

(19) AUSTRALIAN PATENT OFFICE

(54) Title
Method for overexpression of zwitterionic polysaccharides

(51)⁶ International Patent Classification(s)
C12N 15/09 (2006.01) **0BMJP** **A61P**
A61K 31/7088 37/02
(2006.01) 20060101ALI2005122
A61K 38/00 (2006.01) **0BMJP** **C07K**
A61K 48/00 (2006.01) 14/32
A61P 1/00 (2006.01) 20060101ALI2005100
A61P 29/00 (2006.01) **8BMEP** **C07K**
A61P 37/02 (2006.01) 14/47
C07K 14/32 (2006.01) 20060101ALI2005122
C07K 14/47 (2006.01) **0BMJP** **C07K**
C07K 16/18 (2006.01) 20060101ALI2005122
C07K 19/00 (2006.01) 20060101ALI2005122
C12N 1/15 (2006.01) **0BMJP** **C07K**
C12N 1/19 (2006.01) 19/00
C12N 1/21 (2006.01) 20060101ALI2005122
C12N 5/10 (2006.01) **0BMJP** **C12N**
C12P 19/04 (2006.01) 1/15
C12Q 1/04 (2006.01) 20060101ALI2005122
C12N 15/09 **0BMJP** **C12N**
20060101AFI2005122 1/19
0BMJP **A61K** 20060101ALI2005122
31/7088 **0BMJP** **C12N**
20060101ALI2005122 1/21
0BMJP **A61K** 20060101ALI2006052
38/00 **1BMWO** **C12N**
20060101ALI2005122 5/10
0BMJP **A61K** 20060101ALI2005122
48/00 **0BMJP** **C12P**
20060101ALI2005122 19/04
0BMJP **A61P** 20060101ALI2005100
1/00 **8BMEP** **C12Q**
20060101ALI2005122 1/04
0BMJP **A61P** 20060101ALI2005122
29/00 **0BMJP**
20060101ALI2005122 PCT/US03/07775

(21) Application No: 2003218144

(22) Application Date: 2003.03.13

(87) WIPO No: W003/077863

(30) Priority Data

| | | |
|-------------|------------|--------------|
| (31) Number | (32) Date | (33) Country |
| 60/364,168 | 2002.03.13 | US |

(43) Publication Date: 2003.09.29

(43) Publication Journal Date: 2003.11.06

(71) Applicant(s)

The Brigham and Women's Hospital, Inc.

(72) Inventor(s)

Coyne, Michael J, Tzianabos, Arthur O, Comstock, Laurie E, Kasper, Denis L, Weinacht, Katja G

(74) Agent/Attorney

Davies Collison Cave, Level 15 1 Nicholson Street, Melbourne, VIC, 3000

(56) Related Art

Krinos, C. M. et al., 2001, Nature, 414(6863): 555-558

WO 1996/007427

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

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date
25 September 2003 (25.09.2003)

PCT

(10) International Publication Number
WO 2003/077863 A3

(51) International Patent Classification⁷: C12N 15/09, (74) Agent: STEELE, Alan, W.; Wolf, Greenfield & Sacks, 9/00, 9/90, 1/20, 15/00, C07H 21/04 P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(21) International Application Number: PCT/US2003/007775

(22) International Filing Date: 13 March 2003 (13.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/364,168 13 March 2002 (13.03.2002) US

(71) Applicant: THE BRIGHAM AND WOMEN'S HOSPITAL, INC. [US/US]; 75 Francis Street, Boston, MA 02115 (US).

(72) Inventors: COMSTOCK, Laurie, E.; 1 Lilac Lane, Medfield, MA 02052 (US). WEINACHT, Katja, G.; Berliner Strasse 13/2, 79211 Denzlingen (DE). COYNE, Michael, J.; 42 Pinetree Drive, North Attleboro, MA 02760 (US). KASPER, Denis, L.; 48 Monument Avenue, Charlestown, MA 02129 (US). TZIANABOS, Arthur, O.; 37 Hampshire Road, Reading, MA 01867 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

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

Published:
— with international search report

(88) Date of publication of the international search report: 22 July 2004

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

WO 2003/077863 A3

(54) Title: METHOD FOR OVEREXPRESSION OF ZWITTERIONIC POLYSACCHARIDES

(57) Abstract: The present invention is directed to methods for producing and selecting novel mutant strains of *B. fragilis* that constitutively express a particular capsular polysaccharide or only selected capsular polysaccharides; compositions directed to the novel mutant strains of *B. fragilis* that constitutively express a particular capsular polysaccharide or only selected capsular polysaccharides; improved methods for purification of individual capsular polysaccharides; and compositions directed to novel resf2 and inv19 genes and their gene products. Significantly, the present invention provides methods and compositions for overexpressing and purifying immunomodulatory capsular polysaccharide A (PSA) in high yield.

METHOD FOR OVEREXPRESSION OF ZWITTERIONIC POLYSACCHARIDES**Field of the Invention**

The invention relates to compositions and methods for the production and isolation of capsular polysaccharides which have been reported to have immunomodulatory effects. Specifically, the invention relates to compositions and methods for the production and isolation of capsular polysaccharide A (PSA) of *Bacteroides fragilis* (*B. fragilis*).

Background of the Invention

Capsular polysaccharide A (PSA) of *Bacteroides fragilis* NCTC9343 has been reported to be an immunomodulator with therapeutic and preventative applications. U.S. Pat. Nos. 5,679,654 and 5,700,787; Tzianabos AO et al. (2000) *J Biol Chem* 275:6733-40. It was recently reported that in addition to PSA, *B. fragilis* NCTC9343 synthesizes at least seven other capsular polysaccharide (PSB-PSH). Krinos CM et al. (2001) *Nature* 414:555-558. It has also recently been reported that expression of seven of these eight capsular polysaccharides of *B. fragilis* are variable due to phase variation dictated by inversion of DNA segments containing the promoters of each of the polysaccharide biosynthesis loci. Krinos CM et al., *supra*. The fact that this strain synthesizes so many polysaccharides makes purification of PSA from this strain very laborious. This, coupled with the variable expression of PSA due to its phase variation, often results in a very low yield of PSA following extensive procedures for its purification. Scaled-up purification of PSA for preventative or therapeutic applications thus presents technical obstacles.

Summary of the Invention

The invention arises in part from the discovery by the present inventors of methods for controlling the phase variation of seven of the eight known capsular polysaccharides of *B. fragilis*, PSA, PSB, PSD, PSE, PSF, PSG, and PSH. It was surprisingly discovered that inactivation of a novel gene, *res02* (also denoted *mpi*, multiple promoter invertase), has the effect of locking the invertible promoters of each of the capsular polysaccharide biosynthesis loci in either an "on" orientation or an "off" orientation. When an invertible promoter of a capsular polysaccharide is in an on orientation, the promoter is transcriptionally active with respect to the polysaccharide biosynthesis locus associated with the biosynthesis of the

- 2 -

capsular polysaccharide. Conversely, when an invertible promoter of a capsular polysaccharide is in an off orientation, the promoter is transcriptionally inactive with respect to the polysaccharide biosynthesis locus associated with the biosynthesis of the capsular polysaccharide. By locking the promoters and then selecting for bacterial cells constitutively expressing only a particular capsular polysaccharide, or expressing only a restricted set of capsular polysaccharides including the particular capsular polysaccharide, it was discovered according to the instant invention that it was now possible to increase the yield of the particular capsular polysaccharide. Thus, it was discovered according to the instant invention that it is now possible to generate and select mutant strains of *B. fragilis* that constitutively express any combination of selected capsular polysaccharides and not others. Furthermore, using this method it is now possible to generate and select mutant strains of *B. fragilis* that constitutively express only one particular capsular polysaccharide, e.g., PSA, and no other capsular polysaccharide. Because each selected mutant strain is rendered incapable of phase variation, the amount of selected polysaccharide, e.g., PSA, made by each cell is increased compared to wild type. In addition, because a selected mutant strain expresses only a single capsular polysaccharide, or it expresses only a desired combination of capsular polysaccharides, purification of capsular polysaccharide is greatly simplified and more efficient.

The invention also relates in part to a second novel gene, *inv19*, located adjacent to *res02* and believed by the applicants also to be involved in controlling the expression of capsular polysaccharides in *B. fragilis*. Unlike *res02*, inactivation or deletion of *inv19* does not appear to result in locking the phase variable promoters of capsular polysaccharide biosynthesis loci. However, it was discovered according to the instant invention that deletion of both *res02* and *inv19* can produce a genotype different from *res02* deletion alone. For example, in one instance a *res02* deletion alone did not express PSC, while a double *res02/inv19* deletion did express PSC.

As detailed below, the present invention thus includes methods for producing and selecting novel mutant strains of *B. fragilis* that constitutively express only particular capsular polysaccharides; compositions directed to the novel mutant strains of *B. fragilis* that constitutively express only particular capsular polysaccharides; improved methods for purification of individual capsular polysaccharides; and novel compositions directed to the *res02* and *inv19* genes and their gene products.

- 3 -

It is believed that the novel res02 polypeptide encoded by the novel res02 gene functions as an invertase for nucleic acid sequences having specific structural characteristics. The specific structural characteristics include the presence of inverted repeat sequences flanking a central sequence, where the inverted repeat sequences have a particular sequence 5 specifically recognized by the res02 polypeptide.

In one aspect the invention provides an isolated nucleic acid molecule related to res02. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence as set forth in SEQ ID NO:1; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2; (c) a nucleotide 10 sequence which hybridizes under stringent conditions to a complement of (a) or (b); and (d) a nucleotide sequence complementary to any of (a)-(c). In one embodiment the nucleic acid molecule is a nucleotide sequence as set forth in SEQ ID NO:1. In one embodiment the nucleic acid molecule is a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2. A nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 includes a 15 nucleotide sequence that differs from the nucleotide sequence as set forth in SEQ ID NO:1 in codon sequence due to degeneracy of the genetic code.

In another aspect the invention provides an isolated nucleic acid molecule related to res02. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide which is at least 20 about 70 percent identical to a polypeptide as set forth in SEQ ID NO:2, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) a nucleotide sequence encoding an allelic variant of a nucleotide sequence as set forth in SEQ ID NO:1 or (a); (c) a region of the nucleotide sequence of SEQ ID NO:1, (a), or (b) encoding a polypeptide fragment of at least about 9 amino acid residues, wherein the polypeptide 25 fragment has an activity of the encoded polypeptide as set forth in SEQ ID NO:2, or is antigenic; (d) a region of the nucleotide sequence of SEQ ID NO:1, or any of (a)-(c) comprising a fragment of at least about 16 nucleotides; (e) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(d); and (f) a nucleotide sequence complementary to any of (a)-(d).

30 In yet another aspect the invention provides an isolated nucleic acid molecule related to res02. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide as set forth in

- 4 -

SEQ ID NO:2 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 which has a C- and/or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (f) a nucleotide sequence of any of (a)-(e) comprising a fragment of at least about 16 nucleotides; (g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f); and (h) a nucleotide sequence complementary to any of (a)-(e).

In another aspect the invention provides a vector including the res02 nucleic acid molecule of any of the aspects above.

In a further aspect the invention provides a host cell including any of the res02 vectors above.

In another aspect the invention provides an isolated polypeptide related to res02. The isolated polypeptide according to this aspect includes an amino acid sequence as set forth in SEQ ID NO:2.

In another aspect the invention provides an isolated polypeptide related to res02. The isolated polypeptide according to this aspect includes an amino acid sequence selected from the group consisting of: (a) an amino acid sequence for an ortholog of SEQ ID NO:2; (b) an amino acid sequence which is at least about 70 percent identical to the amino acid sequence of SEQ ID NO:2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) a fragment of the amino acid sequence set forth in SEQ ID NO:2 comprising at least about 9 amino acid residues, wherein the fragment has an activity of the polypeptide set forth in SEQ ID NO:2, or is antigenic; and (d) an amino acid sequence for an allelic variant of the amino acid sequence as set forth in SEQ ID NO:2, (a), or (b).

- 5 -

In yet another aspect the invention provides an isolated polypeptide related to rcs02. The isolated polypeptide according to this aspect includes an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as set forth in SEQ ID NO:2 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) the amino acid sequence as set forth in SEQ ID NO:2 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) the amino acid sequence as set forth in SEQ ID NO:2 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (d) the amino acid sequence as set forth in SEQ ID NO:2 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; and (e) the amino acid sequence as set forth in SEQ ID NO:2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2.

In certain embodiments the isolated polypeptide encoded by any of the foregoing res02 nucleic acid molecules has an activity of the polypeptide set forth in SEQ ID NO:2. Specifically, in certain embodiments the activity is promoter invertase activity.

The invention in another aspect provides a selective binding agent or fragment thereof which specifically binds the res02 polypeptide of any of the foregoing aspects. In one embodiment the selective binding agent or fragment thereof specifically binds the polypeptide comprising the amino acid sequence as set forth in SEQ ID NO:2 or a fragment thereof. In one embodiment the selective binding agent is an antibody or fragment thereof.

In yet another aspect the invention provides a fusion polypeptide comprising the res02 polypeptide of any of the foregoing aspects, fused to a heterologous amino acid sequence.

In a further aspect the invention provides an isolated nucleic acid molecule related to inv19. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence as set forth in SEQ ID NO:3; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4; (c) a nucleotide sequence which hybridizes under stringent conditions to a complement of (a) or (b); and (d) a nucleotide sequence complementary to any of (a)-(c). In one embodiment the nucleic acid molecule is a nucleotide sequence as set forth in SEQ ID NO:3. In one embodiment the

- 6 -

nucleic acid molecule is a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4. A nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 includes a nucleotide sequence that differs from the nucleotide sequence as set forth in SEQ ID NO:3 in codon sequence due to degeneracy of the genetic code.

5 In another aspect the invention provides an isolated nucleic acid molecule related to inv19. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide which is at least about 70 percent identical to a polypeptide as set forth in SEQ ID NO:4, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) a nucleotide sequence encoding an allelic variant of a nucleotide sequence as set forth in SEQ ID NO:3 or (a); (c) a region of the nucleotide sequence of SEQ ID NO:3, (a), or (b) encoding a polypeptide fragment of at least about 9 amino acid residues, wherein the polypeptide fragment has an activity of the encoded polypeptide as set forth in SEQ ID NO:4, or is antigenic; (d) a region of the nucleotide sequence of SEQ ID NO:3, or any of (a)-(c)

10 15 comprising a fragment of at least about 16 nucleotides; (e) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(d); and (f) a nucleotide sequence complementary to any of (a)-(d).

In yet another aspect the invention provides an isolated nucleic acid molecule related to inv19. The nucleic acid according to this aspect includes a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 which has a C- and/or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the encoded

- 7 -

polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (f) a nucleotide sequence of any of (a)-(e) comprising a fragment of at least about 16 nucleotides; (g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f); and (h) a nucleotide sequence complementary to any of (a)-(e).

5 In another aspect the invention provides a vector including the inv19 nucleic acid molecule of any of the aspects above.

In a further aspect the invention provides a host cell including any of the inv19 vectors above.

10 In another aspect the invention provides an isolated polypeptide related to inv19. The isolated polypeptide according to this aspect includes an amino acid sequence as set forth in SEQ ID NO:4.

15 In another aspect the invention provides an isolated polypeptide related to inv19. The isolated polypeptide according to this aspect includes an amino acid sequence selected from the group consisting of: (a) an amino acid sequence for an ortholog of SEQ ID NO:4; (b) an amino acid sequence which is at least about 70 percent identical to the amino acid sequence of SEQ ID NO:4, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) a fragment of the amino acid sequence set forth in SEQ ID NO:4 comprising at least about 9 amino acid residues, wherein the fragment has an activity of the polypeptide set forth in SEQ ID NO:4, or is antigenic; and (d) an amino acid sequence for an allelic variant 20 of the amino acid sequence as set forth in SEQ ID NO:4, (a), or (b).

25 In yet another aspect the invention provides an isolated polypeptide related to inv19. The isolated polypeptide according to this aspect includes an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as set forth in SEQ ID NO:4 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) the amino acid sequence as set forth in SEQ ID NO:4 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) the amino acid sequence as set forth in SEQ ID NO:4 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (d) the amino acid sequence as set forth in SEQ ID NO:4 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; and (e) the amino acid sequence as set forth in SEQ ID NO:4 with at least one modification selected from the group consisting of amino acid

- 8 -

substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4.

5 In certain embodiments the isolated polypeptide encoded by any of the foregoing inv19 nucleic acid molecules has an activity of the polypeptide set forth in SEQ ID NO:4.

The invention in another aspect provides a selective binding agent or fragment thereof which specifically binds the inv19 polypeptide of any of the foregoing aspects. In one embodiment the selective binding agent or fragment thereof specifically binds the polypeptide comprising the amino acid sequence as set forth in SEQ ID NO:4 or a fragment 10 thereof. In one embodiment the selective binding agent is an antibody or fragment thereof

In yet another aspect the invention provides a fusion polypeptide comprising the inv19 polypeptide of any of the foregoing aspects, fused to a heterologous amino acid sequence.

According to yet another aspect of the invention, a bacterial cell is provided in which 15 res02 expression is disabled. The expression of res02 can be disabled by any of a number of possible mechanisms, including, but not limited to, alteration of the res02 gene sequence by an insertion or deletion mutation that causes either deletion, truncation, or frameshift of the translated product, or disruption of a res02 gene expression sequence, such that res02 is not expressed by the cell.

20 According to yet another aspect of the invention, a bacterial cell is provided in which inv19 expression is disabled. The expression of inv19 can be disabled by any of a number of possible mechanisms, including, but not limited to, alteration of the inv19 gene sequence by an insertion or deletion mutation that causes either deletion, truncation, or frameshift of the translated product, or disruption of an inv19 gene expression sequence, such that inv19 is not 25 expressed by the cell.

In another aspect the invention provides a population of bacterial cells stably expressing a specific capsular polysaccharide, or only a limited set of capsular polysaccharides including the specific capsular polysaccharide, selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH. In one embodiment the specific 30 capsular polysaccharide is PSA. In certain embodiments the bacterial cells are *B. fragilis*, including *B. fragilis* NCTC9343. In one embodiment the bacterial cells are *B. fragilis* 9343res02mut44. In another embodiment the bacterial cells are *B. fragilis* 9343res02mut2.

- 9 -

The invention further provides a bacterial cell expressing a capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, wherein a promoter controlling expression of the capsular polysaccharide is locked on. In one preferred embodiment the capsular polysaccharide is PSA. In some embodiments promoters 5 controlling expression of each and every capsular polysaccharide selected from the group consisting of: PSB, PSD, PSE, PSF, PSG, and PSH are locked off. In some embodiments any one or combination of capsular polysaccharides selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH is not expressed. Also in some embodiments a promoter controlling expression of any one or combination of capsular polysaccharides 10 selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, wherein said any one or combination of capsular polysaccharides is not expressed, is locked off. In one embodiment res02 expression is disabled. In one embodiment inv19 expression is disabled. In one embodiment res02 expression and inv19 expression are both disabled. In one embodiment the bacterial cell according to any of the foregoing aspects is *B. fragilis*. In 15 one embodiment the bacterial cell according to any of the foregoing aspects is *B. fragilis* NCTC9343. In one embodiment the bacterial cell is *B. fragilis* 9343res02mut44. In another embodiment the bacterial cell is *B. fragilis* 9343res02mut2.

According to yet another aspect, the invention provides a method for locking a phase-variable promoter of a capsular polysaccharide biosynthesis gene on. The method according 20 to this aspect involves inactivating res02 in a bacterial cell to lock a phase-variable promoter of a biosynthesis gene for a capsular polysaccharide on.

In a further aspect the invention provides a method for affecting expression of a capsular polysaccharide biosynthesis gene. The method according to this aspect involves inactivating inv19 in a bacterial cell to affect expression of a biosynthesis gene for a capsular 25 polysaccharide. In certain embodiments the method further involves selecting a bacterial cell expressing at least one capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH. In certain embodiments the method further involves selecting a bacterial cell expressing only one capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH. In one embodiment the one 30 capsular polysaccharide is PSA.

The invention in another aspect provides an improved method for purifying PSA from a bacterial cell. The improvement involves inactivating res02 in the bacterial cell such that a

- 10 -

promoter for PSA is locked on or off, and selecting a bacterial cell expressing PSA. In one embodiment the bacterial cell is *B. fragilis*. In one embodiment the cell is *B. fragilis* NCTC9343. In one particular embodiment the bacterial cell is *B. fragilis* 9343res02mut44. In another particular embodiment the bacterial cell is *B. fragilis* 9343res02mut2.

5 The invention according to yet another aspect further provides a method for producing a pure capsular polysaccharide. The method according to this aspect involves growing a population of bacterial cells stably expressing a specific capsular polysaccharide, or only a restricted set of capsular polysaccharides including the specific capsular polysaccharide, selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, and
10 isolating the specific capsular polysaccharide from the population of bacterial cells to produce a pure capsular polysaccharide. In one embodiment a promoter controlling expression of any one or combination of capsular polysaccharides selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, excluding the specific capsular polysaccharide, is locked off. In one embodiment the specific expressed capsular
15 polysaccharide is PSA. In one embodiment promoters controlling expression of each and every capsular polysaccharide selected from the group consisting of: PSB, PSD, PSE, PSF, PSG, and PSH, are locked off. In one embodiment the bacterial cells are *B. fragilis*. In one embodiment the bacterial cells are *B. fragilis* NCTC9343. In a particular embodiment the bacterial cells are *B. fragilis* 9343res02mut44. In another a particular embodiment the
20 bacterial cells are *B. fragilis* 9343res02mut2.

In yet another aspect the invention provides a method of treating or preventing inflammatory bowel disease. The method according to this aspect involves administering to a subject in need of treatment for or prevention of inflammatory bowel disease an effective amount of a population of bacterial cells stably expressing a specific capsular polysaccharide
25 to treat or prevent the inflammatory bowel disease. In one embodiment the specific capsular polysaccharide is PSA. In one embodiment the bacterial cells are *B. fragilis*. In one embodiment the bacterial cells are *B. fragilis* NCTC9343. In one embodiment the bacterial cells are *B. fragilis* 9343res02mut44 (mut44). In another embodiment the bacterial cells are *B. fragilis* 9343res02mut2 (mut2).

30

Brief Description of the Drawings

Figure 1 is the nucleotide sequence of *res02* (SEQ ID NO:1).

- 11 -

Figure 2 is the deduced amino acid sequence of res02 (SEQ ID NO:2).

Figure 3 is the nucleotide sequence of *inv19* (SEQ ID NO:3).

