IMMUNOGENIC COMPOSITION

TECHNICAL FIELD

[001] The present invention relates to antigens from *Clostridium difficile*. In particular the invention relates to recombinant protein antigens comprising fragments of toxin A and/or toxin B. The invention additionally relates to immunogenic compositions or vaccines comprising these antigens, and the use of the vaccines and immunogenic compositions of the invention in prophylaxis or therapy. The invention also relates to methods of immunising using the compositions of the invention, and the use of the compositions of the invention in the manufacture of a medicament.

BACKGROUND

[002] *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*.

BRIEF SUMMARY

[003] 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). The C-terminal domains of toxin A and toxin B comprise repeating units, for example the C-terminal domain of toxin A is made up of contiguous repeating units (Dove *et al Infect. Immun.* 58:480-499 (1990)), for this reason the C-terminal domain may be referred to as the 'repeating domain'. These repeat portions can be separated further into short repeats (SRs) and long repeats (LRs) as described in Ho *et al* (PNAS 102:18373-18378 (2005)).

[004] The structure of a 127-aa fragment from the C terminus of the toxin A repeat domain has been determined (Ho *et al* PNAS 102:18373-18378 (2005)). This fragment formed a β-solenoid-like fold, composed predominantly of β strands with a low proportion of α helices.

[005] 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.

[006] There is known to be difficulty involved in designing fusion proteins which fold correctly during expression. The polypeptides of the present invention are fusion proteins in which the native β solenoid-like structure is maintained, and which are seen to provide an immune response against both toxin A and toxin B in mice.

[007] In a first aspect of the invention there is provided a polypeptide comprising a first fragment and a second fragment, wherein

(i) the first fragment is a toxin A repeating domain fragment;
(ii) the second fragment is a toxin B repeating domain fragment;
(iii) the first fragment has a first proximal end;
(iv) the second fragment has a second proximal end; and

polypeptide elicits antibodies that neutralize toxin A or toxin B or both

[008] In a second aspect of the invention there is provided a polypeptide comprising:

(i) SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 orSEQ ID NO:35;
(ii) A variant having at least 90%, 95%, 98%, 99% or 100% similarity to SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35; or
(iii) A fragment of at least 250, 280, 300, 350, 380, 400, 430, 450, 480, 500, 530, 550, 580, or 600 amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35.

[009] In a third aspect of the invention there is provided a polynucleotide encoding a polypeptide of the invention.

[010] In a fourth aspect of the invention there is provided a vector comprising a polynucleotide of the invention linked to an inducible promoter.

[011] In a fifth aspect of the invention there is provided a host cell comprising a vector of the invention or a polynucleotide of the invention.

[012] In a sixth aspect of the invention there is provided an immunogenic composition comprising a polypeptide of the invention and a pharmaceutically acceptable excipient.

[013] In a seventh aspect of the invention there is provided a vaccine comprising an immunogenic composition of the invention.

[014] In an eighth aspect of the invention there is provided a use of an immunogenic composition of the invention or a vaccine of the invention in the treatment or prevention of *C.difficile* disease.

[015] In a ninth aspect of the invention there is provided a use of an immunogenic composition of the invention or a vaccine of the invention in the preparation of a medicament for the prevention or treatment of *C.difficile* disease.

[016] In a tenth aspect of the invention there is provided a method of preventing or treating *C.difficile* disease comprising administering an immunogenic composition of the invention or a vaccine of the invention to a patient.

BRIEF DESCRIPTION OF THE DRAWINGS

[017] FIG. 1– Sequence listings of polypeptides of the invention.

[018] FIG. 2 - Pictorial representation of the C-terminal domains of ToxA and ToxB, with the SR repeats depicted as white boxes and the LR boxes depicted as black boxes.

[019] FIG. 3 - Pictorial representation of a junction between the third SR VIII of ToxA and the fourth SR II of Tox B used in Fusion 1.

[020] FIG.4 - Pictorial representation of a junction between the second SR VIII of ToxA and the third SR II of Tox B used in Fusion 2.

[021] FIG.5 - Pictorial representation of a junction between LRVII of ToxA and LRII of ToxB used in Fusion 3 (containing only part of LRVII of ToxA and part of LR II of ToxB).

[022] FIG.6 - Pictorial representation of a junction between the second SR VIII of ToxA and the third SR I of ToxB used in Fusion 4.

[023] FIG.7 - Pictorial representation of a junction comprising a glycine linker between the last residue of the ToxA protein sequence and the beginning of the fourth SRII of ToxB used in Fusion 5.

[024] FIG.8 - Graphs describing the distribution of *C.difficile* ToxA-ToxB fusions 1-5 as determined by sedimentation velocity analytical ultracentrifugation. Panel a) describes the distribution of Fusion 1, panel b) describes the distribution of Fusion 2, panel c) describes the distribution of Fusion 3, panel d) describes the distribution of Fusion 4 and panel e) describes the distribution of Fusion 5.

[025] FIG.9 - Graph describing the Far-UV spectrum of Fusions, 2, 3, 4, and 5 measured using circular dichroism. The spectrum for fusion 2 is represented by a line with the points depicted as small squares, the spectrum for fusion 3 is represented by a line with the points depicted as small diamond shapes, fusion 4 is represented by a line with the points depicted as circles, and fusion 5 is represented by a line with the points depicted as cross shapes.

[026] FIG.10 - Graph describing the near-UV spectrum of Fusions 2, 3, 4, and 5 measured using circular dichroism. The spectrum for fusion 2 is represented by a line with the points depicted as cross shapes, the spectrum for fusion 3 is represented by a line with the points depicted as circles, the spectrum for fusion 4 is represented by a line with the points depicted as triangles, and the spectrum for fusion 5 is represented by a line with the points depicted as small diamond shapes.

[027] FIG.11 – Graph showing anti-ToxA immunogenicity in mice immunised with a fragment of the C-terminus of toxin A (aa 2387-2706), a fragment of the C-terminus of toxin B (aa 1750-2360), or fusions 1, 2, 3, 4 or 5.

[028] FIG.12 - Graph showing hemagglutination inhibition in mice immunised with a fragment of the C-terminus of toxin A (aa 2387-2706), a fragment of the C-terminus of toxin B (aa 1750-2360), or fusions 1, 2, 3, 4 or 5.

[029] FIG.13 – Graph showing anti-ToxB immunogenicity in mice immunised with a fragment of the C-terminus of toxin A (aa 2387-2706), a fragment of the C-terminus of toxin B (aa 1750-2360), or fusions 1, 2, 3, 4 or 5.

[030] FIG.14 – Cytoxicity inhibition titres from mice immunised with a fragment of the C-terminus of toxin A (aa 2387-2706), a fragment of the C-terminus of toxin B (aa 1750-2360), or fusions 1, 2, 3, 4 or 5.

[031] FIG.15 - Graphs describing the distribution of *C.difficile* ToxA-ToxB fusions F52New, F54Gly, F54New and F5ToxB as determined by sedimentation velocity analytical ultracentrifugation. Panel a) describes the distribution of F52New, panel b) describes the

distribution of F54Gly, panel c) describes the distribution of F54New and panel d) describes the distribution of F5ToxB.

[032] FIG.16 - Graph describing the Far-UV spectrum of fusions F52New, F54Gly, F54New and F5ToxB measured using circular dichroism. The spectrum for F52New is represented by a line with the points depicted as double crosses, the spectrum for F54Gly is represented by a line with the points depicted as triangles, F54New is represented by a line with the points depicted as squares, and F5ToxB is represented by a line with the points depicted as cross shapes.

[033] FIG.17 - Graph describing the Near-UV spectrum of fusions F52New, F54Gly, F54New and F5ToxB measured using circular dichroism. The spectrum for F52New is represented by a line with the points depicted as double crosses, the spectrum for F54Gly is represented by a line with the points depicted as triangles, F54New is represented by a line with the points depicted as squares, and F5ToxB is represented by a line with the points depicted as cross shapes.

[034] FIG.18 – Graph showing anti-ToxA ELISA results for mice immunised with the F2, F52New, F54Gly, G54New or F5ToxB fusions.

[035] FIG.19 – Graph showing anti-ToxB ELISA results for mice immunised with the F2, F52New, F54Gly, F54New or F5ToxB fusions.

[036] FIG.20 – Graph showing hemagglutination inhibition in mice immunised with the F2, F52New, F54Gly, F54New or F5ToxB fusions.

[037] FIG.21 – Graph showing cytotoxicity titres in HT29 cells from mice immunised with the F2, F52New, F54Gly, F54New or F5ToxB fusions.

[038] FIG.22 – Graph showing cytotoxicity titres in IMR90 cells from mice immunised with the F2, F52New, F54Gly, F54New or F5ToxB fusions.

