaaaagag
cgtgcagagatttcaatgcttgaaggagcggttttggatattagatacggtgtttcgagaattgcatatagtaaagactttgaaactct
caaagttgattttcttagcaagctacctgaaatgctgaaaatgttcgaagatcgtttatgtcataaaacatatttaaatggtgatcatgt
aacccatcctgacttcatgttgtatgacgctcttgatgttgttttatacatggacccaatgtgcctggatgcgttcccaaaattagtttgttt
taaaaaacgtattgaagctatcccacaaattgataagtacttgaaatccagcaagtatatagcatggcctttgcagggctggcaa
gccacgtttggtggtggcgaccatcctccaaaatcggatctggaagttctgttccaggggcccctgggatccatatggaaattgtg
aatgaagatattctgccgaataatggtctgatgggatactactttaccgatgaacattttaaagatctgaaactgatggcaccgatta
aagatggcaatctgaaatttgaagaaaaaaaagtggataaactgctggataaagataaaagtgatgtgaaaagcattcgttgg
accggtcgtattattccgagcaaagatggtgaatacaccctgagcaccgatcgtgatgatgttctgatgcaggttaataccgaaag
caccattagcaataccctgaaagtgaatatgaaaaaaggcaaagaatataaagtgcgcattgaactgcaggataaaaatctgg
gtagcattgataatctgagcagcccgaatctgtattgggaactggatggtatgaaaaaaatcattccggaagaaaacctgtttctg
cgcgattatagcaatattgaaaaagatgatccgtttattccgaataataacttttttgatccgaaactgatgagcgattgggaagatg
aagatctggataccgataatgataatattccggatagctatgaacgcaatggctataccattaaagatctgattgccgtgaaatgg
gaagatagctttgcagaacagggctataagaaatatgtgagcaattatctggaaagcaataccgcaggcgatccgtataccgat
tatgaaaaagcaagcggcagctttgataaagccattaaaaccgaagcacgtgatccgctggttgcagcatatccgattgttggtg
ttggtatggaaaaactgattattagcaccaatgaacatgcaagcaccgatcagggtaaaaccgttagccgtgcaaccaccaata
gcaaaaccgaaagcaatacagccggtgttagcgttaatgttggttatcagaatggttttaccgccaatgtgaccaccaattatagc
cataccaccgataatagcaccgcagttcaggatagcaatggtgaaagctggaataccggtctgagcattaacaaaggtgaaa
gcgcatatatcaatgccaatgtgcgctattataacaccggcaccgcaccgatgtataaagttaccccgaccaccaatctggttctg
gatggtgataccctgagtaccattaaagcacaagaaaatcagattggcaataatctgagtccgggtgatacctatccgaaaaaa
ggtctgagtccgctggcactgaataccatggatcagtttagcagccgtctgattccgattaactatgatcagctgaaaaaactggat
gccggtaaacaaatcaaactggaaaccacccaggttagcggtaattttggcaccaaaaattcaagcggtcagattgttaccgaa
ggtaatagctggtcagattatatcagccagattgatagcattagcgccagcattattctggatacagaaaatgaaagctatgaacg
tcgtgtgaccgcaaaaaatctgcaggacccggaagataaaacaccggaactgaccattggtgaagcaattgaaaaagcattt
ggtgccaccaaaaaagatggcctgctgtattttaacgatattccgattgatgaaagctgcgtggaactgatttttgatgataataccg
ccaataaaatcaaagatagcctgaaaaccctgagcgacaaaaaaatctataatgtgaaactggaacgcggtatgaatattctg
attaaaaccccgacctattttaccaattttgatgattataacaattatccgagcacttggagcaatgtgaataccaccaatcaggatg
gtctgcagggtagcgcaaataaactgaatggtgaaaccaaaatcaaaattccgatgagcgaactgaaaccgtataaacgttat
gtgtttagcggctatagcaaagatccgctgaccagcaatagcattattgtgaaaatcaaagccaaagaagaaaaaaccgattat
ctggttccggaacagggttataccaaatttagctatgaatttgaaaccaccgaaaaagatagcagtaatattgaaattaccctgatt
ggtagcggcaccacctatctggataatctgagtattaccgaactgaatagcacaccggaaattctggatgaaccggaagtgaaa
attccgaccgatcaagaaattatggatgcccataaaatctattttgccgatctgaactttaatccgagcaccggcaatacctatatta
acggcatgtattttgcaccgacccagaccaataaagaagccctggattatattcagaaatatcgtgttgaagccaccctgcagtat
agcggttttaaagatattggcaccaaagataaagaaatgcgtaattatctgggcgatccgaatcagccgaaaaccaattatgtta

atctgcgcagctattttaccggtggcgaaaacattatgacctacaaaaaactgcgcatttatgccattacaccggatgatcgtgaa
ctgctggttctgagcgttgattaa

SEQ ID 9 – CDTb C40 construct. CDTb'' (minus pro-domain and signal peptide) polypeptide sequence.

LMSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDP
YTDYEKASGSFDKAIKTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKTE
SNTAGVSVNVGYQNGFTANVTTNYSHTTDNSTAVQDSNGESWNTGLSINKGESAYINANV
RYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLSPGDTYPKKGLSPLALNTMDQ
FSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQIDSISASIIL
DTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDN
TANKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQDGLQG
SANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSY
EFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFN
PSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPK
TNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVD

SEQ ID 10 – C44 construct. CDTa mutation E428Q polypeptide sequence.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNELAD
VNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLTVYRRSGPQEFGLT
LTSPEYDFNKLENIDAFKSKWEGQALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGAYL
SAIPGYAGqYEVLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

SEQ ID 11 – CDTa mutation E428Q polynucleotide sequence.

atggtttgcaataccacctataaagcaccgattgaacgtccggaagattttctgaaagataaagaaaaagccaaagaatggga
acgcaaagaagcagaacgtattgaacagaaactggaacgtagcgaaaaagaagcactggaaagctacaaaaaagatagc
gtggaaatttcaaaatatagccagacccgcaattatttctatgattatcagattgaagccaatagccgtgaaaaagaatataaaga
actgcgcaatgccattagcaaaaacaaaattgataaaccgatgtatgtgtattatttcgaaagtccggaaaaatttgcctttaacaa
agtgattcgcaccgaaaatcagaatgaaattagcctggaaaaattcaatgaatttaaagaaaccattcagaataaactgtttaaa
caggatggctttaaagatatttcactgtatgaaccgggtaaaggtgatgaaaaaccgacaccgctgctgatgcatctgaaactgc
ctcgtaataccggtatgctgccgtataccaataccaataatgttagcaccctgattgaacagggctatagcatcaaaattgataaa
attgtgcgcattgtgattgatggcaaacattatatcaaagccgaagccagcgttgtttcaagcctggattttaaagatgatgtgagca

aaggcgatagctggggtaaagcaaactataatgattggagcaataaactgaccccgaatgaactggcagatgtgaatgattat
atgcgtggtggttataccgccattaacaattatctgattagcaatggtccggtgaataatccgaatccggaactggatagcaaaatt
accaatattgaaaatgccctgaaacgcgaaccgattccgaccaatctgaccgtttatcgtcgtagcggtccgcaagaatttggtct
gaccctgaccagtccggaatatgactttaacaaactggaaaatattgatgcctttaaaagcaaatgggaaggtcaggcactgag
ctatccgaactttattagcaccagcattggtagcgttaatatgagcgcatttgccaaacgtaaaattgtgctgcgtattaccattccga
aaggtagtccgggtgcatatctgagcgcaattccgggttatgccggtCaatatgaagttctgctgaatcatggcagcaaattcaaa
attaacaaaattgatagctataaagatggcaccattaccaaactgattgttgatgcaaccctgattccgtaa

SEQ ID 12 – C54 construct. CDTa mutation E430Q polypeptide sequence.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNELAD
VNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLTVYRRSGPQEFGLT
LTSPEYDFNKLENIDAFKSKWEGQALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGAYL
SAIPGYAGEYqVLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

SEQ ID 13 – CDTa N terminal domain (residue 44 to residue 240) polypeptide sequence.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKI

SEQ ID 14 – C49 construct. CDTa Nterminal domain without signal peptide, with the linker existing between the N-term domain and the C-term domain (containing the enzymatic activity). This construct covers the fragment from amino acid 44 to aa 268 polypeptide sequence.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVS

SEQ ID 15 – C50 construct. CDTa without signal peptide and the linker existing between the Nterminal and Cterminal domains of CDTa. This construct covers the fragment from aa 44 to aa 260 polypeptide sequence.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSS

SEQ ID NO:16 – Polypeptide sequence of CDTb with pro-domain removed (CDTb')
EIVNEDILPNNGLMGYYFTDEHFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKSDVKSIRWTG
RIIPSKDGEYTLSTDRDDVLMQVNTESTISNTLKVNMKKGKEYKVRIELQDKNLGSIDNLSSP
NLYWELDGMKKIIPEENLFLRDYSNIEKDDPFIPNNNFFDPKLMSDWEDEDLDTDNDNIPDS
YERNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKAIKTEAR
DPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTA
NVTTNYSHTTDNSTAVQDSNGESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNL
VLDGDTLSTIKAQENQIGNNLSPGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGK
QIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQIDSISASIILDTENESYERRVTAKNLQD
PEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDNTANKIKDSLKTLSDKKIYN
VKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELK
PYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGS
GTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQ
TNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIM
TYKKLRIYAITPDDRELLVLSVD

SEQ ID NO:17 – Polypeptide sequence of CDTb with pro-domain removed (CDTb')
catatggaaattgtgaatgaagatattctgccgaataatggtctgatgggatactactttaccgatgaacattttaaagatctgaaac
tgatggcaccgattaaagatggcaatctgaaatttgaagaaaaaaaaagtggataaactgctggataaagataaaagtgatgtg
aaaagcattcgttggaccggtcgtattattccgagcaaagatggtgaatacaccctgagcaccgatcgtgatgatgttctgatgca
ggttaataccgaaagcaccattagcaataccctgaaagtgaatatgaaaaaaggcaaagaatataaagtgcgcattgaactgc
aggataaaaatctgggtagcattgataatctgagcagcccgaatctgtattgggaactggatggtatgaaaaaaatcattccgga
agaaaacctgtttctgcgcgattatagcaatattgaaaaagatgatccgtttattccgaataataactttttgatccgaaactgatga
gcgattgggaagatgaagatctggataccgataatgataatattccggatagctatgaacgcaatggctataccattaaagatctg
attgccgtgaaatgggaagatagctttgcagaacagggctataagaaatatgtgagcaattatctggaaagcaataccgcaggc
gatccgtataccgattatgaaaaagcaagcggcagctttgataaagccattaaaaccgaagcacgtgatccgctggttgcagca
tatccgattgttggtgttggtatggaaaaactgattattagcaccaatgaacatgcaagcaccgatcagggtaaaaccgttagccg

