



US 20230226175A1

(19) **United States**(12) **Patent Application Publication**
KEMMLER et al.(10) **Pub. No.: US 2023/0226175 A1**(43) **Pub. Date: Jul. 20, 2023**(54) **MODIFIED EXOTOXIN A PROTEINS**(30) **Foreign Application Priority Data**(71) Applicant: **GlaxoSmithKline Biologicals SA,**
Rixensart (BE)

Jun. 25, 2020 (EP) 20182138.6

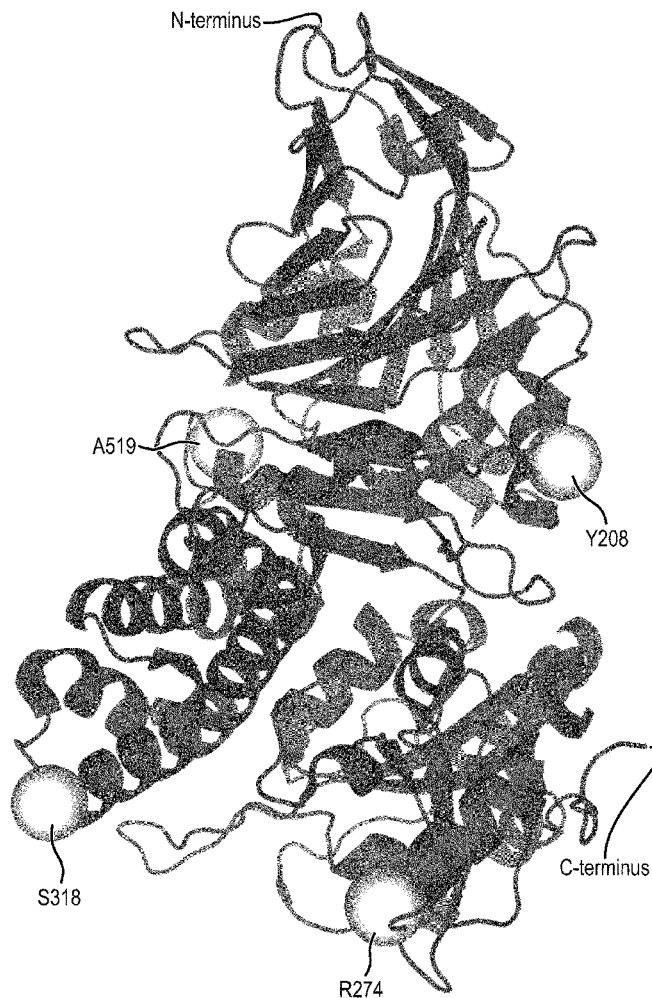
Jun. 25, 2020 (EP) 20182139.4

(72) Inventors: **Stefan Jochen KEMMLER**, Zurich
(CH); **Christian LIZAK**, Zurich (CH);
Gerald Johann POSCH, Zurich (CH);
Julien Laurent QUEBATTE, Zurich
(CH)**Publication Classification**(51) **Int. Cl.****A61K 39/385** (2006.01)**A61K 39/108** (2006.01)**C12N 9/10** (2006.01)(52) **U.S. Cl.**CPC **A61K 39/385** (2013.01); **A61K 39/0266**
(2013.01); **C12N 9/1077** (2013.01); **C12Y**
204/02036 (2013.01)(21) Appl. No.: **18/001,498**(22) PCT Filed: **Jun. 17, 2021**(86) PCT No.: **PCT/EP2021/066336**

§ 371 (c)(1),

(2) Date: **Dec. 12, 2022****Related U.S. Application Data**(60) Provisional application No. 63/043,883, filed on Jun.
25, 2020.(57) **ABSTRACT**

The present invention relates to the field of modified proteins, immunogenic compositions and vaccines comprising the modified proteins, their manufacture and the use of such compositions in medicine. More particularly, it relates to a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein. The modified EPA can be used as a carrier protein for other antigens, particularly saccharide antigens or other antigens lacking T cell epitopes.

Specification includes a Sequence Listing.

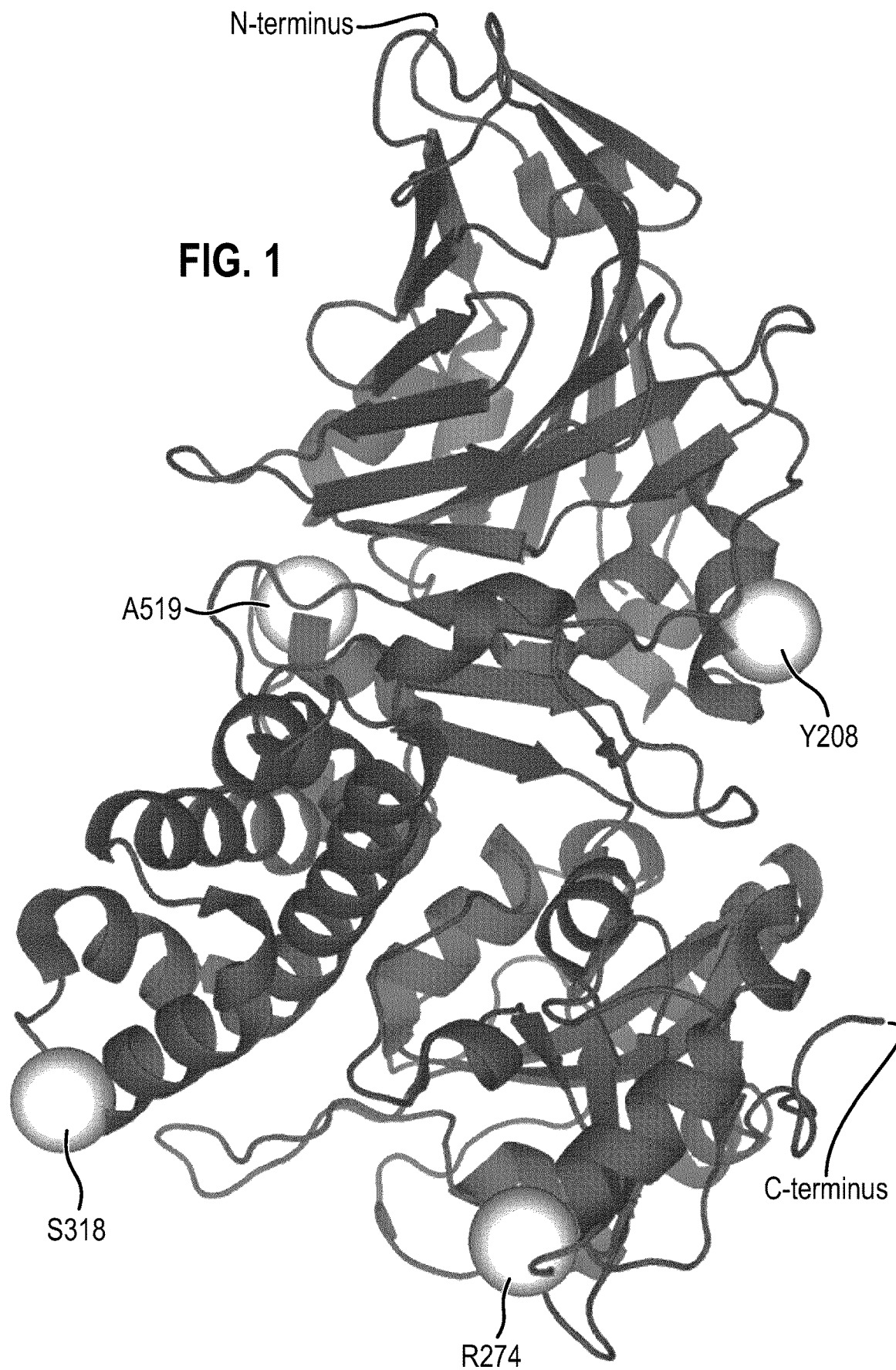


FIG. 2

EPA detox	AAAAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQVLHYSMVLEGGNDALKLAI	60
EPA mut_Y208	AAAAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQVLHYSMVLEGGNDALKLAI	60
EPA mut_R274	AAAAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQVLHYSMVLEGGNDALKLAI	60
EPA mut_S318	AAAAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQVLHYSMVLEGGNDALKLAI	60
EPA mut_A519	AAAAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQVLHYSMVLEGGNDALKLAI	60

EPA detox	DNALSITSDGLTIRLEGGVEPNKPVRSYTRQARGSWSLNWLVPICHEKPSNIKVFIHEL	120
EPA mut_Y208	DNALSITSDGLTIRLEGGVEPNKPVRSYTRQARGSWSLNWLVPICHEKPSNIKVFIHEL	120
EPA mut_R274	DNALSITSDGLTIRLEGGVEPNKPVRSYTRQARGSWSLNWLVPICHEKPSNIKVFIHEL	120
EPA mut_S318	DNALSITSDGLTIRLEGGVEPNKPVRSYTRQARGSWSLNWLVPICHEKPSNIKVFIHEL	120
EPA mut_A519	DNALSITSDGLTIRLEGGVEPNKPVRSYTRQARGSWSLNWLVPICHEKPSNIKVFIHEL	120

EPA detox	NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ	180
EPA mut_Y208	NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ	180
EPA mut_R274	NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ	180
EPA mut_S318	NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ	180
EPA mut_A519	NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ	180

FIG. 2(contd)

EPA detox	PRREKRWSEWASGKVLCLLDPLDGVNY-----LAQQRCLDDTTWEGKIYRVLAGNPAK	234
EPA mut_Y208	PRREKRWSEWASGKVLCLLDPLDGVNY KDQONATK LAQQRCLDDTTWEGKIYRVLAGNPAK	240
EPA mut_R274	PRREKRWSEWASGKVLCLLDPLDGVNY-----LAQQRCLDDTTWEGKIYRVLAGNPAK	234
EPA mut_S318	PRREKRWSEWASGKVLCLLDPLDGVNY-----LAQQRCLDDTTWEGKIYRVLAGNPAK	234
EPA mut_A519	PRREKRWSEWASGKVLCLLDPLDGVNY-----LAQQRCLDDTTWEGKIYRVLAGNPAK	234
	*****	*****
EPA detox	HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT----- R HRQPRGWEQLEQCG	288
EPA mut_Y208	HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLETFT-----RHRQPRGWEQLEQCG	294
EPA mut_R274	HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT KDQONATK RHRQPRGWEQLEQCG	294
EPA mut_S318	HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT-----RHRQPRGWEQLEQCG	288
EPA mut_A519	HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT-----RHRQPRGWEQLEQCG	288
	*****	*****
	*****	*****
EPA detox	YPVQRLVALYLAARLSWNQVDQVIRNALA----- S PGSGDLGEAIREQPEQARLALTTL	342
EPA mut_Y208	YPVQRLVALYLAARLSWNQVDQVIRNALA-----SPGSGDLGEAIREQPEQARLALTTL	348
EPA mut_R274	YPVQRLVALYLAARLSWNQVDQVIRNALA-----SPGSGDLGEAIREQPEQARLALTTL	348
EPA mut_S318	YPVQRLVALYLAARLSWNQVDQVIRNALA KDQONATK PGSGDLGEAIREQPEQARLALTTL	348
EPA mut_A519	YPVQRLVALYLAARLSWNQVDQVIRNALA-----SPGSGDLGEAIREQPEQARLALTTL	342
	*****	*****

FIG. 2(contd)

EPA detox	AAAESERFVRQGTGNDEAGAAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEEFLG	402
EPA_mut_Y208	AAAESERFVRQGTGNDEAGAAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEEFLG	408
EPA_mut_R274	AAAESERFVRQGTGNDEAGAAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEEFLG	408
EPA_mut_S318	AAAESERFVRQGTGNDEAGAAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEEFLG	408
EPA_mut_A519	AAAESERFVRQGTGNDEAGAAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEEFLG	402

EPA detox	DGGDVSFSTRGTQNTVERLLQAHRQLEERGIVFVGYHGTFLEAAQSI VFGGVRARSQDL	462
EPA_mut_Y208	DGGDVSFSTRGTQNTVERLLQAHRQLEERGIVFVGYHGTFLEAAQSI VFGGVRARSQDL	468
EPA_mut_R274	DGGDVSFSTRGTQNTVERLLQAHRQLEERGIVFVGYHGTFLEAAQSI VFGGVRARSQDL	468
EPA_mut_S318	DGGDVSFSTRGTQNTVERLLQAHRQLEERGIVFVGYHGTFLEAAQSI VFGGVRARSQDL	468
EPA_mut_A519	DGGDVSFSTRGTQNTVERLLQAHRQLEERGIVFVGYHGTFLEAAQSI VFGGVRARSQDL	462

EPA detox	DAIWGRFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA	519
EPA_mut_Y208	DAIWGRFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA	525
EPA_mut_R274	DAIWGRFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA	525
EPA_mut_S318	DAIWGRFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA	525
EPA_mut_A519	DAIWGRFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA	522

EPA detox	---	APEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVWIPSAIPTDPRN	576
EPA mut_Y208	---	APEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVWIPSAIPTDPRN	582
EPA mut_R274	---	APEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVWIPSAIPTDPRN	582
EPA mut_S318	---	APEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVWIPSAIPTDPRN	582
EPA mut_A519	ATKA	APEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVWIPSAIPTDPRN	582

EPA detox		VGGDLPSSIPDKEQAISALPDYASQPGKPPREDLK	612
EPA mut_Y208		VGGDLPSSIPDKEQAISALPDYASQPGKPPREDLK	618
EPA mut_R274		VGGDLPSSIPDKEQAISALPDYASQPGKPPREDLK	618
EPA mut_S318		VGGDLPSSIPDKEQAISALPDYASQPGKPPREDLK	618
EPA mut_A519		VGGDLPSSIPDKEQAISALPDYASQPGKPPREDLK	618

FIG. 2(contd)

FIG. 3

Klebsiella pneumoniae O-antigen

Number of glycosites	1	1	1	1	1	1	2
Glycosite position	Y208	K240	R274	S318	A376	A519	K240+ A376

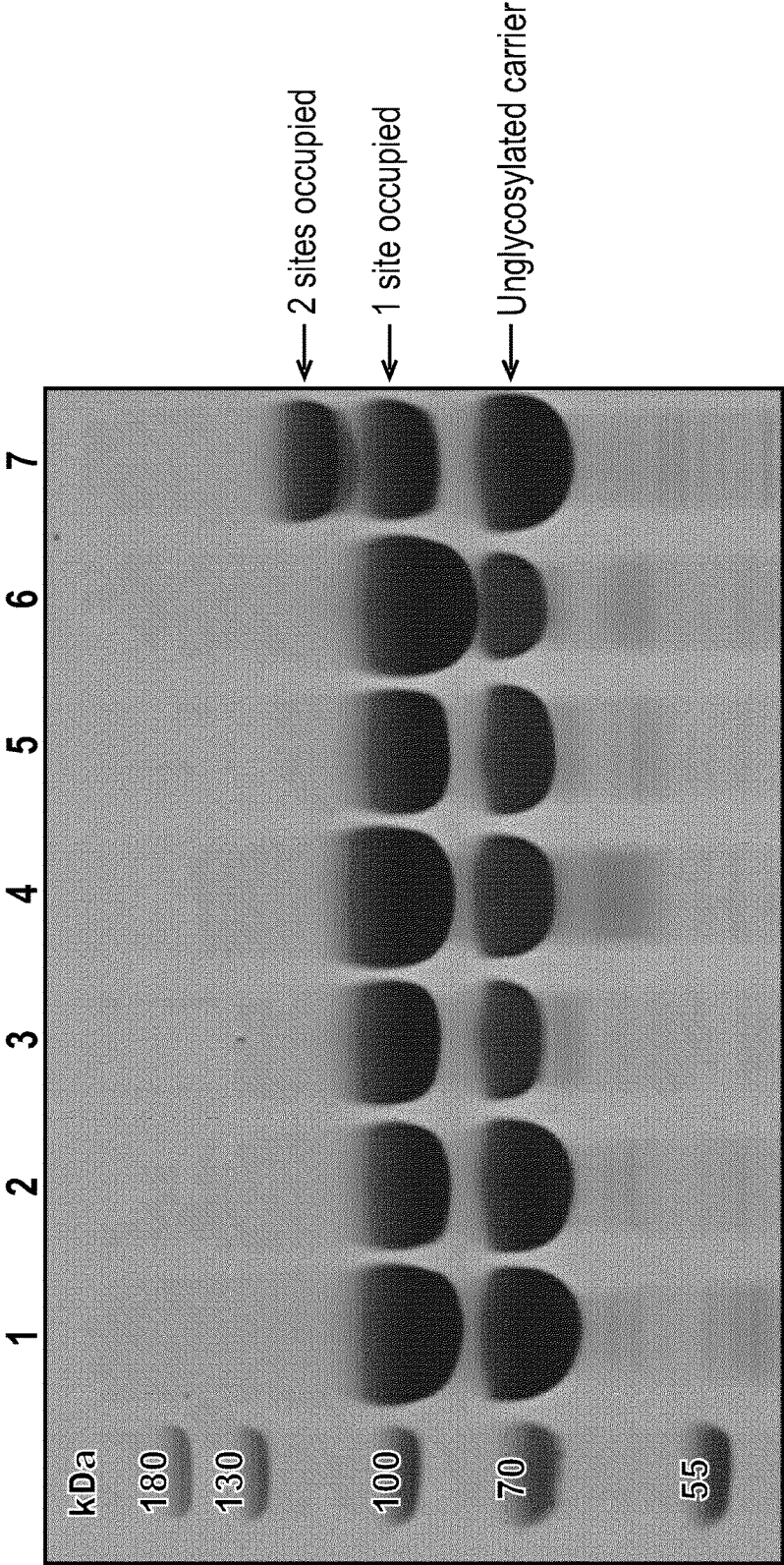


FIG. 4

Klebsiella pneumoniae O-antigen

Number of glycosites	2	2	2	2	2	2	3
Glycosite position	K240+ A376	Y208+ R274	Y208+ S318	Y208+ A519	R274+ S318	R274+ A519	Y208+ R274+ A519

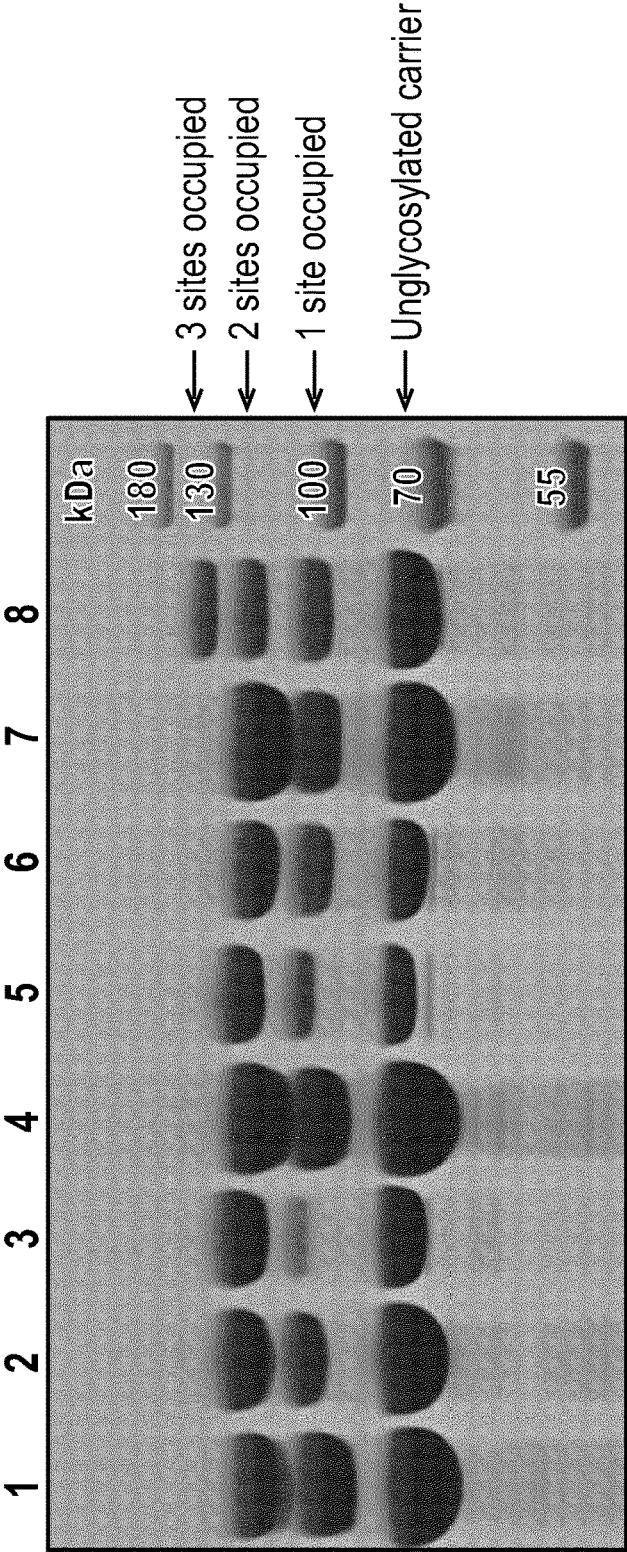


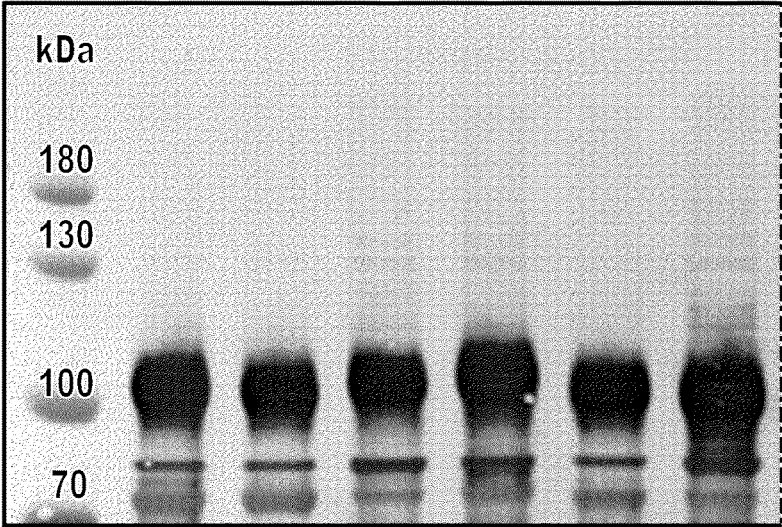
FIG. 5

Klebsiella pneumoniae O-antigen

Number of glycosites		1	1	1	1	1	1
Glycosite position		Y208	K240	R274	S318	A376	A519

123456

anti-KpO



anti-EPA

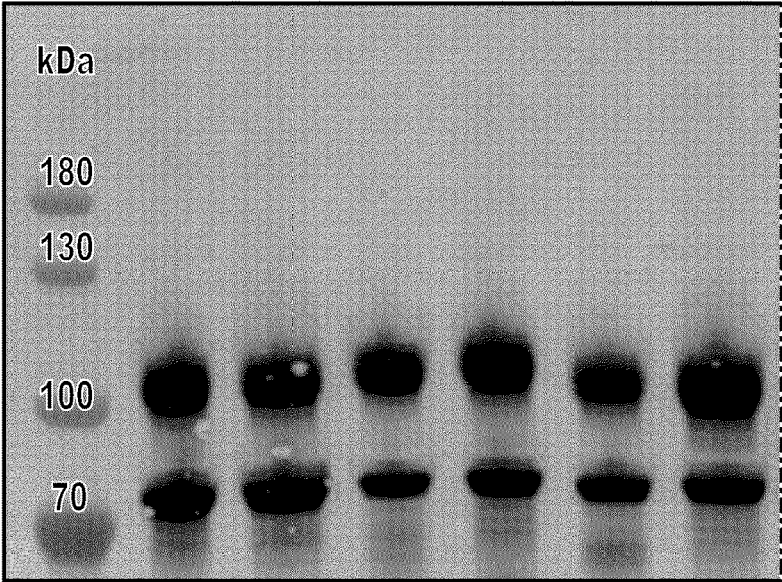
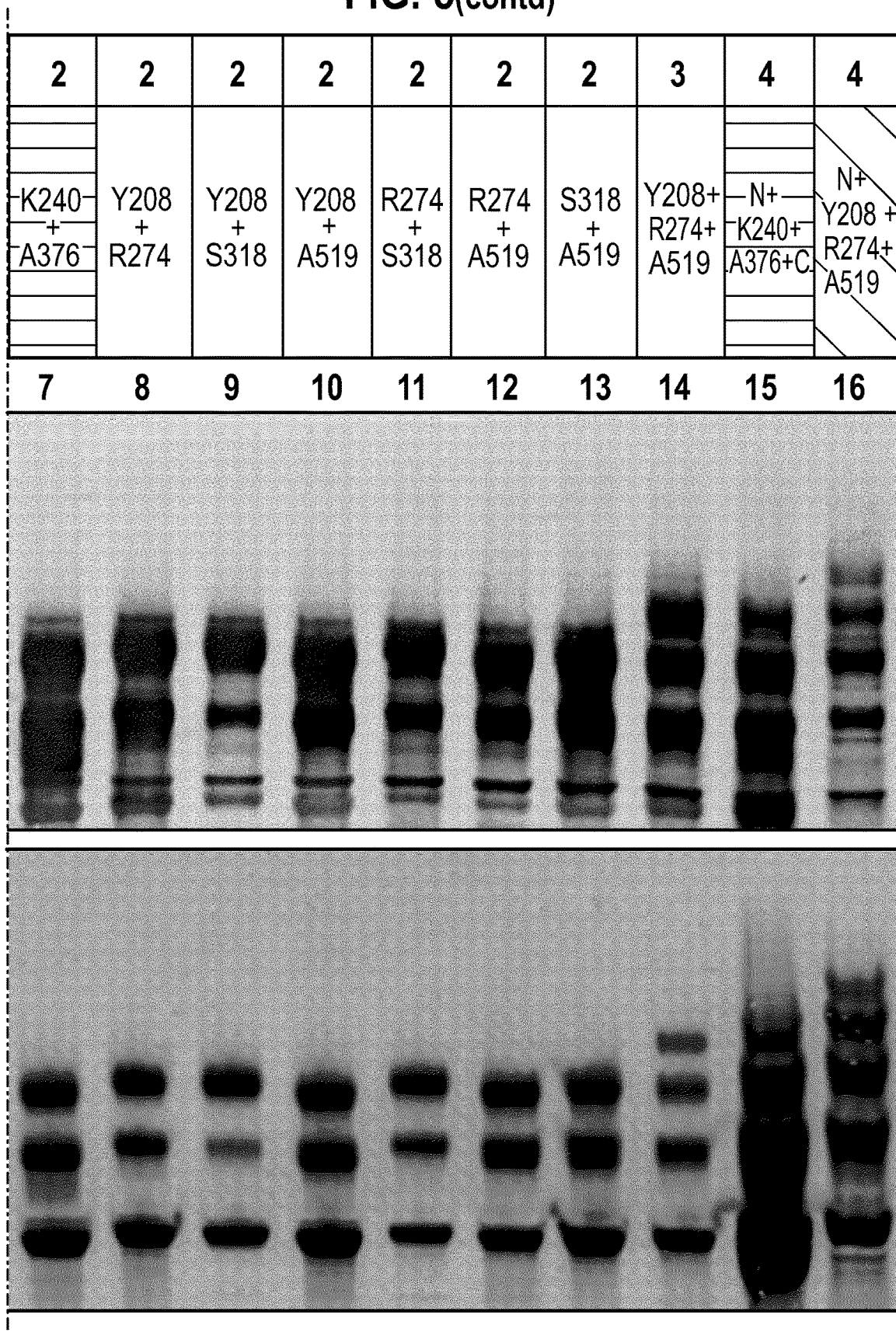
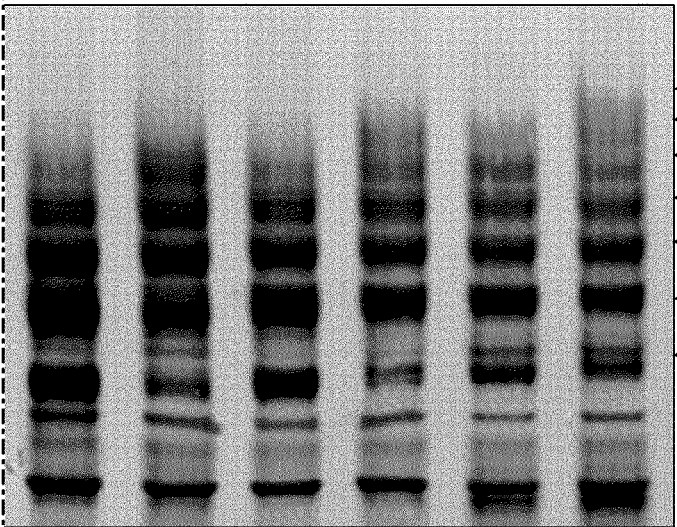


FIG. 5(contd)

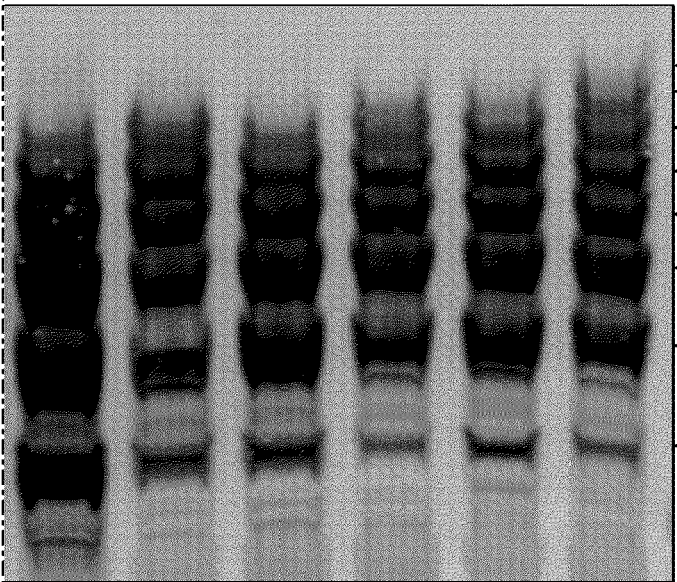


5	5	5	6	6	7
N+ Y208+ R274+ A519 +C	N+ Y208+ S318+ A519 +C	N+ R274+ S318+ A519 +C	N+ Y208+ R274+ S318+ A519 +C	N+ Y208+ K240+ R274+ S318+ A519	N+ Y208+ K240+ R274+ S318+ A519 +C

17 18 19 20 21 22



← 7 sites occupied
← 6 sites occupied
← 5 sites occupied
← 4 sites occupied
← 3 sites occupied
← 2 sites occupied
← 1 site occupied



← 7 sites occupied
← 6 sites occupied
← 5 sites occupied
← 4 sites occupied
← 3 sites occupied
← 2 sites occupied
← 1 site occupied
← Unglycosylated carrier

FIG. 5(contd)

FIG. 6

Klebsiella pneumoniae O-antigen

5	5	5	6	6	7	4
N+ Y208 + R274+ A519 +C	N+ Y208 + S318+ A519 +C	N+ R274 + S318+ A519 +C	N+ Y208+ R274 + S318+ A519 +C	N+ Y208+ K240+ R274 + S318+ A519	N+ Y208+ K240+ R274 + S318+ A519 +C	N+ Y208 + R274+ A519

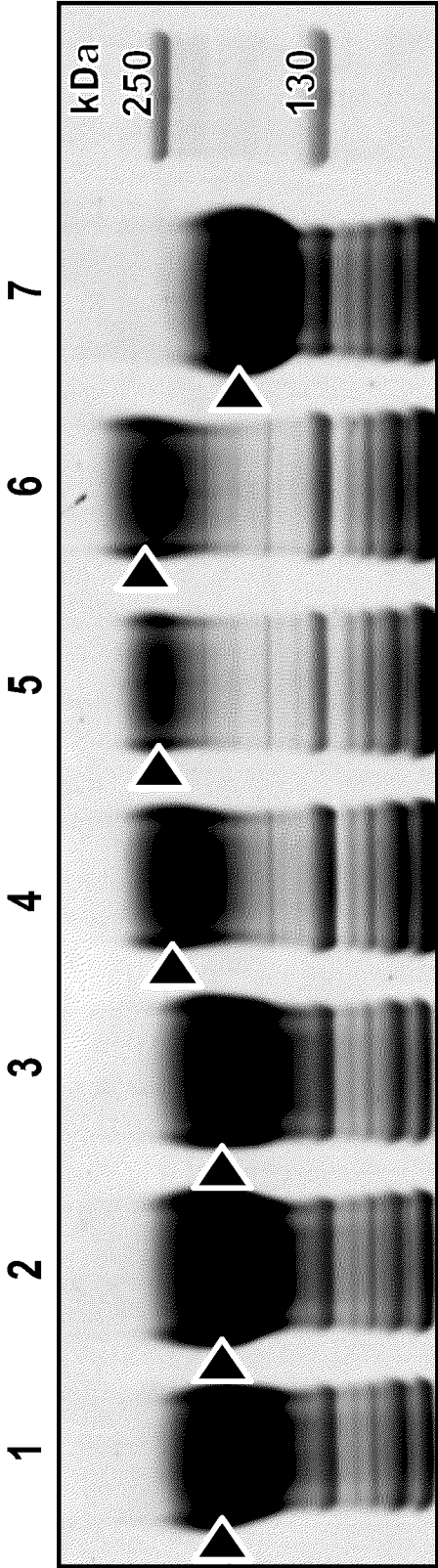


FIG. 7

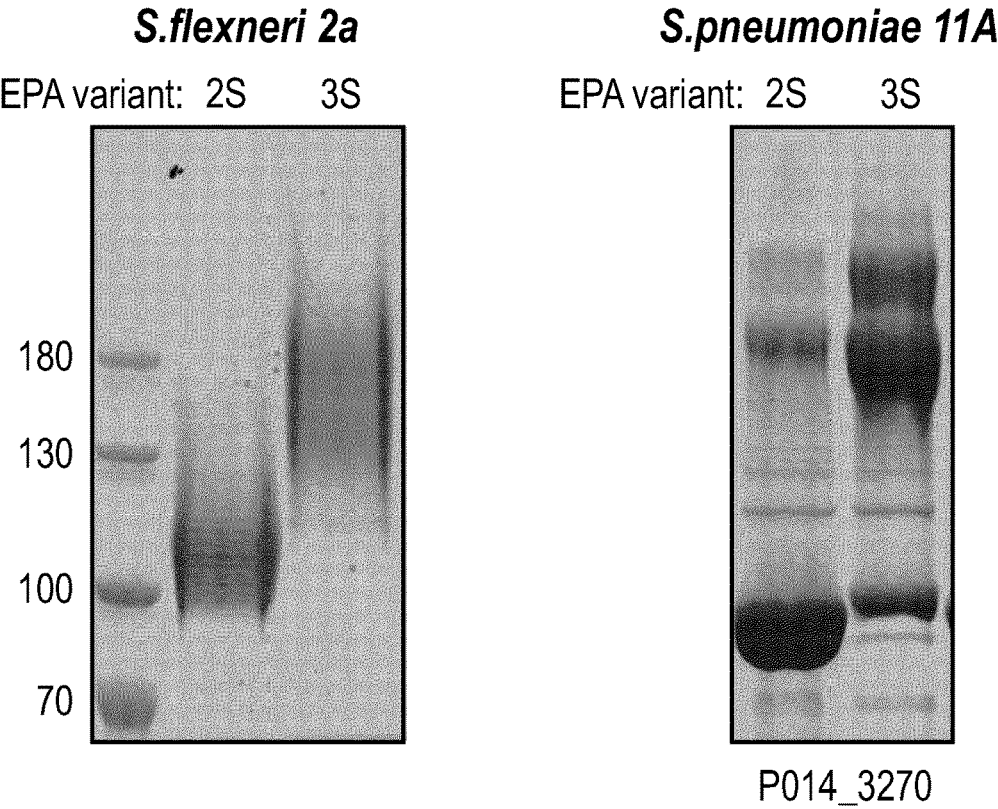


FIG. 8

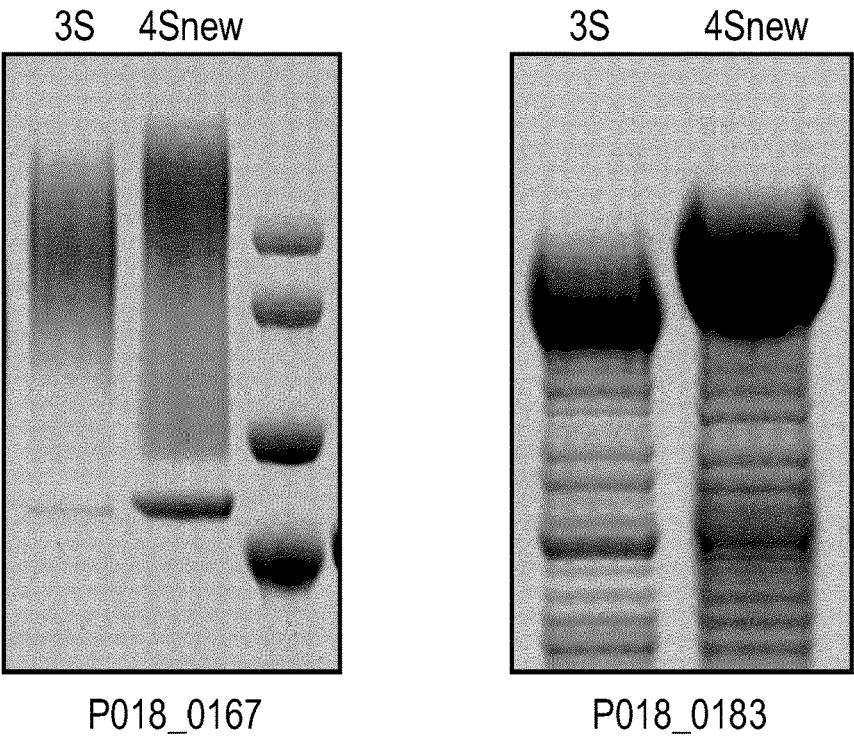
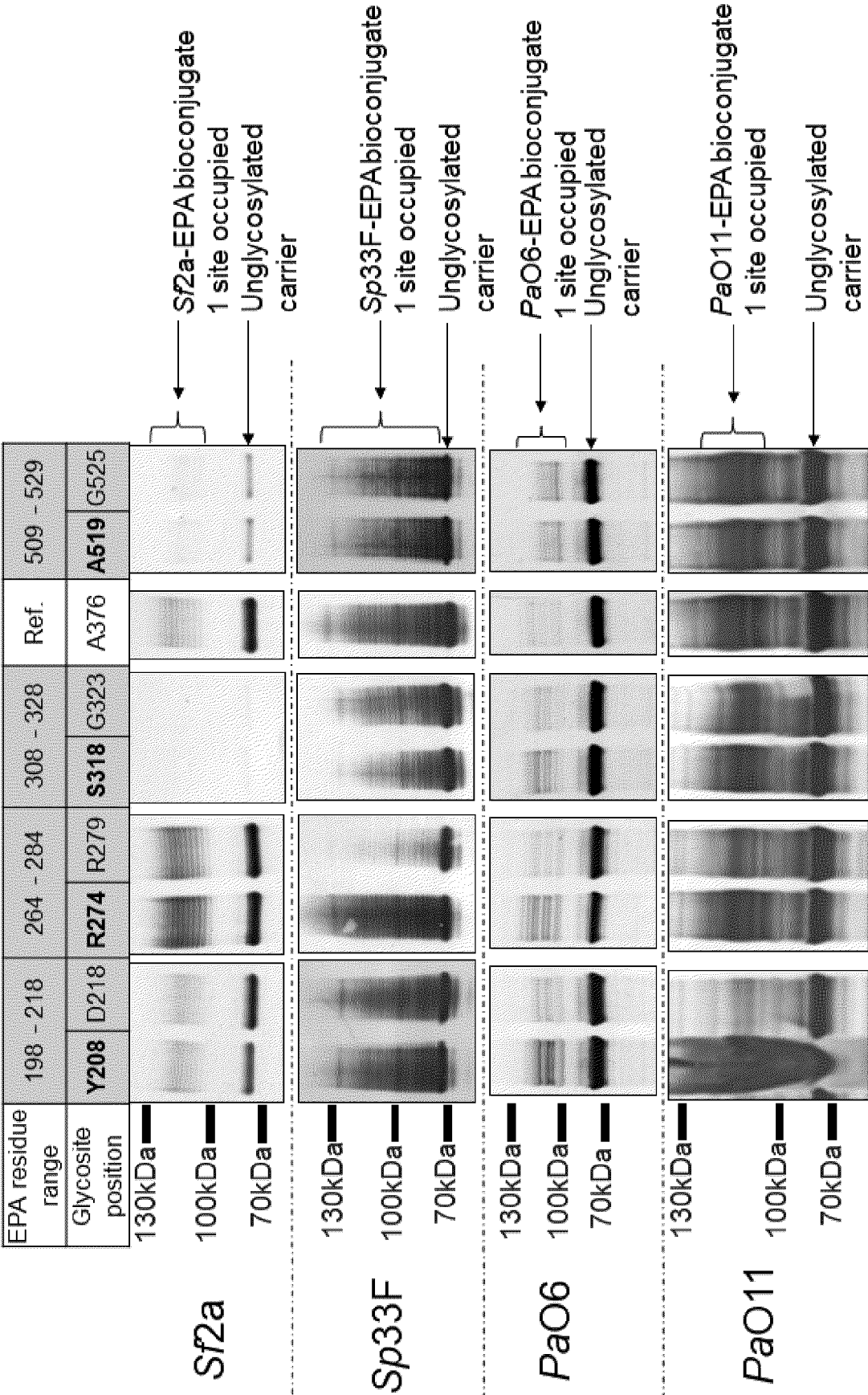


FIG. 9



MODIFIED EXOTOXIN A PROTEINS

FIELD OF THE INVENTION

[0001] The present invention relates to the field of modified proteins, immunogenic compositions and vaccines comprising the modified proteins, their manufacture and the use of such compositions in medicine. More particularly, it relates to a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein. The modified EPA can be used as a carrier protein for other antigens, particularly saccharide antigens or other antigens lacking T cell epitopes.