DETAILED DESCRIPTION

POLYPEPTIDES

[039] The invention relates to a polypeptide comprising a first fragment and a second fragment, wherein

(i) the first fragment is a toxin A repeating domain fragment;
(ii) the second fragment is a toxin B repeating domain fragment;
(iii) the first fragment has a first proximal end;
(iv) the second fragment has a second proximal end; and

wherein the first fragment and the second fragment are adjacent to one another and wherein the polypeptide elicits antibodies that neutralize toxin A or toxin B or both.

[040] The term polypeptide refers to a contiguous sequence of amino acids.

[041] The term 'toxin A repeating domain' refers to the C-terminal domain of the toxin A protein from *C.difficile*, comprising repeated sequences. This 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:1.

[042] 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:2.

[043] 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.

[044] The invention therefore the term 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:1 or a variant with 90%, 95%, 98%, 99% or 100% sequence identity to amino acids 1834-2366 of SEQ ID NO:2. In one embodiment a 'variant' is a polypeptide that varies from the referent polypeptides by conservative amino acid substitutions, whereby a residue is substituted by another with the same physico-chemical properties. Typically such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln, and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. In one embodiment a 'fragment' is a polypeptide which comprises a contiguous portion of at least 250 amino acids of a polypeptide.

[045] 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.

[046] The term 'fragment' of a polypeptide or protein refers to a contiguous portion of at least 100, 200, 230, 250, 300, 350, 380, 400, 450, 480, 500, 530, 550, 580 or 600 amino acids from that polypeptide or protein. The term 'first fragment' refers to a contiguous portion of at least 100, 250, 300, 350, 380, 400, 450, 480, 500, 530, 550, 580 or 600 amino acids of the toxin A repeating domain. The term 'second fragment' refers to a contiguous portion of at least 100, 200, 230, 250, 280, 300, 350, 400, 450 or 500 amino acids of the toxin B repeating domain.

[047] The term 'first proximal end' refers to the end of the first fragment (Tox A fragment) which is covalently linked to the second fragment (ToxB fragment) or covalently linked to a linker sequence between the first and second fragment and is closest to the second fragment in primary structure. The term 'second proximal end' refers to the end of the second fragment which is covalently linked to the first fragment (ToxA fragment) or covalently linked to a linker sequence between the first and second fragment and is closest to the first fragment in primary structure.

[048] The polypeptide may be part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence which contains sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production. Furthermore, addition of exogenous polypeptide or lipid tail or polynucleotide sequences to increase the immunogenic potential of the final molecule is also considered.

[049] The fragments may be positioned such that the N-terminus of the first fragment is adjacent to the C-terminus of the second fragment, alternatively the C-terminus of the first fragment may be adjacent to the N-terminus of the second fragment, or the C-terminus of the first fragment may be adjacent to the C-terminus of the second fragment, or the N-terminus of the first fragment may be adjacent to the N-terminus of the second fragment.

[050] The word 'adjacent' means separated by less than or exactly 20, 15, 10, 8, 5, 2, 1 or 0 amino acids in the primary structure.

[051] The polypeptide of the invention elicits antibodies that neutralise toxin A or toxin B or both. In one embodiment the polypeptide elicits antibodies that neutralise toxin A. In a further embodiment the polypeptide elicits antibodies that neutralise toxin B. In a further embodiment the polypeptide elicits antibodies that neutralise toxin A and toxin B.

[052] Whether a polypeptide elicits antibodies against a toxin can be measured by immunising mice with an immunogenic composition comprising the polypeptide, 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. An example of this technique can be found in example 6. The polypeptide of the invention elicits antibodies that neutralise toxin A 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.

[053] In a further embodiment the polypeptide 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.

[054] Whether a polypeptide 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 polypeptide 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 polypeptide survives. A polypeptide elicits a protective immune response if a mammal immunised with the polypeptide 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*.

[055] The native structure of the C-terminal domains from toxins A and B consist of an extended β solenoid-like structure. This structure consists of primarily β sheet structures, with a minority of α helical structures as seen in Ho et al (PNAS 102:18373-18378 (2005)). The secondary structures present can be determined using circular dichroism. For example measuring the shape and the magnitude of the CD spectra in the far-UV region (190-250nm) and comparing the results with those of known structures. This can be carried out using an optical path of 0.01cm from 178 to 250nm, with a 1nm resolution and bandwidth on a Jasco J-720 spectropolarimeter, for example as seen in example 5 below.

[056] In one embodiment the first fragment comprises less than 25%, 23%, 20%, 18%, 15%, 10%, or 7% alpha helical secondary structure. In one embodiment the second fragment comprises less than 28%, 25%, 23%, 20%, 18%, 15%, 10%, or 7% alpha helical secondary structure. In a further embodiment both the first fragment and the second fragment comprise less than 28%, 25%, 23%, 20%, 18%, 15%, 10%, or 7% alpha helical secondary structure.

[057] In one embodiment the first fragment comprises more than 20%, 25%, 28%, 30%, 33%, 35%, 38%, 40%, or 42% beta sheet structure. In one embodiment the second fragment comprises more than 20%, 25%, 28%, 30%, 33%, 35%, 38%, 40%, or 42% beta sheet structure. In a further embodiment both the first fragment and the second fragment comprises more than 20%, 25%, 28%, 30%, 33%, 35%, 38%, 40%, or 42% beta sheet structure.

[058] Figure 2 depicts the organisation of the C-terminal domains of ToxA and ToxB. The C-terminal domain of toxin A is made up of 8 repeat portions (designated repeat portion I, repeat portion II, repeat portion III, repeat portion IV, repeat portion V, repeat portion VI, repeat portion VII and repeat portion VIII) each of these repeat portions can be further divided into short repeats (SRs) which are depicted as white boxes in figure 2 and long repeats (LRs) which are depicted as black boxes in figure 2 (except for Tox A repeat portion VIII which does not have a long repeat). Each of the long repeats has some structural and sequence similarity to the other long repeats.

Similarly the short repeats have some sequence and structural similarity to one another. The C-terminal domain of toxin B is made up of 5 repeat portions subdivided into SRs and LRs. Each repeat portion contains one LR and between 2 and 5 SRs (except for Tox B repeat portion V which does not have a long repeat). For the purposes of the disclosure the phrase 'a repeat portion' refers to one of the eight repeat portions of ToxA (designated I, II, III, IV, V, VI, VII, and VIII) or one of the five repeat portions of ToxB (designated I, II, III, IV or VI). As used herein the term 'first repeat portion' refers to a repeat portion (or partial repeat portion) from the toxin A repeating domain. The term 'second repeat portion' refers to a repeat portion (or partial repeat portion) from the toxin B repeating domain . For the purposes of the disclosure the term 'long repeat' refers to one of the LR domains depicted as black boxes in Figure 2. For the purposes of the disclosure the term 'short repeat' refers to one of the SR domains depicted as white boxes in Figure 2.

[059] Thus for example, repeat portion I of ToxA contains three SRs and one LR, which can be referred to as the first SRI of ToxA, the second SRI of ToxA, the third SRI of ToxA and the LRI of ToxA, respectively.

[060] The first proximal end is considered to be within a 'repeat portion' if the first fragment ends in an amino acid that is within that repeat portion (i.e., the first proximal end contains only part of the repeat portion sequence). Similarly the second proximal end is considered to be within a 'repeat portion' if the second fragment ends in an amino acid that is within that repeat portion. For example the first proximal end is within 'a repeat portion I of ToxA' if the first fragment ends with any one of amino acids 1832-1924 (inclusive) of VPI10463 or their equivalent in another strain. The first proximal end is not within a short repeat-long repeat-short repeat portion if the first fragment ends with an amino acid that is not within that short-repeat-long repeat-short repeat portion.

