Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3124312 A1 2020/06/25

(21) 3 124 312

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2019/11/05
(87) Date publication PCT/PCT Publication Date: 2020/06/25
(85) Entrée phase nationale/National Entry: 2021/06/18
(86) N° demande PCT/PCT Application No.: US 2019/059893
(87) N° publication PCT/PCT Publication No.: 2020/131236
(30) Priorité/Priority: 2018/12/21 (US62/783,971)

(51) Cl.Int./Int.Cl. *A61K 38/02* (2006.01), *A61K 39/02* (2006.01), *A61K 48/00* (2006.01), *C07K 14/195* (2006.01)

(71) Demandeur/Applicant:
VAXNEWMO LLC, US

(72) Inventeur/Inventor:
HARDING, CHRISTIAN, US

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : MOTIFS DE RECONNAISSANCE DE GLYCOSYLATION A LIAISON O
(54) Title: O-LINKED GLYCOSYLATION RECOGNITION MOTIFS

A.

$ComP_{ADP1}$ fragment: $_{75}$CVGVQEISA**S**NATTNVATATC$_{95}$ (SEQ ID NO: 26)

(57) Abrégé/Abstract:
Provided herein are glycoproteins containing O-linked glycosylation recognition motifs, and methods of making, for example, for use in the production of conjugate vaccines.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 June 2020 (25.06.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/131236 A1

(51) International Patent Classification:
A61K 38/02 (2006.01) *A61K 48/00* (2006.01)
A61K 39/02 (2006.01) *C07K 14/195* (2006.01)

(21) International Application Number:
PCT/US2019/059893

(22) International Filing Date:
05 November 2019 (05.11.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/783,971 21 December 2018 (21.12.2018) US

(71) Applicant: **VAXNEWMO LLC** [US/US]; 4340 Duncan Ave., Suite 202, St. Louis, Missouri 63110 (US).

(72) Inventor: **HARDING, Christian**; c/o Vaxnewmo LLC, 4340 Duncan Ave., Suite 202, St. Louis, Missouri 63110 (US).

(74) Agent: **HOLTZ, William A.** et al.; Thompson Coburn LLP, One US Bank Plaza, St. Louis, Missouri 63101 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

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

WO 2020/131236 A1

(54) Title: *O*-LINKED GLYCOSYLATION RECOGNITION MOTIFS

A.

$ComP_{ADP1}$ fragment: $_{75}$CVGVQEISASNATTNVATATC$_{95}$ (SEQ ID NO: 26)

(57) Abstract: Provided herein are glycoproteins containing *O*-linked glycosylation recognition motifs, and methods of making, for example, for use in the production of conjugate vaccines.

O-LINKED GLYCOSYLATION RECOGNITION MOTIFS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This PCT application claims the benefit of U.S. Provisional Appl. No. 62/783,971, filed on December 21, 2018.

[0002] This application is related to U.S. Appl. No. 15/553,733, filed August 25, 2017, which is a U.S. national stage application of PCT/CA2016/050208, filed February 26, 2016, which claims the benefit of U.S. Provisional Appl. No. 62/121,439, filed on February 26, 2015.

[0003] This application is also related to PCT/US2019/037251, filed June 14, 2019, which claims the benefit of U.S. Provisional Appl. No. 62/685,970, filed on June 16, 2018 and U.S. Provisional Appl. No. 62/783,971, filed on December 21, 2018.

GOVERNMENT FUNDING STATEMENT

[0004] This invention was made with government support under the R41 AI142928-01 grant awarded by the National Institute for Allergy and Infectious Disease (NIAID). The Government has certain rights in the invention.

BACKGROUND

[0005] The first, general protein glycosylation pathway in bacteria, the *N*-linked glycosylation system of *Campylobacter jejuni*, was discovered two decades ago (Szymanski CM, et al. (1999) Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* 32(5):1022-1030). Since then, many diverse prokaryotic glycosylation systems have been characterized, including *O*-linked glycosylation systems that have no homologous counterparts in eukaryotic organisms (Iwashkiw JA, et al. (2013) Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol* 89(1):14-28). Shortly after these discoveries, glycosylation pathways were recombinantly introduced into *E. coli* creating the field of bacterial glycoengineering (Wacker M, *et al.* (2002) N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* 298(5599):1790-1793). Bacterial glycoengineering is an emerging biotechnological tool that

harnesses prokaryotic glycosylation systems for the generation of recombinantly glycosylated proteins using *E. coli* or other Gram-negative organisms as a host. Currently, glycoengineering utilizes two broad approaches to recombinantly glycosylate proteins, both of which can generate *N*- or *O*-linkages: oligosaccharyltransferase (OTase)-dependent and OTase-independent.

[0006] Protein glycosylation, or the covalent attachment of carbohydrates to proteins, is a ubiquitous posttranslational modification. For the most part, protein glycosylation is characterized as either *N*-linked with glycans attached to asparagine residues, or as *O*-linked with glycans attached to serine or threonine residues. While the importance of eukaryotic glycosylation has been and continues to be a source of intensive research, prokaryotic glycosylation has only recently grabbed the attention of the scientific community with the discovery of a general *N*-linked protein glycosylation system in the ε-proteobacterium *Campylobacter jejuni* (Szymanski CM, et al. (1999) Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* 32(5):1022-1030). Since the initial *C. jejuni* discovery, prokaryotic glycosylation systems have been described across a plethora of Gram-negative and Gram-positive bacteria and been shown to contribute towards normal bacterial physiology as well as pathogenesis (Iwashkiw JA, et al. (2013) Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol* 89(1):14-28; Nothaft H & Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8(11):765-778); Schaffer C & Messner P (2017) Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* 41(1):49-91). Given the straightforward nature of prokaryotic genetics, it was only a matter of time before protein glycosylation systems were engineered and exploited for the production of designer glycoproteins in a process termed "bacterial glycoengineering".

[0007] Much like eukaryotic glycosylation, bacteria have evolved an *N*-linked OTase pathway, but also employ *O*-linked OTase systems that are unique to prokaryotic organisms. OTase-independent glycosylation occurs in the cytoplasm and relies on glycosyltransferases to transfer monosaccharides from nucleotide activated precursors for the sequential assembly of glycoproteins. Both OTase-dependent and -independent pathways are exploited for bioconjugating carbohydrates to proteins.

[0008] Bacterial surface polysaccharides are some of the first, and most abundant, microbial components encountered by the immune system during infection (Comstock LE & Kasper DL (2006) Bacterial glycans: key mediators of diverse host immune responses. *Cell* 126(5):847-850).

These polysaccharides, usually in the form of capsule or O antigen attached to lipid A, serve a multitude of purposes, including protecting microbial organisms from external threats and immune clearance. Given their abundance on invading organisms as well as their biochemical distinctness from eukaryotic carbohydrates, some microbial surface polysaccharides have been used as antigens for vaccine development. However, when polysaccharides are used alone in vaccine formulations, they usually act as T-cell independent antigens and therefore do not stimulate immunoglobulin class switching and long-term B cell memory. Moreover, polysaccharide vaccines alone do not elicit protection in vulnerable groups like infants and children under two years of age. This poor immune response can be overcome by covalently attaching a polysaccharide to a protein carrier in a process known as conjugation (De Gregorio E & Rappuoli R (2014) From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nat Rev Immunol* 14(7):505-514).

[0009] Traditionally, glycoconjugate vaccines are synthesized using a semi-synthetic approach where the polysaccharide is extracted from the target bacterium, purified, chemically modified and covalently linked to a carrier protein. This approach has resulted in the commercial licensure of multiple glycoconjugate vaccines to prevent colonization and infection by *Haemophilus influenzae* type B, and multiple serotypes of *Streptococcus pneumoniae* and *Neisseria meningiditis*. For detailed reviews on semi-synthetic or synthetic glycoconjugate vaccine production please refer to the following excellent review article (Berti F & Adamo R (2018) Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification. *Chem Soc Rev* 47(24):9015-9025). Although conjugate vaccines produced chemically have seen immense commercial success (the glycoconjugate vaccine Prevnar 13 has been Pfizer's best-selling product from 2015-2018 with over 24 billion USD in sales), their manufacturing processes are not without drawbacks; including, batch to batch variation, heterogenous product formation, large scale production of pathogenic organisms, and high manufacturing costs (Frasch CE (2009) Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* 27(46):6468-6470).

[0010] Over the last two decades, alternative strategies for producing glycoconjugate vaccines have emerged. These techniques are broad in their approach with some yielding vaccines closer to commercial licensure than others. Specifically, the advent of *in vivo* bacterial conjugations for manufacturing glycoconjugate vaccines have produced some of the most clinically advanced products to date. Commonly referred to as bioconjugation or protein glycan

coupling technology (PGCT), the *in vivo* conjugation of polysaccharides to proteins for glycoconjugate vaccine production relies on OTases (Frasch CE (2009) Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* 27(46):6468-6470). It is generally considered that bioconjugation represents a simplification of the production and manufacturing process of glycoconjugate vaccines (Rappuoli R, De Gregorio E, & Costantino P (2019) On the mechanisms of conjugate vaccines. *Proc Natl Acad Sci U S A* 116(1):14-16).

[0011] Both *N*-linking and *O*-linking OTases have been employed for biologically conjugating polysaccharides to carrier proteins for glycoconjugate vaccine production. Regardless of which OTase is employed, biological conjugations in any Gram-negative bacterium rely on three components: a genetic locus or loci that encode(s) for the polysaccharide biosynthesis proteins, a carrier protein to be glycosylated, and an OTase to transfer the desired carbohydrate to the carrier protein. While these three components are required, they do not necessarily need to be on three separate plasmids.

[0012] Recently, a third class of *O*-linking OTase was employed for bioconjugate vaccine production (Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891). Much like the only other known *O*-linking OTases, PilO and PglL, this third class of OTase, termed PglS, naturally glycosylates a pilin like protein, ComP (Schulz BL, *et al.* (2013) Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. *PLoS One* 8(5):e62768). A follow up study demonstrated that PglS was indeed a pilin specific OTase, likely, only glycosylating ComP as no other glycoproteins were identified using a comprehensive glycoprotein screening approach (Harding CM, *et al.* (2015) Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol Microbiol* 96(5):1023-1041). Originally characterized as a PglL ortholog from the environmental bacterium *Acinetobacter baylyi* strain ADP1, PglS is in fact phylogenetically distinct from PglL proteins. Strains of *Acinetobacter* that encode for a PglS protein also encode for a PglL protein, which has been shown to act as the general OTase glycosylating at least seven membrane-associated proteins in a manner similar to *Neisseria* species (Iwashkiw JA, *et al.* (2012) Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathog* 8(6):e1002758). In addition, some strains of *Acinetobacter* also encode for PilO

OTases, making *Acinetobacter* the only known genera of bacteria carrying genes for all three *O*-OTase families (PilO, PglL, and PglS) (Harding CM, *et al.* (2015) Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol Microbiol* 96(5):1023-1041; Iwashkiw JA, *et al.* (2012) Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathog* 8(6):e1002758).

[0013] Aside from phylogenetic differences, PglS glycosylates its cognate pilin at a unique serine site that is not conserved when compared to the site of glycosylation for PilE (the pilin target of PglL) or PilA (the pilin target for PilO), and is not contained within an LCR (Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891). However, the most notable difference lies in the polysaccharide substrates PglS transfers. PglS is the only known OTase, both *N*- or *O*-linking, capable of transferring polysaccharides with glucose at the reducing end. Many pathogens, like *Streptococcus pneumoniae* (Geno KA, *et al.* (2015) Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin Microbiol Rev* 28(3):871-899), Group B *Streptococcus* (Carboni F, *et al.* (2017) Structure of a protective epitope of group B Streptococcus type III capsular polysaccharide. *Proc Natl Acad Sci U S A* 114(19):5017-5022), and *Klebsiella pneumoniae* (Pan YJ, *et al.* (2015) Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of Klebsiella spp. *Sci Rep* 5:15573), produce capsules that contain polysaccharides with glucose at the reducing and are thus potential targets for PglS dependent bioconjugate vaccine development. Indeed, PglS was used to generate a polyvalent pneumococcal bioconjugate vaccine against serotypes 8, 9V, and 14 (all contain glucose at the reducing end) using the natural acceptor, ComP, as a carrier protein. In addition, a fragment of ComP lacking its first 28 amino acids was also able to serve as a glycotag when translationally fused to the C-terminus of exotoxin A of *P. aeruginosa* paving the way for incorporation of more conventional vaccine carriers in the PglS bioconjugation system (Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891).

SUMMARY

[0014] This disclosure provides for a bioconjugate comprising an oligo- or polysaccharide covalently linked to a fusion protein: wherein the fusion protein comprises a ComP protein (ComP) glycosylation tag; wherein the ComP glycosylation tag comprises both a cysteine residue corresponding to the conserved cysteine residue at position 71 of SEQ ID NO: 2 (ComP110264: ENV58402.1) and a cysteine residue corresponding to the conserved cysteine residue at position 93 of SEQ ID NO: 2 or both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 (ComPADP1: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1; and wherein the fusion protein is glycosylated with the oligo- or polysaccharide on the ComP glycosylation tag at a serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 or position 84 of SEQ ID NO: 1. In certain embodiments, the ComP glycosylation tag does not comprise a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 (ComP110264: ENV58402.1). In certain embodiments, the fusion protein of the bioconjugate does not comprise, in relationship to the ComP glycosylation tag, a methionine residue at a position that would correspond to or correspond about to the conserved methionine residue at position 104 of SEQ ID NO: 2 (ComP110264: ENV58402.1). In certain embodiments, the bioconjugate is a conjugate vaccine.