BACKGROUND TO THE INVENTION

[0002] Protein glycosylation is a common posttranslational modification in bacteria by which glycans are covalently attached to surface proteins, flagella, or pili, for example. Glycoproteins play roles in adhesion, stabilization of proteins against proteolysis, and evasion of the host immune response. Two protein glycosylation mechanisms are distinguished by the mode in which the glycans are transferred to proteins: one mechanism involves the transfer of carbohydrates directly from nucleotide-activated sugars to acceptor proteins (used in e.g. protein O-glycosylation in the Golgi apparatus of eukaryotic cells and flagellin O-glycosylation in some bacteria). A second mechanism involves the preassembly of a polysaccharide onto a lipid-carrier (by glycosyltransferases) which is then transferred to a protein acceptor by an oligosaccharyltransferase (OTase) (Faridmoayer et al., J. Bacteriology, pp. 8088-8098, 2007). This second mechanism is used in, e.g. N-glycosylation in the endoplasmic reticulum of eukaryotic cells, the well-characterized N-linked glycosylation system of *Campylobacter jejuni*, and the more recently characterized O-linked glycosylation systems of *Neisseria meningitidis*, *Neisseria gonococcus*, and *Pseudomonas aeruginosa*. For O-linked glycosylation (O-glycosylation), glycans are generally attached to a serine or threonine residue on the protein acceptor. For N-linked glycosylation (N-glycosylation), glycans are generally attached to an asparagine residue on the protein acceptor. It is possible to reconstitute the N-glycosylation of *C. jejuni* proteins by recombinantly expressing the *pgl* locus and acceptor glycoprotein in *E. coli* at the same time (Wacker et al. (2002) Science 298, 1790-1793).

[0003] WO2006/119987 (Aebi et al.) describes proteins, as well as means and methods for producing proteins, with efficiency for N-glycosylation in prokaryotic organisms in vivo. It further describes the introduction of N-glycans into recombinant proteins for modifying immunogenicity, stability, biological, prophylactic and/or therapeutic activity of said proteins, and the provision of a host cell that displays recombinant N-glycosylated proteins of the present invention on its surface. In addition, it describes a recombinant N-glycosylated protein comprising one or more of the following optimized amino acid sequence(s): D/E-X-N-Z-S/T, wherein X and Z may be any natural amino acid except Pro. The introduction of such optimized amino acid sequence(s) into proteins leads to proteins that are N-glycosylated by an oligosaccharyl transferase in these introduced positions.

[0004] Conjugate vaccines (vaccines comprising a carrier protein covalently linked to an immunogenic antigen) have been a successful approach for vaccination against a variety of bacterial infections. Conjugation of T-independent anti-

gens, for example saccharides, to carrier proteins has long been established as a way of enabling T-cell help to become part of the immune response for a normally T-independent antigen. In this way, an immune response can be enhanced by allowing the development of immune memory and boostability of the response. To increase conjugate vaccine production efficiency, in vivo methods (to produce a "bio-conjugate vaccine") have been in development. These in vivo methods leverage the N-glycosylation and O-glycosylation systems discussed above. For example, WO2009/104074 describes a *Shigella* bioconjugate vaccine comprising a protein carrier comprising exotoxin of *Pseudomonas aeruginosa* (EPA) that has been modified to contain at least one consensus sequence D/E-X-N-Z-S/T and WO2017/035181 describes *E. coli* O-antigens covalently bound to a detoxified exotoxin A of *P. aeruginosa* (EPA) carrier protein. However, it has been found that certain antigens, are less immunogenic than other antigens when conjugated to EPA carrier protein.

[0005] Exotoxin A of *Pseudomonas aeruginosa* (also known as "EPA", or "ETA"), is a secreted bacterial toxin, a member of the ADP-ribosyltransferasetoxin family. The native protein is a single polypeptide chain of 613 amino acids (67kDa). The protein consists of three main domains, domain Ia and b (receptor-binding domain), domain II (transmembrane domain) and domain III (catalytic NAD-ribosyltransferasedomain) (Allured et al., Proc. Natl. Acad. Sci. USA Vol. 83, pp. 1320-1324, March 1986). The last four residues (400-404) of Domain II together with Domain III (405-613) form the catalytic subunit of the toxin with ADP-ribosyltransferase activity (Siegal et al., 1989 J. Biol. Chem. 264, 14256-14261). A mutant form of *Pseudomonas aeruginosa* exotoxin A (ETA) carrying a deletion of glutamic acid-553, an important active-site residue, was expressed in an ETA-negative strain of *P. aeruginosa* and shown to be exported from the cells as efficiently as wild-type ETA. The mutant protein, purified from the culture medium, was devoid of ADP-ribosyltransferase activity (Kileen et al. Biochimica et Biophysica Acta, 1138 (1992) 162-166).

[0006] There exists a need for further EPA carrier proteins suitable for use in conjugation in vivo. There exists a need for further EPA carrier proteins so that there is a choice of EPA carrier proteins for different saccharide antigens. There also exists a need for further EPA carrier proteins which comprise at least one consensus sequence D/E-X-N-Z-S/T and display favourable or improved properties, e.g. increased glycosylation efficiency.

SUMMARY OF THE INVENTION

[0007] The present invention provides modified EPA proteins comprising at least one consensus sequence for glycosylation (e.g. D/E-X-N-Z-S/T) for use in conjugation to an antigen (e.g. bacterial polysaccharide). In the modified EPA proteins of the invention, the glycosylation consensus sequences are introduced into specific regions of the EPA carrier protein. The present inventors have found that the position of the consensus sequence in the EPA amino acid sequence can increase glycosylation efficiency and/or optimize the operation of the N-glycosylation site. Modified EPA proteins of the invention show higher site occupancy, higher sugar:protein ratio and/or higher yield. Differences in protein glycosylation may also influence the biological activity, antigenicity, stability and/or half-life of the protein.

In addition, increased glycosylation can assist the purification of proteins by chromatography. In a specific embodiment, the modified EPA proteins described herein are modified such that the number of glycosylation sites in the carrier proteins is optimized. This may allow for lower concentrations of the EPA protein to be administered in its conjugate form, e.g. in an immunogenic composition or vaccine. The present invention also provides further EPA carrier proteins so that there is a choice of EPA carrier proteins with different numbers of consensus sequence sites for different saccharide antigens. The number of glycosites may be selected to optimise the sugar:protein ratio. For example, for shorter saccharide antigens (glycans) a higher number of glycosites (using the modified EPA carrier proteins of the invention as described herein) may be used to increase the sugar:protein ratio.

[0008] Accordingly, there is provided in one aspect of the present invention, a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0009] There is also provided, a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid

residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0010] According to a further aspect of the invention, there is provided a conjugate (e.g. bioconjugate) comprising a modified EPA protein of the invention linked to an antigen (e.g. a saccharide antigen, optionally a bacterial polysaccharide antigen).

[0011] According to a further aspect of the invention, there is provided a polynucleotide encoding a modified EPA protein of the invention.

[0012] According to a further aspect of the invention, there is provided a vector comprising a polynucleotide encoding a modified EPA protein of the invention.

[0013] According to a further aspect of the invention, there is provided a host cell comprising:

[0014] i) one or more nucleotide sequences comprising polysaccharide synthesis genes, optionally for producing a bacterial polysaccharide antigen (e.g. an O-antigen from a Gram negative bacterium optionally from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium optionally from *Streptococcus pneumoniae* or *Staphylococcus aureus*) or a yeast polysaccharide antigen or a mammalian polysaccharide antigen, optionally integrated into the host cell genome;

[0015] ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid;

[0016] iii) a nucleotide sequence that encodes a modified EPA protein of the invention, optionally within a plasmid.

[0017] According to a further aspect of the invention, there is provided a process for producing a bioconjugate that comprises (or consists of) a modified EPA protein linked to a polysaccharide, said process comprising: (i) culturing the host cell of the invention under conditions suitable for the production of glycoproteins and (ii) isolating the bioconjugate produced by said host cell, optionally isolating the bioconjugate from a periplasmic extract from the host cell.

[0018] According to a further aspect of the invention, there is provided an immunogenic composition comprising a conjugate (e.g. bioconjugate) of the invention and optionally a pharmaceutically acceptable excipient and/or carrier.

[0019] According to a further aspect of the invention, there is provided a vaccine comprising an immunogenic composition of the invention and optionally an adjuvant.

[0020] According to a further aspect of the invention, there is provided a method of inducing an immune response in a subject (e.g. human), the method comprising administering a therapeutically or prophylactically effective amount of a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention or a vaccine of the invention to a subject (e.g. human) in need thereof.

[0021] According to a further aspect of the invention, there is provided a conjugate (e.g. bioconjugate) of the

invention, an immunogenic composition of the invention, or a vaccine of the invention, for use in inducing an immune response in a subject (e.g. human).

[0022] According to a further aspect of the invention, there is provided a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention, or a vaccine of the invention, for use in the manufacture of a medicament inducing an immune response in a subject (e.g. human).

DESCRIPTION OF DRAWINGS/FIGURES

[0023] FIG. 1 shows 3D structure of EPA presented as cartoon and positions Y208, R274, S318 and A519 selected for insertion of glycosites presented as spheres.

[0024] FIG. 2 shows sequence alignment of EPA variants containing single glycosite at positions Y208, R274, S318 and A519 compared to EPA without any glycosite. EPA_detox (SEQ ID NO: 1), EPA_mut_Y208 (SEQ ID NO: 34), EPA_mutR274 (SEQ ID NO: 35), EPA_mut_S318 (SEQ ID NO: 36), and EPA_mut_A519 (SEQ ID NO: 37).

[0025] FIG. 3 shows SDS-PAGE analysis of IMAC (Immobilized metal affinity chromatography) enriched periplasmic extract of *E. coli* strains producing KpO-antigen (*Klebsiella pneumoniae* O-antigen) polysaccharide and expressing PglB and EPA variants with a glycosite KDQNTAK (SEQ ID NO: 4) introduced at the following positions: Y208 (lane 1), K240 (lane 2), R274 (lane 3), S318 (lane 4), A376 (lane 5), A519 (lane 6), and K240 and A376 (lane 7). The bands corresponding to the unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one and two occupied glycosites are labelled.

[0026] FIG. 4 shows SDS-PAGE analysis of IMAC enriched periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with a glycosite KDQNTAK (SEQ ID NO: 4) introduced at the following positions: K240 and A376 (lane 1), Y208 and R274 (lane 2), Y208 and S318 (lane 3), Y208 and A519 (lane 4), R274 and S318 (lane 5), R274 and A519 (lane 6), S318 and A519 (lane 7), and Y208 and R274 and A519 (lane 8). The bands corresponding to the unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one, two and three occupied glycosites are labelled.

[0027] FIG. 5 shows immunoblot analyses of periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with 1 to 7 glycosites introduced at the following positions: Y208 (lane 1), K240 (lane 2), R274 (lane 3), S318 (lane 4), A376 (lane 5), A519 (lane 6), and K240 and A376 (lane 7), Y208 and R274 (lane 8), Y208 and S318 (lane 9), Y208 and A519 (lane 10), R274 and S318 (lane 11), R274 and A519 (lane 12), S318 and A519 (lane 13), Y208 and R274 and A519 (lane 14), N-terminal glycotag and K240 and A376 and C-terminal glycotag (lane 15), N-terminal glycotag and Y208 and R274 and A519 (lane 16), N-terminal glycotag and Y208 and R274 and A519 and C-terminal glycotag (lane 17), N-terminal glycotag and Y208 and S318 and A519 and C-terminal glycotag (lane 18), N-terminal glycotag and R274 and S318 and A519 and C-terminal glycotag (lane 19), N-terminal glycotag and Y208 and R274 and S318 and A519 and C-terminal glycotag (lane 20), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 (lane 21), and N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 and C-terminal glycotag (lane 22). The upper panel represents the immunoblot probed with anti-

KpO-antigen anti-serum, while the bottom panel represents the immunoblot probed with anti-EPA antibody. The bands corresponding to the unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one to seven occupied glycosites are labelled.

[0028] FIG. 6 shows SDS-PAGE analysis of periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with a glycosite KDQNTAK (SEQ ID NO: 4) introduced at the following positions: N-terminal glycotag and Y208 and R274 and A519 and C-terminal glycotag (lane 1), N-terminal glycotag and Y208 and S318 and A519 and C-terminal glycotag (lane 2), N-terminal glycotag and R274 and S318 and A519 and C-terminal glycotag (lane 3), N-terminal glycotag and Y208 and R274 and S318 and A519 and C-terminal glycotag (lane 4), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 (lane 5), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 and C-terminal glycotag (lane 6), and N-terminal glycotag and Y208 and R274 and A519 (lane 7). The bands corresponding to KpO-antigen-EPA bioconjugates are labelled with arrows. The first row of the table indicates the number of glycosites.

[0029] FIG. 7 shows SDS-PAGE analysis of IMAC enriched periplasmic extracts of *E. coli* strain producing Sf2a *Shigella flexneri* 2a (left) or Sp11A *Streptococcus pneumoniae* 11A (right) polysaccharide and expressing PglB and EPA variants with a glycosite introduced at K240 and a second glycosite at A376 (lane 1) or three glycosites at positions Y208, R274, and A519 (lane 2).

[0030] FIG. 8 shows SDS-PAGE analysis of IMAC enriched periplasmic extracts of *E. coli* strain producing two different *Klebsiella pneumoniae* O-antigen polysaccharides (left and right) and expressing PglB and EPA variants with glycosites at positions Y208 and R274 and A519 (lane 1), or with N-terminal glycotag and glycosites at positions Y208 and R274 and A519 (lane 2).

[0031] FIG. 9 shows immunoblot analyses of periplasmic extract of *E. coli* strains producing Sf2a, Sp33F, PaO6 and PaO11 antigen polysaccharide (bottom panel) and expressing PglB and EPA variants with 1 glycosite KDQNTAK (SEQ ID NO: 4) introduced at the following positions: Y208 (lane 1), D218 (lane 2), R274 (lane 3), R279 (lane 4), S318 (lane 5), G323 (lane 6), A376 (lane 7), A519 (lane 8), and G525 (lane 9). The bands corresponding to EPA bioconjugates and unglycosylated EPA carrier protein are labelled with arrows.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0032] Carrierprotein: a protein which may be covalently attached to an antigen (e.g. saccharide antigen, such as a bacterial polysaccharide antigen) to create a conjugate (e.g. bioconjugate). A carrier protein activates T-cell mediated immunity in relation to the antigen to which it is conjugated.

[0033] EPA: Exotoxin A of *Pseudomonas aeruginosa* (also known as “Exotoxin of *P. aeruginosa*”, “EPA”, or “ETA”)

[0034] Anyaminoacidexceptproline(pro,P): refers to an amino acid selected from the group consisting of alanine (ala, A), arginine (arg, R), asparagine (asn, N), aspartic acid (asp, D), cysteine (cys, C), glutamine (gln, Q), glutamic acid

(glu, E), glycine (gly, G), histidine (his, H), isoleucine (ile, I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), serine (ser, S), threonine (thr, T), tryptophan (trp, W), tyrosine (tyr, Y), and valine (val, V).

[0035] Naturally occurring amino acid residues: amino acids that are naturally incorporated into polypeptides. In particular, the 20 amino acids encoded by the universal genetic code: alanine (ala, A), arginine (arg, R), asparagine (asn, N), aspartic acid (asp, D), cysteine (cys, C), glutamine (gln, Q), glutamic acid (glu, E), glycine (gly, G), histidine (his, H), isoleucine (ile, I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), proline (pro, P), serine (ser, S), threonine (thr, T), tryptophan (trp, W), tyrosine (tyr, Y), and valine (val, V).

[0036] Glycosyltransferases (GTFs, Gtfs): enzymes that establish glycosidic linkages. Glycosyltransferases are enzymes that catalyze the formation of the glycosidic linkage to form a glycoside. For example, they catalyze the transfer of saccharide moieties from an activated nucleotide sugar (also known as the “glycosyl donor”) to a nucleophilic glycosyl acceptor molecule, the nucleophile of which can be oxygen-carbon-, nitrogen-, or sulfur-based.

[0037] O-Antigens (also known as O-specific polysaccharides or O-side chains): a component of the surface lipopolysaccharide (LPS) of Gram-negative bacteria. Examples include O-antigens from *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*.

[0038] Lipopolysaccharide (LPS): large molecules consisting of a lipid and a polysaccharide composed joined by a covalent bond.

[0039] Capsular polysaccharide (CP): polysaccharide found on the bacterial cell wall. Examples include capsular polysaccharide from *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Neisseria meningitidis* and *Staphylococcus aureus*.

[0040] wzy: a polysaccharide polymerase gene encoding an enzyme which catalyzes polysaccharide polymerization. The encoded enzyme transfers oligosaccharide units to the non-reducing end forming a glycosidic bond.

[0041] waalL: a O antigen ligase gene encoding a membrane bound enzyme. The encoded enzyme transfers undecaprenyl-diphosphate (UPP)-bound O antigen to the lipid A core oligosaccharide, forming lipopolysaccharide.

[0042] As used herein, the term “conjugate” refers to carrier protein covalently linked to an antigen.

[0043] As used herein, the term “bioconjugate” refers to conjugate between a protein (e.g. a carrier protein) and an antigen (e.g. a saccharide antigen, such as a bacterial polysaccharide antigen) prepared in a host cell background, wherein host cell machinery links the antigen to the protein (e.g. N-linked glycosylation).

[0044] As used herein, the term “modified protein” means a protein that is altered (in one or more way) as compared to wild type (e.g. a “modified EPA protein” excludes a wild type EPA protein).

[0045] As used herein, the term “immunogenic fragment” means a portion of an antigen smaller than the whole, that is capable of eliciting a humoral and/or cellular immune response in a host animal, e.g. human, specific for that fragment. Fragments of a protein can be produced using techniques known in the art, e.g. recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a

terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Typically, fragments comprise at least 10, 20, 30, 40 or 50 contiguous amino acids of the full length sequence. Fragments may be readily modified by adding or removing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40 or 50 amino acids from either or both of the N and C termini. A fragment of a modified EPA protein still comprises the recited modifications that are made to the EPA protein.

[0046] As used herein, the term “conservative amino acid substitution” involves substitution of a native amino acid residue with a non-native residue such that there is little or no effect on the size, polarity, charge, hydrophobicity, or hydrophilicity of the amino acid residue at that position, and without resulting in decreased immunogenicity. For example, these may be substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Conservative amino acid modifications to the sequence of a polypeptide (and the corresponding modifications to the encoding nucleotides) may produce polypeptides having functional and chemical characteristics similar to those of a parental polypeptide.

[0047] As used herein, the term “deletion” is the removal of one or more amino acid residues from the protein sequence. Typically, no more than about from 1 to 6 residues (e.g. 1 to 4 residues) are deleted at any one site within the protein molecule.

[0048] As used herein, the terms “insertion” or “addition” (including other tenses thereof such as “inserted”) means the addition of one or more non-native amino acid residues in the protein sequence or, as the context requires, addition of one or more non-native nucleotides in the polynucleotide sequence. Typically, no more than about from 1 to 10 residues, (e.g. 1 to 7 residues, 1 to 6 residues, or 1 to 4 residues) are inserted at any one site within the protein molecule.

[0049] As used herein, the term “added next to” is the addition of one or more non-native amino acid residues in the protein sequence at a position adjacent to the referenced amino acid or amino acid region. For example, “added next to one or more amino acids between amino acid residues 198-218” means the addition at a position adjacent to any one of amino acid residues 198-218 (including adjacent to amino acid residues 198 or 218).

[0050] As used herein, the term “glycosite” refers to an amino acid sequence recognized by a bacterial oligosaccharyltransferase, e.g. PglB of *C. jejuni*.

[0051] A “consensus sequence” is a sequence have a specific structure and/or function. As used herein, the term “consensus sequence” is a sequence comprising a glycosite. A consensus sequence may be selected from: a five amino acid consensus sequence D/E-X-N-Z-S/T (SEQ ID NO: 2), a seven amino acid consensus sequence K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)).

[0052] As used herein, the term “introduced at” is used herein to reference the location and manner of inserting a consensus sequence into an amino acid sequence. A glycosite which is introduced at an N-terminal or C-terminal position of a protein may be added next to the amino acid sequence at the N-terminus or C-terminus, whereas a consensus sequence (or glycosite) which is introduced at a

specific amino acid residue within the protein, e.g. Y208, may be substituted for that amino acid.

[0053] Unless specifically stated otherwise, providing a numeric range (e.g. “25-30”) is inclusive of endpoints (i.e. includes the values 25 and 30). For example, “between amino acids 198 to 218 . . . of SEQ ID NO: 1” refers to position in the amino acid sequence between amino acid 198 and amino acid 218 of SEQ ID NO: 1 including both amino acids 198 and 218.

[0054] The terms “identical” or percent “identity” refer to nucleotide sequences or amino acid sequences that are the same or have a specified percentage of nucleotide residues or amino acid residues that are the same (e.g. 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identity over a specified region), when compared and aligned for maximum correspondence using, for example, sequence comparison algorithms or by manual alignment and visual inspection. Identity between polypeptides may be calculated by various algorithms. In general, when calculating percentage identity the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting “gaps” in either one or both sequences, to enhance the degree of alignment. For example the Needleman Wunsch algorithm (Needleman and Wunsch 1970, J. Mol. Biol. 48: 443-453) for global alignment, or the Smith Waterman algorithm (Smith and Waterman 1981, J. Mol. Biol. 147: 195-197) for local alignment may be used, e.g. using the default parameters (Smith Waterman uses BLOSUM 62 scoring matrix with a Gap opening penalty of 10 and a Gap extension penalty of 1). A preferred algorithm is described by Dufresne et al. in Nature Biotechnology in 2002 (vol. 20, pp. 1269-71) and is used in the software GenePAST (Genome Quest Life Sciences, Inc. Boston, Mass.). The GenePAST “percent identity” algorithm finds the best fit between the query sequence and the subject sequence, and expresses the alignment as an exact percentage. GenePAST makes no alignment scoring adjustments based on considerations of biological relevance between query and subject sequences. Identity between two sequences is calculated across the entire length of both sequences and is expressed as a percentage of the reference sequence (e.g. SEQ ID NO: 1 of the present invention).

[0055] As used herein the term “recombinant” means artificial or synthetic. In certain embodiments, a “recombinant protein” refers to a protein that has been made using recombinant nucleotide sequences (nucleotide sequences introduced into a host cell). In certain embodiments, the nucleotide sequence that encodes a “recombinant protein” is heterologous to the host cell.

[0056] As used herein the terms “isolated” or “purified” mean a protein, conjugate (e.g. bioconjugate), polynucleotide, or vector in a form not found in nature. This includes, for example, a protein, conjugate (e.g. bioconjugate), polynucleotide, or vector having been separated from host cell or organism (including crude extracts) or otherwise removed from its natural environment. In certain embodiments, an isolated or purified protein is a protein essentially free from all other polypeptides with which the protein is innately associated (or innately in contact with).

[0057] As used herein, the term “subject” refers to an animal, in particular a mammal such as a primate (e.g. human).

[0058] As used herein, the term “effective amount,” in the context of administering a therapy (e.g. an immunogenic

composition or vaccine of the invention) to a subject refers to the amount of a therapy which has a prophylactic and/or therapeutic effect(s). In certain embodiments, an “effective amount” refers to the amount of a therapy which is sufficient to achieve one, two, three, four, or more of the following effects: (i) reduce or ameliorate the severity of a bacterial infection or symptom associated therewith; (ii) reduce the duration of a bacterial infection or symptom associated therewith; (iii) prevent the progression of a bacterial infection or symptom associated therewith; (iv) cause regression of a bacterial infection or symptom associated therewith; (v) prevent the development or onset of a bacterial infection, or symptom associated therewith; (vi) prevent the recurrence of a bacterial infection or symptom associated therewith; (vii) reduce organ failure associated with a bacterial infection; (viii) reduce hospitalization of a subject having a bacterial infection; (ix) reduce hospitalization length of a subject having a bacterial infection; (x) increase the survival of a subject with a bacterial infection; (xi) eliminate a bacterial infection in a subject; (xii) inhibit or reduce a bacterial replication in a subject; and/or (xiii) enhance or improve the prophylactic or therapeutic effect(s) of another therapy.

[0059] The term “comprises” is open-ended and means “includes.” Thus, unless the context requires otherwise, the word “comprises” or “has”, and variations thereof (including “comprise” and “comprising” or “have” and “having”, respectively), will be understood to imply the inclusion of a stated compound(s), molecule(s), composition(s), or steps, but not to the exclusion of any other compound(s), molecule(s), composition(s), or steps. The terms “comprising” and “having” when used as a transition phrase herein are open-ended whereas the term “consisting of” when used as a transition phrase herein is closed (i.e., limited to that which is listed and nothing more). In certain embodiments and for readability, the word “is” may be used as a substitute for “consists of” or “consisting of”. The abbreviation, “e.g.” is derived from the Latin *exempli gratia*, and is used herein to indicate a non-limiting example. Thus, the abbreviation “e.g.” is synonymous with the term “for example”.

EPA Protein

[0060] Exotoxin A of *Pseudomonas aeruginosa* (also known as “EPA”, or “ETA”), is a secreted bacterial toxin, a member of the ADP-ribosyltransferasetoxin family. An EPA protein useful in the invention can be produced by methods known in the art in view of the present disclosure, see for example Ihssen et al. (2010) Microbial Cell Factories 9:61, WO 2006/119987, WO 2009/104074 and WO2015124769A1. Exotoxin A from *Pseudomonas aeruginosa* strain PA103 was cloned and sequenced by Gray et al. (1984) Proc. Natl. Acad. Sci. USA Vol. 81, pp. 2645-2649. Comparison of the deduced NH₂-terminal amino acid sequence with that determined by sequence analysis of the secreted protein indicated that EPA was made as a 638 amino acid precursor from which a highly hydrophobic leader peptide of 25 amino acids is removed during the secretion process (see FIG. 1 of Gray et al. (1984)). Sequences within the EPA structural gene appear to be well-conserved from strain to strain (Vasil et al. (1986) Infection and Immunity, May 1986 pages 538 to 548).

[0061] Because EPA is a toxin, it needs to be detoxified (i.e. rendered non-toxic to a mammal, e.g. human, when provided at a dosage suitable for protection) before it can be

administered in vivo. A modified EPA protein of the invention may be genetically detoxified (i.e. by mutation). The genetically detoxified sequences may remove undesirable activities such as ADP-ribosyltransferase activity, in order to reduce the toxicity, whilst retaining the ability to induce anti-EPA protective and/or neutralizing antibodies following administration to a human. The genetically detoxified sequences may maintain their immunogenic epitopes. A modified EPA protein may be genetically detoxified by one or more point mutations. For example, detoxification can be achieved by mutating and deleting catalytically essential residues, such as substitution of leucine 552 to valine (L552V) and by deletion of glutamic acid-553 (Δ E553), according to Lukac et al. (1988), Infect Immun, 56: 3095-3098, and Ho et al. (2006), Hum Vaccin, 2:89-98. Detoxification can be achieved by mutating/deleting the catalytically essential residues L552V Δ E553 using quick change mutagenesis (Stratagene) and phosphorylated oligonucleotides 5'-GAAGGCGGGCGCGTGACCATCTCTCGGC (SEQ. ID NO. 38) and 5'-GCCGAGAAATGGT-CACGCGCCCGCCTTC (SEQ. ID NO. 39) resulting in construct pGVXN70. Detoxification can be measured by determining the inhibition of ADP-ribosyltransferase and cytotoxic activity according to the methodology described in Lukac et al. (1988), Infect Immun, 56: 3095-3098, and references cited therein, namely Douglas et al (1987) J. Bacteriol 169: 4962-4966 and Douglas et al (1987). A detoxified EPA has ADP-ribosyltransferase and cytotoxic activities lower than wild-type EPA, suitably the same as or less than that of the modified EPA described in Lukac et al

(1988) i.e. Δ E553 EPA (EPA having deletion of glutamic acid-553). Accordingly, the modified EPA protein of the invention may be the amino acid sequence of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1). The modified EPA protein may comprise substitution of leucine 552 to valine (L552V) and deletion of glutamine 553 (Δ E553) with reference to the amino acid sequence of SEQ ID NO: 1 (or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1). The modified EPA protein of the invention may be the amino acid sequence of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1) comprising substitution of leucine 552 to valine (L552V). The modified EPA protein of the invention, may be the amino acid sequence of SEQ ID NO: 1 (or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1) comprising deletion of glutamine 553 (Δ E553). Preferably, the modified EPA protein of the invention, is modified by substitution of leucine 552 to valine (L552V) and deletion of glutamine 553 (Δ E553). Preferably, the modified EPA protein of the invention, is the amino acid sequence of SEQ ID NO: 1 (or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1) comprising substitution of leucine 552 to valine (L552V) and deletion of glutamine 553 (Δ E553). 1.

[0062] In one aspect of the invention, EPA protein has the amino acid sequence of SEQ ID NO: 1:

EPA sequence SEQ ID NO: 1

AEEAFDLWNECAKACVLDLKDGVRSRMSVDPAPAIADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNLVLPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLDPLDGVVNYLAQQRCLDDTWE
 GKIYRVLAGNPAKHDLIDKPTVISHRHLFPEGGLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGYPVQRLV
 ALYLAARLSWNQVDQVIRNALASPGSGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAASADWS
 LTCPVAAGECAGPADSGDALLERNYPTGAFLGDGDDVSFSTRGTQNWTVRLLQAHRLQLEERGYPVFGYHGT
 FLEAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVVPRWSLPGFYRTGL
 TLAAPAAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAI
 SALPDYASQPGKPPREDLK

EPA sequence (amino acids 1 to 612 with numbering) SEQ ID NO: 1

10	20	30	40	50	60
AEEAFDLWNE	CAKACVLDLK	DGVRSRMSV	DPAIADTNGQ	GVLYHSMVLE	GGNDALKLAI
70	80	90	100	110	120
DNALSITSDG	LTIRLEGGVE	PNKPVRYSYT	RQARGSWSLN	WLVPIGHEKP	SNIKVFIHEL
130	140	150	160	170	180
NAGNQLSHMS	PIYTIEMGDE	LLAKLARDAT	FFVRAHESNE	MQPTLAISHA	GSVSVVMAQAQ
190	200	210	220	230	240
PRREKRSEW	ASGKVLCLLD	PLDGVVNYLA	QQRCLDDTW	EGKIYRVLAG	NPAKHDLIDK
250	260	270	280	290	300
PTVISHRLHF	PEGGLAALTA	AHQACHLPLE	AFTRHRQPRG	WEQLEQCGYP	VQRLVALYLA
310	320	330	340	350	360
ARLSWNQVDQ	VIRNALASPG	SGDGLGEAIR	EQPEQARLAL	TAAAESERF	VRQGTGNDEA

- continued

```

      370      380      390      400      410      420
GAASADV VSL TCPVAAGECA GPADSGDALL ERNYPTGAEF LGDGGDV SFS TRGTQNW TVE

      430      440      450      460      470      480
RLQAH RQLE ERGYV FVG YH GTFLEAAQSI VFGGVRARSQ DLDAIWRGFY IAGDPALAYG

      490      500      510      520      530      540
YAQDQEP DAR GRIRNGALLR VYVPRWSLPG FYRTGLTLAA PEAAGEVERL IGHPLPLRLD

      550      560      570      580      590      600
AITGPEEEGG RVTILGWPLA ERTVVIPSAI PTDPRNVGGD LDPSSIPDKE QAISALPDYA

      610
SQPGKPPRED LK

```

[0063] The term “modified EPA protein” refers to a EPA amino acid sequence (for example, having a amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1), which EPA amino acid sequence has been modified by the addition, substitution or deletion of one or more amino acids (for example, by addition of a consensus sequence(s) selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) and/or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)); and/or by substitution of one or more amino acids by a consensus sequence(s) selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3)). For example, a modified EPA protein may be an EPA amino acid sequence of SEQ ID NO: 1 which has been modified in that the amino acid sequence comprises one (or more) consensus sequence (s) selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) and/or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)). As used herein, in consensus sequences of the present invention X and Z are independently any amino acid except proline; preferably, X is Q (glutamine) and Z is A (alanine). The modified EPA protein may also comprise further modifications (additions, substitutions, deletions). Preferably, the modified EPA protein of the invention comprises substitution of leucine 552 to valine (L552V) and deletion of glutamine 553 (Δ E553) with reference to the amino acid sequence of SEQ ID NO: 1 (or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1). In an embodiment, the modified EPA protein of the invention is a non-naturally occurring EPA protein (i.e. not native). A modified EPA protein of the invention may have an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 80% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 85% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 90% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 91% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 92% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 93% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 94% identical to SEQ ID NO: 1. A

modified EPA protein of the invention may have an amino acid sequence at least 95% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 96% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 97% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 98% identical to SEQ ID NO: 1. A modified EPA protein of the invention may have an amino acid sequence at least 99% identical to SEQ ID NO: 1.