Figure 4 is the deduced amino acid sequence of *inv19* (SEQ ID NO:4).

Figure 5 is a diagram showing the construction of pKGW10 and mutation strategy

5 used in deleting res02 from the *B. fragilis* NCTC9343 chromosome. The direction of translation of res02 is from right to left in the figure.

Figure 6 is the nucleotide sequence of Left Flank, *inv19-D1*→ through ←*inv19-D2*, 1,958 bp (SEQ ID NO:5).

Figure 7 is the nucleotide sequence of Right Flank, *inv19-D5*→ through ←*inv19-D6*,

10 2,540 bp (SEQ ID NO:6).

Figure 8 is a composite of Western immunoblot images demonstrating that *mut2* (lane 2) and *mut44* (lane 4) express PSA, but not PSB, PSC, PSD, PSE, PSG, or PSH. Small amounts of PSF were expressed by each of these mutant strains, but not by another res02 mutant, *res02mut8*. Wild-type (lane 1) and *res02mut8* (lane 3) are also shown for 15 comparison.

Detailed Description

The technology described in this application has utilized the genetic information that the inventors have discovered about *B. fragilis* NCTC9343 to circumvent both the difficulty 20 of purification of one capsular polysaccharide from the other seven polysaccharides, and the low yield of a given capsular polysaccharide due to the phase variation in its expression. Importantly, the technology described in this application circumvents both the difficulty of purification of PSA from the other seven polysaccharides, and the low yield of PSA due to the phase variation in its expression.

25 In featured aspects of the invention, *res02* is inactivated so that phase variation of the capsular polysaccharides is eliminated. The lack of *res02* activity renders the invertible promoters for the polysaccharide biosynthesis loci locked in either the on or the off orientation. Selection of particular phenotypes then permits selection of mutants with the promoter of a selected capsular polysaccharide locked on and the promoters of other capsular 30 polysaccharides locked off.

The terms “*res02*”, “*res02* gene”, and “*res02* nucleic acid molecule” refer to a nucleic acid molecule comprising or consisting of a nucleotide sequence as set forth in SEQ ID

- 12 -

NO:1, a nucleotide sequence encoding the polypeptide as set forth in SEQ ID NO:2, and nucleic acid molecules as defined herein.

The terms “*inv19*”, “*inv19 gene*”, and “*inv19 nucleic acid molecule*” refer to a nucleic acid molecule comprising or consisting of a nucleotide sequence as set forth in SEQ ID

5 NO:3, a nucleotide sequence encoding the polypeptide as set forth in SEQ ID NO:4, and nucleic acid molecules as defined herein.

The terms “*res02*”, “*res02 polypeptide*”, and “*res02 gene product*” refer to a polypeptide comprising the amino acid sequence of SEQ ID NO:2 and related polypeptides.

10 Related polypeptides include *res02* polypeptide fragments, *res02* polypeptide orthologs, *res02* polypeptide variants, and *res02* polypeptide derivatives, which possess at least one activity of the polypeptide as set forth in SEQ ID NO:2.

The terms “*inv19*”, “*inv19 polypeptide*”, and “*inv19 gene product*” refer to a polypeptide comprising the amino acid sequence of SEQ ID NO:4 and related polypeptides.

15 Related polypeptides include *inv19* polypeptide fragments, *inv19* polypeptide orthologs, *inv19* polypeptide variants, and *inv19* polypeptide derivatives, which possess at least one activity of the polypeptide as set forth in SEQ ID NO:4.

Methods and Compositions for Overexpressing Capsular Polysaccharides

A bacterial cell in which “*res02 expression is disabled*” is a viable bacterial cell in

20 which the level of functional *res02* polypeptide is negligible compared to the level of functional *res02* polypeptide normally expressed by the bacterial cell under the same conditions. The expression can be disabled by the introduction or presence of a mutation in the *res02* gene that results in a nonfunctional *res02* gene product or no functional *res02* gene product. Such mutations include those involving at least one missense mutation, nonsense 25 mutation, truncation mutation, insertion mutation, or deletion mutation, or any combination thereof, involving the *res02* open reading frame (ORF). In some embodiments mutations resulting in disabled *res02* expression include at least one truncation, insertion, or deletion mutation, or any combination thereof, involving the *res02* promoter. In some embodiments *res02* expression is disabled through manipulation of factors upstream of the expression of 30 *res02*. For example, transcriptional activators, signal transduction, and global regulators can influence expression of *res02*. See, for example, Gally DL et al. (1993) *J Bacteriol* 175:6186-93; Blomfield IC et al. (1993) *J Bacteriol* 175:27-36; Dorman CJ et al. (1987) *J*

- 13 -

Bacteriol 169:3840-43; Eisenstein BI et al. (1987) *Proc Natl Acad Sci USA* 84:6506-10. In one embodiment, a bacterial cell in which res02 expression is disabled has a chromosomal deletion of the res02 gene.

For example, a bacterial cell in which res02 expression is disabled can be a cell in
5 which there is a deletion of any one or combination of the first, second, third, fourth, fifth, sixth, and so on, nucleotides shown in SEQ ID NO:1. In one embodiment the deletion occurs nearer to the 5' end than to the 3' end. For example, deletion of the first, second or third nucleotide shown in SEQ ID NO:1 obliterates the "atg" start codon at position 1 and moves the first "atg" start codon to position 67 of SEQ ID NO:1, generating an ORF that encodes
10 amino acids 23-197 shown in SEQ ID NO:2. As another example, deletion of the fourth nucleotide shown in SEQ ID NO:1 maintains the original "atg" start codon at position 1 but introduces a number of adjacent and nearby stop codons.

As a further example, a bacterial cell in which res02 expression is disabled can be a cell in which there is an insertion of one or more nucleotides following any one or
15 combination of the first, second, third, fourth, fifth, sixth, and so on, nucleotides shown in SEQ ID NO:1. For example, the insertion can be an insertion of an in-frame stop codon (taa, tag, tga) in SEQ ID NO:1. As a further example, the insertion can be a polynucleotide, e.g., a polynucleotide sequence that encodes an exogenous gene product. In one embodiment the insertion occurs nearer to the 5' end than to the 3' end.

20 The deletion or insertion can be accomplished by homologous recombination.

Homologous recombination is a technique originally developed for targeting genes to induce or correct mutations in transcriptionally active genes. Kucherlapati RS (1989) *Prog Nucleic Acid Res Mol Biol* 36:301-10. The basic technique was developed as a method for introducing specific mutations into specific regions of the mammalian genome (Thomas KR
25 et al. (1986) *Cell* 44:419-28; Thomas KR et al. (1987) *Cell* 51:503-12; Doetschman T et al. (1988) *Proc Natl Acad Sci USA* 85:8583-87) or to correct specific mutations within defective genes (Doetschman T et al. (1987) *Nature* 330:576-78). Exemplary homologous recombination techniques are described in U.S. Pat. No. 5,272,071; European Patent Nos. 9193051 and 505500; and PCT/US90/07642 (PCT Pub No. WO 91/09955).

30 Through homologous recombination, the DNA sequence to be inserted into the genome can be directed to a specific region of the gene of interest by attaching it to targeting DNA. The targeting DNA is a nucleotide sequence that is complementary (homologous) to a

- 14 -

region of the genomic DNA. Small pieces of targeting DNA that are complementary to a specific region of the genome are put in contact with the parental strand during the DNA replication process. It is a general property of DNA that has been inserted into a cell to hybridize, and therefore, recombine with other pieces of endogenous DNA through shared 5 homologous regions. If this complementary strand is attached to an oligonucleotide that contains a mutation or a different sequence or an additional nucleotide, it too is incorporated into the newly synthesized strand as a result of the recombination. As a result of the proofreading function, it is possible for the new sequence of DNA to serve as the template. Thus, the transferred DNA is incorporated into the genome.

10 Homologous recombination can also be used to delete a region of genomic DNA. In this instance the targeting DNA is a nucleotide sequence that is complementary (homologous) to regions of the genomic DNA flanking that which is to be deleted. The targeting DNA is usually at least one kilobase (1 kb) long for each flanking region, and more preferably it is closer to 2 kb for each flanking region. If this complementary strand is put in contact with 15 the parental strand during the DNA replication process, it is incorporated into the newly synthesized strand as a result of recombination. As a result of the proofreading function, it is possible for the new sequence of DNA, lacking the original genomic sequence between the homologous flanking regions, to serve as the template. Thus, the transferred DNA is incorporated into the genome.

20 Deletion by homologous recombination involves double crossover allelic exchange. In this method a first recombination event occurs with one flank region, and then a second recombination event occurs with the opposite flank. As a result, double crossover homologous exchange results in two groups of cells, those which are deletion mutants and those which are wild type. The determination of whether a particular cell is wild type or a 25 deletion mutant can be accomplished by screening for the wild type gene or wild type gene product, or direct screening for the chromosomal deletion.

A bacterial cell in which "inv19 expression is disabled" is a viable bacterial cell in which the level of functional inv19 polypeptide is negligible compared to the level of functional inv19 polypeptide normally expressed by the bacterial cell under the same 30 conditions. The expression can be disabled by the introduction or presence of a mutation in the inv19 gene that results in a nonfunctional inv19 gene product or no functional inv19 gene product. Such mutations include those involving at least one missense mutation, nonsense

- 15 -

mutation, truncation mutation, insertion mutation, or deletion mutation, or any combination thereof, involving the *inv19* open reading frame (ORF). In some embodiments mutations resulting in disabled *inv19* expression include at least one truncation, insertion, or deletion mutation, or any combination thereof, involving the *inv19* promoter. In some embodiments 5 *inv19* expression is disabled through manipulation of factors upstream of the expression of *inv19*. For example, transcriptional activators, signal transduction, and global regulators can influence expression of *res02*. See, for example, Gally DL et al. (1993) *J Bacteriol* 175:6186-93; Blomfield IC et al. (1993) *J Bacteriol* 175:27-36; Dorman CJ et al. (1987) *J Bacteriol* 169:3840-43; Eisenstein BI et al. (1987) *Proc Natl Acad Sci USA* 84:6506-10. In 10 one embodiment, a bacterial cell in which *inv19* expression is disabled has a chromosomal deletion of the *inv19* gene.

For example, a bacterial cell in which *inv19* expression is disabled can be a cell in which there is a deletion of any one or combination of the first, second, third, fourth, fifth, sixth, and so on, nucleotides shown in SEQ ID NO:3. In one embodiment the deletion occurs 15 nearer to the 5' end than to the 3' end.

As a further example, a bacterial cell in which *inv19* expression is disabled can be a cell in which there is an insertion of one or more nucleotides following any one or combination of the first, second, third, fourth, fifth, sixth, and so on, nucleotides shown in 20 SEQ ID NO:3. For example, the insertion can be an insertion of an in-frame stop codon (taa, tag, tga) in SEQ ID NO:3. As a further example, the insertion can be a polynucleotide, e.g., a polynucleotide sequence that encodes an exogenous gene product. In one embodiment the insertion occurs nearer to the 5' end than to the 3' end.

The deletion or insertion can be accomplished by homologous recombination.

A "population of bacterial cells" is a culture of bacterial cells. The culture can be 25 liquid culture, semi-solid culture, e.g., in gelatin, or a culture on solid medium, e.g., on nutrient-supplemented agar. Examples of these various types of bacterial culture are well-known by those of skill in the art. A liquid culture can be in a test tube, flask, roller bottle, bioreactor, or other suitable container, preferably with controlled conditions of temperature, aeration, and agitation. The volume of a liquid culture can range from less than 1 mL to 10 L 30 or more. In one embodiment the population of bacterial cells derives from a single bacterium, i.e., it is a clone.

- 16 -

A "bacterial cell expressing a capsular polysaccharide" refers to the detectable phenotype of a bacterial cell with respect to the capsular polysaccharide. The phenotype can be concordant or discordant with genotype. For example, a bacterial cell with a promoter for PSE locked in the on orientation may or may not express PSE. Of course, a bacterial cell

5 with a promoter for PSE locked in the off orientation generally will not express PSE. The genotype can conveniently be assessed by polymerase chain reaction, restriction endonuclease digestion, direct sequencing, or any other method suitable for determining or inferring the sequence of a relevant segment of chromosomal DNA or RNA. The phenotype can conveniently be assessed by Western immunoblot, immunoaffinity, or other suitable

10 assay using antibodies or other capsular polysaccharide binding agents specific for the particular capsular polysaccharide to be assayed. The antibodies can be in the form of monospecific or defined specificity antisera, polyclonal antibodies, monoclonal antibodies, polysaccharide-specific binding fragments of the foregoing antibodies, and derivatives thereof.

15 A "bacterial cell stably expressing a specific capsular polysaccharide" refers to a bacterial cell expressing a particular capsular polysaccharide without phase variation of its expression. A "population of bacterial cells stably expressing a specific capsular polysaccharide" refers to a population of bacterial cells expressing a particular capsular polysaccharide without phase variation of its expression. In certain embodiments the specific

20 capsular polysaccharide is the only capsular polysaccharide expressed by the bacterial cell or by the population of bacterial cells. The absence of phase variation can reflect the inactivation of *revO2* such that the normally invertible promoter for the polysaccharide biosynthesis gene is locked in its on orientation.

A "promoter controlling expression of a capsular polysaccharide" as used herein

25 refers to a nontranscribed genetic element associated with and controlling the transcription of a capsular polysaccharide biosynthesis gene. The capsular polysaccharide biosynthesis gene can occur as part of a polycistronic capsular polysaccharide biosynthesis locus. A given promoter can regulate transcription of one or more capsular polysaccharide biosynthesis genes within the associated capsular polysaccharide biosynthesis gene locus. In one

30 embodiment the promoter includes inverted repeat regions separated by intervening sequence. In one embodiment the promoter is contained between inverted repeat regions in the intervening sequence and is subject to inversion such that in one orientation the promoter

- 17 -

is transcriptionally active ("on") with respect to the capsular polysaccharide biosynthesis gene, while in the opposite or flipped orientation the promoter is transcriptionally inactive ("off") with respect to the capsular polysaccharide biosynthesis gene. The inversion of the invertible promoter region between the inverted repeat regions is subject to control by a sequence-specific enzyme termed a recombinase or invertase. In one embodiment, the recombinase or invertase is res02 or inv19. The promoters for seven of the eight known capsular polysaccharides of *B. fragilis*, PSA, PSB, PSD, PSE, PSF, PSG, and PSH, have been reported to be flanked by inverted repeat regions and are subject to inversion. Krinos CM et al. (2001) *Nature* 414:555-58.

The downstream inverted repeat of each the seven capsular polysaccharides in *B. fragilis* is shown in **Table 1**. Sequence shown in bold print represents a consensus res02 recognition sequence necessary for promoter inversion.

Table 1. Characteristics of the inverted repeat (IR) regions upstream of the seven polysaccharide (PS) biosynthesis loci.

| PS LOCUS | SEQUENCE OF THE DOWNSTREAM IR | BASEPAIRS BETWEEN IRS | SEQ ID NO: |
|----------|--|-----------------------|------------|
| PSA | acgaacg ttttt gaaaca | 193 | 7 |
| PSB | acgaacg ttttt gaaaca | 181 | 7 |
| PSD | tagacgatcg tctatt gaaaca | 189 | 8 |
| PSE | acgaacg ttttt gaaaca | 168 | 7 |
| PSF | ttaaacg aacgtctatt gaaacact | 187 | 9 |
| PSG | gttcaaata tagacgacgtt | 174 | 10 |
| PSH | acgaacg ttttt gaaaca | 192 | 7 |
| PSC | None | | |

A "phase-variable promoter of a capsular polysaccharide biosynthesis gene" refers to a promoter for a capsular polysaccharide biosynthesis gene that, as just described, includes inverted repeat regions separated by intervening sequence and is subject to inversion such that in one orientation the promoter is transcriptionally active ("on") with respect to the capsular polysaccharide biosynthesis genes, while in the opposite orientation the promoter is transcriptionally inactive ("off") with respect to the capsular polysaccharide biosynthesis genes. Phase variation results in the variable expression of the capsular polysaccharide in response to the orientation of the invertible promoter. This phase variation normally can

- 18 -

occur within a given cell over time, for reasons that are not yet understood. Seven of the eight known capsular polysaccharides of *B. fragilis*, PSA, PSB, PSD, PSE, PSF, PSG, and PSH, are subject to phase variation. For the eight known capsular polysaccharides, the phenotype of a given cell theoretically can vary over time among any of 2^8 (256) phenotypes.

5 However, as disclosed herein, the phase variation among the capsular polysaccharides can be fixed or locked through disabling of the promoter inversion mechanism that underlies the phase variation.

A promoter controlling expression of a capsular polysaccharide is "locked on" when the invertible promoter is in its transcriptionally active orientation and it cannot invert to the 10 transcriptionally inactive orientation. The promoter can be locked on because a sequence-specific enzyme that normally inverts the promoter is not present or is otherwise disabled. Alternatively, the promoter can be locked on because at least one inverted repeat flanking the invertible region of the promoter is altered, e.g., deleted, so that the inversion is not possible.

A promoter controlling expression of a capsular polysaccharide is "locked off" when 15 the invertible promoter is in its transcriptionally inactive orientation and it cannot invert to the transcriptionally active orientation. The promoter can be locked off because a sequence-specific enzyme that normally inverts the promoter is not present or is otherwise disabled. Alternatively, the promoter can be locked off because at least one inverted repeat flanking the invertible region of the promoter is altered, e.g., deleted, so that the inversion is not possible.

20 The phrase "inactivating res02 in a bacterial cell to lock a phase-variable promoter of a biosynthesis gene for a capsular polysaccharide" refers to any intervention that renders inversion of a phase-variable promoter of a biosynthesis gene for a capsular polysaccharide of the bacterial cell by res02 gene product impossible. The intervention can typically involve deletion of the res02 gene as described herein. Other interventions are also contemplated, 25 including introduction of other res02 gene mutations that result in nonfunctional res02 gene products, introduction of res02 antisense nucleic acid or other agent that bind to res02 nucleic acid molecule, as well as introduction into the bacterial cell of agents that can interfere with the function of otherwise functional res02 gene product.

The phrase "inactivating inv19 in a bacterial cell to lock a phase-variable promoter of 30 a biosynthesis gene for a capsular polysaccharide" refers to any intervention that renders inversion of a phase-variable promoter of a biosynthesis gene for a capsular polysaccharide of the bacterial cell by inv19 gene product impossible. The intervention can typically involve

- 19 -

deletion of the inv19 gene as described herein. Other interventions are also contemplated, including introduction of other inv19 gene mutations that result in nonfunctional inv19 gene products, introduction of inv19 antisense nucleic acid or other agent that bind to inv19 nucleic acid molecule, as well as introduction into the bacterial cell of agents that can 5 interfere with the function of otherwise functional inv19 gene product.

A "pure capsular polysaccharide" refers to a capsular polysaccharide of the invention that (1) has been separated from at least about 50 percent of proteins, lipids, carbohydrates, or other materials with which it is naturally found when total capsular polysaccharide is isolated from the source cells, and (2) is substantially free of all other capsular polysaccharides. 10 "Substantially free of all other capsular polysaccharides" means the capsular polysaccharide of interest represents at least 80 percent of all capsular polysaccharides present. Preferably, the capsular polysaccharide of interest represents at least 85 percent, more preferably at least 90 percent, even more preferably at least 95 percent, and most preferably at least 98 percent, of all capsular polysaccharides present. In one embodiment, a pure capsular polysaccharide 15 of the present invention is substantially free from any contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or research use.

"Inflammatory bowel disease" refers to a group of chronic inflammatory disorders of unknown or autoimmune cause involving the gastrointestinal tract. These disorders are well 20 known in the medical arts and include two principal categories, ulcerative colitis and Crohn's disease. Ulcerative colitis characteristically occurs as continuous lesions in the colon involving the mucosa alone, whereas Crohn's disease characteristically occurs as discontinuous lesions anywhere in the gastrointestinal tract, involving inflammation of all layers of the bowel wall. While the etiology of inflammatory bowel disease remains 25 uncertain, there is evidence to suggest there may be an underlying infectious etiology, and effective treatments include the use of local or systemic immunomodulatory agents. See Glickman RM in: Harrison's Principles of Internal Medicine, 14th Ed., Fauci AS et al., eds., New York: McGraw-Hill, 1998, Chapter 286.

It is believed by the inventors that live or viable bacteria overexpressing PSA can be 30 used to treat or prevent inflammatory bowel disease. The bowel is normally colonized by bacteria of many species, including *B. fragilis*. Therefore introduction into the bowel of live or viable bacteria overexpressing PSA is expected to be well tolerated. However, due to the

- 20 -

previously described beneficial effects of PSA in inflammatory bowel disease, believed to be related to the ability of PSA to induce the anti-inflammatory cytokine interleukin-10 (IL-10), and the ability these bacteria to overexpress PSA, it is expected that these live or viable bacteria can be introduced into the bowel and thus treat and prevent inflammatory bowel
5 disease. The live or viable bacteria overexpressing PSA can be administered to subjects in need of treatment for inflammatory bowel disease by mouth or per rectum. Their effect can be determined by following disease activity in the usual manner. Furthermore, the live or
10 viable bacteria overexpressing PSA can be administered in conjunction with any other therapeutic agent useful in the treatment of inflammatory bowel disease, except, of course, antibiotics to which the live or viable bacteria overexpressing PSA are sensitive. Viable
15 bacteria specifically include lyophilized bacteria that are capable of growth upon their return to suitable conditions.

Res02 and Inv19 Nucleic Acids and Polypeptides

15 The term "polypeptide allelic variant" refers to one or several possible naturally occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or a population of organisms.

20 The term "isolated nucleic acid molecule" refers to a nucleic acid molecule of the invention that (1) has been separated from at least about 50 percent of proteins, lipids, carbohydrates, or other materials with which it is naturally found when total nucleic acid is isolated from the source cells, (2) is not linked to all or a portion of a polynucleotide to which the "isolated nucleic acid molecule" is linked in nature, (3) is operably linked to a polynucleotide to which it is not linked in nature, or (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated nucleic acid molecule of the present
25 invention is substantially free from any other nucleic acid molecule(s) or other contaminants that are found in its natural environment that would interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use.