[0015] In certain aspects of this disclosure, the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11], or a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions, wherein the variant maintains both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 (ComPADP1: AAC45886.1) and a cysteine residue corresponding to the

conserved cysteine residue at position 95 of SEQ ID NO: 1; and wherein the variant maintains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1.

[0016] This disclosure provides for a ComP glycosylation tag comprising an isolated fragment of a ComP protein, wherein the fragment comprises a serine residue corresponding to the conserved serine residue at position 84 in SEQ ID NO: 1 (ComPADP1: AAC45886.1) and both a cysteine residue corresponding to the conserved cysteine residue at position 71 of SEQ ID NO: 2 (ComP110264: ENV58402.1) and a cysteine residue corresponding to the conserved cysteine residue at position 93 of SEQ ID NO: 2 or both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 (ComPADP1: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1. In certain embodiments, the ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag does not comprise a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 (ComP110264: ENV58402.1). In certain embodiments, the ComP glycosylation tag of Claim 42, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 (ComP110264: ENV58402.1).

[0017] Provided for herein is a fusion protein comprising a ComP glycosylation tag of this disclosure.

[0018] Also provided for herein is a method of *in vivo* conjugation of an oligo- or polysaccharide to an acceptor polypeptide, the method comprising covalently linking the oligo- or polysaccharide to the acceptor polypeptide with a PglS oligosaccharyltransferase (OTase), wherein the acceptor polypeptide comprises the ComP glycosylation tag of this disclosure; optionally, wherein the ComP glycosylation tag is linked to a heterologous carrier protein.

[0019] Also provided for herein is a host cell comprising (a) a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide; (b) a PglS OTase; and (3) an acceptor polypeptide comprising the ComP glycosylation tag of this disclosure.

[0020] Also provided for herein is an isolated nucleic acid encoding the ComP glycosylation tag and/or the fusion protein of this disclosure and a host cell comprising said isolated nucleic acid.

[0021] Also provided for herein is a composition comprising the conjugate vaccine or the fusion protein of thisi disclosure, and an adjuvant.

[0022] A method of inducing a host immune response against a bacterial pathogen comprising administering to a subject in need of the immune response an effective amount of the conjugate vaccine, the fusion protein, or a composition of this disclosure.

[0023] Also provided for herein is a method of preventing or treating a bacterial disease and/or infection in a subject comprising administering to a subject in need thereof the conjugate vaccine, the fusion protein, or a composition of this disclosure.

[0024] Also provided for herein is a method of producing a pneumococcal conjugate vaccine against pneumococcal infection comprising isolating the bioconjugate or glycosylated fusion protein of this disclosure and combining the isolated conjugate vaccine or isolated glycosylated fusion protein with an adjuvant.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] **Figure 1A,B.** **Figure 1A** and **Figure 1B** show that the cysteine residues flanking immediately serine 84 in ComP from *Acinetobacter baylyi* ADP1 ($ComP_{ADP1}$) contribute to PglS dependent glycosylation and ComP stability. **(A)** illustrates the amino acid sequence of $ComP_{ADP1}$ from amino acid residues 75 to 95 with the two cysteine residues flanking serine 84, the site of PglS dependent glycosylation. **(B)** shows that point mutational exchange of either cysteine 75, cysteine 95, or both cysteine 75 and 95 to alanine, glycine, or serine negatively affects ComP stability and blocks glycosylation of serine 84 by PglS with the *Campylobacter jejuni* heptasaccharide. Western blot analysis of *E. coli* whole cell lysates co-expressing PglS, the *C. jejuni* heptasaccharide, and a variant of $ComP_{ADP1}$. *E. coli* strains expressing the single mutants C95A, C95G, and C95S as well as the double mutants C75A/C95A, C75A/C95G, C75A/C95S, C75G/C95A, C75G/C95G, and C75G/C95S all had ComP levels that were below the level of detection indicating the inherent instability of these mutant proteins and the importance of the Cysteine 75 and Cysteine 95.

[0026] **Figure 2.** **Figure 2** shows a schematic of the recombinant fusion protein containing a C-terminal fragment of ComP from *Acinetobacter soli* strain 110264 (herein referred to as $ComP_{110264}$).

[0027] **Figure 3A,B**. **Figure 3A** and **Figure 3B** show $PglS_{ADP1}$ glycosylating recombinant fusion proteins composed of fragments of $ComP_{110264}$ containing the cysteine residues in position 71 and 93 that flank the previously established site of glycosylation at serine 82. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8

capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, and cysteine 93. Specifically, fusion proteins C1, D1, and E1 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS_{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0028] **Figure 4A,B**. **Figure 4A** and **Figure 4B** show $PglS_{ADP1}$ glycosylating recombinant fusion proteins composed of fragments of $ComP_{110264}$ containing the cysteine residues in position 71 and 93 that flank the previously established site of glycosylation at serine 82. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, and cysteine 93. Specifically, fusion proteins E2, F2, G2, H2, A3, B3, and C3 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS_{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0029] **Figure 5A,B**. **Figure 5A** and **Figure 5B** show $PglS_{ADP1}$ glycosylating recombinant fusion proteins composed of fragments of $ComP_{110264}$ containing the cysteine residues in position 71 and 93 that flank the previously established site of glycosylation at serine 82. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, and cysteine 93. Specifically, fusion proteins

D4, E4, F4, G4, A5, B5, D5, and E5 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS_{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0030] **Figure 6A,B**. **Figure 6A** and **Figure 6B** show $PglS_{ADP1}$ glycosylating recombinant fusion proteins composed of fragments of $ComP_{110264}$ containing the cysteine residues in position 71 and 93 that flank the previously established site of glycosylation at serine 82. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, and cysteine 93. Specifically, fusion proteins F5 and H6 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS_{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0031] **Figure 7A,B**. **Figure 7A** and **Figure 7B** show $PglS_{ADP1}$ glycosylation being blocked by the methionine at position 104 even in the presence of the cysteine 71, serine 82, and cysteine 93. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, cysteine 93 and lacked methionine 104. Specifically, fusion proteins B7, C7, D7, E7, F7, A8, and B8 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be

efficiently glycosylated by $PglS^{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0032] **Figure 8A,B**. **Figure 8A** and **Figure 8B** show that $PglS_{ADP1}$ glycosylation of serine 82 is not blocked by the presence of multiple methionine residues 5' of cysteine 71 and cysteine 93. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, cysteine 93 and lacked methionine 104. Specifically, fusion proteins A10 and B10 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS^{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0033] **Figure 9A,B**. **Figure 9A** and **Figure 9B** show that $PglS_{ADP1}$ glycosylation of serine 82 is not blocked by the presence of multiple methionine residues 5' of cysteine 71 and cysteine 93. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, cysteine 93 and lacked methionine 104. Specifically, fusion proteins C10, D10, F10, G10, H10, A11, B11, and C11 were found to be glycosylated as indicated by the immunoreactive bands running at a higher molecular weight. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS^{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0034] **Figure 10A,B**. **Figure 10A** and **Figure 10B** shows that $PglS_{ADP1}$ glycosylation of serine 82 is blocked by the presence methionine at position 104. **(A)** Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. $PglS_{ADP1}$ was only able to glycosylate those recombinant fusion proteins that contained fragments of $ComP_{110264}$ that contained cysteine 71, serine 82, cysteine 93 and lacked methionine 104. Specifically, none of the fusion proteins were found to be glycosylated by $PglS_{ADP1}$. The "+" sample (SEQ ID NO: 29) acts as a positive control as this fusion protein containing the $ComP_{110264}$ fragment consisting of amino acids 29 to 145 has previously been shown to be efficiently glycosylated by $PglS^{ADP1}$. **(B)** Table format defining the fragment of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence or absence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0035] **Figure 11A,B.** **Figure 11A** and **Figure 11B** show fragments of $ComP_{110264}$ displaying efficient glycosylation by $PglS_{ADP1}$ with the serotype 8 pneumococcal capsular polysaccharide. Western blot analysis of *E. coli* whole cell lysates co-expressing $PglS_{ADP1}$, the pneumococcal serotype 8 capsular polysaccharide, and a fusion protein that contains a fragment of $ComP_{110264}$. Western blots were run in duplicate and probed with either the anti-exotoxin A antisera (**A**) or anti-His antisera (**B**). The different $ComP_{110264}$ fragments all showed similar levels of glycosylation as indicated by the immunoreactive bands running at a higher molecular weight. All fragments contain cysteine 71, serine 82, and cysteine 93.

[0036] **Figure 11C.** **Figure 11C** shows in table format the fragments of $ComP_{110264}$ used for recombinant fusion glycosylation experiment and summarizing western blot observations for the presence of glycosylation. For illustrative purposes, serine 82, the site of known PglS dependent glycosylation, is in bold underlined font.

[0037] **Figure 12A,B,C**. **Figure 12A, Figure 12B,** and **Figure 12C** show that N-terminal or C-terminal O-linked glycosylation motifs translationally fused to the EPA carrier protein are glycosylated in the presence of $PglS_{ADP1}$. **(A)** Figure legend defining the features of each EPA carrier fusion protein used for this experiment. Six different fusion proteins were employed as denoted by the presence of a single *O*-linked glycosylation tag or a double glycosylation tag. **(B)** The D5 and D5' $ComP_{110264}$ amino acid fragment sequences. **(C)** Western blot analysis of *E. coli* whole cell lysates co-expressing the pneumococcal CPS8 and a fusion carrier protein in the

presence or absence of $PglS_{ADP1}$. The D5 and D5' glycosylation motifs, whether N-terminal, C-terminal, in tandem or both N- and C-terminal were all glycosylated only in the presence of $PglS_{ADP1}$.

[0038] **Figure 13A,B,C.** **Figure 13A, Figure 13B,** and **Figure 13C** show that two $ComP_{110264}$ fragments translationally fused in tandem at the C-terminus of a carrier protein are glycosylated with high molecular weight polysaccharides. Fusion proteins were purified from *E. coli* cells co-expressing the pneumococcal serotype 8 capsular polysaccharide in the presence or absence of $PglS_{ADP1}$. (**A**) Western blot analysis of Nickel affinity purified EPA fusion proteins probed with the anti-His antibody shows both the unglycoyslated EPA carrier protein and the higher molecular weight EPA carrier protein glycosylated with the pneumococcal CPS8. (**B**) Western blot analysis of Nickel affinity purified EPA fusion proteins probed with the anti-CPS8 antibody shows the presence of the CPS8 polysaccharide only in samples that co-expressed $PglS_{ADP1}$. In addition, EPA carrier proteins containing two $ComP_{110264}$ fragments lacking the first 28 amino acids ($ComP\Delta28_{110264}$) separated by either a glycine-glycine-glycine-serine (GGGS; SEQ ID NO: 23) or proline-alanine-proline-alanine-proline (PAPAP; SEQ ID NO: 25) linker are glycosylated with high molecular weight pneumococcal CPS8. (**C**) Merged western blot images of **13A** and **13B** showing both anti-His (red channel) and anti-CPS8 (green channel).

[0039] **Figure 14A,B,C.** **Figure 14A, Figure 14B,** and **Figure 14C** show that PglS (**C**), but not PglB (**B**) or PglL (**A**), can conjugate pneumococcal CPS14 to its cognate acceptor/carrier protein. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged acceptor protein variants.

[0040] **Figure 15.** **Figure 15** shows that PglS from *A. baylyi* ADP1 ($PglS_{ADP1}$) can transfer multiple pneumococcal capsular polysaccharides to ComP from *A. baylyi* ADP1 ($ComP_{ADP1}$). Western blot analysis on purified $ComP_{ADP1}$ variants probing for hexa-histidine tagged $ComP_{ADP1}$ variants and either pneumococcal CPS8 (left), CPS9V (middle), or CPS14 (right). Co-localization of the anti-His signals with the anti-glycan signals indicates that $ComP_{ADP1}$ was glycosylated with the correct pneumococcal polysaccharide. The asterisk indicates samples that were treated with proteinase K for 2 hours.

[0041] **Figure 16A,B.** **Figure 16A** and **Figure 16B** show that $PglS_{ADP1}$ can transfer the K1 and K2 capsular polysaccharides of *K. pneumoniae* to $ComP_{ADP1}$. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged $ComP_{ADP1}$ variants and RNA polymerase. RNA polymerase was used as a loading control.

[0042] **Figure 17A.** **Figure 17A** shows mass spectrometry of CPS14-$ComP_{ADP1}$ identified a single glycosylated peptide. ISASNATTNVATAT (SEQ ID NO: 22).

[0043] **Figure 17B.** **Figure 17B** shows mass spectrometry of CPS14-$ComP_{ADP1}$ identified a single glycosylated peptide.

[0044] **Figure 18.** **Figure 18** shows Serine 84 of $ComP_{ADP1}$ is the site of PglS dependent glycosylation. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged $ComP_{ADP1}$ variants and the *Campylobacter jejuni* heptasaccharide. The $ComP[S84A]_{ADP1}$ variant was expressed; however, was not glycosylated as indicated by the absence of any reactive bands probing with the anti-hR6 heptasaccharide antisera.

[0045] **Figure 19.** **Figure 19** lists ComP ortholog amino acid sequences. The site of predicted glycosylation is **bolded**, flanked by a predicted disulfide bond (<u>underlined</u>) linking the predicted alpha beta loop to the beta strand region.