[0064] The present invention provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, selected from specific amino acid residues within the EPA protein (consensus sequence sites). These consensus sequence sites are independently selected from (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the present invention provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues

198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. The numbering of the amino acid residues as specified herein, for example in (i) to (iv) above, refers to the amino acid position in SEQ ID NO: 1 (or where an amino acid sequence is at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 to an equivalent position to that of SEQ ID NO: 1 if this sequence was lined up with an amino acid sequence of SEQ ID NO: 1 in order to maximise the sequence identity between the two sequences).

[0065] The present invention also provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. The numbering of the amino acid residues as specified herein, for example in (i) to (iv) above, refers to the amino acid position in SEQ ID NO: 1 (or where an amino acid sequence is at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 to an equivalent position to that of SEQ ID NO: 1 if this sequence was lined up with an amino acid sequence of SEQ ID NO: 1 in order to maximise the sequence identity between the two sequences).

[0066] The present invention also provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises two (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein the two (or more) consensus sequences have each been added next to or substituted for one or more amino acids, selected from specific amino acid residues within the EPA protein (consensus sequence sites). These consensus sequence sites are independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the present invention provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises two (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the two (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519), and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240), of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. The present invention also provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1,

modified in that the amino acid sequence comprises two (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein the two (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519), of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0067] The two (or more) consensus sequences may also be independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519), and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0068] The present invention also provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises three (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein the three (or more) consensus sequences have each been added next to or substituted for one or more amino acids, selected from specific amino acid residues within the EPA protein (consensus sequence sites). These consensus sequence sites are independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279,

e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the present invention provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises three (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the three (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0069] The three (or more) consensus sequences may also be independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519), and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g.

amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0070] The present invention also provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises four (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein the four (or more) consensus sequences have each been added next to or substituted for one or more amino acids, selected from specific amino acid residues within the EPA protein (consensus sequence sites). These consensus sequence sites are independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the present invention provides a modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises four (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the four (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i)

one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0071] The four (or more) consensus sequences may also be independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519), and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

EPA sequence (with numbering) with amino acids Y208, K240, R274, S318, and A519 underlined

SEQ ID NO: 1

10	20	30	40	50	60
AEEAFDLWNE	CAKACVLDLK	DGVRSSRMVS	DPAIADTNGQ	GVLHYSMVLE	GGNDALKLAI
70	80	90	100	110	120
DNALSITSDG	LTIRLEGGVE	PNKPVRYST	RQARGSWSLN	WLVPIGHEKP	SNIKVFIHEL
130	140	150	160	170	180
NAGNQLSHMS	PIYTIEMGDE	LLAKLARDAT	FFVRAHESNE	MQPTLAISHA	GVSVMQAQAQ
190	200	210	220	230	240
PRREKRWSEW	ASGKVLCLLD	PLDGVNYLA	QQRCLNDDTW	EGKIYRVLAG	NPAKHDLDTK
250	260	270	280	290	300
PTVISHRLHF	PEGGSLAALT	AHQACHLPLE	AFTRHRQPRG	WEQLEQCQGP	VQRLVALYLA
310	320	330	340	350	360
ARLSWNQVDQ	VIRNALASPG	SGGDLGEAIR	EQPEQARLAL	TLAAAESERF	VRQGTGNDEA

-continued

```

      370      380      390      400      410      420
GAASADVLSL TCPVAAGECA GPADSGDALL ERNYPTGAEF LGDGGDVSPS TRGTQNWTVL

      430      440      450      460      470      480
RLAQHRQLE ERGYVFGYH GTFLEAAQSI VFGGVRARSQ DLDAIWRGFY IAGDPALAYG

      490      500      510      520      530      540
YAQDQEPDAR GRIRNGALLR VYVPRWSLPG FYRTGLTLAA PEAAGEVERL IGHPLPLRLD

      550      560      570      580      590      600
AITGPEEEGG RVTILGWPLA ERTVVIPSAI PTDPRNVGGD LDPSSIPDKE QAISALPDYA

      610
SQPGKPPRED LK

```

[0072] In an embodiment, the modified EPA protein of the invention may be derived from an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 which is an immunogenic fragment and/or a variant of SEQ ID NO: 1.

[0073] In an embodiment, the modified EPA protein of the invention may be derived from an immunogenic fragment of SEQ ID NO: 1 comprising at least about 15, at least about 20, at least about 40, or at least about 60 contiguous amino acid residues of SEQ ID NO: 1, wherein said polypeptide is capable of eliciting an antibodies which bind to SEQ ID NO: 1.

[0074] Native EPA is known to consist of three distinct structural domains (Allured et al., Proc. Natl. Acad. Sci. USA Vol. 83, pp. 1320-1324, March 1986):

[0075] Domain I, is an antiparallel β -structure. It includes residues 1-252 and residues 365-404. It has 17 β -strands. The first 13 strands form the structural core of an elongated β -barrel. Following strand 13 of domain I, the peptide chain traverses one face of the barrel, leading into the second domain.

[0076] Domain II (residues 253-364) is composed of six consecutive α -helices with one disulfide linking helix A and helix B. Helices B and E are approximately 30 Å in length; helices C and D are approximately 15 Å long.

[0077] Domain III is comprised of the carboxyl-terminal third of the molecule, residues 405-613. The most notable structural feature of domain III is its extended cleft. The domain has a less regular secondary structure than domains I and II.

An immunogenic fragment of EPA protein of the invention may be generated by removing and/or modifying one or more of these domains. In an embodiment, the immunogenic fragment of SEQ ID NO: 1 may comprise the amino acid residues of Domain I (residues 1-252 and residues 365-404) of SEQ ID NO: 1. In another embodiment, the immunogenic fragment of SEQ ID NO: 1 may comprise the amino acid residues of Domain II (residues 253-364) of SEQ ID NO: 1. In another embodiment, the immunogenic fragment of SEQ ID NO: 1 may comprise at least the amino acid residues of Domain II (residues 253-364) of SEQ ID NO: 1 and Domain III (residues 405-612) of SEQ ID NO: 1. In another embodiment, the immunogenic fragment of SEQ ID NO: 1 may comprise the amino acid residues of Domain I (residues 1-252 and residues 365-404) of SEQ ID NO: 1 and Domain II (residues 253-364) of SEQ ID NO: 1. In another embodiment, the immunogenic fragment of SEQ ID NO: 1 may comprise at least the amino acid residues of Domain II (residues 253-364) of SEQ ID NO: 1 and Domain III (residues 405-612) of SEQ ID NO: 1.

[0078] In EPA there are eight cysteines form forming disulfides in sequential order: Cys-11 forms a disulfide with Cys-15, Cys-197 forms a disulfide with Cys-214, Cys-265 forms a disulfide with Cys-287, and Cys-372 forms a disulfide with Cys-379. Suitably, the immunogenic fragment of SEQ ID NO: 1 comprises the eight cysteines of SEQ ID NO: 1: Cys-11, Cys-15, Cys-197, Cys-214, Cys-265, Cys-287, Cys-372 and Cys-379 (i.e. these residues are not modified). Thus, suitably the modified EPA protein of the invention comprises the eight cysteines of SEQ ID NO: 1: Cys-11, Cys-15, Cys-197, Cys-214, Cys-265, Cys-287, Cys-372 and Cys-379 or equivalent cysteines within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0079] In an embodiment, the modified EPA protein of the invention may be derived from an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 which is a variant of SEQ ID NO: 1 and differs from SEQ ID NO: 1 by the deletion and/or addition and/or substitution of one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 amino acids). Amino acid substitution may be conservative or non-conservative. In one aspect, amino acid substitution is conservative. Substitutions, deletions, additions or any combination thereof may be combined in a single variant so long as the variant is an immunogenic polypeptide. In an embodiment, the modified EPA protein of the present invention may be derived from a variant of SEQ ID NO: 1 in which 1 to 10, 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1 amino acids of SEQ ID NO: 1 have been substituted or deleted.

[0080] Suitably the immunogenic fragment and/or a variant of SEQ ID NO: 1 comprises a B-cell or T-cell epitope. Such epitopes may be predicted using a combination of 2D-structure prediction, e.g. using the PSIPRED program (from David Jones, Brunel Bioinformatics Group, Dept. Biological Sciences, Brunel University, Uxbridge UB8 3PH, UK) and antigenic index calculated on the basis of the method described by Jameson and Wolf (CABIOS 4:181-186 [1988]).

[0081] In modified EPA proteins of the invention one or more consensus sequences have each been added next to, or substituted for one or more amino acids of SEQ ID NO: 1 or a EPA amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In an embodiment of the invention, one or more amino acids (e.g. 1-7 amino acids, e.g. one amino acid) of the EPA amino acid sequence (for example, having an amino acid sequence of SEQ ID NO: 1 or a EPA amino acid sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1)

is/are substituted by a five amino acid D/E-X-N-Z-S/T (SEQ ID NO: 2) or by a seven amino acid K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4) also referred to as “KDQNA TK”) consensus sequence. For example, a single amino acid in the EPA amino acid sequence (e.g. SEQ ID NO: 1) may be substituted (i.e. replaced) with a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence. Alternatively, 2, 3, 4, 5, 6 or 7 amino acids in the EPA amino acid sequence (e.g. SEQ ID NO: 1 or a EPA amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1) may be substituted (i.e. replaced) with a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence. Preferably, at an internal consensus sequence site (i.e. a consensus sequence site that is within the EPA amino acid sequence rather than added next to the N-terminal or C-terminal amino acid), a single amino acid in the EPA amino acid sequence (e.g. SEQ ID NO: 1) is substituted (i.e. replaced) with a K-D-Q-N-A-T-K (SEQ ID NO: 4) consensus sequence. The classical 5 amino acid glycosylation consensus sequence (D/E-X-N-Z-S/T (SEQ ID NO: 2)) may also be extended by 1-5 other amino acid residues either side of the consensus sequence for more efficient glycosylation. For example, an extended consensus sequence may be J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5). As used herein, J and U are independently 1 to 5 naturally occurring amino acid residues, preferably J and U are independently 1 to 5 amino acid residues independently selected from glycine and/or serine, e.g. B may be G-S-G-G and U may be G-S-G-G. For example, an extended consensus sequence may be G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25). Preferably, an extended consensus sequence, such as J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) or G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25) is used where the consensus sequence is added next to the N-terminal or C-terminal amino acid of the EPA protein.

[0082] A combination of consensus sequences selected from: a five amino acid consensus sequence D/E-X-N-Z-S/T (SEQ ID NO: 1), a seven amino acid consensus sequence K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) and an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be used. For example, 1, 2, 3, 4 or 5, amino acids within the carrier protein amino acid sequence may each independently be substituted (i.e. replaced) with a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequence, wherein X and Z are independently any amino acid except proline (preferably wherein X is Q (glutamine), Z is A (alanine)) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4), and 1 or 2 consensus sequences J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) wherein J and U are independently 1 to 5 naturally occurring amino acid residues (preferably J and U are independently 1 to 5 amino acid residues independently selected from glycine and/or serine, e.g. G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25)) may be added next to the N-terminal or C-terminal amino acids of the carrier protein. Thus, a carrier protein may comprise 1, 2, 3, 4 or 5 consensus sequences selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline (preferably wherein X is Q (glutamine), Z is A (alanine)), and the carrier protein may further comprise 1 or 2 extended consensus sequences J-D/E-X-N-Z-S/

T-U (SEQ ID NO: 5) wherein J and U are independently 1 to 5 naturally occurring amino acid residues (preferably J and U are independently 1 to 5 amino acid residues independently selected from glycine and/or serine, e.g. G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25). Introduction of one or more consensus sequence(s) selected from: a five amino acid consensus sequence D/E-X-N-Z-S/T (SEQ ID NO: 2), a seven amino acid consensus sequence K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) and/or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) according to the present invention enables the modified EPA protein to be glycosylated. Thus, the present invention also provides a modified EPA protein of the invention wherein the modified EPA protein is glycosylated.

Position of the Consensus Sequence(s)

[0083] As described above, the present inventors have found that the position of the consensus sequence at specific regions/amino acids in the EPA amino acid sequence can increase glycosylation efficiency and/or optimize the operation of the N-glycosylation site.

[0084] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) between amino acids 198-218 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. More particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 203-213 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be substituted for amino acid residue Y208 or D218 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Even more particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 205-211 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Preferably, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be substituted for amino acid residue Y208 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0085] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) between amino acids 264-284 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%,

position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0087] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) between amino acids 509-529 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. More particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 514-524 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be substituted for amino acid residue A519 or G525 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Even more particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 516-522 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Preferably, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be substituted for amino acid residue A519 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0088] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) between amino acids 230-250 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. More particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 235-245 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Even more particularly, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence may be added next to or substituted for one or more amino acids residues between amino acids 237-243 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Preferably, a D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) consensus sequence

may be substituted for amino acid residue K240 of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0089] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to or substituted for one or more amino acids residue(s) within the N-terminal 10 amino acids of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. More particularly, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to or substituted for one or more amino acids residue(s) within the N-terminal 5 amino acids of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Preferably, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to the N-terminal amino acid of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0090] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to or substituted for one or more amino acids residue(s) within the C-terminal 10 amino acids of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. More particularly, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to or substituted for one or more amino acids residue(s) within the C-terminal 5 amino acids of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Preferably, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence (e.g. J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5)) may be added next to the C-terminal amino acid of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0091] In the modified EPA protein of the invention, the consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, D218, R274, R279, S318,

[0092] G323, A519 and G525; e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318 and A519) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In the modified EPA protein of the invention, two or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, D218, R274, R279, S318, G323, A519, G525 and K240; e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In the modified EPA protein of the invention, three or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, D218, R274, R279, S318, G323, A519, G525 and K240; e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In the modified EPA protein of the invention, four or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, D218, R274, R279, S318, G323, A519, G525 and K240; e.g. each consensus sequence may be substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. In the modified EPA protein of the invention, five or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, D218, R274, R279, S318, G323, A519, G525 and K240; e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0093] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) (e.g. G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25)) may

be added next to or substituted for one or more amino acids residue(s) at the N-terminus of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the modified EPA protein of the invention may comprise a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, added next to, or substituted for, one or more amino acids, at the N-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, added next to the N-terminal amino acid of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0094] In modified EPA proteins of the invention, a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) or an extended consensus sequence J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) (e.g. G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25)) may be added next to or substituted for one or more amino acids residue(s) at the C-terminus of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus, the modified EPA protein of the invention may comprise a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, added next to, or substituted for, one or more amino acids, at the C-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, added next to the C-terminal amino acid of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0095] In the modified EPA protein of the invention, three or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) and a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next

to or substituted for one or more amino acids residue(s) at the N-terminus (e.g. added next to the N-terminal amino acid) and a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) C-terminus (e.g. added next to the C-terminal amino acid) of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0096] In the modified EPA protein of the invention, four or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may each be independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) and a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) at the N-terminus (e.g. added next to the N-terminal amino acid) and/or a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) C-terminus (e.g. added next to the C-terminal amino acid) of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0097] In the modified EPA protein of the invention, five or more consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) may be substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318, A519 and K240) and a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) at the N-terminus (e.g. added next to the N-terminal amino acid) and/or a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) may be added next to or substituted for one or more amino acids residue(s) C-terminus (e.g. added next to the C-terminal amino acid) of SEQ ID NO: 1 or an equivalent position in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0098] The modified EPA protein of the invention may comprise at least one consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, has been added next to, or substituted for: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), or (iii) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid

residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may contain a consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), which has been added next to, or substituted for one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208) of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may contain a consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), which has been added next to, or substituted for one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318) of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may contain a consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), which has been added next to, or substituted for one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0099] The modified EPA protein of the invention may contain two consensus sequences. The modified EPA protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises two consensus sequences, e.g. wherein two consensus sequences are added next to or substituted for two amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein of the invention may contain two consensus sequences, optionally substituted for amino acid residues selected from: (i) Y208 and R274, (ii) Y208 and S318, (iii) Y208 and A519, (iv) R274 and S318, (v) R274 and A519, or (vi) S318 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues Y208 and R274 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 6. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues Y208 and S318 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an

amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 28. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues Y208 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 29. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues R274 and S318 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 30. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues R274 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 31. The modified EPA protein of the invention may contain two consensus sequences substituted for amino acid residues S318 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 7.

[0100] The modified EPA protein of the invention may contain three consensus sequences. The modified EPA protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises three consensus sequences, e.g. wherein three consensus sequences are added next to or substituted for three independently selected amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein of the invention may contain three consensus sequences, optionally substituted for amino acid residues Y208, R274 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 7.

[0101] The modified EPA protein of the invention may contain four consensus sequences. The modified EPA protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises four consensus sequences, e.g. wherein four consensus sequences are added next to or substituted for four independently selected amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein

of the invention may contain four consensus sequences, optionally substituted for amino acid residues Y208, R274, A519 and added next to the N-terminal amino acid of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, the modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 8.

[0102] The modified EPA protein of the invention may contain five consensus sequences. The modified EPA protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises five consensus sequences, e.g. wherein five consensus sequences are added next to or substituted for five independently selected amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein of the invention may contain five consensus sequences, optionally selected from: substitution of amino acid residue Y208, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus (i.e. added at the N-terminus) and addition at the C-terminus (i.e. added at the C-terminus) of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, a modified EPA protein may comprise (or consist of) (i) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 9, (ii) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 10, or (iii) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 11.

[0103] The modified EPA protein of the invention may contain six consensus sequences. The modified EPA protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises six consensus sequences, e.g. wherein six consensus sequences are added next to or substituted for six independently selected amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein of the invention may contain six consensus sequences, optionally selected from: substitution of amino acid residue Y208, substitution of amino acid residue K240, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus (i.e. added at the N-terminus) and addition at the C-terminus (i.e. added at the C-terminus) of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, a modified EPA protein of the invention may comprise (or consist of) (i) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 12, or (ii) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 13.

[0104] The modified EPA protein of the invention may contain seven consensus sequences. The modified EPA

protein may have an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 modified in that the amino acid sequence comprises seven consensus sequences, e.g. wherein seven consensus sequences are added next to or substituted for seven independently selected amino acid residues of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. Thus the modified EPA protein of the invention may contain seven consensus sequences, optionally substitution of amino acid residues Y208, K240, R274, S318, A519, addition at the N-terminus (i.e. added at the N-terminus) and addition at the C-terminus (i.e. added at the C-terminus) of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1. For example, a modified EPA protein of the invention may comprise (or consist of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 14.

[0105] It will be understood by a person skilled in the art, that reference to “between amino acids . . .” (for example “between amino acids 198-218”) is referring to the amino acid number counting consecutively from the N-terminus of the amino acid sequence, for example “between amino acids 198 to 218 . . . of SEQ ID NO: 1” refers to position in the amino acid sequence between amino acid 198 and amino acid 218 of SEQ ID NO: 1 including both amino acids 198 and 218. Thus, in an embodiment where “a consensus sequence selected from D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) (e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4)) has been added next to or substituted for one or more amino acids between amino acid residues 198-218”, the consensus sequence may have been added next to or substituted for any one (or more) of amino acid numbers 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218 in SEQ ID NO: 1. A person skilled in the art will understand that when the EPA amino acid sequence is a variant and/or fragment of an amino acid sequence of SEQ ID NO: 1, such as an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, the reference to “between amino acids . . .” refers to a the position that would be equivalent to the defined position, if this sequence was lined up with an amino acid sequence of SEQ ID NO: 1 in order to maximise the sequence identity between the two sequences (Sequence alignment tools are not limited to Clustal Omega (www.ebi.ac.uk/ClustalOmega), MUSCLE (www.ebi.ac.uk/MUSCLE), or T-coffee (www.tcoffee.org). In one aspect, the sequence alignment tool used is Clustal Omega (www.ebi.ac.uk/ClustalOmega).

[0106] The amino acid numbers referred to herein correspond to the amino acids in SEQ ID NO: 1 and as described above, a person skilled in the art can determine equivalent amino acid positions in an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1 by alignment. The addition or deletion of amino acids from the variant and/or fragment of SEQ ID NO: 1 could lead to a difference in the actual amino acid position of the consensus sequence in the mutated sequence, however, by lining the mutated sequence up with the reference sequence, the amino acid in an equivalent position to the corresponding amino acid in the reference sequence

can be identified and hence the appropriate position for addition or substitution of the consensus sequence can be established.

[0107] The modified EPA protein of the invention may be an isolated modified EPA protein. The modified EPA protein of the invention may be a recombinant modified EPA protein. The modified EPA protein of the invention may be an isolated recombinant modified EPA protein.

Consensus Sequence

[0108] The modified EPA protein of the invention comprises a D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) or J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) consensus sequence, wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues. The classical 5 amino acid glycosylation consensus sequence (D/E-X-N-Z-S/T (SEQ ID NO: 2)) may be extended by 1-5 other amino acid residues either side of the consensus sequence for more efficient glycosylation J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) (e.g. G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G (SEQ ID NO: 25)). The classical 5 amino acid glycosylation consensus sequence (D/E-X-N-Z-S/T (SEQ ID NO: 2)) may be extended by lysine residues for more efficient glycosylation (e.g. K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3)). Thus consensus sequences in the modified EPA protein of the invention may comprises (or consist) of a D/E-X-N-Z-S/T (SEQ ID NO: 2) consensus sequence.

[0109] In the modified EPA protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) or J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) wherein X is Q (glutamine) and Z is A (alanine). In the modified EPA protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X is Q (glutamine) and Z is A (alanine). In an embodiment, the consensus sequence is D/E-X-N-Z-S/T (SEQ ID NO: 2), wherein X is Q (glutamine) and Z is A (alanine), e.g. D-Q-N-A-T (SEQ ID NO: 25) also referred to as “DQNAT”. In an embodiment, the consensus sequence is K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X is Q (glutamine) and Z is A (alanine), e.g. K-D-Q-N-A-T-K (SEQ ID NO: 4) also referred to as “KDQNATK”. In the modified EPA protein of the invention, the consensus sequence(s) may be selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) or J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) wherein X is Q (glutamine), Z is A (alanine), J and U are independently 1 to 5 amino acid residues independently selected from glycine and/or serine.

[0110] In an embodiment, the modified EPA protein of the invention comprises at least two D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention comprises at least three D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention comprises at least four D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention comprises at least five D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention comprises at least

six D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention comprises at least seven D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention contains three to seven D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention contains four to seven D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences. In an embodiment, the modified EPA protein of the invention contains five to seven D/E-X-N-Z-S/T (SEQ ID NO: 2) or K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) consensus sequences.

[0111] Introduction of such glycosylation sites can be accomplished by, e.g. adding new amino acids to the primary structure of the protein (i.e. the glycosylation sites are added, in full or in part), or by mutating existing amino acids in the protein in order to generate the glycosylation sites (i.e. amino acids are not added to the protein, but selected amino acids of the protein are mutated so as to form glycosylation sites). In an embodiment, the consensus sequence(s) are recombinantly introduced into the EPA amino acid sequence of SEQ ID NO: 1 or a EPA amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0112] The modified EPA protein of the invention may further comprise a “peptide tag” or “tag”, i.e. a sequence of amino acids that allows for the isolation and/or identification of the modified EPA protein. For example, adding a tag to a modified EPA protein of the invention can be useful in the purification of that protein and, hence, the purification of conjugate (e.g. bioconjugate) vaccines comprising the tagged modified EPA protein. Exemplary tags that can be used herein include, without limitation, histidine (HIS) tags (e.g. hexa histidine-tag, or 6XHis-Tag), FLAG-TAG, and HA tags. In one embodiment, the tag is a hexa-histidine tag. The tags used herein are removable, e.g. removal by chemical agents or by enzymatic means, once they are no longer needed, e.g. after the protein has been purified. Thus, the modified EPA protein of the invention may further comprise a peptide tag. Optionally the peptide tag is located at the C-terminus of the amino acid sequence. Optionally the peptide tag comprises six histidine residues at the C-terminus of the amino acid sequence. In one aspect, the modified EPA protein of the invention comprises (or consists of) an amino acid sequence which is at least 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 100% identical to any one of the sequences of SEQ ID NOs: 6 to 14 and a peptide tag (e.g. six histidine residues at the C-terminus of the amino acid sequence).

[0113] In an embodiment, the modified EPA protein of the invention comprises a signal sequence which is capable of directing the EPA protein to the periplasm of a host cell (e.g. bacterium). Signal sequences, including periplasmic signal sequences, are usually removed during translocation of the protein into, for example, the periplasm by signal peptidases (i.e. a mature protein is a protein from which at least the signal sequence has been removed). The signal sequence may be from *E. coli* flagellin (FlgI) [MKFLSALILLVT-TAAQA (SEQ ID NO: 15)], *E. coli* outer membrane porin A (OmpA) [MKKTAIAVALAGFATVAQA (SEQ ID NO: 16)], *E. coli* maltose binding protein (MalE) [MKIKTGA-

RILALSALTMMFSASALA (SEQ ID NO: 17)], *Erwin carotovorus* pectate lyase (PelB) [MKYLLPTAAAGLLL-LAAQPAMA (SEQ ID NO: 18)], heat labile *E. coli* enterotoxin LTIIb [MSFKKIIKAFVIMAALVSVQAH (SEQ ID NO: 19)], *Bacillus subtilis* endoxylanase XynA [MFKFKKKFLVGLTAAAFMSISMFSATASA (SEQ ID NO: 20)], *E. coli* DsbA [MKKIWLALAGLVLAFSASA (SEQ ID NO: 21)], ToIB [MKQALRVAFGFLILWASVLHA (SEQ ID NO: 22)] or SipA [MKMNKKVLLTST-MAASLLSVASVQAS (SEQ ID NO: 23)]. In a specific embodiment, the signal sequence is from *E. coli* DsbA [MKKIWLALAGLVLAFSASA (SEQ ID NO: 21)].

[0114] Thus, the present invention provides a modified EPA protein, wherein the amino acid sequence further comprises a signal sequence which is capable of directing the EPA protein to the periplasm of a host cell (e.g. bacterium), optionally said signal sequence being DsbA (SEQ ID NO: 21). A signal peptide of the protein DsbA from *E. coli* can be genetically fused to the N-terminus of the mature EPA sequence. For example, a plasmid derived from pEC415 [Schulz, H., Hennecke, H., and Thony-Meyer, L., Science, 281, 1197-1200, 1998] containing the DsbA signal peptide code followed by a RNase sequence can be digested (NdeI to EcoRI) to keep the DsbA signal and remove the RNase insert. EPA is then amplified using PCR (forward oligo 5'-AAGCTAGCGCCGCCGAGGAAGCCTTCGACC (SEQ. ID NO. 32) and reverse oligo 5'-AAGAA TTCTCAGTGGTGGTGGTGGTGCTTCAGGTCCT CGCGCGGCGG (SEQ. ID NO. 33)) and digested NheI/EcoRI and ligated to replace the RNase sequence removed previously. The resulting construct (pGVXN69) encodes a protein product with an DsbA signal peptide, the mature EPA sequence and a hexa-histag.

[0115] A further aspect of the invention is a polynucleotide encoding a modified EPA protein of the invention. For example, a polynucleotide encoding a modified EPA protein, having a nucleotide sequence that encodes a polypeptide with an amino acid sequence that is at least 97%, 98%, 99% or 100% identical to any one of SEQ ID NOS: 6 to 14. For example, a nucleotide sequence according to SEQ ID NO: 40 or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 40. For example, a nucleotide sequence according to SEQ ID NO: 41 or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 41. For example, a nucleotide sequence according to SEQ ID NO: 42 or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 42.

[0116] For example, a nucleotide sequence according to SEQ ID NO: 43 or a nucleotide sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 43. The nucleotide sequence comprises nucleotides encoding for amino acids corresponding to one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3). For example, encoding for a modified EPA protein having a consensus sequence substituted at one or more of positions Y208, R274, S318 and/or A519 in an amino acid sequence of SEQ ID NO: 1 or at an equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0117] A vector comprising such a polynucleotide is a further aspect of the invention.

Conjugates

[0118] The present invention also provides a conjugate (e.g. bioconjugate) comprising (or consisting of) a modified EPA protein of the invention linked to an antigen (e.g. a saccharide antigen, optionally a bacterial polysaccharide antigen). The antigen may be a bacterial polysaccharide antigen, or a yeast polysaccharide antigen, or a mammalian polysaccharide antigen.

[0119] In an embodiment, the conjugate comprises a conjugate (e.g. bioconjugate) comprising (or consisting of) a modified EPA protein of the invention covalently linked to an antigen (e.g. a saccharide antigen, optionally a bacterial polysaccharide antigen), wherein the antigen is linked (either directly or through a linker). In an embodiment, the antigen is directly linked to the modified EPA protein of the invention. In an embodiment, the antigen is directly linked to an amino acid residue of the modified EPA protein.

[0120] In an embodiment, the modified EPA protein is covalently linked to the antigen through a chemical linkage obtainable using a chemical conjugation method (i.e. the conjugate is produced by chemical conjugation). The chemical conjugation method may be selected from the group consisting of carbodiimide chemistry, reductive amination, cyanation chemistry (for example CDAP chemistry), maleimide chemistry, hydrazide chemistry, ester chemistry, and N-hydroxysuccinimide chemistry. Conjugates can be prepared by direct reductive amination methods as described in, US200710184072 (Hausdorff) U.S. Pat. No. 4,365,170 (Jennings) and U.S. Pat. No. 4,673,574 (Anderson). Other methods are described in EP-0-161-188, EP-208375 and EP-0-477508. The conjugation method may alternatively rely on activation of the saccharide with 1-cyano-4-dimethylamino pyridinium tetrafluoroborate (CDAP) to form a cyanate ester. Such conjugates are described in PCT published application WO 93/15760 Uniformed Services University and WO 95/08348 and WO 96/29094. See also Chu C. et al. Infect. Immunity, 1983 245 256.

[0121] In general the following types of chemical groups on a modified EPA protein can be used for coupling/conjugation:

[0122] A) Carboxyl (for instance via aspartic acid or glutamic acid). In one embodiment this group is linked to amino groups on saccharides directly or to an amino group on a linker with carbodiimide chemistry e.g. with EDAC.

[0123] B) Amino group (for instance via lysine). In one embodiment this group is linked to carboxyl groups on saccharides directly or to a carboxyl group on a linker with carbodiimide chemistry e.g. with EDAC. In another embodiment this group is linked to hydroxyl groups activated with CDAP or CNBr on saccharides directly or to such groups on a linker; to saccharides or linkers having an aldehyde group; to saccharides or linkers having a succinimide ester group.

[0124] C) Sulphydryl (for instance via cysteine). In one embodiment this group is linked to a bromo or chloro acetylated saccharide or linker with maleimide chemistry. In one embodiment this group is activated/modified with bis diazobenzidine.

[0125] D) Hydroxyl group (for instance via tyrosine). In one embodiment this group is activated/modified with bis diazobenzidine.

[0126] E) Imidazolyl group (for instance via histidine). In one embodiment this group is activated/modified with bis diazobenzidine.

[0127] F) Guanidyl group (for instance via arginine).

[0128] G) Indolyl group (for instance via tryptophan).

[0129] On a saccharide, in general the following groups can be used for a coupling: OH, COOH or NH₂. Aldehyde groups can be generated after different treatments such as: periodate, acid hydrolysis, hydrogen peroxide, etc.

[0130] Conjugates can be purified by any method known in the art for purification of a protein, for example, by chromatography (e.g. ion exchange, anionic exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. See, e.g., Saraswat et al., 2013, Biomed. Res. Int. ID0312709 (p. 1-18); see also the methods described in WO 2009/104074. The actual conditions used to purify a particular conjugate will depend, in part, on the synthesis strategy (e.g., synthetic production vs. recombinant production) and on factors such as net charge, hydrophobicity, and/or hydrophilicity of the bioconjugate.

[0131] In an embodiment, the amino acid residue on the modified EPA protein to which the antigen is linked is selected from the group consisting of: Ala, Arg, Asp, Cys, Gly, Glu, Gln, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val. Optionally, the amino acid is: an amino acid containing a terminal amine group, a lysine, an arginine, a glutaminic acid, an aspartic acid, a cysteine, a tyrosine, a histidine or a tryptophan. In an embodiment, the amino acid residue on the modified EPA protein to which the antigen is linked is not an asparagine residue and in this case, the conjugate is typically produced by chemical conjugation. Alternatively, the antigen is linked to an amino acid on the modified EPA protein selected from asparagine, aspartic acid, glutamic acid, lysine, cysteine, tyrosine, histidine, arginine or tryptophan (e.g. asparagine) and in the case of asparagine, the conjugate may be a bioconjugate (for example an enzymatic conjugation using a oligosaccharyl-transferase such as PglB). In an embodiment, the amino acid residue on the modified EPA protein to which the antigen is linked is an asparagine residue. Preferably, the amino acid residue on the modified EPA protein to which the antigen is linked is part of the consensus sequence, e.g. the asparagine in D/E-X-N-Z-S/T (SEQ ID NO: 2), K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) or J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5) consensus sequence.

[0132] The conjugate of the invention may be a conjugate of a recombinant modified EPA protein (e.g. chemical conjugate or bioconjugate). The conjugate of the invention may be a conjugate of an isolated recombinant modified EPA protein and a recombinant antigen, e.g. recombinant saccharide (i.e. bioconjugate).

Antigens

[0133] The antigen may be a saccharide antigen, e.g. a bacterial polysaccharide, for example an O-antigen or a capsular polysaccharide, a yeast polysaccharide or a mammalian polysaccharide. Polysaccharides comprise 2 or more monosaccharides, typically greater than 10 monosaccharides. In an embodiment, the antigen in a conjugate (e.g. bioconjugate) of the invention is a bacterial polysaccharide selected from a *Shigella* species, *Pseudomonas* species, *Klebsiella* species, *Streptococcus* species, or *Staphylococcus* species (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, or *Staphylococcus aureus*). In an embodiment, the antigen is a bacterial polysaccharide

antigen (e.g. an O-antigen from a Gram negative bacterium, optionally from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium, optionally *Streptococcus pneumoniae* or *Staphylococcus aureus*). In an embodiment, the antigen is an O-antigen from a Gram negative bacterium. In an embodiment, the antigen in a conjugate (e.g. bioconjugate) of the invention is a bacterial polysaccharide selected from from a *Shigella* species, *Klebsiella* species, or *Streptococcus* species (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Klebsiella pneumoniae* or *Streptococcus pneumoniae*). In an embodiment, the antigen in a conjugate (e.g. bioconjugate) of the invention is a bacterial polysaccharide selected from *Shigella flexneri*, *Klebsiella pneumoniae* and *Streptococcus pneumoniae*). In an embodiment, the antigen is a bacterial polysaccharide from *Klebsiella pneumoniae*. Thus, the present invention provides a conjugate (e.g. bioconjugate) comprising a modified EPA protein of the invention linked to an antigen wherein the antigen is a saccharide, optionally a bacterial polysaccharide (e.g. from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae* or *Staphylococcus aureus*). In an embodiment, the antigen is an O-antigen. In another embodiment, the antigen is a capsular polysaccharide.

[0134] In certain embodiments, the antigen is an O-antigen e.g. from a Gram-negative bacterium. In certain embodiments, the antigen is an O-antigen from *Salmonella* species, *Shigella* species, *Pseudomonas* species or *Klebsiella* species. In certain embodiments, the antigen is an O-antigen from *Shigella* species, *Pseudomonas* species or *Klebsiella* species (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, or *Klebsiella pneumoniae*). In an embodiment, the antigen is an O-antigen from *Shigella dysenteriae*, *Shigella flexneri* or *Shigella sonnei*. For example, the antigen may be an O-antigen from *S. dysenteriae* type 1, *S. sonnei*, and *S. flexneri* type 6, and *S. flexneri* 2a and 3a 0 (Dmitriev, B.A., et al. Somatic Antigens of *Shigella* Eur J. Biochem, 1979. 98: p. 8; Liu et al Structure and genetics of *Shigella O* antigens FEMS Microbiology Review, 2008. 32: p. 27).