[061] The amino acid positions of each domain has been defined for toxin A and toxin B from strain VPI10463 (ATCC43255). These are as follows

Name		Start position	End position
ToxA_I	SR1	1832	1852
	SR2	1853	1873
	SR3	1874	1893
	LR	1894	1924
ToxA_II	SR1	1925	1944
	SR2	1945	1965
	SR3	1966	1986
	SR4	1987	2007
	SR5	2008	2027
	LR	2028	2058
ToxA_III	SR1	2059	2078

	SR2	2079	2099
	SR3	2100	2120
	SR4	2121	2141
	SR5	2142	2161
	LR	2162	2192
ToxA_IV	SR1	2193	2212
	SR2	2213	2233
	SR3	2234	2253
	SR4	2254	2275
	LR	2276	2306
ToxA_V	SR1	2307	2326
	SR2	2327	2347
	SR3	2348	2368
	SR4	2369	2389
	SR5	2390	2409
	LR	2410	2440
ToxA_VI	SR1	2441	2460
	SR2	2461	2481
	SR3	2482	2502
	SR4	2503	2522
	LR	2523	2553
ToxA_VII	SR1	2554	2573
	SR2	2574	2594
	SR3	2595	2613
	LR	2614	2644
ToxA_VIII	SR1	2645	2664
	SR2	2665	2686

	SR3	2687	2710
ToxB_I	SR1	1834	1854
	SR2	1855	1876
	SR3	1877	1896
	LR	1897	1926
ToxB_II	SR1	1927	1946
	SR2	1947	1967
	SR3	1968	1987
	SR4	1988	2007
	SR5	2008	2027
	LR	2028	2057
ToxB_III	SR1	2058	2078
	SR2	2079	2099
	SR3	2100	2119
	SR4	2120	2139
	SR5	2140	2159
	LR	2160	2189
ToxB_IV	SR1	2190	2212
	SR2	2213	2233
	SR3	2234	2253
	SR4	2254	2273
	SR5	2274	2293
	LR	2294	2323
ToxB_V	SR1	2324	2343
	SR2	2344	2366

Table 1

[062] For this reason the term 'repeat portion' may refer to amino acids 1832-1924, 1925-2058, 2059-2192, 2193-2306, 2307-2440, 2441-2553, 2554-2644 or 2645-2710 of toxin A (SEQ

ID NO:1), or amino acids 1834-1926, 1927-2057, 2058-2189, 2190-2323 or 2324-2366 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*.

[063] For this reason the term 'short repeat' may refer to amino acids 1832-1852, 1853-1873, 1874-1893, 1925-1944 1945-1965, 1966-1986, 1987-2007, 2008-2027, 2059-2078, 2079-2099, 2100-2120, 2121-2141, 2142-2161, 2193-2212, 2213-2233, 2234-2253, 2254-2275, 2307-2326, 2327-2347, 2348-2368, 2369-2389, 2390-2409, 2441-2460, 2461-2481, 2482-2502, 2503-2522, 2554-2573, 2574-2594, 2595-2613, 2645-2664, 2665-2686 or 2687-2710 of toxin A (SEQ ID NO:1) or amino acids 1834-1854, 1855-1876, 1877-1896, 1927-1946, 1947-1967, 1968-1987, 1988-2007, 2008-2027, 2058-2078, 2079-2099, 2100-2119, 2120-2139, 2140-2159, 2190-2212, 2213-2233, 2234-2253, 2254-2273, 2274-2293, 2324-2343 or 2344-2366 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*.

[064] Similarly the term 'long repeat' may refer to amino acids 1894-1924, 2028-2058, 2162-2192, 2276-2306, 2410-2440, 2523-2553 or 2614-2644 of toxin A (SEQ ID NO:1) or amino acids 1897-1926, 2028-2057, 2160-2189 or 2294-2323 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*.

[065] Similarly the term 'short repeat-long repeat-short repeat portion' may refer to amino acids 1874-1944, 2008-2078, 2142-2212, 2254-2326, 2390-2460, 2503-2573, or 2595-2664 of toxin A (SEQ ID NO:1) or amino acids 1877-1946, 2008-2078, 2140-2212 or 2274-2343 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*. The term 'does not disrupt a short repeat-long repeat-short repeat portion' means that the proximal end is in a region which does not disrupt the structure of the short-repeat-long repeat-short repeat portion, in general this means that the proximal end is not within a long repeat and not within the short repeats making up a short repeat-long repeat-short repeat portion, except the proximal end may be in the region of 1, 2, 3, 4, 5 or 6 amino acids of the short repeat which are furthest away from the long repeat in sequence. In an embodiment the term 'does not disrupt a short-repeat-long repeat-short repeat portion' means that the proximal end is not within the short repeat-long repeat-short repeat portion.

[066] In one embodiment the first proximal end is within a short repeat. In one embodiment the second proximal end is within a short repeat. In one embodiment the first proximal end and the second proximal end are within a short repeat. In one embodiment the first proximal end does not disrupt a short-repeat long repeat-short repeat portion. In one embodiment the second proximal end does not disrupt a short-repeat long repeat-short repeat portion. In one embodiment the first proximal end and the second proximal end do not disrupt a short-repeat long repeat-short repeat portion.

[067] In one embodiment the first proximal end is not within amino acids 1878-1940, 2146-2208, 2012-2074, 2258-2322, 2394-2456, 2507-2569, 2599-2660 or 2593-2660 of toxin A (SEQ ID NO:1) or their equivalents in a different strain of *C.difficile*. In a second embodiment the second proximal end is not within amino acids 1881-1942, 2012-2074, 2144-2208, or 2278-2339 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*. In a further embodiment the first proximal end is not within amino acids 1878-1940, 2146-2208, 2012-2074, 2258-2322, 2394-2456, 2507-2569, 2599-2660 or 2593-2660 of toxin A (SEQ ID NO:1) or their equivalents in a different strain of *C.difficile* and the second proximal end is not within amino acids 1881-1942, 2012-2074, 2144-2208, or 2278-2339 of toxin B (SEQ ID NO:2) or their equivalents in a different strain of *C.difficile*.

[068] In one embodiment the first proximal end is within repeat portion V (amino acids 2307-2440 of SEQ ID NO:1 or their equivalent in a different strain), VI (amino acids 2441-2553 of SEQ ID NO:1 or their equivalent in a different strain), VII (amino acids 2554-2644 of SEQ ID NO:1or their equivalent in a different strain) or VIII (amino acids 2645-2710 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A. In a further embodiment the first proximal end is within repeat portion VII (amino acids 2554-2644 of SEQ ID NO:1 or their equivalent in a different strain)

of toxin A. In a further embodiment the first proximal end is within repeat portion VIII (amino acids 2645-2710 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A.

[069] In one embodiment the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain), II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain), or iii (amino acids 2058-2189 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the second proximal end is within repeat portion II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B.

[070] In one embodiment the first proximal end is within repeat portion VIII (amino acids 2645-2710 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A and the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain)of toxin B. In a further embodiment the first proximal end is within repeat portion VIII (amino acids 2645-2710 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A and the second proximal end is within repeat portion II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain)of toxin B. In a further embodiment the first proximal end is within repeat portion VII (amino acids 2554-2644 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A and the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the first proximal end is within repeat portion VII (amino acids 2554-2644 of SEQ ID NO:1 or their equivalent in a different strain) of toxin A and the second proximal end is within repeat portion II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the first proximal end is within repeat portion VI amino acids 2441-2553 of SEQ ID NO:1 or their equivalent in a different strain) and the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the first proximal end is within repeat portion VI amino acids 2441-2553 of SEQ ID NO:1 or their equivalent in a different strain) and the second proximal end is within repeat portion II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the first proximal end is within repeat portion V (amino acids 2307-2440 of SEQ ID NO:1 or their equivalent in a different strain) and the second proximal end is within repeat portion I (amino acids 1834-1926 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B. In a further embodiment the first proximal end is within repeat portion V (amino acids 2307-2440 of SEQ ID NO:1 or their equivalent in a different strain) and the second proximal end is within repeat portion II (amino acids 1927-2057 of SEQ ID NO:2 or their equivalent in a different strain) of toxin B.

[071] In one embodiment the first proximal end is within amino acids 2690-2710, or 2695-2710, or 2700-2710 of SEQ ID NO:1 or their equivalent in a different strain. In a further embodiment the first proximal end is within amino acids 2670-2700, or 2675-2695, or 2680-2690 of SEQ ID NO:1 or their equivalent in a different strain. In one embodiment the second proximal end is within amino acids 1860-1878 of Toxin B or their equivalents in a different strain. In one embodiment the second proximal end is within amino acids 1950-1980, 1955-1975 or 1960-1970 of SEQ ID NO:2 or their equivalent in a different strain. In a further embodiment the second proximal end is within amino acids 1978-2008, 1983-2003 or 1988-1998 of SEQ ID NO:2 or their equivalent in a different strain. In a further embodiment the second proximal end is within amino acids 1860-1878, 1854-1876, 1857-1887, 1862-1882, or 1867-1877 of SEQ ID NO:2 or their equivalent in a different strain.

[072] In one embodiment the first fragment consists of an entire toxin A repeating domain (amino acids 1832-2710). In one embodiment the second fragment consists of an entire toxin B repeating domain (amino acids 1833-2366).