25 The term "nucleic acid sequence" or "nucleic acid molecule" refers to a DNA or RNA sequence. The term encompasses molecules formed from any of the natural bases (adenine, cytosine, guanine, thymine, uracil) as well as base analogs of DNA and RNA such as, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinyl-cytosine, pseudoisocytosine, 5-(carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-

- 21 -

carboxymethylaminomethyl-2-thiouracil, 5-carboxy-methylaminomethyluracil, dihydrouracil, inosine, N6-iso-pentenyladenine, 1-methyladenine, 1-methylpseudoouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonyl-methyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, 2,6-diaminopurine, and any combination thereof. In some embodiments the nucleic acid sequence or nucleic acid molecule can have a modified backbone, e.g., a backbone characterized by at least one internucleoside linkage that is other than a phosphodiester bond. In one embodiment the modified backbone includes at least one stabilized internucleoside linkage, e.g., a phosphorthioate or phosphorodithioate linkage.

15 The term "vector" is used to refer to any molecule (e.g., nucleic acid, plasmid, or virus) used to transfer coding information to a host cell.

The term "expression vector" refers to a vector that is suitable for transformation of a host cell and contains nucleic acid sequences that direct and/or control the expression of inserted heterologous nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and RNA splicing, if introns are present.

20 The term "operably linked" is used herein to refer to an arrangement of gene expression sequence and coding sequence wherein transcriptional and/or translational control elements of the gene expression sequence and open reading frame of the coding sequence are covalently linked in such a way as to place the expression or transcription and/or translation of the coding sequence under the influence or control of the gene expression sequence. Thus, two DNA sequences are said to be operably linked if induction of a promoter in the 5' gene expression sequence results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequence, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a gene expression sequence would be operably linked to a res02 or inv19 nucleic acid sequence if the gene expression sequence

- 22 -

were capable of effecting transcription of the res02 or inv19 nucleic acid sequence such that the resulting transcript is translated into the desired protein or polypeptide. A gene expression sequence need not be contiguous with the coding sequence, so long as it functions correctly. Thus, for example, intervening untranslated yet transcribed sequences can be 5 present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence. Furthermore, an enhancer can be upstream or downstream of the coding sequence, and the enhancer need not be contiguous with the coding sequence.

The term "host cell" is used to refer to a cell which has been transformed, or is 10 capable of being transformed with a nucleic acid sequence and then of expressing a selected gene of interest. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic make-up to the original parent, so long as the selected gene is present.

The term "res02 polypeptide fragment" refers to a polypeptide that comprises a 15 truncation at the amino-terminus and/or a truncation at the carboxyl-terminus of the polypeptide as set forth in SEQ ID NO:2. The term "res02 polypeptide fragment" also refers to amino-terminal and/or carboxyl-terminal truncations of res02 polypeptide orthologs, res02 polypeptide derivatives, or res02 polypeptide variants, or to amino-terminal and/or carboxyl-terminal truncations of the polypeptides encoded by res02 polypeptide allelic variants or 20 res02 polypeptide splice variants. In certain embodiments, truncations and/or deletions comprise about 10 amino acids, or about 20 amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The polypeptide fragments so produced will comprise contiguous amino acids numbering about 187, or about 177, or about 147, or about 122, or about 97, or down to about 9, including 25 every integer therebetween. Such res02 polypeptide fragments can optionally comprise an amino-terminal methionine residue. It will be appreciated that such fragments can be used, for example, to generate antibodies to res02 polypeptides.

The term "inv19 polypeptide fragment" refers to a polypeptide that comprises a truncation at the amino-terminus and/or a truncation at the carboxyl-terminus of the 30 polypeptide as set forth in SEQ ID NO:4. The term "inv19 polypeptide fragment" also refers to amino-terminal and/or carboxyl-terminal truncations of inv19 polypeptide orthologs, inv19 polypeptide derivatives, or inv19 polypeptide variants, or to amino-terminal and/or carboxyl-

- 23 -

terminal truncations of the polypeptides encoded by inv19 polypeptide allelic variants or inv19 polypeptide splice variants. In certain embodiments, truncations and/or deletions comprise about 10 amino acids, or about 20 amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The 5 polypeptide fragments so produced will comprise contiguous amino acids numbering about 296, or about 286, or about 256, or about 231, or about 206, or down to about 9, including every integer therebetween. Such inv19 polypeptide fragments can optionally comprise an amino-terminal methionine residue. It will be appreciated that such fragments can be used, for example, to generate antibodies to inv19 polypeptides.

10 The term "res02 polypeptide ortholog" refers to a polypeptide from another species that corresponds to res02 polypeptide amino acid sequence as set forth in SEQ ID NO:2.

The term "inv19 polypeptide ortholog" refers to a polypeptide from another species that corresponds to inv19 polypeptide amino acid sequence as set forth in SEQ ID NO:4.

15 The term "res02 polypeptide variants" refers to res02 polypeptides comprising amino acid sequences having one or more amino acid sequence substitutions, deletions (such as internal deletions and/or res02 polypeptide fragments), and/or additions (such as internal additions and/or res02 fusion polypeptides) as compared to the res02 polypeptide amino acid sequence set forth in SEQ ID NO:2. Variants can be naturally occurring (e.g., res02 polypeptide allelic variants and res02 polypeptide orthologs) or artificially constructed. Such 20 res02 polypeptide variants can be prepared from the corresponding nucleic acid molecules having a DNA sequence that varies accordingly from the DNA sequence as set forth in SEQ ID NO:1. In certain embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 1 to 100, or more than 100 amino acid substitutions, insertions, additions and/or deletions, 25 wherein the substitutions can be conservative, or non-conservative, or any combination thereof.

The term "inv19 polypeptide variants" refers to inv19 polypeptides comprising amino acid sequences having one or more amino acid sequence substitutions, deletions (such as internal deletions and/or inv19 polypeptide fragments), and/or additions (such as internal 30 additions and/or inv19 fusion polypeptides) as compared to the inv19 polypeptide amino acid sequence set forth in SEQ ID NO:4. Variants can be naturally occurring (e.g., inv19 polypeptide allelic variants and inv19 polypeptide orthologs) or artificially constructed. Such

- 24 -

inv19 polypeptide variants can be prepared from the corresponding nucleic acid molecules having a DNA sequence that varies accordingly from the DNA sequence as set forth in SEQ ID NO:3. In certain embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 5 1 to 100, or more than 100 amino acid substitutions, insertions, additions and/or deletions, wherein the substitutions can be conservative, or non-conservative, or any combination thereof.

The term "res02 polypeptide derivatives" refers to the polypeptide as set forth in SEQ ID NO:2, res02 polypeptide fragments, res02 polypeptide orthologs, or res02 polypeptide 10 variants, as defined herein, that have been chemically modified. The term "res02 polypeptide derivatives" also refers to the polypeptides encoded by res02 polypeptide allelic variants, as defined herein, that have been chemically modified.

The term "inv19 polypeptide derivatives" refers to the polypeptide as set forth in SEQ ID NO:4, inv19 polypeptide fragments, inv19 polypeptide orthologs, or inv19 polypeptide 15 variants, as defined herein, that have been chemically modified. The term "inv19 polypeptide derivatives" also refers to the polypeptides encoded by inv19 polypeptide allelic variants, as defined herein, that have been chemically modified.

The term "res02 fusion polypeptide" refers to a fusion of one or more amino acids (such as a heterologous protein or peptide) at the amino- or carboxyl-terminus of the 20 polypeptide as set forth in SEQ ID NO:2, res02 polypeptide fragments, res02 polypeptide orthologs, res02 polypeptide variants, or res02 derivatives, as defined herein. The term "res02 fusion polypeptide" also refers to a fusion of one or more amino acids at the amino- or carboxyl-terminus of the polypeptide encoded by res02 polypeptide allelic variants or res02 polypeptide splice variants, as defined herein.

25 The term "inv19 fusion polypeptide" refers to a fusion of one or more amino acids (such as a heterologous protein or peptide) at the amino- or carboxyl-terminus of the polypeptide as set forth in SEQ ID NO:4, inv19 polypeptide fragments, inv19 polypeptide orthologs, inv19 polypeptide variants, or inv19 derivatives, as defined herein. The term "inv19 fusion polypeptide" also refers to a fusion of one or more amino acids at the amino- or 30 carboxyl-terminus of the polypeptide encoded by inv19 polypeptide allelic variants or inv19 polypeptide splice variants, as defined herein.

- 25 -

The term "biologically active res02 polypeptides" refers to res02 polypeptides having at least one activity characteristic of the polypeptide comprising the amino acid sequence of SEQ ID NO:2. In one embodiment the activity is promoter invertase activity. In addition, a res02 polypeptide can be active as an immunogen; that is, the res02 polypeptide contains at least one epitope to which antibodies may be raised.

The term "biologically active inv19 polypeptides" refers to inv19 polypeptides having at least one activity characteristic of the polypeptide comprising the amino acid sequence of SEQ ID NO:4. In addition, an inv19 polypeptide can be active as an immunogen; that is, the inv19 polypeptide contains at least one epitope to which antibodies may be raised.

The term "isolated polypeptide" refers to a polypeptide of the present invention that (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates, or other materials with which it is naturally found when isolated from the source cell, (2) is not linked (by covalent or noncovalent interaction) to all or a portion of a polypeptide to which the "isolated polypeptide" is linked in nature, (3) is operably linked (by covalent or noncovalent interaction) to a polypeptide with which it is not linked in nature, or (4) does not occur in nature. Preferably, the isolated polypeptide is substantially free from any other polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or research use.

The term "identity," as known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between nucleic acid molecules or polypeptides, as the case may be, as determined by the match between strings of two or more nucleotide or two or more amino acid sequences. "Identity" measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (i.e., "algorithms"). Percent identity with respect to nucleic acid molecules can be conveniently determined using a sequence alignment algorithm or computer program such as GAP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm. Percent identity with respect to polypeptides can be conveniently determined using a sequence alignment algorithm or computer program such as GAP, BLASTP, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

- 26 -

The term "similarity" is a related concept, but in contrast to "identity," "similarity" refers to a measure of relatedness which includes both identical matches and conservative substitution matches. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and 5 similarity would both be 50%. If in the same example, there are five more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% (15/20). Therefore, in cases where there are conservative substitutions, the percent similarity between two polypeptides will be higher than the percent identity between those two polypeptides.

10 Except as used in connection with the term "isolated", the term "naturally occurring" or "native" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to materials as they are found in nature and not manipulated by man. When used in connection with the term "isolated", the term "naturally occurring" or "native" when used in connection with biological materials such as nucleic acid 15 molecules, polypeptides, host cells, and the like, refers to isolated materials sharing wild type structural features as they are found in nature. For example, an isolated naturally occurring res02 polypeptide is in one embodiment an isolated wild type res02 polypeptide having the sequence of SEQ ID NO:2. The term "non-naturally occurring" or "non-native" as used herein refers to a material that is not found in nature or that has been structurally modified or 20 synthesized by man.

The term "effective amount" refers to that amount of a substance that is useful in accomplishing the purpose for which it is used. The term "therapeutically effective amount" refers to that amount of a substance that is useful in accomplishing the therapeutic effect for which it is used. Thus a therapeutically effective amount of a substance is that amount of the 25 substance that, when administered to a subject to treat or prevent a condition or disease of the subject, prevents the onset of, alleviates the symptoms of, or stops the progression of the condition or disease of the subject.

The term "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refers to one or more formulation materials suitable for accomplishing 30 or enhancing the delivery of an agent as a pharmaceutical composition.

The term "antigen" refers to a molecule or a portion of a molecule capable of being bound by a selective binding agent, such as an antibody, and additionally capable of being

- 27 -

used in an animal to produce antibodies capable of binding to an epitope of that antigen. An antigen can have one or more epitopes.

The term "selective binding agent" refers to a molecule or molecules having specificity for an antigen, e.g., a res02 or inv19 polypeptide. As used herein, the terms 5 "specific" and "specificity" refer to the ability of the selective binding agents to bind to an antigen of interest and not to bind to other antigens.

The term "transduction" is used to refer to the transfer of genes from one bacterium to another, usually by a phage.

The term "transfection" is used to refer to the uptake of foreign or exogenous DNA by 10 a cell, and a cell has been "transfected" when the exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are well known in the art and are disclosed herein. See, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual (2nd ed., Cold Spring Harbor Laboratory Press, 1989). Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells.

15 The term "transformation" as used herein refers to a change in a cell's genetic characteristics, and a cell has been transformed when it has been modified to contain a new DNA. For example, a cell is transformed where it is genetically modified from its native state. Following transfection or transduction, the transforming DNA can recombine with that of the cell by physically integrating into a chromosome of the cell, can be maintained 20 transiently as an episomal element without being replicated, or can replicate independently as a plasmid. A cell is considered to have been stably transformed when the DNA is replicated with the division of the cell.

Relatedness of Nucleic Acid Molecules and/or Polypeptides

25 It is understood that related nucleic acid molecules include allelic variants of the nucleic acid molecule of either SEQ ID NO:1 or SEQ ID NO:3, and include sequences which are complementary to any of the above nucleotide sequences. Related nucleic acid molecules also include a nucleotide sequence encoding a polypeptide comprising or consisting essentially of a substitution, modification, addition and/or deletion of one or more amino acid 30 residues compared to the polypeptide in either SEQ ID NO:2 or SEQ ID NO:4.

Related nucleic acid molecules also include fragments of res02 or inv19 nucleic acid molecules which encode a polypeptide of at least about 9 contiguous amino acids, or about 50

- 28 -

amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids, or more than about 200 amino acid residues of the res02 or inv19 polypeptide of either SEQ ID NO:2 or SEQ ID NO:4.

In addition, related res02 or inv19 nucleic acid molecules also include those

- 5 molecules which comprise nucleotide sequences which hybridize under moderately or highly stringent conditions as defined herein with the fully complementary sequence of the res02 or inv19 nucleic acid molecule of either SEQ ID NO:1 or SEQ ID NO:3, respectively, or of a molecule encoding a polypeptide, which polypeptide comprises the amino acid sequence as shown in either SEQ ID NO:2 or SEQ ID NO:4, or of a nucleic acid fragment as defined
- 10 herein, or of a nucleic acid fragment encoding a polypeptide as defined herein. Hybridization probes can be prepared using the res02 or inv19 sequences provided herein to screen cDNA, genomic or synthetic DNA libraries for related sequences. Regions of the DNA and/or amino acid sequence of res02 or inv19 polypeptide that exhibit significant identity to known sequences are readily determined using sequence alignment algorithms as described herein
- 15 and those regions can be used to design probes for screening.

The term "nucleic acid fragment" refers to a nucleic acid molecule sharing in common with a sequence from which it derives any sequence of at least 16 contiguous bases that is at least one base shorter than the sequence from which it derives. The nucleic acid fragment can be 16, 17, 18, 19, 20, and so on, bases long, up to but not including the total

- 20 number of bases of the sequence from which it derives. For example, in some embodiments a nucleic acid fragment of SEQ ID NO:1 is a 16-mer defined by bases 1-16, 2-17, 3-18, 4-19, and so on through bases 579-594 of SEQ ID NO:1. As a further example, in some embodiments a nucleic acid fragment of SEQ ID NO:1 is a 17-mer defined by bases 1-17, 2-18, 3-19, 4-120, and so on through bases 578-594 of SEQ ID NO:1. In certain preferred
- 25 embodiments a nucleic acid fragment encodes an immunogenic peptide of a corresponding polypeptide encoded by the full length nucleic acid sequence from which the fragment is derived. Since immunogenic peptides are generally recognized to be at least 9 amino acids long, the preferred nucleic acid fragments encoding the immunogenic peptides are at least 27 bases long. Nucleic acid fragments of the invention are also useful as primers for polymerase
- 30 chain reaction amplification of target DNA and for use as probes in DNA hybridization. In some embodiments a nucleic acid fragment further contains noncontiguous sequence, flanked by contiguous sequence. For example, the fragment can add at least one base, delete at least

- 29 -

one base, or change at least one base, and still be useful as a primer or probe. Such methods are well known in the art for the purposes of introducing a restriction endonuclease site to a PCR amplification product, introducing a mutation within a specific sequence, etc.

The term "highly stringent conditions" refers to those conditions that are designed to

5 permit hybridization of DNA strands whose sequences are highly complementary, and to exclude hybridization of significantly mismatched DNAs. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of "highly stringent conditions" for hybridization and washing are 0.015 M sodium chloride, 0.0015 M sodium citrate at 65-68°C or 0.015 M
10 sodium chloride, 0.0015 M sodium citrate, and 50% formamide at 42°C. See Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd ed., Cold Spring Harbor Laboratory Press, 1989); Anderson et al., Nucleic Acid Hybridisation: A Practical Approach Ch. 4 (IRL Press Limited).

15 More stringent conditions (such as higher temperature, lower ionic strength, higher formamide, or other denaturing agent) can also be used, although the rate of hybridization will be affected. Other agents can be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinyl-pyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or
20 another non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization experiments are usually carried out at pH 6.8-7.4; however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH. See Anderson et al.,
25 Nucleic Acid Hybridisation: A Practical Approach Ch. 4 (IRL Press Limited).

Factors affecting the stability of DNA duplex include base composition, length, and degree of base pair mismatch. Hybridization conditions can be adjusted by one skilled in the art in order to accommodate these variables and allow DNAs of different sequence relatedness to form hybrids. The melting temperature of a perfectly matched DNA duplex
30 can be estimated by the following equation:

$$T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log[\text{Na}^+]) + 0.41(\%G+C) - 600/N - 0.72(\%\text{formamide})$$

- 30 -

where N is the length of the duplex formed, $[Na^+]$ is the molar concentration of the sodium ion in the hybridization or washing solution, and %G+C is the percentage of (guanine+cytosine) bases in the hybrid. For imperfectly matched hybrids, the melting 5 temperature is reduced by approximately 1°C for each 1% mismatch.

The term "moderately stringent conditions" refers to conditions under which a DNA duplex is able to form with a greater degree of base pair mismatching than could occur under "highly stringent conditions". Examples of typical "moderately stringent conditions" are 0.015 M sodium chloride, 0.0015 M sodium citrate at 50-65°C or 0.015 M sodium chloride, 10 0.0015 M sodium citrate, and 20% formamide at 37-50°C. By way of example, "moderately stringent conditions" of 50°C in 0.015 M sodium ion will allow about a 21% mismatch.

It will be appreciated by those skilled in the art that there is no absolute distinction between "highly stringent conditions" and "moderately stringent conditions." For example, at 0.015 M sodium ion (no formamide), the melting temperature of perfectly matched long 15 DNA is about 71°C. With a wash at 65°C (at the same ionic strength), this would allow for approximately a 6% mismatch. To capture more distantly related sequences, one skilled in the art can simply lower the temperature or raise the ionic strength.

A good estimate of the melting temperature in 1M NaCl, e.g., in 6X salt sodium citrate (SSC), for oligonucleotide probes up to about 20 nucleotides is given by:

20

$$T_m = 2^\circ C \text{ per A-T base pair} + 4^\circ C \text{ per G-C base pair}$$

High stringency washing conditions for oligonucleotides are usually at a temperature of 0-5°C below the T_m of the oligonucleotide in 6X SSC, 0.1% SDS.

25 In another embodiment, related nucleic acid molecules comprise or consist of a nucleotide sequence that is at least about 70 percent identical to the nucleotide sequence as shown in either SEQ ID NO:1 or SEQ ID NO:3, or comprise or consist essentially of a nucleotide sequence encoding a polypeptide that is at least about 70 percent identical to the polypeptide as set forth in either SEQ ID NO:2 or SEQ ID NO:4. In certain embodiments, 30 the nucleotide sequences are about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical to the nucleotide sequence as shown in either SEQ ID NO:1 or SEQ ID NO:3, or the nucleotide sequences encode a

- 31 -

polypeptide that is about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical to the polypeptide sequence as set forth in either SEQ ID NO:2 or SEQ ID NO:4. Related nucleic acid molecules encode polypeptides possessing at least one activity of the polypeptide set forth in either SEQ ID 5 NO:2 or SEQ ID NO:4.

Differences in the nucleic acid sequence can result in conservative and/or non-conservative modifications of the amino acid sequence relative to the amino acid sequence of either SEQ ID NO:2 or SEQ ID NO:4. Conservative modifications to the amino acid sequence of either SEQ ID NO:2 or SEQ ID NO:4 (and the corresponding modifications to 10 the encoding nucleotides) will produce a polypeptide having functional and chemical characteristics similar to those of res02 or inv19 polypeptides. In contrast, substantial modifications in the functional and/or chemical characteristics of res02 or inv19 polypeptides can be accomplished by selecting substitutions in the amino acid sequence of either SEQ ID NO:2 or SEQ ID NO:4 that differ significantly in their effect on maintaining (a) the structure 15 of the molecular backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

For example, a "conservative amino acid substitution" can involve a substitution of a 20 native amino acid residue with a normative residue such that there is little or no effect on the polarity or charge of the amino acid residue at that position. Furthermore, any native residue in the polypeptide can also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis". Conservative amino acid substitutions also encompass non-naturally occurring amino acid residues that are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics, and 25 other reversed or inverted forms of amino acid moieties.

Naturally occurring residues can be divided into the following classes based on 30 common side chain properties: (1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile; (2) neutral hydrophilic: Cys, Ser, Thr; (3) acidic: Asp, Glu; (4) basic: Asn, Gln, His, Lys, Arg; (5) residues that influence chain orientation: Gly, Pro; and (6) aromatic: Trp, Tyr, Phe.

Thus, for example, conservative amino acid substitutions can involve the exchange of 35 a member from one of these classes for another member from the same class. Non-

- 32 -

conservative amino acid substitutions can involve the exchange of a member of one of these classes for a member from another class.

In making such changes, the hydropathy index of amino acids can be considered.

Each amino acid has been assigned a hydropathy index on the basis of its hydrophobicity and 5 charge characteristics. The hydropathy indices are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

10 The importance of the amino acid hydropathy index in conferring interactive biological function on a protein is generally understood in the art. Kyte J et al. (1982) *J Mol Biol* 157:105-32. It is known that certain amino acids can be substituted for other amino acids having a similar hydropathy index or score and still retain a similar biological activity. In making changes based upon the hydropathy index, the substitution of amino acids whose 15 hydropathy indices are within +2 is preferred, those which are within +1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biologically functionally equivalent protein or peptide thereby created is intended for use in immunological 20 embodiments, as in the present case. The greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, i.e., with a biological property of the protein.

The following hydrophilicity values have been assigned to these amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 \pm 1); glutamate (+3.0 \pm 1); serine (+0.3); 25 asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 \pm 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); and tryptophan (-3.4). In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are particularly 30 preferred, and those within ± 0.5 are even more particularly preferred. One can also identify epitopes from primary amino acid sequences on the basis of hydrophilicity. These regions are also referred to as "epitopic core regions".

- 33 -

Desired amino acid substitutions (whether conservative or non-conservative) can be determined by those skilled in the art at the time such substitutions are desired. For example, amino acid substitutions can be used to identify important residues of the res02 or inv19 polypeptide, or to increase or decrease the affinity of the res02 or inv19 polypeptides 5 described herein. Exemplary amino acid substitutions are set forth in **Table 2**.