[0046] **Figure 20.** **Figure 20** shows that $PglS_{ADP1}$, but not $PglS_{110264}$, efficiently glycosylates both its cognate $ComP_{ADP1}$ as well as $ComP_{110264}$ from *A. soli* CIP 110264. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged ComP variants and RNA polymerase. RNA polymerase was used as a loading control.

[0047] **Figure 20.** **Figure 21** shows that $PglS_{ADP1}$ efficiently glycosylates DsbA-$ComP\Delta28_{110264}$ fusions but not DsbA-$ComP\Delta28_{ADP1}$ fusions. All fusions either had a triple alanine peptide (AAA; SEQ ID NO: 24) or glycine-glycine-glycine-serine peptide (GGGS; SEQ ID NO: 23) linking DsbA to either a hexa-histidine tagged $ComP\Delta28_{110264}$ or $ComP\Delta28_{ADP1}$. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged ComP variants and RNA polymerase. RNA polymerase was used as a loading control.

[0048] **Figure 22.** **Figure 22** shows that $PglS_{ADP1}$ efficiently glycosylates MBP-$ComP\Delta28_{110264}$ fusions but not MBP-$ComP\Delta28_{ADP1}$ fusions. All fusions either had a triple alanine peptide (AAA; SEQ ID NO: 24) or glycine-glycine-glycine-serine peptide (GGGS; SEQ ID NO: 23) linking maltose binding protein (MBP) to either a hexa-histidine tagged $ComP\Delta28_{110264}$ or $ComP\Delta28_{ADP1}$. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged ComP variants and RNA polymerase. RNA polymerase was used as a loading control.

[0049] **Figure 23.** **Figure 23** $PglS_{ADP1}$, but not $PglS_{110264}$, efficiently EPA-GGGS-$ComP\Delta28_{110264}$ fusions. Western blot analysis on *E. coli* whole cell lysates or periplasmic extracts probing for hexa-histidine tagged ComP variants. EPA-GGGS – exotoxin A with a

glycine-glycine-glycine-serine peptide (GGGS; SEQ ID NO: 23) linking a hexa-histidine tagged ComP$\Delta 28_{110264}$ variant.

[0050] Figure 24. **Figure 24** shows amino acid sequences of representative ComP$\Delta 28_{110264}$ fusion proteins.

[0051] Figure 25A,B,C. **Figure 25A, Figure 25B, and Figure 25C** show that a monovalent CPS14-ComP$_{ADP1}$ bioconjugate vaccine induces serotype specific IgG antibodies.

[0052] Figure 26. **Figure 26** shows that a trivalent bioconjugate vaccine against serotypes 8, 9V, and 14 induces serotype specific IgG titers at comparable levels to Prevnar 13.

[0053] Figure 27. **Figure 27** lists ComP $\Delta 28$ ortholog amino acid sequences in which the amino acids corresponding to the 28 N-terminal amino acids of SEQ ID NO: 1 (ComP$_{ADP1}$: AAC45886.1) have been removed. The site of predicted glycosylation is **bolded**, flanked by a predicted disulfide bond (<u>underlined</u>) linking the predicted alpha beta loop to the beta strand region.

[0054] Figure 28. **Figure 28** shows an alignment of a region ComP sequences including the serine (S) residue (boxed) corresponding to the serine residue at position 84 of SEQ ID NO: 1 (ComP$_{ADP1}$: AAC45886.1).

[0055] Figure 29. **Figure 29** shows higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested CPS14-ComP bioconjugates. GluC digested CPS14-ComP was subjected to HCD fragmentation enabling the confirmation of a semi-GluC derived single peptide attached to a glycan with the CPS14 repeating subunit. Additional glycopeptides were also observed decorated with extended glycans corresponding to up to four tetrasaccharide repeat units.

[0056] Figure 30. **Figure 30** shows higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested ComP glycosylated with the *C. jejuni* heptasaccharide (ComP-Glycan$_{Cj}$). GluC digested ComP-Glycan$_{Cj}$ was subjected to HCD fragmentation enabling the confirmation of a single peptide attached to a glycan with the CPS14 repeating subunit. Low collision energies regimes were undertaken to confirm the glycosylation of the peptide ISASNATTNVATAT (SEQ ID NO: 22) with a 1380.53 Da glycan corresponding to 6*HexNAc,1*Hexose.

[0057] Figure 31. **Figure 31** shows higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested ComP glycosylated with the C. jejuni heptasaccharide (ComP-Glycan$_{Cj}$). GluC digested ComP-Glycan$_{Cj}$ was subjected to HCD fragmentation enabling

the confirmation of a single peptide attached to a glycan with the CPS14 repeating subunit. High collision energies regimes were undertaken to confirm the glycosylation of the peptide ISASNATTNVATAT (SEQ ID NO: 22) with a 1380.53 Da glycan corresponding to 6*HexNAc,1*Hexose.

[0058] **Figure 32A-I. Figure 32A-I** shows that the oligosaccharyltransferase PglS can glycosylate the acceptor protein ComP with the pneumococcal CPS14 polysaccharide. *E. coli* SDB1 cells co-expressing an acceptor protein (DsbA, AcrA, or ComP), an OTase (PglL, PglB, or PglS), and the CPS14 polysaccharide were analyzed for protein glycosylation via western blot analysis of the affinity purified acceptor proteins. (**A-C**): DsbA purified from SDB1 cells in the presence or absence of PglL. (**A**): Anti-His channel probing for hexa-histidine tagged DsbA. (**B**): Anti-glycan channel probing for CPS14. (**C**): Merged images for panels A and B. (**D-F**): AcrA purified from SDB1 cells in the presence or absence of PglB. (**D**): Anti-His channel probing for hexa-histidine tagged AcrA. (**E**): Anti-glycan channel probing for CPS14. (**F**): Merged images for panels D and E. (**G-I**): ComP purified from SDB1 cells in the presence or absence of PglS. (**G**): Anti-His channel probing for hexa-histidine tagged ComP. (**H**): Anti-glycan channel probing for CPS14. (**I**): Merged images for panels G and H. The asterisk indicates samples that were proteinase K treated for 1h at 55°C.

[0059] **Figure 33A,B.** **Figure 33A** and **Figure 33B** show higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested CPS14-ComP bioconjugates. GluC digested CPS14-ComP was subjected to HCD fragmentation enabling the confirmation of a single peptide attached to a glycan with the CPS14 repeating subunit. High collision energies (**A**) and low collision energies (**B**) regimes were undertaken to confirm the glycosylation of the peptide ISASNATTNVATAT (SEQ ID NO: 22) with a 1378.47 Da glycan corresponding to HexNAc2Hexose6.

[0060] **Figure 34A-F.** **Figure 34A-F** shows Western blot analysis of CPS8-ComP and CPS9V-ComP glycoproteins. *E. coli* SDB1 cells were prepared co-expressing ComP, PglS, and either the pneumococcal CPS8 or CPS9V. Affinity purified glycosylated ComP from each strain was analyzed for protein glycosylation via western blot analysis. (**A-C**): Western blot analysis of CPS8-ComP bioconjugates compared against ComP alone (**A**): Anti-His channel probing for hexa-histidine tagged ComP purified from SDB1 expressing CPS8 in the presence or absence of PglS. (**B**): Anti-glycan channel probing for CPS8. (**C**): Merged images for panels A and B. (**D-F**): Western blot analysis of CPS9V-ComP bioconjugates compared against ComP alone (**D**):

Anti-His channel probing for hexa-histidine tagged ComP purified from SDB1 expressing CPS9V in the presence or absence of PglS. (**E**): Anti-glycan channel probing for CPS9V. (**F**): Merged images for panels D and E. The asterisk indicates samples that were proteinase K treated for 1h at 55˚C.

[0061] **Figure 35A-F.** **Figure 35A-F** shows IgG responses of mice vaccinated with ComP, PREVNAR 13®, a monovalent CPS14-ComP bioconjugate and a trivalent CPS8-/CPS9V-/CPS14-ComP bioconjugate. Groups of mice were vaccinated with ComP alone, PREVNAR 13®, a monovalent CPS14-ComP bioconjugate vaccine, or a CPS8-/CPS9V-/CPS14-ComP bioconjugate vaccine. Sera was collected on day 49 and analyzed for serotype specific IgG responses via ELISA compared against sera collected on day 0. (**A-C**): No detectable increases in IgG responses were detected in placebo vaccinated mice for serotypes 8 (**A**), 9V (**B**), or 14 (**C**). (D-F): PREVNAR 13® vaccinated mice did not have detectable IgG responses titer increases to serotype 8 (**D**), but did have IgG responses increases in IgG titers specific to serotype 9V (**E**) and 14 (**F**). Unpaired t-tests (Mann-Whitney) were performed to statistically analyze pre-immune sera from day 49 sera. P values for each case tested were **** p=0.0001. Each dot represents a single vaccinated mouse. Error bars indicate the standard deviation of the mean.

[0062] **Figure 35G-L.** **Figure 35G-L** shows IgG responses of mice vaccinated with ComP, PREVNAR 13®, a monovalent CPS14-ComP bioconjugate and a trivalent CPS8-/CPS9V-/CPS14-ComP bioconjugate. Groups of mice were vaccinated with ComP alone, PREVNAR 13®, a monovalent CPS14-ComP bioconjugate vaccine, or a CPS8-/CPS9V-/CPS14-ComP bioconjugate vaccine. Sera was collected on day 49 and analyzed for serotype specific IgG responses via ELISA compared against sera collected on day 0. (**G-I**): Mice vaccinated with a CPS14-ComP bioconjugate vaccine did not have IgG responses detectable increases in IgG titers specific to serotypes 8 (**G**) or 9V (**H**), but did have IgG responses statistically significant IgG titer increases to serotype 14 (**I**). (**J-L**): Trivalent CPS8-/CPS9V-/CPS14-ComP bioconjugate vaccinated mice all had statistically significant IgG responses increases in IgG titers to serotypes 8 (**J**), 9V (**K**), and 14 (**L**). Unpaired t-tests (Mann-Whitney) were performed to statistically analyze pre-immune sera from day 49 sera. P values for each case tested were **** p=0.0001. Each dot represents a single vaccinated mouse. Error bars indicate the standard deviation of the mean.

[0063] **Figure 36A,B.** **Figure 36A** and **Figure 36B** shows bactericidal activity of sera from vaccinated mice against *S. pneumoniae* serotypes 8 and 14. Opsonophagocytosis assays

(OPA) of sera from mice vaccinated with either buffer control, PREVNAR 13®, or bioconjugate vaccine against both *S. pneumoniae* serotypes 8 (**A**) and 14 (**B**). Serotype-specific commercial rabbit anti-*S. pneumoniae* sera were used as positive controls. A 5% (v/v) sample serum and a bacterial MOI of 0.01 were added to fresh whole blood from naive mice to perform the assay. Viable bacterial counts were performed after 4 h of incubation. To determine bacterial killing, viable bacterial counts from tubes incubated with sample sera were compared to those incubated with control naive mouse sera. Results are expressed as percent bacterial killing for individual mice, with horizontal bars representing the standard deviation of the mean.

[0064] Figure 37A,B. **Figure 37A** and **Figure 37B** shows analysis of EPA glycosylation with the CPS8 capsular polysaccharide. Western blot analysis of EPA-CPS8 bioconjugates compared against EPA alone. (**A** - Left panel) Anti-His channel probing for hexa-histidine tagged EPA purified from SDB1 expressing CPS8 in the presence or absence of PglS. (**A** – Middle panel) Anti-glycan channel probing for CPS8. (**A** – Right panel) Merged images for left and middle panels. (**B**): EPA-CPS8 separated on a SDS polyacrylamide gel stained with Coomassie.

[0065] Figure 37C. **Figure 37C** shows intact protein mass spectrometry analysis showing the MS1 mass spectra for purified EPA-CPS8. The EPA fusion protein has a theoretical mass of 79,526.15 Daltons and can be observed as the peak at 79,514.76. The EPA fusion protein was also observed in multiple states of increasing mass corresponding to the CPS8 repeating subunit, which has a theoretical mass of 662 Daltons. Varying glycoforms of the EPA-CPS8 were observed and are denoted by "$g^{numeric}$" where "g" stands for glycoform and the "numeric" corresponds to the number of repeating CPS8 subunits. The EPA fusion protein was modified with up to 11 repeating subunits of the CPS8 glycan. Panel **D** provides a zoomed in view of the varying EPA-CPS8 glycoforms.

[0066] Figure 37D. **Figure 37D** provides a zoomed in view of the varying EPA-CPS8 glycoforms from **Figure 37C**.