[0135] In an embodiment, the antigen is an O-antigen from *Pseudomonas aeruginosa*. For example, the antigen may be an O-antigen from *Pseudomonas aeruginosa* serotypes 1-20 (Raymond et al., J Bacteriol. 2002 184(13):3614-22). In an embodiment, the antigen is an O-antigen from *Klebsiella pneumoniae*.

[0136] In certain embodiments, the antigen is a capsular polysaccharide from *Neisseria meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In certain embodiments, the antigen is a capsular polysaccharide from *Streptococcus* species or *Staphylococcus* species. (e.g. *Streptococcus pneumoniae* or *Staphylococcus aureus*). In an embodiment, the antigen is a capsular polysaccharide from *Staphylococcus aureus*. For example, the antigen may be a capsular polysaccharide from *Staphylococcus aureus* type 5 and 8. In an embodiment, the antigen is a capsular polysaccharide from *Streptococcus pneumoniae*.

Host cell

[0137] The present invention also provides a host cell comprising:

[0138] i) one or more nucleotide sequences comprising polysaccharide synthesis genes, optionally for producing a bacterial polysaccharide antigen (e.g. an O-antigen from a Gram negative bacterium optionally from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium optionally from *Streptococcus pneumoniae* or *Staphylococcus aureus*) or a yeast polysaccharide antigen or a mammalian polysaccharide antigen, optionally integrated into the host cell genome;

[0139] ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid;

[0140] iii) a nucleotide sequence that encodes a modified EPA protein of the invention, optionally within a plasmid.

[0141] Disclosures of methods for making such host cells which are capable of producing bioconjugates are found in WO 06/119987, WO 09/104074, WO 11/62615, WO 11/138361, WO 14/57109, WO14/72405 and WO16/20499.

[0142] Host cells that can be used to produce the bioconjugates of the invention, include archaea, prokaryotic host cells, and eukaryotic host cells. In certain embodiments, the host cell is a non-human host cell. Exemplary prokaryotic host cells for use in production of the bioconjugates of the invention include *Escherichia* species, *Shigella* species, *Klebsiella* species, *Xanthomonas* species, *Salmonella* species, *Yersinia* species, *Lactococcus* species, *Lactobacillus* species, *Pseudomonas* species, *Corynebacterium* species, *Streptomyces* species, *Streptococcus* species, *Staphylococcus* species, *Bacillus* species, and *Clostridium* species. Preferably, the host cell is *E. coli* (e.g. *E. coli* K12 W3110).

[0143] Host cells may be modified to delete or modify genes in the host cell genetic background (genome) that compete or interfere with the synthesis of the polysaccharide of interest (e.g. compete or interfere with one or more heterologous polysaccharide synthesis genes that are recombinantly introduced into the host cell). These genes can be deleted or modified in the host cell background (genome) in a manner that makes them inactive/dysfunctional (i.e. the host cell nucleotide sequences that are deleted/modified do not encode a functional protein or do not encode a protein whatsoever). In an embodiment, when nucleotide sequences are deleted from the genome of the host cells of the invention, they are replaced by a desirable sequence, e.g. a sequence that is useful for glycoprotein production. Exemplary genes that can be deleted in host cells (and, in some cases, replaced with other desired nucleotide sequences) include genes of host cells involved in glycolipid biosynthesis, such as *waaL* (see, e.g. Feldman et al. 2005, PNAS USA 102:3016-3021), the O antigen cluster (*rfb* or *wb*), enterobacterial common antigen cluster (*wec*), the lipid A core biosynthesis cluster (*waa*), galactose cluster (*gal*), arabinose cluster (*ara*), colonic acid cluster (*wc*), capsular polysaccharide cluster, undecaprenol-pyrophosphate biosynthesis genes (e.g. *uppS* (Undecaprenyl pyrophosphate synthase), *uppP* (Undecaprenyl diphosphatase)), Und-P recycling genes, metabolic enzymes involved in nucleotide activated sugar biosynthesis, enterobacterial common antigen cluster, and prophage O antigen modification clusters like the *gtrABS* cluster. In an embodiment, one or more of the *waaL* gene, *gtrA* gene, *gtrB* gene, *gtrS* gene, or a gene or genes from the *wec* cluster or a gene, or a gene or genes from

the colonic acid cluster (*wc*), or a gene or genes from the *rfb* gene cluster are deleted or functionally inactivated from the genome of a prokaryotic host cell of the invention. In another embodiment, one or more of the *waaL* gene, *gtrA* gene, *gtrB* gene, *gtrS* gene, or a gene or genes from the *wec* cluster or a gene or genes from the *rfb* gene cluster are deleted or functionally inactivated from the genome of a prokaryotic host cell of the invention. In a specific embodiment the host cell of the invention is *E. coli*, wherein the native enterobacterial common antigen cluster (*ECA*, *wec*) with the exception of *wecA*, the colanic acid cluster (*wca*), and the O16-antigen cluster have been deleted. In addition, the native lipopolysaccharide O-antigen ligase *waaL* may be deleted from the host cell of the invention. In addition, the native *gtrA* gene, *gtrB* gene and *gtrS* gene, may be deleted from the host cell of the invention.

[0144] The host cells of the present invention are engineered to comprise heterologous nucleotide sequences. The host cells of the present invention are engineered to comprise a nucleotide sequence that encodes a modified EPA protein of the invention, optionally within a plasmid. The host cells of the invention also comprise one or more nucleotide sequences comprising polysaccharide synthesis genes. Thus, host cells of the invention can produce a bioconjugate comprising an antigen, for example a saccharide antigen (e.g. a bacterial, yeast or mammalian polysaccharide antigen) which is attached to a modified EPA protein of the invention. One or more heterologous nucleotide sequences may encode for the polysaccharide synthesis proteins to produce the bacterial polysaccharide antigen, yeast polysaccharide antigen or mammalian polysaccharide antigen. Thus the present invention also provides a host cell comprising:

[0145] i) one or more heterologous nucleotide sequences comprising polysaccharide synthesis genes for producing a bacterial polysaccharide antigen (e.g. an O-antigen from a Gram negative bacterium optionally from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium optionally from *Streptococcus pneumoniae* or *Staphylococcus aureus*) or a yeast polysaccharide antigen or a mammalian polysaccharide antigen, optionally integrated into the host cell genome;

[0146] ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid;

[0147] iii) a nucleotide sequence that encodes a modified EPA protein of the invention, optionally within a plasmid.

[0148] The host cells of the invention may comprise one or more nucleotide sequences sufficient for producing a saccharide antigen (e.g. a bacterial polysaccharide antigen), in particular for producing a saccharide antigen (e.g. a bacterial polysaccharide antigen) that is heterologous to the host cell. For example, where the host cell is *E. coli* the host cell may comprise one more nucleotide sequences comprising polysaccharide synthesis genes sufficient for producing a bacterial polysaccharide antigen of a bacteria which is not an *E. coli* polysaccharide antigen. The bacterial polysaccharide antigen may be an O-antigen or a capsular polysaccharide antigen. Thus the present invention also provides a host cell comprising:

[0149] i) one or more nucleotide sequences comprising polysaccharide synthesis genes, for producing a bacterial polysaccharide antigen (e.g. an O-antigen from a Gram negative bacterium, optionally from *Shigella dysenteriae*,

Shigella flexneri, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium optionally from *Streptococcus pneumoniae* or *Staphylococcus aureus*), optionally integrated into the host cell genome;

[0150] ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid;

[0151] iii) a nucleotide sequence that encodes a modified EPA protein of the invention, optionally within a plasmid.

[0152] Polysaccharide synthesis genes encode proteins involved in synthesis of a polysaccharide (polysaccharide synthesis proteins). In an embodiment, the host cells may comprise one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from a Gram negative bacterium selected from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium selected from *Streptococcus pneumoniae* and *Staphylococcus aureus*. In another embodiment, the host cells may comprise one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from a Gram negative bacterium selected from *Shigella flexneri* and *Klebsiella pneumoniae*, or a capsular polysaccharide from a Gram positive bacterium selected from *Streptococcus pneumoniae* and *Staphylococcus aureus*.

Host Cells for Production of a Bacterial Polysaccharide Antigen

[0153] The host cells of the invention may comprise one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen. In certain embodiments, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Salmonella* species, *Shigella* species, *Pseudomonas* species or *Klebsiella* species. In certain embodiments, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Shigella* species, *Pseudomonas* species or *Klebsiella* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei* or *Klebsiella pneumoniae*). In certain embodiments, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Shigella* species or *Klebsiella* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei* or *Klebsiella pneumoniae*). In an embodiment, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Shigella dysenteriae*, *Shigella flexneri* or *Shigella sonnei*. For example, the host cell may comprise one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *S. dysenteriae* type 1, *S. sonnei*, and *S. flexneri* type 6, and *S. flexneri* 2a and 3a 0 (Dmitriev, B. A., et al Somatic Antigens of *Shigella* Eur J. Biochem, 1979. 98: p. 8; Liu et al Structure and genetics of *Shigella* O antigens FEMS Microbiology Review, 2008. 32: p. 27). In an embodiment, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Pseudomonas aeruginosa* e.g. *Pseudomonas aeruginosa* serotypes 1-20. In an embodiment, the host cell

comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen from *Klebsiella pneumoniae*.

[0154] The host cells of the invention may comprise one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide. In certain embodiments, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide from *N. meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In certain embodiments, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide from *Streptococcus* species, or *Staphylococcus* species. (e.g. *Streptococcus pneumoniae* or *Staphylococcus aureus*). In an embodiment, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide from *Staphylococcus aureus*, e.g. from *Staphylococcus aureus* type 5 and 8. In an embodiment, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide from *Streptococcus pneumoniae*.

Host Cells Comprising Heterologous Nucleotide Sequences for Producing a Bacterial Polysaccharide Antigen

[0155] The host cells of the present invention may naturally express one or more nucleotide sequences comprising polysaccharide synthesis genes for production of a saccharide antigen (e.g. a bacterial polysaccharide antigen), or the host cells may be engineered to express one or more such nucleotide sequences. For example, host cells of the present invention may utilize endogenous or heterologous glycosyltransferases for sequential assembly of oligosaccharides in the cytosol (cytosolic glycosyltransferases). Heterologous nucleotide sequences (e.g. nucleotide sequences that encode carrier proteins and/or nucleotide sequences that encode other proteins, e.g. proteins involved in glycosylation) can be introduced into the host cells of the invention using methods such as electroporation, chemical transformation by heat shock, natural transformation, phage transduction, and conjugation. In specific embodiments, heterologous nucleotide sequences are introduced into the host cells of the invention using a plasmid, e.g. the heterologous nucleotide sequences are expressed in the host cells by a plasmid (e.g. an expression vector). In another specific embodiment, heterologous nucleotide sequences are introduced into the host cells of the invention using the method of insertion described in WO14/037585. In certain embodiments, the host cell of the present invention comprises one or more nucleotide sequences comprising polysaccharide synthesis genes which are heterologous to the host cell. In certain embodiments, one or more of said nucleotide sequences comprising polysaccharide synthesis genes which are heterologous to the host cell are integrated into the genome of the host cell. The heterologous nucleotide sequences may encode, without limitation, glycosyltransferases, oligosaccharyl transferases, epimerases, flippases, and/or polymerases. In certain embodiments, the host cells of the invention comprise one or more heterologous nucleotide sequences encoding glycosyltransferase(s). Said glycosyl-

transferase(s) can be derived from, e.g. *Escherichia* species, *Shigella* species, *Klebsiella* species, *Salmonella* species, *Pseudomonas* species, *Streptococcus* species, or *Staphylococcus* species.

[0156] The host cells of the invention may comprise one or more heterologous nucleotide sequences comprising polysaccharide synthesis genes for producing an O-antigen. In certain embodiments, the host cell comprises one or more nucleotide sequences from *Salmonella* species, *Shigella* species, *Pseudomonas* species or *Klebsiella* species that encode polysaccharide synthesis proteins for producing an O-antigen. In certain embodiments, the host cell comprises one or more nucleotide sequences from *Shigella* species, *Pseudomonas* species or *Klebsiella* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, or *Klebsiella pneumoniae*) that encode polysaccharide synthesis proteins for producing an O-antigen. In certain embodiments, the host cell comprises one or more nucleotide sequences from *Shigella* species or *Klebsiella* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, or *Klebsiella pneumoniae*) that encode polysaccharide synthesis proteins for producing an O-antigen. In an embodiment, the host cell comprises one or more nucleotide sequences from *Shigella dysenteriae*, *Shigella flexneri* or *Shigella sonnei*. that encode polysaccharide synthesis proteins for producing an O-antigen. For example, the host cell may comprise one or more nucleotide sequences from *S. dysenteriae* type 1, *S. sonnei*, and *S. flexneri* type 6, and *S. flexneri* 2a and 3a that encode polysaccharide synthesis proteins for producing an O-antigen. In an embodiment, the host cell comprises one or more nucleotide sequences from *Pseudomonas aeruginosa*, e.g. *Pseudomonas aeruginosa* serotypes 1-20, that encode polysaccharide synthesis proteins for producing an O-antigen. In an embodiment, the host cell comprises one or more nucleotide sequences from *Klebsiella pneumoniae* that encode polysaccharide synthesis proteins for producing an O-antigen. The nucleotide sequences that encode an O-antigen may be an rfb cluster. As used herein, rfb cluster refer to a gene cluster that encodes enzymatic machinery capable of synthesis of an O antigen. The host cells may comprise an rfb gene cluster from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*.

[0157] The host cells of the invention may comprise one or more heterologous nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular saccharide. In certain embodiments, the host cell comprises one or more nucleotide sequences from *N. meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus* that encode polysaccharide synthesis proteins for producing a capsular saccharide. In certain embodiments, the host cell comprises one or more nucleotide sequences from *Streptococcus* species, or *Staphylococcus* species. (e.g. *Streptococcus pneumoniae* or *Staphylococcus aureus*) that encode polysaccharide synthesis proteins for producing a capsular polysaccharide. In an embodiment, the host cell comprises one or more nucleotide sequences from *Staphylococcus aureus*, e.g. from *Staphylococcus aureus* type 5 and 8, that encode polysaccharide synthesis proteins for producing a capsular polysaccharide.

In an embodiment, the host cell comprises one or more nucleotide sequences comprising polysaccharide synthesis genes for producing a capsular polysaccharide from *Streptococcus pneumoniae*. The nucleotide sequences may be a capsular polysaccharide gene cluster. The host cells may comprise a capsular polysaccharide gene cluster from a *Streptococcus* strain (e.g. *S. pneumoniae*, *S. pyrogens*, *S. agalactiae*), a *Staphylococcus* strain (e.g. *S. aureus*). The capsular polysaccharide gene cluster for *Streptococcus pneumoniae* maps between dexB and aliA in the pneumococcal chromosome (Lhull et al., 1999, J. Exp. Med. 190, 241-251). There are typically four relatively conserved genes: (wzg), (wzh), (wza), (wze) at the 5' end of the capsular polysaccharide gene cluster (Jiang et al., 2001, Infect. Immun. 69, 1244-1255). Also included in the capsular polysaccharide gene cluster of *S. pneumoniae* are wzx (polysaccharide flippase gene) and wzy (polysaccharide polymerase gene). The CP gene clusters of all 90 *S. pneumoniae* serotypes have been sequenced by Sanger Institute (http://www.sanger.ac.uk/Projects/S_pneumoniae/CPS/), and wzx and wzy of 89 serotypes have been annotated and analyzed (Kong et al., 2005, J. Med. Microbiol. 54, 351-356). The capsular biosynthetic genes of *S. pneumoniae* are further described in Bentley et al. (PLoS Genet. 2006 March; 2(3): e31 and the sequences are provided in GenBank. Thus, in an embodiment the host cells of the invention may further comprise a nucleotide sequence encoding a polymerase (e.g. wzy), a flippase (e.g. wzx) and optionally a nucleotide sequence encoding and/or a chain length regulator (e.g. wzz).

[0158] In certain embodiments, the host cells may also comprise heterologous nucleotide sequences that are located outside of an rfb cluster or a capsular polysaccharide cluster. For example, nucleotide sequences encoding glycosyltransferases and acetyltransferases that are found outside of rfb clusters or capsular polysaccharide clusters and that modify recombinant polysaccharides can be introduced into the host cells.

Oligosaccharyl Transferase

[0159] N-linked protein glycosylation (the addition of carbohydrate molecules to an asparagine residue in the polypeptide chain of the target protein) is the most common type of post-translational modification occurring in the endoplasmic reticulum of eukaryotic organisms. The process is accomplished by the enzymatic oligosaccharyltransferase complex (OST) responsible for the transfer of a preassembled oligosaccharide from a lipid carrier (dolichol phosphate) to an asparagine residue of a nascent protein within the conserved sequence Asn-X-Ser/Thr (where X is any amino acid except proline) in the Endoplasmic reticulum.

[0160] It has been shown that a bacterium, the food-borne pathogen *Campylobacter jejuni*, can also N-glycosylate its proteins (Wacker et al. Science. 2002; 298(5599):1790-3) due to the fact that it possesses its own glycosylation machinery. The machinery responsible of this reaction is encoded by a cluster called "pgl" (for protein glycosylation). The *C. jejuni* glycosylation machinery can be transferred to *E. coli* to allow for the glycosylation of recombinant proteins expressed by the *E. coli* cells. Previous studies have demonstrated how to generate *E. coli* strains that can perform N-glycosylation (see, e.g. Wacker et al. Science. 2002; 298 (5599):1790-3; Nita-Lazar et al. Glycobiology. 2005; 15(4):

361-7; Feldman et al. Proc Natl Acad Sci U S A. 2005; 102(8):3016-21; Kowarik et al. EMBO J. 2006; 25(9):1957-66; Wacker et al. Proc Natl Acad Sci USA. 2006; 103(18):7088-93; International Patent Application Publication Nos. WO2003/074687, WO2006/119987, WO 2009/104074, and WO/2011/06261, and WO2011/138361).

[0161] The host cells of the present invention comprise a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid. In a specific embodiment, the oligosaccharyl transferase is an oligosaccharyl transferase from *Campylobacter*. In another specific embodiment, the oligosaccharyl transferase is a pglB, optionally from *Campylobacter jejuni* (i.e. pglB; see, e.g. Wacker et al. 2002, Science 298:1790-1793; see also, e.g. NCBI Gene ID: 3231775, UniProt Accession No. 086154) SEQ ID NO: 24:

```
MLKKEYLKNPYLVLFAMII LAYVSVFCRFYVWVWASEFNEYFFNNQLM
IISNDGYAFAEGARDMIAGHPQNDLSYGGSSLSALTYWLYKITPFSFE
SIILYMSTFLSSLVVIPTILLANEYKRPLMGFVAALLASIANSYNRTM
SGYYDTMDLVIVLPMFILFPMVRMILKKDFFSLIALPLFIGIYLWWYPS
SYTLNVALIGLFLIYTLIFHRKEKIFYIAVILSSLTSLNIAWFYQSAII
VILFALFALEQKRLNFMIIIGILGSATLIFLILSGGVDPILYQLKFYIFR
SDESANLTQGFMYFNVNQTIQEVENVDLSEPMRRISGSEIVFLFSLFGF
VWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMAFGFLLSEF
KAIMVKKYSQLTSNVCIVFATILTLAPVFIHIYNYKAPT VFSQNEASLL
NQLKNIANREDYVVTWWDYGYPVRYSDVKTLVDGGKHLGKDNFFPSFA
LSKDEQAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNVD
LHLASLSKPDFKIDIPKIRDIYLYMPARMSLIFSIVASFHINLDIGVL
DKPEIHSIAYPLDVKNGEIYLSNGVVLSDDFRSFKIGDNWVSVNSIVEI
NSIKQGEYKITPIDDKAQFYIFYLKDSAIPIYAQFILMDKTMFNSAYVQM
FPLGNYDKNLFDLVINSRDAKVFKLKI
```

[0162] Thus host cells of the present invention may comprise a nucleotide sequence encoding pglB, optionally pglB from *Campylobacter jejuni*, optionally a nucleotide sequence encoding pglB from *Campylobacter jejuni* having a sequence at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 24, optionally within a plasmid.

Polymerase

[0163] Host cells of the present invention may also comprise a nucleotide sequence that encodes a polymerase (e.g. wzy). In an embodiment, the polymerase (e.g. wzy) is introduced into a host cell of the invention (i.e. the polymerase is heterologous to the host cell). In an embodiment, the polymerase is a bacterial polymerase. In an embodiment, the polymerase is a capsular polysaccharide polymerase (e.g. wzy) or an O antigen polymerase (e.g. wzy). In an embodiment, the polymerase is an O-antigen polysaccharide polymerase (e.g. wzy), e.g. from *Shigella* species, *Pseudomonas* species or *Escherichia* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, or *E. coli*). In an embodiment, the polymerase is a capsular polysaccharide polymerase (e.g. wzy),

e.g. from *N. meningitidis* serogroup A (MenA), *N. meningitidis* serogroup C (MenC), *N. meningitidis* serogroup Y (MenY), *N. meningitidis* serogroup W (MenW), *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In an embodiment, the polymerase is a capsular polysaccharide polymerase (e.g. wzy) of *Streptococcus pneumoniae*. Said wzy polymerase may be incorporated (e.g. inserted into the genome or expressed by a plasmid) in said host cell as part of a rfb cluster or capsular polysaccharide cluster. Thus, a host cell of the invention may further comprise a nucleotide sequence encoding a heterologous wzy polymerase.

Flippases

[0164] A host cell of the invention may also comprise a nucleotide sequence encoding a flippase (e.g. wzx), e.g. a heterologous flippase. Flippases translocate wild type repeating units and/or their corresponding engineered (hybrid) repeat units from the cytoplasm into the periplasm of host cells (e.g. *E. coli*). In an embodiment, the flippase is a bacterial flippase, e.g. a flippase of the polysaccharide biosynthetic pathway of interest. In a specific embodiment, the host cell of the invention comprises a nucleotide sequence encoding a flippase (e.g. wzx gene) of a polysaccharide biosynthetic pathway of a *Streptococcus* species, *Shigella* species, *Escherichia* species, *Pseudomonas* species, or *Staphylococcus* species. (e.g. *Streptococcus pneumoniae*, *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *E. coli*, *Pseudomonas aeruginosa*, or *Staphylococcus aureus*. In an embodiment, the flippase is a capsular polysaccharide flippase (e.g. wzx) of *Streptococcus pneumoniae*. Other flippases that can be introduced into the host cells of the invention are for example from *Campylobacter jejuni* (e.g. pglK).

Accessory Enzymes

[0165] In an embodiment, nucleotide sequences encoding one or more accessory enzymes are introduced into the host cells of the invention. Thus, a host cell of the invention may further comprise one or more of these accessory enzymes. Such nucleotide sequences encoding one or more accessory enzymes can be either plasmid-borne or integrated into the genome of the host cells of the invention. Exemplary accessory enzymes include, without limitation, epimerases (see e.g. WO2011/062615), branching, modifying (e.g. to add cholins, glycerolphosphates, pyruvates), amidating, chain length regulating, acetylating, formylating, polymerizing enzymes. Thus a host cell of the invention may also comprise a nucleotide sequence encoding a chain length regulator (e.g. wzz), e.g. a heterologous chain length regulator. In an embodiment, the chain length regulator is a capsular polysaccharide chain length regulator (e.g. wzz) of *Streptococcus pneumoniae*.

Bioconjugates

[0166] The present invention provides a bioconjugate comprising a modified EPA protein of the invention linked to an antigen (e.g. a bacterial polysaccharide antigen or a yeast polysaccharide antigen or a mammalian polysaccharide antigen). In a specific embodiment, said antigen is an O-antigen or a capsular polysaccharide. In an embodiment, the antigen is an O-antigen from a Gram negative bacterium. In an embodiment, the present invention provides a biocon-

jugate comprising a modified EPA protein of the invention linked to an antigen wherein the antigen is a saccharide, optionally a bacterial polysaccharide (e.g. from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae* or *Staphylococcus aureus*). In an embodiment, the present invention provides a bioconjugate comprising a modified EPA protein of the invention linked to an antigen wherein the antigen is a bacterial polysaccharide (e.g. from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Klebsiella pneumoniae*, or *Streptococcus pneumoniae*). In another embodiment, the present invention provides a bioconjugate comprising a modified EPA protein of the invention linked to an antigen wherein the antigen is a bacterial polysaccharide from *Shigella flexneri*, *Klebsiella pneumoniae* or *Streptococcus pneumoniae*. The antigen is linked to an amino acid on the modified EPA protein selected from asparagine, aspartic acid, glutamic acid, lysine, cysteine, tyrosine, histidine, arginine or tryptophan (e.g. asparagine). Bioconjugates, as described herein, have advantageous properties over chemical conjugates of antigen-carrier protein, in that they require less chemicals in manufacture and are more consistent in terms of the final product generated.

[0167] A further aspect of the invention is a process for producing a bioconjugate that comprises (or consists of) a modified EPA protein linked to a saccharide, said process comprising (i) culturing the host cell of the invention under conditions suitable for the production of glycoproteins and (ii) isolating the bioconjugate produced by said host cell, optionally isolating the bioconjugate from a periplasmic extract from the host cell. For example, bioconjugates can be made using the shakeflask process, e.g. in a LB shake flask. In aspect of the invention, a fed-batch process for the production of recombinant glycosylated proteins in bacteria can be used to produce bioconjugates of the invention. The aim is to increase glycosylation efficiency and recombinant protein yield per cell and while maintaining simplicity and reproducibility in the process. Bioconjugates of the present invention can be manufactured on a commercial scale by developing an optimized manufacturing method using typical *E. coli* production processes. Various types of feed strategies, such as batch, chemostat and fed-batch can be used.

[0168] The bioconjugates of the invention can be purified for example, by chromatography (e.g. ion exchange, anionic exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of proteins. See, e.g. Saraswat et al. 2013, Biomed. Res. Int. ID #312709 (p. 1-18); see also the methods described in WO 2009/104074. Further, the bioconjugates may be fused to heterologous polypeptide sequences described herein or otherwise known in the art to facilitate purification.

Analytical Methods

[0169] Various methods can be used to analyze the structural compositions and sugar chain lengths of the bioconjugates of the invention and to determine glycosylation site usage.

[0170] Hydrazinolysis can be used to analyze glycans. First, polysaccharides are released from their protein carriers by incubation with hydrazine according to the manufacturer's instructions (Ludger Liberate Hydrazinolysis Glycan Release Kit, Oxfordshire, UK). The nucleophile hydrazine

attacks the glycosidic bond between the polysaccharide and the carrier protein and allows release of the attached glycans. N-acetyl groups are lost during this treatment and have to be reconstituted by re-N-acetylation. The free glycans are purified on carbon columns and subsequently labeled at the reducing end with the fluorophor 2-amino benzamide. See Bigge J C, Patel T P, Bruce J A, Goulding P N, Charles S M, Parekh R B: Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. Anal Biochem 1995, 230(2):229-238. The labeled polysaccharides are separated on a GlycoSep-N column (GL Sciences) according to the HPLC protocol of Royle et al. See Royle L, Mattu T S, Hart E, Langridge J I, Merry A H, Murphy N, Harvey D J, Dwek R A, Rudd P M: An analytical and structural database provides a strategy for sequencing O-glycans from microgram quantities of glycoproteins. Anal Biochem 2002, 304(1):70-90. The resulting fluorescence chromatogram indicates the polysaccharide length and number of repeating units. Structural information can be gathered by collecting individual peaks and subsequently performing MS/MS analysis. Thereby the monosaccharide composition and sequence of the repeating unit can be confirmed and additionally in homogeneity of the polysaccharide composition can be identified. Alternatively, high mass MS and size exclusion HPLC can be applied to measure the size of the complete bioconjugates.

[0171] Yield may be measured as carbohydrate amount derived from a liter of bacterial production culture grown in a bioreactor under controlled and optimized conditions. After purification of bioconjugate, the carbohydrate yields can be directly measured by either the anthrone assay or ELISA using carbohydrate specific antisera. Indirect measurements are possible by using the protein amount (measured by BCA, Lowry, or bardford assays) and the glycan length and structure to calculate a theoretical carbohydrate amount per gram of protein. In addition, yield can also be measured by drying the glycoprotein preparation from a volatile buffer and using a balance to measure the weight.

[0172] Various methods can be used to analyze the conjugates of the invention including, for example, SDS-PAGE or capillary gel electrophoresis. Polymer length is defined by the number of repeat units that are linearly assembled. This means that the typical ladder like pattern is a consequence of different repeat unit numbers that compose the glycan. Thus, two bands next to each other in SDS

[0173] PAGE (or other techniques that separate by size) differ by only a single repeat unit. These discrete differences are exploited when analyzing glycoproteins for glycan size: the unglycosylated carrier protein and the bioconjugate with different polymer chain lengths separate according to their electrophoretic mobilities. The first detectable repeat unit number (n_1) and the average repeat unit number ($n_{average}$) present on a bioconjugate are measured. These parameters can be used to demonstrate batch to batch consistency or polysaccharide stability, for example.

[0174] Glycosylation site usage may be quantified by, for example, glycopeptide LC-MS/MS: conjugates are digested with protease(s), and the peptides are separated by a suitable chromatographic method (C18, Hydrophilic interaction HPLC HILIC, GlycoSepN columns, SE HPLC, AE HPLC), and the different peptides are identified using MS/MS. This method can be used with or without previous sugar chain shortening by chemical (smith degradation) or enzymatic methods. Quantification of glycopeptide peaks using UV

detection at 215 to 280 nm allows relative determination of glycosylation site usage. In another embodiment, site usage may be quantified by size exclusion HPLC: Higher glycosylation site usage is reflected by an earlier elution time from a SE HPLC column. In yet another embodiment, site usage may be quantified by quantitative densitometry of purified bioconjugates stained with Coomassie Brilliant Blue following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Immunogenic Compositions and Vaccines

[0175] The conjugates (e.g. bioconjugate), of the invention are particularly suited for inclusion in immunogenic compositions and vaccines.

[0176] The present invention provides an immunogenic composition comprising a conjugate (e.g. bioconjugate) of the invention, and optionally a pharmaceutically acceptable excipient and/or carrier.

[0177] Immunogenic compositions comprise an immunologically effective amount of the modified EPA protein or conjugate (e.g. bioconjugate) of the invention, as well as any other components. By “immunologically effective amount”, it is meant that the administration of that amount to an individual, either as a single dose or as part of a series is effective for treatment or prevention. This amount varies depending on the health and physical condition of the individual to be treated, age, the degree of protection desired, the formulation of the vaccine and other relevant factors.

[0178] Pharmaceutically acceptable excipients and carriers are described, for example, in Remington's Pharmaceutical Sciences, by E. W. Martin, Mack Publishing Co. Easton, Pa., 5th Edition (1975). Pharmaceutically acceptable excipients can include a buffer, such as a phosphate buffer (e.g. sodium phosphate). Pharmaceutically acceptable excipients can include a salt, for example sodium chloride. Pharmaceutically acceptable excipients can include a solubilizing/stabilizing agent, for example, polysorbate (e.g. TWEEN 80). Pharmaceutically acceptable excipients can include a preservative, for example 2-phenoxyethanol or thiomersal. Pharmaceutically acceptable excipients can include a carrier such as water or saline.

[0179] Also provided is a method of making the immunogenic composition of the invention comprising the step of mixing the modified EPA protein or the conjugate (e.g. bioconjugate) of the invention with a pharmaceutically acceptable excipient and/or carrier.

[0180] The present invention also provides an immunogenic composition (e.g., a vaccine composition) optionally comprising an adjuvant.

[0181] The term “adjuvant” refers to a compound that when administered in conjunction with or as part of an immunogenic composition of vaccine of the invention augments, enhances and/or boosts the immune response to modified EPA protein conjugate/bioconjugate, but when the compound is administered alone does not generate an immune response to the modified EPA protein conjugate/bioconjugate. Adjuvants can enhance an immune response by several mechanisms including, e.g. lymphocyte recruitment, stimulation of B and/or T cells, and stimulation of macrophages. Specific examples of adjuvants include, but are not limited to, aluminum salts (alum) (such as aluminum hydroxide, aluminum phosphate, and aluminum sulfate), 3 De-O-acylated monophosphoryl lipid A (MPL) (see United

Kingdom Patent GB2220211), MF59 (Novartis), AS01 (GlaxoSmithKline), and saponins, such as QS21 (see Kensil et al. in Vaccine Design: The Subunit and Adjuvant Approach (eds. Powell & Newman, Plenum Press, NY, 1995); U.S. Pat. No. 5,057,540). In some embodiments, the adjuvant is Freund's adjuvant (complete or incomplete). Other adjuvants are oil in water emulsions (such as squalene or peanut oil), optionally in combination with immune stimulants, such as monophosphoryl lipid A (see Stoute et al. N. Engl. J. Med. 336, 86-91 (1997)).

[0182] Also provided is a method of making the immunogenic composition of the invention comprising the step of mixing the modified EPA protein or the conjugate (e.g. bioconjugate) of the invention with a pharmaceutically acceptable excipient and/or carrier and an adjuvant. Vaccine preparation is generally described in Vaccine Design (“The subunit and adjuvant approach” (eds Powell M. F. & Newman M. J.) (1995) Plenum Press New York).

[0183] The immunogenic compositions of the invention can be included in a container, pack, or dispenser together with instructions for administration.

[0184] The immunogenic compositions or vaccines of the invention can be stored before use, e.g.

[0185] the compositions can be stored frozen (e.g. at about -20°C . or at about -70°C .); stored in refrigerated conditions (e.g. at about 4°C .); or stored at room temperature. The immunogenic compositions or vaccines of the invention may be stored in solution or lyophilized. In an embodiment, the solution is lyophilized in the presence of a sugar such as sucrose, trehalose or lactose. In another embodiment, the vaccines of the invention are lyophilized and extemporaneously reconstituted prior to use.

Administration and Dosage

[0186] Immunogenic compositions or vaccines of the invention may be used to protect or treat a subject (e.g. mammal), by means of administering said immunogenic composition or vaccine via systemic or mucosal route. These administrations may include injection via the intramuscular (IM), intraperitoneal, intradermal (ID) or subcutaneous (SC) routes; or via mucosal administration to the oral/alimentary, respiratory, genitourinary tracts.

[0187] In one aspect, the immunogenic composition or vaccine of the invention is administered by the intramuscular delivery route. Intramuscular administration may be to the thigh or the upper arm. Injection is typically via a needle (e.g. a hypodermic needle), but needle-free injection may alternatively be used. A typical intramuscular dose is 0.5 ml.

[0188] In another aspect, the immunogenic composition or vaccine of the invention is administered by the intradermal administration. Human skin comprises an outer “horny” cuticle, called the stratum corneum, which overlays the epidermis. Underneath this epidermis is a layer called the dermis, which in turn overlays the subcutaneous tissue. The conventional technique of intradermal injection, the “mantoux procedure”, comprises steps of cleaning the skin, and then stretching with one hand, and with the bevel of a narrow gauge needle (26 to 31 gauge) facing upwards the needle is inserted at an angle of between 10 to 15° . Once the bevel of the needle is inserted, the barrel of the needle is lowered and further advanced whilst providing a slight pressure to elevate it under the skin. The liquid is then injected very slowly thereby forming a bleb or bump on the skin surface, followed by slow withdrawal of the needle.

[0189] In another aspect, the immunogenic composition or vaccine of the invention is administered by the intranasal administration. Typically, the immunogenic composition or vaccine is administered locally to the nasopharyngeal area, e.g. without being inhaled into the lungs. It is desirable to use an intranasal delivery device which delivers the immunogenic composition or vaccine formulation to the nasopharyngeal area, without or substantially without it entering the lungs. Suitable devices for intranasal administration of the vaccines according to the invention are spray devices. Suitable commercially available nasal spray devices include ACCUSPRAY™ (Becton Dickinson).

[0190] The amount of conjugate (e.g. bioconjugate) in each immunogenic composition or vaccine dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccines. Such amount will vary depending upon which specific immunogen is employed and how it is presented. The content of conjugate (e.g. bioconjugate) will typically be in the range 1-100 µg, suitably 5-50 µg.

Prophylactic and Therapeutic Uses

[0191] The present invention also provides an immunogenic composition of the invention, or the vaccine of the invention, for use in medicine.

[0192] The present invention provides a method of inducing an immune response in a subject (e.g. human), the method comprising administering a therapeutically or prophylactically effective amount of a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention or a vaccine of the invention, to a subject (e.g. human) in need thereof. The present invention also provides a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention or a vaccine of the invention, for use in inducing an immune response in a subject (e.g. human). The present invention also provides a conjugate (e.g. bioconjugate) of the invention, the immunogenic composition of the invention or the vaccine of the invention for use in the manufacture of a medicament for inducing an immune response in a subject (e.g. human).