Table 2. Amino Acid Substitutions.

| ORIGINAL RESIDUES | EXEMPLARY SUBSTITUTIONS | PREFERRED SUBSTITUTIONS |
|-------------------|---|-------------------------|
| Ala | Val, Leu, Ile | Val |
| Arg | Lys, Gln, Asn | Lys |
| Asn | Gln | Gln |
| Asp | Glu | Glu |
| Cys | Ser, Ala | Ser |
| Gln | Asn | Asn |
| Glu | Asp | Asp |
| Gly | Pro, Ala | Ala |
| His | Asn, Gln, Lys, Arg | Arg |
| Ile | Leu, Val, Met, Ala, Phe, Norleucine | Leu |
| Leu | Ile, Val, Met, Ala, Phe, Norleucine | Ile |
| Lys | Arg, 1,4 Diamino-butyric Acid, Gln, Asn | Arg |
| Met | Leu, Phe, Ile | Leu |
| Phe | Leu, Val, Ile, Ala, Tyr | Leu |
| Pro | Ala, Gly | Gly |
| Ser | Thr, Ala, Cys | Thr |
| Thr | Ser | Ser |
| Trp | Tyr, Phe | Tyr |
| Tyr | Trp, Phe, Thr, Ser | Phe |
| Val | Ile, Met, Leu, Phe, Ala, Norleucine | Leu |

A skilled artisan will be able to determine suitable variants of the polypeptide as set 10 forth in either SEQ ID NO:2 or SEQ ID NO:4 using well-known techniques. For identifying suitable areas of the molecule that can be changed without destroying biological activity, one skilled in the art can target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other species are

- 34 -

known, one skilled in the art can compare the amino acid sequence of a res02 or inv19 polypeptide to such similar polypeptides. With such a comparison, one can identify residues and portions of the molecules that are conserved among similar polypeptides. It will be appreciated that changes in areas of the res02 or inv19 molecule that are not conserved 5 relative to such similar polypeptides would be less likely to adversely affect the biological activity and/or structure of a res02 or inv19 polypeptide. One skilled in the art would also know that, even in relatively conserved regions, one can substitute chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions). Therefore, even areas that may be important for biological activity or 10 for structure can be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides that are important for activity or structure. In view of such a comparison, one can predict the importance of amino acid residues in a res02 or inv19 15 polypeptide that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art can opt for chemically similar amino acid substitutions for such predicted important amino acid residues of res02 or inv19 polypeptides.

In other embodiments, related nucleic acid molecules comprise or consist of a nucleotide sequence encoding a polypeptide as set forth in either SEQ ID NO:2 or SEQ ID 20 NO:4 with at least one amino acid insertion and wherein the polypeptide has an activity of the polypeptide set forth in either SEQ ID NO:2 or SEQ ID NO:4, or a nucleotide sequence encoding a polypeptide as set forth in either SEQ ID NO:2 or SEQ ID NO:4 with at least one amino acid deletion and wherein the polypeptide has an activity of the polypeptide set forth in either SEQ ID NO:2 or SEQ ID NO:4. Related nucleic acid molecules also comprise or 25 consist of a nucleotide sequence encoding a polypeptide as set forth in either SEQ ID NO:2 or SEQ ID NO:4 wherein the polypeptide has a carboxyl- and/or amino-terminal truncation and further wherein the polypeptide has an activity of the polypeptide set forth in either SEQ ID NO:2 or SEQ ID NO:4. Related nucleic acid molecules also comprise or consist of a nucleotide sequence encoding a polypeptide as set forth in either SEQ ID NO:2 or SEQ ID 30 NO:4 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, carboxyl-terminal truncations, and

- 35 -

amino-terminal truncations and wherein the polypeptide has an activity of the polypeptide set forth in either SEQ ID NO:2 or SEQ ID NO:4.

In addition, the polypeptide comprising the amino acid sequence of either SEQ ID NO:2 or SEQ ID NO:4, or other res02 or inv19 polypeptide, can be fused to a homologous 5 polypeptide to form a homodimer or to a heterologous polypeptide to form a heterodimer. Heterologous peptides and polypeptides include, but are not limited to: an epitope to allow for the detection and/or isolation of a res02 or inv19 fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor 10 protein; an enzyme or portion thereof which is catalytically active; a polypeptide or peptide which promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region; and a polypeptide which has a therapeutic activity different from the polypeptide comprising the amino acid sequence as set forth in either SEQ ID NO:2 or SEQ ID NO:4, or other res02 or inv19 15 polypeptide.

Fusions can be made either at the amino-terminus or at the carboxyl-terminus of the polypeptide comprising the amino acid sequence set forth in either SEQ ID NO:2 or SEQ ID NO:4, or other res02 or inv19 polypeptide. Fusions can be direct with no linker or adapter molecule, or they can be through a linker or adapter molecule. A linker or adapter molecule 20 can be one or more amino acid residues, typically from about 20 to about 50 amino acid residues. A linker or adapter molecule can also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for the separation of the fused moieties. It will be appreciated that once constructed, the fusion polypeptides can be derivatized according to the methods described herein.

25 In a further embodiment of the invention, the polypeptide comprising the amino acid sequence of either SEQ ID NO:2 or SEQ ID NO:4, or other res02 or inv19 polypeptide, is fused to one or more domains of an Fc region of human IgG. Antibodies comprise two functionally distinct parts, a variable domain known as "Fab" that binds an antigen, and a constant domain known as "Fc" that is involved in effector functions such as complement 30 activation and attack by phagocytic cells. An Fc has a long serum half-life, whereas an Fab is short-lived. Capon DJ et al. (1989) *Nature* 337:525-31. When constructed together with a therapeutic protein, an Fc domain can provide longer half-life or incorporate such functions

- 36 -

as Fc receptor binding, protein A binding, complement fixation, and perhaps even placental transfer.

In one example, a human IgG hinge, CH2, and CH3 region can be fused at either the amino-terminus or carboxyl-terminus of the res02 polypeptides using methods known to the skilled artisan. In another example, a human IgG hinge, CH2, and CH3 region can be fused at either the amino-terminus or carboxyl-terminus of a res02 polypeptide fragment.

The resulting fusion polypeptide may be purified by use of a Protein A affinity column. Peptides and proteins fused to an Fc region have been found to exhibit a substantially greater half-life in vivo than the unfused counterpart. Also, a fusion to an Fc region allows for dimerization/multimerization of the fusion polypeptide. The Fc region can be a naturally occurring Fc region, or it can be altered to improve certain qualities, such as therapeutic qualities, circulation time, or reduced aggregation.

Identity and similarity of related nucleic acid molecules and polypeptides are readily calculated by known methods. Such methods include, but are not limited to those described in Computational Molecular Biology (A. M. Lesk, ed., Oxford University Press 1988); Biocomputing: Informatics and Genome Projects (D. W. Smith, ed., Academic Press 1993); Computer Analysis of Sequence Data (Part 1, A. M. Griffin and H. G. Griffin, eds., Humana Press 1994); G. von Heinle, Sequence Analysis in Molecular Biology (Academic Press 1987); Sequence Analysis Primer (M. Gribskov and J. Devereux, eds., M. Stockton Press 1991); and Carillo et al. (1988) *SIAM J Applied Math* 48:1073.

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are described in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux J et al. (1984) *Nucleic Acids Res* 12(1 Pt 1):387-95; Genetics Computer Group, University of Wisconsin, Madison, Wis.), BLASTP, BLASTN, and FASTA (Altschul SF et al. (1990) *J Mol Biol* 215:403-10). The BLASTX program is publicly available from the National Center for Biotechnology Information (NCBI) and other sources (Altschul et al., BLAST Manual (NCB NLM NIH, Bethesda, MD); Altschul SF et al. (1990) *J Mol Biol* 215:403-10). The well-known Smith Waterman algorithm can also be used to determine identity.

- 37 -

Certain alignment schemes for aligning two amino acid sequences can result in the matching of only a short region of the two sequences, and this small aligned region can have very high sequence identity even though there is no significant relationship between the two full-length sequences. Accordingly, in one embodiment, the selected alignment method (e.g., 5 GAP program) will result in an alignment that spans at least 50 contiguous amino acids of the claimed polypeptide.

For example, using the computer algorithm GAP (Genetics Computer Group, University of Wisconsin, Madison, Wis.), two polypeptides for which the percent sequence identity is to be determined are aligned for optimal matching of their respective amino acids 10 (the "matched span," as determined by the algorithm). A gap opening penalty (which is calculated as 3X the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the particular comparison matrix) and a gap extension penalty (which is usually 0.1X the gap opening penalty), as well as a comparison matrix such as 15 PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix is also used by the algorithm. See: Dayhoff et al., 5 Atlas of Protein Sequence and Structure (Supp. 3 1978) (PAM250 comparison matrix); Henikoff S et al. (1992) *Proc Natl Acad Sci USA* 89:10915-19 (BLOSUM 62 comparison matrix).

Preferred parameters for polypeptide sequence comparison include the following:

20 Algorithm: Needleman SB et al. (1970) *J Mol Biol* 48:443-53; Comparison matrix: BLOSUM 62 (Henikoff et al, supra); Gap Penalty: 12; Gap Length Penalty: 4; Threshold of Similarity: 0.

The GAP program is useful with the above parameters. The aforementioned 25 parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparison include the following: Algorithm: Needleman et al., supra; Comparison matrix: matches=+10, mismatch=0; Gap Penalty: 50; Gap Length Penalty: 3.

The GAP program is also useful with the above parameters. The aforementioned 30 parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties, comparison matrices, and thresholds of similarity can be used, including those set forth in the

- 38 -

Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will be apparent to those of skill in the art and will depend on the specific comparison to be made, such as DNA-to-DNA, protein-to-protein, protein-to-DNA; and additionally, whether the comparison is between given pairs of sequences (in which case 5 GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

Nucleic Acid Molecules

The nucleic acid molecules encoding a polypeptide comprising the amino acid 10 sequence of a res02 or inv19 polypeptide can readily be obtained in a variety of ways including, without limitation, chemical synthesis, genomic library screening, expression library screening, and/or PCR amplification of genomic DNA.

Recombinant DNA methods used herein are generally those set forth in Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd ed., Cold Spring Harbor Laboratory Press, 15 1989) and/or Current Protocols in Molecular Biology (Ausubel et al., eds., Green Publishers Inc. and Wiley and Sons 1994). The invention provides for nucleic acid molecules as described herein and methods for obtaining such molecules.

Where a gene encoding the amino acid sequence of a res02 or inv19 polypeptide has been identified from one species, all or a portion of that gene can be used as a probe to 20 identify orthologs or related genes from the same species. The probes or primers can be used to screen genomic DNA from various bacteria believed to express the res02 or inv19 polypeptide. In addition, part or all of a nucleic acid molecule having the sequence as set forth in either SEQ ID NO:1 or SEQ ID NO:3 can be used to screen a genomic library to identify and isolate a gene encoding the amino acid sequence of a res02 or inv19 polypeptide. 25 Typically, conditions of moderate or high stringency will be employed for screening to minimize the number of false positives obtained from the screening.

Nucleic acid molecules encoding the amino acid sequence of res02 or inv19 polypeptides can also be identified by expression cloning which employs the detection of 30 positive clones based upon a property of the expressed protein. Typically, nucleic acid libraries are screened by the binding an antibody or other binding partner (e.g., receptor or ligand) to cloned proteins that are expressed and displayed on a host cell surface. The

- 39 -

antibody or binding partner is modified with a detectable label to identify those cells expressing the desired clone.

Recombinant expression techniques conducted in accordance with the descriptions set forth below can be followed to produce these polynucleotides and to express the encoded 5 polypeptides. For example, by inserting a nucleic acid sequence that encodes the amino acid sequence of a res02 or inv19 polypeptide into an appropriate vector, one skilled in the art can readily produce large quantities of the desired nucleotide sequence. The sequences can then be used to generate detection probes or amplification primers. Alternatively, a 10 polynucleotide encoding the amino acid sequence of a res02 or inv19 polypeptide can be inserted into an expression vector. By introducing the expression vector into an appropriate host, the encoded res02 or inv19 polypeptide can be produced in large amounts.

Another method for obtaining a suitable nucleic acid sequence is the polymerase chain reaction (PCR). In this method, DNA template is prepared from an appropriate source. Two primers, typically complementary to two separate regions of DNA encoding the amino 15 acid sequence of a res02 or inv19 polypeptide, are then added to the template DNA along with a polymerase such as Taq polymerase, and the polymerase amplifies the DNA region between the two primers.

Another means of preparing a nucleic acid molecule encoding the amino acid sequence of a res02 or inv19 polypeptide is chemical synthesis using methods well known to 20 the skilled artisan such as those described by Engels et al. (1989) *Angew Chem Intl Ed* 28:716-34. These methods include, *inter alia*, the phosphotriester, phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the amino acid sequence of a res02 or inv19 polypeptide will 25 be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these methods. The fragments can then be ligated, together to form the full-length nucleotide sequence of a res02 or inv19 gene. Other methods known to the skilled artisan can be used as well.

In certain embodiments, nucleic acid variants contain codons which have been altered 30 for optimal expression of a res02 or inv19 polypeptide in a given host cell. Particular codon alterations will depend upon the res02 or inv19 polypeptide and host cell selected for expression. Such "codon optimization" can be carried out by a variety of methods, for

- 40 -

example, by selecting codons which are preferred for use in highly expressed genes in a given host cell. Computer algorithms which incorporate codon frequency tables such as "Eco_high.Cod" for codon preference of highly expressed bacterial genes can be used and are provided by the University of Wisconsin Package Version 9.0 (Genetics Computer Group, 5 Madison, Wis.). Other useful codon frequency tables include "Celegans_high.cod," "Celegans_low.cod," "Drosophila_high.cod," "Human_high.cod," "Maize_high.cod," and "Yeast_high.cod."

In some cases, it may be desirable to prepare nucleic acid molecules encoding res02 or inv19 polypeptide variants. Nucleic acid molecules encoding variants can be produced 10 using site directed mutagenesis, PCR amplification, or other appropriate methods, where the primer(s) have the desired point mutations (see Sambrook et al., *supra*, and Ausubel et al., *supra*, for descriptions of mutagenesis techniques). Chemical synthesis using methods described by Engels et al., *supra*, can also be used to prepare such variants. Other methods known to the skilled artisan may be used as well.

15

Vectors and Host Cells

A nucleic acid molecule encoding the amino acid sequence of a res02 or inv19 polypeptide is inserted into an appropriate expression vector using standard ligation techniques. The vector is typically selected to be functional in the particular host cell 20 employed (i.e., the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). A nucleic acid molecule encoding the amino acid sequence of a res02 or inv19 polypeptide can be amplified/expressed in prokaryotic, yeast, insect (baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend in part on whether a res02 or inv19 polypeptide is to be post- 25 translationally modified (e.g., glycosylated and/or phosphorylated). If so, yeast, insect, or mammalian host cells are preferable. For a review of expression vectors, see *Meth. Enz.*, vol. 185 (D. V. Goeddel, ed., Academic Press 1990).

Typically, expression vectors used in any of the host cells will contain sequences for 30 plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as "flanking sequences" in certain embodiments will typically include one or more of the following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a

- 41 -

complete intron sequence containing a donor and acceptor splice site, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element. For expression of res02 or inv19, expression vectors used in any of the host cells typically will include one or more of the following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element.

10 Optionally, the vector can contain a "tag"-encoding sequence, i.e., an oligonucleotide molecule located at the 5' or 3' end of the res02 or inv19 polypeptide coding sequence; the oligonucleotide sequence encodes polyHis (such as hexaHis), or another "tag" such as FLAG, HA (hemagglutinin influenza virus), or myc for which commercially available antibodies exist. This tag is typically fused to the polypeptide upon expression of the polypeptide, and 15 can serve as a means for affinity purification of the res02 or inv19 polypeptide from the host cell. Affinity purification can be accomplished, for example, by column chromatography using antibodies against the tag as an affinity matrix. Optionally, the tag can subsequently be removed from the purified res02 or inv19 polypeptide by various means such as using certain peptidases for cleavage.

20 Flanking sequences can be homologous (i.e., from the same species and/or strain as the host cell), heterologous (i.e., from a species other than the host cell species or strain), hybrid (i.e., a combination of flanking sequences from more than one source), or synthetic, or the flanking sequences can be native sequences which normally function to regulate res02 or inv19 polypeptide expression. As such, the source of a flanking sequence can be any 25 prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequence is functional in, and can be activated by, the host cell machinery.

30 Flanking sequences useful in the vectors of this invention can be obtained by any of several methods well known in the art. Typically, flanking sequences useful herein, other than the res02 or inv19 gene flanking sequences, will have been previously identified by mapping and/or by restriction endonuclease digestion and can thus be isolated from the proper source using the appropriate restriction endonucleases. In some cases, the full

- 42 -

nucleotide sequence of a flanking sequence may be known. Here, the flanking sequence can be synthesized using the methods described herein for nucleic acid synthesis or cloning.

Where all or only a portion of the flanking sequence is known, it can be obtained using PCR and/or by screening a genomic library with a suitable oligonucleotide and/or

5 flanking sequence fragment from the same or another species. Where the flanking sequence is not known, a fragment of DNA containing a flanking sequence can be isolated from a larger piece of DNA that may contain, for example, a coding sequence or even another gene or genes. Isolation can be accomplished by restriction endonuclease digestion to produce the proper DNA fragment followed by isolation using agarose gel purification, Qiagen® column chromatography (Chatsworth, Calif.), or other methods known to the skilled artisan. The 10 selection of suitable enzymes to accomplish this purpose will be readily apparent to one of ordinary skill in the art.

An origin of replication is typically a part of those prokaryotic expression vectors purchased commercially, and the origin aids in the amplification of the vector in a host cell.

15 Amplification of the vector to a certain copy number can, in some cases, be important for the optimal expression of a res02 or inv19 polypeptide. If the vector of choice does not contain an origin of replication site, one may be chemically synthesized based on a known sequence, and ligated into the vector. For example, the origin of replication from the plasmid pBR322 (New England Biolabs, Beverly, Mass.) is suitable for most gram-negative bacteria and 20 various origins (e.g., SV40, polyoma, adenovirus, vesicular stomatitis virus (VSV), or papillomaviruses such as HPV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors (for example, the SV40 origin is often used only because it contains the early promoter).

25 A transcription termination sequence is typically located 3' of the end of a polypeptide coding region and serves to terminate transcription. Usually, a transcription termination sequence in prokaryotic cells is a G-C rich fragment followed by a poly-T sequence. While the sequence is easily cloned from a library or even purchased commercially as part of a vector, it can also be readily synthesized using methods for nucleic acid synthesis such as 30 those described herein.

A selectable marker gene element encodes a protein necessary for the survival and growth of a host cell grown in a selective culture medium. Typical selection marker genes

- 43 -

5 encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, tetracycline, or kanamycin for prokaryotic host cells; (b) complement auxotrophic deficiencies of the cell; or (c) supply critical nutrients not available from complex media. A neomycin resistance gene can also be used for selection in prokaryotic and eukaryotic host cells.

Other selection genes can be used to amplify the gene that will be expressed. Amplification is the process wherein genes that are in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable selectable markers for mammalian 10 cells include dihydrofolate reductase (DHFR) and thymidine kinase. The mammalian cell transformants are placed under selection pressure wherein only the transformants are uniquely adapted to survive by virtue of the selection gene present in the vector. Selection pressure is imposed by culturing the transformed cells under conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to 15 the amplification of both the selection gene and the DNA that encodes a res02 or inv19 polypeptide. As a result, increased quantities of res02 or inv19 polypeptide are synthesized from the amplified DNA.

A ribosome binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). 20 The element is typically located 3' to the promoter and 5' to the coding sequence of a res02 or inv19 polypeptide to be expressed. The Shine-Dalgarno sequence is varied but is typically a polypurine (i.e., having a high A-G content). Many Shine-Dalgarno sequences have been identified, each of which can be readily synthesized using methods set forth herein and used in a prokaryotic vector.

25 A leader, or signal, sequence can be used to direct a res02 or inv19 polypeptide out of the host cell. Typically, a nucleotide sequence encoding the signal sequence is positioned in the coding region of a res02 or inv19 nucleic acid molecule, or directly at the 5' end of a res02 or inv19 polypeptide coding region. Many signal sequences have been identified, and any of those that are functional in the selected host cell can be used in conjunction with a 30 res02 or inv19 nucleic acid molecule. Additionally, a signal sequence can be chemically synthesized using methods described herein. In most cases, the secretion of a res02 or inv19 polypeptide from the host cell via the presence of a signal peptide will result in the removal

- 44 -

of the signal peptide from the secreted res02 or inv19 polypeptide. The signal sequence can be a component of the vector, or it can be a part of a res02 or inv19 nucleic acid molecule that is inserted into the vector.

Included within the scope of this invention is the use of either a nucleotide sequence 5 encoding a native res02 or inv19 polypeptide without a signal sequence or a nucleotide sequence encoding a heterologous signal sequence joined to a res02 or inv19 polypeptide coding region. The heterologous signal sequence selected should be one that is recognized and processed, i.e., cleaved by a signal peptidase, by the host cell. For prokaryotic host cells, the signal sequence is provided by a prokaryotic signal sequence selected, for example, from 10 the group of the alkaline phosphatase, penicillinase, or heat-stable enterotoxin II leaders. For yeast secretion, the signal sequence can be provided by the yeast invertase, alpha factor, or acid phosphatase leaders. In mammalian cell expression any suitable mammalian signal sequence can be employed.

In some cases, such as where glycosylation is desired in a eukaryotic host cell 15 expression system, one can manipulate the various presequences to improve glycosylation or yield. For example, one can alter the peptidase cleavage site of a particular signal peptide, or add pro-sequences, which also can affect glycosylation. The final protein product can have, in the -1 position (relative to the first amino acid of the mature protein) one or more additional amino acids incident to expression, which may not have been totally removed. For 20 example, the final protein product can have one or two amino acid residues found in the peptidase cleavage site, attached to the amino-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired res02 or inv19 polypeptide, if the enzyme cuts at such area within the mature polypeptide.