[0067] Figure 38A,B. **Figure 38A** and **Figure 38B** shows analysis of immune responses to ComP-CPS8 and EPA-CPS8 bioconjugates in mice. (**A**): Titers of CPS8 IgG antibodies in mice immunized with CPS8 bioconjugate vaccines. Mouse groups were as follows: EPA (n= 9, mice vaccinated with 5 μg of total protein), ComP-CPS8 (n=10, mice vaccinated with 5 μg total polysaccharide), and EPA-CPS8 (n=10, mice vaccinated with 100 ng of total polysaccharide). All mice were immunized with 100 μL of a vaccine diluted 1:1 with Imject Alum Adjuvant on days 1, 14, and 28. Sera were collected on day 4. For the titration, ELISA plates were coated with

whole cell serotype 8 pneumococci and incubated with 2-fold serial dilutions of sera. Titers for individual mice are shown, with horizontal bars representing the standard error of the mean. Statistically significant titers compared to the EPA placebo group are denoted with asterisk and were determined using Kruskal-Wallis one-way Anova. **, P=0.0223 and ****, P<0.0001. For analysis and representation purposes, negative titer values (<100) were given an arbitrary value of 10. (**B**): Opsonophagocytosis killing of *S. pneumoniae* serotype 8 by day 42 sera from mice immunized with ComP-CPS8 and EPA-CPS8 bioconjugate vaccines. The same mouse groups described for the IgG titers were employed for the OPA.A 40% (vol/vol) sample of serum and bacterial MOI of 0.01 were added to fresh whole blood from naïve mice to perform the assay. Results are expressed as percent bacterial killing for individual mice, with horizontal bars representing the standard error of the mean. Statistically significant killing compared to the EPA placebo group is denoted with asterisk and were determined using Kruskal-Wallis one-way Anova. **, P= 0.0015.

[0068] **Figure 39A,B,C.** **Figure 39A, Figure 39B,** and **Figure 39C** shows that a conserved and homologous serine is believed to be the site of glycosylation in ComP proteins from *A. baylyi* ADP1 and *A. soli* 110264. Serines 82 and 84 of $ComP_{ADP1}$ and the homologous serines 79 and 82 of $ComP_{110264}$ were mutated to an alanine and probed for glycosylation in the presence of PglS and the serotype 8 capsular polysaccharide. (**A-C**) SDB1 cells expressing ComP variants in the presence of PglS and CPS8 were probed via western blotting for protein glycosylation. (**A**) Anti-His channel probing for ComP expression and glycosylation. (**B**) Anti-glycan channel probing for CPS8. (**C**) Merged image for panels A and B.

[0069] **Figure 40A,B,C.** **Figure 40A, Figure 40B,** and **Figure 40C** shows an analysis of EPA glycosylation with the *Klebsiella pneumoniae* K1 and K2 capsular polysaccharides. Western blot analysis of purified the (**A**) non-glycosylated EPA, (**B**) EPA glycosylated with the *K. pneumoniae* K1 capsular polysaccharide, or (**C**) EPA glycosylated with the *K. pneumoniae* K2 capsular polysaccharide. The "g^0" denotes the non-glycosylated EPA fusion and "g^n" denotes the EPA fusion glycosylated with different sized K1 or K2 repeating subunits as depicted in panel B or C, respectively.

[0070] **Figure 41A,B.** **Figure 41A** and **Figure 41B** shows intact protein mass spectrometry analysis showing the MS1 mass spectra for purified EPA-K2. The EPA fusion protein has a theoretical mass of 79,526.15 Daltons and can be observed as the peak at 79,518.73. The EPA fusion protein was also observed in multiple states of increasing mass corresponding to

the *K. pneumoniae* K2 capsular polysaccharide repeating subunit, which has a theoretical mass of 662 Daltons. (**A**) Varying glycoforms of the EPA-K2 were observed and are denoted by "$g^{numeric}$" where "g" stands for glycoform and "numeric" corresponds to the number of repeating K2 subunits. The EPA fusion protein was modified with up to 11 repeating subunits of the K2 capsule. (**B**) A zoomed in view of A is also provided.

DETAILED DESCRIPTION

[0071] To the extent necessary to provide descriptive support, the subject matter and/or text of the appended claims is incorporated herein by reference in their entirety.

[0072] It will be understood by all readers of this written description that the exemplary aspects and embodiments described and claimed herein may be suitably practiced in the absence of any recited feature, element or step that is, or is not, specifically disclosed herein.

Definitions.

[0073] It is to be noted that the term "a" or "an" entity refers to one or more of that entity; for example, "a polysaccharide," is understood to represent one or more polysaccharides. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

[0074] Furthermore, "and/or" where used herein is to be taken as specific disclosure of each of the specified features or components with or without the other. Thus, the term and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following embodiments: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[0075] It is understood that wherever aspects are described herein with the language "comprising" or "comprises" otherwise analogous aspects described in terms of "consisting of," "consists of," "consisting essentially of," and/or "consists essentially of," and the like are also provided.

[0076] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure is related.

[0077] Numeric ranges are inclusive of the numbers defining the range. Even when not explicitly identified by "and any range in between," or the like, where a list of values is recited,

e.g., 1, 2, 3, or 4, unless otherwise stated, the disclosure specifically includes any range in between the values, *e.g.*, 1 to 3, 1 to 4, 2 to 4, etc.

[0078] The headings provided herein are solely for ease of reference and are not limitations of the various aspects or aspects of the disclosure, which can be had by reference to the specification as a whole.

[0079] As used herein, the term "non-naturally occurring" substance, composition, entity, and/or any combination of substances, compositions, or entities, or any grammatical variants thereof, is a conditional term that explicitly excludes, but only excludes, those forms of the substance, composition, entity, and/or any combination of substances, compositions, or entities that are well-understood by persons of ordinary skill in the art as being "naturally-occurring," or that are, or might be at any time, determined or interpreted by a judge or an administrative or judicial body to be, "naturally-occurring."

[0080] As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids are included within the definition of "polypeptide," and the term "polypeptide" can be used instead of, or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-standard amino acids. A polypeptide can be derived from a natural biological source or produced by recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It can be generated in any manner, including by chemical synthesis.

[0081] A "protein" as used herein can refer to a single polypeptide, i.e., a single amino acid chain as defined above, but can also refer to two or more polypeptides that are associated, e.g., by disulfide bonds, hydrogen bonds, or hydrophobic interactions, to produce a multimeric protein.

[0082] By an "isolated" polypeptide or a fragment, variant, or derivative thereof is intended a polypeptide that is not in its natural milieu. No particular level of purification is required. For example, an isolated polypeptide can be removed from its native or natural environment.

Recombinantly produced polypeptides and proteins expressed in host cells are considered isolated as disclosed herein, as are recombinant polypeptides that have been separated, fractionated, or partially or substantially purified by any suitable technique.

[0083] As used herein, the term "non-naturally occurring" polypeptide, or any grammatical variants thereof, is a conditional term that explicitly excludes, but only excludes, those forms of the polypeptide that are well-understood by persons of ordinary skill in the art as being "naturally-occurring," or that are, or might be at any time, determined or interpreted by a judge or an administrative or judicial body to be, "naturally-occurring."

[0084] Disclosed herein are certain binding molecules, or antigen-binding fragments, variants, or derivatives thereof. Unless specifically referring to full-sized antibodies such as naturally-occurring antibodies, the term "binding molecule" encompasses full-sized antibodies as well as antigen-binding fragments, variants, analogs, or derivatives of such antibodies, e.g., naturally-occurring antibody or immunoglobulin molecules or engineered antibody molecules or fragments that bind antigen in a manner similar to antibody molecules.

[0085] As used herein, the term "binding molecule" refers in its broadest sense to a molecule that specifically binds an antigenic determinant. As described further herein, a binding molecule can comprise one of more "binding domains." As used herein, a "binding domain" is a two- or three-dimensional polypeptide structure that cans specifically bind a given antigenic determinant, or epitope. A non-limiting example of a binding molecule is an antibody or fragment thereof that comprises a binding domain that specifically binds an antigenic determinant or epitope. Another example of a binding molecule is a bispecific antibody comprising a first binding domain binding to a first epitope, and a second binding domain binding to a second epitope.

[0086] The terms "antibody" and "immunoglobulin" can be used interchangeably herein. An antibody (or a fragment, variant, or derivative thereof as disclosed herein comprises at least the variable domain of a heavy chain and at least the variable domains of a heavy chain and a light chain. Basic immunoglobulin structures in vertebrate systems are relatively well understood. See, e.g., Harlow et al., Antibodies: A Laboratory Manual, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988).

[0087] Binding molecules, e.g., antibodies or antigen-binding fragments, variants, or derivatives thereof include, but are not limited to, polyclonal, monoclonal, human, humanized, or chimeric antibodies, single chain antibodies, epitope-binding fragments, e.g., Fab, Fab' and F(ab')2, Fd, Fvs, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv),

fragments comprising either a VL or VH domain, fragments produced by a Fab expression library. ScFv molecules are known in the art and are described, e.g., in US patent 5,892,019. Immunoglobulin or antibody molecules encompassed by this disclosure can be of any type (e.g., IgG, IgE, IgM, IgD, IgA, and IgY), class (e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2) or subclass of immunoglobulin molecule.

[0088] By "specifically binds," it is meant that a binding molecule, e.g., an antibody or fragment, variant, or derivative thereof binds to an epitope via its antigen binding domain, and that the binding entails some complementarity between the antigen binding domain and the epitope. According to this definition, a binding molecule is said to "specifically bind" to an epitope when it binds to that epitope, via its antigen-binding domain more readily than it would bind to a random, unrelated epitope. The term "specificity" is used herein to qualify the relative affinity by which a certain binding molecule binds to a certain epitope. For example, binding molecule "A" can be deemed to have a higher specificity for a given epitope than binding molecule "B," or binding molecule "A" can be said to bind to epitope "C" with a higher specificity than it has for related epitope "D."

[0089] The term "bispecific antibody" as used herein refers to an antibody that has binding sites for two different antigens within a single antibody molecule. It will be appreciated that other molecules in addition to the canonical antibody structure can be constructed with two binding specificities. It will further be appreciated that antigen binding by bispecific antibodies can be simultaneous or sequential. Triomas and hybrid hybridomas are two examples of cell lines that can secrete bispecific antibodies. Bispecific antibodies can also be constructed by recombinant means. (Ströhlein and Heiss, Future Oncol. 6:1387-94 (2010); Mabry and Snavely, IDrugs. 13:543-9 (2010)). A bispecific antibody can also be a diabody.

[0090] The term "polynucleotide" is intended to encompass a singular nucleic acid as well as plural nucleic acids, and refers to an isolated nucleic acid molecule or construct, e.g., messenger RNA (mRNA) or plasmid DNA (pDNA). A polynucleotide can comprise a conventional phosphodiester bond or a non-conventional bond (e.g., an amide bond, such as found in peptide nucleic acids (PNA)). The term "nucleic acid" refers to any one or more nucleic acid segments, e.g., DNA or RNA fragments, present in a polynucleotide. By "isolated" nucleic acid or polynucleotide is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, a recombinant polynucleotide encoding a polypeptide subunit contained in a vector is considered isolated as disclosed herein. Further examples of an

isolated polynucleotide include recombinant polynucleotides maintained in heterologous host cells or purified (partially or substantially) polynucleotides in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of polynucleotides. Isolated polynucleotides or nucleic acids further include such molecules produced synthetically. In addition, polynucleotide or a nucleic acid can be or can include a regulatory element such as a promoter, ribosome binding site, or a transcription terminator.

[0091] As used herein, a "non-naturally occurring" polynucleotide, or any grammatical variants thereof, is a conditional definition that explicitly excludes, but only excludes, those forms of the polynucleotide that are well-understood by persons of ordinary skill in the art as being "naturally-occurring," or that are, or that might be at any time, determined or interpreted by a judge or an administrative or judicial body to be, "naturally-occurring."

[0092] In certain embodiments, the polynucleotide or nucleic acid is DNA. In other embodiments, a polynucleotide can be RNA.

[0093] A "vector" is nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker gene and other genetic elements known in the art.

[0094] A "transformed" cell, or a "host" cell, is a cell into which a nucleic acid molecule has been introduced by molecular biology techniques. As used herein, the term transformation encompasses those techniques by which a nucleic acid molecule can be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration. A transformed cell or a host cell can be a bacterial cell or a eukaryotic cell.

[0095] The term "expression" as used herein refers to a process by which a gene produces a biochemical, for example, a polypeptide. The process includes any manifestation of the functional presence of the gene within the cell including, without limitation, gene knockdown as well as both transient expression and stable expression. It includes without limitation transcription of the gene into messenger RNA (mRNA), and the translation of such mRNA into polypeptide(s). If the final desired product is a biochemical, expression includes the creation of that biochemical and any precursors. Expression of a gene produces a "gene product." As used herein, a gene product can be either a nucleic acid, e.g., a messenger RNA produced by transcription of a gene, or a polypeptide that is translated from a transcript. Gene products

described herein further include nucleic acids with post transcriptional modifications, e.g., polyadenylation, or polypeptides with post translational modifications, e.g., methylation, glycosylation, the addition of lipids, association with other protein subunits, proteolytic cleavage, and the like.

[0096] As used herein the terms "treat," "treatment," or "treatment of" (e.g., in the phrase "treating a subject") refers to reducing the potential for disease pathology, reducing the occurrence of disease symptoms, e.g., to an extent that the subject has a longer survival rate or reduced discomfort. For example, treating can refer to the ability of a therapy when administered to a subject, to reduce disease symptoms, signs, or causes. Treating also refers to mitigating or decreasing at least one clinical symptom and/or inhibition or delay in the progression of the condition and/or prevention or delay of the onset of a disease or illness.

[0097] By "subject" or "individual" or "animal" or "patient" or "mammal," is meant any subject, particularly a mammalian subject, for whom diagnosis, prognosis, or therapy is desired. Mammalian subjects include humans, domestic animals, farm animals, sports animals, and zoo animals, including, e.g., humans, non-human primates, dogs, cats, guinea pigs, rabbits, rats, mice, horses, cattle, bears, and so on.