tgcaaccaccaatagcaaaaccgaaagcaatacagccggtgttagcgttaatgttggttatcagaatggttttaccgccaatgtga
ccaccaattatagccataccaccgataatagcaccgcagttcaggatagcaatggtgaaagctggaataccggtctgagcatta
acaaaggtgaaagcgcatatatcaatgccaatgtgcgctattataacaccggcaccgcaccgatgtataaagttaccccgacca
ccaatctggttctggatggtgataccctgagtaccattaaagcacaagaaaatcagattggcaataatctgagtccgggtgatacc
tatccgaaaaaaggtctgagtccgctggcactgaataccatggatcagtttagcagccgtctgattccgattaactatgatcagctg
aaaaaactggatgccggtaaacaaatcaaactggaaaccacccaggttagcggtaattttggcaccaaaaattcaagcggtca
gattgttaccgaaggtaatagctggtcagattatatcagccagattgatagcattagcgccagcattattctggatacagaaaatga
aagctatgaacgtcgtgtgaccgcaaaaaatctgcaggacccggaagataaaacaccggaactgaccattggtgaagcaatt
gaaaaagcatttggtgccaccaaaaaagatggcctgctgtattttaacgatattccgattgatgaaagctgcgtggaactgatttttg
atgataataccgccaataaaatcaaagatagcctgaaaaccctgagcgacaaaaaaatctataatgtgaaactggaacgcgg
tatgaatattctgattaaaaccccgacctattttaccaattttgatgattataacaattatccgagcacttggagcaatgtgaataccac
caatcaggatggtctgcagggtagcgcaaataaactgaatggtgaaaccaaaatcaaaattccgatgagcgaactgaaaccg
tataaacgttatgtgtttagcggctatagcaaagatccgctgaccagcaatagcattattgtgaaaatcaaagccaaagaagaaa
aaaccgattatctggttccggaacagggttataccaaatttagctatgaatttgaaaccaccgaaaaagatagcagtaatattgaa
attaccctgattggtagcggcaccacctatctggataatctgagtattaccgaactgaatagcacaccggaaattctggatgaacc
ggaagtgaaaattccgaccgatcaagaaattatggatgcccataaaatctattttgccgatctgaactttaatccgagcaccggca
atacctatattaacggcatgtattttgcaccgacccagaccaataaagaagccctggattatattcagaaatatcgtgttgaagcca
ccctgcagtatagcggttttaaagatattggcaccaaagataaagaaatgcgtaattatctgggcgatccgaatcagccgaaaa
ccaattatgttaatctgcgcagctattttaccggtggcgaaaacattatgacctacaaaaaactgcgcatttatgccattacaccgg
atgatcgtgaactgctggttctgagcgttgattaa

SEQ ID NO:18 – sequence of Fusion 1 (F1)

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDAN
NIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYF
SYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNK
FLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAAT
GWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANN
IEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYY
FNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRF
LYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVF
KGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFM
PDTAMAAAGGLFEIDGVIYFFGVDGVKAPGFVSINDNKHYFDDSGVMKVGYTEIDGKHFYF
AENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWK
DLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQN
IDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETY
TIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNEN

GEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLD
EKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISE

SEQ ID NO:19 – sequence of Fusion 2 (F2)

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDAN
NIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYF
SYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNK
FLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAAT
GWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANN
IEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYY
FNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRF
LYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVF
KGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFM
PDTAMAAAGGLNQIGDYKYYFNSDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHFY
FAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGW
KDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQ
NIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGET
YTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNE
NGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDL
DEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISE

SEQ ID NO:20 – sequence of Fusion 3 (F3)

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDAN
NIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYF
SYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNK
FLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAAT
GWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANN
IEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYY
FNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRF
LYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVF
KGSNGFEYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFD
EDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDN
GIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDME
NESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIE

DKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYI
AATGSVIIDGEEYYFDPDTAQLVISE

SEQ ID NO:21 – sequence of Fusion 4 (F4)

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDAN
NIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYF
SYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNK
FLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAAT
GWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANN
IEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYY
FNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRF
LYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVF
KGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFM
PDTAMAAAGGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFTGKL
IIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQIGDYKYYFNSDGVMQKGFV
SINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEE
GEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFND
DGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANT
VNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKGINLI
DDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPD
GFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQL
VISE

SEQ ID NO:22 – sequence of Fusion 5 (F5)

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDAN
NIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYF
SYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNK
FLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEAAT
GWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANN
IEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYY
FNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRF
LYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVF
KGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFM
PDTAMAAAGGLFEIDGVIYFFGVDGVKAPGIYGGGFVSINDNKHYFDDSGVMKVGYTEIDG
KHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAV

VGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIE
SGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYY
FGETYTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNY
YFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTG
WLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISE

SEQ ID NO: 23 – nucleotide sequence of F54 Gly

ATGGCAACCGGTTGGCAGACCATCGATGGCAAAAAATATTATTTTAATACCAACACCGC
AATTGCAAGCACCGGCTATACCATTATCAACGGCAAACACTTTTATTTTAACACCGACG
GCATTATGCAGATTGGTGTGTTTAAAGGTCCGAACGGCTTTGAATACTTTGCACCGGCA
AATACCGATGCCAATAATATTGAAGGCCAGGCCATTCTGTATCAGAATGAATTTCTGAC
CCTGAACGGCAAAAAATACTACTTTGGCAGCGATAGCAAAGCAGTTACCGGTTGGCGC
ATCATCAACAATAAGAAATATTACTTCAACCCGAATAATGCAATTGCAGCAATTCATCTG
TGCACCATTAACAACGACAAATATTATTTCAGCTATGACGGTATTCTGCAGAATGGCTAC
ATTACCATCGAACGCAACAACTTTTATTTCGATGCCAACAACGAAAGCAAAATGGTGAC
CGGTGTTTTCAAAGGCCCTAATGGTTTTGAGTATTTCGCTCCGGCAAACACCCATAATA
ACAACATTGAAGGTCAGGCGATCGTTTATCAGAACAAATTCCTGACGCTGAATGGTAAG
AAATACTATTTCGATAATGACAGCAAAGCCGTGACCGGCTGGCAGACAATTGACGGGA
AGAAATATTACTTTAATCTGAATACCGCAGAAGCAGCAACCGGTTGGCAAACGATCGAC
GGTAAAAAGTACTACTTCAACCTGAACACAGCCGAAGCAGCCACAGGATGGCAGACTA
TTGATGGAAAAAAATACTATTTCAACACCAACACCTTTATTGCATCTACCGGTTATACCA
GCATTAACGGTAAACATTTCTACTTCAACACCGATGGTATCATGCAGATCGGCGTTTTCA
AAGGTCCAAATGGTTTCGAATACTTTGCCCCTGCCAATACAGATGCAAATAACATCGAG
GGTCAGGCAATCCTGTACCAAAACAAATTTCTGACCCTGAATGGGAAAAAATATTACTTT
GGTAGCGATTCTAAAGCCGTTACCGGTCTGCGTACCATTGATGGTAAAAAATACTACTT
TAATACGAATACAGCCGTTGCGGTTACAGGCTGGCAGACCATTAACGGGAAAAAATACT
ATTTTAACACAAATACCAGCATTGCCTCAACGGGTTATACCATTATTTCGGGTAAACACT
TCTACTTTAATACCGATGGTATTATGCAAATCGGAGTCTTTAAAGGACCTGATGGGTTCG
AATATTTTGCGCCTGCGAACACTGATGCGAACAATATCGAAGGACAGGCAATCCGCTAT
CAGAATCGCTTTCTGTATCTGCACGACAACATCTATTATTTTGGCAACAATTCAAAAGCA
GCCACCGGCTGGGTTACAATTGATGGCAACCGCTACTATTTCGAACCGAATACCGCAAT
GGGTGCAAATGGCTACAAAACCATCGATAATAAAAATTTCTATTTTCGCAACGGTCTGC
CGCAGATCGGGGTATTTAAAGGTAGCAACGGCTTCGAATACTTCGCTCCAGCGAATAC
GGACGCGAACAATATTGAGGGTCAAGCGATTCGTTATCAAAACCGTTTTCTGCATCTGC
TGGGCAAAATCTACTACTTTGGCAATAACAGTAAAGCAGTTACTGGATGGCAGACAATC
AATGGTAAAGTGTACTATTTTATGCCGGATACCGCCATGGCAGCAGCCGGTGGTCTGTT

TGAAATTGATGGCGTGATCTATTTTTTTGGTGTGGATGGTGTTAAAGCACCGGGAATAT
ACGGTGGTACCGGCTTTGTGACCGTGGGTGATGATAAATACTATTTCAATCCGATTAAC
GGTGGTGCAGCGAGCATTGGCGAAACCATCATCGATGACAAAAACTATTATTTCAACCA
GAGCGGTGTGCTGCAGACCGGTGTGTTTAGCACCGAAGATGGCTTTAAATATTTTGCG
CCAGCGAACACCCTGGATGAAAACCTGGAAGGCGAAGCGATTGATTTTACCGGCAAAC
TGATCATCGATGAAAACATCTATTACTTCGATGATAACTATCGTGGTGCGGTGGAATGG
AAAGAACTGGATGGCGAAATGCATTATTTTTCTCCGGAAACCGGTAAAGCGTTTAAAGG
CCTGAACCAGATCGGCGATTACAAATACTACTTCAACAGCGATGGCGTGATGCAGAAA
GGCTTTGTGAGCATCAACGATAACAAACACTATTTCGATGATAGCGGTGTGATGAAAGT
GGGCTATACCGAAATTGATGGCAAACATTTCTACTTCGCGGAAAACGGCGAAATGCAGA
TTGGCGTGTTCAATACCGAAGATGGTTTCAAATACTTCGCGCACCATAACGAAGATCTG
GGTAACGAAGAAGGCGAAGAAATTAGCTATAGCGGCATCCTGAACTTCAACAACAAAAT
CTACTACTTTGATGATAGCTTTACCGCGGTGGTGGGCTGGAAAGATCTGGAAGATGGC
AGCAAATATTATTTCGATGAAGATACCGCGGAAGCGTATATTGGCCTGAGCCTGATTAA
CGATGGCCAGTACTATTTTAACGATGATGGCATTATGCAGGTGGGTTTCGTGACCATTA
ATGATAAAGTGTTCTATTTCAGCGATAGCGGCATTATTGAAAGCGGCGTGCAGAACATT
GATGATAACTACTTCTACATCGATGATAACGGCATTGTGCAGATCGGCGTTTTTGATAC
CAGCGATGGCTACAAATATTTCGCACCGGCCAATACCGTGAACGATAACATTTATGGCC
AGGCGGTGGAATATAGCGGTCTGGTGCGTGTGGGCGAAGATGTGTATTATTTCGGCGA
AACCTATACCATCGAAACCGGCTGGATTTATGATATGGAAAACGAAAGCGATAAATATTA
CTTTAATCCGGAAACGAAAAAAGCGTGCAAAGGCATTAACCTGATCGATGATATCAAAT
ACTATTTTGATGAAAAAGGCATTATGCGTACCGGTCTGATTAGCTTCGAAAACAACAACT
ATTACTTCAACGAAAACGGTGAAATGCAGTTCGGCTACATCAACATCGAAGATAAAATG
TTCTACTTCGGCGAAGATGGTGTTATGCAGATTGGTGTTTTTAACACCCCGGATGGCTT
CAAATACTTTGCCCATCAGAATACCCTGGATGAAAATTTCGAAGGTGAAAGCATTAACTA
TACCGGCTGGCTGGATCTGGATGAAAAACGCTACTACTTCACCGATGAATACATTGCGG
CGACCGGCAGCGTGATTATTGATGGCGAAGAATACTACTTCGATCCGGATACCGCGCA
GCTGGTGATTAGCGAACATCATCATCATCACCAT

SEQ ID NO: 24 – amino acid of F54Gly

MATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTD
ANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKY
YFSYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQ
NKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYY
FNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEY
FAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTG

WQTINGKKYYFNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIE
GQAIRYQNRFLYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKT
IDNKNFYFRNGLPQIGVFKGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNS
KAVTGWQTINGKVYYFMPDTAMAAAGGLFEIDGVIYFFGVDGVKAPGIYGGTGFVTVGDDK
YYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFT
GKLIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQIGDYKYYFNSDGVMQK
GFVSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLG
NEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYY
FNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAP
ANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKG
INLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFN
TPDGFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDT
AQLVISEHHHHHH