The expression and cloning vectors of the present invention will typically contain a 25 promoter that is recognized by the host organism and operably linked to the molecule encoding the res02 or inv19 polypeptide. Promoters are untranscribed sequences located upstream (i.e., 5') to the start codon of a structural gene (generally within about 100 to 1000 bp) that control the transcription of the structural gene. Promoters are conventionally grouped into one of two classes: inducible promoters and constitutive promoters. Inducible 30 promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. Constitutive promoters, on the other hand, initiate continual gene

- 45 -

product production; that is, there is little or no control over gene expression. A large number of promoters, recognized by a variety of potential host cells, are well known. A suitable promoter is operably linked to the DNA encoding res02 or inv19 polypeptide by removing the promoter from the source DNA by restriction enzyme digestion and inserting the desired 5 promoter sequence into the vector. The native res02 or inv19 promoter sequence can be used to direct amplification and/or expression of a res02 or inv19 nucleic acid molecule. A heterologous promoter is preferred, however, if it permits greater transcription and higher yields of the expressed protein as compared to the native promoter, and if it is compatible with the host cell system that has been selected for use.

10 Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems; alkaline phosphatase; a tryptophan (trp) promoter system; and hybrid promoters such as the tac promoter. Other known bacterial promoters are also suitable. Their sequences have been published, thereby enabling one skilled in the art to ligate them to the desired DNA sequence, using linkers or adapters as needed to supply any 15 useful restriction sites.

Suitable promoters for use with yeast hosts are also well known in the art. Yeast enhancers are advantageously used with yeast promoters. Suitable promoters for use with mammalian host cells are well known and include, but are not limited to, those obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as 20 Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), retroviruses, hepatitis-B virus and Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, for example, heat-shock promoters and the actin promoter.

Additional promoters which may be of interest in controlling res02 or inv19 gene 25 expression include, but are not limited to: the SV40 early promoter region (Benoist C et al. (1981) *Nature* 290:304-10); the CMV promoter; the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto T et al. (1980) *Cell* 22:787-97); the herpes thymidine kinase promoter (Wagner MJ et al. (1981) *Proc Natl Acad Sci USA* 78:1441-45); the regulatory sequences of the metallothionein gene (Brinster RL et al. (1982) *Nature* 30: 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa- 30 Kamaroff et al. (1978) *Proc Natl Acad Sci USA* 75:3727-31); or the tac promoter (DeBoer et al. (1983) *Proc Natl Acad Sci USA* 80:21-25). Also of interest are the following animal

- 46 -

transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region which is active in pancreatic acinar cells (Swift et al. (1984) *Cell* 38:639-46; Omitz et al. (1986) *Cold Spring Harbor Symp Quant Biol* 50:399-409; MacDonald (1987) *Hepatology* 7:425-515); the insulin gene control region 5 which is active in pancreatic beta cells (Hanahan (1985) *Nature* 315:115-22); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl R et al. (1984) *Cell* 38:647-58; Adams JM et al. (1985) *Nature* 318:533-38; Alexander WS et al. (1987) *Mol Cell Biol* 7:1436-44); the mouse mammary tumor virus control region which is 10 active in testicular, breast, lymphoid and mast cells (Leder P et al. (1986) *Cell* 45:485-95); the albumin gene control region which is active in liver (Pinkert CA et al. (1987) *Genes Dev* 1:268-76); the alpha-fetoprotein gene control region which is active in liver (Krumlauf R et al. (1985) *Mol Cell Biol* 5:1639-48; Hammer RE et al. (1987) *Science* 235:53-58); the alpha 15 1-antitrypsin gene control region which is active in the liver (Kelsey GD et al. (1987) *Genes Dev* 1:161-71); the beta-globin gene control region which is active in myeloid cells (Magram J et al. (1985) *Nature* 315:338-40; Kollias G et al. (1986) *Cell* 46:89-94); the myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead C et al. (1987) *Cell* 48:703-12); the myosin light chain-2 gene control region which is active in skeletal muscle (Shani M (1985) *Nature* 314:283-86); and the gonadotropin releasing 20 hormone gene control region which is active in the hypothalamus (Mason AJ et al. (1986) *Science* 234:1372-78).

An enhancer sequence can be inserted into the vector to increase the transcription of a DNA encoding a res02 or inv19 polypeptide of the present invention by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300 bp in length, that act on the promoter to increase transcription. Enhancers are relatively orientation and position 25 independent. They have been found 5' and 3' to the transcription unit. Several enhancer sequences available from mammalian genes are known (e.g., globin, elastase, albumin, alpha-fetoprotein and insulin). Typically, however, an enhancer from a virus will be used. The SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are exemplary enhancing elements for the activation of eukaryotic 30 promoters. While an enhancer can be spliced into the vector at a position 5' or 3' to a res02 or inv19 nucleic acid molecule, it is typically located at a site 5' from the promoter.

- 47 -

Expression vectors of the invention can be constructed from a starting vector such as a commercially available vector. Such vectors may or may not contain all of the desired flanking sequences. Where one or more of the flanking sequences described herein are not already present in the vector, they can be individually obtained and ligated into the vector.

5 Methods used for obtaining each of the flanking sequences are well known to one skilled in the art.

Preferred vectors for practicing this invention are those which are compatible with bacterial, insect, and mammalian host cells. Such vectors include, *inter alia*, pCRII, pCR3, and pcDNA3.1 (Invitrogen, San Diego, CA), pBSII (Stratagene, La Jolla, CA), pET15
10 (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacI, Invitrogen), pDSR-alpha (PCT Pub. No. WO 90/14363) and pFastBacDual (Gibco-BRL, Grand Island, NY).

Additional suitable vectors include, but are not limited to, cosmids, plasmids, or modified viruses, but it will be appreciated that the vector system must be compatible with
15 the selected host cell. Such vectors include, but are not limited to plasmids such as Bluescript® plasmid derivatives (a high copy number ColE1-based phagemid, Stratagene Cloning Systems, La Jolla, CA), PCR cloning plasmids designed for cloning Taq-amplified PCR products (e.g., TOPO™ TA Cloning® Kit, PCR2.1.RTM. plasmid derivatives, Invitrogen, Carlsbad, CA), and mammalian, yeast or virus vectors such as a baculovirus
20 expression system (pBacPAK plasmid derivatives, Clontech, Palo Alto, CA).

After the vector has been constructed and a nucleic acid molecule encoding a res02 or inv19 polypeptide has been inserted into the proper site of the vector, the completed vector can be inserted into a suitable host cell for amplification and/or polypeptide expression. The transformation of an expression vector for a res02 or inv19 polypeptide into a selected host
25 cell can be accomplished by well known methods including methods such as transfection, infection, calcium chloride, electroporation, microinjection, lipofection, DEAE-dextran method, or other known techniques. The method selected will in part be a function of the type of host cell to be used. These methods and other suitable methods are well known to the skilled artisan, and are set forth, for example, in Sambrook et al., *supra*.

30 Host cells can be prokaryotic host cells (such as *E. coli*) or eukaryotic host cells (such as a yeast, insect, or vertebrate cell). The host cell, when cultured under appropriate conditions, synthesizes a res02 or inv19 polypeptide which can subsequently be collected

- 48 -

from the culture medium (if the host cell secretes it into the medium) or directly from the host cell producing it (if it is not secreted). The selection of an appropriate host cell will depend upon various factors, such as desired expression levels, polypeptide modifications that are desirable or necessary for activity (such as glycosylation or phosphorylation) and ease of 5 folding into a biologically active molecule.

A number of suitable host cells are known in the art and many are available from the American Type Culture Collection (ATCC), Manassas, VA. Examples include, but are not limited to, mammalian cells, such as Chinese hamster ovary cells (CHO), CHO DHFR(-) cells (Urlaub G et al. (1980) *Proc Natl Acad Sci USA* 97:4216-20), human embryonic kidney 10 (HEK) 293 or 293T cells, or 3T3 cells. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production, and purification are known in the art. Other suitable mammalian cell lines, are the monkey COS-1 and COS-7 cell lines, and the CV-1 cell line. Further exemplary mammalian host cells include primate cell lines and rodent cell lines, including transformed cell lines. Normal 15 diploid cells, cell strains derived from in vitro culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include but are not limited to, mouse neuroblastoma N2A cells, HeLa, mouse L-929 cells, 3T3 lines derived from Swiss, BALB/c or NIH mice, BHK or HaK hamster cell lines. 20 Each of these cell lines is known by and available to those skilled in the art of protein expression.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of *E. coli* (e.g., HB101, DH5 α , DH10, and MC1061) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis*, *Pseudomonas* 25 spp., other *Bacillus* spp., *Streptomyces* spp., and the like can also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for the expression of the polypeptides of the present invention. Preferred yeast cells include, for example, *Saccharomyces cerevisiae* and *Pichia pastoris*.

Additionally, where desired, insect cell systems can be utilized in the methods of the 30 present invention. Such systems are described, for example, in Kitts PA et al. (1993) *Biotechniques* 14:810-17; Lucklow VA (1993) *Curr Opin Biotechnol* 4:564-72; and Lucklow VA et al. (1993) *J Virol* 67:4566-79. Preferred insect cells are Sf-9 and Hi5 (Invitrogen).

- 49 -

One can also use transgenic animals to express glycosylated res02 or inv19 polypeptides. For example, one can use a transgenic milk-producing animal (a cow or goat, for example) and obtain the present glycosylated polypeptide in the animal milk. One can also use plants to produce res02 or inv19 polypeptides, however, in general, the glycosylation occurring in plants is different from that produced in mammalian cells, and can result in a glycosylated product which is not suitable for human therapeutic use.

Polypeptide Production

Host cells comprising a res02 or inv19 polypeptide expression vector can be cultured using standard media well known to the skilled artisan. The media will usually contain all nutrients necessary for the growth and survival of the cells. Suitable media for culturing *E. coli* cells include, for example, Luria Broth (LB) and/or Terrific Broth (TB). Suitable media for culturing eukaryotic cells include Roswell Park Memorial Institute medium 1640 (RPMI 1640), Minimal Essential Medium (MEM) and/or Dulbecco's Modified Eagle Medium (DMEM), all of which can be supplemented with serum and/or growth factors as necessary for the particular cell line being cultured. A suitable medium for insect cultures is Grace's medium supplemented with yeastolate, lactalbumin hydrolysate, and/or fetal calf serum as necessary.

Typically, an antibiotic or other compound useful for selective growth of transfected or transformed cells is added as a supplement to the media. The compound to be used will be dictated by the selectable marker element present on the plasmid with which the host cell was transformed. For example, where the selectable marker element is kanamycin resistance, the compound added to the culture medium will be kanamycin. Other compounds for selective growth include ampicillin, tetracycline, and neomycin.

The amount of a res02 or inv19 polypeptide produced by a host cell can be evaluated using standard methods known in the art. Such methods include, without limitation, Western immunoblot analysis, SDS-polyacrylamide gel electrophoresis, non-denaturing gel electrophoresis, high performance liquid chromatography (HPLC) separation, immunoprecipitation, and/or activity assays such as DNA binding gel shift assays.

If a res02 or inv19 polypeptide has been designed to be secreted from the host cells, the majority of polypeptide may be found in the cell culture medium. If, however, the rcs02

- 50 -

or inv19 polypeptide is not secreted from the host cells, it will be present in the cytoplasm and/or the nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells).

For a res02 or inv19 polypeptide situated in the host cell cytoplasm and/or nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells), the intracellular material 5 (including inclusion bodies for gram-negative bacteria) can be extracted from the host cell using any standard technique known to the skilled artisan. For example, the host cells can be lysed to release the contents of the periplasm/cytoplasm by French press, homogenization, and/or sonication followed by centrifugation.

If a res02 or inv19 polypeptide has formed inclusion bodies in the cytosol, the 10 inclusion bodies can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with a chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as dithiothreitol at alkaline pH or tris carboxyethyl phosphine at acid pH to release, break apart, and solubilize 15 the inclusion bodies. The solubilized res02 or inv19 polypeptide can then be analyzed using gel electrophoresis, immunoprecipitation, or the like. If it is desired to isolate the res02 or inv19 polypeptide, isolation can be accomplished using standard methods such as those described herein and in Marston FA et al. (1990) *Methods Enzymol* 182:264-76.

In some cases, a res02 or inv19 polypeptide may not be biologically active upon 20 isolation. Various methods for "refolding" or converting the polypeptide to its tertiary structure and generating disulfide linkages can be used to restore biological activity. Such methods include exposing the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used 25 at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases the refolding/oxidation solution will also contain a reducing agent or the reducing agent plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridges. Some of the commonly used redox couples include cysteine/cystamine, 30 glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol (DTT)/dithiane DTT, and 2-mercaptoethanol (β ME)/dithio- β (ME). In many instances, a cosolvent can be used or can be needed to increase the efficiency of the refolding, and the more common reagents used for

- 51 -

this purpose include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

If inclusion bodies are not formed to a significant degree upon expression of a res02 or inv19 polypeptide, then the polypeptide will be found primarily in the supernatant after 5 centrifugation of the cell homogenate. The polypeptide can be further isolated from the supernatant using methods such as those described herein.

The purification of a res02 or inv19 polypeptide from solution can be accomplished using a variety of techniques. If the polypeptide has been synthesized such that it contains a tag such as Hexahistidine (res02 or inv19 polypeptide/hexaHis) or other small peptide such as 10 FLAG (Eastman Kodak Co., New Haven, CT) or myc (Invitrogen, Carlsbad, CA) at either its carboxyl- or amino-terminus, it can be purified in a one-step process by passing the solution through an affinity column where the column matrix has a high affinity for the tag.

For example, polyhistidine binds with great affinity and specificity to nickel. Thus, an affinity column of nickel (such as the Qiagen® nickel columns) can be used for 15 purification of res02 or inv19 polypeptide/polyHis. See, e.g., Current Protocols in Molecular Biology 10.11.8 (Ausubel et al., eds., Green Publishers, Inc. and Wiley and Sons, 1993).

Additionally, res02 or inv19 polypeptides can be purified through the use of a monoclonal antibody that is capable of specifically recognizing and binding to a res02 or inv19 polypeptide.

20 Other suitable procedures for purification include, without limitation, affinity chromatography, immunoaffinity chromatography, ion exchange chromatography, molecular sieve chromatography, HPLC, electrophoresis (including native gel electrophoresis) followed by gel elution, and preparative isoelectric focusing ("Isoprime" machine/technique, Hoefer Scientific, San Francisco, CA). In some cases, two or more purification techniques can be 25 combined to achieve increased purity.

Res02 or inv19 polypeptides can also be prepared by chemical synthesis methods (such as solid phase peptide synthesis) using techniques known in the art such as those set forth by Merrifield et al. (1963) *J Am Chem Soc* 85:2149; Houghten RA et al. (1985) *Proc Natl Acad Sci USA* 82:5131-35; and Stewart and Young, *Solid Phase Peptide Synthesis* 30 (Pierce Chemical Co., 1984). Such polypeptides can be synthesized with or without a methionine on the amino-terminus. Chemically synthesized res02 or inv19 polypeptides can be oxidized using methods set forth in these references to form disulfide bridges. Chemically

- 52 -

synthesized res02 or inv19 polypeptides are expected to have comparable biological activity to the corresponding res02 or inv19 polypeptides produced recombinantly or purified from natural sources, and thus can be used interchangeably with a recombinant or natural res02 or inv19 polypeptide.

5 Another means of obtaining res02 or inv19 polypeptide is via purification from biological samples such as capsular polysaccharide-expressing bacterial cells in which the res02 or inv19 polypeptide is naturally found. Such purification can be conducted using methods for protein purification as described herein. The presence of the res02 or inv19 polypeptide during purification can be monitored, for example, using an antibody prepared
10 against recombinantly produced res02 or inv19 polypeptide or peptide fragments thereof.

A number of additional methods for producing nucleic acids and polypeptides are known in the art, and the methods can be used to produce polypeptides having specificity for res02 or inv19 polypeptide. See, e.g., Roberts RW et al. (1997) *Proc Natl Acad Sci USA* 94:12297-302, which describes the production of fusion proteins between an mRNA and its
15 encoded peptide. See also, Roberts RW (1999) *Curr Opin Chem Biol* 3:268-73.

Additionally, U.S. Pat. No. 5,824,469 describes methods for obtaining oligonucleotides capable of carrying out a specific biological function. The procedure involves generating a heterogeneous pool of oligonucleotides, each having a 5' randomized sequence, a central preselected sequence, and a 3' randomized sequence. The resulting heterogeneous pool is
20 introduced into a population of cells that do not exhibit the desired biological function.

Subpopulations of the cells are then screened for those that exhibit a predetermined biological function. From that subpopulation, oligonucleotides capable of carrying out the desired biological function are isolated.

U.S. Pat. Nos. 5,763,192; 5,814,476; 5,723,323; and 5,817,483 describe processes for
25 producing peptides or polypeptides. This is done by producing stochastic genes or fragments thereof, and then introducing these genes into host cells which produce one or more proteins encoded by the stochastic genes. The host cells are then screened to identify those clones producing peptides or polypeptides having the desired activity.

Another method for producing peptides or polypeptides is described in
30 PCT/US98/20094 (WO99/15650) filed by Athersys, Inc. Known as "Random Activation of Gene Expression for Gene Discovery" (RAGE-GD), the process involves the activation of endogenous gene expression or over-expression of a gene by *in situ* recombination methods.

- 53 -

For example, expression of an endogenous gene is activated or increased by integrating a regulatory sequence into the target cell which is capable of activating expression of the gene by non-homologous or illegitimate recombination. The target DNA is first subjected to radiation, and a genetic promoter inserted. The promoter eventually locates a break at the 5 front of a gene, initiating transcription of the gene. This results in expression of the desired peptide or polypeptide.

It will be appreciated that these methods can also be used to create comprehensive res02 or inv19 polypeptide expression libraries, which can subsequently be used for high throughput phenotypic screening in a variety of assays, such as biochemical assays, cellular 10 assays, and whole organism assays (e.g., plant, mouse, etc.).

Selective Binding Agents

The term "selective binding agent" refers to a molecule that has specificity for one or more antigens selected from res02 polypeptide, inv19 polypeptide, PSA, PSB, PSC, PSD, 15 PSE, PSF, PSG, PSH, and fragments thereof. Suitable selective binding agents include, but are not limited to, antibodies and derivatives thereof, polypeptides, and small molecules. Suitable selective binding agents can be prepared using methods known in the art.

Selective binding agents such as antibodies and antibody fragments that bind res02 polypeptide, inv19 polypeptide, PSA, PSB, PSC, PSD, PSE, PSF, PSG, PSH, and fragments 20 thereof, are within the scope of the present invention. The antibodies can be polyclonal including monospecific polyclonal; monoclonal (MAbs); recombinant; chimeric; humanized, such as CDR-grafted; human; single chain; and/or bispecific; as well as fragments, variants, or derivatives thereof. Antibody fragments include those portions of the antibody that bind to an epitope of an antigen. Examples of such fragments include Fab, F(ab') and F(ab')₂ 25 fragments generated by enzymatic cleavage of full-length antibodies. Other binding fragments include those generated by recombinant DNA techniques, such as the expression of recombinant plasmids containing nucleic acid sequences encoding antibody variable regions.

Polyclonal antibodies directed toward an antigen generally are produced in animals 30 (e.g., rabbits or mice) by means of multiple subcutaneous or intraperitoneal injections of antigen and an adjuvant. It may be useful to conjugate an antigen to a carrier protein that is immunogenic in the species to be immunized, such as keyhole limpet hemocyanin, albumin,

- 54 -

bovine thyroglobulin, or soybean trypsin inhibitor. Also, aggregating agents such as alum are used to enhance the immune response. After immunization, the animals are bled and the serum is assayed for antigen-specific antibody titer.

Monoclonal antibodies directed toward an antigen are produced using any method 5 that provides for the production of antibody molecules by continuous cell lines in culture. Examples of suitable methods for preparing monoclonal antibodies include the hybridoma methods of Kohler G et al. (1975) *Nature* 256:495-97 and the human B-cell hybridoma method (Kozbor D et al. (1984) *J Immunol* 133:3001-5; Brodeur et al., *Monoclonal Antibody Production Techniques and Applications* 51-63 (Marcel Dekker, Inc., 1987). Also provided 10 by the invention are hybridoma cell lines that produce monoclonal antibodies reactive with res02 polypeptide, inv19 polypeptide, PSA, PSB, PSC, PSD, PSE, PSF, PSG, PSH, and fragments thereof.

The antibodies of the invention can be employed in any known assay method, such as 15 competitive binding assays, direct and indirect sandwich assays, Western immunoblot assays, and immunoprecipitation assays (Sola, *Monoclonal Antibodies: A Manual of Techniques*, CRC Press, Inc., 1987, pp. 147-158) for the detection and quantitation of antigen. The antibodies will bind antigen with an affinity that is appropriate for the assay method being employed.

For diagnostic applications, in certain embodiments, anti-res02, or anti-inv19, or anti- 20 capsular polysaccharide antibodies can be labeled with a detectable moiety. The detectable moiety can be any one that is capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety can be a radioisotope, such as ^3H , ^{14}C , ^{32}P , ^{35}S , or ^{125}I , a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, 25 rhodamine, or luciferin; or an enzyme, such as alkaline phosphatase, beta-galactosidase, or horseradish peroxidase (Bayer EA et al. (1990) *Methods Enzymol* 184:138-60).

Competitive binding assays rely on the ability of a labeled standard (e.g., a res02 or 30 inv19 polypeptide, a capsular polysaccharide, or an immunologically reactive portion thereof) to compete with the test sample analyte (a res02 or inv19 polypeptide or a capsular polysaccharide) for binding with a limited amount of anti-res02, anti-inv19, or anti-capsular polysaccharide antibody. The amount of an antigen in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies typically are

- 55 -

insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte which remain unbound.

Sandwich assays typically involve the use of two antibodies, each capable of binding

- 5 to a different immunogenic portion, or epitope, of the protein to be detected and/or quantitated. In a sandwich assay, the test sample analyte is typically bound by a first antibody which is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three-part complex. See, e.g., U.S. Pat. No. 4,376,110. The second antibody can itself be labeled with a detectable moiety (direct sandwich assays)
- 10 or can be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assays). For example, one type of sandwich assay is an enzyme-linked immunosorbent assay (ELISA), in which case the detectable moiety is an enzyme.

Examples

- 15 The following examples are illustrative only and are not intended to limit the scope of the invention in any way.

Example 1. Bacterial strain and isolation of *B. fragilis* polysaccharide.

- 20 *B. fragilis* NCTC9343 was originally obtained from the National Collection of Type Cultures (London, England), stored at -80°F in yeast both until used, and grown anaerobically as previously described. Pantosti A et al. (1991) *Infect Immun* 59:2075-2082. The capsular polysaccharide from *B. fragilis* NCTC9343 was isolated by hot phenol/water extraction and subsequent purification of PSA and PSB performed as previously described. Tzianabos A et al. (1992) *J Biol Chem* 267:18230-18235.