[073] In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1968-1987 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion II of toxin B (amino acids 1947-1967 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion I of toxin B (amino acids 1877-1896 of SEQ ID NO:2 or their equivalents in a different strain. In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion I of toxin B (amino acids 1855-1876 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VIII of toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VIII of toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1968-1987 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VIII of toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion II of toxin B (amino acids 1947-1967 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VIII or toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion I of Toxin B (amino acids 1877-1896 of toxin B or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portionVIII of toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion I of toxin B (amino acids 1855-1876 of SEQ ID NO:2 or their equivalents in a different strain. In one embodiment the first proximal end is within short repeat 2 of repeat portion VIII of toxin A (amino acids 2665-2686 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion vii of toxin A (amino acids 2574-2594 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion vii of toxin A (amino acids 2574-2594 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1668-1987 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion vii of toxin A (amino acids 2574-2594 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion II of toxin B (amino acids 1947-1967 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VII of toxin A (amino acids 2574-2594 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion I of toxin B (amino acids 1855-1876 of SEQ

ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VII of toxin A (amino acids 2574-2594 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VI of toxin A (amino acids 2482-2502 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VI of toxin A (amino acids 2482-2502 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1968-1987 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VI of toxin A (amino acids 2482-2502 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion II of toxin B (amino acids 1947-1967 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VI of toxin A (amino acids 2482-2502 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion I of toxin B (amino acids 1855-1876 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 3 of repeat portion VI of toxin A (amino acids 2482-2502 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VI of toxin A (amino acids 2461-2481 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VI of toxin A (amino acids 2461-2481 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1968-1987 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VI of toxin A (amino acids 2461-2481 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion II of toxin B (amino acids 1947-1967 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VI of toxin A (amino acids 2461-2481 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 2 of repeat portion I of toxin B (amino acids 1855-1876 of SEQ ID NO:2 or their equivalents in a different strain). In one embodiment the first proximal end is within short repeat 2 of repeat portion VI of toxin A (amino acids 2461-2481 of SEQ ID NO:1 or their equivalents in a different strain) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854 of SEQ ID NO:2 or their equivalents in a different strain).

[074] In one embodiment the first proximal end is within amino acids 2690-2710, or 2695-2710, or 2700-2710 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within amino acids 1950-1980, 1955-1975 or 1960-1970 of SEQ ID NO:2 or their equivalent in a different strain. In one embodiment the first proximal end is within amino acids 2690-2710, or 2695-2710, or 2700-2710 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within amino acids 1978-2008, 1983-2003 or 1988-1998 of SEQ ID NO:2 or their equivalent in a different strain. In one embodiment the first proximal end is within amino acids 2690-2710, or 2695-2710, or 2700-2710 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within amino acids 1857-1887, 1862-1882, or 1867-1877 of SEQ ID NO:2 or their equivalent in a different strain. In one embodiment the first proximal end is within amino acids 2670-2700, or 2675-2695, or 2680-2690 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within amino acids 1950-1980, 1955-1975 or 1960-1970 of SEQ ID NO:2 or their equivalent in a different strain. In one embodiment the first proximal end is within amino acids 2670-2700, or 2675-2695, or 2680-2690 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within

amino acids 1978-2008, 1983-2003 or 1988-1998 of SEQ ID NO:2 or their equivalent in a different strain. In one embodiment the first proximal end is within amino acids 2670-2700, or 2675-2695, or 2680-2690 of SEQ ID NO:1 or their equivalent in a different strain and the second proximal end is within amino acids 1857-1887, 1862-1882, 1860-1878 or 1867-1877 of SEQ ID NO:2 or their equivalent in a different strain.

[075] In one embodiment the first fragment comprises at least 100, 200, 300, 400 or 450 amino acids. In one embodiment the second fragment comprises at least 100, 200, 300 or 400 amino acids.

[076] In one embodiment the polypeptide further comprises a linker. This linker may be between the first proximal end and the second proximal end, alternatively the linker may link the distal ends of the first fragment and/or the second fragment to a further sequence of amino acids.

[077] A peptide linker sequence may be employed to separate the first fragment and second fragment . 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 fragment and/or the second fragments; and (3) the lack of hydrophobic or charged residues that might react with the Tox A and/or ToxB functional epitopes. Peptide linker sequences may 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.

[078] In one embodiment the linker comprises between 1-19, 1-15, 1-10, 1-5, 1-2, 5-20, 5-15, 5-15, 10-20, or 10-15 amino acids. In one embodiment the linker is a glycine linker, the linker may comprise multiple (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 18 or 19) contiguous glycine residues, or alternatively the linker may comprise some glycine residues and some residues of other amino acids such as alanine. In a further embodiment the linker comprises a single glycine residue.

[079] In an embodiment the polypeptide of the invention is part of a larger fusion protein. The fusion proteins may further comprise amino acids encoding an immunogenic portion of a further protein antigen. For example the fusion protein may further comprise an immunogenic portion of a protein antigen obtained or derived from a bacterium selected from the group consisting of *S.pneumoniae*, *H.influenzae*, *N.meningitidis*, *E.coli*, *M.cattarhalis*, *C.tentani*, *C.diphtheriae*, *B.pertussis*, *S.epidermidis*, enterococci, *S.aureus*, and *Pseudomonas aeruginosa*. In this case the linker may be between the first fragment or the second fragment and a further immunogenic portion of a protein antigen.

[080] The term "immunogenic portion thereof" or 'immunogenic fragment' refers to a fragment of a polypeptide wherein the fragment comprises an epitope that is recognized by cytotoxic T lymphocytes, helper T lymphocytes or B cells. Suitably, the immunogenic portion will comprise at least 30%, suitably at least 50%, especially at least 75% and in particular at least 90% (e.g. 95% or 98%) of the amino acids in the reference sequence. The immunogenic portion will suitably comprise all of the epitope regions of the reference sequence.

[081] In one embodiment the polypeptide comprises an immunogenic fragment of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25 or SEQ ID NO27. In one embodiment the polypeptides comprises an immunogenic fragment of at least 500, 550, 600, 650, 700, 750, 780, 800, 830, 850, 880, 900, 920, or 950 amino acids of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25 or SEQ ID NO27. In a further embodiment the polypeptide comprises a variant of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25 or SEQ ID NO27, in a further embodiment the polypeptide comprises a variant having at least 80%, 85%, 90%, 92%, 95%, 98%, 99%, or 100% sequence identity to SEQ ID NO:3-SEQ ID NO:7.

[082] In one embodiment the polypeptide comprises more than 450, 475, 500, 525, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825 or 850 amino acids from toxin A. In one embodiment the polypeptide comprises less than 850, 825, 800, 775, 750, 725, 700, 675, 650, 625, or 600 amino acids from toxin A. In one embodiment the polypeptide comprises more than 350, 375, 400, 425, 450, 475, 500 or 525 amino acids from toxin B. In one embodiment the polypeptide comprises less than 525, 500, 475, or 450 amino acids from toxin B.

[083] The term 'identity' is known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heine, G., Academic Press, 1987; andSequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the Needle program BLASTP, BLASTN (Altschul, S.F. et al., J. Molec. Biol. 215: 403-410 (1990), and FASTA(Pearson and Lipman Proc. Natl. Acad. Sci. USA 85; 2444-2448 (1988). The BLAST family of programs is publicly available from NCBI and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

[084] Parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Henikoff and Henikoff,

Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 10

Gap extension penalty:0.5

A program useful with these parameters is publicly available as the 'needle' program from EMBOSS package (Rice P.et al, Trends in Genetics 2000 col.16(6):276-277). The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

[085] In order to determine the identity of a reference sequence to SEQ ID NO:1, in one embodiment the sequence identity is calculated over the entire length of the reference sequence. In a further embodiment the sequence identity is calculated over the entire length of the sequence in SEQ ID NO:1. In order to determine the identity of a reference sequence to SEQ ID NO:2, in one embodiment the sequence identity is calculated over the entire length of the reference sequence. In a further embodiment the sequence identity is calculated over the entire length of the sequence in SEQ ID NO:2.

[086] In a further aspect of the invention there is provided a polypeptide comprising (i) SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35, (ii) a variant having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99% or 100% identity to SEQ ID NO:10-19; or (iii) a fragment of at least 100, 200, 230, 250, 300, 350, 380, 400, 450, 480, 500, 530, 550, 580 or 600 amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID

NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 orSEQ ID NO:35. In a further embodiment the polypeptide comprises SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35, ii) a variant having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99% or 100% identity to SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35; or (iii) a fragment of at least 100, 200, 230, 250, 300, 350, 380, 400, 450, 480, 500, 530, 550, 580 or 600 amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35. In a further embodiment the polypeptide comprises SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 orSEQ ID NO:35, ii) a variant having at least 80%, 85%, 88%, 90%, 92%, 95%, 98%, 99% or 100% identity to SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 orSEQ ID NO:35; or (iii) a fragment of at least 100, 200, 230, 250, 300, 350, 380, 400, 450, 480, 500, 530, 550, 580 or 600 amino acids of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33 or SEQ ID NO:34 or SEQ ID NO:35.