SEQ ID NO: 25 – nucleotide sequence of F54 New

ATGGCAACCGGTTGGCAGACCATCGATGGCAAAAAATATTATTTTAATACCAACACCGC
AATTGCAAGCACCGGCTATACCATTATCAACGGCAAACACTTTTATTTTAACACCGACG
GCATTATGCAGATTGGTGTGTTTAAAGGTCCGAACGGCTTTGAATACTTTGCACCGGCA
AATACCGATGCCAATAATATTGAAGGCCAGGCCATTCTGTATCAGAATGAATTTCTGAC
CCTGAACGGCAAAAAATACTACTTTGGCAGCGATAGCAAAGCAGTTACCGGTTGGCGC
ATCATCAACAATAAGAAATATTACTTCAACCCGAATAATGCAATTGCAGCAATTCATCTG
TGCACCATTAACAACGACAAATATTATTTCAGCTATGACGGTATTCTGCAGAATGGCTAC
ATTACCATCGAACGCAACAACTTTTATTTCGATGCCAACAACGAAAGCAAAATGGTGAC
CGGTGTTTTCAAAGGCCCTAATGGTTTTGAGTATTTCGCTCCGGCAAACACCCATAATA
ACAACATTGAAGGTCAGGCGATCGTTTATCAGAACAAATTCCTGACGCTGAATGGTAAG
AAATACTATTTCGATAATGACAGCAAAGCCGTGACCGGCTGGCAGACAATTGACGGGA
AGAAATATTACTTTAATCTGAATACCGCAGAAGCAGCAACCGGTTGGCAAACGATCGAC
GGTAAAAAGTACTACTTCAACCTGAACACAGCCGAAGCAGCCACAGGATGGCAGACTA
TTGATGGAAAAAAATACTATTTCAACACCAACACCTTTATTGCATCTACCGGTTATACCA
GCATTAACGGTAAACATTTCTACTTCAACACCGATGGTATCATGCAGATCGGCGTTTTCA
AAGGTCCAAATGGTTTCGAATACTTTGCCCCTGCCAATACAGATGCAAATAACATCGAG
GGTCAGGCAATCCTGTACCAAAACAAATTTCTGACCCTGAATGGGAAAAAATATTACTTT
GGTAGCGATTCTAAAGCCGTTACCGGTCTGCGTACCATTGATGGTAAAAAATACTACTT
TAATACGAATACAGCCGTTGCGGTTACAGGCTGGCAGACCATTAACGGGAAAAAATACT
ATTTTAACACAAATACCAGCATTGCCTCAACGGGTTATACCATTATTTCGGGTAAACACT
TCTACTTTAATACCGATGGTATTATGCAAATCGGAGTCTTTAAAGGACCTGATGGGTTCG
AATATTTTGCGCCTGCGAACACTGATGCGAACAATATCGAAGGACAGGCAATCCGCTAT

CAGAATCGCTTTCTGTATCTGCACGACAACATCTATTATTTTGGCAACAATTCAAAAGCA
GCCACCGGCTGGGTTACAATTGATGGCAACCGCTACTATTTCGAACCGAATACCGCAAT
GGGTGCAAATGGCTACAAAACCATCGATAATAAAAATTTCTATTTTCGCAACGGTCTGC
CGCAGATCGGGGTATTTAAAGGTAGCAACGGCTTCGAATACTTCGCTCCAGCGAATAC
GGACGCGAACAATATTGAGGGTCAAGCGATTCGTTATCAAAACCGTTTTCTGCATCTGC
TGGGCAAAATCTACTACTTTGGCAATAACAGTAAAGCAGTTACTGGATGGCAGACAATC
AATGGTAAAGTGTACTATTTTATGCCGGATACCGCCATGGCAGCAGCCGGTGGTCTGTT
TGAAATTGATGGCGTGATCTATTTTTTTGGTGTGGATGGTGTTAAAGCAGTTACCGGCTT
TGTGACCGTGGGTGATGATAAATACTATTTCAATCCGATTAACGGTGGTGCAGCGAGCA
TTGGCGAAACCATCATCGATGACAAAAACTATTATTTCAACCAGAGCGGTGTGCTGCAG
ACCGGTGTGTTTAGCACCGAAGATGGCTTTAAATATTTTGCGCCAGCGAACACCCTGGA
TGAAAACCTGGAAGGCGAAGCGATTGATTTTACCGGCAAACTGATCATCGATGAAAACA
TCTATTACTTCGATGATAACTATCGTGGTGCGGTGGAATGGAAAGAACTGGATGGCGAA
ATGCATTATTTTTCTCCGGAAACCGGTAAAGCGTTTAAAGGCCTGAACCAGATCGGCGA
TTACAAATACTACTTCAACAGCGATGGCGTGATGCAGAAAGGCTTTGTGAGCATCAACG
ATAACAAACACTATTTCGATGATAGCGGTGTGATGAAAGTGGGCTATACCGAAATTGAT
GGCAAACATTTCTACTTCGCGGAAAACGGCGAAATGCAGATTGGCGTGTTCAATACCGA
AGATGGTTTCAAATACTTCGCGCACCATAACGAAGATCTGGGTAACGAAGAAGGCGAA
GAAATTAGCTATAGCGGCATCCTGAACTTCAACAACAAAATCTACTACTTTGATGATAGC
TTTACCGCGGTGGTGGGCTGGAAAGATCTGGAAGATGGCAGCAAATATTATTTCGATGA
AGATACCGCGGAAGCGTATATTGGCCTGAGCCTGATTAACGATGGCCAGTACTATTTTA
ACGATGATGGCATTATGCAGGTGGGTTTCGTGACCATTAATGATAAAGTGTTCTATTTCA
GCGATAGCGGCATTATTGAAAGCGGCGTGCAGAACATTGATGATAACTACTTCTACATC
GATGATAACGGCATTGTGCAGATCGGCGTTTTTGATACCAGCGATGGCTACAAATATTT
CGCACCGGCCAATACCGTGAACGATAACATTTATGGCCAGGCGGTGGAATATAGCGGT
CTGGTGCGTGTGGGCGAAGATGTGTATTATTTCGGCGAAACCTATACCATCGAAACCG
GCTGGATTTATGATATGGAAAACGAAAGCGATAAATATTACTTTAATCCGGAAACGAAAA
AAGCGTGCAAAGGCATTAACCTGATCGATGATATCAAATACTATTTTGATGAAAAAGGCA
TTATGCGTACCGGTCTGATTAGCTTCGAAAACAACAACTATTACTTCAACGAAAACGGT
GAAATGCAGTTCGGCTACATCAACATCGAAGATAAAATGTTCTACTTCGGCGAAGATGG
TGTTATGCAGATTGGTGTTTTTAACACCCCGGATGGCTTCAAATACTTTGCCCATCAGAA
TACCCTGGATGAAAATTTCGAAGGTGAAAGCATTAACTATACCGGCTGGCTGGATCTGG
ATGAAAAACGCTACTACTTCACCGATGAATACATTGCGGCGACCGGCAGCGTGATTATT
GATGGCGAAGAATACTACTTCGATCCGGATACCGCGCAGCTGGTGATTAGCGAACATC
ATCATCATCACCAT

SEQ ID NO: 26 amino acid sequence of F54 New

MATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTD
ANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKY
YFSYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQ
NKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEA
ATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDA
NNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKK
YYFNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQN
RFLYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIG
VFKGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYF
MPDTAMAAAGGLFEIDGVIYFFGVDGVKAVTGFVTVGDDKYYFNPINGGAASIGETIIDDKN
YYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFTGKLIIDENIYYFDDNYRGAVE
WKELDGEMHYFSPETGKAFKGLNQIGDYKYYFNSDGVMQKGFVSINDNKHYFDDSGVMK
VGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYY
FDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFY
FSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVR
VGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLI
SFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFE
GESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISEHHHHHH

SEQ ID NO: 27 nucleotide sequence of F5 ToxB

ATGGCAACCGGTTGGCAGACCATCGATGGCAAAAAATATTATTTTAATACCAACACCGC
AATTGCAAGCACCGGCTATACCATTATCAACGGCAAACACTTTTATTTTAACACCGACG
GCATTATGCAGATTGGTGTGTTTAAAGGTCCGAACGGCTTTGAATACTTTGCACCGGCA
AATACCGATGCCAATAATATTGAAGGCCAGGCCATTCTGTATCAGAATGAATTTCTGAC
CCTGAACGGCAAAAAATACTACTTTGGCAGCGATAGCAAAGCAGTTACCGGTTGGCGC
ATCATCAACAATAAGAAATATTACTTCAACCCGAATAATGCAATTGCAGCAATTCATCTG
TGCACCATTAACAACGACAAATATTATTTCAGCTATGACGGTATTCTGCAGAATGGCTAC
ATTACCATCGAACGCAACAACTTTTATTTCGATGCCAACAACGAAAGCAAAATGGTGAC
CGGTGTTTTCAAAGGCCCTAATGGTTTTGAGTATTTCGCTCCGGCAAACACCCATAATA
ACAACATTGAAGGTCAGGCGATCGTTTATCAGAACAAATTCCTGACGCTGAATGGTAAG
AAATACTATTTCGATAATGACAGCAAAGCCGTGACCGGCTGGCAGACAATTGACGGGA
AGAAATATTACTTTAATCTGAATACCGCAGAAGCAGCAACCGGTTGGCAAACGATCGAC
GGTAAAAAGTACTACTTCAACCTGAACACAGCCGAAGCAGCCACAGGATGGCAGACTA
TTGATGGAAAAAAATACTATTTCAACACCAACACCTTTATTGCATCTACCGGTTATACCA
GCATTAACGGTAAACATTTCTACTTCAACACCGATGGTATCATGCAGATCGGCGTTTTCA
AAGGTCCAAATGGTTTCGAATACTTTGCCCCTGCCAATACAGATGCAAATAACATCGAG
GGTCAGGCAATCCTGTACCAAAACAAATTTCTGACCCTGAATGGGAAAAAATATTACTTT