[0193] Also provided herein are methods of inducing an immune response in a subject against a bacterium, comprising administering to the subject a conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention. The conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention can be used to induce an immune response against a bacterium, e.g. *Shigella* species, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *N. meningitidis*, *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In an embodiment, the conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention can be used to induce an immune response against a bacterium, e.g. *Streptococcus* species, *Shigella* species, *Pseudomonas* species, *Klebsiella* species, or *Staphylococcus* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, or *Staphylococcus aureus*). In one embodiment, said subject has bacterial infection at the time of administration. In another embodiment, said subject does not have a bacterial infection at the time of administration.

[0194] Also provided herein are methods of inducing the production of opsonophagocytic antibodies in a subject against a bacterium, comprising administering to the subject a conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention. The conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention can be used to induce the production of opsonophagocytic antibodies in a subject against a bacterium, e.g. *Shigella* species, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *N. meningitidis*, *H. influenzae* type b (Hib), Group B *Streptococcus* (GBS), *Streptococcus pneumoniae*, or *Staphylococcus aureus*. In an embodiment, the conjugate (e.g. bioconjugate) of the invention an immunogenic composition of the invention or a vaccine of the invention can be used to induce the production of opsonophagocytic antibodies in a subject against a bacterium, e.g. *Streptococcus* species, *Shigella* species, *Pseudomonas* species, *Klebsiella* species, or *Staphylococcus* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae* or *Staphylococcus aureus*).

[0195] The present invention also provides methods of treating and/or preventing a yeast or bacterial infection in a subject comprising administering to the subject a conjugate (e.g. bioconjugate) of the invention. The conjugate (e.g. bioconjugate) may be in the form of an immunogenic composition or vaccine. Thus the present invention provides a method of treating and/or preventing a yeast or bacterial infection in a subject (e.g. human), the method comprising administering a therapeutically or prophylactically effective amount of a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention or a vaccine of the invention, to a subject (e.g. human) in need thereof. The present invention also provides a conjugate (e.g. bioconjugate) of the invention, an immunogenic composition of the invention or a vaccine of the invention, for use in treating and/or preventing a yeast or bacterial infection in a subject (e.g. human). The present invention also provides a conjugate (e.g. bioconjugate) of the invention, the immunogenic composition of the invention or the vaccine of the invention for use in the manufacture of a medicament for treating and/or preventing a yeast or bacterial infections in a subject (e.g. human). In a specific embodiment, the immunogenic composition or vaccine of the invention is used in the prevention of infection of a subject by a bacterium. Bacterial infections that can be treated and/or prevented using the conjugate (e.g. bioconjugate) of the invention include those caused by *N. meningitidis*, *H. influenzae* type b (Hib), *Streptococcus* species, *Shigella* species, *Pseudomonas* species, *Klebsiella* species, or *Staphylococcus* species. (e.g. *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae* or *Staphylococcus aureus*).

Embodiments of the invention are further described in the subsequent numbered paragraphs:

[0196] 1. A modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises one (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z

are independently any amino acid except proline, wherein the one (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213; or one or more amino acids between 205-211; e.g. amino acid residue D218; e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279; or one or more amino acids between amino acid residues 271-277, e.g. amino acid residue R279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323; or one or more amino acids between amino acid residues 315-321, e.g. amino acid residue G323, e.g. amino acid residue S318), and (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; or one or more amino acids between amino acid residues 516-522; e.g. amino acid residue G525, e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0197] 2. The modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) of paragraph 1 having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises two (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the two (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519), and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240), of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0198] 3. A modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein of paragraph 1 having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises three (or more) consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X

and Z are independently any amino acid except proline, wherein the three (or more) consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between amino acid residues 264-284 (e.g. one or more amino acids between amino acid residues 269-279, e.g. amino acid residue R274), (iii) one or more amino acids between amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), (iv) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 (e.g. one or more amino acids between amino acid residues 235-245; e.g. amino acid residue K240) or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0199] 4. The modified EPA protein of any of paragraphs 1 to 3, wherein the consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), have each been independently substituted for one or more amino acids (e.g. each consensus sequence is substituted for a single amino acid residue, such as a single amino acid residue selected from Y208, R274, S318 and A519) of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0200] 5. The modified EPA protein of any of paragraphs 1 to 4, wherein a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, has been added next to, or substituted for, one or more amino acids, at the N-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0201] 6. The modified EPA protein of any of paragraphs 1 to 5, wherein a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, has been added next to, or substituted for, one or more amino acids, at the C-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0202] 7. The modified EPA protein of any of paragraphs 1 to 6, wherein at least one consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, has been added next to, or substituted for: (i) one or more amino acids between amino acid residues 198-218 (e.g. one or more amino acids between amino acid residues 203-213, e.g. amino acid residue Y208), (ii) one or more amino acids between

amino acid residues 308-328 (e.g. one or more amino acids between amino acid residues 313-323, e.g. amino acid residue S318), or (iii) one or more amino acids between amino acid residues 509-529 (e.g. one or more amino acids between amino acid residues 514-524; e.g. amino acid residue A519) of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0203] 8. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains two consensus sequences, optionally substituted for amino acid residues selected from: (i) Y208 and R274, (ii) Y208 and S318, (iii) Y208 and A519, (iv) R274 and S318, (v) R274 and A519, or (vi) S318 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0204] 9. The modified EPA protein of paragraph 8 wherein the modified EPA protein contains two consensus sequences substituted for amino acid residues Y208 and R274 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 6:

```
AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVL
EGGNDALKLAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSW
LNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLAR
DATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRWSEWASGKVL
CLLDPLDGVYNKDQNAKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDI
KPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTKDQNAKHKRQPRGWE
QLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDLGEAIRE
QPEQARLALTLLAAEESERFVRQGTGNDEAGAASADVSLTCPVAAGECA
GPADSGDALLERNYPTGAFLDGGDVFSFSTRGTQNWTVRLLQHRQL
EERGVYFVGHYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALA
YGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLAAPEAAGEV
ERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTTVVIPSAPITDPRN
VGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK
```

[0205] 10. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains three consensus sequences, optionally substituted for amino acid residues Y208, R274 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0206] 11. The modified EPA protein of paragraph 10 wherein the modified EPA protein contains three consensus sequences substituted for amino acid residues Y208, R274 and A519 of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 7:

```
AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVL
EGGNDALKLAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSW
LNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLAR
DATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRWSEWASGKVL
CLLDPLDGVYNKDQNAKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDI
KPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTKDQNAKHKRQPRGWE
QLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDLGEAIRE
QPEQARLALTLLAAEESERFVRQGTGNDEAGAASADVSLTCPVAAGECA
GPADSGDALLERNYPTGAFLDGGDVFSFSTRGTQNWTVRLLQHRQL
EERGVYFVGHYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALA
YGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLKQNAKAP
EAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTTVVIPSAP
PTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK
```

[0207] 12. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains four consensus sequences, optionally substituted for amino acid residues Y208, R274, A519 and added next to the N-terminal amino acid of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0208] 13. The modified EPA protein of paragraph 12, wherein the modified EPA protein contains four consensus sequences substituted for amino acid residues Y208, R274, A519 and added next to the N-terminal amino acid of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 8:

```
GSGGGDQNAATGSGGKLAEAFDLWNECAKACVLDLKDGVRSRMSVDP
AIADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIRLEGGVEP
NKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMS
PIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQA
QPRREKRWSEWASGKVLCLLDPLDGVYNKDQNAKLAQQRCLDDTWEG
KIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEA
FTKDQNAKHKRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRN
ALASPGSGDLGEAIREQPEQARLALTLLAAEESERFVRQGTGNDEAGAA
SADVSLTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVFSFSTR
GTQNWTVRLLQHRQLEERGVYFVGHYHGTFLAAQSIIVFGGVRARSQD
LDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPG
FYRTGLTLKQNAKAPAEAGEVERLIGHPLPLRLDAITGPEEEGGRVT
ILGWPLAERTTVVIPSAPITDPRNVGGDLDPSSIPDKEQAISALPDYASQ
PGKPPREDLK
```

[0209] 14. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains five

consensus sequences, optionally selected from: substitution of amino acid residue Y208, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0210] 15. The modified EPA protein of paragraph 14, wherein the modified EPA protein contains five consensus sequences selected from: substitution of amino acid residue Y208, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 9, 10 or 11:

SEQ NO: 9:

GSGGG**QDNAT**GSGGGAAEEAFDLWNECAKACVLDLKDGVSSRMSVDP
ADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPI
YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGSVVMAQAQ
RREKRWSEWASGKVLCLLDPLDGVYN**KDNATKLA**QQRCLDDTWEGKI
YRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT
KDNATKKHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNAL
ASPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAASA
DVVSLTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRGT
QNWTVERRLLQAHRLQLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLD
AIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVPRWSLPGFY
RTGLTL**KDNATKA**PEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTIL
GWPLAERTTVVIPSIAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPG
KPPREDLGSGGG**QDNAT**GSGG

SEQ ID NO: 10:

GSGGG**QDNAT**GSGGGAAEEAFDLWNECAKACVLDLKDGVSSRMSVDP
ADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPI
YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGSVVMAQAQ
RREKRWSEWASGKVLCLLDPLDGVYN**KDNATKLA**QQRCLDDTWEGKI
YRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT
RHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNAL**AKDQNA**
TKSPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAAS
ADVVS LTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRG
TQNWTVERRLLQAHRLQLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLD
DAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVPRWSLPGF

-continued

YRTGLTL**KDNATKA**PEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTI
LGWPLAERTTVVIPSIAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQ
GKPPREDLGSGGG**QDNAT**GSGG
SEQ ID NO: 11:
GSGGG**QDNAT**GSGGGAAEEAFDLWNECAKACVLDLKDGVSSRMSVDP
ADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPI
YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGSVVMAQAQ
RREKRWSEWASGKVLCLLDPLDGVYN**KDNATKLA**QQRCLDDTWEGKI
YRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT**KDNAT**
KHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNAL**AKDQNA**
TKSPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAAS
ADVVS LTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRG
TQNWTVERRLLQAHRLQLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLD
DAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVPRWSLPGF
YRTGLTL**KDNATKA**PEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTI
LGWPLAERTTVVIPSIAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQ
GKPPREDLGSGGG**QDNAT**GSGG

[0211] 16. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains six consensus sequences, optionally selected from: substitution of amino acid residue Y208, substitution of amino acid residue K240, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0212] 17. The modified EPA protein of paragraph 16, wherein the modified EPA protein contains six consensus sequences selected from: substitution of amino acid residue Y208, substitution of amino acid residue K240, substitution of amino acid residue R274, substitution of amino acid residue S318, substitution of amino acid residue A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 12 or 13:

SEQ ID NO: 12:

GSGGG**QDNAT**GSGGGAAEEAFDLWNECAKACVLDLKDGVSSRMSVDP
ADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPI
YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGSVVMAQAQ
RREKRWSEWASGKVLCLLDPLDGVYN**KDNATKLA**QQRCLDDTWEGKI
YRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFT

-continued

KDQ¹NATK²HRQ³PRGWEQ⁴LEQ⁵CGYPVQ⁶RLVALYLAARLSWNQVDQ⁷VIRNAL
AKDQ¹NATK²SPGSGGDLGEAIREQ³PEQARLALT⁴LAAAESERFVRQ⁵GTGND
EAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLGDGGDV
SFSTRGTQ¹NWTV²VERLLQ³AHRQ⁴LEERGYPV⁵FGYHGT⁶FLEAAQSI⁷VFGGVR
ARSQDLDAIWRGFIAGDPALAYGYAQDQEPDARGIRNGALLRVYVPR
WSLPGFYRTGLTLKDQ¹NATKAPEAAGEVERLIGHPLPLRLDAITGPEEE
GGRVTILGWPLAERTV¹VIPSAIPTDPRNVGGDLDPSSIPDKEQAISALP
DYASQPGKPPREDLGSGGGDQ¹NATGSGG

SEQ ID NO: 13

GSGGGDQ¹NATGSGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPAI
ADTNGQGV¹LHYSVM²LEGGNDALKLAIDNALSITS³DGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVP¹IGHEKPSNIKVFIHELNAGNQLSHMSPI
YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ¹P
RREKRWSEWASGKVLCLLDPLDGVYNKDQ¹NATKLAQ²QRCNLDDT³WEGKI
YRVLAGNPAKHDL¹DKDNNNSTPT²TVISHRLHFPEGGSLAALTAHQACHL
PLEAFTKDQ¹NATK²HRQ³PRGWEQ⁴LEQ⁵CGYPVQ⁶RLVALYLAARLSWNQVDQ⁷
VIRNALAKDQ¹NATK²SPGSGGDLGEAIREQ³PEQARLALT⁴LAAAESERFVR
QGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFL
GDGGDV¹SFSTRGTQ²NWTV³VERLLQ⁴AHRQ⁵LEERGYPV⁶FGYHGT⁷FLEAAQSI
VFGGVRARSQDLDAIWRGFIAGDPALAYGYAQDQEPDARGIRNGALL
RVYVPRWSLPGFYRTGLTLKDQ¹NATKAPEAAGEVERLIGHPLPLRLDAI
TGPEEEGGRVTILGWPLAERTV¹VIPSAIPTDPRNVGGDLDPSSIPDKEQ
AISALPDYASQPGKPPREDLK

[0213] 18. The modified EPA protein of any of paragraphs 1 to 7, wherein the modified EPA protein contains seven consensus sequences, optionally substitution of amino acid residues Y208, K240, R274, S318, A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

[0214] 19. The modified EPA protein of paragraph 8, wherein the modified EPA protein contains seven consensus sequences, substituted for amino acid residues Y208, K240, R274, S318, A519, addition at the N-terminus and addition at the C-terminus of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, optionally comprising (or consisting of) an amino acid sequence which is at least 95%, 96%, 97%, 98%, 99% or 100% identical to SEQ ID NO: 14:

SEQ ID NO: 14:

GSGGGDQ¹NATGSGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPAI
ADTNGQGV¹LHYSVM²LEGGNDALKLAIDNALSITS³DGLTIRLEGGVEPNK
PVRYSYTRQARGSWSLNWLVP¹IGHEKPSNIKVFIHELNAGNQLSHMSPI

-continued

YTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ¹P
RREKRWSEWASGKVLCLLDPLDGVYNKDQ¹NATKLAQ²QRCNLDDT³WEGKI
YRVLAGNPAKHDL¹DKDNNNSTPT²TVISHRLHFPEGGSLAALTAHQACHL
PLEAFTKDQ¹NATK²HRQ³PRGWEQ⁴LEQ⁵CGYPVQ⁶RLVALYLAARLSWNQVDQ⁷
VIRNALAKDQ¹NATK²SPGSGGDLGEAIREQ³PEQARLALT⁴LAAAESERFVR
QGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFL
GDGGDV¹SFSTRGTQ²NWTV³VERLLQ⁴AHRQ⁵LEERGYPV⁶FGYHGT⁷FLEAAQSI
VFGGVRARSQDLDAIWRGFIAGDPALAYGYAQDQEPDARGIRNGALL
RVYVPRWSLPGFYRTGLTLKDQ¹NATKAPEAAGEVERLIGHPLPLRLDAI
TGPEEEGGRVTILGWPLAERTV¹VIPSAIPTDPRNVGGDLDPSSIPDKEQ
AISALPDYASQPGKPPREDLGSGGGDQ¹NATGSGG

[0215] 20. The modified EPA protein of any one of paragraphs 1 to 19, wherein X is Q (glutamine) and Z is A (alanine).

[0216] 21. The modified EPA protein of any of paragraphs 1 to 20, wherein the amino acid sequence comprises substitution of leucine 552 to valine (L552V) (or at a position equivalent to L552 of SEQ ID NO:1) and deletion of glutamine 553 (Δ E553) (or at a position equivalent to E553 of SEQ ID NO:1).

[0217] 22. The modified EPA protein of any of paragraphs 1 to 21, wherein the amino acid sequence further comprises a peptide tag, optionally said peptide tag comprising six histidine residues and optionally said peptide tag located at the C-terminus of the amino acid sequence.

[0218] 23. The modified EPA protein of any of paragraphs 1 to 22, wherein the amino acid sequence further comprises a signal sequence which is capable of directing the EPA protein to the periplasm of a host cell (e.g. bacterium), optionally said signal sequence being DsbA (SEQ ID NO: 21).

[0219] 24. A conjugate (e.g. a bioconjugate) comprising a modified EPA protein of any of paragraphs 1 to 23 linked to an antigen (e.g. a saccharide antigen, optionally a bacterial polysaccharide antigen).

[0220] 25. The conjugate according to paragraph 24, wherein the modified EPA protein is covalently linked to an antigen through a chemical linkage obtainable using a chemical conjugation method, either directly or via a linker.

[0221] 26. The conjugate (e.g. bioconjugate) of paragraph 24, wherein the antigen is covalently linked to an amino acid on the modified EPA protein selected from asparagine, aspartic acid, glutamic acid, lysine, cysteine, tyrosine, histidine, arginine or tryptophan (e.g. asparagine).

[0222] 27. The conjugate (e.g. bioconjugate) of any one of paragraphs 24 to 26, wherein the antigen is a saccharide, optionally a bacterial polysaccharide (e.g. from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae* or *Staphylococcus aureus*), optionally an O-antigen from a Gram negative bacterium.

[0223] 28. A polynucleotide encoding the modified EPA protein of any of paragraphs 1 to 23.

[0224] 29. A vector comprising the polynucleotide of paragraph 28.

[0225] 30. A host cell comprising:

- [0226] i) one or more nucleotide sequences comprising polysaccharide synthesis genes, optionally for producing a bacterial polysaccharide antigen (e.g. an O-antigen from a Gram negative bacterium optionally from *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* or a capsular polysaccharide from a Gram positive bacterium optionally from *Streptococcus pneumoniae* or *Staphylococcus aureus*) or a yeast polysaccharide antigen or a mammalian polysaccharide antigen, optionally integrated into the host cell genome;
 - [0227] ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, optionally within a plasmid;
 - [0228] iii) a nucleotide sequence that encodes a modified EPA protein according to any of paragraphs 1 to 23, optionally within a plasmid.
- [0229] 31. A host cell according to paragraph 30 further comprising a nucleotide sequence encoding a polymerase (e.g. wzy), a flippase (e.g. wz4 and optionally a nucleotide sequence encoding and/or a chain length regulator (e.g. wzz).
- [0230] 32. The host cell according to paragraph 30 or paragraph 31 wherein the oligosaccharyl transferase is a Pg/B, optionally derived from *Campylobacter jejuni*.
- [0231] 33. The host cell according to any of paragraphs 30 to 31, wherein the host cell is *E. coli* (e.g. *E. coli* K12 W3110).
- [0232] 34. A process for producing a bioconjugate that comprises a modified EPA protein linked to a polysaccharide, said process comprising (i) culturing the host cell of any one of paragraphs 30 to 33 under conditions suitable for the production of glycoproteins and (ii) isolating the bioconjugate, optionally isolating the bioconjugate from a periplasmic extract from the host cell.
- [0233] 35. An immunogenic composition comprising a conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, and optionally a pharmaceutically acceptable excipient and/or carrier.
- [0234] 36. A vaccine comprising the immunogenic composition of paragraph 35 and optionally an adjuvant.
- [0235] 37. A method of inducing an immune response in a subject (e.g. human), the method comprising administering a therapeutically or prophylactically effective amount of the conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic composition of paragraph 35 or the vaccine of paragraph 36, to a subject (e.g. human) in need thereof.
- [0236] 38. The conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic composition of paragraph 35 or the vaccine of paragraph 36, for use in inducing an immune response in a subject (e.g. human).
- [0237] 39. The conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic composition of paragraph 35 or the vaccine of paragraph 36, for use in the manufacture of a medicament for inducing an immune response in a subject (e.g. human).
- [0238] 40. A method of treating and/or preventing a yeast or bacterial infection in a subject (e.g. human), the method comprising administering a therapeutically or prophylactically effective amount of the conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic

composition of paragraph 35 or the vaccine of paragraph 36, to a subject (e.g. human) in need thereof.

[0239] 41. The conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic composition of paragraph 35 or the vaccine of paragraph 36, for use in treating and/or preventing a yeast or bacterial infection in a subject (e.g. human).

[0240] 42. The conjugate (e.g. bioconjugate) of any of paragraphs 24 to 27, the immunogenic composition of paragraph 35 or the vaccine of paragraph 36, for use in the manufacture of a medicament for treating and/or preventing a yeast or bacterial infection in a subject (e.g. human). In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only, and are not to be construed as limiting the scope of the invention in any manner.

EXAMPLES

[0241] Material and Methods

[0242] Engineering of EPA for Glycosylation with Antigenic Glycans

[0243] In order to predict suitable positions for insertion of glycosites, the crystal structure of EPA was analyzed. 67 solvent accessible amino acid residues were selected for site directed mutagenesis. As a template for mutagenesis the genetically detoxified EPA containing mutations L552VΔE553 was used (Lukac et al (1988), Infect Immun, 56: 3095-3098, and Ho et al. (2006), Hum Vaccin, 2:89-98). This gene was cloned with the DsbA signal sequence at the N-terminus and a His₆ tag at C-terminus into a plasmid derived from pEC415 (Schulz, H., Hennecke, H., Thony-Meyer, L., Prototype of a heme chaperone essential for cytochrome c maturation. In *Science* 281, 1197-1200, 1998). Each selected amino acid residue was substituted with the glycosylation sequon KDQNATK (SEQ ID NO: 4) leading to 67 EPA variants in total, each containing a single glycosylation site (glycosite). For generation of EPA variants containing two, three and four glycosites additional rounds of mutagenesis in the available and selected single-site EPA variants was performed. For generation of EPA variants containing more than four glycosites gene synthesis was applied. Out of 67 tested variants, 4 positions were selected for further combinations presented in this work: Y208, R274, S318 and A519. These mutations were found to be advantageous because (i) they did not reduce protein expression level, indicating that the overall protein structure was not influenced i.e. destabilized by insertion of the glycosite at these positions, and (ii) the selected positions were found to provide efficient glycosylation i.e. high site occupancy. Additional positions in the region close to Y208, R274, S318 and A519 were also investigated for suitability for glycosite insertion (D218, R279, G232 and G525).

[0244] Glycosylation Tests with Engineered EPA Containing One or More Glycosites

[0245] The 67 EPA variants containing a single inserted glycosite were tested for in vivo glycosylation efficiency using various antigenic glycans. For such glycosylation tests, *E. coli* strain W3110 ΔwaaL was transformed with three plasmids: a pEC415 plasmid carrying an EPA variant, a plasmid expressing PglB and with a plasmid expressing enzyme for the biosynthesis of the polysaccharide of interest (as described below). The non-pathogenic *E. coli* K12 strain W3110 was obtained from the Coli Genetic Stock Center (Yale University, New Haven (Conn.), USA, product num-

ber CGSC #4474). In some cases, glycosylation test were performed with an *E. coli* strain in which the cluster of genes for polysaccharide biosynthesis was integrated into the *E. coli* genome (see WO2014/057109 and WO2015/052344 for further details relating to integration), which allowed transformation only with two plasmids expressing EPA variant and PglB.

[0246] The selection criteria for EPA variants with single glycosite included the total expression level and the level of produced glycoconjugate, the later indicating suitability of glycosite position for modification by PglB.

[0247] For the data set presented in this work, the used *E. coli* strains are derivatives of strain W3110, which include a deletion in the lipopolysaccharide O-antigen ligase gene *waaL*, the deletion or replacement of the O16 O-antigen cluster *rfb* and the replacement of a genomic cluster with the cluster responsible for the biosynthesis of the wanted recombinant glycan of *Klebsiella pneumoniae* O-antigen (KpO-antigen) *Shigella flexneri* 2a (Sf2a) or *Streptococcus pneumoniae* 11A (Sp11A) and 33F (Sp33F) capsular polysaccharides and *Pseudomonas aeruginosa* O-antigens (PaO6 and PaO11). The *E. coli* strain used for the recombinant production of the *K. pneumoniae* O-antigen in FIG. 6 and FIG. 8 (P018-0183) glycan contains in addition genomic deletions of potentially interfering elements: genes *wzzE-wzcG* from the enterobacterial common antigen (ECA) *wec* cluster, the colanic acid *wca* cluster, *gtrABS* and *wzzB* genes involved in the O16 biosynthesis.

[0248] The *E. coli* strain producing KpO-antigen for FIG. 1-5 or KpO-antigen for FIG. 8 (P018_0167) glycan was transformed with a pEC415 plasmid carrying an EPA variant and a plasmid expressing PglB. To prepare a pre-culture, 5 ml TB (Terrific Broth) medium containing 10 mM MgCl₂ and appropriate antibiotics was inoculated with a streak of colonies from the transformation plate and grown at 37° C. o/n (overnight). The pre-culture was used to inoculate 50 ml of supplemented TB medium in a shake flask to give a starting OD₆₀₀=0.1. The cultures were grown at 37° C., with 200 rpm shaking until reaching OD₆₀₀=0.8-1 and then induced by addition of 0.001% arabinose (EPA) and 0.1 mM IPTG

[0249] (PglB). The expression and glycosylation of EPA variants was continued at 37° C. o/n.

[0250] The *E. coli* strain producing KpO-antigen for FIG. 6 and FIG. 8 (P018-0183) was transformed with a plasmid expressing both EPA variant and PglB. These cultures were induced only by addition of 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression and glycosylation of EPA variants was continued at 37° C. o/n.

[0251] Periplasmic Extract Preparation

[0252] The amount of cells from o/n cultures corresponding to OD₆₀₀=60 (measured using a spectrophotometer) was harvested by centrifugation. The cell pellets were resuspended in 1.5 ml of lysis buffer (30 mM Tris-HCl pH 8.5, 1 mM EDTA (Ethylenediaminetetraacetic acid), 20% sucrose) and lysozyme was added to a final concentration of 1 mg/ml. The suspensions were incubated with slight shaking for 25 minutes at 4° C. and then centrifuged at 16,000 rcf for 10 min. After centrifugation, the supernatant corresponding to periplasmic extract (PPE) was transferred to a fresh tube.

[0253] Enrichment of Periplasmic Extract by Immobilized Metal Affinity Chromatography (IMAC)

[0254] In order to enrich periplasmic extracts with EPA variants and allow more direct read-out by SDS-PAGE, the

His-tagged EPA variants were purified using one-step purification on Ni-NTA (Nickel Nitrilo-triacetic Acid) agarose. 1 ml of PPE was mixed with 250 μl of 5× binding buffer (150 mM Tris-HCl pH 8.0, 50 mM imidazole, 2.5 M NaCl, 20 mM MgCl₂) followed by addition of 200 μl of pre-equilibrated Ni-NTA slurry and incubated with slight shaking for 30 min. After that the resin was washed with 1× binding buffer (30 mM Tris pH 8.0, 10 mM imidazole, 500 mM NaCl) and the bound protein eluted with elution buffer (30 mM Tris pH 8.0, 500 mM imidazole, 200 mM NaCl). The IMAC enriched PPE was analysed by SDS-PAGE (Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4". *Nature*. 227 (5259): 680-685. Bibcode:1970Natur.227.680L. doi:10.1038/227680a0. ISSN 0028-0836. PMID 5432063). Unglycosylated carrier and EPA glycoconjugates glycosylated at 1 and more positions were detected on the gel by Coomassie staining (Fazekas de St. Groth, S.; Webster, R. G.; Datyner, A. (1963). "Two new staining procedures for quantitative estimation of proteins on electrophoretic strips". *Biochimica et Biophysica Acta*. 71: 377-391. doi:10.1016/0006-3002(63)91092-8. PMID 18421828).

[0255] Western Blot Analysis of Periplasmic Extract

[0256] Periplasmic extracts were also analysed by immunoblots against EPA (Sigma-Aldrich, Cat. number P2318) and against polysaccharide attached to EPA. For detection of KpO-antigen, anti-serum against a K-capsular mutant of *Klebsiella pneumoniae* was used.

[0257] The results of these experiments are shown in FIGS. 3 to 6 and as described in the following Examples.

Example 1: (FIG. 3)

[0258] SDS-PAGE analysis was carried out on IMAC enriched periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with a glycosite KDQNATK (SEQ ID NO: 4) introduced at the following positions into SEQ ID NO: 1: Y208 (lane 1), K240 (lane 2), R274 (lane 3), S318 (lane 4), A376 (lane 5), A519 (lane 6), and K240 and A376 (lane 7). The bands shown in FIG. 3 correspond to the unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one and two occupied glycosites.

[0259] Conclusion:

[0260] Glycosylation of EPA with KpO-antigen at each of the new positions was confirmed to be equally good or superior compared to positions K240 and A376. In particular, Y208, S318 and A519 look superior compared to positions K240 and A376. Y208 has higher total expression and all 3 have higher amount of conjugates than positions K240 and A376.

Example 2: (FIG. 4)

[0261] SDS-PAGE analysis was carried out on IMAC enriched periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with a glycosite KDQNATK (SEQ ID NO: 4) introduced at the following positions: K240 and A376 (lane 1, SEQ ID NO: 26), Y208 and R274 (lane 2, SEQ ID NO: 6), Y208 and S318 (lane 3, SEQ ID NO: 28), Y208 and A519 (lane 4; SEQ ID NO: 29), R274 and S318 (lane 5; SEQ ID NO: 30), R274 and A519 (lane 6; SEQ ID NO: 31), S318 and A519 (lane 7, SEQ ID NO: 7), and Y208 and R274 and A519 (lane 8). The bands shown in FIG. 4 correspond to the

unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one, two and three occupied glycosites.

[0262] Conclusion:

[0263] Glycosylation of EPA-2S variants (i.e. variants of EPA having 2 glycosylation sites added) containing glycosites at the new positions with KpO-antigen was confirmed to be equally good or superior compared to EPA-2S with glycosites at positions K240 and A376. From the gel shown in FIG. 4, Y208+A519 and S318+A519 have higher total protein and all 2S combinations have higher ratio of double to single glycosylated EPA. EPA-3S (i.e. variants of EPA having 3 glycosylation sites added) with glycosites at positions Y208, R274 and A519 was also well glycosylated with KpO-antigen.

Example 3: (FIG. 5)

[0264] Immunoblot analysis was carried out on periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide and expressing PglB and EPA variants with 1 to 7 glycosites introduced at the following positions: Y208 (lane 1), K240 (lane 2), R274 (lane 3), S318 (lane 4), A376 (lane 5), A519 (lane 6), and K240 and A376 (lane 7), Y208 and R274 (lane 8, SEQ ID NO: 6), Y208 and S318 (lane 9), Y208 and A519 (lane 10), R274 and S318 (lane 11), R274 and A519 (lane 12), S318 and A519 (lane 13), Y208 and R274 and A519 (lane 14, SEQ ID NO: 7), N-terminal glycotag and K240 and A376 and C-terminal glycotag (lane 15), N-terminal glycotag and Y208 and R274 and A519 (lane 16, SEQ ID NO: 8), N-terminal glycotag and Y208 and R274 and A519 and C-terminal glycotag (lane 17; SEQ ID NO: 9), N-terminal glycotag and Y208 and S318 and A519 and C-terminal glycotag (lane 18, SEQ ID NO: 10), N-terminal glycotag and R274 and S318 and A519 and C-terminal glycotag (lane 19, SEQ ID NO: 11), N-terminal glycotag and Y208 and R274 and S318 and A519 and C-terminal glycotag (lane 20, SEQ ID NO: 12), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 (lane 21, SEQ ID NO: 13), and N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 and C-terminal glycotag (lane 22, SEQ ID NO: 14). In FIG. 5, the upper panel represents the immunoblot probed with anti-KpO-antigen anti-serum, while the bottom panel represents the immunoblot probed with anti-EPA antibody. The bands correspond to the unglycosylated EPA carrier, and to KpO-antigen-EPA bioconjugates with one to seven occupied glycosites.

[0265] Conclusion:

[0266] This demonstrated that glycosylation at up to 7 glycosites is possible by combining the glycosite positions when using a variant of EPA which includes the combination of seven consensus sequences.

Example 4: (FIG. 6)

[0267] SDS-PAGE analysis was carried out on periplasmic extract of *E. coli* strains producing KpO-antigen polysaccharide (of a different serotype to that presented in FIG. 1-5, Examples 1-3) and expressing PglB and EPA variants with a glycosite KDQNA TK (SEQ ID NO: 4) introduced at the following positions: N-terminal glycotag and Y208 and R274 and A519 and C-terminal glycotag (lane 1, SEQ ID NO: 9), N-terminal glycotag and Y208 and S318 and A519 and C-terminal glycotag (lane 2, SEQ ID NO: 10), N-terminal glycotag and R274 and S318 and A519 and C-terminal glycotag (lane 3, SEQ ID NO: 11), N-terminal glycotag

and Y208 and R274 and S318 and A519 and C-terminal glycotag (lane 4, SEQ ID NO: 12), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 (lane 5, SEQ ID NO: 13), N-terminal glycotag and Y208 and K240 and R274 and S318 and A519 and C-terminal glycotag (lane 6, SEQ ID NO: 14), and N-terminal glycotag and Y208 and R274 and A519 (lane 7, SEQ ID NO: 8). In FIG. 6, the bands corresponding to KpO-antigen-EPA bioconjugates are labelled with arrows.

[0268] Conclusion:

[0269] Another *Klebsiella pneumoniae* serotype O-antigen can be used to show glycosylation efficiency of EPA having more than 3 glycosites.

Example 5: (FIG. 7)

[0270] SDS-PAGE analysis was carried out on IMAC enriched periplasmic extracts of *E. coli* strain producing Sf2a (see FIG. 7—left) or Sp11A (see FIG. 7—right) polysaccharide and expressing PglB and EPA variants with a glycosite introduced at K240 and a second glycosite at A376 (lane 1, SEQ ID NO: 26) or three glycosites at positions Y208, R274, and A519 (lane 2, SEQ ID NO: 7).

[0271] Results:

TABLE 1

Sugar quantification in purified Sf2a-EPA bioconjugates		
Parameter	Sf2a-EPA2S (Y208, R274 and A519)	Sf2a-EPA3S (K240 and A376)
Yield [mg PS/L FV]	14	90
Sugar/Protein ratio [%]	22	48
Site occupancy [%]	93:7:0	0:43:57
Mono:Di:Tri		

[0272] Conclusion: EPA3S (Y208, R274 and A519) lead to higher site occupancy, higher sugar:protein ratio and higher sugar yield compared to EPA2S (K240 and A376). The data shows an increase in di-glycosylation as well as tri-glycosylation with EPA3S compared to EPA2S.

Example 6: (FIG. 8)

[0273] SDS-PAGE analysis was carried out on IMAC enriched periplasmic extracts of *E. coli* strain producing two different *Klebsiella pneumoniae* O-antigen polysaccharides (left and right) and expressing PglB and EPA variants with glycosites at positions Y208 and R274 and A519 (lane 1, SEQ ID NO: 7), or with N-terminal glycotag and glycosites at positions Y208 and R274 and A519 (lane 2, SEQ ID NO: 8).

[0274] Conclusion: Addition of an N-terminal glycosite to the 3-site EPA (Y208, R274 and A519) increased modal molecular weight of the conjugate, indicating high site occupancy of this additional site.

Example 7: (FIG. 9)

[0275] Immunoblot analysis was carried out on periplasmic extract of *E. coli* strains producing Sf2a-, Sp33F, PaO6 or PaO11 antigen polysaccharide and expressing PglB and EPA variants with 1 glycosite introduced at the following positions: Y208 (lane 1), D218 (lane 2), R274 (lane 3), R279 (lane 4), S318 (lane 5), G323 (lane 6), A376 (lane 7), A519 (lane 8), and G525 (lane 9). In FIG. 9, the panel with Sf2a represents the immunoblot probed with anti-EPA antibody,

while the other three panels represent the immunoblot probed with anti-His antibody. The bands correspond to the unglycosylated EPA carrier, and to O-antigen-EPA bioconjugates with one occupied glycosite.

[0276] Conclusion:

[0277] This comparative data demonstrated that besides successful glycosylation at the positions Y208, R274, S318 and A519, it is possible to introduce glycosites at additional positions within the sequence range 198-218 (example position D218), 264-284 (example position R279), 308-328 (example position G323) and 509-529 (example position G525). Glycosylation at these additional positions is in most of the cases efficient, but not superior than at the primarily chosen position within the specified range. When comparing

glycosylation at the positions D218, R279, G323 and G525 with the position A376, it can be concluded that there is dependence on the type of antigen polysaccharide. Overall, it seems that the additional positions D218, R279, G323 and G525 are not always superior than A376, but in many cases they show equal glycosylation level. These additional sites for glycosylation offer further possibilities for combination with other glycosylation sites, which can lead to higher total sugar to protein ratio i.e. higher glycan yield and ultimately reduce the amount of bioconjugate that needs to be used for immunizations. Having more positions suitable for glycosylation also increases flexibility in bioconjugate design and allows to select the best combination for each antigen polysaccharide.