[087] In one embodiment the polypeptide comprises more than 450, 475, 500, 525, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825 or 850 amino acids from toxin A. In one embodiment the polypeptide comprises less than 850, 825, 800, 775, 750, 725, 700, 675, 650, 625, or 600 amino acids from toxin A. In one embodiment the polypeptide comprises more than 350, 375, 400, 425, 450, 475, 500 or 525 amino acids from toxin B. In one embodiment the polypeptide comprises less than 525, 500, 475, or 450 amino acids from toxin B.

[088] In a further embodiment the polypeptide elicits neutralising antibodies that neutralise toxin A or toxin B or both. In a further embodiment the polypeptide elicits antibodies that neutralise toxin A. In a further embodiment the polypeptide elicits antibodies that neutralise toxin B. In a further embodiment the polypeptide elicits antibodies that neutralise toxin A and toxin B. The polypeptide of the invention elicits antibodies that neutralise toxin A 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.

[089] In a further embodiment the polypeptide 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, monkey, non-human primate and human. In one embodiment the mammalian host is a mouse. In a further embodiment the mammalian host is a human.

[090] Whether a polypeptide 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 polypeptide 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 polypeptide survives. A polypeptide elicits a protective immune response if a mammal immunised with the polypeptide survives at least 10%, 20%, 30%, 50%, 70%, 80%, 90%, or 100% longer than a reference mammal which has not been

immunised after challenge with *C.difficile*. In one embodiment the polypeptide of the invention elicits a protective immune response against strains of *C.difficile* in a mammal selected from the group consisting of mouse, guinea pig, monkey and human. In one embodiment the mammal is a mouse, in a further embodiment the mammal is a human.

[091] The native structure of the C-terminal (repeat) domains from toxins A and B consist of an extended β solenoid-like structure. This structure consists of primarily β sheet structures, with a minority of α helical structures as seen in Ho et al (PNAS 102:18373-18378 (2005)). The secondary structures present can be determined using circular dichroism. For example measuring the shape and the magnitude of the CD spectra in the far-UV region (190-250nm) and comparing the results with those of known structures. This can be carried out using an optical path of 0.01cm from 178 to 250nm, with a 1nm resolution and bandwidth on a Jasco J-720 spectropolarimeter, for example as seen in example 5 below.

[092] In one embodiment the polypeptide comprises less than 25%, 23%, 20%, 28%, 15%, 10%, or 7% alpha helical secondary structure. In a further embodiment the polypeptide comprises more than 20%, 25%, 28%, 30%, 33%, 35%, 38%, 40%, or 42% beta sheet structure.

POLYNUCLEOTIDES

[093] The invention further provides a polynucleotide encoding a polypeptide of the invention. For the purposes of the invention the term 'polynucleotide(s)' generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA including single and double-stranded regions/forms.

[094] The term "polynucleotide encoding a peptide" as used herein encompasses polynucleotides that include a sequence encoding a peptide or polypeptide of the invention. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the peptide or polypeptide (for example, polynucleotides interrupted by integrated phage, an integrated insertion sequence, an integrated vector sequence, an integrated transposon sequence, or due to RNA editing or genomic DNA reorganization) together with additional regions, that also may contain coding and/or non-coding sequences.

[095] It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal similarity to the nucleotide sequence of any *native* (i.e. naturally occuring) gene. Nonetheless, polynucleotides that *vary* due to differences in codon usage are specifically contemplated by the present invention, for example polynucleotides that are optimized for human and/or primate and/or *E.coli* codon selection.

[096] Sequences encoding a desired polypeptide may be synthesized, 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 ASI 431 A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

[097] Moreover, the polynucleotide sequences of the present invention 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.

VECTORS

[098] In a further aspect of the invention the present invention relates vector comprising a polynucleotide of the invention linked to an inducible promoter such that when the promoter is induced a polypeptide encoded by the polynucleotide is expressed.

[099] A further aspect of the invention comprises said vector wherein the inducible promoter is activated by addition of a sufficient quantity of IPTG (Isopropyl β-D-1-thiogalactopyranoside) preferably to the growth medium. Optionally this is at a concentration of between 0.1 and 10mM, 0.1 and 5mM, 0.1 and 2.5mM, 0.2 and 10mM, 0.2 and 5mM, 0.2 and 2.5mM, 0.4 and 10mM, 1 and 10mM, 1 and 5mM, 2.5 and 10mM, 2.5 and 5mM, 5 and 10mM. Alternatively the promoter may be induced by a change in temperature or pH.

HOST CELLS

[0100] For recombinant production of the polypeptides of the invention, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis, et al., BASIC METHODS IN MOLECULAR BIOLOGY, (1986) and Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, conjugation, transduction, scrape loading, ballistic introduction and infection.

[0101] Representative examples of appropriate hosts include gram negative bacterial cells, such as cells of, *E. coli*, Acinetobacter, Actinobacillus, Bordetella, Brucella, Campylobacter, Cyanobacteria, Enterobacter, Erwinia, Franciscella, Helicobacter, hemophilus, Klebsiella, Legionella, Moraxella, Neisseria, Pasteurella, Proteus, Pseudomonas, Salmonella, Serratia, Shigella, Treponema, Vibrio, Yersinia. In one embodiment the host cell is an Escherichia coli cell. Alternatively gram positive bacterial cells may also be used. A great variety of expression systems can be used to produce the polypeptides of the invention. In one embodiment the vector is derived from bacterial plasmids. Generally any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, (supra).

IMMUNOGENIC COMPOSITIONS AND VACCINES

[0102] There is further provided an immunogenic composition comprising a polypeptide of the invention and a pharmaceutically acceptable excipient.

[0103] In one embodiment the immunogenic composition further comprises an adjuvant. The choice of a suitable adjuvant to be mixed with bacterial toxins or conjugates made using the processes of the invention is within the knowledge of the person skilled in the art. Suitable adjuvants include an aluminium salt such as aluminium hydroxide gel or aluminum phosphate or alum, but may also be other metal salts such as those of calcium, magnesium, iron or zinc, or

may be an insoluble suspension of acylated tyrosine, or acylated sugars, cationically or anionically derivatized saccharides, or polyphosphazenes.

[0104] 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.cattarhalis*, tetanus, diphtheria, 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).

[0105] There is further provided a vaccine comprising the immunogenic composition, this vaccine may further comprise a pharmaceutically acceptable excipient.

[0106] 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.

[0107] 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.

[0108] 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.

[0109] 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.

[0110] A further aspect of the invention is an immunogenic composition of the invention for use in the treatment or prevention of *C.difficile* disease. In one embodiment there is provided an immunogenic composition of the invention for use in the treatment or prevention of primary and/or recurrence episodes of *C.difficile* disease.

[0111] A further aspect of the invention is use of the immunogenic composition or vaccine or kit of the invention in the manufacture of a medicament for the treatment or prevention of *C.difficile* disease. In one embodiment there is provided an immunogenic composition of the invention for

use in the manufacture of a medicament for the treatment or prevention of primary and/or recurrence episodes of *C.difficile* disease.

[0112] Around" or "approximately" are defined as within 10% more or less of the given figure for the purposes of the invention.

[0113] 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."

[0114] Embodiments herein relating to "vaccine compositions" of the invention are also applicable to embodiments relating to "immunogenic compositions" of the invention, and vice versa.

[0115] 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).

[0116] 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.

[0117] All references or patent applications cited within this patent specification are incorporated by reference herein in their entirety.

[0118] 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

Example 1: Design of five *C.difficile* ToxA-ToxB fusions

[0119] Fusion proteins containing fragments of the C-terminal repeating domains of ToxA and ToxB were designed. These fusions contained a fragment of the C-terminal repeating domain of ToxA and a fragment of the C-terminal repeating domain of ToxB, and a junction between the C-terminal end of the ToxA fragment and the N terminal end of the ToxB fragment. Two strategies were devised, in the first strategy; the fusion was designed such that the long solenoid structure was maintained at the junction between the two fragments. In the second strategy, the two fragments of the fusions are separated by a linker to allow their independent correct folding.

[0120] The C-terminal part of ToxA and B is composed of repeated sequences: short repeats (SR) and long repeats (LR) (PNAS 2005 vol 102 : 18373-18378).

[0121] The partial known 3D structure for the C-terminal domain of ToxA (PNAS 2005 Greco et al., vol 102 : 18373-18378 ; Nature Structural & Molecular biology 2006 vol 13(5) : 460-461 ; PDB codes : 2F6E, 2G7C and 2QJ6).

[0122] The inventors predicted that there are two kinds of important interactions between residues of the C-terminal part of ToxA and ToxB. The first interaction is occurring between residues contained in a LR and its preceding SR and is important to maintain the solenoid-like structure. The second type of interaction occurs between residues contained in a LR and the following SR and this interaction is mediating the carbohydrate-binding function of the toxin.

[0123] A new "structural-functional" repeat SR-LR-SR was defined. The structure of this repeat was maintained intact in our designed fusions.