GGTAGCGATTCTAAAGCCGTTACCGGTCTGCGTACCATTGATGGTAAAAAATACTACTT
TAATACGAATACAGCCGTTGCGGTTACAGGCTGGCAGACCATTAACGGGAAAAAATACT
ATTTTAACACAAATACCAGCATTGCCTCAACGGGTTATACCATTATTTCGGGTAAACACT
TCTACTTTAATACCGATGGTATTATGCAAATCGGAGTCTTTAAAGGACCTGATGGGTTCG
AATATTTTGCGCCTGCGAACACTGATGCGAACAATATCGAAGGACAGGCAATCCGCTAT
CAGAATCGCTTTCTGTATCTGCACGACAACATCTATTATTTTGGCAACAATTCAAAAGCA
GCCACCGGCTGGGTTACAATTGATGGCAACCGCTACTATTTCGAACCGAATACCGCAAT
GGGTGCAAATGGCTACAAAACCATCGATAATAAAAATTTCTATTTTCGCAACGGTCTGC
CGCAGATCGGGGTATTTAAAGGTAGCAACGGCTTCGAATACTTCGCTCCAGCGAATAC
GGACGCGAACAATATTGAGGGTCAAGCGATTCGTTATCAAAACCGTTTTCTGCATCTGC
TGGGCAAAATCTACTACTTTGGCAATAACAGTAAAGCAGTTACTGGATGGCAGACAATC
AATGGTAAAGTGTACTATTTTATGCCGGATACCGCCATGGCAGCAGCCGGTGGTCTGTT
TGAAATTGATGGCGTGATCTATTTTTTTGGTGTGGATGGTGTTAAAGCAGTGAGCGGTC
TGATTTATATTAACGATAGCCTGTATTACTTTAAACCACCGGTGAATAACCTGATTACCG
GCTTTGTGACCGTGGGTGATGATAAATACTATTTCAATCCGATTAACGGTGGTGCAGCG
AGCATTGGCGAAACCATCATCGATGACAAAAACTATTATTTCAACCAGAGCGGTGTGCT
GCAGACCGGTGTGTTTAGCACCGAAGATGGCTTTAAATATTTTGCGCCAGCGAACACC
CTGGATGAAAACCTGGAAGGCGAAGCGATTGATTTTACCGGCAAACTGATCATCGATGA
AAACATCTATTACTTCGATGATAACTATCGTGGTGCGGTGGAATGGAAAGAACTGGATG
GCGAAATGCATTATTTTTCTCCGGAAACCGGTAAAGCGTTTAAAGGCCTGAACCAGATC
GGCGATTACAAATACTACTTCAACAGCGATGGCGTGATGCAGAAAGGCTTTGTGAGCAT
CAACGATAACAAACACTATTTCGATGATAGCGGTGTGATGAAAGTGGGCTATACCGAAA
TTGATGGCAAACATTTCTACTTCGCGGAAAACGGCGAAATGCAGATTGGCGTGTTCAAT
ACCGAAGATGGTTTCAAATACTTCGCGCACCATAACGAAGATCTGGGTAACGAAGAAG
GCGAAGAAATTAGCTATAGCGGCATCCTGAACTTCAACAACAAAATCTACTACTTTGATG
ATAGCTTTACCGCGGTGGTGGGCTGGAAAGATCTGGAAGATGGCAGCAAATATTATTTC
GATGAAGATACCGCGGAAGCGTATATTGGCCTGAGCCTGATTAACGATGGCCAGTACT
ATTTTAACGATGATGGCATTATGCAGGTGGGTTTCGTGACCATTAATGATAAAGTGTTCT
ATTTCAGCGATAGCGGCATTATTGAAAGCGGCGTGCAGAACATTGATGATAACTACTTC
TACATCGATGATAACGGCATTGTGCAGATCGGCGTTTTTGATACCAGCGATGGCTACAA
ATATTTCGCACCGGCCAATACCGTGAACGATAACATTTATGGCCAGGCGGTGGAATATA
GCGGTCTGGTGCGTGTGGGCGAAGATGTGTATTATTTCGGCGAAACCTATACCATCGA
AACCGGCTGGATTTATGATATGGAAAACGAAAGCGATAAATATTACTTTAATCCGGAAAC
GAAAAAAGCGTGCAAAGGCATTAACCTGATCGATGATATCAAATACTATTTTGATGAAAA
AGGCATTATGCGTACCGGTCTGATTAGCTTCGAAAACAACAACTATTACTTCAACGAAAA
CGGTGAAATGCAGTTCGGCTACATCAACATCGAAGATAAAATGTTCTACTTCGGCGAAG
ATGGTGTTATGCAGATTGGTGTTTTTAACACCCCGGATGGCTTCAAATACTTTGCCCATC

AGAATACCCTGGATGAAAATTTCGAAGGTGAAAGCATTAACTATACCGGCTGGCTGGAT
CTGGATGAAAAACGCTACTACTTCACCGATGAATACATTGCGGCGACCGGCAGCGTGA
TTATTGATGGCGAAGAATACTACTTCGATCCGGATACCGCGCAGCTGGTGATTAGCGAA
CATCATCATCATCACCAT

SEQ ID NO: 28 amino acid sequence of F5 ToxB

MATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTD
ANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKY
YFSYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQ
NKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEA
ATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDA
NNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKK
YYFNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQN
RFLYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIG
VFKGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYF
MPDTAMAAAGGLFEIDGVIYFFGVDGVKAVSGLIYINDSLYYFKPPVNNLITGFVTVGDDKYY
FNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEGEAIDFTGK
LIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQIGDYKYYFNSDGVMQKGF
VSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNE
EGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFN
DDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPAN
TVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKGINL
IDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTP
DGFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQ
LVISEHHHHHH

SEQ ID NO: 29 - nucleotide sequence of F52 new

ATGGCAACCGGTTGGCAGACCATCGATGGCAAAAAATATTATTTTAATACCAACACCGC
AATTGCAAGCACCGGCTATACCATTATCAACGGCAAACACTTTTATTTTAACACCGACG
GCATTATGCAGATTGGTGTGTTTAAAGGTCCGAACGGCTTTGAATACTTTGCACCGGCA
AATACCGATGCCAATAATATTGAAGGCCAGGCCATTCTGTATCAGAATGAATTTCTGAC
CCTGAACGGCAAAAAATACTACTTTGGCAGCGATAGCAAAGCAGTTACCGGTTGGCGC
ATCATCAACAATAAGAAATATTACTTCAACCCGAATAATGCAATTGCAGCAATTCATCTG
TGCACCATTAACAACGACAAATATTATTTCAGCTATGACGGTATTCTGCAGAATGGCTAC
ATTACCATCGAACGCAACAACTTTTATTTCGATGCCAACAACGAAAGCAAAATGGTGAC
CGGTGTTTTCAAAGGCCCTAATGGTTTTGAGTATTTCGCTCCGGCAAACACCCATAATA
ACAACATTGAAGGTCAGGCGATCGTTTATCAGAACAAATTCCTGACGCTGAATGGTAAG

AAATACTATTTCGATAATGACAGCAAAGCCGTGACCGGCTGGCAGACAATTGACGGGA
AGAAATATTACTTTAATCTGAATACCGCAGAAGCAGCAACCGGTTGGCAAACGATCGAC
GGTAAAAAGTACTACTTCAACCTGAACACAGCCGAAGCAGCCACAGGATGGCAGACTA
TTGATGGAAAAAAATACTATTTCAACACCAACACCTTTATTGCATCTACCGGTTATACCA
GCATTAACGGTAAACATTTCTACTTCAACACCGATGGTATCATGCAGATCGGCGTTTTCA
AAGGTCCAAATGGTTTCGAATACTTTGCCCCTGCCAATACAGATGCAAATAACATCGAG
GGTCAGGCAATCCTGTACCAAAACAAATTTCTGACCCTGAATGGGAAAAAATATTACTTT
GGTAGCGATTCTAAAGCCGTTACCGGTCTGCGTACCATTGATGGTAAAAAATACTACTT
TAATACGAATACAGCCGTTGCGGTTACAGGCTGGCAGACCATTAACGGGAAAAAATACT
ATTTTAACACAAATACCAGCATTGCCTCAACGGGTTATACCATTATTTCGGGTAAACACT
TCTACTTTAATACCGATGGTATTATGCAAATCGGAGTCTTTAAAGGACCTGATGGGTTCG
AATATTTTGCGCCTGCGAACACTGATGCGAACAATATCGAAGGACAGGCAATCCGCTAT
CAGAATCGCTTTCTGTATCTGCACGACAACATCTATTATTTTGGCAACAATTCAAAAGCA
GCCACCGGCTGGGTTACAATTGATGGCAACCGCTACTATTTCGAACCGAATACCGCAAT
GGGTGCAAATGGCTACAAAACCATCGATAATAAAAATTTCTATTTTCGCAACGGTCTGC
CGCAGATCGGGGTATTTAAAGGTAGCAACGGCTTCGAATACTTCGCTCCAGCGAATAC
GGACGCGAACAATATTGAGGGTCAAGCGATTCGTTATCAAAACCGTTTTCTGCATCTGC
TGGGCAAAATCTACTACTTTGGCAATAACAGTAAAGCAGTTACTGGATGGCAGACAATC
AATGGTAAAGTGTACTATTTTATGCCGGATACCGCCATGGCAGCAGCCGGTGGTCTGTT
TGAAATTGATGGCGTGATCTATTTTTTTGGTGTGGATGGTGTTAAAGCAGTGAAAGGCC
TGAACCAGATCGGCGATTACAAATACTACTTCAACAGCGATGGCGTGATGCAGAAAGG
CTTTGTGAGCATCAACGATAACAAACACTATTTCGATGATAGCGGTGTGATGAAAGTGG
GCTATACCGAAATTGATGGCAAACATTTCTACTTCGCGGAAAACGGCGAAATGCAGATT
GGCGTGTTCAATACCGAAGATGGTTTCAAATACTTCGCGCACCATAACGAAGATCTGGG
TAACGAAGAAGGCGAAGAAATTAGCTATAGCGGCATCCTGAACTTCAACAACAAAATCT
ACTACTTTGATGATAGCTTTACCGCGGTGGTGGGCTGGAAAGATCTGGAAGATGGCAG
CAAATATTATTTCGATGAAGATACCGCGGAAGCGTATATTGGCCTGAGCCTGATTAACG
ATGGCCAGTACTATTTTAACGATGATGGCATTATGCAGGTGGGTTTCGTGACCATTAAT
GATAAAGTGTTCTATTTCAGCGATAGCGGCATTATTGAAAGCGGCGTGCAGAACATTGA
TGATAACTACTTCTACATCGATGATAACGGCATTGTGCAGATCGGCGTTTTTGATACCA
GCGATGGCTACAAATATTTCGCACCGGCCAATACCGTGAACGATAACATTTATGGCCAG
GCGGTGGAATATAGCGGTCTGGTGCGTGTGGGCGAAGATGTGTATTATTTCGGCGAAA
CCTATACCATCGAAACCGGCTGGATTTATGATATGGAAAACGAAAGCGATAAATATTACT
TTAATCCGGAAACGAAAAAAGCGTGCAAAGGCATTAACCTGATCGATGATATCAAATAC
TATTTTGATGAAAAAGGCATTATGCGTACCGGTCTGATTAGCTTCGAAAACAACAACTAT
TACTTCAACGAAAACGGTGAAATGCAGTTCGGCTACATCAACATCGAAGATAAAATGTT
CTACTTCGGCGAAGATGGTGTTATGCAGATTGGTGTTTTTAACACCCCGGATGGCTTCA

AATACTTTGCCCATCAGAATACCCTGGATGAAAATTTCGAAGGTGAAAGCATTAACTATA
CCGGCTGGCTGGATCTGGATGAAAAACGCTACTACTTCACCGATGAATACATTGCGGC
GACCGGCAGCGTGATTATTGATGGCGAAGAATACTACTTCGATCCGGATACCGCGCAG
CTGGTGATTAGCGAACATCATCATCATCACCAT

SEQ ID NO: 30 – amino acid sequence of F52 New

MATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTD
ANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKY
YFSYDGILQNGYITIERNNFYFDANNESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQ
NKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNLNTAEA
ATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDA
NNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKK
YYFNTNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQN
RFLYLHDNIYYFGNNSKAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIG
VFKGSNGFEYFAPANTDANNIEGQAIRYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYF
MPDTAMAAAGGLFEIDGVIYFFGVDGVKAVKGLNQIGDYKYYFNSDGVMQKGFVSINDNKH
YFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYS
GILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQV
GFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYG
QAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFNPETKKACKGINLIDDIKYYF
DEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPDGFKYFA
HQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISEHHH
HHH