25

Example 2. Identification of *res02* and *inv19*.

- 30 Hypothesizing that the DNA inversions of the seven promoter regions would be controlled by specific proteins that are involved in recombining DNA, a search of *B. fragilis* NCTC9343 genomic sequences provided by the microbial pathogen group at the Sanger Centre was performed, looking for open reading frames (ORFs) with homology to *fimB* and *fimE*, two genes of *E. coli* that are involved in inverting DNA. Five ORFs were retrieved. These five *B. fragilis* ORFs were then used to reprobe the database, and twenty-five

- 56 -

homologs were retrieved and given temporary designations of *inv1* - *inv25*. Three additional homologs have subsequently been retrieved and given temporary designations of *inv26* - *inv28*.

Based on our data, we knew that the gene product that is involved in inverting the polysaccharide promoters would be conserved in all strains. Therefore, DNA hybridizations were performed to probe a collection of *B. fragilis* strains with internal portions of each of these 25 ORFs.

Seven of the original twenty-five ORFs (and nine of the twenty-eight ORFs) were found to be conserved within *B. fragilis*. Analysis of the sequence surrounding one of these conserved genes, *inv19*, demonstrated a divergently transcribed gene of the serine site-specific recombinase family which was designated *res02* (now also denoted *mpi*) and was also shown by hybridization to be conserved among *B. fragilis*.

Fifteen out of fifteen *B. fragilis* strains tested probed positive with internal portion of *res02* (Table 3). Twenty-five out of twenty-six *B. fragilis* strains tested probed positive with internal portion of *inv19*; the one negative strain was 1279155I (Table 4).

Table 3. *res02* probe-positive *B. fragilis* strains.

| | | |
|--------|-------|-------|
| 1279-2 | 45703 | CM3 |
| 13141 | B110 | CM11 |
| 17905 | B117 | CM12 |
| 2429 | B124 | PA5 |
| 26877 | B272 | US398 |

Table 4. *inv19* probe-positive *B. fragilis* strains.

| | | | | |
|-----------|------------|-------|----------|-----------|
| 12775L1II | 1285531I | 2429 | B117 | CM11 |
| 1277810I | 1287245I | 17905 | B124 | CM12 |
| 1279-2 | 12905-23V | 26877 | B272 | IL89375II |
| 1281262I | 1291662III | 45703 | B356772I | PA5 |
| 1284-2 | 13141 | B110 | CM33 | US398 |

20

Several of these conserved genes were cloned into a vector along with the PSA promoter invertible region and analyzed for their ability to invert the PSA promoter region. *Res02* was demonstrated to bring about inversion of the PSA promoter region.

Example 3. Method of deleting res02 from the *B. fragilis* 9343 chromosome.

In order to delete the *res02* gene from the *B. fragilis* 9343 chromosome, homologous recombination with a double crossover event was employed to replace the full length copy of *res02* with a deleted copy. To do this, plasmid pKGW10 was created. In order to construct this plasmid, 9343 chromosomal DNA was used as a template to amplify the DNA flanking the region to be deleted (Figure 5). Primer pairs inv19D-1 plus inv19D-2 and inv19D-5 plus inv19D-6 were used in two separate PCRs. Table 5 shows that primers inv19D-1 and inv19D-6 each incorporate a *Bam*HI restriction endonuclease site (shown underlined), while primers inv19D-2 and inv19D-5 each incorporate an *Nco*I restriction endonuclease site (shown underlined).

Table 5. Sequences of the primers used to make the *res02* deletion.

| | | | |
|----------|---|---|----------------|
| inv19-D1 | → | 5'- <u>ccggatcc</u> aggtaactgataactccgggtactcc-3' | (SEQ ID NO:11) |
| inv19-D2 | ← | 5'- <u>atccatggccgg</u> tttatgaaaacgtatgttta-3' | (SEQ ID NO:12) |
| inv19-D5 | → | 5'-cgccatggtttccgtacttactctcaaataagc-3' | (SEQ ID NO:13) |
| inv19-D6 | ← | 5'- <u>gggatcc</u> atgacatgataatggggaaagg-3' | (SEQ ID NO:14) |

Upon amplification, the resulting amplification products were: Left Flank, inv19-D1→ through ←inv19-D2, 1,958 bp (Figure 6, SEQ ID NO:5); and Right Flank, inv19-D5→ through ←inv19-D6, 2,540 bp (Figure 7, SEQ ID NO:6).

These primers were placed so that 534 bp of the 594 bp *res02* open reading frame were deleted. The resulting PCR products were digested with *Bam*HI and *Nco*I (restriction sites constructed into the primers) and ligated in a three-way reaction with a *B. fragilis* suicide vector that encoded for erythromycin resistance. This ligation mixture was transformed into *E. coli* and the transformants were tested by both PCR and plasmid digestion for the correct ligation of the flanks. The plasmid containing the correct ligation of the flanks was named pKGW10.

pKGW10 was conjugally transferred from *E. coli* to *B. fragilis* 9343 and the cointegrate (containing all of the DNA of pKGW10 integrated into the chromosome) resulting from homologous recombination was selected by erythromycin resistance (encoded by pKGW10). The cointegrate was passaged in nonselective medium to allow for resolution of the cointegrate by the second recombination event (where either the mutant copy or the wild type copy of *res02* was lost along with the intervening plasmid). Bacteria were plated

- 58 -

onto medium without antibiotics (approximately 200 colonies per plate). The colonies were replica plated to medium containing erythromycin and the erythromycin-sensitive (Em^s) colonies were selected (these Em^s colonies represent those that only have the mutant or wild type copy of *res02*, but not both). These colonies were screened by PCR for those containing 5 the mutant genotype. PCR was used to determine that each of the seven polysaccharide invertible promoter regions was locked and unable to invert in each of the *res02* mutants. Deletion was also confirmed by Southern blot in each case.

Example 4. Monospecific antisera for PSA-PSH.

10 Monospecific antisera for each of the eight known capsular polysaccharides of *B. fragilis* were prepared as previously described. Comstock LE et al. (1999) *Infect Immun* 67:3525-32; Coyne MJ et al. (2000) *Infect Immun* 68:6176-81; Coyne M et al. (2001) *Infect Immun* 69:4342-50; Krinos CM et al. (2001) *Nature* 414:555-8. These monospecific antisera were used in Western immunoblot phenotype analyses of *res02* mutants.

15 Example 5. Deletion of *res02* locks capsular polysaccharide promoters on or off.

The *res02* open reading frame was deleted from the 9343 chromosome by double crossover allelic exchange, resulting in several mutants that all had a chromosomal deletion of *res02*. Analysis of these mutants demonstrated that each of the seven polysaccharide 20 promoter inversion regions was locked in a single orientation, demonstrating that the Res02 product is involved in the DNA inversions.

Several of these mutants had the PSA and PSE promoter regions locked on and the promoter regions for PSB, PSD, PSF, PSG, and PSH locked off. Therefore, it was expected that this strain would constitutively express PSA, PSC (the promoter of PSC does not 25 undergo inversions) and PSE, but not express PSB, PSD, PSF, PSG, or PSH.

Figure 8 shows the results of phenotypic analysis of one of these mutants, 9343res02mu44 (mut44). This analysis demonstrated that mut44 indeed synthesized PSA in high quantities (as least ten times more purified PSA is isolated compared to a typical grow-up with wild-type). It was also shown that mut44 was unable to produce six of the other 30 seven capsular polysaccharides, including PSC and PSE (as determined by Western blot analysis, immunoelectrophoresis, and analysis of purified capsular polysaccharide). It was unexpectedly found that, despite the fact that the promoter for the polysaccharide

- 59 -

biosynthesis locus of PSF was turned off, mut44 expressed small amounts of PSF (Figure 8). This observation could be consistent with the existence of a secondary promoter that is unaffected by DNA inversions. However, another mutant, mut8, was discovered to express only PSH and no PSF (Figure 8). The small amount of PSF expressed by mut 44 does not 5 significantly interfere with the purification of PSA from this strain.

Another *resO2* mutant, 9343res02mut2 (mut2), was isolated that also had all of the polysaccharide promoter flip regions locked off except for the PSA promoter region, which was locked on. The phenotype of this mutant was found to be the same as mut44, i.e., found to overexpress PSA and to express small amounts of PSF, but none of the other seven known 10 capsular polysaccharides PSB-PSH.

Example 6. mut44 overexpresses PSA.

By deletion of the open reading frame designated *resO2*, *B. fragilis* strains have been created that overexpress PSA compared to wild type and that are devoid of most or all the 15 other seven known capsular polysaccharide of this strain. These strains make a sufficient amount of PSA to be easily purified to make them attractive for large scale purification of the potent zwitterionic polysaccharide, PSA, for commercial interests.

The yield of total capsular polysaccharide (all eight polysaccharides together) isolated from wild type *B. fragilis* NCTC9343 ranges from 6.26-21.9 mg/liter of culture. After 20 extensive methods for purification of PSA from the other polysaccharides, the yield of pure PSA isolated from wild type *B. fragilis* NCTC9343 ranges from 0-3.1 mg/liter of culture, with an average of 1.56 mg/liter of culture. In the first large grow-up (16 liters) of mut44 there was a yield of 21.8 mg pure PSA per liter of culture. Not only was the yield 14 times greater than from wild type, but the extraction and purification methods were much easier 25 because the PSA did not have to be isolated from the other capsular polysaccharides. Unlike wild type, the yield of PSA from this mutant is expected to be consistently high since expression of PSA is no longer undergoing phase variation due to the promoter being unable to flip off.

30 Example 7. Large scale purification of PSA from mut44.

A 16 liter batch fermentation of mut44 was grown in supplemented basal medium. Following overnight growth, the bacteria were pelleted and resuspended in 667 ml of dH₂O at

- 60 -

68°C. Glass beads were added to the resuspended pellet and placed in a 4 liter water bath. 667 ml of 75% phenol (prewarmed to 68°C) was added to the mixture and stirred for at least 30 min. The mixture was then stirred overnight in the cold room. The mixture was then centrifuged at 8000 rpm for 20 min. The top aqueous phase was removed (approx. 500ml) 5 and added to 500ml of ether in a separatory funnel. The mixture was shaken and allowed to separate for 20 min. The bottom phase was retained. The sample was placed in a rotary evaporator in a 60°C waterbath for ether evaporation and sample concentration. The sample was placed in dialysis tubing and dialyzed against 10 liters of water with six changes of buffer. 1M Tris was added to make it 6.5% of the volume. MgCl₂ and CaCl₂ were added to a 10 final concentration of 20mM. RNase was added to bring the concentration to 3.33 mg/ml and DNase was added to 0.07 mg/ml. The sample was digested overnight at 37°C. The pH of the sample was adjusted to 7.5 with 10M NaOH. Pronase was added to 0.33 mg/ml and incubated overnight. Fresh pronase was added and incubated for an additional 2 hours. 15 EDTA was added to make the solution 50mM and mixed for 30 min. The sample was precipitated using 5 volumes of -20°C ethanol. The sample was resuspended in 3% deoxycholate and applied to a sepharose 400 column. Column fractions were monitored by silver-stained SDS-PAGE for separation of high molecular weight capsular polysaccharide and low molecular weight LPS. PSA purity was tested by immunoelectrophoresis and NMR analysis.

20

Example 8. Deletion of *inv19*.

The *inv19* open reading frame was deleted from the 9343 chromosome by double crossover allelic exchange, resulting in several mutants that all have a chromosomal deletion of *inv19*.

25

Biological Deposit

A deposit of mut44 was made with American Type Culture Collection, Manassas, VA on March 11, 2002, under the description *Bacteroides fragilis*: 9343res02mut44, and assigned to Patent Deposit Designation PTA-4135.

30

2003218144 23 Jul 2008

- 61 -

All of the references, patents and patent publications identified or cited herein are incorporated, in their entirety by reference.

Although this invention has been described with respect to specific embodiments, the details of these embodiments are not to be construed as limitations. Various equivalents, changes and modifications may be made without departing from the spirit and scope of this invention, and it is understood that such equivalent embodiments are part of this invention.

Throughout this specification and claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers or steps but not the exclusion of any other integer or group of integers.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

We claim:

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence as set forth in SEQ ID NO:1; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2; (c) a nucleotide sequence which hybridizes under stringent conditions to a complement of (a) or (b); and (d) a nucleotide sequence complementary to any of (a)-(c).
2. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide which is at least about 70 percent identical to a polypeptide as set forth in SEQ ID NO:2, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) a nucleotide sequence encoding an allelic variant of a nucleotide sequence as set forth in SEQ ID NO: 1 or (a); (c) a region of the nucleotide sequence of SEQ ID NO:1, (a), or (b) encoding a polypeptide fragment of at least about 9 amino acid residues, wherein the polypeptide fragment has an activity of the encoded polypeptide as set forth in SEQ ID NO:2, or is antigenic; (d) a region of the nucleotide sequence of SEQ ID NO:1, or any of (a)-(c) comprising a fragment of at least about 16 nucleotides; (e) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(d); and (f) a nucleotide sequence complementary to any of (a)-(d).
3. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 which has a C- and/or N-terminal truncation,

wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (f) a nucleotide sequence of any of (a)-(e) comprising a fragment of at least about 16 nucleotides; (g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f); and (h) a nucleotide sequence complementary to any of (a)-(e).

4. An isolated polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2.

5. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) an amino acid sequence for an ortholog of SEQ ID NO:2; (b) an amino acid sequence which is at least about 70 percent identical to the amino acid sequence of SEQ ID NO:2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) a fragment of the amino acid sequence set forth in SEQ ID NO:2 comprising at least about 9 amino acid residues, wherein the fragment has an activity of the polypeptide set forth in SEQ ID NO:2, or is antigenic; and (d) an amino acid sequence for an allelic variant of the amino acid sequence as set forth in SEQ ID NO:2, (a), or (b).

6. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as set forth in SEQ ID NO:2 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (b) the amino acid sequence as set forth in SEQ ID NO:2 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (c) the amino acid sequence as set forth in SEQ ID NO:2 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; (d) the amino acid sequence as set forth in SEQ ID

NO:2 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2; and (e) the amino acid sequence as set forth in SEQ ID NO:2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2.

7. An isolated polypeptide encoded by the nucleic acid molecule of any of claims 1, 2, or 3, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:2.

8. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence as set forth in SEQ ID NO:3; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4; (c) a nucleotide sequence which hybridizes under stringent conditions to a complement of (a) or (b); and (d) a nucleotide sequence complementary to any of (a)-(c).

9. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide which is at least about 70 percent identical to a polypeptide as set forth in SEQ ID NO:4, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) a nucleotide sequence encoding an allelic variant of a nucleotide sequence as set forth in SEQ ID NO:3 or (a); (c) a region of the nucleotide sequence of SEQ ID NO:3, (a), or (b) encoding a polypeptide fragment of at least about 9 amino acid residues, wherein the polypeptide fragment has an activity of the encoded polypeptide as set forth in SEQ ID NO:4, or is antigenic; (d) a region of the nucleotide sequence of SEQ ID NO:3, or any of (a)-(c) comprising a fragment of at least about 16 nucleotides; (e) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(d); and (f) a nucleotide sequence complementary to any of (a)-(d).

10. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide as set forth in

2003218144 23 Jul 2008

SEQ ID NO:4 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (d) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 which has a C- and/or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (e) a nucleotide sequence encoding a polypeptide as set forth in SEQ ID NO:4 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (f) a nucleotide sequence of any of (a)-(e) comprising a fragment of at least about 16 nucleotides; (g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f); and (h) a nucleotide sequence complementary to any of (a)-(e).

11. A vector comprising the nucleic acid molecule of any one of claims 1-3 or 8-10.
12. A host cell comprising the vector of claim 11.
13. An isolated polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:4.
14. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) an amino acid sequence for an ortholog of SEQ ID NO:4; (b) an amino acid sequence which is at least about 70 percent identical to the amino acid sequence of SEQ ID NO:4, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) a fragment of the amino acid sequence set forth in SEQ ID NO:4 comprising at least about 9 amino acid residues, wherein the fragment has an activity

2003218144 23 Jul 2008

of the polypeptide set forth in SEQ ID NO:4, or is antigenic; and (d) an amino acid sequence for an allelic variant of the amino acid sequence as set forth in SEQ ID NO:4, (a), or (b).

15. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as set forth in SEQ ID NO:4 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (b) the amino acid sequence as set forth in SEQ ID NO:4 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (c) the amino acid sequence as set forth in SEQ ID NO:4 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; (d) the amino acid sequence as set forth in SEQ ID NO:4 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4; and (e) the amino acid sequence as set forth in SEQ ID NO:4 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4.

16. An isolated polypeptide encoded by the nucleic acid molecule of any of claims 8, 9, or 10, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO:4.

17. An isolated selective binding agent or fragment thereof which specifically binds the polypeptide of any of claims 4-6 or 13-15.

18. The selective binding agent or fragment thereof of claim 17 that specifically binds the polypeptide comprising the amino acid sequence as set forth in SEQ ID NO:2 or SEQ ID NO:4 or a fragment thereof.

19. The selective binding agent of claim 17 or 18 that is an antibody or fragment thereof.

20. A fusion polypeptide comprising the polypeptide of any of claims 4-6 or 13-15 fused to a heterologous amino acid sequence.
21. An isolated bacterial cell in which *res02* expression is disabled.
22. An isolated bacterial cell in which *inv19* expression is disabled.
23. A bacterial cell as claimed in claim 21 or 22, stably expressing a specific capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH.
24. A bacterial cell as claimed in claim 21 or 22, expressing a capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, wherein a promoter controlling expression of the capsular polysaccharide is locked on.
25. The bacterial cell according to claim 23 or 24, wherein the capsular polysaccharide is PSA.
26. The bacterial cell according to claim 25, further wherein promoters controlling expression of each and every capsular polysaccharide selected from the group consisting of: PSB, PSD, PSE, PSF, PSG, and PSH are locked off.
27. The bacterial cell according to claim 24, further wherein any one or combination of capsular polysaccharides selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH is not expressed.
28. The bacterial cell according to claim 24, further wherein a promoter controlling expression of any one or combination of capsular polysaccharides selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, is locked off.

2003218144 23 Jul 2008

- 68 -

29. The bacterial cell according to any one of claims 23-28, wherein res02 expression is disabled.

30. The bacterial cell according to any one of claims 21, 22 or 23-28, wherein the bacterial cell is *B. fragilis*.

31. The bacterial cell according to any one of claims 21, 22, or 23-28, wherein the bacterial cell is *B. fragilis* NCTC9343.

32. The bacterial cell according to any one of claims 21, 23, 24 or 25, wherein the bacterial cell is *B. fragilis* 9343res02mut44.

33. The bacterial cell according to any one of claims 21, 23, 24 or 25, wherein the bacterial cell is *B. fragilis* 9343res02mut2.

34. A method for locking a phase-variable promoter of a capsular polysaccharide biosynthesis gene on, comprising:
inactivating res02 in a bacterial cell to lock a phase-variable promoter of a biosynthesis gene for a capsular polysaccharide on.

35. The method according to claim 34, further comprising:
selecting a bacterial cell expressing at least one capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH.

36. The method according to claim 34, further comprising:
selecting a bacterial cell expressing only one specific capsular polysaccharide selected from the group consisting of PSA, PSB, PSD, PSE, PSF, PSG, and PSH.

37. A method for purifying PSA from a bacterial cell, the method comprising:
inactivating res02 in said bacterial cell such that a promoter for PSA is locked on or off;
and selecting a bacterial cell expressing PSA.

38. The method according to any one of claims 34-37, wherein the bacterial cell is *B. fragilis*.

39. The method according to any one of claims 34-37, wherein the cell is *B. fragilis* NCTC9343.

40. A method for producing a pure capsular polysaccharide, comprising:
growing a population of bacterial cells as claimed in claim 21 or 22 stably expressing a specific capsular polysaccharide selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH; and
isolating the specific capsular polysaccharide from the population of bacterial cells to produce a pure capsular polysaccharide.

41. The method according to claim 40, wherein a promoter controlling expression of any one or combination of capsular polysaccharides selected from the group consisting of: PSA, PSB, PSD, PSE, PSF, PSG, and PSH, excluding the specific capsular polysaccharide, is locked off.

42. The method according to claim 40 or 41, wherein the specific capsular polysaccharide is PSA.

43. The method according to claim 42, wherein promoters controlling expression of each and every capsular polysaccharide selected from the group consisting of: PSB, PSD, PSE, PSF, PSG, and PSH, are locked off.

44. A method for treating or preventing inflammatory bowel disease, comprising:
administering to a subject in need of treatment for or prevention of inflammatory bowel disease an effective amount of a population of bacterial cells as claimed in claim 21 or 22, stably expressing a specific capsular polysaccharide to treat or prevent the inflammatory bowel disease.

2003218144 23 Jul 2008

- 70 -

45. The method according to claim 36 or 44, wherein the specific capsular polysaccharide is PSA.

46. The method according to any one of claims 40-45, wherein the population of bacterial cells is *B. fragilis*.

47. The method according to claim 46, wherein the population of bacterial cells is *B. fragilis* NCTC9343.

48. The method according to claim 46, wherein the population of bacterial cells is *B. fragilis* 9343res02mut44.

49. The method according to claim 46, wherein the population of bacterial cells is *B. fragilis* 9343res02mut2.

50. An isolated nucleic acid molecule according to any one of claims 1-3 or 8-10, or an isolated polypeptide according to anyone of claims 4-7 or 13-16, or a vector according to claim 11, or a host cell according to claim 12, or a selective binding agent or fragment thereof according to any one of claims 17-19, or a fusion polypeptide according to claim 20, or a bacterial cell according to any one of claims 21-33, or a method according to any one of claims 34-49 substantially as hereinbefore described with reference to the Figures and/or Examples.