[0098] The term "pharmaceutical composition" refers to a preparation that is in such form as to permit the biological activity of the active ingredient to be effective, and that contains no additional components that are unacceptably toxic to a subject to which the composition would be administered. Such composition can be sterile.

[0099] An "effective amount" of an antibody as disclosed herein is an amount sufficient to carry out a specifically stated purpose. An "effective amount" can be determined empirically and in a routine manner, in relation to the stated purpose.

Overview.

[0100] Conjugate vaccines, consisting of a polysaccharide linked to a protein, are lifesaving prophylactics. Traditionally, conjugate vaccines are manufactured using chemical methodologies. However, *in vivo* bacterial conjugations have emerged as manufacturing alternatives. *In vivo* conjugation (bioconjugation) is reliant upon an oligosaccharyltransferase to attach polysaccharides to proteins. Currently, the oligosaccharyltransferases employed for bioconjugations are not suitable for the generation of conjugate vaccines when the polysaccharides contain glucose at the reducing end. This limitation has enormous implications as ~75% of *Streptococcus pneumoniae* capsules contain glucose as the reducing end sugar.

Disclosed herein is the use of an O-linked oligosaccharyltransferase to generate the first ever polyvalent pneumococcal bioconjugate vaccine with polysaccharides containing glucose at their reducing end. Pneumococcal bioconjugates were immunogenic, protective, and rapidly produced with recombinant techniques. Certain aspects disclosed herein provide for the engineering, characterization, and immunological responses of a polyvalent pneumococcal bioconjugate vaccine using the natural acceptor protein ComP as a vaccine carrier as well as a monovalent pneumococcal bioconjugate vaccine using a conventional vaccine carrier; e.g., in certain aspects, containing the *Pseudomonas aeruginosa* exotoxin A protein. This establishes a platform to overcome limitations of other conjugating enzymes enabling the development of bioconjugate vaccines for many important human and animal pathogens.

[0101] Even with the introduction and implementation of pneumococcal conjugate vaccines over the last two decades, ~1.5 million deaths are still attributed to *S. pneumoniae* each year. This is due in part to the 90+ serotypes of *S. pneumoniae* and the complex manufacturing methods required to synthesize pneumococcal conjugate vaccines. Together these factors hinder global distribution and development of broader, more protective variations of the vaccines. To expedite development and lower manufacturing costs, disclosed herein is a platform for developing conjugate vaccines, for example pneumococcal conjugate vaccines, using *in vivo* conjugation. This streamlined process has the potential to complement existing manufacturing pipelines or completely bypass the dependency on chemical conjugation methodologies, enabling the production of a more comprehensive conjugate vaccines.

[0102] Traditional, chemical conjugate vaccine synthesis is considered complex, costly, and laborious (Frasch, C.E. *Vaccine* 27, 6468-6470 (2009)) however, *in vivo* conjugation has been thoroughly progressing as a viable biosynthetic alternative (Huttner, A. et al. *Lancet Infect Dis* 17, 528-537 (2017)). These strides are best highlighted by the successes of GlycoVaxyn, (now LimmaTech Biologics AG an independent company with direct ties to GlaxoSmithKline), a clinical stage biopharmaceutical company with multiple bioconjugate vaccines in various phases of clinical trials, one of which (Flexyn2a) has just completed a Phase 2b challenge study. Although GlycoVaxyn has been at the forefront of the *in vivo* conjugation revolution, the ability to glycosylate carrier/acceptor proteins with polysaccharides containing glucose (Glc) as the reducing end sugar has been elusive and, expectedly, has stymied the development of a pneumococcal bioconjugate vaccine.

[0103] The oligosaccharyltransferase PglS – previously referred to as PglL by Schulz *et al.* (PMID23658772) and $PglL_{ComP}$ by Harding *et al.* 2015 (PMID 26727908) – was only recently characterized as a functional OTase (Schulz, B.L. et al. *PLoS One* 8, e62768 (2013)). Subsequent mass spectrometry studies on total glycopeptides demonstrated that PglS does not act as a general PglL-like OTase, glycosylating multiple periplasmic and outer membrane proteins (Harding, C.M. et al. *Mol Microbiol* 96, 1023-1041 (2015)). In fact, the genome of *A. baylyi* ADP1 encodes for two OTase, a PglL-like ortholog (UniProtKB/Swiss-Prot: Q6FFS6.1), which acts as the general OTase and PglS (UniProtKB/Swiss-Prot: Q6F7F9.1), which glycosylates a single protein, ComP (Harding, C.M. et al. *Mol Microbiol* 96, 1023-1041 (2015)).

[0104] ComP is orthologous to type IV pilin proteins, like PilA from *Pseudomonas aeruginosa* and PilE from *Neisseria meningiditis,* both of which are glycosylated by the OTases TfpO (Castric, P. *Microbiology* 141 (Pt 5), 1247-1254 (1995)) and PglL (Power, P.M. et al. *Mol Microbiol* 49, 833-847 (2003)), respectively. Although TfpO and PglL also glycosylate their cognate pilins at serine residues, the sites of glycosylation differ between each system. TfpO glycosylates its cognate pilin at a C-terminal serine residue (Comer, J.E., Marshall, M.A., Blanch, V.J., Deal, C.D. & Castric, P. *Infect Immun* 70, 2837-2845 (2002)), which is not present in ComP. PglL glycosylates PilE at an internal serine located at position 63 (Stimson, E. et al. *Mol Microbiol* 17, 1201-1214 (1995)). ComP also contains serine residues near position 63 and the surrounding residues show moderate conservation to PilE from *N. meningiditis*. Comprehensive glycopeptide analysis, however, revealed this serine and the surrounding residues were not the site of glycosylation in ComP. PglS glycosylates ComP at a single serine residue located at position corresponding to the conserved serine at position 84 of $ComP_{ADP1}$: AAC4588631 (SEQ ID NO: 1) (also corresponding to the conserved serine at position 82 of $ComP_{110264}$: ENV58402.1 (SEQ ID NO: 2)), which is a novel glycosylation site not previously found within the type IV pilin superfamily. The ability of PglS to transfer polysaccharides containing glucose as the reducing end sugar coupled with the identification of a novel site of glycosylation within the pilin superfamilies demonstrates that PglS is a functionally distinct OTase from PglL and TfpO.

[0105] **PglS, but not PglB or PglL, transferred polysaccharides containing glucose at their reducing end to the acceptor protein ComP.** Two classes of OTases, PglB and PglL, have previously been employed for *in vivo* conjugation (Feldman, M.F. et al. *Proc Natl Acad Sci USA* 102, 3016-3021 (2005); Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S. &

Feldman, M.F. *J Bacteriol* 189, 8088-8098 (2007)). PglB, the first OTase described, preferentially transfers glycans containing an acetamido-group at the C-2 position of the reducing end (i.e. *N*-acetylglucosamine), as it is believed to play a role in substrate recognition (Wacker, M. et al. *Proc Natl Acad Sci USA* 103, 7088-7093 (2006)). However, polysaccharides with galactose (Gal) at the reducing end, such as the *S. enterica* Typhimurium O antigen, can be transferred by an engineered PglB variant (Ihssen, J. et al. *Open Biol* 5, 140227 (2015)). The second described OTase, PglL from *N. meningiditis*, has more relaxed substrate specificity than PglB, naturally transferring polysaccharides with an acetamido-group at the C-2 position as well as polysaccharides containing galactose (Gal) at the reducing end (Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S. & Feldman, M.F. *J Bacteriol* 189, 8088-8098 (2007); Pan, C. et al. *MBio* 7 (2016)). However, there is no evidence available for PglB or PglL mediated transfer of polysaccharides containing glucose (Glc) at the reducing end, which is of particular interest given that the majority of pneumococcal CPSs contain glucose at the reducing end (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)). The ability of PglB and PglL to transfer the pneumococcal serotype 14 capsular polysaccharide (CPS14) to their cognate glycosylation targets, AcrA (Wacker, M. et al. *Science* 298, 1790-1793 (2002)) and DsbA (Vik, A. et al. *Proc Natl Acad Sci USA* 106, 4447-4452 (2009)), respectively, was tested. As seen in **Figure 14A** and **Figure 14B**, both acceptor proteins were expressed; however, no evidence for CPS14 glycosylation to either acceptor protein was observed.

[0106] *Acinetobacter* species have been describes as containing three O-linked OTases; a general PglL OTase responsible for glycosylating multiple proteins, and two pilin-specific OTases (Harding, C.M. *Mol Microbiol* 96, 1023-1041 (2015)). The first pilin-specific OTase is an ortholog of TfpO (also known as PilO) and is not employed for *in vivo* conjugation systems due to its inability to transfer polysaccharides with more than one repeating unit (Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen J.S. & Feldman, M.F. *J Bacteriol* 189, 8088-8098 (2007)). The second pilin specific OTase, PglS glycosylates a single protein, the type IV pilin ComP[28]. A bioinformatic analysis indicated that PglS is the archetype of a distinct family of OTases. Given that PglS represents a new class of *O*-OTase, its ability to transfer pneumococcal CPS14 to its cognate acceptor protein, ComP (Harding, C. M. et al. *Mol Microbiol* 96, 1023-1041 (2015)) was tested. As seen in **Figure 14C**, co-expression of the CPS14 biosynthetic locus in conjunction with PglS and a hexa-his tagged variant of ComP resulted in a typical ladder-like pattern of bands compatible with protein glycosylation when analyzed via western blotting

(**Figure 14B**). The higher molecular weight, modal distribution of signals is indicative of protein glycosylation with repeating glycan subunits of increasing molecular weight. Together, these results indicate that, unlike the previously characterized OTases, PglS is able to transfer polysaccharides with glucose at the reducing end.

[0107] There are more than 90 serotypes of *S. pneumoniae* (Geno, K. A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)). Many increasingly prevalent serotypes, like serotypes 8, 22F, and 33F are not included in currently licensed vaccines. Therefore, the versatility was tested of PglS to generate a multivalent pneumococcal bioconjugate vaccine against two serotypes included in Prevnar 13 (serotype 9V and 14) and one serotype not included (serotype 8) (Package Insert-Prevnar 13 - FDA, on the world wide web at fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM201669.pdf)). Importantly, all of three of these capsular polysaccharides contain glucose as the reducing end sugar (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)). As seen in **Figure 15**, western blot analysis of affinity purified proteins from whole cells co-expressing PglS, a hexa-his tagger ComP variant, and either CPS8, CPS9V, or CPS14 resulted in the generation CPS-specific bioconjugates. Moreover, antisera specific to either the CPS8, CPS9V, or CPS14 antigens also reacted to the anti-His reactive bands, indicating that ComP-His was glycosylated with the correct polysaccharides. To confirm that the material purified was not contaminated with lipid-linked polysaccharides, the samples were treated with proteinase K and observed a loss of signal when analyzed via western blotting, confirming that the bioconjugates were proteinaceous.

[0108] Therefore, it was demonstrated that PglS can transfer *S. pneumoniae* polysaccharides to ComP, wherein PglB and PglL could not. Specifically, PglS is the only OTase in the known universe capable of transferring polysaccharides with glucose at the reducing end. In certain aspects, PglS can be used to transfer any lipid-linked oligosaccharide or polysaccharide (collectively referred to herein as "oligo- or polysaccharide") containing glucose at the reducing end to ComP or a fusion protein containing a fragment of ComP.

[0109] **PglS can transfer capsular polysaccharides of *Klebsiella* to ComP.** *Klebsiella pneumonia* (*K. pneumoniae*), a Gram negative opportunistic human pathogen, produces a capsular polysaccharide known to be important for virulence. To date at least 79 antigenically distinct capsular polysaccharides have been described for *Klebsiella* species (Pan, Y.J. et al. *Sci Rep* 5, 15573 (2015)). Furthermore, *K. pneumoniae* is known to produce at least 59 of the 77 capsular polysaccharides, more than half of which contain glucose as the reducing end sugar

(Pan, Y.J. et al. *Sci Rep* 5, 15573 (2015)). To determine if PglS could transfer *K. pneumoniae* capsular polysaccharides to ComP, the genes encoding for the proteins required for the synthesis of either the K1 or the K2 capsular polysaccharides were cloned into the IPTG inducible pBBR1MCS-2 vector (Kovach, M.E. et al. *Gene* 166, 175-176 (1995)). The K1 capsule gene locus was cloned from *K. pneumoniae* NTUH K-2044, a previously characterized K1 capsule producing strain (Wu, K.M. et al. *J Bacteriol* 191, 4492-4501 (2009)). The K2 capsule gene locus was cloned from *K. pneumoniae* 52.145, a previously characterized K2 capsule producing strain (Lery, L. M. et al. *BMC Biol* 12, 41 (2014)). The K1 or the K2 capsular polysaccharide expressing plasmids were then individually introduced into *E. coli* co-expressing PglS OTase and the acceptor protein ComP from a separate plasmid vector. To enhance expression of K1 and K2 specific polysaccharides, the *K. pneumoniae* transcriptional activator *rmpA* from *K. pneumoniae* NTUH K-2044 was subsequently cloned into pACT3 (Dykxhoorn, D.M., St Pierre, R. & Linn, T. *Gene* 177, 133-136 (1996)), a low copy, IPTG inducible vector as it has previously been characterized as a regulator of capsule in *K. pneumoniae* (Arakawa, Y. et al. *Infect Immun* 59, 2043-2050 (1991)); Yeh, K.M. et al. *J Clin Microbiol* 45, 466-471 (2007)). Introduction of the *rmpA* gene into *E. coli* strains co-expressing PglS and hexa-his tagged ComP variant and either the K1 or K2 capsular polysaccharides from *K. pneumoniae*, resulted robust expression and detection of higher molecular ComP bioconjugates as indicated by the typical ladder-like pattern of bands compatible with protein glycosylation when analyzed via western blotting (**Figure 16B**). The modal distribution of signals is indicative of protein glycosylation with repeating glycan subunits of increasing molecular weight. Thus collectively, PglS was able to glycosylate ComP with the K1 and K2 capsular polysaccharides from *K. pneumonia*. Increased efficiency of conjugation was observed with co-expression of the transcriptional activator *rmpA* from *K. pneumoniae*.