SEQ ID NO: 31 – amino acid sequence of Toxin A

MSLISKEELIKLAYSIRPRENEYKTILTNLDEYNKLTTNNNENKYLQLKKLNESIDVFMN
KYKTSSRNRALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWADI
NAEYNIKLWYDSEAFLVNTLKKAIVESSTTEALQLLEEEIQNPQFDNMKFYKKRMEFIYD
RQKRFINYYKSQINKPTVPTIDDIIKSHLVSEYNRDETVLESYRTNSLRKINSNHGIDIR
ANSLFTEQELLNIYSQELLNRGNLAAASDIVRLLALKNFGGVYLDVDMLPGIHSDLFKTI
SRPSSIGLDRWEMIKLEAIMKYKKYINNYTSENFDKLDQQLKDNFKLIIESKSEKSEIFS
KLENLNVSDLEIKIAFALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNQHLNPAIESD
NNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQVGFMPEARSTISLSGPGAYASAYYDF
INLQENTIEKTLKASDLIEFKFPENNLSQLTEQEINSLWSFDQASAKYQFEKYVRDYTGG
SLSEDNGVDFNKNTALDKNYLLNNKIPSNNVEEAGSKNYVHYIIQLQGDDISYEATCNLF
SKNPKNSIIIQRNMNESAKSYFLSDDGESILELNKYRIPERLKNKEKVKVTFIGHGKDEF

NTSEFARLSVDSLSNEISSFLDTIKLDISPKNVEVNLLGCNMFSYDFNVEETYPGKLLLS
IMDKITSTLPDVNKNSITIGANQYEVRINSEGRKELLAHSGKWINKEEAIMSDLSSKEYI
FFDSIDNKLKAKSKNIPGLASISEDIKTLLLDASVSPDTKFILNNLKLNIESSIGDYIYY
EKLEPVKNIIHNSIDDLIDEFNLLENVSDELYELKKLNNLDEKYLISFEDISKNNSTYSV
RFINKSNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGNLLDNIQLDHTSQVNT
LNAAFFIQSLIDYSSNKDVLNDLSTSVKVQLYAQLFSTGLNTIYDSIQLVNLISNAVNDT
INVLPTITEGIPIVSTILDGINLGAAIKELLDEHDPLLKKELEAKVGVLAINMSLSIAAT
VASIVGIGAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSESKKYGPLKT
EDDKILVPIDDLVISEIDFNNNSIKLGTCNILAMEGGSGHTVTGNIDHFFSSPSISSHIP
SLSIYSAIGIETENLDFSKKIMMLPNAPSRVFWWETGAVPGLRSLENDGTRLLDSIRDLY
PGKFYWRFYAFFDYAITTLKPVYEDTNIKIKLDKDTRNFIMPTITTNEIRNKLSYSFDGA
GGTYSLLLSSYPISTNINLSKDDLWIFNIDNEVREISIENGTIKKGKLIKDVLSKIDINK
NKLIIGNQTIDFSGDIDNKDRYIFLTCELDDKISLIIEINLVAKSYSLLLSGDKNYLISN
LSNTIEKINTLGLDSKNIAYNYTDESNNKYFGAISKTSQKSIIHYKKDSKNILEFYNDST
LEFNSKDFIAEDINVFMKDDINTITGKYYVDNNTDKSIDFSISLVSKNQVKVNGLYLNES
VYSSYLDFVKNSDGHHNTSNFMNLFLDNISFWKLFGFENINFVIDKYFTLVGKTNLGYVE
FICDNNKNIDIYFGEWKTSSSKSTIFSGNGRNVVVEPIYNPDTGEDISTSLDFSYEPLYG
IDRYINKVLIAPDLYTSLININTNYYSNEYYPEIIVLNPNTFHKKVNINLDSSSFEYKWS
TEGSDFILVRYLEESNKKILQKIRIKGILSNTQSFNKMSIDFKDIKKLSLGYIMSNFKSF
NSENELDRDHLGFKIIDNKTYYYDEDSKLVKGLININNSLFYFDPIEFNLVTGWQTINGK
KYYFDINTGAALTSYKIINGKHFYFNNDGVMQLGVFKGPDGFEYFAPANTQNNNIEGQAI
VYQSKFLTLNGKKYYFDNNSKAVTGWRIINNEKYYFNPNNAIAAVGLQVIDNNKYYFNPD
TAIISKGWQTVNGSRYYFDTDTAIAFNGYKTIDGKHFYFDSDCVVKIGVFSTSNGFEYFA
PANTYNNNIEGQAIVYQSKFLTLNGKKYYFDNNSKAVTGLQTIDSKKYYFNTNTAEAATG
WQTIDGKKYYFNTNTAEAATGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQ
IGVFKGPNGFEYFAPANTDANNIEGQAILYQNEFLTLNGKKYYFGSDSKAVTGWRIINNK
KYYFNPNNAIAAIHLCTINNDKYYFSYDGILQNGYITIERNNFYFDANNESKMVTGVFKG
PNGFEYFAPANTHNNNIEGQAIVYQNKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNL
NTAEAATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFY
FNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVT
GLRTIDGKKYYFNTNTAVAVTGWQTINGKKYYFNTNTSIASTGYTIISGKHFYFNTDGIM
QIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRFLYLHDNIYYFGNNSKAATGWVTIDG
NRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVFKGSNGFEYFAPANTDANNIEGQAI
RYQNRFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFMPDTAMAAAGGLFEIDGVIYFFGV
DGVKAPGIYG

SEQ ID NO: 32 – amino acid sequence of Toxin B

MSLVNRKQLEKMANVRFRTQEDEYVAILDALEEYHNMSENTVVEKYLKLKDINSLTDIYI
DTYKKSGRNKALKKFKEYLVTEVLELKNNNLTPVEKNLHFVWIGGQINDTAINYINQWKD
VNSDYNVNVFYDSNAFLINTLKKTVVESAINDTLESFRENLNDPRFDYNKFFRKRMEIIY
DKQKNFINYYKAQREENPELIIDDIVKTYLSNEYSKEIDELNTYIEESLNKITQNSGNDV
RNFEEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMYLDVDMLPGIQPDLFES
IEKPSSVTVDFWEMTKLEAIMKYKEYIPEYTSEHFDMLDEEVQSSFESVLASKSDKSEIF
SSLGDMEASPLEVKIAFNSKGIINQGLISVKDSYCSNLIVKQIENRYKILNNSLNPAISE
DNDFNTTTNTFIDSIMAEANADNGRFMMELGKYLRVGFFPDVKTTINLSGPEAYAAAYQD
LLMFKEGSMNIHLIEADLRNFEISKTNISQSTEQEMASLWSFDDARAKAQFEEYKRNYFE
GSLGEDDNLDFSQNIVVDKEYLLEKISSLARSSERGYIHYIVQLQGDKISYEAACNLFAK
TPYDSVLFQKNIEDSEIAYYYNPGDGEIQEIDKYKIPSIISDRPKIKLTFIGHGKDEFNT
DIFAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLLGCNMFSYSINVEETYPGKLLLKVK
DKISELMPSISQDSIIVSANQYEVRINSEGRRELLDHSGEWINKEESIIKDISSKEYISF
NPKENKITVKSKNLPELSTLLQEIRNNSNSSDIELEEKVMLTECEINVISNIDTQIVEER
IEEAKNLTSDSINYIKDEFKLIESISDALCDLKQQNELEDSHFISFEDISETDEGFSIRF
INKETGESIFVETEKTIFSEYANHITEEISKIKGTIFDTVNGKLVKKVNLDTTHEVNTLN
AAFFIQSLIEYNSSKESLSNLSVAMKVQVYAQLFSTGLNTITDAAKVVELVSTALDETID
LLPTLSEGLPIIATIIDGVSLGAAIKELSETSDPLLRQEIEAKIGIMAVNLTTATTAIIT
SSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVVDYFKHVSLVETEGVFTLLD
DKIMMPQDDLVISEIDFNNNSIVLGKCEIWRMEGGSGHTVTDDIDHFFSAPSITYREPHL
SIYDVLEVQKEELDLSKDLMVLPNAPNRVFAWETGWTPGLRSLENDGTKLLDRIRDNYEG
EFYWRYFAFIADALITTLKPRYEDTNIRINLDSNTRSFIVPIITTEYIREKLSYSFYGSG
GTYALSLSQYNMGINIELSESDVWIIDVDNVVRDVTIESDKIKKGDLIEGILSTLSIEEN
KIILNSHEINFSGEVNGSNGFVSLTFSILEGINAIIEVDLLSKSYKLLISGELKILMLNS
NHIQQKIDYIGFNSELQKNIPYSFVDSEGKENGFINGSTKEGLFVSELPDVVLISKVYMD
DSKPSFGYYSNNLKDVKVITKDNVNILTGYYLKDDIKISLSLTLQDEKTIKLNSVHLDES
GVAEILKFMNRKGNTNTSDSLMSFLESMNIKSIFVNFLQSNIKFILDANFIISGTTSIGQ
FEFICDENDNIQPYFIKFNTLETNYTLYVGNRQNMIVEPNYDLDDSGDISSTVINFSQKY
LYGIDSCVNKVVISPNIYTDEINITPVYETNNTYPEVIVLDANYINEKINVNINDLSIRY
VWSNDGNDFILMSTSEENKVSQVKIRFVNVFKDKTLANKLSFNFSDKQDVPVSEIILSFT
PSYYEDGLIGYDLGLVSLYNEKFYINNFGMMVSGLIYINDSLYYFKPPVNNLITGFVTVG
DDKYYFNPINGGAASIGETIIDDKNYYFNQSGVLQTGVFSTEDGFKYFAPANTLDENLEG
EAIDFTGKLIIDENIYYFDDNYRGAVEWKELDGEMHYFSPETGKAFKGLNQIGDYKYYFN
SDGVMQKGFVSINDNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFA
HHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTAVVGWKDLEDGSKYYFDEDTAEAYIG
LSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFYIDDNGIVQIG
VFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESD

KYYFNPETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIED
KMFYFGEDGVMQIGVFNTPDGFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYI
AATGSVIIDGEEYYFDPDTAQLVISE

SEQ ID NO: 33 – amino acid sequence of CDTb" C39 when expressed in fusion with GST.

LMSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKA
IKTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTANVTTNYSH
TTDNSTAVQDSNGESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLS
PGDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQ
IDSISASIILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDNT
ANKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIP
MSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLD
NLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEAT
LQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVD

Remarks:

- The protein tested in the cytotoxicity assay was obtained after cleavage of the GST by PreScission protease
- Following experimental results, it is demonstrated that the mature CDTb (without SP and pro-domain) starts at the Ser^{212} (in red and underlined in the sequence).