SEQUENCE LISTINGS

SEQ ID NO: 1 EPA sequence from *Pseudomonas aeruginosa*
 AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHNLNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLLDPLDGVVNYLAQQRCLDDTWE
 GKIYRVLAGNPAKHDLIDKPTVISHRLHFPPEGGLAALTAHQACHLPLEAFTTRHRQPRGWEQLEQCGYPVQRLV
 ALYLAARLSWNQVDQVIRNALASPGSGDGLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAASADWS
 LTCPVAAGCAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNWTVRLLQAHRRQLEERGVYFVGYHGT
 FLEAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVPRWSLPGFYRTGL
 TLAAPAEAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPIAIPDPRNVGGDLDPSSIPDKEQA
 ISALPDYASQPGKPPREDLK

SEQ ID NO: 2 Consensus sequence (artificial sequence)
 D/E-X-N-Z-S/T

SEQ ID NO: 3 Consensus sequence (artificial sequence)
 K-D/E-X-N-Z-S/T-K

SEQ ID NO: 4 Consensus sequence (artificial sequence)
 K-D-Q-N-A-T-K

SEQ ID NO: 5 Consensus sequence (artificial sequence)
 J-D/E-X-N-Z-S/T-U

SEQ ID NO: 6 Modified EPA sequence with consensus sequences inserted at Y208 + R274
 (artificial sequence)
 AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHNLNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLLDPLDGVVNYKDNATKLAQQRCLN
 LDDTWEKGKIYRVLAGNPAKHDLIDKPTVISHRLHFPPEGGLAALTAHQACHLPLEAFTKDQATKHRQPRGWEQ
 LEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDGLGEAIREQPEQARLALTAAAESERFVRQGTGN
 DEAGAASADVVSLTCTPVAAGCAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNWTVRLLQAHRRQLE
 ERGVYFVGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVP
 RWSLPGFYRTGLTLAAPEAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPIAIPDPRNVGGD
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 7 Modified EPA sequence with consensus sequences inserted at Y208 + R274 + A519
 (artificial sequence)
 AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHNLNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLLDPLDGVVNYKDNATKLAQQRCLN
 LDDTWEKGKIYRVLAGNPAKHDLIDKPTVISHRLHFPPEGGLAALTAHQACHLPLEAFTKDQATKHRQPRGWEQ
 LEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDGLGEAIREQPEQARLALTAAAESERFVRQGTGN
 DEAGAASADVVSLTCTPVAAGCAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNWTVRLLQAHRRQLE
 ERGVYFVGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGIRNGALLRVYVP
 RWSLPGFYRTGLTKDQATKAPEAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPIAIPDPRNVGGD
 RNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 8 Modified EPA sequence with consensus sequences inserted at N-terminal +
 Y208 + R274 + A519 (artificial sequence)
 LKLAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHNLNAGNQLSHMS
 PIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLLDPLDGV
 VYKDNATKLAQQRCLDDTWEKGKIYRVLAGNPAKHDLIDKPTVISHRLHFPPEGGLAALTAHQACHLPLEAFT
 TKDQATKHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDGLGEAIREQPEQARLA
 LTAAAESERFVRQGTGNDEAGAASADWSLTCTPVAAGCAGPADSGDALLERNYPTGAFLDGGDVSFSTRGT
 QNWTVRLLQAHRRQLEERGVYFVGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQ
 PDARGIRNGALLRVYVPRWSLPGFYRTGLTKDQATKAPEAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPIAIPDPRNVGGD
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

-continued

SEQUENCE LISTINGS

SEQ ID NO: 9 Modified EPA sequence with consensus sequences inserted at N-terminal + Y208 + R274 + A519 + C-terminal (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NKDQNTATKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTK
DQNATKHKRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALT
LAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLDGGDVFSFSTRGTQ
NWTVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPD
ARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGW
PLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLGSGGGDQNTATSGGG

SEQ ID NO: 10 Modified EPA sequence with consensus sequences inserted at N-terminal + Y208 + S318 + A519 + C-terminal (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NKDQNTATKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTK
HRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALAKDQNTATKSPGSGGDLGEAIREQPEQARLALT
TLAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLDGGDVFSFSTRGT
QNTWVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEP
DARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILG
WPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLGSGGGDQNTATSGGG

SEQ ID NO: 11 Modified EPA sequence with consensus sequences inserted at N-terminal+ R274 + S318 + A519 + C-terminal (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NYLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTKDQNTATK
HRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALAKDQNTATKSPGSGGDLGEAIREQPEQARLALT
TLAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLDGGDVFSFSTRGT
QNTWVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEP
DARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILG
WPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLGSGGGDQNTATSGGG

SEQ ID NO: 12 Modified EPA sequence with consensus sequences inserted at N-terminal + Y208 + R274 + S318 + A519 + C-terminal (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NKDQNTATKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTK
DQNATKHKRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALAKDQNTATKSPGSGGDLGEAIREQ
EQARLALTTLAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFLDGGDV
FSFSTRGTQNTWVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGY
AQDQEPDARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAITGPEEEGG
RVITILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLGSGGGDQNTATSGGG

SEQ ID NO: 13 Modified EPA sequence with consensus sequences inserted at N-terminal + Y208 + K240 + R274 + S318 + A519 (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NKDQNTATKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKDNNSPTPTVISHRLHFPEGGSLAALTAHQACHL
PLEAFTKDQNTATKHKRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALAKDQNTATKSPGSGGDLG
EAIREQPEQARLALTTLAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFL
DGGDVFSFSTRGTQNTWVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGD
PALAYGYAQDQEPDARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAIT
GPEEEGGRVTILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 14 Modified EPA sequence with consensus sequences inserted at N-terminal + Y208 + K240 + R274 + S318 + A519+C-terminal (artificial sequence)
GSGGGDQNTATSGGGGAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHSMVLEGGNDALK
LAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIY
TIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPREKRWSEWASGKVLCLLDPLDGVY
NKDQNTATKLAQQRCLDDTWEGKIYRVLAGNPAKHDLDIKDNNSPTPTVISHRLHFPEGGSLAALTAHQACHL
PLEAFTKDQNTATKHKRQPRGWEQLEQCGYVPVQRLVALYLAARLSWNQVDQVIRNALAKDQNTATKSPGSGGDLG
EAIREQPEQARLALTTLAAAESERFVRQGTGNDEAGAASADVSLTCPVAAGECAGPADSGDALLERNYPTGAEFL
DGGDVFSFSTRGTQNTWVERLLQAHRLQLEERGYPVFGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGD
PALAYGYAQDQEPDARGRIRNGALLRVVVPWWSLPGFYRTGLTLKDQNTATKAPEAAGEVERLIGHPLPLRLDAIT
GPEEEGGRVTILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLGSGGGDQ
NATSGGG

-continued

SEQUENCE LISTINGS

SEQ ID NO: 15 *E. coli* flagellin (FlgI) signal sequence
MIKFLSALILLVTTAAQA

SEQ ID NO: 16 *E. coli* outer membrane porin A (OmpA) signal sequence
MKKTAIAIAVALAGPATVAQA

SEQ ID NO: 17 *E. coli* maltose binding protein (MalE) signal sequence
MKIKTGARILALSALTMMFSASALA

SEQ ID NO: 18 *Erwinia carotovora* pectate lyase (PelB) signal sequence
MKYLLPTAAAGLLLLAAQPAMA

SEQ ID NO: 19 heat labile *E. coli* enterotoxin LTIIb signal sequence
MSFKKIIKAFVIMAALVSVQAHA

SEQ ID NO: 20 *Bacillus subtilis* endoxylanase XynA signal sequence
MPKFKKKFLVGLTAAPFMSIMFSATASA

SEQ ID NO: 21 *E. coli* DsbA signal sequence
MKKIWLALAGLVLAFSASA

SEQ ID NO: 22 *E. coli* TolB signal sequence
MKQALRVAFGLILWASVLHA

SEQ ID NO: 23 *Streptococcus agalactiae* SipA signal sequence
MKMNKKVLLTSTMAASLLSVASVQAS

SEQ ID NO: 24 pglB from *Campylobacter jejuni*
MLKKEYLKNPVLVLFAMIILAYVFSVFCRFYVWVWASEFNEYFFNNQLMIISNDGYAFAEGARDMIAGFHQPND
LSYYGSSLSALTYWLYKITPPFSFESIILYMSTFLSSLVVIPTILLANEYKRPLMGFVAALLASIANSYNRTMSGYYD
TDMLVIVLPMFILFFMVRMILKKDFSLIALPLFIGIYLWYWPSSYTLNVALIGLFLIYTLIFHRKEKIFYIAVILSSLT
LSNIAWFYQSAIIVILFALFALEQKRLNFMIIIGILGSATLIFLILSGGVDPIYQLKFYIFRSDSANLTQGFMYFNVN
QTIQEVENVLDSEFMRRISGSEIVFLFSLFGFVWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMALGFGFL
LSEFKAIMVKYSQSLTSNVCIVFATILTLAPVFIHIYNYKAPTVPFSQNEASLLNQLKNIANREDYVVTWWDYGYVP
RYYSDVKTLVDGGKHLGKDNFFPSFALSKEQAAANMARLSVEYTEKSFYAPQNDILKTDILQAMMKDYNQSNV
DLFLASLSKPDFKIDTPKTRDIYLYMPARMSLIFSTVASFSFINLDTGVLDKPTTFSTAYPLDVKNGEIYLSNGVVLS
DDFRSFKIGDNWVSVNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQMFFLGNY
DKNLFDLVINSRDAKVFKLKI

SEQ ID NO: 25 Consensus sequence (artificial)
G-S-G-G-G-D/E-X-N-Z-S/T-G-S-G-G

SEQ ID NO: 26 Modified EPA sequence with consensus sequences inserted at K240 + A376
(artificial sequence)
AEEAFDLWNECAKACVLDLKDGVRRSSRMSVDPAIADTNGQGVLHYSMVLEGGNDALKLAIDNALSITS DGLTIR
LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARDA
TFFVRAHESNEMQPTLAISHAGVSVMMAQAQPRREKRWSEWASGKVLCLDPLDGVYNYLAQQRCNLDDTWE
GKIYRVLAGNPAKHDLDIKDNNNSTPTVISHLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGY
PVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAA
SADVSLTQCPVAKDNRTKGECAAGPADSGDALLERNYPTGAEFLGDGDDVSFSTRGTQNTWVERLLQAHRLQ
ERGVYFVGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVP
RWSLPGFYRTGLTLAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVVIIPSAIPTDPRNVGGD
LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 27 Modified EPA sequence with consensus sequences inserted at N-terminal +
K240 + A376 + C-terminal (artificial sequence)
GSGGGDQNGATGSGGGKLAEAFDLWNECAKACVLDLKDGVRRSSRMSVDPAIADTNGQGVLHYSMVLEGGNDA
LKLAIIDNALSITS DGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMS
PIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVMMAQAQPRREKRWSEWASGKVLCLDPLDGV
VYNYLAQQRCNLDDTWEKGIYRVLAGNPAKHDLDIKDNNNSTPTVISHLHFPEGGSLAALTAHQACHLPLEAF
TRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTAAAE
ESERFVRQGTGNDEAGAA SADVSLTQCPVAKDNRTKGECAAGPADSGDALLERNYPTGAEFLGDGDDVSFSTR
GTQNTWVERLLQAHRLQERGVYFVGYHGTFLAAQSI VFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQ
EPDARGRIRNGALLRVYVPRWVSLPGFYRTGLTLAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAE
RTVVIIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLKLGSGGGDQNGAT

SEQ ID NO: 28 Modified EPA sequence with consensus sequences inserted at Y208 + S318
(artificial sequence)
AEEAFDLWNECAKACVLDLKDGVRRSSRMSVDPAIADTNGQGVLHYSMVLEGGNDALKLAIDNALSITS DGLTIR
LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARDA
TFFVRAHESNEMQPTLAISHAGVSVMMAQAQPRREKRWSEWASGKVLCLDPLDGVYNYKQNAATKLAQQRCN
LDDTWEKGIYRVLAGNPAKHDLDIKPTVISHLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGY
PVQRLVALYLAARLSWNQVDQVIRNALAKDQNAATKPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGN
DEAGAA SADVSLTQCPAAGECAGPADSGDALLERNYPTGAEFLGDGDDVSFSTRGTQNTWVERLLQAHRLQ

- continued

SEQUENCE LISTINGS

ERGVYVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVP
 RWSLPGFYRTGLTLAAPEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRNVGGD
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 29 Modified EPA sequence with consensus sequences inserted at Y208 + A519
 (artificial sequence)

AAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVHLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHKEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLDPLDGVYINKDQATKLAQQRCN
 LDDTWEGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCGY
 PVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAA
 SADVVSILTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNTVERLLQAHRLQLEERGVYV
 VGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPG
 FYRTGLTLKDQATKAPEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRNVGGD
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 30 Modified EPA sequence with consensus sequences inserted at R274 + S318
 (artificial sequence)

AAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVHLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHKEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
 TFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLDPLDGVYNYLAQQRCNLDDTWE
 GKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTKDQATKHRQPRGWEQLEQCGY
 PVQRLVALYLAARLSWNQVDQVIRNALAKDQATKPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGN
 DEAGAAADVVSLTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNTVERLLQAHRLQLE
 ERGVYVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVP
 RWSLPGFYRTGLTLAAPEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRNVGGD
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 31 Modified EPA sequence with consensus sequences inserted at R274 + A519
 (artificial sequence)

AAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVHLHYSMVLEGGNDALKLAIDNALSITSDGLTI
 RLEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHKEKPSNIKVFIHELNAGNQLSHMSPIYTIEMGDELLAKLARD
 ATFFVRAHESNEMQPTLAISHAGVSVVMAQAQPRREKRSEWASGKVLCLDPLDGVYNYLAQQRCNLDDTW
 EGKIYRVLAGNPAKHDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTKDQATKHRQPRGWEQLEQCG
 YPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGA
 ASADVVSILTCPVAAGECAGPADSGDALLERNYPTGAFLDGGDVSFSTRGTQNTVERLLQAHRLQLEERGVYV
 FVGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSL
 PGFYRTGLTLKDQATKAPEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRNVGG
 LDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 32 Forward primer (artificial sequence)
 AAGCTAGCCGCCCGAGGAGCCTTCGACC

SEQ ID NO: 33 Reverse primer (artificial sequence)
 AAGAATCTCAGTGGTGGTGGTGGTGGTTCAGGTCCTCGCGCGCGG

SEQ ID NO: 34 EPA_mut_Y208

AAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVHLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHKEKPSNIKVFIHEL
 NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ
 PRREKRSEWASGKVLCLDPLDGVYINKDQATKLAQQRCNLDDTWEGKIYRVLAGNPAK
 HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCG
 YPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQPEQARLALT
 AAAESERFVRQGTGNDEAGAAADVVVSILTCPVAAGECAGPADSGDALLERNYPTGAFLG
 DGGDVSFSTRGTQNTVERLLQAHRLQLEERGVYVFGYHGTFLAAQSIIVFGGVRARSQDL
 DAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA
 APEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRN
 VGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 36 EPA_mut_S318

AAEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVHLHYSMVLEGGNDALKLAIDNALSITSDGLTIR
 LEGGVEPNKPVRYSYTRQARGSWSLNWLVPIGHKEKPSNIKVFIHEL
 NAGNQLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTLAISHAGVSVVMAQAQ
 PRREKRSEWASGKVLCLDPLDGVYNYLAQQRCNLDDTWEGKIYRVLAGNPAK
 HDLDIKPTVISHRLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWEQLEQCG
 YPVQRLVALYLAARLSWNQVDQVIRNALAKDQATKPGSGGDLGEAIREQPEQARLALT
 AAAESERFVRQGTGNDEAGAAADVVVSILTCPVAAGECAGPADSGDALLERNYPTGAFLG
 DGGDVSFSTRGTQNTVERLLQAHRLQLEERGVYVFGYHGTFLAAQSIIVFGGVRARSQDL
 DAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRWSLPGFYRTGLTLA
 APEAAGEVERLIGHPLPLRLDAITGP EEGGRVTILGWPLAERTVVIPTDPRN
 VGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

-continued

SEQUENCE LISTINGS

SEQ ID NO: 37 EPA_mut_A519

AEEAFDLWNECAKACVLDLKDGVRSRMSVDPADTNGQGVLYHYSMVLEGGNDALKLAI
 DNALSI TSDGLTIRLEGGVEPNKPVRYSYTRQARGSWSLNLVPIGHEKPSNI KVF IHEL
 NAGNQLSHMSPI YTTIEMGDELLAKLARDATFFVRAHESNEMQPTLAI SHAGVSVVMAQAQ
 PRREKRWSEWASGKVLCLLDPLDGVYNYLAQQRCNLDDTWEGKIYRVLAGNPAK
 HDLDIKPTVISIHRHLHFPEGGSLAALTAHQACHLPLEAFTRHRQPRGWLEQCCG
 YPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDLGEAIREQPEQARLALT
 AAAESERFVRQGTGNDEAGAASADVSLTCTPVAAGCAGPADSGDALLERNYPTGAFLG
 DGGDVSFSTRGTQNWTVERLLQAHRLQLEERGVYFVG YHGT FLEAAQSIVFGGVRARSQDL
 DAIWRGFYIAGDPALAYGYAQDQEPDARGIRIRNGALLRVYVPRWSLPGFYRTGLTLKDQN
 ATKAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRVTILGWPLAERTVVIPSAIPTDPRN
 VGGDLDPSSIPDKEQAISALPDYASQPGKPPREDLK

SEQ ID NO: 38 Primer

GAAGCGGGCGCGTGACCATTCTCGGC

SEQ ID NO: 39 Primer

GCCGAGATGGTCACGCGCCCGCCTTC

SEQ ID NO: 40 Nucleotide sequence EPA with mutation Y208 > KDQNTAK

ATGAAAAAGATTGGCTGGCGCTGGCTGGTTAGTTTAGCGTTAGCGCTAGCGCGCGCCGAGGAAGCCTTC
 GACCTCTGGAACGAATGCGCCAAGGCCTGCGTGCTCGACCTCAAGGACGGCGTGCGTTCCAGCCGCATGAG
 CGTCGACCCCGGCCATCGCCGACACCAACGGCCAGGGCGTGCTGCACTACTCCATGGTCTCGAGGGCGGCA
 ACGACGCGCTCAAGCTGGCCATCGACAACGCCCTCAGCATCACCAGCGACGGCTGACCATCCGCTCGAAG
 GTGGCGTCGAGCCGAACAAGCCGGTGCGCTACAGCTACACGCGCCAGGCGCGCGCAGTTGGTCTGTAAC
 TGGCTGGTGCCGATCGGCCACGAGAAGCCTTCGAACATCAAGGTGTTATCCACGAAGTGAACGCCGGTAAC
 CAGCTCAGCCACATCTCGCCGATCTACACCATCGAGATGGGCGACGAGTTGCTGGCGAAGCTGGCGCGCGA
 TGCCACCTTCTTCGTGCGGCGCGACGAGCAACGAGATGCGAGCCGACGCTCGCCATCAGCCATGCCGGGT
 CAGCGTGGTCATGGCCAGGCCAGCGCGCGGAAAAGCGCTGGAGCGAATGGGCCAGCGCAAGGTGT
 TGTGCTGCTCGACCCGCTGACAGGGGTCTACAAACAAAGATCAGAACGCGACCAAACTCGCCAGCAGCGCT
 GCAACCTCGACGATACCTGGGAAGGCAAGATCTACCGGTGCTCGCCGCAACCCGCGCAAGCATGACCTG
 GACATCAAGCCCACGGTCATCAGTCATCGCTGCATTTCCCGAGGGCGGCGAGCTGGCCGCGCTGACCGCG
 CACAGGGCTGCACTCGCGCTGGAGACCTTACCCGCTCATCGCAGCGCGCGGCTGGGAACAACCTGGAG
 CAGTGGGCTATCCGGTGCAGCGCGCTGGTTCGCTCGCCCTTACCTGGCGGCGCGGTGTCGTGGAACCAAGTCA
 CCAGTGATCCGCAACGCCCTGGCCAGCCCGCGAGCGCGGCGACCTGGGCGAAGCGATCCGCGAGCAGC
 CGGAGCAGGCCGCTGCGCCCTGACCTGGCCGCGCGCGGAGAGCGCTTCGTCGCGGAGGGCACCAGG
 AACGACAGGGCCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 GGGCCCGGCGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 GCGGCGAGCTCAGCTTACGACCCGCGCGCACGCAAGTGGACGGTGGAGCGGCTGCTCCAGGCGCACCGC
 CAACTGGAGGAGCGCGGCTATGTGTTTCGTCGGCTACACGCGACCTTCTCGAAGCGGCGCAAGCATCGT
 TTCGGCGGGGTGCGCGCGCGCGAGCCAGGACCTCGACGCGATCTGGCGCGGTTTCTATATCGCCGGCGATCC
 GCGCTGCGCTACGGCTACGCCAGGACAGGAACCCGACGCGCGCGCGCGGATCCGCAACGGTGCCTGCTG
 TCGGGTCTATGTGCGCGCGCGCGGCTGCGCGGCTTCTACCGCACCGGCTGACCTGGCGCGCGCGGAG
 GCGGCGGCGGAGGTGCAACGGCTGATCGGCCATCCGCTGCGCGCTGCGCCTGGACGCCATACCGGCCCGGA
 GGAGGAAGGGCGGGCGGTGACCATTTCTCGGTGGCGGCTGGCGGAGCGCACCGTGGTATTCCTTCGGCGA
 TCCCCACCGACCCGCGCAACCTCGGCGGCGGACCTCGACCCGTCAGCATCCCCGACAAGGAACAGGCGATCA
 CGCCCTTCCGCGACTACGCCAGCCAGCCCGGCAACCGCGCGCGAGGACTTGAAGCACCACCACCACCACC
 ACTGA

SEQ ID NO: 41 Nucleotide sequence EPA with mutation R274 > KDQNTAK

ATGAAAAAGATTGGCTGGCGCTGGCTGGTTAGTTTAGCGTTAGCGCTAGCGCGCGCCGAGGAAGCCTTC
 GACCTCTGGAACGAATGCGCCAAGGCCTGCGTGCTCGACCTCAAGGACGGCGTGCGTTCCAGCCGCATGAG
 CGTCGACCCCGGCCATCGCCGACACCAACGGCCAGGGCGTGCTGCACTACTCCATGGTCTCGAGGGCGGCA
 ACGACGCGCTCAAGCTGGCCATCGACAACGCCCTCAGCATCACCAGCGACGGCTGACCATCCGCTCGAAG
 GCAGCGTCGAGCGGCGGCTGAGCGCGCTGCGCTACAGCTACACGCGCCAGGCGCGCGCAGTTGGTCTGTAAC
 TGGCTGGTACCGATCGGCCACGAGAAGCCTTCGAACATCAAGGTGTTATCCACGAAGTGAACGCCGGTAAC
 CAGCTCAGCCACATCTCGCCGATCTACACCATCGAGATGGGCGACGAGTTGCTGGCGAAGCTGGCGCGCGA
 TGCCACCTTCTTCGTGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 CAGCGTGGTCATGGCTCAGGCCAGCGCGCGGAAAAGCGCTGGAGCGAATGGGCCAGCGGCAAGGTGT
 TGTGCTGCTCGACCCGCTGACAGGGGTCTACAACTACCTCGCCAGCAGCGCTGCAACCTCGAGATACCT
 GGGAAAGCGAAGATCTACCGGTGCTCGCCGCAACCCGCGCAAGCATGACCTGGACATCAAGCCACGGT
 ATCAGTCATCGCTGCATTTCCCGAGGGCGGCGGCTGGCGCGCTGACCGCGCACAGGCGCTGCCACCTG
 CCGCTGGAGGCTTCACTAAAGATCAGAACGCGACCAACATCGCAGCGCGCGGCTGGGAACAACCTGGAG
 CAGTGGGCTATCCGGTGCAGCGCTGGTTCGCTTACCTGGCGGCGGCGACTGTCGTGGAACCAAGTCA
 CAGGTGATCCGCAACCGCTTGGCCAGCCCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 CGGAGCAGGCCGCTGCGCCCTGACCTGGCCGCGCGCGGAGAGCGAGCGCTTCGTCGCGGAGGGCACCAGG
 AACGACAGGGCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 GGGCCCGGCGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
 GCGGCGAGCTCAGCTTACGACCCGCGGCGACGCAAGTGGACGGTGGAGCGGCTGCTCCAGGCGCACCGC
 CAACTGGAGGAGCGGCTATGTGTTTCGTCGGCTACACGCGACCTTCTCGAAGCGGCGCAAGCATCGT
 TTCGGCGGGGTGCGCGCGCGGAGGACCTCGACGCGATCTGGCGCGGTTTCTATATCGCCGGCGATCC
 GCGCTGCGCTACGGCTACGCCAGGACAGGAACCCGACGCGCGCGCGGATCCGCAACGGTGCCTGCTG
 TCGGGTCTATGTGCGCGCTGAGTCTGCGGGCTTCTACCGCACCGGCTGACCTGGCGCGCGCGGAG

- continued

SEQUENCE LISTINGS

GCGGCGGGCGAGGTCGAACGGCTGATCGGCCATCCGCTGCCGCTGCGCCTGGACGCCATCACCGGCCCGA
 GGAGGAAGGGGGGCGGTGACCATTTCTCGGTGGCCGCTGGCCGAGCGCACCGTGGTGATTCCCTCGGCGA
 TCCCCACGACCCGCGCAACGTGCGCGGCGACCTCGACCCGTCCAGCATCCCCGACAAGGAACAGGCGATCA
 GCGCCCTGCCGACTACGCCAGCCAGCCCGGCAACC GCCGCGCGAGGACTTGAAGCACCACCACCACC
 ACTGA

SEQ ID NO: 42 Nucleotide sequence EPA with mutation S318 > KDQNA TK
 ATGAAAAAGATTGGCTGGCGCTGGCTGGTTTAGTTT TAGCGTTTAGCGCTAGCGCCGCCGAGGAAGCCTTC
 GACCTCTGGAACGAATGCGCCAAGGCCCTGCGTGCTCGACCTCAAGGACGGCGTGCGTTCCAGCCGCATGAG
 CGTCGACCCGGCCATCGCCGACACCAACGGCCAGGGCGTGCTGCACTACTCCATGGTCTCTGGAGGGCGGCA
 ACGACGCGCTCAAGCTGGCCATCGACAACGCCCTCAGCATCACAGCGACGGCTGACCATCCGCTCGAAG
 GCGGCGTCGAGCCGAACAAGCCGGTGGCTACAGCTACACGCCAGCGCGCGGCGAGTTGGTCGCTGAAC
 TGGCTGGTACCGATCGGCCACGAGAGCCCTCGAACATCAAGGTGTTATCCACGAACCTGAACGCCGGTAAC
 CAGCTCAGCCACATGTGCGCGATCTACACCATCGAGATGGGCGACGAGTTGCTGGCGAAGCTGGCGCGCA
 TGCCACCTTCTTCGTAGGGCGCACGAGAGCAACGAGATGCGAGCGACGCTCGCATCAGCCATGCCGGGT
 CAGCGTGGTCATGGCTCAGGCCAGCCGCGCCGGAAAAGCGCTGGAGCGAATGGGCGAGCGCAAGGTGT
 TGTGCTGCTCGACCCGCTGACGGGGTCTACAACTACTCGCCAGCAGCGCTGCAACCTCGACGATACCT
 GGGAAGGCAAGATCTACCGGTGCTCGCCGGCAACCCGGCGAAGCATGACCTGGACATCAAGCCACGGTC
 ATCAGTCATCGCTGCATTTCCTCGAGGGCGCAGCCTGGCCGCGCTGACCGCGCACAGGCCTGCCACCTG
 CCGCTGGAGGCCCTTCACTCGTCATCGCCAGCCGCGCGGCTGGGAACAACCTGGAGCAGTGGGCTATCCGT
 GCAGCGGCTGGTGGCCCTTACTGGCGGCGGACTGTCTGGAACCAAGGTGACCAAGGTGATCCGCAACG
 CCCTGGCCAAAGATCAGAACGCGACCAAACCCGCGAGCGCGGCGACCTGGGCGAAGCGATCCGCGAGCAG
 CCGGAGCAGGCCCGTCTGGCCCTGACCCCTGGCCGCGCCGCGAGAGCGAGCGCTTCGTCCGGCAGGGCACC
 CAACGACGAGGCGCGCGCGCGCAGCGCCGAGTGGTGAGCCTGACCTGCCCGGTGCGCGCGGTGAATGCG
 CGGGCCCGGCGGACAGCGGCGGCGGCTGCTGGAGCGCAACTATCCACTGGCGCGGAGTTCTCTCGCGAC
 GGCGGCGACGTCAGCTTACGACCCGCGGCGACGAGAACCTGGAGCGGTGGAGCGGCTGCTCAGGCGCACCG
 CCAACTGGAGGAGCGCGCTATGTGTTCTGCTGGCTACACGCGCACTTCCTCGAAGCGGCGCAAGCATCGT
 CTTCGGCGGGTGCGCGCGCGCAGCCAGGACCTCGACGCGATCTGGCGCGGTTCTATATCGCCGCGATC
 CGGCGCTGGCTACGGCTACGCCCAGGACCGAGAACCCGACGCGCGCGCGGATCCGCAACCGTGCCCTG
 CTGCGGGTCTATGTGCGCGCTGGAGTCTGCGGGCTTCTACCGCACCGGCTGACCTGGCGCGCGCGGA
 GGCGGCGGGCGAGGTGGAACCGCTGATCGGCCATCCGCTGCGCTGCGCCTGGAGCGCATCAACGGCCCG
 AGGAGGAAGGCGGCGCGTGACCATTCCTCGGCTGGCGCTGGCCGAGCGCACCGTGGTGATTCCCTCGGCG
 ATCCCCACCGACCCGCGCAACGTGCGGCGCGACCTCGACCCGTCCAGCATCCCCGACAAGGAACAGGCGATC
 AGCGCCCTGCCGACTACGCCAGCCAGCCCGGCAAAACGCCGCGCGAGGACTTGAAGCACCACCACCACC
 CACTGA

SEQ ID NO: 43 Nucleotide sequence EPA with mutation A519 > KDQNA TK
 ATGAAAAAGATTGGCTGGCGCTGGCTGGTTTAGTTT TAGCGTTTAGCGCTAGCGCCGCCGAGGAAGCCTT-
 GACCTCTGGAACGAATGCGCCAAGGCCCTGCGTGCTCGACCTCAAGGACGGCGTGCGTTCCAGCCGCATGAG
 CGTCGACCCGGCCATCGCCGACACCAACGGCCAGGGCGTGCTGCACTACTCCATGGTCTCTGGAGGGCGGCA
 ACGACGCGCTCAAGCTGGCCATCGACAACGCCCTCAGCATCACAGCGACGGCTGACCATCCGCTCGAAG
 GCGGCGTCGAGCCGAACAAGCCGGTGGCTACAGCTACACGCCAGCGCGCGGCGAGTTGGTCGCTGAAC
 TGGCTGGTACCGATCGGCCACGAGAGCCCTCGAACATCAAGGTGTTATCCACGAACCTGAACGCCGGTAAC
 CAGCTCAGCCACATGTGCGCGATCTACACCATCGAGATGGGCGACGAGTTGCTGGCGAAGCTGGCGCGCA
 TGCCACCTTCTTCGTAGGGCGCACGAGAGCAACGAGATGCGACCGACGCTCGCATCAGCCATGCCGGGT
 CAGCGTGGTCATGGCTCAGGCCAGCCGCGCCGGAAAAGCGCTGGAGCGAATGGGCGAGCGCAAGGTGT
 TGTGCTGCTCGACCCGCTGGACGGGGTCTACAACTACTCGCCAGCAGCGCTGCAACCTCGAGATACCT
 GGGAAGGCAAGATCTACCGGTGCTCGCCGGCAACCCGGCGAAGCATGACCTGGACATCAAGCCACGGTC
 ATCAGTCATCGCTGCATTTCCTCGAGGGCGGCGAGCCTGGCCGCGCTGACCGCGCACAGGCCTGCCACCTG
 CCGCTGGAGGCCCTTCACTCGTCATCGCCAGCCGCGCGGCTGGGAACAACCTGGAGCAGTGGGCTATCCGT
 GCAGCGGCTGGTGGCCCTTACTGGCGGCGGACTGTCTGGAACCAAGGTGACCAAGGTGATCCGCAACG
 CCCTGGCCAGCCCGGCGAGCGGCGGCGACCTGGGCGAAGCGATCCGCGAGCAGCCGAGCAGGCCGCTCTG
 GCCCTGACCCCTGGCGCGCGCGAGAGCGAGCGCTTCGTCCGGCAGGGCACCAGCAACGACAGGCGCGCG
 GGCGACGCGCGAGCTGGTGAGCCTGACCTGCCCGGTGCGCGCGGTGAATGCGCGGGCCCGGCGGACAGCG
 GCGAGCCCTGCTGGAGCGCAACTATCCACTGGCGCGGAGTTCTCGCGACGCGCGCGACGTGACGTTT
 AGCACCCGCGGCGACGAGAATGGACGTTGGAGCGGCTGCTCCAGGCGCACCGCAACTGGAGGAGCGCGG
 CTATGTGTTCTGCGCTACACGCGACCTTCCTCGAAGCGCGCAAGCATCGTCTTCGGCGGGTGCGCGC
 GCGAGCCAGGACCTCGACGCGATCTGGCGCGGTTCTATATCGCCGCGATCCGGCGCTGGCTACGGCT
 ACGCCAGGACCTGAGAGCGCAACGTGCGGCGGCGGATCCGCAACGGTCCCTGCTGCGGGTCTATGTGCG
 CGCTGGAGTCTGCGGGCTTCTACCGCACCGGCTGACCTGAAAGATCAGAACGCGACCAAGCGCCGGA
 GGCGGCGGGCGAGGTGCAACGGCTGATCGGCCATCCGCTGCGCTGCGCCTGGACGCCATCAACGGCCCG
 AGGAGGAAGGCGGCGCGTGACCATTCCTCGGCTGGCGCTGGCCGAGCGCACCGTGGTGATTCCCTCGGCG
 ATCCCCACCGACCCGCGCAACGTGCGGCGGCGACCTCGACCCGTCCAGCATCCCCGACAAGGAACAGGCGATC
 AGCGCCCTGCCGACTACGCCAGCCAGCCCGGCAAAACGCCGCGCGAGGACTTGAAGCACCACCACCACC
 CACTGA

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 43

<210> SEQ ID NO 1

<211> LENGTH: 612

<212> TYPE: PRT

<213> ORGANISM: *Pseudomonas aeruginosa*

<400> SEQUENCE: 1

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1 5 10 15

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
20 25 30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
35 40 45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
50 55 60

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

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85 90 95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100 105 110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
115 120 125

Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
130 135 140

Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
145 150 155 160

Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
165 170 175

Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
180 185 190

Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr
195 200 205

Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile
210 215 220

Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys
225 230 235 240

Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu
245 250 255

Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Ala Phe
260 265 270

Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys Gly
275 280 285

Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu Ser
290 295 300

Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Ser Pro Gly
305 310 315 320

Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala
325 330 335

Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg
340 345 350

-continued

Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val
 355 360 365
 Ser Leu Thr Cys Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp
 370 375 380
 Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe
 385 390 395 400
 Leu Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn
 405 410 415
 Trp Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg
 420 425 430
 Gly Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln
 435 440 445
 Ser Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala
 450 455 460
 Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly
 465 470 475 480
 Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly
 485 490 495
 Ala Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr
 500 505 510
 Arg Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu
 515 520 525
 Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly
 530 535 540
 Pro Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala
 545 550 555 560
 Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn
 565 570 575
 Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala
 580 585 590
 Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg
 595 600 605
 Glu Asp Leu Lys
 610

<210> SEQ ID NO 2
 <211> LENGTH: 5
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Consensus sequence
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (1)..(1)
 <223> OTHER INFORMATION: Xaa can be Asp or Glu
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (2)..(2)
 <223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
 except proline
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (4)..(4)
 <223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
 except proline
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (5)..(5)
 <223> OTHER INFORMATION: Xaa can be Ser or Thr

-continued

<400> SEQUENCE: 2

Xaa Xaa Asn Xaa Xaa
1 5

<210> SEQ ID NO 3

<211> LENGTH: 7

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Consensus sequence

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (2)..(2)

<223> OTHER INFORMATION: Xaa can be Asp or Glu

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (3)..(3)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (5)..(5)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (6)..(6)

<223> OTHER INFORMATION: Xaa can be Ser or Thr

<400> SEQUENCE: 3

Lys Xaa Xaa Asn Xaa Xaa Lys
1 5

<210> SEQ ID NO 4

<211> LENGTH: 7

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Consensus sequence

<400> SEQUENCE: 4

Lys Asp Gln Asn Ala Thr Lys
1 5

<210> SEQ ID NO 5

<211> LENGTH: 7

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Consensus sequence

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (1)..(1)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (2)..(2)

<223> OTHER INFORMATION: Xaa can be Asp or Glu

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (3)..(3)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline

<220> FEATURE:

<221> NAME/KEY: MISC_FEATURE

<222> LOCATION: (5)..(5)