[0110] PglS can transfer *K. pneumoniae* polysaccharides to ComP. Given that most *K. pneumoniae* capsular polysaccharides contain glucose as the reducing end sugar, the only other commercially licensed OTases (PglB and PglL) should be unable to generate conjugate vaccines using these polysaccharides. Moreover, co-expression of the transcriptional activator, RmpA, with the capsule gene cluster enhanced capsule expression to detectably levels. In certain aspects, the method for producing *Klebsiella* conjugates can be used to generate a pan Klebsiella conjugate vaccine encompassing all serotypes – including other species such as *K. varricola, K. michiganensis,* and *K. oxytoca.*

[0111] Mass spectrometry and site directed mutagenesis confirm PglS is an *O*-linked OTase and reveal that ComP is glycosylated at a serine residue corresponding to position 84 of $ComP_{ADP1}$. *N*-glycosylation in bacteria generally occurs within the sequon D-X-N-S-T (SEQ ID NO: 21), where X is any amino acid but proline (Kowarik, M. et al. *EMBO J* 25, 1957-1966 (2006)). On the contrary, *O*-glycosylation does not seem to follow a defined sequon. Most *O*-glycosylation events in bacterial proteins occur in regions of low complexity (LCR), rich in serine, alanine, and proline (Vik, A. et al. *Proc Natl Acad Sci USA* 106, 4447-4452 (2009)). Alternatively, some pilins are *O*-glycosylated at a C-terminal serine residue (Comer, J.E., Marshall, M.A., Blanch, V.J., Deal, C.D. & Castric, P. *Infect Immun* 70, 2837-2845 (2002)). ComP does not appear to have an obvious LCR or a C-terminal serine residue homologous to those found in other pilin like proteins and therefore mass spectrometry was employed to determine the site(s) of glycosylation. Purified CPS14-ComP bioconjugates were subjected to proteolytic digestion, ZIC-HILIC glycopeptide enrichment, and multiple MS analyses. As seen in **Figure 17A** and **Figure 17B**, a single glycopeptide consisting of the peptide ISASNATTNVATAT (SEQ ID NO: 22) was identified attached to a glycan that matched the published CPS14 composition (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)). To enable confirmation of both the peptide and attached glycan sequences, multiple collision energies regimes were performed to confirm the glycosylation of the semi-GluC derived peptide ISASNATTNVATAT (SEQ ID NO: 22) with a 1378.47 Da glycan corresponding to $HexNAc_2Hexose_6$ (**Figure 17B**). Additional glycopeptides were also observed decorated with extended glycans corresponding to up to four tetrasaccharide repeat units (**Figure 29**).

[0112] It was previously shown that *Acinetobacter* species predominantly glycosylate proteins at serine residues and thus it was hypothesized that either serine (S) 82 or 84–as numbered in SEQ ID NO: 1–was the site of glycosylation (Scott, N.E. et al. *Mol Cell Proteomics* 13, 2354-2370 (2014)). To determine which serine residue was the site of glycosylation, these serine residues were individually mutated to alanine (A) and the glycosylation status of both mutant proteins was analyzed. For this experiment, the biosynthetic locus for the *C. jejuni* heptasaccharide was employed as the donor glycan, as glycosylation is readily detectable with the hR6 anti-glycan antisera as well as by an increase in electrophoretic mobility (Schwarz, F. et al. *Nat Chem Biol* 6, 264-266 (2010)). As shown in **Figure 18**, wild type hexa-his tagged ComP was glycosylated with the *C. jejuni* heptasaccharide as indicated by its increased electrophoretic mobility and co-localization with hR6 antisera signal when co-expressed with PglS. MS analysis

also confirmed the presence of the *C. jejuni* heptasaccharide on the identical semi-GluC derived peptide ISASNATTNVATAT (SEQ ID NO: 22) modified by CPS14 (**Figure 30** and **Figure 31**). As a negative control, a catalytically inactive PglS mutant (H324A) was generated, that when co-expressed with the *C. jejuni* heptasacchride glycan was unable to glycosylate wild type ComP. Site directed mutagenesis was performed and it was observed that glycosylation of ComP with the *C. jejuni* heptasaccharide was abolished in the ComP[S84A] mutant, whereas ComP[S82A] was glycosylated at wild-type levels. Together, these results indicate that ComP is singly glycosylated at serine 84 (as numbered in SEQ ID NO: 1) by PglS, which is a unique site that is different than other previously characterized pilin like proteins. This corresponds to serine 82 as numbered in SEQ ID NO: 2.

[0113] Bioinformatic features of ComP pilin orthologs. ComP was first described as a factor required for natural transformation in *Acinetobacter baylyi* ADP1 (Porstendorfer, D., Drotschmann, U. & Averhoff, B. *Appl Environ Microbiol* 63, 4150-4157 (1997)). In a subsequent study, it was demonstrated that ComP from *A. baylyi* ADP1 (herein referred to as $ComP_{ADP1}$) was glycosylated by a novel OTase, PglS, located immediately downstream of ComP, and not the general OTase PglL located elsewhere on the chromosome (Harding, C.M. et al. Mol Microbiol 96, 1023-1041 (2015)). The $ComP_{ADP1}$ protein (NCBI identifier AAC45886.1) belongs to a family of proteins called type IV pilins. Specifically, ComP shares homology to type IVa major pilins (Giltner, C.L., Nguyen, Y. & Burrows, L.L. *Microbiol Mol Biol Rev* 76, 740-772 (2012)). Type IVa pilins share high sequence homology at their N-terminus, which encode for the highly conserved leader sequence and N-terminal alpha helix; however, the C-terminus display remarkable divergences across genera and even within species (Giltner, C.L., Nguyen, Y. & Burrows, L.L. *Microbiol Mol Biol Rev* 76, 740-772 (2012)). To help differentiate ComP orthologs from other type IVa pilin proteins, such as, PilA from *A. baumannii, P. aeruginosa,* and *Haemophilus influenzae* as well as PilE from *Neisseria* species (Pelicic, V. *Mol Microbiol* 68, 827-837 (2008)), a BLASTp analysis was performed comparing the primary amino acid sequence of $ComP_{ADP1}$ against all proteins from bacteria in the *Acinetobacter* genus. Expectedly, many *Acinetobacter* type IVa pilin orthologs, including $ComP_{ADP1}$, share high homology at their N-termini; however, very few proteins display high sequence conservation across the entire amino acid sequence of ComP. At least six ComP orthologs (**Figure 19**) were identified based on the presence of the conserved serine at position 84 relative to $ComP_{ADP1}$ as well as a conserved disulfide bond flanking the site of predicted glycosylation connecting the predicted alpha beta

loop to the beta strand region (Giltner, C.L., Nguyen, Y. & Burrows, L.L. *Microbiol Mol Biol Rev* 76, 740-772 (2012)). Furthermore, all six ComP orthologs carry both a *pglS* homolog immediately downstream of the *comP* gene as well as a *pglL* homolog located elsewhere in the chromosome. Together, at least the presence of the conserved serine at position 84, the disulfide loop flanking the site of glycosylation, the presence of a *pglS* gene immediately downstream of *comP*, and the presence of a *pglL* homolog located elsewhere on the chromosome differentiate ComP pilin variants from other type IVa pilin variants.

[0114] Therefore, features common to ComP proteins are disclosed herein that identify ComP orthologs in different *Acinetobacter* species. ComP proteins can be differentiated from other pilins by the presence of the conserved glycosylated serine located at position 84 relative to the ADP1 ComP protein and the presence of a disulfide loop flanking the site of glycosylation. In addition, the presence of a *pglS* homolog immediately downstream of ComP is an indicator of ComP. Further to be classified as a PglS OTase protein rather than a PglL OTase protein, the OTase downstream of ComP must display higher sequence conservation with PglS (ACIAD3337) when compared to PglL (ACIAD0103) in *A. baylyi* ADP1. It is also evident to one of ordinary skill in the art that in any embodiment disclosed herein, a ComP protein comprises and is capable of being glycosylated on a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1).

[0115] **ComP from *A. soli* CIP 110264 is glycosylated by PglS from *A. baylyi* ADP1.** Given the presence of multiple ComP orthologs, whether PglS from *A. baylyi* ADP1 was able to glycosylate a divergent ComP protein was investigated. The ComP protein from *A. soli* CIP 110264 ($ComP_{110264}$) is 71% identical at the amino acid level when compared to $ComP_{ADP1}$. However, consistent with the features above, $ComP_{110264}$ contains the predicted disulfide bridge between the predicted alpha-beta loop and the second beta strand as well as the conserved serine located at position 84 relative to $ComP_{ADP1}$. Moreover, a PglS ortholog can be found immediately downstream of $ComP_{110264}$. To determine whether PglS from *A. baylyi* ADP1 ($PglS_{ADP1}$) could glycosylate $ComP_{110264}$, $PglS_{ADP1}$ was cloned into pACT3 and $ComP_{110264}$ into pEXT20 (Dykxhoorn, D.M., St Pierre, R. & Linn, T. *Gene* 177, 133-136 (1996)) and these plasmids were introduced into *E. coli* expressing the serotype 8 capsular polysaccharide (CPS8) from *S. pneumoniae*. Further, the converse experiment was performed by cloning and expressing PglS from *A. soli* CIP 110264 ($PglS_{110264}$) with $ComP_{ADP1}$. As seen in **Figure 20**, $PglS_{110264}$ minimally glycosylated its cognate acceptor pilin $ComP_{110264}$ as indicated by higher molecular weight ComP

pilin variants when compared to whole cell lysates lacking $PglS_{110264}$. Based on western blot analysis, $PglS_{110264}$ appeared to not glycosylate $ComP_{ADP1}$. On the other hand, $PglS_{ADP1}$ efficiently glycosylated both $ComP_{ADP1}$ and $ComP_{110264}$ as indicated by the robust increase of His-reactive signals of increasing electrophoretic mobility. Collectively, $PglS_{ADP1}$ appears to be an optimal OTase from heterologous glycosylation in *E. coli* with a unique ability to cross glycosylate multiple ComP substrates. Thus it was demonstrated that PglS proteins from different *Acinetobacter* species can glycosylate divergent, non-native ComP sequences.

[0116] Generation of a soluble, periplasmic fusion protein capable of being glycosylated by PglS. All members of type IVa pilin family are considered membrane proteins as part of their N-terminal alpha helix is embedded within the inner membrane (Giltner, C.L., Nguyen, Y. & Burrows, L.L. *Microbiol Mol Biol Rev* 76, 740-772 (2012)). Therefore, in order to generate soluble variants of ComP that are able to be glycosylated by PglS, translational fusions were constructed of truncated ComP fragment proteins onto three different carrier proteins. The carrier proteins, DsbA and MalE (also known as maltose binding protein–MBP) from *E. coli*, were selected as suitable carriers as both have been previously shown to facilitate periplasmic localization and solubility of acceptor proteins fused at their C-termini (Malik, A. *Biotech* 6, 44 (2016)). Exotoxin A from *Pseudomonas aeruginosa* (EPA) was also selected as it has been previously shown to act as an immunogenic carrier protein in other conjugate vaccine formulations (Ravenscroft, N. et al. *Glycobiology* 26, 51-62 (2016)). Fusion proteins consisted of a leader sequence, carrier protein, a short linker peptide, a ComP variant without the first 28 amino acids, and a hexa-histidine tag. The first 28 amino acids of $ComP_{ADP1}$ and $ComP_{110264}$ were removed as these amino acids contain the leader sequence as well as the hydrophobic region of the N-terminal alpha helix predicted to be embedded into the inner membrane. Fusion constructs were then introduced into *E. coli* expressing the pneumococcal serotype 8 capsular polysaccharide (CPS8) and either pACT3 alone or pACT3 carrying $pglS_{110264}$ or $pglS_{ADP1}$. As seen in **Figure 21**, *E. coli* cells expressing either DsbA-AAA-$ComP\Delta28_{110264}$ or DsbA-GGGS-$ComP\Delta28_{110264}$ in combination with $PglS_{ADP1}$ demonstrated detectable levels of glycosylation as indicated by the modal distribution of his reactive signals of increasing electrophoretic mobility. *E. coli* cells expressing fusions containing $ComP\Delta28_{ADP1}$ did not demonstrate any detectable glycosylation. The same glycosylation pattern was observed for *E. coli* cells expressing maltose binding protein (MBP) fusions. Specifically, as seen in **Figure 22**, *E. coli* cells expressing either MBP-AAA-$ComP\Delta28_{110264}$ or MBP-GGGS-$ComP\Delta28_{110264}$ in combination with $PglS_{ADP1}$

demonstrated detectable levels of glycosylation as indicated by the modal distribution of anti-His reactive signals; whereas, fusions with ComPΔ28$_{ADP1}$ were only minimally glycosylated. Lastly, to demonstrate that a previously established carrier protein used for conjugate vaccine formulations could be glycosylated by PglS with the pneumococcal CPS8, a fusion protein was engineered containing the DsbA signal peptide sequence fused to EPA. The ComPΔ28$_{110264}$ peptide was then fused with glycine-glycine-glycine-serine (GGGS; SEQ ID NO: 23) linker to the C-terminus of EPA and tested for glycosylation in the presence and absence of PglS$_{ADP1}$ in both whole cell extracts and in periplasmic extracts. As seen in **Figure 23**, EPA-GGGS-ComPΔ28$_{110264}$ constructs were found to be glycosylated in both the whole cell extract and periplasmic extracts of cells co-expressing the CPS8 glycan and PglS$_{ADP1}$ as indicated by the modal distribution of anti-His reactive signals. No detectable glycosylation was observed in samples lacking a PglS ortholog or in the samples expressing PglS$_{110264}$. Collectively, PglS$_{ADP1}$ is an optimal OTase for transferring polysaccharides containing glucose at the reducing end to truncated ComP fusion proteins. Specific amino acid sequences for each fusion construct are shown in **Figure 24**.