SEQ ID NO: 34 – amino acid sequence of CdtB receptor binding domain with linker in N-term of sequence, from aa 620-876 (C52)

MTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEE
KTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIY
FADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLR
SYFTGGENIMTYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 35 – Nucleotide sequence of C52

ATGACCAATTTTGATGATTATAACAATTATCCGAGCACTTGGAGCAATGTGAATACCACCAATCAGGATGGTCTG
CAGGGTAGCGCAAATAAACTGAATGGTGAAACCAAAATCAAAATTCCGATGAGCGAACTGAAACCGTATAAACGT
TATGTGTTTAGCGGCTATAGCAAAGATCCGCTGACCAGCAATAGCATTATTGTGAAAATCAAAGCCAAAGAAGAA
AAAACCGATTATCTGGTTCCGGAACAGGGTTATACCAAATTTAGCTATGAATTTGAAACCACCGAAAAAGATAGC
AGTAATATTGAAATTACCCTGATTGGTAGCGGCACCACCTATCTGGATAATCTGAGTATTACCGAACTGAATAGC

ACACCGGAAATTCTGGATGAACCGGAAGTGAAAATTCCGACCGATCAAGAAATTATGGATGCCCATAAAATCTAT
TTTGCCGATCTGAACTTTAATCCGAGCACCGGCAATACCTATATTAACGGCATGTATTTTGCACCGACCCAGACC
AATAAAGAAGCCCTGGATTATATTCAGAAATATCGTGTTGAAGCCACCCTGCAGTATAGCGGTTTTAAAGATATT
GGCACCAAAGATAAAGAAATGCGTAATTATCTGGGCGATCCGAATCAGCCGAAAACCAATTATGTTAATCTGCGC
AGCTATTTTACCGGTGGCGAAAACATTATGACCTACAAAAAACTGCGCATTTATGCCATTACACCGGATGATCGT
GAACTGCTGGTTCTGAGCGTTGATGGCGGTCACCACCATCATCATCATTAA

SEQ ID NO: 36 – amino acid sequence of CdtB receptor binding domain without linker in N-term of sequence, from aa 636-876 (C55)

MNTTNQDGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSY
EFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYIN
GMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLR
IYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 37 – nucleotide sequence of C55

ATGAATACCACCAATCAGGATGGTCTGCAGGGTAGCGCAAATAAACTGAATGGTGAAACCAAAATCAAAATTCCGATGAG
CGAACTGAAACCGTATAAACGTTATGTGTTTAGCGGCTATAGCAAAGATCCGCTGACCAGCAATAGCATTATTGTGAAAA
TCAAAGCCAAAGAAGAAAAAACCGATTATCTGGTTCCGGAACAGGGTTATACCAAATTTAGCTATGAATTTGAAACCACC
GAAAAAGATAGCAGTAATATTGAAATTACCCTGATTGGTAGCGGCACCACCTATCTGGATAATCTGAGTATTACCGAACT
GAATAGCACACCGGAAATTCTGGATGAACCGGAAGTGAAAATTCCGACCGATCAAGAAATTATGGATGCCCATAAAATCT
ATTTTGCCGATCTGAACTTTAATCCGAGCACCGGCAATACCTATATTAACGGCATGTATTTTGCACCGACCCAGACCAAT
AAAGAAGCCCTGGATTATATTCAGAAATATCGTGTTGAAGCCACCCTGCAGTATAGCGGTTTTAAAGATATTGGCACCAA
AGATAAAGAAATGCGTAATTATCTGGGCGATCCGAATCAGCCGAAAACCAATTATGTTAATCTGCGCAGCTATTTTACCG
GTGGCGAAAACATTATGACCTACAAAAAACTGCGCATTTATGCCATTACACCGGATGATCGTGAACTGCTGGTTCTGAGC
GTTGATGGCGGTCACCACCATCATCATCATTAA

SEQ ID NO: 38 – amino acid sequence of CDTb prodomain sequence (long, aa43-211) (C58)

MEIVNEDILPNNGLMGYYFTDEHFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKSDVKSIRWTGRIIPSKDGEYTL
STDRDDVLMQVNTESTISNTLKVNMKKGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPEENLFLRDYS
NIEKDDPFIPNNNFFDPKLM

SEQ ID NO: 39 – amino acid sequence of CDTb prodomain sequence (short, aa43-186) (C59)

MEIVNEDILPNNGLMGYYFTDEHFKDLKLMAPIKDGNLKFEEKKVDKLLDKDKSDVKSIRWTGRIIPSKDGEYTL
STDRDDVLMQVNTESTISNTLKVNMKKGKEYKVRIELQDKNLGSIDNLSSPNLYWELDGMKKIIPEENLF

SEQ ID NO: 40 – amino acid sequence of Fusion CDTa N-term with linker (aa44-268) to CDTb receptor binding domain with linker in N term of sequence (aa621-876) (C60)

CDTa part of the fusion is underlined.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISL
EKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTL
IEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDTNFDDYNNYPSTWSNVNTTNQDGL
QGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYT
KFSYEFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKI
YFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYL
GDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 41 – amino acid sequence of Fusion CDTa N-term with linker (aa44-268) to CDTb receptor binding domain without linker in N term of sequence (aa636-876) (C61)

CDTa part of the fusion is underlined.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISL
EKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTL
IEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDNTTNQDGLQGSANKLNGETKIKIP
MSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNI
EITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYI
NGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSY
FTGGENIMTYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 42 – amino acid sequence of Fusion CDTa N-term without linker (aa44-260) to CDTb receptor binding domain with linker in N term of sequence (aa621-876) (C62)

CDTa part of the fusion is underlined.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISL
EKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTL
IEQGYSIKIDKIVRIVIDGKHYIKAEASVVSTNFDDYNNYPSTWSNVNTTNQDGLQGSANKL
NGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFE
TTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNF

NPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPK
TNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 43 – amino acid sequence of Fusion CDTa N-term without linker (aa44-260) to CDTb receptor binding domain without linker in N term of sequence (aa636-876) (C63)

CDTa part of the fusion is underlined.

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISL
EKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTL
IEQGYSIKIDKIVRIVIDGKHYIKAEASVVSNTTNQDGLQGSANKLNGETKIKIPMSELKPY
KRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGS
GTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAP
TQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENI
MTYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 44 – amino acid sequence of Fusion F2- CDTb receptor binding domain with linker in N term of sequence (aa621-876) (C64)

F2 sequence is underlined.

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNEF
LTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYFSYDGILQNGYITIERNNFYFDANN
ESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAE
AATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFE
YFAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYYFN
TNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRFLYLHDNIYYFGNNS
KAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVFKGSNGFEYFAPANTDANNIEGQAIRYQN
RFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFMPDTAMAAAGGLNQIGDYKYYFNSDGVMQKGFVSINDNKHYFD
DSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTA
VVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFY
IDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFN
PETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPD
GFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISETNFDDYNNY
PSTWSNVNTTNQDGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQG
YTKFSYEFETTEKDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPST

GNTYINGMYFAPTQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIM
TYKKLRIYAITPDDRELLVLSVDGGHHHHHH

SEQ ID NO: 45 – amino acid sequence of Fusion of F2 to CDTb receptor binding domain without linker in N term of sequence (aa636-876) with 2 heterogeneous Gly residues between F2 and CTDb sequences (C65)

F2 sequence is underlined.

MGWQTIDGKKYYFNTNTAIASTGYTIINGKHFYFNTDGIMQIGVFKGPNGFEYFAPANTDANNIEGQAILYQNEF
LTLNGKKYYFGSDSKAVTGWRIINNKKYYFNPNNAIAAIHLCTINNDKYYFSYDGILQNGYITIERNNFYFDANN
ESKMVTGVFKGPNGFEYFAPANTHNNNIEGQAIVYQNKFLTLNGKKYYFDNDSKAVTGWQTIDGKKYYFNLNTAE
AATGWQTIDGKKYYFNLNTAEAATGWQTIDGKKYYFNTNTFIASTGYTSINGKHFYFNTDGIMQIGVFKGPNGFE
YFAPANTDANNIEGQAILYQNKFLTLNGKKYYFGSDSKAVTGLRTIDGKKYYFNTNTAVAVTGWQTINGKKYYFN
TNTSIASTGYTIISGKHFYFNTDGIMQIGVFKGPDGFEYFAPANTDANNIEGQAIRYQNRFLYLHDNIYYFGNNS
KAATGWVTIDGNRYYFEPNTAMGANGYKTIDNKNFYFRNGLPQIGVFKGSNGFEYFAPANTDANNIEGQAIRYQN
RFLHLLGKIYYFGNNSKAVTGWQTINGKVYYFMPDTAMAAAGGLNQIGDYKYYFNSDGVMQKGFVSINDNKHYFD
DSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEEGEEISYSGILNFNNKIYYFDDSFTA
VVGWKDLEDGSKYYFDEDTAEAYIGLSLINDGQYYFNDDGIMQVGFVTINDKVFYFSDSGIIESGVQNIDDNYFY
IDDNGIVQIGVFDTSDGYKYFAPANTVNDNIYGQAVEYSGLVRVGEDVYYFGETYTIETGWIYDMENESDKYYFN
PETKKACKGINLIDDIKYYFDEKGIMRTGLISFENNNYYFNENGEMQFGYINIEDKMFYFGEDGVMQIGVFNTPD
GFKYFAHQNTLDENFEGESINYTGWLDLDEKRYYFTDEYIAATGSVIIDGEEYYFDPDTAQLVISEGGNVNTTNQ
DGLQGSANKLNGETKIKIPMSELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTE
KDSSNIEITLIGSGTTYLDNLSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAP
TQTNKEALDYIQKYRVEATLQYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITP
DDRELLVLSVDGGHHHHHH

SEQ ID NO: 46 – amino acid sequence of CDTa without signal peptide, with two mutations (E428Q, E430Q, aa 44-463) (C67)

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQTRNYFYDYQIEAN
SREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLEKFNEFKETIQNKLFKQDGFKDISLY
EPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDV
SKGDSWGKANYNDWSNKLTPNELADVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLT
VYRRSGPQEFGLTLTSPEYDFNKLENIDAFKSKWEGQALSYPNFISTSIGSVNMSAFAKRKIVLRITIPKGSPGA
YLSAIPGYAGqYqVLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

SEQ ID NO: 47 – nucleotide sequence of C67

ATGGTTTGCAATACCACCTATAAAGCACCGATTGAACGTCCGGAAGATTTTCTGAAAGATAAAGAAAAAGCCAAA
GAATGGGAACGCAAAGAAGCAGAACGTATTGAACAGAAACTGGAACGTAGCGAAAAAGAAGCACTGGAAAGCTAC
AAAAAAGATAGCGTGGAAATTTCAAAATATAGCCAGACCCGCAATTATTTCTATGATTATCAGATTGAAGCCAAT
AGCCGTGAAAAAGAATATAAAGAACTGCGCAATGCCATTAGCAAAAACAAAATTGATAAACCGATGTATGTGTAT
TATTTCGAAAGTCCGGAAAAATTTGCCTTTAACAAAGTGATTCGCACCGAAAATCAGAATGAAATTAGCCTGGAA
AAATTCAATGAATTTAAAGAAACCATTCAGAATAAACTGTTTAAACAGGATGGCTTTAAAGATATTTCACTGTAT
GAACCGGGTAAAGGTGATGAAAAACCGACACCGCTGCTGATGCATCTGAAACTGCCTCGTAATACCGGTATGCTG
CCGTATACCAATACCAATAATGTTAGCACCCTGATTGAACAGGGCTATAGCATCAAAATTGATAAAATTGTGCGC
ATTGTGATTGATGGCAAACATTATATCAAAGCCGAAGCCAGCGTTGTTTCAAGCCTGGATTTTAAAGATGATGTG
AGCAAAGGCGATAGCTGGGGTAAAGCAAACTATAATGATTGGAGCAATAAACTGACCCCGAATGAACTGGCAGAT
GTGAATGATTATATGCGTGGTGGTTATACCGCCATTAACAATTATCTGATTAGCAATGGTCCGGTGAATAATCCG
AATCCGGAACTGGATAGCAAAATTACCAATATTGAAAATGCCCTGAAACGCGAACCGATTCCGACCAATCTGACC
GTTTATCGTCGTAGCGGTCCGCAAGAATTTGGTCTGACCCTGACCAGTCCGGAATATGACTTTAACAAACTGGAA
AATATTGATGCCTTTAAAAGCAAATGGGAAGGTCAGGCACTGAGCTATCCGAACTTTATTAGCACCAGCATTGGT
AGCGTTAATATGAGCGCATTTGCCAAACGTAAAATTGTGCTGCGTATTACCATTCCGAAAGGTAGTCCGGGTGCA
TATCTGAGCGCAATTCCGGGTTATGCCGGTCAATATCAGGTTCTGCTGAATCATGGCAGCAAATTCAAAATTAAC
AAAATTGATAGCTATAAAGATGGCACCATTACCAAACTGATTGTTGATGCAACCCTGATTCCGTAA

SEQ ID NO: 48 – amino acid sequence of CDTa without signal peptide, with seven mutations (R345A, Q350A, N385A, R402A, S388F, E428Q, E430Q, aa 44-463) (C69)