1/5

SEQ ID NO:1

atggtaatagcttatttggagactaaggacggaaaaacatgttttggtcaatcagaaggaa 60
gagatttatgcgttgcagagaagaatgggttgcgttgcatacgatgttgcacacagagacc 120
gtaagcggaaagcgtggccacaaaagacagaaagctatcagatgttgcatacgatgttgc 180
ccccgggatatactgttgcacggatgttgcagatgttgcacactgttgcagatgttgc 240
atgactatctgttgcatacgatgttgcacccaaagaggcc 300
tatgttgcgttgcacccaaagaggcc 360
gaaatagaagaaaggacactgttgcgttgcacccaaagaggcc 420
ggaaaggactgttgcgttgcacccaaagaggcc 480
aataaaggccgttgcgttgcacccaaagaggcc 540
aagatgggggtatccagaacaacccgttgcgttgcgttgcgttgcgttgcgttgcgttgc 594

Fig. 1

SEQ ID NO:2

| | |
|---|-----|
| MVIAYLRVSTEKQFLANQKEEIMRFAEKNGNLSIDKNYTETVSGSVSTKDRKLSELLKRMH | 60 |
| PGDTLIVITEISRLSRTLLEMTILNFCIKKQVVLYSTKEGYVFQDDINSKVLGFAGLMA | 120 |
| BIERNLISMRTKEALARRKQEGMTLGRKKGDTPKIKLLRANKRVLTKELDKGTTYSELAE | 180 |
| KMGVSRTTLFRFMKTMY | 197 |

Fig. 2

SEQ ID NO:3

Fig. 3

SUBSTITUTE SHEET (RULE 26)

SEQ ID NO:4

| | |
|---|-----|
| MEIEKFIFKSLARKAKLGGGRYSTANTYLTLHSFQKFAGKASLT | 60 |
| LNGKRYNTISLYMVRMLRSICNQASEQNIASLNTRELFENVFIGNEPTAKRAISPV | 120 |
| LEADFSKNSRLDFARDLFLLSFYLRGIPFVDLVHLRKTDVQGNMLVYFRQKTG | 180 |
| HNCAKVLRLKYASLCKESVYLLPVISAAGEEGHKQYRSALRVYNKRLNQISGILKLK | 240 |
| TSYVARHSHWATTALQKGVPVSVISAGMGHASEKVTTIYLASFDNKTLSNANKVIAVRF | 300 |
| KKEEEE | 306 |

Fig. 4

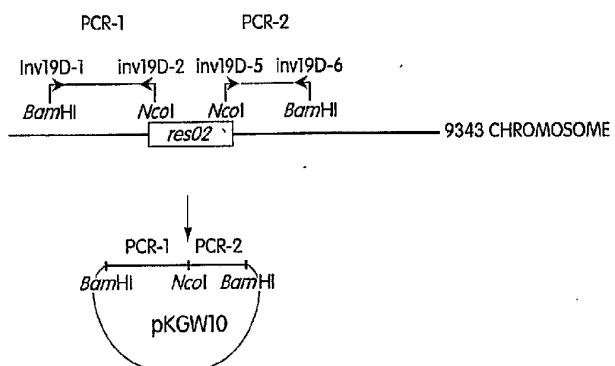


Fig. 5

SUBSTITUTE SHEET (RULE 26)

SEQ ID NO:5

```

agtactgataactccggtaactccggactgtbaaaagaaaatgcggagaagctccca 60
gggacaagacgttaaacgtacccgtgaaaaatgacacgggttctcgacaaattgcgg 120
aaccgaacttattacccatgttaggcataaaagacaaaaggaaacgcattgtcaa 180
catcggtgacgtaaaatacaaaatggctacaagccgctctcaacacagagaaggac 240
atctacccgttccatcgaaacccgttgcacccctcaactgcctgcagccgtagaagg 300
ccaggaaggacagactgttacgttaaccatccgttccggacaaaaggcagcttgc 360
caagaaaaatcgttgaagtcagcgcgataaaaggaaactggcaggcgttagc 420
gttacccgttccatcgacccgttgcacttgcgtatgaccaaaaaggaaatgtatt 480
tcaacacacgtaaacaagcgtatcataggggtaaaggctgtaaagacaggcttata 540
ccgcataaaaagtccgtctccccccggagtgaacacggacttctgttcttga 600
accattcaaaaaaaaaaaagattttcaacagcgcacatgcctttcataatcggctct 660
gagtaatctccggcaacagctgtataaggtacttccgttcccaatcaccacta 720
ccgcacgcgcacgcgtccggccacgggtcgtcagccatctcgcgcatagtctcgt 780
ogaagtccgaaaaggcgaaatccgcacacggatcagtttgcataccctgtcg 840
agaagcgtccctgcgcacacggatcagtttgcataccacacggatccttca 900
ttccggctgcacatattgaatttgcacccgaaatggcgcacacccggatccagac 960
tcggggaaaatttgcacaaatatttgcacatccgttgcataattttgc 1020
aatacttttgcacccgcgtcaatgcgggaccccttccaaacctgtataatgc 1080
tcagtttgcgggtgtccctgtgaaatttgcgttgcataattgtataactctaa 1140
atttactatattctaaacaaatccgtgtcaagaacttgcgttgcataattctaa 1200
aaataataattgtatgaaaacttattgcacggatggacacacgcgtcaaaaacttgc 1260
gttacccgttccatcgatgcgttgcacatcgttgcgttgcgttgcacc 1320
ggaaagcgttacattgcctcagcgtcatctcgcatttgcgttgcataatgc 1380
catcgaaatcatcttccatcgttaacccggcgggatctgttgcgttgcgttgcacc 1440
cgcaggccgtatcatcgatctgttgcgttgcgttgcgttgcgttgcacc 1500
catcgaaatcatccgttgcgttgcgttgcgttgcgttgcgttgcacc 1560
aggctattctatggacccgttgcgttgcgttgcgttgcgttgcgttgcacc 1620
cggcatccgcgcgttgcgttgcgttgcgttgcgttgcgttgcacc 1680
ctacgttataatatgcgttgcgttgcgttgcgttgcgttgcgttgcacc 1740
gtcccttcgttgcgttgcgttgcgttgcgttgcgttgcgttgcacc 1800
attgtatgcgttgcgttgcgttgcgttgcgttgcgttgcacc 1860
tctcgttgcgttgcgttgcgttgcgttgcgttgcacc 1920
ccccctaaaccgttactaatacatcgatccgttgcgttgcgttgcacc 1958

```

Fig. 6

SEQ ID NO:6

| | |
|--|------|
| tttccgtacttactctcaaataagctattaccataattcatgttttaatgattaata | 60 |
| caccacgaaaaaaacggctattcatcaaatacgggacactttttacgttcccttttct | 120 |
| catgcacactgggtattttggaaacttccatcgctatacatttatgttatttttttttct | 180 |
| taactatttccagcatatttccaatctgtcactcaaataattttttattataaaacccgtg | 240 |
| ttcttgaacacactaaaaaagaacaagaaaatggaaatagaaaaattcattaaatcttttag | 300 |
| caagaaaaagegaagtttagccgggggttacagcacagccaaatcccttacacttgc | 360 |
| acagtttcaagaatgttgcggggaaagccttactgtactttttaaagagatcactccgaga | 420 |
| gtatacaggtaacggacaatctaattcccaacggggaaacgggtacaacacgatctgc | 480 |
| tctacatgcgcgttgcgttccatctgcatacggcatcgaggcagaacatagtttgc | 540 |
| tcaacaccccgagatgttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 600 |
| ccatctaccggcttccatttcccgctgtcgaagcagatttgcgttgcgttgcgttgc | 660 |
| tcgattttggccgcgcaccccttttgcgttgcgttgcgttgcgttgcgttgcgttgc | 720 |
| acctggatcatctccgcacggatgtgcgttgcgttgcgttgcgttgcgttgcgttgc | 780 |
| aaacgggacgcaacttacggtaatcatagaaaaactgcgttgcgttgcgttgcgttgc | 840 |
| atgcgtcgcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 900 |
| aggggcacacggatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 960 |
| ccggaaatctggaaatggactccgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1020 |
| ccacggccctgcacaaagggttccgggttccgttgcgttgcgttgcgttgcgttgc | 1080 |
| cagagaagggtgcataatctggatgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1140 |
| ataaaaaatggatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1200 |
| tcttttacttatttagtaatagaacacaggatcttttgcgttgcgttgcgttgcgttgc | 1260 |
| aataattgaaactccacaaaatgataatttttctataaaaatggattataacc | 1320 |
| agttgaatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1380 |
| gtttgtatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1440 |
| ataatccaggccctgttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1500 |
| gctattccgcatttattactaataatggatgcgttgcgttgcgttgcgttgcgttgc | 1560 |
| taagggaacaaaattgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1620 |
| tatcgcaacatcttctaaatgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1680 |
| aaaaggattgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1740 |
| accctatt | 1800 |
| gtatttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1860 |
| cgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 1920 |
| ttgttt | 1980 |
| ttacttt | 2040 |
| atccctgaagtccatctgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2100 |
| aatagcaatataatcgacttaccatcaaatgttgcgttgcgttgcgttgcgttgc | 2160 |
| gttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2220 |
| ggaaaacccggtaagattgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2280 |
| tgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2340 |
| ctgtatacgcttttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2400 |
| gagcttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2460 |
| aaccggattcaggaaatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2520 |
| ttcccccattatctatgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc | 2540 |

Fig. 7

SUBSTITUTE SHEET (RULE 26)

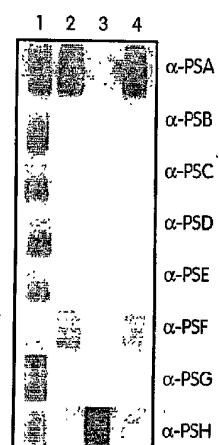


Fig. 8

SUBSTITUTE SHEET (RULE 26)

SEQUENCE LISTING

<110> The Brigham and Women's Hospital, Inc.

<120> METHOD FOR OVEREXPRESSION OF ZWITTERIONIC POLYSACCHARIDES

<130> B00801.70279

<150> US 60/364,168

<151> 2001-03-13

<160> 14

<170> PatentIn version 3.1

<210> 1

<211> 594

<212> DNA

<213> Bacteroides fragilis

<400> 1
atggtaatag cttatggag agtaagtacg gaaaaacagt ttttggctaa tcagaaggaa 60
gagattatgc gattgcaga gaagaatggg ttgtcgattt acaagtggta cacagagacc 120
gttageggaa gcgtgagcac aaaagacaga aagttatcg agttattgaa gagaatgcatt 180
cccgggata cactgattgt aacggagatt tcgagattga gccgtacact gctcgagatt 240
atgactatcc tgaattttg tattaaagaag caggtagtgc tctatagcac caaagaggc 300
tatgtgttcc aggacgacat caacagcaag gtgtcgatgc tccgttccgg actgtatggcg 360
gaaatagaaa ggaacctgtat ttcgatgcgt accaaagaag ctctcgacg cagaaagcag 420
gaaggaaatga ctttaggccc aaagaaaggg gatacgccta aaataaaatt gctgcgtgcc 480
aataagcgcg tacttaccaa agaacttgac aaaggaacta cttactcgga attggcggag 540
aagatggggg tatccagaac aaccctgttc cgggttatga aaacgtatgtat 594

<210> 2

<211> 197

<212> PRT

<213> Bacteroides fragilis

<400> 2

Met Val Ile Ala Tyr Leu Arg Val Ser Thr Glu Lys Gln Phe Leu Ala
1 5 10 15

Asn Gln Lys Glu Glu Ile Met Arg Phe Ala Glu Lys Asn Gly Leu Ser
20 25 30

Ile Asp Lys Trp Tyr Thr Glu Thr Val Ser Gly Ser Val Ser Thr Lys
35 40 45

Asp Arg Lys Leu Ser Glu Leu Leu Lys Arg Met His Pro Gly Asp Thr
 50 55 60

Leu Ile Val Thr Glu Ile Ser Arg Leu Ser Arg Thr Leu Leu Glu Ile
 65 70 75 80

Met Thr Ile Leu Asn Phe Cys Ile Lys Lys Gln Val Val Leu Tyr Ser
 85 90 95

Thr Lys Glu Gly Tyr Val Phe Gln Asp Asp Ile Asn Ser Lys Val Leu
 100 105 110

Gly Phe Ala Phe Gly Leu Met Ala Glu Ile Glu Arg Asn Leu Ile Ser
 115 120 125

Met Arg Thr Lys Glu Ala Leu Ala Arg Arg Lys Gln Glu Gly Met Thr
 130 135 140

Leu Gly Arg Lys Lys Gly Asp Thr Pro Lys Ile Lys Leu Leu Arg Ala
 145 150 155 160

Asn Lys Arg Val Leu Thr Lys Glu Leu Asp Lys Gly Thr Thr Tyr Ser
 165 170 175

Glu Leu Ala Glu Lys Met Gly Val Ser Arg Thr Thr Leu Phe Arg Phe
 180 185 190

Met Lys Thr Met Tyr
 195

<210> 3
 <211> 921
 <212> DNA
 <213> Bacteroides fragilis

<400> 3
 atggaaatag aaaaattcat taaatcttta gcaagaaaag cgaagttagg cggcggttac 60
 agcacagcca atacctacat ctacactttg cacagtttc agaagtttgc gggaaaagcc 120
 tcactgactt ttgaagagat cactcccgag agtataagg agtacgagca atacttaatc 180
 ctcaacggga aacgggtacaa cacgatctcg ctctacatgc gcatgttgcg ttccatctgc 240
 aatcaggccat cggagccgaa catagctcg ctcaacaccc gcgagctgtt tgagaatgtt 300
 tttatccgca acgaacccac tgccaagcgt gccatctcac ccgtccatcat ttcccgctg 360

ctcgaagcag atttcagcaa gaacagccgg ctcgattttg cccgcgacct cttcttgcta 420
 agcttctacc tgagggaaat cccgttgtc gacctggatc atctccgcaa gaccgtgtg 480
 caggaaaca tgctcgatca tttccgcag aaaaacggac agcaacttac ggtaatcata 540
 gaaaactgcg ccaaagtatc cttgcgtaaat tatgcctcgc ttgcggaaaga atccgtctat 600
 ctgtgtcccg tcatcagcgc agccggagag gaggggcaca agcagttaccg aagtgcattg 660
 agggtataca acaaaccgtt caaccagata tccggaaatac tggaaattgaa gactccgtg 720
 acctcttatac tggcacgcgc cagttggcg accacggccc tgcagaaagg ggttccgggt 780
 tcagtatca gtgcaggaat ggggcatgtc tcaagaaagg tgcacatacat ttatctggca 840
 tcttttgcata acaaaaacgtt cagtaacgc aataaaaaag tgattgcgc cgtgagattt 900
 aagaaagagg aggaggaggatg a 921

<210> 4
 <211> 306
 <212> PRT
 <213> Bacteroides fragilis

<400> 4

Met Glu Ile Glu Lys Phe Ile Lys Ser Leu Ala Arg Lys Ala Lys Leu
 1 5 10 15

Gly Gly Arg Tyr Ser Thr Ala Asn Thr Tyr Leu Tyr Thr Leu His Ser
 20 25 30

Phe Gln Lys Phe Ala Gly Lys Ala Ser Leu Thr Phe Glu Ile Thr
 35 40 45

Pro Glu Ser Ile Lys Glu Tyr Glu Gln Tyr Leu Ile Leu Asn Gly Lys
 50 55 60

Arg Tyr Asn Thr Ile Ser Leu Tyr Met Arg Met Leu Arg Ser Ile Cys
 65 70 75 80

Asn Gln Ala Ser Glu Gln Asn Ile Ala Ser Leu Asn Thr Arg Glu Leu
 85 90 95

Phe Glu Asn Val Phe Ile Gly Asn Glu Pro Thr Ala Lys Arg Ala Ile
 100 105 110

Ser Pro Val Leu Ile Ser Arg Leu Leu Glu Ala Asp Phe Ser Lys Asn
 115 120 125

Ser Arg Leu Asp Phe Ala Arg Asp Leu Phe Leu Leu Ser Phe Tyr Leu
 130 135 140

Arg Gly Ile Pro Phe Val Asp Leu Val His Leu Arg Lys Thr Asp Val
 145 150 155 160

Gln Gly Asn Met Leu Val Tyr Phe Arg Gln Lys Thr Gly Gln Gln Leu
 165 170 175

Thr Val Ile Ile Glu Asn Cys Ala Lys Val Ile Leu Arg Lys Tyr Ala
 180 185 190

Ser Leu Cys Lys Glu Ser Val Tyr Leu Leu Pro Val Ile Ser Ala Ala
 195 200 205

Gly Glu Glu Gly His Lys Gln Tyr Arg Ser Ala Leu Arg Val Tyr Asn
 210 215 220

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

Thr Ser Tyr Val Ala Arg His Ser Trp Ala Thr Thr Ala Leu Gln Lys
 245 250 255

Gly Val Pro Val Ser Val Ile Ser Ala Gly Met Gly His Ala Ser Glu
 260 265 270

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

Asn Ala Asn Lys Lys Val Ile Ala Ala Val Arg Phe Lys Lys Glu Glu
 290 295 300

Glu Glu
 305

<210> 5
 <211> 1958
 <212> DNA
 <213> Bacteroides fragilis

<400> 5
 agtactgata actccgggtga ctcccgccac tgtaaaaagaa aatgccggag aagctccccca 60
 gggacaagac gtaaacgcta cggtgaaaaa tgacacggtg ttatcgaca aattgcgggt 120

| | | | | | | |
|-------------|--------------|-------------|-------------|-------------|-------------|------|
| aaccgaacctt | attacctcca | ttgttaggcga | taaagacaaa | gcggaaagcca | ttgtcaaagc | 180 |
| catcggtgac | gtaaaataca | aagttaggcta | caagccggt | ctcaacacag | agaaggacag | 240 |
| catctacctt | gctttagate | cgaaccgtt | gacccttcaa | ctgcctgcag | ccgtagaagg | 300 |
| ccaggaagga | cagactgtta | cgttaaccat | ttcgctccg | gacaaggca | gctttgttta | 360 |
| caagaaaaat | cagtgaaagt | tgaagctcg | cgccgataaa | gtggaaactgg | caggcgtage | 420 |
| ggtacctgtt | cctcagaccc | tgttcgactt | cggatgacc | aaaaagaagt | gattgcgtgt | 480 |
| tcaacacacg | taacaacgcg | tatctatagg | gggtaaagcc | tgtaaaagaca | ggcttataac | 540 |
| ccgcatgaaa | aagtccgtgc | tctccccccc | ggagtgaaaca | cggaactttt | gttccttgaa | 600 |
| accattcaaa | aaaaaaagat | tatttcacag | caggcaatgc | cttttcataa | tcggcgttcct | 660 |
| gagtaatctc | cgccacaaggc | tctgtataag | ctactttccc | gtctttccca | atcaccacta | 720 |
| cogcagcgcg | cagcactccg | gccagccgtc | cgtcagccat | cctcacgcca | tagctctgt | 780 |
| cgaagtccga | aaaggcgaaa | tccgacaacg | gaatcacgtt | ttcgatacc | tctgtgtgc | 840 |
| agaagcgtcc | ctgcgcacaaac | ggcaagtctt | tcgaaatggc | caataccacg | gtatccatca | 900 |
| ttccggctgc | cattttattt | aatttacgca | ccgaagtggc | gcacacacgg | gtatccagac | 960 |
| tccggaaaat | attcagaaca | atattttac | catttcagatc | tttttagtgcg | aaagaagata | 1020 |
| aatcaactttt | caccagctcg | aaatcgggag | ccacccccc | aacctgtata | aattcgcacaa | 1080 |
| tcagctttac | cggttgcct | ttgaaatttg | ttgttgcata | aattgataac | tctaaagtttt | 1140 |
| atttactata | ttctaaacaa | tccgtgcaag | aactttgttc | acgtggaca | ataatctaaa | 1200 |
| aaataaaaatt | gatatggaaa | catttattcg | cgagatggaa | cacccgtca | aaaactggtg | 1260 |
| gttatctctt | attctggta | ttctgtacat | catcgtggct | ctctgtctgc | tatcgccacc | 1320 |
| gggaagcagt | tacattgccc | tcagcgtcat | cttcagcatt | tcgtgtcgta | taagtggtat | 1380 |
| catcgaaatc | atcttcctca | tcagtaaccg | gcgaggcattc | tcgtccctggg | gatggtacct | 1440 |
| cgcaggccgt | atcatcgatc | tgatctttag | catctacgt | gtagectatc | cgctgtcg | 1500 |
| catggaaagtc | ataccgttca | tagtgcctt | ctggatgtat | ttccgcgggtt | tctccgcac | 1560 |
| aggctattct | atggacctgt | agcgttatgg | caccctgtgg | tggggatggt | acatgggatt | 1620 |
| cggcatccctc | gccatcattt | tttcgtgtat | catctgtgg | cagccggccg | tagtgcctt | 1680 |
| ctacgttata | tatatgtgg | catttacttt | cctgtatcata | ggattttcc | gtgtcatgtt | 1740 |
| gttcctcgaa | ctgaaaagcc | ttcataaaacg | atcaacggtg | atgaacggta | aatgataaac | 1800 |
| attgaatgaa | cccttattca | ccacagatta | cacagattaa | cacagatatt | ttaattaaac | 1860 |

| | | | | | | |
|----------------------------|-------------|--------------|-------------|--------------|-------------|------|
| tctcagtaaa | ataaatttatt | aatctgtgag | aatctgtgta | atctgcggtg | aattatgact | 1920 |
| ccccaaacccg | ttactaatac | atcgffffca | taaaacccg | | | 1958 |
| <210> 6 | | | | | | |
| <211> 2540 | | | | | | |
| <212> DNA | | | | | | |
| <213> Bacteroides fragilis | | | | | | |
| <400> 6 | | | | | | |
| ttttccgtac | ttactctcaa | ataagctatt | accataattc | atgtttttaa | atgattaata | 60 |
| caccacgaaa | aaacggctat | tcattcaaat | acgggacact | tttttacgtt | cctttttct | 120 |
| catgccactt | gggtattttct | gaaactttca | ttcgtctata | catttatgtt | attgattttt | 180 |
| tactaatttc | cagcatattt | tccaatctgt | cactcaaat | cttttttatt | ataaaccgtg | 240 |
| ttcttgaaca | cactaaaaag | aacaagaaaa | tggaaataga | aaaatttatt | aaatctttag | 300 |
| caagaaaagc | gaagtttaggc | gggcgttaca | gcacagccaa | tacccatctc | tacactttgc | 360 |
| acagttttca | gaagtttgcg | ggaaaaggct | cactgacttt | tgaagagatc | actcccgaga | 420 |
| gtatcaagga | gtacgagcaa | tacttaatcc | tcaacgggaa | acggatcacac | acgatctcgc | 480 |
| tctacatgcg | catgttgcgt | tccatctgca | atcaggcattc | ggagcagaac | atagcttcgc | 540 |
| tcaacaccccg | cgagctgttt | gagaatgttt | ttatcgccaa | cgaaccact | gc当地cggtg | 600 |
| ccatctcacc | cgccctcatt | tcgcgttgc | tcgaagcaga | tttcagcaag | aacagccggc | 660 |
| tcgattttgc | ccgcgcaccc | ttcttgcata | gcttctacct | gaggggaaatc | ccgtttgtcg | 720 |
| acctgggtaca | tctccgcaag | accgatgtgc | agggaaacat | gctcgtttat | ttccgcgcaga | 780 |
| aaacgggaca | gcaacttacg | gtaatcatag | aaaactgcgc | caaagtgtatc | ttgcgttaatg | 840 |
| atgcctcgct | ttgcaaagaa | tccgtctata | tgctgcccgt | catacgcga | gc当地ggagg | 900 |
| agggggaccaa | gcagttaccga | agtgcattga | gggtatacaa | caaacgcctc | aaccagatat | 960 |
| ccggaaatact | gaaatttgaag | actccgcgt | ccttttatgt | ggcacgcac | agttgggca | 1020 |
| ccacggccct | gcagaaaaggg | gttccggttt | cagtgtatc | tgc当地gtat | gggoatgttt | 1080 |
| cagagaaggt | gacatacatt | tatctggcat | cttttgataa | caaaaacgcctc | agtaacgc当地 | 1140 |
| ataaaaaaagt | gatttgc当地 | gtgagatttta | agaaaagagga | ggaggaggta | taatagctgt | 1200 |
| tctcttactt | attaggtat | agaacagatt | cattttttttt | atcgctgc当地 | aaatagagat | 1260 |
| aataatttggaa | actccacaaa | caaaaatgtata | atttcttttc | tataaaaatgt | gattataacc | 1320 |
| agttgaaatgt | tcagtttggaa | ataatttttatt | cacttaatag | aaatattatgt | cataattcc | 1380 |
| gtttgatgtta | attactcaaa | caggatgttta | caatttgcaaa | taatttgaca | tcagaattat | 1440 |