[0117] Immunization with a glycosylated ComP bioconjugate elicits an immune response. T-cell dependent immune responses to conjugate vaccines are characterized by the secretion of high affinity IgG1 antibody (Avci, F.Y., Li, X., Tsuji, M. & Kasper, D.L. *Nat Med* 17, 1602-1609 (2011)). The immunogenicity of a CPS14-ComP bioconjugate in a murine vaccination model was evaluated. As seen in **Figure 25A**, sera collected from mice vaccinated with a CPS14-ComP bioconjugate had a significant increase in CPS14 specific IgG titers but not IgM titers. Further, secondary HRP-tagged anti-IgG subtype antibodies were employed to determine which of the IgG subtypes had elevated titers. As seen in **Figure 25B**, IgG1 titers appeared to be higher than the other subtypes.

[0118] Next, a second vaccination trial was performed comparing the immunogenicity of a trivalent CPS8-, CPS9V-, and CPS14-ComP bioconjugate to the current standard of care, PREVNAR 13®. Serotypes 9V and 14 are included in PREVNAR 13® and elevated IgG titers could be seen in PREVNAR 13® immunized mice against these two serotypes (**Figure 26**). The monovalent immunization against serotype 14 also showed significant induction of serotype specific IgG titers, which were similar to the preliminary immunization (**Figure 25** and **Figure 26**). Mice receiving the trivalent bioconjugate, all had elevations in serotype specific IgG titers when compared to control as expected, day 49 sera have shown much more elevated IgG tires for

serotypes 8 and 14 compared to serotype 9V. Nevertheless, IgG titers against 9V were still significantly higher than the placebo (**Figure 26**).

[0119] Provide herein are bioconjugates comprising an oligo- or polysaccharide linked to a fusion protein. In certain embodiments, the oligo- or polysaccharide is covalently linked to the fusion protein. In certain embodiments, the fusion protein comprises a ComP protein (ComP). In certain other embodiments, the fusion protein comprises a glycosylation tag of a ComP protein (as described in detail elsewhere herein).

[0120] As disclosed herein, it has been discovered that ComP is glycosylated on a serine (S) residue. This serine residue is conserved in ComP proteins and corresponds to position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1). This serine residue also corresponds to position 82 of SEQ ID NO: 2 (ComP110264: ENV58402.1) (**Figure 39A, B, and C**). Thus, in certain aspects, a fusion protein (and thus the bioconjugate) is glycosylated with an oligo- or polysaccharide on a ComP glycosylation tag thereof at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) or corresponding to the serine redisue at position 82 of SEQ ID NO: 2. **Figure 28** shows an alignment of a region of ComP sequences including the serine (S) residue (boxed) corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), which is conserved across the ComP sequences. In certain embodiments, in order to be able to be glycosylated, the ComP glycosylation tag comprises both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1. Or, similarly described, in certain embodiments, in order to be able to be glycosylated, the ComP glycosylation tag comprises both a cysteine residue corresponding to the conserved cysteine residue at position 71 of SEQ ID NO: 2 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 93 of SEQ ID NO: 2.

[0121] In certain embodiments of a bioconjugate of this disclosure, the oligo- or polysaccharide comprises a glucose at its reducing end.

[0122] One of ordinary skill in the art would recognize that by aligning ComP sequences with SEQ ID NO: 1, (e.g., either full sequences or partial sequences) the conserved serine residue of a non-SEQ ID NO: 1 ComP protein disclosed herein, corresponding to the serine residue at position 84 of SEQ ID NO: 1, can be identified. Further, one of ordinary skill in the art would recognize that by aligning ComP sequences with SEQ ID NO: 1, other residues, regions, and/or

features corresponding to residues, regions, and/or features of SEQ ID NO: 1 as referred to herein can be identified in the non-SEQ ID NO: 1 ComP sequence and referenced in relation to SEQ ID NO:1. And, while reference is generally made herein to SEQ ID NO: 1, by analogy, reference can similarly be made to any residue, region, feature and the like of any ComP sequence disclosed herein, for example, in reference to SEQ ID NO: 2.

[0123] A ComP protein is a protein that has been identified as ComP protein consistent with the description provided herein. For example, representative examples of ComP proteins include, but are not limited to: AAC45886.1 ComP [*Acinetobacter sp.* ADP1]; ENV58402.1 hypothetical protein F951_00736 [*Acinetobacter soli* CIP 110264]; APV36638.1 competence protein [*Acinetobacter soli* GFJ-2]; PKD82822.1 competence protein [*Acinetobacter radioresistens* 50v1]; SNX44537.1 type IV pilus assembly protein PilA [*Acinetobacter puyangensis* ANC 4466]; and OAL75955.1 competence protein [Acinetobacter sp. SFC]. In certain aspects, a ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1). SEQ ID NO: 1 comprises a leader sequence of 28 amino acids. In certain aspects, a ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$) that do not include the 28 amino acid leader sequence but do contain a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1). In certain aspects, a ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$) that does not include the 28 amino acid leader sequence but does contain a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1). In certain aspects, the ComP protein comprises SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$). In certain aspects, the ComP protein is SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4

($ComP_{50v1}$: PKD82822.1), SEQ ID NO: 5 ($ComP_{4466}$: SNX44537.1), or SEQ ID NO: 6 ($ComP_{SFC}$: OAL75955.1).

[0124] In certain aspects, the bioconjugate is produced *in vivo* in a host cell such as by any of the methods of production disclosed herein. In certain aspects, the bioconjugate is produced in a bacterial cell, a fungal cell, a yeast cell, an avian cell, an algal cell, an insect cell, or a mammalian cell. In certain aspects, the bioconjugate is produced in a cell free system. Examples of the use of a cell free system utilizing OTases other than PglS can be found in WO2013/067523A1, which in incorporated herein by reference.

[0125] It has been discovered that a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1) can have an inhibitory effect on glycosylation when present in a ComP glycosylation tag even though the full length ComP protein comprising this methionine residue is glycosylated. Thus, in certain embodiments, the ComP glycosylation tag of this disclosure does not comprise a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). For example, in certain embodiments, such methionine residue in a ComP amino acid sequence is substituted with another amino acid that does not exhibit an inhibitory effect or is deleted from the ComP glycosylation tag amino acid sequence. In certain embodiments, the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). For example, in certain embodiments, the amino acid sequence of the ComP glycosylation tag ends with the residue corresponding to position 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, or 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). One of ordinary skill in the art would recognize that a fusion protein comprising a ComP glycosylation tag likewise would not comprise a methionine residue at a position corresponding to or corresponding about to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1) in relation to the ComP glycosylation tag, even if the methionine residue is attributed to a sequence of the fusion protein not as belonging to the ComP glycosylation tag sequence. For example, in certain embodiments, the fusion protein of the bioconjugate does not comprise, in relationship to the ComP glycosylation tag, a methionine residue at a position that would correspond to or correspond about to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). In certain embodiments, the fusion protein of the bioconjugate does not comprise, in relationship to the ComP glycosylation tag, a methionine

residue at a position that would correspond to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

[0126] A ComP glycosylation tag of the current disclosure is generally not a full length ComP protein. In certain embodiments of any ComP glycosylation tag described herein, the ComP glycosylation tag has a length of between 18 and 50 amino acids in length, for example, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48 ,49, or 50 amino acids in length. In certain embodiments, the glycosylation tag has length of between 21 and 45 amino acids in length. In certain embodiments, the glycosylation tag has a length of between 23 and 45 amino acids in length.

[0127] The ComP glycosylation tag of the current disclosure can be a fragment, a variant, or a variant fragment of a ComP protein as described anywhere herein. In certain embodiments, the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$). For example, in certain embodiments, the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$) or SEQ ID NO: 8 ($ComP\Delta28_{110264}$). In certain embodiments, the ComP protein comprises SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$). Further, in certain embodiments, the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4 (Com_{P50v1}: PKD82822.1), SEQ ID NO: 5 ($ComP_{4466}$: SNX44537.1), or SEQ ID NO: 6 ($ComP_{SFC}$: OAL75955.1). For example, in certain embodiments, the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) or SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). Further, in certain embodiments, the ComP protein comprises SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4 (Com_{P50v1}: PKD82822.1), SEQ ID NO: 5 ($ComP_{4466}$: SNX44537.1), or SEQ ID NO: 6 ($ComP_{SFC}$: OAL75955.1).

[0128] In certain embodiments, a ComP glycosylation tag of the current disclosure can be defined as comprising or consisting of the amino acid consensus sequence of SEQ ID NO: 27:

$X_1X_2GTX_5X_6X_7X_8X_9X_{10}X_{11}X_{12}CX_{14}GVX_{17}X_{18}IX_{20}X_{21}X_{22}ASX_{25}X_{26}TX_{28}NVX_{31}$
$X_{32}AX_{34}CX_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}X_{44}$ (SEQ ID NO: 27)

wherein:
X_1 is V, A, or no amino acid;
X_2 is A, G, T, or no amino acid;
X_5 is P, S, or Q;
X_6 is S, M, or I;
X_7 is T, P, or V;
X_8 is A, S, or T;
X_9 is G, N, S, or T;
X_{10} is N or no amino acid;
X_{11} is S, G, or A;
X_{12} is S or N;
X_{14} is V, T, or A;
X_{17} is Q, T, or E;
X_{18} is E, Q, or T;
X_{20} is S, N, A, or G;
X_{21} is S or no amino acid;
X_{22} is G or no amino acid;
X_{25} is N, S, or A;
X_{26} is A, S, or K;
X_{28} is T, S, or K;
X_{31} is A or E;
X_{32} is T or S;
X_{34} is T, Q, or A;
X_{36} is G, S, or T;
X_{37} is A, G, or D;
X_{38} is S, L, or A;
X_{39} is S, G, D, or T;
X_{40} is A, V, or G;
X_{41} is G, I, or V;
X_{42} is Q, T, or I;
X_{43} is I, V, T, or L; and
X_{44} is I, T, or V.

[0129] In certain embodiments, a ComP glycosylation tag comprises or consists of a fragment of the amino acid consensus sequence of SEQ ID NO: 27, wherein the fragment retains the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27. In certain embodiments, a ComP glycosylation tag comprises or consists of a variant of the amino acid consensus sequence of SEQ ID NO: 27 or a fragment thereof, having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions, however, wherein the variant maintains the

cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27. In certain embodiments, the amino acid substitution is a conservative amino acid substitution. As disclosed herein, in certain embodiments, a ComP glycosylation tag comprising SEQ ID NO: 27 does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). Further, in certain embodiments, the amino acid sequence of a ComP glycosylation tag comprising SEQ ID NO: 27 does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 44 of SEQ ID NO: 27. In certain embodiments, a ComP glycosylation tag comprising or consisting of the amino acid consensus sequence of SEQ ID NO: 27 or fragment and/or variant thereof is not more than 25, 30, 40, 45, or 50 amino acids in length.

[0130] In certain embodiments, a ComP glycosylation tag of the current disclosure can be defined as comprising or consisting of the amino acid consensus sequence of SEQ ID NO: 28:

$CX_2GVX_5X_6IX_8X_9X_{10}ASX_{13}X_{14}TX_{16}NVX_{19}X_{20}AX_{22}C$ (SEQ ID NO: 28)

wherein:

X_2 is V, T, or A, optionally V;
X_5 is Q, T, or E, optionally Q;
X_6 is E, Q, or T;
X_8 is S, N, A, or G;
X_9 is S or no amino acid;
X_{10} is G or no amino acid;
X_{13} is N, S, or A, optionally N;
X_{14} is A, S, or K, optionally A;
X_{16} is T, S, or K;
X_{19} is A or E, optionally A;
X_{20} is T or S, optionally T; or
X_{22} is T, Q, or A, optionally T.

[0131] In certain embodiments, a ComP glycosylation tag comprises or consists of a variant of the amino acid consensus sequence of SEQ ID NO: 28 having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions, however, wherein the variant maintains the cysteine residue at position 1 of SEQ ID NO: 28, the cysteine residue at position 23 of SEQ ID NO: 28, and the serine residue at position 12 of SEQ ID NO: 28. In certain embodiments, the amino acid substitution is a conservative amino acid substitution.