[0124] Figure 2 represents the C-terminal domains of ToxA and ToxB and the defined "SR-LR-SR" box.

[0125] The positions of the short (SR) and long repeats (LR) of ToxA and ToxB repeats are presented in table 1.

[0126] A list of the "SR-LR-SR" boxes contained in the C-terminal domain of ToxA and ToxB is presented in Table 2.

Name	Start position	End position
ToxA_1	1874	1944
ToxA_2	2008	2078
ToxA_3	2142	2212
ToxA_4	2254	2326
ToxA_5	2390	2460
ToxA_6	2503	2573
ToxA_7	2595	2664
ToxB_1	1877	1946
ToxB_2	2008	2078
ToxB_3	2140	2212
ToxB_4	2274	2343

Table 2

[0127] Finally, the number of SRs between two LRs will be maintained in the designed fusions to keep the long solenoid-like structure.

[0128] Before the design of junctions for the fusions, two working hypotheses were defined: first hypothesis, the shorter the fusions, the better the probability for the fusions to be stably over expressed; second hypothesis, according to the concept of "SR-LR-SR" boxes, the start position has to be chosen in order to ensure a correct folding of the first SR of this previously defined SR-LR-SR box. Thus the fusions start at the beginning of the SR that precedes the SR-LR-SR box. Using these two hypothesis, three start positions were analysed: residue 2370, 2234 and 2121 of ToxA.

[0129] The start position 2370 was excluded. The start position 2234 was also excluded because one of the residues involved in interactions important for the protein structural stability is not conserved. So, it was decided that all the designed fusion will begin at residue 2121 of ToxA.

[0130] All fusions will end at the last residue of ToxB.

[0131] Four fusions (F1-4) were designed in order to maintain the entire fusion in a long solenoid-like structure between the two fusion fragments.

[0132] The fusions 1 (F1) and 2 (F2) were designed using the same hypothesis. All SR protein sequences of ToxA and ToxB had been compared using a multiple alignment software (ClustalW - Thompson JD *et al.* (1994) *Nucleic Acids Res.*, 22, 4673-4680). The more similar sequences were the third SR VIII of ToxA and the third SR II of ToxB and third SR III of ToxB. In order to make a choice between these two SR of ToxB, a structural homology modelling (using the SwissModel interface - Arnold K *et al.* (2006) *Bioinformatics*, 22, 195-201) was performed on the C-terminal part of ToxB using the

known 3D structure of partial ToxA C-terminal domain (PDB code : 2QJ6). Using the third SR VIII of ToxA, the best local structural superposition (performed using SwissPDBViewer - Guex N *et al.* (1997), *Electrophoresis* 18, 2714-2723) was obtained with the third SR II of ToxB. So, two junctions were designed : the first one is between the third SR VIII of ToxA and the fourth SR II of ToxB (F1) and the second one is between the second SR VIII of ToxA and the third SR II of ToxB (F2). These junctions are presented in figure 3 and 4 respectively.

[0133] To design the fusion 3 (F3), a global structural superposition was performed between both the known structure of the partial C-terminal domain of ToxA and the predicted structure of C-terminal domain of ToxB (using SwissModel and SwissPDBViewer softwares). The best superposition was found between LR VII of ToxA and LR II of ToxB. So, it was decided to make a junction in this similar LR. The junction was performed firstly in a region where the sequence is conserved between ToxA and ToxB, after that in order to keep in the ToxA part of the fusion, the residues in interaction with the preceding SR and lastly, in order to keep in the ToxB part, the residues in interaction with the following SR. This junction is shown in figure 5.

[0134] For the design of fusion 4 (F4), the C-terminal domain of ToxB was divided in 4 fragments and a more precise homology modelling (SwissModel) was performed on them. The split was realised in order to keep intact the "SR-LR-SR" boxes (each domain finishes at the end of the SR that follows a LR). A structural superposition between the predicted structures of these fragment and the known 3D structure of ToxA was made and the best structural surperposition was obtained for the third SR of ToxB (SR I) and the last SR of ToxA (third SR VIII). So, the junction was done between the second SR VIII of ToxA and the third SRIof ToxB. This design is presented in figure 6.

[0135] The last fusion (F5) was designed in order to allow an independent correct folding of the two fragments of the fusion. The linker was added between the last residue of the ToxA protein sequence and the beginning of the fourth SR II of ToxB (always taking into account the importance of an intact "SR-LR-SR" box). Only one exogenous residue (Glycine) was added as linker and located between two existing Glycines. Thus, the linker can also be described as composed of 3 Glycines surrounding by known (for ToxA) and predicted (for ToxB) beta-strand. This last design is shown in figure 7.

Example 2: Cloning expression and purification of the fusion proteins

Expression plasmid and recombinant strain

[0136] Genes encoding the fusion proteins of partial C-terminal domains of ToxA and ToxB (SEQ ID NO:3, 4, 5, 6 and 7) and a His tag were 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* strain BLR (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.).

Host strain:

[0137] BLR(DE3). 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.

[0138] Genotype : *E.coli* BLR::DE3 strain, F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (*DE3*) Δ(srl-recA)306::Tn*10* (Tet^R)

Expression of the recombinant proteins:

[0139] An *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.

[0140] 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.

[0141] 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.

[0142] 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:

[0143] The bacterial pellet was resuspended in 20 mM bicine buffer (pH 8.0) containing 500 mM NaCl and a mixture of protease inhibitor (Complete, Roche). Bacteria were lysed using a French Press system 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation for example at 20 000g for 30 min at 4°C.

[0144] The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a GE column (15 ml for example) (Ni loaded) preequilibrated with the same buffer used to bacterial resuspension. After loading on the column, the column was washed with the same buffer. Elution was performed using a 20mM bicine buffer (pH 8.0) containing 500 mM NaCl and different concentrations of imidazole (5-600 mM). After gel analysis, more pure fractions were selected, concentrated and loaded on SEC chromatography for further purification step.

[0145] Fractions containing the fusion proteins were selected on the basis of purity by SDS-PAGE and dialyzed against bicine buffer (20mM Bicine, 150 mM NaCl,with or without 5mM EDTA pH8.0), Protein concentration was determined using DC Protein Assay of BioRad. Proteins were thus pooled, sterile-filtered on 0.22 µm, stored at -80°C.

[0146] Alternatively, IMAC purification was preceded by a DEAE purification step using 2mM bicine buffer (pH 8.0) for loading and washing, and eluted using a gradient with the same buffer but with 1M NaCl added.

Example 3 – Cloning expression and purification of the separate *C.difficile* Tox A and Tox B fragments

Expression plasmid and recombinant strain.

[0147] Genes encoding the protein fragments of ToxA and ToxB (SEQ ID NO:8 and SEQ ID NO:9) and a His tag were 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* strain BLR (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.).

Host strain:

[0148] BLR(DE3). 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.

[0149] Genotype : *E.coli* BLR::DE3 strain, F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (*DE3*) Δ(srl-recA)306::Tn*10* (Tet^R)

Expression of the recombinant proteins:

[0150] 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.

[0151] 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.

[0152] 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.

[0153] 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:

[0154] The bacterial pellet was resuspended in 20 mM bicine buffer (pH 8.0) containing 500 mM NaCl supplemented by a mixture of protease inhibitor (Complete without EDTA, Roche cat 11873580001) and benzonase. (Roche cat 1.01695.0001). Bacteria were lysed using a French Press system 2 X 20 000 PSI. Soluble (supernatant) and insoluble (pellet) components were separated by centrifugation at 34 000g or 48 000g for 25-30 min at 4°C. Supernatant was harvested and filtrated on 0.22 µm filter.

[0155] The 6-His tagged-protein was purified under native conditions on IMAC. The soluble components were loaded on a GE column (for example 15ml) (Ni loaded) pre-equilibrated with the same buffer used to bacterial resuspension. After loading, the column was washed with the same buffer.

For ToxA:

[0156] Elution was performed using a 20mM bicine buffer (pH 8.0) containing 500 mM NaCl and different concentrations of imidazole (5-100 mM). After gel analysis, more pure fractions were selected, concentrated and loaded on SEC chromatography (SUPERDEX^{TM} 75) for further purification step in the same buffer without imidazole.

For ToxB:

[0157] A second wash was performed with 20mM bicine buffer (pH 8.0) containing 500 mM NaCl and 0.5 % deoxycholate or same buffer with 150 mM NaCl. Elution was performed using a 20mM bicine buffer (pH 8.0) containing 500 mM NaCl and different concentrations of imidazole (10-500 mM). After gel analysis, more pure fractions were selected, supplemented with 5 mM EDTA and loaded on SEC chromatography (SUPERDEX^{TM} 200) for further purification step in same buffer with 5 mM EDTA.