WO 03/077863

PCT/US03/07775

ataatccagc ccctgttta tggtagtta acctctaaa gttatattc atatctttc 1500
gctatccgc atttattac ctaataagta agggatctt ttttttattt acctaataag 1550
taaggaaaca aattgcaatg cacacagcaa gatggtagtt attcaaacat taacgacaac 1620
tatcgcaac atttctaaa gtacagtatg aaacaggtat tgccgttcaa taaagtattt 1680
aaaaggatgg tattcacgg agatcttattt cttttgcattt gacacccctt gtccttgc 1740
accctattgg ggagcaatt ttttgcagat ccatttattt actctacttcc ccaagttactg 1800
gtattgtca acttatgtca cctggtagt aacatgtctt caggtatcat attgcacccg 1860
cgtgttagtac gtcccgagca aatcgatgg cgtgcatttac gcaacagtgc gggacacgccc 1920
ttgtttttt cctgcgcgtt cacccttggaa aacttggca tcccttccgc ccgtttttt 1980
ttactgttctt acattgtgtt cactctgtt tttttttttt accgggttattt gttccgaaag 2040
atcctgaagt cctatgtaa gcatggaggc aactccgcgca gcatcatttctt ggtggaaagc 2100
aatagcaata taatcgaaactt ctaccatcaa atgacggacg acgtcacttcc cggattccgt 2160
gtcatoggtt acttgacgca ccagccaggc agccgcatttcc cggaaaagggtt gaaactatctg 2220
ggaaaaccccg gtaagattgtt ggaccgcctg aagcaggaggg gggatcgagca ggttttattgt 2280
tgcctgcctt cggcccgccg cgaagagattt ctcccccattca tggactatttgc gaaaatcac 2340
ctgatacgtt ttttgcgtt ccccaacgttgc cgcgcgttcc tggactatttgc gaaaatcac 2400
gagctcctgg gcaacgttgc cgtactctgc atccgcaggc agccgcgttgc ttttgcggaa 2460
aaccgattca ggaagcgtgtt gttcgacatc gctttctgc tcttgcgttgc ttttgcggaa 2520
ttcccccattttt tctatgtcat 2540

<210> 7

<211> 19

<212> DNA

<213> Bacteroides fragilis

<400> 7

acgaacgtttt ttttggaaaca

19

<210> 8

<211> 22

<212> DNA

<213> Bacteroides fragilis

<400> 8

tagacgatcg ttttatttggaaa ca

22

<210> 9

<211> 25

WO 03/077863

PCT/US03/07775

<212> DNA

<213> Bacteroides fragilis

<400> 9

ttaaacgaac gtctattgaa acaet

25

<210> 10

<211> 20

<212> DNA

<213> Bacteroides fragilis

<400> 10

gttcaaatacg acgaacgttt

20

<210> 11

<211> 32

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 11

cggatccag tactgataac tccggtgact cc

32

<210> 12

<211> 32

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 12

atccatggcc ggtttatgaa aacgatgtat ta

32

<210> 13

<211> 34

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 13

cggatccatgtt ttccgtactt actctcaaat aagc

34

<210> 14

<211> 32

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

-8-

<400> 14
ggggatccat gacatagata atggggaaaga gg

32

1

- 9 -

SEQUENCE LISTING

<110> The Brigham and Women's Hospital, Inc.

<120> METHOD FOR OVEREXPRESSION OF ZWITTERIONIC POLYSACCHARIDES

<130> B00801.70279

<150> US 60/364,168
<151> 2001-03-13

<160> 14

<170> PatentIn version 3.1

<210> 1
<211> 594
<212> DNA
<213> Bacteroides fragilis

<400> 1
atggtaatg cttatttgag agtaagtacg gaaaaacagt ttttggctaa tcagaaggaa 60
gagattatgc gatttgcaaga gaagaatggg ttgtcgattt acaagtggta cacagagacc 120
gtaaggcgaa gcgtgagcac aaaagacaga aagctatcag agttattgaa gagaatgcat 180
cccggggata cactgattgt aacggagatt tcgagattga gccgtacact gtcgagatt 240
atgactatcc tgaattttt tattaagaag caggtatgtc tctatagcac caaagagggc 300
tatgtgtttc aggacgacat caacagcaag gtgtgtggat tcgcgttcgg actgtatggcg 360
gaaatagaaa ggaacctgat ttgcgtgcgt accaaagaag ctctcgacg cagaaagcag 420
gaaggaatga ctttaggccc aaagaaaaggg gatacgccctaa aataaaaatt gctgcgtgcc 480
aataagcgcg tacttaccaa agaacttgac aaaggaacta cttactcgga attggcggag 540
aagatggggg tatccagaac aaccctgttc cggttatga aaacgtatgtat 594

<210> 2
<211> 197
<212> PRT
<213> Bacteroides fragilis

<400> 2

Met Val Ile Ala Tyr Leu Arg Val Ser Thr Glu Lys Gln Phe Leu Ala
1 5 10 15

Asn Gln Lys Glu Glu Ile Met Arg Phe Ala Glu Lys Asn Gly Leu Ser
20 25 30

Ile Asp Lys Trp Tyr Thr Glu Thr Val Ser Gly Ser Val Ser Thr Lys
35 40 45

Asp Arg Lys Leu Ser Glu Leu Leu Lys Arg Met His Pro Gly Asp Thr
 50 55 60

Leu Ile Val Thr Glu Ile Ser Arg Leu Ser Arg Thr Leu Leu Glu Ile
 65 70 75 80

Met Thr Ile Leu Asn Phe Cys Ile Lys Lys Gln Val Val Leu Tyr Ser
 85 90 95

Thr Lys Glu Gly Tyr Val Phe Gln Asp Asp Ile Asn Ser Lys Val Leu
 100 105 110

Gly Phe Ala Phe Gly Leu Met Ala Glu Ile Glu Arg Asn Leu Ile Ser
 115 120 125

Met Arg Thr Lys Glu Ala Leu Ala Arg Arg Lys Gln Glu Gly Met Thr
 130 135 140

Leu Gly Arg Lys Lys Gly Asp Thr Pro Lys Ile Lys Leu Leu Arg Ala
 145 150 155 160

Asn Lys Arg Val Leu Thr Lys Glu Leu Asp Lys Gly Thr Thr Tyr Ser
 165 170 175

Glu Leu Ala Glu Lys Met Gly Val Ser Arg Thr Thr Leu Phe Arg Phe
 180 185 190

Met Lys Thr Met Tyr
 195

<210> 3
 <211> 921
 <212> DNA
 <213> Bacteroides fragilis
 <400> 3
 atggaaatag aaaaattcat taaatcttta gcaagaaaag cgaagtttagg cggcggtac 60
 agcacagcca atacctacct ctacactttg cacagtttc agaagtttgc gggaaaagcc 120
 tcactgactt ttgaagagat cactcccgag agtatacaagg agtacgagca atacttaatc 180
 ctcaacggga aacggtaacaa cacgatctcg ctctacatgc gcatgttgcg ttccatctgc 240
 aatcaggcat cggagcagaa catagttcg ctcaacaccc gcgagctgtt tgagaatgtt 300
 ttatcgca acgaacccac tgccaaagcgt gccatctcac ccgtcctcat ttcccgctg 360

ctcgaagcag atttcagcaa gaacagccgg ctcgatttg cccgcaccc cttcttgcta 420
 agcttctacc tgagggaaat cccgttgtc gacctggtaat ctcgcggaa gaccgatgtg 480
 caggaaaca tgctcggttta tttccgcccgg aaaaacggac agcaacttac ggtaatcata 540
 gaaaactcg cccaaatgtat cttgcgttaag tatgcctcgcc tttgcaaaaga atccgtctat 600
 ctgcgtcccg tcacgcgc agccggagag gagggcaca agcagtaccg aagtgcattg 660
 agggatataca acaaacgcct caaccagata tccggaatac tgaaattgaa gactccgctg 720
 acctcttgc tggcacgcca cagttggcgcc accacggccc tgcaaaaaagg ggttccgggtt 780
 tcagtgatca gtgcaggaat ggggcatgtc tcagagaagg tgacatacat ttatctggca 840
 tcttttgcata acaaaacgcct cagtaacgc aataaaaaag tgattgcgc cgtgagattt 900
 aagaaagagg aggaggaggta a 921

<210> 4
 <211> 306
 <212> PRT
 <213> Bacteroides fragilis

<400> 4

Met Glu Ile Glu Lys Phe Ile Lys Ser Leu Ala Arg Lys Ala Lys Leu
 1 5 10 15

Gly Gly Arg Tyr Ser Thr Ala Asn Thr Tyr Leu Tyr Thr Leu His Ser
 20 25 30

Phe Gln Lys Phe Ala Gly Lys Ala Ser Leu Thr Phe Glu Glu Ile Thr
 35 40 45

Pro Glu Ser Ile Lys Glu Tyr Glu Gln Tyr Leu Ile Leu Asn Gly Lys
 50 55 60

Arg Tyr Asn Thr Ile Ser Leu Tyr Met Arg Met Leu Arg Ser Ile Cys
 65 70 75 80

Asn Gln Ala Ser Glu Gln Asn Ile Ala Ser Leu Asn Thr Arg Glu Leu
 85 90 95

Phe Glu Asn Val Phe Ile Gly Asn Glu Pro Thr Ala Lys Arg Ala Ile
 100 105 110

Ser Pro Val Leu Ile Ser Arg Leu Leu Glu Ala Asp Phe Ser Lys Asn
 115 120 125

Ser Arg Leu Asp Phe Ala Arg Asp Leu Phe Leu Leu Ser Phe Tyr Leu
 130 135 140

Arg Gly Ile Pro Phe Val Asp Leu Val His Leu Arg Lys Thr Asp Val
 145 150 155 160

Gln Gly Asn Met Leu Val Tyr Phe Arg Gln Lys Thr Gly Gln Gln Leu
 165 170 175

Thr Val Ile Ile Glu Asn Cys Ala Lys Val Ile Leu Arg Lys Tyr Ala
 180 185 190

Ser Leu Cys Lys Glu Ser Val Tyr Leu Leu Pro Val Ile Ser Ala Ala
 195 200 205

Gly Glu Glu Gly His Lys Gln Tyr Arg Ser Ala Leu Arg Val Tyr Asn
 210 215 220

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

Thr Ser Tyr Val Ala Arg His Ser Trp Ala Thr Thr Ala Leu Gln Lys
 245 250 255

Gly Val Pro Val Ser Val Ile Ser Ala Gly Met Gly His Ala Ser Glu
 260 265 270

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

Asn Ala Asn Lys Val Ile Ala Ala Val Arg Phe Lys Lys Glu Glu
 290 295 300

Glu Glu
 305

<210> 5
 <211> 1958
 <212> DNA
 <213> Bacteroides fragilis

<400> 5
 agtaatgata actccgggtga ctccccccac tgtaaaaagaa aatggccggag aagctcccca 60
 gggacaagac gtaaacgcta cggtgaaaaa tgacacggtg ttcttcgaca aattggccggt 120

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| aaccgaactt | attacctcca | ttgttaggoga | taaagacaaa | gcggaagcca | ttgtcaaagc | 180 |
| catcggtgac | gtaaaataca | aagttaggcta | caagccggct | ctcaacacag | agaaggacag | 240 |
| catctacctt | gttttcgatc | cgaacccgtt | gacccttcaa | ctgcctgcag | ccgtagaagg | 300 |
| ccaggaagga | cagactgtta | ccgtaaccat | ttcgtctccg | gacaaaggca | gttttgccta | 360 |
| caagaaaaat | cagttgaagt | tgaagcttag | cgccgataaa | gtggaactgg | caggcgttagc | 420 |
| ggtacctgtt | cctcagaccc | tgttcgactt | cggatgtacc | aaaaagaagt | gattgcctgt | 480 |
| tcaacacacg | taacacacg | tagtcatagg | gggtaaagcc | tgtaaagaca | ggctttatac | 540 |
| ccgcatgaaa | aagtccgtgc | tatccccccc | ggagtgaaca | cggaatttct | gttcccttgaa | 600 |
| accattcaaa | aaaaaaagat | tatccacag | cagccaaatgc | cttttcataa | tccggctcct | 660 |
| gagtaatctc | cggcacaage | totgtataag | ctactttccc | gtctttccca | atcaccacta | 720 |
| ccgcacgcgc | cagcagtccg | gccagcggtc | cgtcagccat | cctcacgcca | tagctctgt | 780 |
| cgaagtccgaa | aaagcggaaa | tccgacaacg | gaatcacgtt | ttcgtacacc | tctgtcgtgc | 840 |
| agaagcgtcc | ctgcgcaaac | ggcaagtctt | tcgaaatggc | caataccacg | gtatccctca | 900 |
| tccoggctgc | cattttattt | aatttacgca | ccgaagtggc | gcacacacgg | gtatccagac | 960 |
| tcggaaaaat | attcagaaca | atattttac | ctttcagatc | tttttagtgcg | aaagaagata | 1020 |
| aatcactttt | caccagctcg | aatatcggag | ccacctttcc | aacctgtata | aattcgccaa | 1080 |
| tcagctttac | cggttgtcct | ttgaaattt | ttgttgccat | aattgataac | tctaagttt | 1140 |
| atttactata | ttctaaacaa | tcgggtcaag | aactttgttc | acgtatggaca | ataatctaaa | 1200 |
| aaataaaaatt | gatatgaaaa | ctttatttcga | cgagatggaa | cacgcagtca | aaaactggtg | 1260 |
| gttatctctt | attctgggtt | ttctgtacat | catcgtggct | ctctgtctgc | tattcgcacc | 1320 |
| gggaagcagt | tacattggcc | tcagcgtcat | cttcagcatt | tcgtatgtca | taagtggtat | 1380 |
| catcgaaatc | atcttcctca | tcagtaaccc | cgaggcattc | tcgtcctggg | gatggtacct | 1440 |
| cgcaggccgt | atcategatc | tgtatctagg | catctaccc | gtagctatc | cgctgtctcg | 1500 |
| catggaagtc | ataccgttca | tagtgcctt | ctggatgtat | ttcccggtt | tctccgcac | 1560 |
| aggctttat | atggacactga | agcgttatgg | cacccgttag | tggggatgtt | acatgggatt | 1620 |
| cgccatccctc | gccccatctt | gttcgtgtat | cacccgttag | cagccggccg | tagtgcct | 1680 |
| ctacgttata | tatatgtctgg | cattcactt | cctgtatc | ggattttcc | gtgtcatgtt | 1740 |
| gtcccttcgaa | ctgaaaagcc | ttcataaaacg | atcaacggtg | atgaacggta | aatgataaac | 1800 |
| attgaatgaa | cccctattca | ccacagatta | cacagattaa | cacagatatt | ttaattaaac | 1860 |

| | |
|--|------|
| tctcagtaaa ataaatttatt aatctgtgag aatctgtgta atctgcggtg aattatgact | 1920 |
| ccccctaaccg ttactaatac atcggtttca taaaaccgg | 1958 |
| | |
| <210> 6 | |
| <211> 2540 | |
| <212> DNA | |
| <213> <i>Bacteroides fragilis</i> | |
| | |
| <400> 6 | |
| ttttccgtac ttactctcaa ataagctatt accataattc atgttttaa atgattaata | 60 |
| caccacgaaa aaacggctat tcattcaat acgggacact tttttacgtt cttttttttt | 120 |
| catgccactt gggtattttct gaaactttca ttgcgttata catttatgtt attgtttttt | 180 |
| tactaatttc cagcatattt tccaatctgt cactcaaaat cttttttattt ataaaccgtg | 240 |
| ttcttgaaca cactaaaaag aacaagaaaa tggaaataga aaaatttcatt aaatcttttag | 300 |
| caagaaaaagc gaagtttaggc gggcggttaca gcacagccaa tacctacctc tacactttgc | 360 |
| acagttttca gaagtttgcg gaaaaagcct cactgacttt tgaagagatc actcccgaga | 420 |
| gtatcaagga gtacgagcaa tacttaatcc tcaacgggaa acggtacaac acgatctcg | 480 |
| tctacatgcg catgttgcgt tccatctgca atcaggcattc ggagcagaac atagatccgc | 540 |
| tcaacaccccg cgagctgtttt gagaatgttt ttatcgccaa cgaacccact gccaagcgtg | 600 |
| ccatctcacc cgtcccttattt tcccgctgc tcgaagcaga tttcagcaag aacagccggc | 660 |
| tcttttgc ccgcgacccctc ttcttgctaa gcttctacct gaggggaaatc ccgtttgtcg | 720 |
| acctggata tctccgcaag accgatgtgc agggaaacat gctcggttat ttccgcccaga | 780 |
| aaacgggaca gcaacttacg gtaatcatag aaaactgcgc caaagtgtatc ttgcgttaagt | 840 |
| atgcctcgat ttgcaaaagaa tccgtctatac tgctggccgt catcagcgca gcccggagagg | 900 |
| aggggcacaa gcagtgaccga agtgcattga gggtatacaa caaacgcctc aaccagatata | 960 |
| ccggaaatact gaaattgaag actccgctga cctcttatgt ggcacgcccac agttggggcga | 1020 |
| ccacggccctt gcagaaaggg gttccggttt cagtgtatcg tgcaggaaatc gggcatgctt | 1080 |
| cagagaaggt gacatacattt tatctggcat cttttataaa caaaacgcctc agtaacgcaa | 1140 |
| ataaaaaaaat gattccgccc gtgagatataa agaaagagga ggaggaggatgta taatagctgt | 1200 |
| tcttttactt attaggtaat agaacagatt cattttttt atcgctgcaaa aaatagagat | 1260 |
| aataattgtaa actccacaaa caaaatgtata atttcttttc tataaaatgtt gattataacc | 1320 |
| agttgaagta tcaagtttgcataatttattt cacttaatag aaatatttagt cataattccct | 1380 |
| gtttgatgtatcaaa caggagtttca caatttgcaaa taatttgaca tcagaattat | 1440 |

ataatccagc ccctgttta tgtttagtta acctctaaaa gtttatatttc atatctttc 1500
 gctattccgc attctattac ctaataagta agggatcaact ttgttctatt acctaataag 1560
 taagggaca aattgaatg cacacagca gatggtagtt attcaaacat taacgacaac 1620
 tatcgcaaac atttctaaaa gtacagttatg aaacaggtat tgcggttcaa taaagtcat 1680
 aaaaggattt tattcacccg agatcttattt ctcttgaatg gcaccccttgc gtccttgc 1740
 accctattgg ggagcaaaatt ttttgcagat ccatttcatc actcacttcc ccaagttactg 1800
 gtattgtca actttagtcta cctgggttgc aacatgttctt caggtatcat attgcaccgc 1860
 cgtgttagtac gtcccgagca aatcgatgg cgtgccttac gcaacagtgc gggacacgccc 1920
 ttgtttttt cctgcgcgcctt caccttttgc aacttcggca tcccttccgc ccgccttttc 1980
 ttactgttctt acatggcgat cactctgttgc ttsgtttgc accggttattt gttccgcaag 2040
 atccctgaatg cctatcgtaa gcatggaggc aactcccgca gcatcatttgc ggtggaaagc 2100
 aatagcaata taatcgaaact ctaccatcaa atgacggacg acgtcacttcc cggattccgt 2160
 gtcatcggtt acttttgcgca ccagccaggc agccgccttcc ccggaaaagggtt gaactatctg 2220
 ggaaaaccccg gtaagattgtt ggaccgccttgc aagcaggaggag gagtcgagca ggtttattgt 2280
 tgccctgcctt ccggcccgccag cgaagagattt ctccccatca tgcactatttgc cggaaaatcact 2340
 ctgatacgctt ttttcgtgtt ccccaacgttgc cgcagcttgc tgaacggcgcatgtacttcc 2400
 gagtcctcgg gcaacgtgcc cgtacttgc atccggcagg agccgccttgc ttttgcggaa 2460
 aaccgatttca ggaagcggtt gttcgacatc gcttttcttgc tcttttttgc ttgcaccctc 2520
 ttccccattttatgtcat 2540

<210> 7
 <211> 19
 <212> DNA
 <213> Bacteroides fragilis

<400> 7
 acgaacgtttt tttgaaaca 19

<210> 8
 <211> 22
 <212> DNA
 <213> Bacteroides fragilis

<400> 8
 tagacgatcg tctattgaaa ca 22

<210> 9
 <211> 25

<212> DNA
<213> Bacteroides fragilis

<400> 9
ttaaacgaac gtctattgaa acact 25

<210> 10
<211> 20
<212> DNA
<213> Bacteroides fragilis

<400> 10
gttcaaatag acgaacgttt 20

<210> 11
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 11
ccggatccag tactgataac tccgggtgact cc 32

<210> 12
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 12
atccatggcc ggtttatgaa aacgatgtat ta 32

<210> 13
<211> 34
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 13
cgccatggtt ttccgtactt actctcaaat aagc 34

<210> 14
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 14
ggggatccat gacatagata atgggaaaga gg

32

- 9 -