<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline

<220> FEATURE:

-continued

<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: Xaa can be Ser or Thr
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid

<400> SEQUENCE: 5

Xaa Xaa Xaa Asn Xaa Xaa Xaa
1 5

<210> SEQ ID NO 6
<211> LENGTH: 624
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
inserted at Y208+R274

<400> SEQUENCE: 6

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1 5 10 15

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
20 25 30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
35 40 45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
50 55 60

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

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85 90 95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100 105 110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
115 120 125

Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
130 135 140

Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
145 150 155 160

Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
165 170 175

Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
180 185 190

Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Lys
195 200 205

Asp Gln Asn Ala Thr Lys Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp
210 215 220

Thr Trp Glu Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys
225 230 235 240

His Asp Leu Asp Ile Lys Pro Thr Val Ile Ser His Arg Leu His Phe
245 250 255

Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln Ala Cys His
260 265 270

Leu Pro Leu Glu Ala Phe Thr Lys Asp Gln Asn Ala Thr Lys His Arg

-continued

275					280					285					
Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln
290						295					300				
Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val
305					310					315					320
Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp
				325					330					335	
Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu
			340					345					350		
Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly
			355				360					365			
Asn	Asp	Glu	Ala	Gly	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	
	370					375					380				
Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala
385						390					395				400
Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly
				405				410						415	
Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu
			420					425					430		
Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe
			435				440					445			
Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe
	450					455					460				
Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly
465						470					475				480
Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp
				485				490						495	
Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg
			500					505					510		
Val	Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu
			515				520					525			
Thr	Leu	Ala	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu	Ile	Gly
	530					535					540				
His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu	Glu	Glu
545						550					555				560
Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg	Thr	Val
				565				570						575	
Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly	Gly	Asp
			580					585					590		
Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser	Ala	Leu
	595						600					605			
Pro	Asp	Tyr	Ala	Ser	Gln	Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp	Leu	Lys
	610					615					620				

<210> SEQ ID NO 7

<211> LENGTH: 630

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
inserted at Y208+R274+A519

<400> SEQUENCE: 7

-continued

Ala	Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val	1	5	10	15
Leu	Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro	20	25	30	
Ala	Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val	35	40	45	
Leu	Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu	50	55	60	
Ser	Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu	65	70	75	80
Pro	Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser	85	90	95	
Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn	100	105	110	
Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His	115	120	125	
Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys	130	135	140	
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu	145	150	155	160
Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met	165	170	175	
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser	180	185	190	
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys	195	200	205	
Asp	Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	210	215	220	
Thr	Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	225	230	235	240
His	Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	245	250	255	
Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	260	265	270	
Leu	Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn	Ala	Thr	Lys	His	Arg	275	280	285	
Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	290	295	300	
Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	305	310	315	320
Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp	325	330	335	
Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	340	345	350	
Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	355	360	365	
Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	370	375	380	
Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	385	390	395	400
Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly				

-continued

405	410	415
Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp Thr Val Glu		
420	425	430
Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe		
435	440	445
Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val Phe		
450	455	460
Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly		
465	470	475
Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp		
485	490	495
Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg		
500	505	510
Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu		
515	520	525
Thr Leu Lys Asp Gln Asn Ala Thr Lys Ala Pro Glu Ala Ala Gly Glu		
530	535	540
Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile		
545	550	555
Thr Gly Pro Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro		
565	570	575
Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro		
580	585	590
Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu		
595	600	605
Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro		
610	615	620
Pro Arg Glu Asp Leu Lys		
625	630	

<210> SEQ ID NO 8
 <211> LENGTH: 647
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+Y208+R274+A519

<400> SEQUENCE: 8

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Gly Lys		
1	5	10
Leu Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys		
20	25	30
Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp		
35	40	45
Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met		
50	55	60
Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala		
65	70	75
Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val		
85	90	95
Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly		
100	105	110

-continued

Ser	Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser
	115						120					125			
Asn	Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser
130					135						140				
His	Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala
145				150						155					160
Lys	Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn
			165						170					175	
Glu	Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val
			180					185					190		
Met	Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala
		195					200					205			
Ser	Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn
210					215						220				
Lys	Asp	Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp
225					230					235					240
Asp	Thr	Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala
			245						250					255	
Lys	His	Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His
		260						265					270		
Phe	Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys
		275					280						285		
His	Leu	Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn	Ala	Thr	Lys	His
290						295					300				
Arg	Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val
305					310					315					320
Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln
			325						330					335	
Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly
		340						345					350		
Asp	Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala
	355						360					365			
Leu	Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr
370						375					380				
Gly	Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr
385					390					395					400
Cys	Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp
			405						410					415	
Ala	Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp
			420					425					430		
Gly	Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val
		435					440					445			
Glu	Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val
450					455						460				
Phe	Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val
465					470					475					480
Phe	Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg
			485					490						495	
Gly	Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln
			500					505					510		
Asp	Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu

-continued

515	520	525
Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg Thr Gly		
530	535	540
Leu Thr Leu Lys Asp Gln Asn Ala Thr Lys Ala Pro Glu Ala Ala Gly		
545	550	555
Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala		
565	570	575
Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp		
580	585	590
Pro Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp		
595	600	605
Pro Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys		
610	615	620
Glu Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys		
625	630	635
Pro Pro Arg Glu Asp Leu Lys		
645		

<210> SEQ ID NO 9
 <211> LENGTH: 658
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+Y208+R274+A519+C-terminal

<400> SEQUENCE: 9

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Gly Ala		
1	5	10
Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val Leu		
20	25	30
Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro Ala		
35	40	45
Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val Leu		
50	55	60
Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu Ser		
65	70	75
Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu Pro		
85	90	95
Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser Trp		
100	105	110
Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn Ile		
115	120	125
Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His Met		
130	135	140
Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys Leu		
145	150	155
Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu Met		
165	170	175
Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met Ala		
180	185	190
Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser Gly		
195	200	205

-continued

Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys	Asp
210						215					220				
Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr
225					230					235					240
Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His
			245						250					255	
Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	Pro
			260					265					270		
Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	Leu
		275					280					285			
Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn	Ala	Thr	Lys	His	Arg	Gln
290						295					300				
Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg
305					310					315					320
Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp
			325						330					335	
Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu
			340					345					350		
Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr
		355					360					365			
Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn
370						375					380				
Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro
385					390					395					400
Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	Leu
			405						410					415	
Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	Gly
			420				425						430		
Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	Arg
		435					440					445			
Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	Val
450						455					460				
Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	Gly
465					470					475					480
Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	Phe
			485					490						495	
Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	Gln
			500				505						510		
Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	Val
		515					520					525			
Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	Thr
530						535					540				
Leu	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val
545					550					555					560
Glu	Arg	Leu	Ile	Gly	His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr
			565						570					575	
Gly	Pro	Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	
		580					585					590			
Ala	Glu	Arg	Thr	Val	Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg
		595					600					605			
Asn	Val	Gly	Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln

-continued

610	615	620	
Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro			
625	630	635	640
Arg Glu Asp Leu Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser			
	645	650	655
Gly Gly			

<210> SEQ ID NO 10
 <211> LENGTH: 659
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+Y208+S318+A519+C-terminal

<400> SEQUENCE: 10

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Gly Ala			
1	5	10	15
Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val Leu			
	20	25	30
Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro Ala			
	35	40	45
Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val Leu			
	50	55	60
Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu Ser			
	65	70	75
Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu Pro			
	85	90	95
Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser Trp			
	100	105	110
Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn Ile			
	115	120	125
Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His Met			
	130	135	140
Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys Leu			
	145	150	155
Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu Met			
	165	170	175
Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met Ala			
	180	185	190
Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser Gly			
	195	200	205
Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Lys Asp			
	210	215	220
Gln Asn Ala Thr Lys Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr			
	225	230	235
Trp Glu Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His			
	245	250	255
Asp Leu Asp Ile Lys Pro Thr Val Ile Ser His Arg Leu His Phe Pro			
	260	265	270
Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln Ala Cys His Leu			
	275	280	285
Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln			

-continued

290					295					300						
Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	
305					310					315				320		
Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	
			325						330					335		
Leu	Ala	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ser	Pro	Gly	Ser	Gly	Gly	Asp	
			340					345					350			
Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	
		355					360					365				
Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	
	370					375					380					
Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	
385						390					395				400	
Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	
			405						410					415		
Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	
			420					425					430			
Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	
	435						440					445				
Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	
	450					455					460					
Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	
465						470					475				480	
Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	
			485					490						495		
Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	
			500				505						510			
Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	
		515					520					525				
Val	Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	
	530					535					540					
Thr	Leu	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ala	Pro	Glu	Ala	Ala	Gly	Glu	
545						550					555				560	
Val	Glu	Arg	Leu	Ile	Gly	His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	
			565					570						575		
Thr	Gly	Pro	Glu	Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	
		580					585					590				
Leu	Ala	Glu	Arg	Thr	Val	Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	
		595					600					605				
Arg	Asn	Val	Gly	Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	
	610					615					620					
Gln	Ala	Ile	Ser	Ala	Leu	Pro	Asp	Tyr	Ala	Ser	Gln	Pro	Gly	Lys	Pro	
625						630					635				640	
Pro	Arg	Glu	Asp	Leu	Gly	Ser	Gly	Gly	Gly	Asp	Gln	Asn	Ala	Thr	Gly	
			645					650						655		

Ser Gly Gly

<210> SEQ ID NO 11

<211> LENGTH: 659

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

-continued

<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
inserted at N-terminal+R274+S318+A519+C-terminal

<400> SEQUENCE: 11

Gly	Ser	Gly	Gly	Gly	Asp	Gln	Asn	Ala	Thr	Gly	Ser	Gly	Gly	Gly	Ala	1	5	10	15
Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val	Leu	20	25	30	
Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro	Ala	35	40	45	
Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val	Leu	50	55	60	
Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu	Ser	65	70	75	80
Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu	Pro	85	90	95	
Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser	Trp	100	105	110	
Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn	Ile	115	120	125	
Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His	Met	130	135	140	
Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys	Leu	145	150	155	160
Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu	Met	165	170	175	
Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met	Ala	180	185	190	
Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser	Gly	195	200	205	
Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Tyr	Leu	210	215	220	
Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	Trp	Glu	Gly	Lys	Ile	Tyr	225	230	235	240
Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	Asp	Leu	Asp	Ile	Lys	Pro	245	250	255	
Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	Pro	Glu	Gly	Gly	Ser	Leu	Ala	260	265	270	
Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	Leu	Pro	Leu	Glu	Ala	Phe	Thr	275	280	285	
Lys	Asp	Gln	Asn	Ala	Thr	Lys	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu	Gln	290	295	300	
Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	305	310	315	320
Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	325	330	335	
Leu	Ala	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ser	Pro	Gly	Ser	Gly	Gly	Asp	340	345	350	
Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	355	360	365	
Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	370	375	380	

-continued

Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys
 385 390 395 400
 Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala
 405 410 415
 Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly
 420 425 430
 Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp Thr Val Glu
 435 440 445
 Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe
 450 455 460
 Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val Phe
 465 470 475 480
 Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly
 485 490 495
 Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp
 500 505 510
 Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg
 515 520 525
 Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu
 530 535 540
 Thr Leu Lys Asp Gln Asn Ala Thr Lys Ala Pro Glu Ala Ala Gly Glu
 545 550 555 560
 Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile
 565 570 575
 Thr Gly Pro Glu Glu Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro
 580 585 590
 Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro
 595 600 605
 Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu
 610 615 620
 Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro
 625 630 635 640
 Pro Arg Glu Asp Leu Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly
 645 650 655
 Ser Gly Gly

<210> SEQ ID NO 12

<211> LENGTH: 665

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+Y208+R274+S318+A519+C-terminal

<400> SEQUENCE: 12

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Ala
 1 5 10 15
 Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val Leu
 20 25 30
 Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro Ala
 35 40 45
 Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val Leu
 50 55 60

-continued

Glu 65	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu	Ser	
					70					75					80	
Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu	Pro	
				85					90					95		
Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser	Trp	
			100					105					110			
Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn	Ile	
	115						120					125				
Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His	Met	
	130					135					140					
Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys	Leu	
	145				150					155					160	
Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu	Met	
			165					170						175		
Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met	Ala	
			180					185					190			
Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser	Gly	
	195					200						205				
Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys	Asp	
	210					215					220					
Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	
	225				230					235					240	
Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	
			245					250						255		
Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	Pro	
	260							265					270			
Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	Leu	
	275						280					285				
Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn	Ala	Thr	Lys	His	Arg	Gln	
	290					295					300					
Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	
	305				310					315				320		
Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	
			325					330						335		
Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ser	
		340						345					350			
Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	
	355					360						365				
Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	
	370					375					380					
Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	
	385				390					395				400		
Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	
			405					410						415		
Ala	Asp	Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	
			420					425					430			
Glu	Phe	Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	
	435						440					445				
Gln	Asn	Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	
	450					455					460					

-continued

Glu Arg Gly Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala
 465 470 475 480
 Ala Gln Ser Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu
 485 490 495
 Asp Ala Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala
 500 505 510
 Tyr Gly Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg
 515 520 525
 Asn Gly Ala Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly
 530 535 540
 Phe Tyr Arg Thr Gly Leu Thr Leu Lys Asp Gln Asn Ala Thr Lys Ala
 545 550 555 560
 Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro
 565 570 575
 Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg Val
 580 585 590
 Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile Pro Ser
 595 600 605
 Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp Pro Ser
 610 615 620
 Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala
 625 630 635 640
 Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Gly Ser Gly Gly Gly
 645 650 655
 Asp Gln Asn Ala Thr Gly Ser Gly Gly
 660 665

<210> SEQ ID NO 13

<211> LENGTH: 658

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+Y208+K240+R274+S318+A519

<400> SEQUENCE: 13

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Gly Ala
 1 5 10 15
 Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val Leu
 20 25 30
 Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro Ala
 35 40 45
 Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val Leu
 50 55 60
 Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu Ser
 65 70 75 80
 Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu Pro
 85 90 95
 Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser Trp
 100 105 110
 Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn Ile
 115 120 125
 Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His Met
 130 135 140

-continued

Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys	Leu	145	150	155	160
Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu	Met	165	170	175	
Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met	Ala	180	185	190	
Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser	Gly	195	200	205	
Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys	Asp	210	215	220	
Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	225	230	235	240
Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	245	250	255	
Asp	Leu	Asp	Ile	Lys	Asp	Asn	Asn	Asn	Ser	Thr	Pro	Thr	Val	Ile	Ser	260	265	270	
His	Arg	Leu	His	Phe	Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	275	280	285	
His	Gln	Ala	Cys	His	Leu	Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn	290	295	300	
Ala	Thr	Lys	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	305	310	315	320
Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	325	330	335	
Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Lys	Asp	340	345	350	
Gln	Asn	Ala	Thr	Lys	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	355	360	365	
Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	370	375	380	
Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	385	390	395	400
Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala	405	410	415	
Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg	420	425	430	
Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser	435	440	445	
Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln	450	455	460	
Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	Val	Gly	Tyr	His	465	470	475	480
Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	Gly	Gly	Val	Arg	485	490	495	
Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	Phe	Tyr	Ile	Ala	500	505	510	
Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	Gln	Glu	Pro	Asp	515	520	525	
Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	Val	Tyr	Val	Pro	530	535	540	

-continued

Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	Thr	Leu	Lys	Asp
545					550					555					560
Gln	Asn	Ala	Thr	Lys	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu
				565						570				575	
Ile	Gly	His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu
			580					585					590		
Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg
		595					600					605			
Thr	Val	Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly
	610					615					620				
Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser
625					630					635					640
Ala	Leu	Pro	Asp	Tyr	Ala	Ser	Gln	Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp
			645						650					655	

Leu Lys

<210> SEQ ID NO 14

<211> LENGTH: 671

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences inserted at N-terminal+Y208+K240+R274+S318+A519+C-terminal

<400> SEQUENCE: 14

Gly	Ser	Gly	Gly	Gly	Asp	Gln	Asn	Ala	Thr	Gly	Ser	Gly	Gly	Gly	Ala
1				5					10					15	
Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val	Leu
		20					25					30			
Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro	Ala
		35				40					45				
Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val	Leu
	50				55					60					
Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu	Ser
65			70						75					80	
Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu	Pro
		85					90						95		
Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser	Trp
		100					105						110		
Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn	Ile
	115					120						125			
Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His	Met
	130				135					140					
Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys	Leu
145				150					155					160	
Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu	Met
		165						170						175	
Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met	Ala
		180					185						190		
Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser	Gly
	195					200						205			
Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys	Asp
210					215						220				

-continued

Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr
225					230					235					240
Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His
			245						250					255	
Asp	Leu	Asp	Ile	Lys	Asp	Asn	Asn	Asn	Ser	Thr	Pro	Thr	Val	Ile	Ser
			260					265					270		
His	Arg	Leu	His	Phe	Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala
		275					280					285			
His	Gln	Ala	Cys	His	Leu	Pro	Leu	Glu	Ala	Phe	Thr	Lys	Asp	Gln	Asn
	290					295					300				
Ala	Thr	Lys	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys
305					310					315					320
Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu
				325					330					335	
Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Lys	Asp
			340					345					350		
Gln	Asn	Ala	Thr	Lys	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala
		355					360					365			
Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala
	370					375					380				
Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala
385					390					395					400
Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala
				405					410					415	
Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg
			420					425					430		
Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser
		435					440					445			
Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln
	450					455					460				
Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	Val	Gly	Tyr	His
465					470					475					480
Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	Gly	Gly	Val	Arg
				485					490					495	
Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	Phe	Tyr	Ile	Ala
			500					505					510		
Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	Gln	Glu	Pro	Asp
		515					520					525			
Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	Val	Tyr	Val	Pro
	530					535					540				
Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	Thr	Leu	Lys	Asp
545					550					555					560
Gln	Asn	Ala	Thr	Lys	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu
				565					570					575	
Ile	Gly	His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu
			580					585					590		
Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg
		595					600					605			
Thr	Val	Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly
	610					615					620				
Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser

-continued

625	630	635	640
Ala Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp			
	645	650	655
Leu Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly			
	660	665	670

<210> SEQ ID NO 15
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 15

Met Ile Lys Phe Leu Ser Ala Leu Ile Leu Leu Leu Val Thr Thr Ala
1 5 10 15

Ala Gln Ala

<210> SEQ ID NO 16
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 16

Met Lys Lys Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala
1 5 10 15

Thr Val Ala Gln Ala
20

<210> SEQ ID NO 17
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 17

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala
20 25

<210> SEQ ID NO 18
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: *Erwinia carotovora*s

<400> SEQUENCE: 18

Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1 5 10 15

Ala Gln Pro Ala Met Ala
20

<210> SEQ ID NO 19
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 19

Met Ser Phe Lys Lys Ile Ile Lys Ala Phe Val Ile Met Ala Ala Leu
1 5 10 15

Val Ser Val Gln Ala His Ala
20

-continued

<210> SEQ ID NO 20
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: *Bacillus subtilis*

<400> SEQUENCE: 20

Met Phe Lys Phe Lys Lys Lys Phe Leu Val Gly Leu Thr Ala Ala Phe
1 5 10 15

Met Ser Ile Ser Met Phe Ser Ala Thr Ala Ser Ala
20 25

<210> SEQ ID NO 21
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 21

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser
1 5 10 15

Ala Ser Ala

<210> SEQ ID NO 22
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: *E. coli*

<400> SEQUENCE: 22

Met Lys Gln Ala Leu Arg Val Ala Phe Gly Phe Leu Ile Leu Trp Ala
1 5 10 15

Ser Val Leu His Ala
20

<210> SEQ ID NO 23
<211> LENGTH: 26
<212> TYPE: PRT
<213> ORGANISM: *Streptococcus agalactiae*

<400> SEQUENCE: 23

Met Lys Met Asn Lys Lys Val Leu Leu Thr Ser Thr Met Ala Ala Ser
1 5 10 15

Leu Leu Ser Val Ala Ser Val Gln Ala Ser
20 25

<210> SEQ ID NO 24
<211> LENGTH: 713
<212> TYPE: PRT
<213> ORGANISM: *Campylobacter jejuni*

<400> SEQUENCE: 24

Met Leu Lys Lys Glu Tyr Leu Lys Asn Pro Tyr Leu Val Leu Phe Ala
1 5 10 15

Met Ile Ile Leu Ala Tyr Val Phe Ser Val Phe Cys Arg Phe Tyr Trp
20 25 30

Val Trp Trp Ala Ser Glu Phe Asn Glu Tyr Phe Phe Asn Asn Gln Leu
35 40 45

Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp
50 55 60

-continued

Met	Ile	Ala	Gly	Phe	His	Gln	Pro	Asn	Asp	Leu	Ser	Tyr	Tyr	Gly	Ser	65	70	75	80
Ser	Leu	Ser	Ala	Leu	Thr	Tyr	Trp	Leu	Tyr	Lys	Ile	Thr	Pro	Phe	Ser	85	90	95	
Phe	Glu	Ser	Ile	Ile	Leu	Tyr	Met	Ser	Thr	Phe	Leu	Ser	Ser	Leu	Val	100	105	110	
Val	Ile	Pro	Thr	Ile	Leu	Leu	Ala	Asn	Glu	Tyr	Lys	Arg	Pro	Leu	Met	115	120	125	
Gly	Phe	Val	Ala	Ala	Leu	Leu	Ala	Ser	Ile	Ala	Asn	Ser	Tyr	Tyr	Asn	130	135	140	
Arg	Thr	Met	Ser	Gly	Tyr	Tyr	Asp	Thr	Asp	Met	Leu	Val	Ile	Val	Leu	145	150	155	160
Pro	Met	Phe	Ile	Leu	Phe	Phe	Met	Val	Arg	Met	Ile	Leu	Lys	Lys	Asp	165	170	175	
Phe	Phe	Ser	Leu	Ile	Ala	Leu	Pro	Leu	Phe	Ile	Gly	Ile	Tyr	Leu	Trp	180	185	190	
Trp	Tyr	Pro	Ser	Ser	Tyr	Thr	Leu	Asn	Val	Ala	Leu	Ile	Gly	Leu	Phe	195	200	205	
Leu	Ile	Tyr	Thr	Leu	Ile	Phe	His	Arg	Lys	Glu	Lys	Ile	Phe	Tyr	Ile	210	215	220	
Ala	Val	Ile	Leu	Ser	Ser	Leu	Thr	Leu	Ser	Asn	Ile	Ala	Trp	Phe	Tyr	225	230	235	240
Gln	Ser	Ala	Ile	Ile	Val	Ile	Leu	Phe	Ala	Leu	Phe	Ala	Leu	Glu	Gln	245	250	255	
Lys	Arg	Leu	Asn	Phe	Met	Ile	Ile	Gly	Ile	Leu	Gly	Ser	Ala	Thr	Leu	260	265	270	
Ile	Phe	Leu	Ile	Leu	Ser	Gly	Gly	Val	Asp	Pro	Ile	Leu	Tyr	Gln	Leu	275	280	285	
Lys	Phe	Tyr	Ile	Phe	Arg	Ser	Asp	Glu	Ser	Ala	Asn	Leu	Thr	Gln	Gly	290	295	300	
Phe	Met	Tyr	Phe	Asn	Val	Asn	Gln	Thr	Ile	Gln	Glu	Val	Glu	Asn	Val	305	310	315	320
Asp	Leu	Ser	Glu	Phe	Met	Arg	Arg	Ile	Ser	Gly	Ser	Glu	Ile	Val	Phe	325	330	335	
Leu	Phe	Ser	Leu	Phe	Gly	Phe	Val	Trp	Leu	Leu	Arg	Lys	His	Lys	Ser	340	345	350	
Met	Ile	Met	Ala	Leu	Pro	Ile	Leu	Val	Leu	Gly	Phe	Leu	Ala	Leu	Lys	355	360	365	
Gly	Gly	Leu	Arg	Phe	Thr	Ile	Tyr	Ser	Val	Pro	Val	Met	Ala	Leu	Gly	370	375	380	
Phe	Gly	Phe	Leu	Leu	Ser	Glu	Phe	Lys	Ala	Ile	Met	Val	Lys	Lys	Tyr	385	390	395	400
Ser	Gln	Leu	Thr	Ser	Asn	Val	Cys	Ile	Val	Phe	Ala	Thr	Ile	Leu	Thr	405	410	415	
Leu	Ala	Pro	Val	Phe	Ile	His	Ile	Tyr	Asn	Tyr	Lys	Ala	Pro	Thr	Val	420	425	430	
Phe	Ser	Gln	Asn	Glu	Ala	Ser	Leu	Leu	Asn	Gln	Leu	Lys	Asn	Ile	Ala	435	440	445	
Asn	Arg	Glu	Asp	Tyr	Val	Val	Thr	Trp	Trp	Asp	Tyr	Gly	Tyr	Pro	Val	450	455	460	
Arg	Tyr	Tyr	Ser	Asp	Val	Lys	Thr	Leu	Val	Asp	Gly	Gly	Lys	His	Leu				

-continued

465	470	475	480
Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gln			
	485	490	495
Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser			
	500	505	510
Phe Tyr Ala Pro Gln Asn Asp Ile Leu Lys Thr Asp Ile Leu Gln Ala			
	515	520	525
Met Met Lys Asp Tyr Asn Gln Ser Asn Val Asp Leu Phe Leu Ala Ser			
	530	535	540
Leu Ser Lys Pro Asp Phe Lys Ile Asp Thr Pro Lys Thr Arg Asp Ile			
	545	550	555
Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu Ile Phe Ser Thr Val Ala			
	565	570	575
Ser Phe Ser Phe Ile Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe			
	580	585	590
Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu Ile Tyr			
	595	600	605
Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys Ile			
	610	615	620
Gly Asp Asn Val Val Ser Val Asn Ser Ile Val Glu Ile Asn Ser Ile			
	625	630	635
Lys Gln Gly Glu Tyr Lys Ile Thr Pro Ile Asp Asp Lys Ala Gln Phe			
	645	650	655
Tyr Ile Phe Tyr Leu Lys Asp Ser Ala Ile Pro Tyr Ala Gln Phe Ile			
	660	665	670
Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gln Met Phe Phe			
	675	680	685
Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val Ile Asn Ser Arg			
	690	695	700
Asp Ala Lys Val Phe Lys Leu Lys Ile			
	705	710	

```

<210> SEQ ID NO 25
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Consensus sequence
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: Xaa can be Asp or Glu
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (9)..(9)
<223> OTHER INFORMATION: Xaa can be any naturally occurring amino acid
except proline
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (10)..(10)
<223> OTHER INFORMATION: Xaa can be Ser or Thr

<400> SEQUENCE: 25
Gly Ser Gly Gly Gly Xaa Xaa Asn Xaa Xaa Gly Ser Gly Gly

```

-continued

1	5	10
<210> SEQ ID NO 26		
<211> LENGTH: 624		
<212> TYPE: PRT		
<213> ORGANISM: Artificial		
<220> FEATURE:		
<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences inserted at K240+A376		
<400> SEQUENCE: 26		
Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val		
1	5	10 15
Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro		
	20	25 30
Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val		
	35	40 45
Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu		
	50	55 60
Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu		
65	70	75 80
Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser		
	85	90 95
Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn		
	100	105 110
Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His		
	115	120 125
Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys		
	130	135 140
Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu		
145	150	155 160
Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met		
	165	170 175
Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser		
	180	185 190
Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr		
	195	200 205
Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile		
	210	215 220
Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys		
225	230	235 240
Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg Leu His Phe		
	245	250 255
Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln Ala Cys His		
	260	265 270
Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg Gly Trp Glu		
	275	280 285
Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr		
	290	295 300
Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn		
305	310	315 320
Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg		
	325	330 335

-continued

Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu
 340 345 350
 Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala
 355 360 365
 Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala Lys Asp Gln
 370 375 380
 Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala
 385 390 395 400
 Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly
 405 410 415
 Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp Thr Val Glu
 420 425 430
 Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe
 435 440 445
 Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val Phe
 450 455 460
 Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly
 465 470 475 480
 Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp
 485 490 495
 Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg
 500 505 510
 Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu
 515 520 525
 Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly
 530 535 540
 His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu
 545 550 555 560
 Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr Val
 565 570 575
 Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp
 580 585 590
 Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu
 595 600 605
 Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys
 610 615 620

<210> SEQ ID NO 27

<211> LENGTH: 652

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at N-terminal+K240+A376+C-terminal

<400> SEQUENCE: 27

Gly Ser Gly Gly Gly Asp Gln Asn Ala Thr Gly Ser Gly Gly Lys
 1 5 10 15
 Leu Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys
 20 25 30
 Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp
 35 40 45
 Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met
 50 55 60

-continued

Val	Leu	Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	
65					70					75					80	
Leu	Ser	Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	
			85						90					95		
Glu	Pro	Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	
		100						105					110			
Ser	Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	
	115						120					125				
Asn	Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	
130						135					140					
His	Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	
145					150					155					160	
Lys	Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	
			165						170					175		
Glu	Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	
		180						185					190			
Met	Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	
	195					200					205					
Ser	Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	
	210					215					220					
Tyr	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	Trp	Glu	Gly	Lys	
225					230					235					240	
Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	Asp	Leu	Asp	Ile	
			245					250						255		
Lys	Asp	Asn	Asn	Asn	Ser	Thr	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	
		260						265					270			
Phe	Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	
	275						280					285				
His	Leu	Pro	Leu	Glu	Ala	Phe	Thr	Arg	His	Arg	Gln	Pro	Arg	Gly	Trp	
	290					295					300					
Glu	Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	
305					310					315					320	
Tyr	Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	
			325					330						335		
Asn	Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	Ile	
		340						345					350			
Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	Ala	
	355						360					365				
Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	Gly	
	370					375				380						
Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Lys	Asp	
385					390					395					400	
Gln	Asn	Arg	Thr	Lys	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	
			405						410					415		
Ala	Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	
		420						425					430			
Gly	Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	
	435						440					445				
Glu	Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	
	450					455					460					

-continued

Phe	Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val
465					470					475					480
Phe	Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg
				485					490					495	
Gly	Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln
			500					505					510		
Asp	Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu
		515					520					525			
Arg	Val	Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly
	530					535					540				
Leu	Thr	Leu	Ala	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu	Ile
545					550					555					560
Gly	His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu	Glu
				565					570					575	
Glu	Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg	Thr
			580					585					590		
Val	Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly	Gly
		595					600					605			
Asp	Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser	Ala
	610					615					620				
Leu	Pro	Asp	Tyr	Ala	Ser	Gln	Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp	Leu
625					630					635					640
Lys	Leu	Gly	Ser	Gly	Gly	Gly	Asp	Gln	Asn	Ala	Thr				
				645				650							

<210> SEQ ID NO 28

<211> LENGTH: 624

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
inserted at N-terminal+K240+A376+C-terminal

<400> SEQUENCE: 28

Ala	Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val
1				5				10						15	
Leu	Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro
		20					25						30		
Ala	Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val
	35				40						45				
Leu	Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu
	50			55						60					
Ser	Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu
65				70					75					80	
Pro	Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser
			85					90						95	
Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn
		100					105						110		
Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His
		115				120						125			
Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys
	130					135					140				
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu
145					150					155					160

-continued

Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met
				165					170					175	
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser
			180					185					190		
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys
		195					200					205			
Asp	Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp
	210					215					220				
Thr	Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys
	225				230					235					240
His	Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe
				245					250					255	
Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His
			260					265					270		
Leu	Pro	Leu	Glu	Ala	Phe	Thr	Arg	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu
		275					280					285			
Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr
	290					295					300				
Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn
	305				310					315					320
Ala	Leu	Ala	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Pro	Gly	Ser	Gly	Gly	Asp
			325						330					335	
Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu
			340					345					350		
Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly
		355					360					365			
Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys
	370					375					380				
Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala
	385				390					395					400
Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly
			405					410					415		
Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu
		420						425					430		
Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe
		435					440					445			
Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe
	450					455					460				
Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly
	465				470					475					480
Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp
			485					490						495	
Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg
		500						505					510		
Val	Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu
		515					520					525			
Thr	Leu	Ala	Ala	Pro	Glu	Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu	Ile	Gly
	530					535					540				
His	Pro	Leu	Pro	Leu	Arg	Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu	Glu	Glu
	545				550					555					560

-continued

Gly	Gly	Arg	Val	Thr	Ile	Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg	Thr	Val
				565					570					575	
Val	Ile	Pro	Ser	Ala	Ile	Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly	Gly	Asp
			580					585					590		
Leu	Asp	Pro	Ser	Ser	Ile	Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser	Ala	Leu
		595					600					605			
Pro	Asp	Tyr	Ala	Ser	Gln	Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp	Leu	Lys
	610					615					620				

<210> SEQ ID NO 29

<211> LENGTH: 624

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
inserted at Y208+A519

<400> SEQUENCE: 29

Ala	Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val
1				5					10					15	
Leu	Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro
		20						25					30		
Ala	Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val
		35				40						45			
Leu	Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu
	50				55						60				
Ser	Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu
65				70					75					80	
Pro	Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser
			85						90					95	
Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn
		100						105					110		
Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His
		115					120					125			
Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys
	130					135					140				
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu
145				150						155					160
Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met
			165						170					175	
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser
			180					185					190		
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys
		195					200					205			
Asp	Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp
	210					215					220				
Thr	Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys
225				230						235				240	
His	Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe
			245					250						255	
Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His
			260					265					270		
Leu	Pro	Leu	Glu	Ala	Phe	Thr	Arg	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu
		275					280						285		

-continued

Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr
 290 295 300
 Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn
 305 310 315 320
 Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg
 325 330 335
 Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu
 340 345 350
 Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala
 355 360 365
 Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala Ala Gly Glu
 370 375 380
 Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr
 385 390 395 400
 Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly Gly Asp Val Ser Phe Ser
 405 410 415
 Thr Arg Gly Thr Gln Asn Trp Thr Val Glu Arg Leu Leu Gln Ala His
 420 425 430
 Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe Val Gly Tyr His Gly Thr
 435 440 445
 Phe Leu Glu Ala Ala Gln Ser Ile Val Phe Gly Gly Val Arg Ala Arg
 450 455 460
 Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp
 465 470 475 480
 Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg
 485 490 495
 Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg Val Tyr Val Pro Arg Trp
 500 505 510
 Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu Thr Leu Lys Asp Gln Asn
 515 520 525
 Ala Thr Lys Ala Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly
 530 535 540
 His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu
 545 550 555 560
 Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr Val
 565 570 575
 Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp
 580 585 590
 Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu
 595 600 605
 Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys
 610 615 620

<210> SEQ ID NO 30

<211> LENGTH: 624

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

 <223> OTHER INFORMATION: Modified EPA sequence with consensus sequences
 inserted at R274+S318

<400> SEQUENCE: 30

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val

-continued

1	5	10	15
Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro	20	25	30
Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val	35	40	45
Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu	50	55	60
Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu	65	70	75
Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser	85	90	95
Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn	100	105	110
Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His	115	120	125
Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys	130	135	140
Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu	145	150	155
Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met	165	170	175
Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser	180	185	190
Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr	195	200	205
Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile	210	215	220
Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys	225	230	235
Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu	245	250	255
Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Ala Phe	260	265	270
Thr Lys Asp Gln Asn Ala Thr Lys His Arg Gln Pro Arg Gly Trp Glu	275	280	285
Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr	290	295	300
Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val Ile Arg Asn	305	310	315
Ala Leu Ala Lys Asp Gln Asn Ala Thr Lys Pro Gly Ser Gly Gly Asp	325	330	335
Leu Gly Glu Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu	340	345	350
Thr Leu Ala Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly	355	360	365
Asn Asp Glu Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys	370	375	380
Pro Val Ala Ala Gly Glu Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala	385	390	395
Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly	405	410	415