[0158] Fractions containing ToxA or ToxB fragments were selected on the basis of purity by SDS-PAGE and dialyzed against bicine buffer (20mM Bicine, 150 mM NaCl, pH8.0), protein concentration was determined using RCDC Protein Assay of BioRad. Proteins were thus pooled, sterile-filtered on 0.22 µm, stored at -80°C.

Example 4 – Molecular weight evaluation of the five *C.difficile* ToxA-ToxB fusions

[0159] Analytical ultracentrifugation is 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 42 000RPM after the AN-60Ti rotor had been equilibrated to 15°C.
    a. F1 fusion protein, 500µg/ml, 20mM Bicine, 150mM NaCl, pH8,0
    b. F2 fusion protein, 500µg/ml, 20mM Bicine, 150mM NaCl, pH8,0
    c. F3 fusion protein, 500µg/ml, 20mM Bicine, 150mM NaCl, pH8,0
    d. F4 fusion protein, 500µg/ml, 20mM Bicine, 150mM NaCl, pH8,0
    e. F5 fusion protein, 500µg/ml, 20mM Bicine, 150mM NaCl, pH8,0
2. For data collection, 160 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. The molecular weight of the different species was 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.

[0160] Figure 8 describes the distribution of the ToxA-ToxB fusions as determined by sedimentation velocity analytical ultracentrifugation.

[0161] The molecular weight of the major species detected from the C(S) distribution of all five ToxA-ToxB fusion proteins corresponds to their monomeric form. The best fit frictional ratios determined for the five fusions are all between 2 and 2,2. This may indicate that the proteins are present in solution as an elongated form, which would be consistent with the protein structure.

Example 5– evaluation of secondary and tertiary structures of *C.difficile* ToxA-ToxB fusions by circular dichroism and fluorescence spectroscopy

[0162] Circular dichroism is used to determine the secondary structure composition of a protein by measuring the difference in the absorption of left-handed polarized light versus right-handed polarized light which is due to structural asymmetry. The shape and the magnitude of the CD spectra in the far-UV region (190-250nm) are different whether a protein exhibits a beta-sheet, alpha-helix or random coil structure. The relative abundance of each secondary structure type in a given protein sample can be calculated by comparison to reference spectra.

[0163] The tertiary structure of a protein sample can be assessed by the evaluation of the immobilisation of the aromatic amino acids. The observation of a CD signal in the near-UV region (250-50nm) may be attributable to the polarization of phenylalanine, tyrosine and tryptophane residues and is a good indication that the protein is folded into a well defined structure.

[0164] The following protocol was used:

1. Far UV spectra are measured using an optical path of 0,01cm from 178 to 250nm, with a 1nm resolution and bandwidth on a Jasco J-720 spectropolarimeter. Temperature of the cell is

maintained at 23°C by a Peltier thermostated RTE-111 cell block. A nitrogen flow of 10L/min is maintained during the measurements.

2. Near-UV spectra are measured using an optical path of 0,01cm from 250 to 300nm, with a 1nm resolution and bandwidth on a Jasco J-720 spectropolarimeter. Temperature of the cell is maintained at 23°C by a Peltier thermostated RTE-111 cell block. A nitrogen flow of 6L/min is maintained during the measurements.

[0165] The observation of the far-UV spectra (figure 9) for all five ToxA-ToxB fusion proteins suggests a weak content of alpha helix structures and a high content of beta sheet structures. Also, all proteins exhibited a maximum at 230nm, which is unusual for soluble globular proteins. This particularity has been well characterized in the literature and is associated with a small group of proteins known for their absence of alpha helix and their high content in beta sheet and aromatic amino acids (Zsila, Analytical Biochemistry,391(2009) 154-156). Those particularities are coherent with the structure that is expected for the ToxA-ToxB fusion proteins. Crystal structures of 13 proteins exhibiting the characteristic CD spectra with a positive signal at 230nm were compared (Protein Data Bank). The average secondary structure content of those proteins is 42% beta sheet ±9% and 7% alpha helix ±6%. This strongly indicates that the spectral signature of the ToxA-ToxB fusion proteins is diagnostic of a high beta sheet and low alpha helix containing protein.

[0166] The observation of the shape of the near-UV spectra (figure 10) for all five fusion proteins indicates that at least some of the aromatic amino acids are immobilised, which is a strong indication of a compact and specific tertiary structure. Moreover, the treatment of the protein with a denaturing concentration of urea caused the disappearance of the near-UV signal, which is an additional indication that this characteristic spectra was due to protein folding.

Example 6 – Immunisation of mice with Tox A or Tox B fragments and ToxA-ToxB fusions

[0167] Balb/C mice were immunized with the constructs described in examples 2 and 3.

Mice immunization

[0168] Groups of 15 female Balb/c mice were immunized IM at days 0, 14 and 28 with 3µg or10 µg of the separate fragments of toxA and toxB (see example 2) as well as with ToxA-ToxB fusions proteins (see example 3) adjuvanted with AS03B. A control group of 10 mice was vaccinated with AS03B alone.

[0169] Anti-ToxA and anti-ToxB ELISA titers were determined in individual sera collected at day 42 (post III).

[0170] Hemagglutination inhibition titers were determined in pooled Post III sera.

Anti-ToxA and anti-ToxB ELISA response: Protocol

[0171] Samples of the toxA or toxB fragments were coated at 1 µg/ml 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 with agitation. After washing, bound murine antibody was detected using Jackson ImmunoLaboratories Inc. peroxidase-conjugated affiniPure Goat Anti-Mouse IgG (H+L) (ref: 115-035-003) 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.

[0172] The level of anti-ToxA or anti-ToxB antibodies present in the sera was expressed in mid-point titers. A GMT was calculated for the 15 samples in each treatment group (10 for the control group).

Hemagglutination inhibition assay: Protocol

[0173] Serial twofold dilutions of mice pooled antisera (25µl) were performed in phosphate buffered saline (PBS) in 96-well U-bottom microplates.

[0174] 25 µl of native Toxin A (0,2 µg/well) were then added and the plates were incubated at room temperature for 30 minutes.

[0175] After incubation, 50 µl of purified rabbit erythrocytes diluted at 2% were added to each well. The plates were incubated at 37°C for 2 hours.

[0176] Plates were analysed visually, with hemagglutination presenting as diffuse red cells in the well and the inhibition of hemagglutination observed as a red point settled in the well.

[0177] The inhibition titers were defined as the reciprocal of the highest dilution of the serum inhibiting hemagglutination.

Cytotoxicity assay

[0178] IMR90 fibroblast cells were cultured at 37°C with 5% CO_2, in EMEM + 10% fetal bovine serum + 1% glutamine + 1% antibiotics (penicillin-streptomycin-amphotericin) and were seeded in 96-well tissue culture plates at a density of 5.10^4 cells/well.

[0179] After 24h, the cell media was removed from the wells.

[0180] Serial twofold dilutions of mice pooled antisera (50µl) were performed in cell media.

[0181] 50 µl of native Toxin B (0.5ng/ml) is then added and the plates incubated at 37°C with 5% CO2 for 24 hours.

[0182] Cells were observed after 24 hours, and the proportion of rounded cells was determined.

[0183] The inhibition titers were defined as the reciprocal of the highest dilution of the serum inhibiting 50% cell rounding.

Results:

[0184] Elisa results, using Tox A antibodies are described in figure 11. Anti-Tox A antibodies were induced after immunization with the ToxA alone but also with each of the 5 fusions.

[0185] The functional properties of these antibodies were tested in the hemagglutination assay. This assay is only adapted for Tox A evaluation as no hemagglutination is observed with ToxB.

[0186] Haemagglutination inhibition titres are described in figure 12. Haemaglutination inhibition was observed with the anti-Tox A fragment sera or sera directed against each of the ToxA-ToxB fusions.

[0187] An ELISA using ToxB antibodies was also performed; the results of this are illustrated in Figure 13. Anti-Tox B antibodies were induced after immunization with the ToxB fragment alone but also with the F2, F3 and F4 fusions.

[0188] Cytotoxicity inhibition titres are described in figure 14. Inhibition titers obtained using sera from mice immunised with the ToxB fragment or the ToxA-ToxB fusions were greater than that obtained using control sera.

Example 7 Design, cloning, expression and purification of 4 further fusion proteins

[0189] Four further fusion proteins were designed using the design principles described in example 1, these were named F54 Gly (SEQ ID NO:21), F54 New (SEQ ID NO:23), F5 ToxB (SEQ ID NO:25) and F52 New (SEQ ID NO:27).

[0190] These fusion proteins were expressed according to the procedure described in example 2.

Example 8 – Molecular weight evaluation of the *C.difficile* ToxA-ToxB fusions described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27

[0191] The molecular weight of the fusions described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27 were determined as described in example 4.

[0192] Figure 15 describes the distribution of these four further fusion proteins as determined by sedimentation velocity analytical ultracentifugation.