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISL
EKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTL
IEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNEL
ADVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLTVY**A**RSGP**A**EF
GLTLTSPEYDFNKLENIDAFKSKWEGQALSYP**A**FI**F**TSIGSVNMSAFAK**A**KIVLRITIPKGS
PGAYLSAIPGYAG**Q**Y**Q**VLLNHGSKFKINKIDSYKDGTITKLIVDATLIP

SEQ ID NO: 49 – nucleotide sequence of C69

ATGGTTTGCAATACCACCTATAAAGCACCGATTGAACGTCCGGAAGATTTTCTGAAAGATAA
AGAAAAAGCCAAAGAATGGGAACGCAAAGAAGCAGAACGTATTGAACAGAAACTGGAACGTA
GCGAAAAAGAAGCACTGGAAAGCTACAAAAAAGATAGCGTGGAAATTTCAAAATATAGCCAG
ACCCGCAATTATTTCTATGATTATCAGATTGAAGCCAATAGCCGTGAAAAAGAATATAAAGA
ACTGCGCAATGCCATTAGCAAAAACAAAATTGATAAACCGATGTATGTGTATTATTTCGAAA
GTCCGGAAAAATTTGCCTTTAACAAAGTGATTCGCACCGAAAATCAGAATGAAATTAGCCTG

GAAAAATTCAATGAATTTAAAGAAACCATTCAGAATAAACTGTTTAAACAGGATGGCTTTAA
AGATATTTCACTGTATGAACCGGGTAAAGGTGATGAAAAACCGACACCGCTGCTGATGCATC
TGAAACTGCCTCGTAATACCGGTATGCTGCCGTATACCAATACCAATAATGTTAGCACCCTG
ATTGAACAGGGCTATAGCATCAAAATTGATAAAATTGTGCGCATTGTGATTGATGGCAAACA
TTATATCAAAGCCGAAGCCAGCGTTGTTTCAAGCCTGGATTTTAAAGATGATGTGAGCAAAG
GCGATAGCTGGGGTAAAGCAAACTATAATGATTGGAGCAATAAACTGACCCCGAATGAACTG
GCAGATGTGAATGATTATATGCGTGGTGGTTATACCGCCATTAACAATTATCTGATTAGCAA
TGGTCCGGTGAATAATCCGAATCCGGAACTGGATAGCAAAATTACCAATATTGAAAATGCCC
TGAAACGCGAACCGATTCCGACCAATCTGACCGTTTATGCACGTAGCGGTCCGGCAGAATTT
GGTCTGACCCTGACCAGTCCGGAATATGACTTTAACAAACTGGAAAATATTGATGCCTTTAA
AAGCAAATGGGAAGGTCAGGCACTGAGCTATCCGGCATTTATTTTCACCAGCATTGGTAGCG
TTAATATGAGCGCATTTGCCAAAGCAAAAATTGTGCTGCGTATTACCATTCCGAAAGGTAGT
CCGGGTGCATATCTGAGCGCAATTCCGGGTTATGCCGGTCAGTATCAGGTTCTGCTGAATCA
TGGCAGCAAATTCAAAATTAACAAAATTGATAGCTATAAAGATGGCACCATTACCAAACTGA
TTGTTGATGCAACCCTGATTCCG

SEQ ID NO: 50 – amino acid sequence of CDTb without signal sequence and prodomain (mature fragment based on MS data) with Ca2+ binding motif mutation (aa212-876, mut Asp-9-11-13 Ala) (C97)

3 mutated residues in this sequence. Three Asp residues were mutated into Ala. They are highlighted in bold and underlined.

MSDWEDEDL**A**T**A**N**A**NIPDSYERNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKAI
KTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTANVTTNYSHT
TDNSTAVQDSNGESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLSP
GDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQI
DSISASIILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDNTA
NKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPM
SELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDN
LSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATL
QYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVDHHHHHH

SEQ ID NO: 51 – amino acid sequence of CDTb with prodomain removed (CDTb'', aa212-876) (C55)

MSDWEDEDLDTDNDNIPDSYERNGYTIKDLIAVKWEDSFAEQGYKKYVSNYLESNTAGDPYTDYEKASGSFDKAI
KTEARDPLVAAYPIVGVGMEKLIISTNEHASTDQGKTVSRATTNSKTESNTAGVSVNVGYQNGFTANVTTNYSHT
TDNSTAVQDSNGESWNTGLSINKGESAYINANVRYYNTGTAPMYKVTPTTNLVLDGDTLSTIKAQENQIGNNLSP
GDTYPKKGLSPLALNTMDQFSSRLIPINYDQLKKLDAGKQIKLETTQVSGNFGTKNSSGQIVTEGNSWSDYISQI
DSISASIILDTENESYERRVTAKNLQDPEDKTPELTIGEAIEKAFGATKKDGLLYFNDIPIDESCVELIFDDNTA
NKIKDSLKTLSDKKIYNVKLERGMNILIKTPTYFTNFDDYNNYPSTWSNVNTTNQDGLQGSANKLNGETKIKIPM
SELKPYKRYVFSGYSKDPLTSNSIIVKIKAKEEKTDYLVPEQGYTKFSYEFETTEKDSSNIEITLIGSGTTYLDN
LSITELNSTPEILDEPEVKIPTDQEIMDAHKIYFADLNFNPSTGNTYINGMYFAPTQTNKEALDYIQKYRVEATL
QYSGFKDIGTKDKEMRNYLGDPNQPKTNYVNLRSYFTGGENIMTYKKLRIYAITPDDRELLVLSVDHHHHHH

SEQ ID NO: 52 – amino acid sequence of CDTa without signal peptide, with five mutations (R345A, Q350A, N385A, R402A, S388F, aa 44-463) (C107)

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEISKYSQ
TRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIRTENQNEISLE
KFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTGMLPYTNTNNVSTLIE
QGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSWGKANYNDWSNKLTPNELAD
VNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIENALKREPIPTNLTVYARSGPAEFGLT
LTSPEYDFNKLENIDAFKSKWEGQALSYPAFIFTSIGSVNMSAFAKAKIVLRITIPKGSPGAYL
SAIPGYAGEYEVLLNHGSKFKINKIDSYKDGTITKLIVDATLIPHHHHHH**

SEQ ID NO: 53 – Polynucleotide sequence of CDTa without signal peptide, with five mutations (R345A, Q350A, N385A, R402A, S388F, aa 44-463) (C107)

ATGGTTTGCAATACCACCTATAAAGCACCGATTGAACGTCCGGAAGATTTTCTGAAAGA
TAAAGAAAAAGCCAAAGAATGGGAACGCAAAGAAGCAGAACGTATTGAACAGAAACTG
GAACGTAGCGAAAAAGAAGCACTGGAAAGCTACAAAAAAGATAGCGTGGAAATTTCAAA
ATATAGCCAGACCCGCAATTATTTCTATGATTATCAGATTGAAGCCAATAGCCGTGAAAA
AGAATATAAAGAACTGCGCAATGCCATTAGCAAAAACAAAATTGATAAACCGATGTATGT
GTATTATTTCGAAAGTCCGGAAAAATTTGCCTTTAACAAAGTGATTCGCACCGAAAATCA
GAATGAAATTAGCCTGGAAAAATTCAATGAATTTAAAGAAACCATTCAGAATAAACTGT
TTAAACAGGATGGCTTTAAAGATATTTCACTGTATGAACCGGGTAAAGGTGATGA
AAAACCGACACCGCTGCTGATGCATCTGAAACTGCCTCGTAATACCGGTATGCTG
CCGTATACCAATACCAATAATGTTAGCACCCTGATTGAACAGGGCTATAGCATCA
AAATTGATAAAATTGTGCGCATTGTGATTGATGGCAAACATTATATCAAAGCCGA
AGCCAGCGTTGTTTCAAGCCTGGATTTTAAAGATGATGTGAGCAAAGGCGATAG
CTGGGGTAAAGCAAACTATAATGATTGGAGCAATAAACTGACCCCGAATGAACT

GGCAGATGTGAATGATTATATGCGTGGTGGTTATACCGCCATTAACAATTATCTG
ATTAGCAATGGTCCGGTGAATAATCCGAATCCGGAACTGGATAGCAAAATTACC
AATATTGAAAATGCCCTGAAACGCGAACCGATTCCGACCAATCTGACCGTTTATG
CACGTAGCGGTCCGGCAGAATTTGGTCTGACCCTGACCAGTCCGGAATATGACTT
TAACAAACTGGAAAATATTGATGCCTTTAAAAGCAAATGGGAAGGTCAGGCACT
GAGCTATCCGGCATTTATTTTCACCAGCATTGGTAGCGTTAATATGAGCGCATTT
GCCAAAGCAAAAATTGTGCTGCGTATTACCATTCCGAAAGGTAGTCCGGGTGCA
TATCTGAGCGCAATTCCGGGTTATGCCGGTgAaTATgAaGTTCTGCTGAATCATGG
CAGCAAATTCAAAATTAACAAAATTGATAGCTATAAAGATGGCACCATTACCAA
ACTGATTGTTGATGCAACCCTGATTCCGCACCACCATCATCATCATTAATAA

SEQ ID NO:54 - Amino acid sequence of CDTa without signal peptide, with six mutations (R345A, Q350A, N385A, R402A, S388F, E430Q, aa 44-463) (C108)

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEI
SKYSQTRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIR
TENQNEISLEKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTG
MLPYTNTNNVSTLIEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSW
GKANYNDWSNKLTPNELADVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIE
NALKREPIPTNLTVYARSGPAEFGLTLTSPEYDFNKLENIDAFKSKWEGQALSYPAFIF
TSIGSVNMSAFAKAKIVLRITIPKGSPGAYLSAIPGYAGEYQVLLNHGSKFKIN
KIDSYKDGTITKLIVDATLIPHHHHHH**

SEQ ID NO:55 – Polynucleotide sequence of CDTa without signal peptide, with six mutations (R345A, Q350A, N385A, R402A, S388F, E430Q, aa 44-463) (C108)

ATGGTTTGCAATACCACCTATAAAGCACCGATTGAACGTCCGGAAGATTTTCTGA
AAGATAAAGAAAAAGCCAAAGAATGGGAACGCAAAGAAGCAGAACGTATTGAA
CAGAAACTGGAACGTAGCGAAAAAGAAGCACTGGAAAGCTACAAAAAAGATAG
CGTGGAAATTTCAAAATATAGCCAGACCCGCAATTATTTCTATGATTATCAGATT
GAAGCCAATAGCCGTGAAAAAGAATATAAAGAACTGCGCAATGCCATTAGCAAA
AACAAAATTGATAAACCGATGTATGTGTATTATTTCGAAAGTCCGGAAAAATTTG
CCTTTAACAAAGTGATTCGCACCGAAAATCAGAATGAAATTAGCCTGGAAAAAT
TCAATGAATTTAAAGAAACCATTCAGAATAAACTGTTTAAACAGGATGGCTTTAA