His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala

-continued

130					135					140					
Lys 145	Leu	Ala	Arg	Asp	Ala 150	Thr	Phe	Phe	Val	Arg 155	Ala	His	Glu	Ser	Asn 160
Glu	Met	Gln	Pro	Thr 165	Leu	Ala	Ile	Ser	His 170	Ala	Gly	Val	Ser	Val 175	Val
Met	Ala	Gln	Ala 180	Gln	Pro	Arg	Arg	Glu 185	Lys	Arg	Trp	Ser	Glu 190	Trp	Ala
Ser	Gly	Lys 195	Val	Leu	Cys	Leu	Leu 200	Asp	Pro	Leu	Asp	Gly 205	Val	Tyr	Asn
Tyr	Leu 210	Ala	Gln	Gln	Arg	Cys 215	Asn	Leu	Asp	Asp	Thr 220	Trp	Glu	Gly	Lys
Ile 225	Tyr	Arg	Val	Leu	Ala 230	Gly	Asn	Pro	Ala	Lys 235	His	Asp	Leu	Asp	Ile 240
Lys	Pro	Thr	Val 245	Ile	Ser	His	Arg	Leu	His 250	Phe	Pro	Glu	Gly	Gly 255	Ser
Leu	Ala	Ala	Leu 260	Thr	Ala	His	Gln	Ala 265	Cys	His	Leu	Pro	Leu 270	Glu	Ala
Phe	Thr	Lys 275	Asp	Gln	Asn	Ala	Thr 280	Lys	His	Arg	Gln	Pro 285	Arg	Gly	Trp
Glu	Gln 290	Leu	Glu	Gln	Cys	Gly 295	Tyr	Pro	Val	Gln	Arg 300	Leu	Val	Ala	Leu
Tyr 305	Leu	Ala	Ala	Arg	Leu 310	Ser	Trp	Asn	Gln	Val 315	Asp	Gln	Val	Ile	Arg 320
Asn	Ala	Leu	Ala 325	Ser	Pro	Gly	Ser	Gly	Gly 330	Asp	Leu	Gly	Glu	Ala 335	Ile
Arg	Glu	Gln	Pro 340	Glu	Gln	Ala	Arg	Leu 345	Ala	Leu	Thr	Leu 350	Ala	Ala	Ala
Glu	Ser	Glu 355	Arg	Phe	Val	Arg	Gln 360	Gly	Thr	Gly	Asn	Asp 365	Glu	Ala	Gly
Ala 370	Ala	Ser	Ala	Asp	Val	Val 375	Ser	Leu	Thr	Cys 380	Pro	Val	Ala	Ala	Gly
Glu 385	Cys	Ala	Gly	Pro	Ala 390	Asp	Ser	Gly	Asp	Ala 395	Leu	Leu	Glu	Arg	Asn 400
Tyr	Pro	Thr	Gly 405	Ala	Glu	Phe	Leu	Gly	Asp 410	Gly	Gly	Asp	Val	Ser 415	Phe
Ser	Thr	Arg	Gly 420	Thr	Gln	Asn	Trp	Thr 425	Val	Glu	Arg	Leu 430	Leu	Gln	Ala
His	Arg	Gln 435	Leu	Glu	Glu	Arg	Gly 440	Tyr	Val	Phe	Val	Gly 445	Tyr	His	Gly
Thr 450	Phe	Leu	Glu	Ala	Ala 455	Gln	Ser	Ile	Val	Phe	Gly 460	Gly	Val	Arg	Ala
Arg 465	Ser	Gln	Asp	Leu	Asp 470	Ala	Ile	Trp	Arg	Gly 475	Phe	Tyr	Ile	Ala	Gly 480
Asp	Pro	Ala	Leu 485	Ala	Tyr	Gly	Tyr	Ala	Gln 490	Asp	Gln	Glu	Pro	Asp 495	Ala
Arg	Gly	Arg	Ile 500	Arg	Asn	Gly	Ala	Leu 505	Leu	Arg	Val	Tyr 510	Val	Pro	Arg
Trp	Ser	Leu 515	Pro	Gly	Phe	Tyr	Arg 520	Thr	Gly	Leu	Thr 525	Leu	Lys	Asp	Gln
Asn 530	Ala	Thr	Lys	Ala	Pro	Glu 535	Ala	Ala	Gly	Glu 540	Val	Glu	Arg	Leu	Ile

-continued

Gly His Pro Leu Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu
545 550 555 560

Glu Gly Gly Arg Val Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr
565 570 575

Val Val Ile Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly
580 585 590

Asp Leu Asp Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala
595 600 605

Leu Pro Asp Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu
610 615 620

Lys
625

<210> SEQ ID NO 32
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Forward primer

<400> SEQUENCE: 32

aagctagcgc cgccgaggaa gccttcgacc

30

<210> SEQ ID NO 33
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Reverse primer

<400> SEQUENCE: 33

aagaattctc agtgggtgggtg gtgggtgggtgc ttcaggtcct cgcgcggcgg

50

<210> SEQ ID NO 34
<211> LENGTH: 618
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: EPA_mut_Y208

<400> SEQUENCE: 34

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
1 5 10 15

Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
20 25 30

Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
35 40 45

Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
50 55 60

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

Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
85 90 95

Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
100 105 110

Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
115 120 125

Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys
130															
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu
145															
Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met
165															
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser
180															
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Lys
195															
Asp	Gln	Asn	Ala	Thr	Lys	Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp
210															
Thr	Trp	Glu	Gly	Lys	Ile	Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys
225															
His	Asp	Leu	Asp	Ile	Lys	Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe
245															
Pro	Glu	Gly	Gly	Ser	Leu	Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His
260															
Leu	Pro	Leu	Glu	Thr	Phe	Thr	Arg	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu
275															
Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr
290															
Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn
305															
Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	Ile	Arg
325															
Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	Ala	Glu
340															
Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	Gly	Ala
355															
Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala	Gly	Glu
370															
Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg	Asn	Tyr
385															
Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser	Phe	Ser
405															
Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln	Ala	His
420															
Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	Val	Gly	Tyr	His	Gly	Thr
435															
Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	Gly	Gly	Val	Arg	Ala	Arg
450															
Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	Phe	Tyr	Ile	Ala	Gly	Asp
465															
Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	Gln	Glu	Pro	Asp	Ala	Arg
485															
Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	Val	Tyr	Val	Pro	Arg	Pro
500															
Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	Thr	Leu	Ala	Ala	Pro	Glu
515															
520															
525															

-continued

Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu	Ile	Gly	His	Pro	Leu	Pro	Leu	Arg
530						535					540				
Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu	Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile
545					550					555					560
Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg	Thr	Val	Val	Ile	Pro	Ser	Ala	Ile
				565					570					575	
Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly	Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile
		580						585					590		
Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser	Ala	Leu	Pro	Asp	Tyr	Ala	Ser	Gln
		595					600					605			
Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp	Leu	Lys						
610						615									

<210> SEQ ID NO 35
 <211> LENGTH: 618
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: EPA_mut_R274

<400> SEQUENCE: 35

Ala	Glu	Glu	Ala	Phe	Asp	Leu	Trp	Asn	Glu	Cys	Ala	Lys	Ala	Cys	Val
1				5					10					15	
Leu	Asp	Leu	Lys	Asp	Gly	Val	Arg	Ser	Ser	Arg	Met	Ser	Val	Asp	Pro
		20						25					30		
Ala	Ile	Ala	Asp	Thr	Asn	Gly	Gln	Gly	Val	Leu	His	Tyr	Ser	Met	Val
		35				40						45			
Leu	Glu	Gly	Gly	Asn	Asp	Ala	Leu	Lys	Leu	Ala	Ile	Asp	Asn	Ala	Leu
	50					55					60				
Ser	Ile	Thr	Ser	Asp	Gly	Leu	Thr	Ile	Arg	Leu	Glu	Gly	Gly	Val	Glu
65				70						75					80
Pro	Asn	Lys	Pro	Val	Arg	Tyr	Ser	Tyr	Thr	Arg	Gln	Ala	Arg	Gly	Ser
			85						90					95	
Trp	Ser	Leu	Asn	Trp	Leu	Val	Pro	Ile	Gly	His	Glu	Lys	Pro	Ser	Asn
		100						105					110		
Ile	Lys	Val	Phe	Ile	His	Glu	Leu	Asn	Ala	Gly	Asn	Gln	Leu	Ser	His
		115				120						125			
Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys
	130					135					140				
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu
145				150						155					160
Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met
			165						170					175	
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser
		180						185					190		
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Tyr
		195					200					205			
Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	Trp	Glu	Gly	Lys	Ile
	210					215					220				
Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	Asp	Leu	Asp	Ile	Lys
225				230						235					240
Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	Pro	Glu	Gly	Gly	Ser	Leu
				245					250						255

-continued

Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	Leu	Pro	Leu	Glu	Ala	Phe	260	265	270
Thr	Lys	Asp	Gln	Asn	Ala	Thr	Lys	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu	275	280	285
Gln	Leu	Glu	Gln	Cys	Gly	Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	290	295	300
Leu	Ala	Ala	Arg	Leu	Ser	Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	305	310	315
Ala	Leu	Ala	Ser	Pro	Gly	Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	Ile	Arg	325	330	335
Glu	Gln	Pro	Glu	Gln	Ala	Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	Ala	Glu	340	345	350
Ser	Glu	Arg	Phe	Val	Arg	Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	Gly	Ala	355	360	365
Ala	Ser	Ala	Asp	Val	Val	Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala	Gly	Glu	370	375	380
Cys	Ala	Gly	Pro	Ala	Asp	Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg	Asn	Tyr	385	390	395
Pro	Thr	Gly	Ala	Glu	Phe	Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser	Phe	Ser	405	410	415
Thr	Arg	Gly	Thr	Gln	Asn	Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln	Ala	His	420	425	430
Arg	Gln	Leu	Glu	Glu	Arg	Gly	Tyr	Val	Phe	Val	Gly	Tyr	His	Gly	Thr	435	440	445
Phe	Leu	Glu	Ala	Ala	Gln	Ser	Ile	Val	Phe	Gly	Gly	Val	Arg	Ala	Arg	450	455	460
Ser	Gln	Asp	Leu	Asp	Ala	Ile	Trp	Arg	Gly	Phe	Tyr	Ile	Ala	Gly	Asp	465	470	475
Pro	Ala	Leu	Ala	Tyr	Gly	Tyr	Ala	Gln	Asp	Gln	Glu	Pro	Asp	Ala	Arg	485	490	495
Gly	Arg	Ile	Arg	Asn	Gly	Ala	Leu	Leu	Arg	Val	Tyr	Val	Pro	Arg	Trp	500	505	510
Ser	Leu	Pro	Gly	Phe	Tyr	Arg	Thr	Gly	Leu	Thr	Leu	Ala	Ala	Pro	Glu	515	520	525
Ala	Ala	Gly	Glu	Val	Glu	Arg	Leu	Ile	Gly	His	Pro	Leu	Pro	Leu	Arg	530	535	540
Leu	Asp	Ala	Ile	Thr	Gly	Pro	Glu	Glu	Glu	Gly	Gly	Arg	Val	Thr	Ile	545	550	555
Leu	Gly	Trp	Pro	Leu	Ala	Glu	Arg	Thr	Val	Val	Ile	Pro	Ser	Ala	Ile	565	570	575
Pro	Thr	Asp	Pro	Arg	Asn	Val	Gly	Gly	Asp	Leu	Asp	Pro	Ser	Ser	Ile	580	585	590
Pro	Asp	Lys	Glu	Gln	Ala	Ile	Ser	Ala	Leu	Pro	Asp	Tyr	Ala	Ser	Gln	595	600	605
Pro	Gly	Lys	Pro	Pro	Arg	Glu	Asp	Leu	Lys							610	615	

<210> SEQ ID NO 36

<211> LENGTH: 618

<212> TYPE: PRT

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: EPA_mut_S318

-continued

<400> SEQUENCE: 36

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val
 1 5 10 15
 Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro
 20 25 30
 Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val
 35 40 45
 Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu
 50 55 60
 Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu
 65 70 75 80
 Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser
 85 90 95
 Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn
 100 105 110
 Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His
 115 120 125
 Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu Leu Ala Lys
 130 135 140
 Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu Ser Asn Glu
 145 150 155 160
 Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser Val Val Met
 165 170 175
 Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu Trp Ala Ser
 180 185 190
 Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val Tyr Asn Tyr
 195 200 205
 Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu Gly Lys Ile
 210 215 220
 Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu Asp Ile Lys
 225 230 235 240
 Pro Thr Val Ile Ser His Arg Leu His Phe Pro Glu Gly Gly Ser Leu
 245 250 255
 Ala Ala Leu Thr Ala His Gln Ala Cys His Leu Pro Leu Glu Ala Phe
 260 265 270
 Thr Arg His Arg Gln Pro Arg Gly Trp Glu Gln Leu Glu Gln Cys Gly
 275 280 285
 Tyr Pro Val Gln Arg Leu Val Ala Leu Tyr Leu Ala Ala Arg Leu Ser
 290 295 300
 Trp Asn Gln Val Asp Gln Val Ile Arg Asn Ala Leu Ala Lys Asp Gln
 305 310 315 320
 Asn Ala Thr Lys Pro Gly Ser Gly Gly Asp Leu Gly Glu Ala Ile Arg
 325 330 335
 Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala Ala Ala Glu
 340 345 350
 Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu Ala Gly Ala
 355 360 365
 Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala Ala Gly Glu
 370 375 380
 Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala Leu Leu Glu Arg Asn Tyr

-continued

385	390	395	400
Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly Gly Asp Val Ser Phe Ser	405	410	415
Thr Arg Gly Thr Gln Asn Trp Thr Val Glu Arg Leu Leu Gln Ala His	420	425	430
Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe Val Gly Tyr His Gly Thr	435	440	445
Phe Leu Glu Ala Ala Gln Ser Ile Val Phe Gly Gly Val Arg Ala Arg	450	455	460
Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly Phe Tyr Ile Ala Gly Asp	465	470	475
Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp Gln Glu Pro Asp Ala Arg	485	490	495
Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg Val Tyr Val Pro Arg Trp	500	505	510
Ser Leu Pro Gly Phe Tyr Arg Thr Gly Leu Thr Leu Ala Ala Pro Glu	515	520	525
Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg	530	535	540
Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg Val Thr Ile	545	550	555
Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile	565	570	575
Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ile	580	585	590
Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln	595	600	605
Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys	610	615	

<210> SEQ ID NO 37
 <211> LENGTH: 618
 <212> TYPE: PRT
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: EPA_mut_A519

<400> SEQUENCE: 37

Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys Ala Cys Val	1	5	10	15
Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser Val Asp Pro	20	25	30	
Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr Ser Met Val	35	40	45	
Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp Asn Ala Leu	50	55	60	
Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly Gly Val Glu	65	70	75	80
Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala Arg Gly Ser	85	90	95	
Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys Pro Ser Asn	100	105	110	
Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln Leu Ser His				

-continued

115					120					125					
Met	Ser	Pro	Ile	Tyr	Thr	Ile	Glu	Met	Gly	Asp	Glu	Leu	Leu	Ala	Lys
	130					135					140				
Leu	Ala	Arg	Asp	Ala	Thr	Phe	Phe	Val	Arg	Ala	His	Glu	Ser	Asn	Glu
145					150					155					160
Met	Gln	Pro	Thr	Leu	Ala	Ile	Ser	His	Ala	Gly	Val	Ser	Val	Val	Met
				165					170					175	
Ala	Gln	Ala	Gln	Pro	Arg	Arg	Glu	Lys	Arg	Trp	Ser	Glu	Trp	Ala	Ser
			180					185					190		
Gly	Lys	Val	Leu	Cys	Leu	Leu	Asp	Pro	Leu	Asp	Gly	Val	Tyr	Asn	Tyr
	195						200					205			
Leu	Ala	Gln	Gln	Arg	Cys	Asn	Leu	Asp	Asp	Thr	Trp	Glu	Gly	Lys	Ile
	210					215					220				
Tyr	Arg	Val	Leu	Ala	Gly	Asn	Pro	Ala	Lys	His	Asp	Leu	Asp	Ile	Lys
225					230					235					240
Pro	Thr	Val	Ile	Ser	His	Arg	Leu	His	Phe	Pro	Glu	Gly	Gly	Ser	Leu
				245					250					255	
Ala	Ala	Leu	Thr	Ala	His	Gln	Ala	Cys	His	Leu	Pro	Leu	Glu	Ala	Phe
				260					265					270	
Thr	Arg	His	Arg	Gln	Pro	Arg	Gly	Trp	Glu	Gln	Leu	Glu	Gln	Cys	Gly
	275						280					285			
Tyr	Pro	Val	Gln	Arg	Leu	Val	Ala	Leu	Tyr	Leu	Ala	Ala	Arg	Leu	Ser
	290					295					300				
Trp	Asn	Gln	Val	Asp	Gln	Val	Ile	Arg	Asn	Ala	Leu	Ala	Ser	Pro	Gly
305					310					315					320
Ser	Gly	Gly	Asp	Leu	Gly	Glu	Ala	Ile	Arg	Glu	Gln	Pro	Glu	Gln	Ala
				325					330						335
Arg	Leu	Ala	Leu	Thr	Leu	Ala	Ala	Ala	Glu	Ser	Glu	Arg	Phe	Val	Arg
				340					345					350	
Gln	Gly	Thr	Gly	Asn	Asp	Glu	Ala	Gly	Ala	Ala	Ser	Ala	Asp	Val	Val
	355					360						365			
Ser	Leu	Thr	Cys	Pro	Val	Ala	Ala	Gly	Glu	Cys	Ala	Gly	Pro	Ala	Asp
	370					375					380				
Ser	Gly	Asp	Ala	Leu	Leu	Glu	Arg	Asn	Tyr	Pro	Thr	Gly	Ala	Glu	Phe
385				390						395					400
Leu	Gly	Asp	Gly	Gly	Asp	Val	Ser	Phe	Ser	Thr	Arg	Gly	Thr	Gln	Asn
				405					410					415	
Trp	Thr	Val	Glu	Arg	Leu	Leu	Gln	Ala	His	Arg	Gln	Leu	Glu	Glu	Arg
			420					425					430		
Gly	Tyr	Val	Phe	Val	Gly	Tyr	His	Gly	Thr	Phe	Leu	Glu	Ala	Ala	Gln
	435					440						445			
Ser	Ile	Val	Phe	Gly	Gly	Val	Arg	Ala	Arg	Ser	Gln	Asp	Leu	Asp	Ala
	450					455					460				
Ile	Trp	Arg	Gly	Phe	Tyr	Ile	Ala	Gly	Asp	Pro	Ala	Leu	Ala	Tyr	Gly
465					470					475					480
Tyr	Ala	Gln	Asp	Gln	Glu	Pro	Asp	Ala	Arg	Gly	Arg	Ile	Arg	Asn	Gly
				485					490					495	
Ala	Leu	Leu	Arg	Val	Tyr	Val	Pro	Arg	Trp	Ser	Leu	Pro	Gly	Phe	Tyr
				500					505				510		
Arg	Thr	Gly	Leu	Thr	Leu	Lys	Asp	Gln	Asn	Ala	Thr	Lys	Ala	Pro	Glu
			515				520					525			

-continued

Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu Pro Leu Arg
530 535 540

Leu Asp Ala Ile Thr Gly Pro Glu Glu Gly Gly Arg Val Thr Ile
545 550 555 560

Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile Pro Ser Ala Ile
565 570 575

Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp Pro Ser Ser Ile
580 585 590

Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp Tyr Ala Ser Gln
595 600 605

Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys
610 615

<210> SEQ ID NO 38

<211> LENGTH: 27

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 38

gaaggcgggc gcgtagccat tctcggc

27

<210> SEQ ID NO 39

<211> LENGTH: 27

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 39

gccgagaatg gtcacgcgcc cgccttc

27

<210> SEQ ID NO 40

<211> LENGTH: 1932

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Nucleotide sequence EPA with mutation Y208 >
KDQNTK

<400> SEQUENCE: 40

atgaaaaaga tttggctggc gctggctggt ttagtttttag cgttttagcgc tagcgccgcc 60

gaggaagcct tcgacctctg gaacgaatgc gccaaaggcct gcgtgctcga cctcaaggac 120

ggcgtgcgtt ccagccgcat gacgctcgac cgggccatcg ccgacaccaa cggccagggc 180

gtgctgcact actccatggt cctggagggc ggcaacgacg cgctcaagct ggccatcgac 240

aacgccctca gcatcaccag cgacggcctg accatccgcc tcgaagggtg cgtcgagccg 300

aacaagccgg tgcgtacag ctacacgcgc caggcgcgcg gcagttggtc gctgaactgg 360

ctggtgcga tcggccacga gaagccttcg aacatcaagg tgttcatcca cgaactgaac 420

gccgtaacc agctcagcca catgtcgccg atctacacca tcgagatggg cgacgagttg 480

ctggcgaagc tggcgcgca tgccaccttc ttcgtcaggc cgacgagag caacgagatg 540

cagccgacgc tcgccatcag ccatgcgggg gtcagcgtgg tcatggccca ggcccagccg 600

cgccgggaaa agcgctggag cgaatgggcc agcggcaagg tggtgtgcct gctcgacccg 660

-continued

ctggacgggg tctacaacaa agatcagaac gcgaccaaac tgcgccagca gcgctgcaac	720
ctcgacgata cctgggaagg caagatctac cgggtgctcg ccggcaaccc ggcgaaagcat	780
gacctggaca tcaageccac ggtcatcagt catcgctgc atttcccga gggcggcagc	840
ctggccgcgc tgaccgcgca ccaggcctgc cacctgccgc tggagacctt caccgctcat	900
cgccagccgc gcgctggga acaactggag cagtgcggt atccggtgca gcggtgggtc	960
gccctctacc tggcgcgcg gctgtcgtgg aaccaggctg accagggtgat ccgcaacgcc	1020
ctggccagcc ccggcagcgg cggcgacctg ggcaagcga tccgcgagca gccggagcag	1080
gccccctctg ccctgacctt ggccgcgcgc gagagcgagc gcttcgtccg gcagggcacc	1140
ggcaacgacg aggcggcgcg gccacgcgc gacgtggtga gcctgacctg cccggtcgcc	1200
gccggtgaat gcgcggggcc ggccgacagc ggcgacgccc tgctggagcg caactatccc	1260
actggcgcg agttcctcgg cgacggcggc gacgtcagct tcagcacccg cggcacgcag	1320
aactggacgg tggagcggt gctccaggcg caccgccaac tggaggagcg cggctatgtg	1380
ttcgtcggct accacggcac ctctctcgaa gcggcgcaaa gcacgtctt cggcggggtg	1440
cgcgcgca gccaggacct cgacgcgac tggcgcggtt tctatatcg cgcgatccg	1500
gcgctggcct acggctacgc ccaggaccag gaacccgacg cgcgcgccg gatccgcaac	1560
ggtgccctgc tgcgggtcta tgtgccgcgc ccgagctgc cgggttcta ccgcaccgc	1620
ctgacctg ccgcgcgga ggccggcggc gaggtcgaa ggctgatcgg ccacccgctg	1680
ccgctgcgcc tggacgcat caccggcccc gaggaggaag gcggcgcgct gaccattctc	1740
ggctggccgc tggccgagcg caccgtggtg attccctcgg cgatccccc cgcaccgcgc	1800
aacgtcgcg gcgacctga cccgtccagc atccccgaca aggaacaggc gatcagcgcc	1860
ctgcgggact acgccagcca gcccggcaaa ccgcgcgcgc aggaactgaa gcaccaccac	1920
caccaccact ga	1932

<210> SEQ ID NO 41

<211> LENGTH: 1932

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: SEQ ID NO: 41 Nucleotide sequence EPA with mutation R274 > KDQNAK

<400> SEQUENCE: 41

atgaaaaaga tttggtggc gctggctggt ttagttttag cgtttagcgc tagcgccgcc	60
gaggaagcct tcgacctctg gaacgaatgc gccaaaggcct gcgtgctcga cctcaaggac	120
ggcgtgcgtt ccagccgcat gagcgtcgac ccggccatcg ccgacaccaa cggccagggc	180
gtgctgcact actccatggt cctggagggc ggcaacgacg cgctcaagct ggccatcgac	240
aacgccccca gcatcaccag cgacggcctg accatccgcc tcgaaggcgg cgctcgagccg	300
aacaagccgg tgcgtacag ctacacgcgc caggcgcgcg gcagttggtc gctgaaatgg	360
ctggtaccga tcggccacga gaagccctcg aacatcaagg tgttcaccca cgaactgaac	420
gccgtaacc agctcagcca catgtcgccg atctacacca tcgagatggg cgacgagttg	480
ctggcgaagc tggcgcgca tggcaccttc ttcgtcaggg cgcacgagag caacgagatg	540
cagccgacgc tcgccatcag ccatgccggg gtcagcgtgg tcatggctca ggcccagccg	600
cgccgggaaa agcgtggag cgaatgggcc agcggcaagg tggtgtgct gctcgacccg	660

-continued

ctggacgggg tctacaacta cctcgcccag cagcgctgca acctcgacga tacctgggaa	720
ggcaagatct accgggtgct cgccggcaac ccggcggaagc atgacctgga catcaagccc	780
acgggtcatca gtcactgcct gcatctcccc gagggcgga gcctggccgc gctgaccgcg	840
caccaggcct gccacctgcc gctggaggcc ttactaaaag atcagaacgc gaccaaact	900
cgccagccgc gcggtggga acaactggag cagtgcggct atccggtgca gcggtggtc	960
gccctctacc tggcgggcg actgtcgtgg aaccaggctg accaggatg ccgcaacgcc	1020
ctggccagcc ccggcagcgg cgccgacctg ggcaagcga tccgcgagca gccggagcag	1080
gcccgtctgg ccctgacct ggccgcccgc gagagcgagc gcttcgtccg gcagggcacc	1140
ggcaacgacg aggcggcgcc ggccagcgcc gacgtgggta gcctgacctg cccggtcgcc	1200
gccggtgaat gcgcgggccc ggccgacagc ggccagcccc tgcgtggagcg caactatccc	1260
actggcgcg agttcctcgg cgacggcgcc gacgtcagct tcagcaccgc cgccacgcag	1320
aactggacgg tggagcggtc gctccaggcg caccgccaac tggaggagcg cggctatgtg	1380
ttcgtcggt accacggcac ctctctcgaa gcggcgcaaa gcatcgtctt cgcgggggtg	1440
cgcgcgcga gccaggacct cgacgcgac tggcgcggtt tctatatcgc cgccgatccg	1500
gcgtggcct acggtacgc ccaggaccag gaaccgcagc cgccggcgcc gatccgaac	1560
ggtgccctgc tgcgggtcta tgtgcgcgc tggagtctgc cgggcttcta ccgaccggc	1620
ctgacctgg ccgcgcgga ggccggcgcc gaggtcgaa ggctgatcgg ccatccgctg	1680
ccgctgcgc tggacgcat caccggcccc gaggaggaag gcggcgcggt gaccattctc	1740
ggctggcgc tggccgagcg caccgtggtg attcctcgg cgatccccc acgaccgcgc	1800
aacgtcgcg gcgacctga cccgtccagc atccccgaca aggaacaggc gatcagcgcc	1860
ctgccggact acgccagcca gcccgcaaaa ccgcgcgcg aggacttgaa gcaccaccac	1920
caccaccact ga	1932

<210> SEQ ID NO 42

<211> LENGTH: 1932

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Nucleotide sequence EPA with mutation S318 >
KDQNTAK

<400> SEQUENCE: 42

atgaaaaaga tttggctggc gctggctggt ttagttttag cgttttagcg tagcgccgcc	60
gagggaagcct tcgacctctg gaacgaatgc gccaaaggcct gcgtgctcga cctcaaggac	120
ggcgtgcgtt ccagccgcat gagcgtcgac ccggccatcg ccgacaccaa cgccagggc	180
gtgctgcact actccatggt cctggagggc ggcaacgacg cgctcaagct ggccatcgac	240
aacgccccta gcatcaccag cgacggcctg accatccgcc tcgaaggcgg cgtcgagccg	300
aacaagccgg tgcgctacag ctacacgcgc caggcgcgcg gcagttgggt gctgaactgg	360
ctggtaccga tcggccacga gaagccctcg aacatcaagg tgttcattcca cgaactgaac	420
gccgtaacc agtcagcca catgtcgccg atctacacca tcgagatggg cgacgagttg	480
ctggcggaagc tggcgcgga tgccaccttc ttcgtcaggg cgcacgagag caacgagatg	540
cagccgacgc tcgccatcag ccatgcgggg gtcagcgtgg tcatggctca ggccagccg	600

-continued

cgccgggaaa agcgctggag cgaatgggcc agcggcaagg tggtgtgcct gctcgacccg	660
ctggacgggg tctacaacta cctcgcccag cagcgctgca acctcgacga tacctgggaa	720
ggcaagatct accgggtgct cgccggcaac cggcggaagc atgacctgga catcaagccc	780
acgggtcatca gtcctgcct gcatctccc gagggcgga ccctggccgc gctgacccgc	840
caccaggcct gccacctgcc gctggaggcc ttcactcgtc atcgccagcc gcgcggctgg	900
gaacaactgg agcagtgcgg ctatccgtg cagcggtgg tcgccccta cctggcgcg	960
cgactgtcgt ggaaccaggc cgaccagggt atccgcaacg ccctggccaa agatcagaac	1020
gcgacaaac cggcgagcgg cgccgacctg ggcggaagcga tccgcgagca gccggagcag	1080
gcccgtctgg ccctgacct ggccggccgc gagagcgagc gcttcgtccg gcagggcacc	1140
ggcaacgacg agggcgccgc ggccagcgcc gacgtggtga ccctgacctg ccgggtcgcc	1200
gcccgtgaat gcgcgggccc ggccgacagc ggcgacgccc tgctggagcg caactatccc	1260
actggcgcg agttcctcgg cgacggcgcc gacgtcagct tcagcacccg cggcacgcag	1320
aactggacgg tggagcggct gctccaggcg caccgccaac tggaggagcg cggtatgtg	1380
ttcgtcggct accacggcac ctctctcgaa gcggcgcaaa gcacgtctt cggcggggtg	1440
cgcgcgcgca gccaggacct cgacgcgac tggcgcggtt tctatatcgc cgcgcatccg	1500
gcgctggcct acggtacgc ccaggaccag gaaccgcagc cgcgcgccg gatccgcaac	1560
ggtgccctgc tgcgggtcta tgtgccgcgc tggagtctgc cgggcttcta ccgcaccggc	1620
ctgacctgg ccgcgcgga ggccgggggc gaggtcgaa ggctgatcgg ccacccgctg	1680
ccgctgcgc tggacgcat caccggcccc gaggaggaag gcggcgcgct gaccattctc	1740
ggctggccgc tggccgagcg caccgtggtg attccctcgg cgatcccccac cgaccgcgc	1800
aacgtcgcg gcgacctga ccgctccagc atccccgaca aggaacaggc gatcagcgcc	1860
ctgccggaact acgccagcca gcccgcaaaa ccgcccgcgc aggaacttga gcaccaccac	1920
caccaccact ga	1932

<210> SEQ ID NO 43

<211> LENGTH: 1932

<212> TYPE: DNA

<213> ORGANISM: Artificial

<220> FEATURE:

<223> OTHER INFORMATION: Nucleotide sequence EPA with mutation A519 >
KDQNATK

<400> SEQUENCE: 43

atgaaaaaga tttggctggc gctggctggt ttagttagc cgttagcgc tagcgccgc	60
gaggaagcct tcgacctctg gaacgaatgc gccaaaggcct gcgtgctcga cctcaaggac	120
ggcgtgcgtt ccagccgcat gacgctcgac ccggccatcg ccgacacca cgccagggc	180
gtgctgcaact actccatggt cctggagggc ggcaacgacg cgctcaagct ggccatcgac	240
aacgccccta gcatcaccag cgacggcctg accatccgcc tcgaaggcgg cgtcgagccg	300
aacaagccgg tgcgtacag ctacacgcgc caggcgcgcg gcagttggtc gctgaactgg	360
ctggtaccga tcggccacga gaagccctcg aacatcaagg tggtcatcca cgaactgaac	420
gcccgttaac agctcagcca catgtcgccg atctacacca tcgagatggg cgacgagttg	480
ctggcgaagc tggcgcgca tgccacctc ttcgtcaggc cgcacgagag caacgagatg	540
cagccgacgc tcgccatcag ccatgcggg gtcagcgtgg tcatggctca ggcccagccg	600

-continued

cgccgggaaa agcgctggag cgaatgggccc agcggcaagg tgttgtgctt gctcgacccc	660
ctggacgggg tctacaacta cctcgcccag cagcgctgca acctcgacga tacctgggaa	720
ggcaagatct accgggtgct cgccggcaac ccggcggaag atgacctgga catcaagccc	780
acggtcatca gtcategctt gcatctcccc gagggcgga gcttggcgcc gctgacgcgc	840
caccaggcct gccacctgct gctggaggcc ttactctgtc atcgccagcc gcgcggctgg	900
gaacaactgg agcagtgccg ctatccgggtg cagcggtctg tcgccctcta cctggcgccg	960
cgactgtcgt ggaaccaggt cgaccaggtg atccgcaacg ccctggccag ccccgccagc	1020
ggcgccgacc tgggcgaagc gatccgcgag cagccggagc aggcccgctt ggccctgacc	1080
ctggccgccc ccgagagcga gcgctctgtc cggcagggca ccggcaacga cgaggccggc	1140
gcgccagcgc ccgacgtggt gacccctgac tgcccggtcg ccgcccgtga atgcgcgggc	1200
ccggcggaac gcggcgagcc cctgctggag cgcaactatc ccaactggcg ggagtctctc	1260
ggcgacggcg gcgacgtcag cttcagcacc cgcggcacgc agaactggac ggtggagcgc	1320
ctgctccagg cgcaccgcca actggaggag cgcggctatg tgttctgtcg ctaccacggc	1380
accttctctg aagcggcgca aagcatctgt ttccggcggg tgccgcgcgc cagccaggac	1440
ctcgacgcga tctggcgccg tttctatata gccggcgatc cggcgctggc ctacggctac	1500
gcccaggacc aggaaccgga cgcgcgcggc cggatccgca acggtgacct gctgcgggtc	1560
tatgtgccgc gctggagtct gccgggcttc taccgcaccg gcctgacctt gaaagatcag	1620
aacgcgacca aagcgcgga ggccggcggc gaggtcgaa ggctgatcgg ccatccgctg	1680
ccgctgcgcc tggacgccat caccggcccc gaggaggaag gcgggcgcgt gaccattctc	1740
ggctggccgc tggccgagcg caccgtggtg attccctcgg cgatccccac cgaccgcgc	1800
aacgtcggcg gcgacctga cccgtccagc atccccgaca aggaacaggc gatcagcgcc	1860
ctgcccgaact acgccagcca gcccgggcaaa ccgcgcgcgc aggaactgaa gcaccaccac	1920
caccaccact ga	1932

1-15. (canceled)

16. A modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises at least one consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the at least one consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218, (ii) one or more amino acids between amino acid residues 264-284, (iii) one or more amino acids between amino acid residues 308-328, and (iv) one or more amino acids between amino acid residues 509-529 of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

17. The modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) of claim **16** having an amino acid sequence of

SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence comprises at least two consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the at least two consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218, (ii) one or more amino acids between amino acid residues 264-284, (iii) one or more amino acids between amino acid residues 308-328, (iv) one or more amino acids between amino acid residues 509-529, and (v) one or more amino acids between amino acid residues 230-250, of SEQ ID NO: 1 or at equivalent position(s) within an amino acid 5 sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

18. A modified EPA (Exotoxin A of *Pseudomonas aeruginosa*) protein of claim **16** having an amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1, modified in that the amino acid sequence com-

prises at least three consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, wherein the at least three consensus sequences have each been added next to or substituted for one or more amino acids, independently selected from: (i) one or more amino acids between amino acid residues 198-218, (ii) one or more amino acids between amino acid residues 264-284, (iii) one or more amino acids between amino acid residues 308-328, (iv) one or more amino acids between amino acid residues 509-529 of SEQ ID NO: 1, and (v) one or more amino acids between amino acid residues 230-250 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

19. The modified EPA protein of claim **16**, wherein the consensus sequence(s) selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3) have each been independently substituted for one or more amino acids of the amino acid sequence of SEQ ID NO: 1 or an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

20. The modified EPA protein of claim **16**, wherein a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, has been added next to, or substituted for, one or more amino acids, at the N-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

21. The modified EPA protein of claim **16**, wherein a further consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and J-D/E-X-N-Z-S/T-U (SEQ ID NO: 5), wherein X and Z are independently any amino acid except proline and J and U are independently 1 to 5 naturally occurring amino acid residues, has been added next to, or substituted for, one or more amino acids, at the C-terminus of SEQ ID NO: 1 or at an equivalent position within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

22. The modified EPA protein of claim **16**, wherein at least one consensus sequence selected from: D/E-X-N-Z-S/T (SEQ ID NO: 2) and K-D/E-X-N-Z-S/T-K (SEQ ID NO: 3), wherein X and Z are independently any amino acid except proline, has been added next to, or substituted for: (i) one or

more amino acids between amino acid residues 198-218, (ii) one or more amino acids between amino acid residues 308-328, or (iii) one or more amino acids between amino acid residues 509-529 of SEQ ID NO: 1 or at equivalent position(s) within an amino acid sequence at least 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to SEQ ID NO: 1.

23. The modified EPA protein of claim **16**, wherein the amino acid sequence comprises substitution of leucine 552 to valine (L552V) (or at a position equivalent to L552 of SEQ ID NO:1) and deletion of glutamine 553 (4E553) (or at a position equivalent to E553 of SEQ ID NO:1).

24. A conjugate comprising a modified EPA protein of claim **16** covalently linked to an antigen.

25. The conjugate of claim **24**, wherein the antigen is a saccharide.

26. A host cell comprising:

- (i) one or more nucleotide sequences comprising polysaccharide synthesis genes, for producing a bacterial polysaccharide antigen or a yeast polysaccharide antigen or a mammalian polysaccharide antigen, integrated into the host cell genome;
- (ii) a nucleotide sequence encoding a heterologous oligosaccharyl transferase, within a plasmid;
- (iii) a nucleotide sequence that encodes a modified EPA protein according to claim **16**, within a plasmid.

27. A process for producing a bioconjugate that comprises a modified EPA protein linked to a polysaccharide, said process comprising (i) culturing the host cell of claim **26** under conditions suitable for the production of glycoproteins and (ii) isolating the bioconjugate, isolating the bioconjugate from a periplasmic extract from the host cell.

28. An immunogenic composition comprising a conjugate of claim **24**, and at least one of a pharmaceutically acceptable excipient or carrier.

29. A vaccine comprising the immunogenic composition of claim **28** and an adjuvant.

30. A method of inducing an immune response in a subject, the method comprising administering a therapeutically or prophylactically effective amount of the conjugate of claim **24** to a subject in need thereof.

31. A method of inducing an immune response in a subject, the method comprising administering a therapeutically or prophylactically effective amount of the immunogenic composition of claim **28**, to a subject in need thereof.

* * * * *