[0193] The molecular weight of the main species determined from the C(S) distribution of all four protein fusions described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27 corresponds to their monomeric form and all proteins exhibit sedimentation properties similar to F1 to F5 fusions.

Example 9 – evaluation of secondary and tertiary structures of *C.difficile* ToxA-ToxB fusions described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27

[0194] The secondary and tertiary structures of the fusions described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27 were assessed according to the method described in example 5. The far UV CD for these fusion proteins can be found in figure 16, and the near UV spectra for these fusions can be found in figure 17.

[0195] Analysis of the near and far UV CD spectra of the proteins described in SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, and SEQ ID NO:27 shows that all four have the same high beta sheet structure than F1 to F5 fusions. In addition, observation of the near UV spectra show no significant difference in the position of the aromatic amino acids in the tertiary structure compared to F1 to F5 fusions.

Example 10 – immunisation of mice with Tox A-Tox B fusions

[0196] Balb/c mice were immunised with the four fusion protein constructs F54 Gly (SEQ ID NO:21), F54 New (SEQ ID NO:23), F5 ToxB (SEQ ID NO:25) and F52 New (SEQ ID NO:27) as described in example 6.

[0197] An ELISA was carried out using the anti-ToxA and anti-ToxB ELISA response:protocol described in example 6 except here the samples of the toxA or toxB fragments were coated at 2µg/ml in phosphate buffered saline on high-binding microtitre plates. A hemagglutination inhibition assay was performed as described in example 6. A toxB cytotoxicity assay was performed as described in example 6. A further toxA cytotoxicity assay was performed as described below.

ToxA cytotoxicity assay

[0198] HT29 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 tissue culture plates at a density of 5.10^4 cells/well.

[0199] After 24h, the cell media was removed from the wells.

[0200] Serial twofold dilutions of mice pooled antisera (50µl) were performed in cell media.

[0201] 50µl of native Toxin B (0.15ng/ml) is then added and the plates incubated at 37°C with 5% CO_2 for 48 hours.

[0202] Cells were observed after 48 hours and the proportion of rounded cells were determined.

[0203] The results of the anti-toxA ELISA, anti-toxB Elisa, Haemagglutination inhibition and cytotoxicity assays are described in Figures 18, 19 ,20, 21 and 22 respectively.

We claim:

1. A polypeptide comprising a first fragment and a second fragment, wherein
   (i) the first fragment is a toxin A repeating domain fragment;
   (ii) the second fragment is a toxin B repeating domain fragment;
   (iii) the first fragment has a first proximal end;
   (iv) the second fragment has a second proximal end; and

   wherein the first fragment and the second fragment are separated by less than or exactly 5 amino acids in the primary structure, wherein the polypeptide elicits antibodies that neutralize both toxin A and toxin B, wherein the first proximal end is within a short repeat, the second proximal end is within a short repeat and the first proximal end and the second proximal end do not disrupt short repeat-long repeat-short repeat portions.
2. The polypeptide of claim 1 wherein the polypeptide elicits a protective immune response in a mammalian host against strains of *C.difficile*.
3. The polypeptide of any one of claims 1-2 wherein the first fragment and/or the second fragment comprise less than 25%, 20%, 18% or 15% alpha helical structure.
4. The polypeptide of any preceding claim wherein the first fragment and/or the second fragment comprise more than 25%, 30%, 35%, 38% or 40% beta sheet structure.
5. The polypeptide of any preceding claim wherein the first proximal end is not within amino acids 1878-1940, 2012-2074, 2146-2208, 2258-2322, 2394-2456, 2507-2569 or 2598-2660 of toxin A.
6. The polypeptide of any preceding claim wherein the second proximal end is not within amino acids 1881-1942, 2012-2074, 2144-2208 or 2278-2339 of toxin B.
7. The polypeptide of any preceding claim wherein the first fragment comprises at least 100, 250, 400 or 450 amino acids.
8. The polypeptide of any preceding claim with the second fragment comprises at least 100, 200, 300 or 400 amino acids.
9. The polypeptide of any preceding claim wherein the first proximal end is within repeat portion VIII (amino acids 2645-2710) of toxin A.
10. The polypeptide of any one of claims 1-9 wherein the first proximal end is within amino acids 2700-2710 or 2680-2690 of toxin A.
11. The polypeptide of any preceding claim wherein the second proximal end is within repeat portion I (amino acids 1834-1926) of toxin B.
12. The polypeptide of any preceding claim wherein the second proximal end is within amino acids 1860-1878, 1854-1876.
13. The polypeptide of any one of claims 1-10 wherein the second proximal end is within repeat portion II (amino acids 1927-2057) of toxin B.
14. The polypeptide of any one of claims 1-10 or 13 wherein the second proximal end is within amino acids 1960-1970, 1988-1998 or 1867-1877 of toxin B.
15. The polypeptide of any one of claims 9, 10, 11 or 12 wherein the first proximal end is within repeat portion VIII (amino acids 2645-2710) of toxin A and wherein the second proximal end is within repeat portion I (amino acids 1834-1926) of toxin B.
16. The polypeptide of any one of claims 9, 10, 13 or 14 wherein the first proximal end is within repeat portion VIII (amino acids 2645-2710) of toxin A and wherein the second proximal end is within repeat portion II (amino acids 1927-2057) of toxin B.
17. The polypeptide of any one of claims 9, 10, 13 or 14 wherein the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007).
18. The polypeptide of any one of claims 9, 10, 13 or 14 wherein the first proximal end is within short repeat 2 of repeat portion VIII (amino acids 2665-2686) of toxin A and the second proximal end is within short repeat 3 of repeat portion II of toxin B (amino acids 1968-1987).
19. The polypeptide of any one of claims 9, 10, 11 or 12 wherein the first proximal is within short repeat 2 of repeat portion VIII of toxin A (amino acids 2665-2686) and the second proximal end is within short repeat 3 of repeat portion I of toxin B (1877-1896).
20. The polypeptide of any one of claims 1-8, 9, 10, 11, 12 and 15 wherein the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2645-2686) and the second proximal end is within short repeat 1 of repeat portion I of toxin B (amino acids 1834-1854).

21. The polypeptide of any one of claims 9, 10, 13 or 14 wherein the first proximal end is within short repeat 3 of repeat portion VIII of toxin A (amino acids 2687-2710) and the second proximal end is within short repeat 4 of repeat portion II of toxin B (amino acids 1988-2007).
22. The polypeptide of claim 16 wherein the first proximal end is within amino acids 2700-2710 of toxin A and the second proximal end is within amino acids 1960-1970 of toxin B.
23. The polypeptide of claims 16 wherein the first proximal end is within amino acids 2680-2690 of toxin A and the second proximal end is within amino acids 1960-1970 of toxin B.
24. The polypeptide of claims 16 wherein the first proximal end is within amino acids 2700-2710 of toxin A and the second proximal end is within amino acids 1988-1998 of toxin B.
25. The polypeptide of claim 16 wherein the first proximal end is within amino acids 2680-2690 and the second proximal end is within amino acids 1860-1878.
26. The polypeptide of any preceding claim wherein the polypeptide comprises an immunogenic fragment of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7 SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25 or SEQ ID NO:27.
27. The polypeptide of any preceding claim wherein the polypeptide comprises an immunogenic fragment of at least 500, 600, 700, 750, 800, 850 or 900 amino acids of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:7 SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25 or SEQ ID NO:27.
28. The polypeptide of any preceding claim wherein the polypeptide comprises more than 450, 475, 500, 525, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825 or 850 amino acids from toxin A.
29. The polypeptide of any preceding claim wherein the polypeptide comprises less than 850, 825, 800, 775, 750, 725, 700, 675, 650, 625, or 600 amino acids from toxin A.
30. The polypeptide of any preceding claim wherein the polypeptide comprises more than 350, 375, 400, 425, 450, 475, 500 or 525 amino acids from toxin B.
31. The polypeptide of any preceding claim wherein the polypeptide comprises less than 525, 500, 475, or 450 amino acids from toxin B.
32. A polynucleotide encoding the polypeptide of any one of claims 1-31.
33. A vector comprising the polynucleotide of claim 32 linked to an inducible promoter.
34. An immunogenic composition comprising the polypeptide of any one of claims 1-31 and a pharmaceutically acceptable excipient.
35. The immunogenic composition of claim 34 further comprising an adjuvant.
36. A vaccine comprising the immunogenic composition of any one of claims 34-35.
37. The immunogenic composition of any one of claims 34-35 or the vaccine of claim 36 for use in the treatment or prevention of *C.difficile* disease.
38. A use of the immunogenic composition of any one of claims 34-35 or the vaccine of claim 36 in the preparation of a medicament for the prevention or treatment of *C.difficile* disease.