AGATATTTCACTGTATGAACCGGGTAAAGGTGATGAAAAACCGACACCGCTGCT
GATGCATCTGAAACTGCCTCGTAATACCGGTATGCTGCCGTATACCAATACCAAT
AATGTTAGCACCCTGATTGAACAGGGCTATAGCATCAAAATTGATAAAATTGTGC
GCATTGTGATTGATGGCAAACATTATATCAAAGCCGAAGCCAGCGTTGTTTCAAG
CCTGGATTTTAAAGATGATGTGAGCAAAGGCGATAGCTGGGGTAAAGCAAACTA
TAATGATTGGAGCAATAAACTGACCCCGAATGAACTGGCAGATGTGAATGATTA
TATGCGTGGTGGTTATACCGCCATTAACAATTATCTGATTAGCAATGGTCCGGTG
AATAATCCGAATCCGGAACTGGATAGCAAAATTACCAATATTGAAAATGCCCTG
AAACGCGAACCGATTCCGACCAATCTGACCGTTTATGCACGTAGCGGTCCGGCA
GAATTTGGTCTGACCCTGACCAGTCCGGAATATGACTTTAACAAACTGGAAAATA
TTGATGCCTTTAAAAGCAAATGGGAAGGTCAGGCACTGAGCTATCCGGCATTTAT
TTTCACCAGCATTGGTAGCGTTAATATGAGCGCATTTGCCAAAGCAAAAATTGTG
CTGCGTATTACCATTCCGAAAGGTAGTCCGGGTGCATATCTGAGCGCAATTCCGG
GTTATGCCGGTgAaTATcAaGTTCTGCTGAATCATGGCAGCAAATTCAAAATTAAC
AAAATTGATAGCTATAAAGATGGCACCATTACCAAACTGATTGTTGATGCAACCC
TGATTCCGCACCACCATCATCATCATTAATAA

SEQ ID NO:56 - Amino acid sequence of CDTa without signal peptide, with six mutations (R345A, Q350A, N385A, R402A, S388F, E428Q, aa 44-463) (C110

MVCNTTYKAPIERPEDFLKDKEKAKEWERKEAERIEQKLERSEKEALESYKKDSVEI
SKYSQTRNYFYDYQIEANSREKEYKELRNAISKNKIDKPMYVYYFESPEKFAFNKVIR
TENQNEISLEKFNEFKETIQNKLFKQDGFKDISLYEPGKGDEKPTPLLMHLKLPRNTG
MLPYTNTNNVSTLIEQGYSIKIDKIVRIVIDGKHYIKAEASVVSSLDFKDDVSKGDSW
GKANYNDWSNKLTPNELADVNDYMRGGYTAINNYLISNGPVNNPNPELDSKITNIE
NALKREPIPTNLTVYARSGPAEFGLTLTSPEYDFNKLENIDAFKSKWEGQALSYPAFIF
TSIGSVNMSAFAKAKIVLRITIPKGSPGAYLSAIPGYAGQYEVLLNHGSKFKIN
KIDSYKDGTITKLIVDATLIPHHHHHH**

Claims

1. An immunogenic composition comprising an isolated *Clostridium difficile* CDTb protein wherein the composition does not further comprise an isolated protein having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% similarity to SEQ ID NO: 1, SEQ ID NO: 31 or SEQ ID NO: 32.

2. The immunogenic composition of claim 1 wherein the composition does not comprise an isolated *Clostridium difficile* CDTa protein.

3. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 3; or

(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:3; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO:3.

4. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein with the signal sequence removed.

5. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein comprises

(i)SEQ ID NO: 7 or SEQ ID NO: 16; or

(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 7 or SEQ ID NO:16; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750 or 800 contiguous amino acids of SEQ ID NO: 7 or SEQ ID NO:16.

6. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein with the prodomain removed.

7. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 9 or SEQ ID NO: 51; or

(ii) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:9 or SEQ ID NO: 51; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 or 650 contiguous amino acids of SEQ ID NO:9 or SEQ ID NO: 51.

8. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein comprising the receptor binding domain.

9. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 34 or SEQ ID NO: 36; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 34 or SEQ ID NO: 36; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150 or 200 contiguous amino acids of SEQ ID NO: 34 or SEQ ID NO: 36.

10. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a mutated CDTb protein incapable of binding to CDTa.

11. The immunogenic composition of claim 10 wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 50; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 50; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO: 50.

12. The immunogenic composition of claim 1 or claims 3-11 comprising an isolated *Clostridium difficile* CDTa protein.

13. The immunogenic composition of claim 12 wherein the isolated *Clostridium difficile* CDTa protein is truncated.

14. The immunogenic composition of claim 13 wherein the truncated *Clostridium difficile* CDTa protein is a truncated CDTa protein which does not comprise the C-terminal domain.

15. The immunogenic composition of claim 14 wherein the truncated *Clostridium difficile* CDTa protein is

(i) SEQ ID NO: 14 or SEQ ID NO: 15

(i) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 14 or SEQ ID NO: 15; or

(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, or 190 contiguous amino acids of SEQ ID NO: 14 or SEQ ID NO: 15 .

16. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a monomer of CDTb.

17. The immunogenic composition of any preceding claim wherein the isolated *Clostridium difficile* CDTb protein is a multimer of CDTb.

18. The immunogenic composition of claim 17 wherein the isolated *Clostridium difficile* CDTb protein is a heptamer of CDTb.

19. An immunogenic composition comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a truncated CDTb protein comprising the receptor binding domain.

20. The immunogenic composition of claim 19 wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 34 or SEQ ID NO: 36; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 34 or SEQ ID NO: 36; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150 or 200 contiguous amino acids of SEQ ID NO: 34 or SEQ ID NO: 36.

21. An immunogenic composition comprising isolated *Clostridium difficile* CDTb protein wherein the isolated *Clostridium difficile* CDTb protein is a mutated CDTb protein incapable of binding to CDTa.

22. The immunogenic composition of claim 21 wherein the isolated *Clostridium difficile* CDTb protein comprises

(i) SEQ ID NO: 50; or

(i) a variant of CDTb having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 50; or

(iii) a fragment of CDTb having at least 30, 50, 80, 100, 120, 150, 200, 250 or 300 contiguous amino acids of SEQ ID NO: 50.

23. The immunogenic composition of claim 19-22 comprising an isolated *Clostridium difficile* CDTa protein comprising

(i) SEQ ID NO: 1; or

(ii) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:1; or

(iii) a fragment of CDTa having at least at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID NO:1.

24. An immunogenic composition comprising isolated *Clostridium difficile* CDTa protein wherein the isolated *Clostridium difficile* CDTa protein is a truncated CDTa protein which does not comprise the C-terminal domain.

25. The immunogenic composition of claim 24 wherein the isolated *Clostridium difficile* CDTa protein is

(i) SEQ ID NO: 14 or SEQ ID NO: 15; or

(ii)a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO:14 or SEQ ID NO: 15; or

(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, or 190 contiguous amino acids of SEQ ID NO:14 or SEQ ID NO: 15.

26. The immunogenic composition of any one of claims 1 and 3-25 comprising an isolated *Clostridium difficile* CDTa protein containing a mutation which reduces its ADP-ribosyltransferase activity.

27. The immunogenic composition of any one of claims 1 and 3-26 comprising an isolated *Clostridium difficile* CDTa protein having a mutation from glutamate to a different amino acid at position 428.

28. The immunogenic composition of claim 27 wherein the isolated *Clostridium difficile* CDTa protein has a mutation from glutamate to glutamine at position 428.

29. The immunogenic composition of any one of claims 1 and 3-28 comprising an isolated *Clostridium difficile* CDTa protein having a mutation from glutamate to a different amino acid at position 430.

30. The immunogenic composition of claim 29 wherein the isolated *Clostridium difficile* CDTa protein has a mutation from glutamate to glutamine at position 430.

31. The immunogenic composition of any one of claims 26 to 30 wherein the isolated *Clostridium difficile* CDTa protein is

(i) SEQ ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54; or

(ii) a variant of CDTa having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54; or

(iii) a fragment of CDTa having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID ID NO: 46; SEQ ID NO: 48; SEQ ID NO: 52; or SEQ ID NO: 54.

32. An immunogenic composition comprising a fusion protein comprising a CDTa protein and a CDTb protein.

33. The immunogenic composition of claim 32 wherein the CDTa protein is truncated.

34. the immunogenic composition of claim 33 wherein the CDTa protein does not comprise the C-terminal domain.

35. The immunogenic composition of any one of claims 32-34 wherein the CDTb protein is truncated.

36. The immunogenic composition of claim 35 wherein the CDTb protein comprises the receptor binding domain.

37. The immunogenic composition of any one of claims 33-37 wherein the fusion protein is

(i) SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43; or

(ii) a variant having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99%, 100% sequence identity to SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43; or

(iii) a fragment having at least 30, 50, 80, 100, 120, 150, 200, 250, 300, 350 or 400 contiguous amino acids of SEQ ID NO：40; SEQ ID NO: 41; SEQ ID NO: 42; or SEQ ID NO: 43.

38. The immunogenic composition of any preceding claim wherein the composition elicits antibodies that neutralize CDTa or CDTb or both.

39. The immunogenic composition of any preceding claim wherein the composition elicits antibodies that neutralize binary toxin.

40. The immunogenic composition of any preceding claim wherein the immunogenic composition further comprises an isolated *Clostridium difficile* toxin A protein and/or an isolated *C.difficile* toxin B protein.

41. The immunogenic composition of claim 40 wherein the immunogenic composition comprises an isolated *Clostridium difficile* toxin A protein and an isolated *C.difficile* toxin B protein wherein the isolated *Clostridium difficile* toxin A protein and the isolated *C.difficile* toxin B protein form a fusion protein.

42. The immunogenic composition of claim 41 wherein the fusion protein is

(i) SEQ ID NO; 18, 19, 20, 21, 22, 24, 26, 28 or 30; or

(ii) a variant having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to SEQ ID NO: 18, 19, 20, 21, 22, 24, 26, 28 or 30; or

(iii) a fragment of at least 800, 850, 900 or 950 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NO: 18, 19, 20, 21, 22, 24, 26, 28 and 30.

43. An immunogenic composition comprising a fusion protein between an isolated *Clostridium difficile* toxin A protein and/or an isolated *Clostridium difficile* toxin B protein fused to a CDTb protein.

44. The immunogenic composition of claim 43 wherein the fusion protein comprises an isolated *Clostridium difficile* toxin A protein and an isolated *Clostridium difficile* toxin B protein fused to a CDTb protein.

45. The immunogenic composition of claim 44 wherein the fusion protein is

(i)SEQ ID NO: 44 or SEQ ID NO: 45; or

(ii) a variant having at least 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to SEQ ID NO: 44 or SEQ ID NO: 45; or

(iii) a fragment of at least 800, 850, 900 or 950 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NO: 44 or SEQ ID NO: 45.

46. The immunogenic composition of any preceding claim wherein the immunogenic composition further comprises an adjuvant.

47. The immunogenic composition of any one of the preceding claims further comprising additional antigens.

48. The immunogenic composition of any one of the preceding claims further comprising a saccharide from *C.difficile*.

49. A vaccine comprising the immunogenic composition of any one of the preceding claims and a pharmaceutically acceptable excipient.

50. The immunogenic composition of any one claims 1-48 or the vaccine of claim 49 for use in the treatment or prevention of *C.difficile* disease.

51. The immunogenic composition of any one of claims 1-48 or the vaccine of claim 49 for use in the treatment or prevention of disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571.

52. The immunogenic composition of any one claims 1-48 or the vaccine of claim 49 for use in the treatment or prevention of disease caused by *C.difficile* strain 078.

53. A use of the immunogenic composition of any one of claims 1-48 or the vaccine of claim 49 in the preparation of a medicament for the prevention or treatment of *C.difficile* disease.

54. The use of claim 53 wherein the disease is a disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571.

55. The use of claim 53 wherein the disease is a disease caused by *C.difficile* strain 078.

56. A method of preventing or treating *C.difficile* disease comprising administering the immunogenic composition of any one of claims 1-48 or the vaccine of claim 49 to a mammalian subject.

57. The method of claim 56 wherein the disease is a disease caused by a strain of *C.difficile* selected from the group consisting of 078, 019, 023, 027, 033, 034, 036, 045, 058, 059, 063, 066, 075, 078, 080, 111, 112, 203, 250 and 571.

58. The method of claim 56 wherein the disease is a disease caused by *C.difficile* strain 078.