[0132] In certain embodiments, a ComP glycosylation tag comprising SEQ ID NO: 28 does not comprise a methionine residue in a position corresponding to the conserved methionine

residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). Further, in certain embodiments, the amino acid sequence of a ComP glycosylation tag comprising SEQ ID NO: 28 does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). In certain embodiments, a ComP glycosylation tag comprising the amino acid consensus sequence of SEQ ID NO: 28 or variant thereof is not more than 25, 30, 40, 45, or 50 amino acids in length.

[0133] In certain embodiments, the ComP glycosylation tag comprises or consists of a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11], wherein the variant maintains both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1 and the variant maintains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1. In certain embodiments, the amino acid substitution is a conservative amino acid substitution. Further, in certain embodiments, the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO:

96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11]. In certain embodiments, such a ComP glycosylation tag comprising one of the above sequences or variants thereof does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). Further, in certain embodiments, the amino acid sequence of such a ComP glycosylation tag comprising one of the above sequences or variants thereof does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1). In certain embodiments, a ComP glycosylation tag comprising an amino acid sequence and/or variant thereof listed above is not more than 25, 30, 40, 45, or 50 amino acids in length. In certain embodiments, a ComP glycosylation tag consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11].

[0134] In certain embodiments, the oligo- or polysaccharide for conjugation to the glycosylation tag, fusion protein, and/or bioconjugate is produced by a bacteria from the genus *Streptococcus*. For example, in certain embodiments, the polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide. Further, in certain embodiments, the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b. In certain other embodiments, the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*. For example, in certain embodiments, the polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varricola*, *Klebsiella michinganenis*, or *Klebsiella oxytoca* capsular polysaccharide. In certain embodiments, the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide. Further, in certain embodiments, the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*.

[0135] In certain embodiments, the bioconjugate is produced *in vivo*. For example, in certain embodiments, the bioconjugate is produced in a bacterial cell.

[0136] As the bioconjugate comprises an oligo- or polysaccharide covalently linked to a fusion protein, in certain applications, it may be advantageous to form a fusion protein with a carrier protein or fragment thereof. In certain embodiments, the carrier protein is one recognized in the art as useful in producing conjugate vaccines. In certain embodiments, when a ComP glycosylation tag fragment is fused to a carrier protein or fragment thereof, the glycosylation tag fragment and thus the fusion protein, can be glycosylated at the conserved serine residue described elsewhere herein. In certain embodiments, the fusion protein comprises a carrier protein selected from the group consisting of diphtheria toxoid CRM197, tetanus toxoid, *Pseudomonas aeruginosa* Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, *Haemophilus influenza* protein D, or a fragment thereof. In certain embodiments, the carrier protein or fragment thereof is linked to the ComP glycosylation tag via an amino acid linker, for example $(GGGS)_n$ (SEQ ID NO: 23), wherein n is at least one or AAA (SEQ ID NO: 24). In order to increase the potential immunogenicity of a ComP fusion protein, it may be advantageous to include more than one glycosylation tag. Thus, in certain embodiments, the fusion protein comprise two or more, three or more, four or more, five or more, six or more, eight or more, ten or more, fifteen or more, or twenty or more ComP glycosylation tags. In certain embodiments, the fusion protein comprises any of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 20 to any of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, or 25 ComP glycosylation tags. In certain embodiments, multiple glycosylation tags are arranged in tandem to one another in the fusion protein. In certain embodiments, multiple glycosylation tags are arranged apart from one another in the fusion protein, for example separated by sequences of carrier protein. In certain embodiments, the glycosylation tag(s) can be, for example, located at the N-terminal end of the carrier protein and/or fusion protein. In certain embodiments, the glycosylation tag(s) can be, for example, located at the C-terminal end of the carrier protein and/or fusion protein. In certain embodiments, the glycosylation tag(s) can be located internally within the carrier protein and/or fusions protein, for example, wherein a glycosylation tag is located between multiple carrier proteins in a fusion protein. In certain embodiments, the multiple carrier proteins can be identical in type or different in type. In certain embodiments, the glycosylation tags can be identical in type or different in type. In certain embodiments, these ComP glycosylation tags are identical. In certain embodiments, at least two of the ComP glycosylation tags differ from each other. In certain

embodiments, at least three, at least four, or at least five of the ComP glycosylation tags all differ from each other. Further, in certain embodiments, none of the ComP glycosylation tags are the same.

[0137] A bioconjugate of this invention may have one of numerous uses including, but not limited to, use as a conjugate vaccine. For example, in certain embodiments, the conjugate vaccine is a vaccine against *Streptococcus pneumoniae* serotype 8, *Streptococcus pneumoniae* serotype 1, *Streptococcus pneumoniae* serotype 2, *Streptococcus pneumoniae* serotype 4, *Streptococcus pneumoniae* serotype 5, *Streptococcus pneumoniae* serotype 6A, *Streptococcus pneumoniae* serotype 6B, *Streptococcus pneumoniae* serotype 7F, *Streptococcus pneumoniae* serotype 9N, *Streptococcus pneumoniae* serotype 9V, *Streptococcus pneumoniae* serotype 10A, *Streptococcus pneumoniae* serotype 11A, *Streptococcus pneumoniae* serotype 12F, *Streptococcus pneumoniae* serotype 14, *Streptococcus pneumoniae* serotype 15B, *Streptococcus pneumoniae* serotype 17F, *Streptococcus pneumoniae* serotype 18C, *Streptococcus pneumoniae* serotype 19F, *Streptococcus pneumoniae* serotype 19A, *Streptococcus pneumoniae* serotype 20, *Streptococcus pneumoniae* serotype 22F, *Streptococcus pneumoniae* serotype 23F, *Streptococcus pneumoniae* serotype 33F, *Klebsiella pneumoniae* serotype K1, *Klebsiella pneumoniae* serotype K2, *Klebsiella pneumoniae* serotype K5, *Klebsiella pneumoniae* serotype K16, *Klebsiella pneumoniae* serotype K20, *Klebsiella pneumoniae* serotype K54, *Klebsiella pneumoniae* serotype K57, *Streptococcus agalactiae* serotype Ia, *Streptococcus agalactiae* serotype Ib, *Streptococcus agalactiae* serotype II, *Streptococcus agalactiae* serotype III, *Streptococcus agalactiae* serotype IV, *Streptococcus agalactiae* serotype V, *Streptococcus agalactiae* serotype VI, *Streptococcus agalactiae* serotype VII, *Streptococcus agalactiae* serotype VIII, *Streptococcus agalactiae* serotype IX, *Streptococcus pyogenes* Group A Carbohydrate, *Enterococcus faecalis* serotype A, *Enterococcus faecalis* serotype B, *Enterococcus faecalis* serotype C, *Enterococcus faecalis* serotype D, *Enterococcus faecium* capsular polysaccharide and lipotechoic acid, *Moraxella catarrhalis* lipooligosaccharide A, *Moraxella catarrhalis* lipooligosaccharide B, *Moraxella catarrhalis* lipooligosaccharide C, and *Staphylococcus aureus* lipotechoic acid. In certain embodiments, the conjugate vaccine is useful because it induces an immune response when administered to a subject. In certain embodiments, the immune response elicits long term memory (memory B and T cells), is an antibody response, and is optionally a serotype-specific antibody response. In certain embodiments, the antibody response is an IgG or IgM response. For example, in certain embodiments the antibody response can be an IgG

response, and in certain embodiments, an IgG1 response. In certain embodiments, the conjugate vaccine generates immunological memory in a subject administered the vaccine.

[0138] Provided for herein is a fusion protein as disclosed in further detail elsewhere herein and comprising a ComP glycosylation tag as disclosed in detail elsewhere herein. In certain embodiments, the fusion protein is glycosylated at a serine residue on the glycosylation tag corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1). In certain embodiments, the fusion protein is glycosylated with an oligo- or polysaccharide. In certain embodiments, the oligo- or polysaccharide is produced by a bacteria from the genus *Streptococcus* such as, for example, a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide. In certain embodiments, the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b. In certain embodiments, the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*, for example, a *Klebsiella pneumoniae*, *Klebsiella varricola*, *Klebsiella michinganenis*, or *Klebsiella oxytoca* capsular polysaccharide. In certain embodiments, the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide. In certain embodiments, the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*. In certain of any embodiments disclosed herein, the oligo- or polysaccharide comprises a glucose at its reducing end. Certain embodiments are drawn a fusion protein wherein the fusion protein is produced *in vivo*. For example, in certain embodiments, the fusion protein is produced in a mammalian cell, fungal cell, yeast cell, insect cell, avian cell, algal cell, or bacterial cell. In certain embodiments, the fusion protein is produced in a bacterial cell, for example, *E. coli*.

[0139] Disclosed herein are methods for the *in vivo* conjugation of an oligo- or polysaccharide to a polypeptide (*in vivo* glycosylation). In certain embodiments, the method comprises covalently linking the oligo- or polysaccharide to the polypeptide with a PglS oligosaccharyltransferase (OTase) (described elsewhere herein). In certain embodiments, the polypeptide comprises a ComP protein or a glycosylation tag thereof. In certain embodiments, the polypeptide comprises a ComP protein or a glycosylation tag thereof linked to a heterologous polypeptide such as a carrier protein. Representative examples of PglS OTases include, but are not limited to $PglS_{110264}$, $PglS_{ADP1}$, $PglS_{GFJ-2}$, $PglS_{50v1}$, $PglS_{4466}$, and $PglS_{SFC}$. ComP proteins are described in detail elsewhere and representative examples include, but are not limited to $ComP_{110264}$, $ComP_{ADP1}$, $ComP_{GFJ-2}$, $ComP_{50v1}$, $ComP_{4466}$, and $ComP_{SFC}$. It will be recognized that while a PglS OTase from an organism would naturally glycosylate the ComP protein from that

organism (e.g., $PglS_{110264}$ glycosylates $ComP_{110264}$) in certain embodiments, a PglS from one organism glycosylates a ComP from a different organism (e.g., $PglS_{ADP1}$ glycosylates $ComP_{110264}$). For example, in certain aspects, the PglS OTase is $PglS_{ADP1}$. In certain embodiments, where the PglS OTase is $PglS_{ADP1}$, the ComP protein glycosylated is not $ComP_{ADP1}$. For example, in certain embodiments where the PglS OTase is $PglS_{ADP1}$, the ComP protein is $ComP_{110264}$. Of course, it will be recognized that a PglS OTase does not naturally glycosylate a ComP protein or a glycosylation tag fragment thereof, even from the same organism as the PglS Otase, when the ComP protein or glycosylation tag fragment thereof is linked to a heterologous carrier protein.

[0140] In certain embodiments for any combination of PglS and ComP, the ComP protein or glycosylation tag fragment thereof is glycosylated at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1).

[0141] In certain embodiments disclosed herein, the *in vivo* glycosylation occurs in a host cell. In certain embodiments, for example, the host cell can be a mammalian cell, fungal cell, yeast cell, insect cell, avian cell, algal cell, or bacterial cell. In certain embodiments, the host cell is a bacterial cell, for example, *E. coli*.

[0142] In certain embodiments, the method comprises culturing a host cell comprising the components necessary for the conjugation of the oligo- or polysaccharide to the polypeptide. In general, these components are the oligosaccharyltransferase, the acceptor polypeptide to be glycosylated, and the oligo- or polysaccharide. In certain embodiments, the method comprises culturing a host cell that comprises: (a) a genetic cluster encoding for the proteins required to synthesize the oligo- or polysaccharide; (b) a PglS OTase; and (3) the acceptor polypeptide. Further, it has been discovered that production of the oligo- or polysaccharide can be enhanced by a transcriptional activator. In certain embodiments, the production of the oligo- or polysaccharide is enhanced by the *K. pneumoniae* transcriptional activator rmpA (*K. pneumoniae* NTUH K-2044) or a homolog of the *K. pneumoniae* transcriptional activator rmpA (*K. pneumoniae* NTUH K-2044). In certain embodiments, the method further comprises expressing and/or providing such a transcriptional activator in the host cell along with the other components.

[0143] In certain embodiments, the carrier protein linked to the ComP glycosylation tag is, for example, diphtheria toxoid CRM197, tetanus toxoid, Pseudomonas aeruginosa Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, Haemophilus influenza protein D, or a fragment thereof.

[0144] Certain embodiments, are directed to a method a conjugate vaccine comprising a bioconjugate of this disclosure or a method of producing such conjugate vaccine.

[0145] Certain embodiments also provide for a host cell comprising the components for *in vivo* glycosylation of an acceptor ComP protein or glycosylation tag fragment thereof. In certain embodiments, a host cell comprises: (a) a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide; (b) a PglS OTase; and (3) an acceptor polypeptide comprising a ComP protein or a glycosylation tag fragment thereof. In certain embodiments, the acceptor polypeptide is a fusion protein. In certain embodiments, the host cell further comprises a transcriptional activator such as described above along with the other components.

[0146] In certain embodiments, a host cell comprises an isolated nucleic acid encoding a PglS OTase. In certain embodiments a host cell comprises an isolated nucleic acid encoding the ComP acceptor polypeptide. In certain embodiments, a host cell comprises a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide. In certain embodiments, a host cell comprises at least two of an isolated nucleic acid encoding a PglS OTase, an isolated nucleic acid encoding the ComP acceptor polypeptide, and genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide. In embodiments aspects, a host cell comprises a nucleic acid encoding a PglS OTase of one organism and a nucleic acid encoding the ComP acceptor polypeptide from a different organism.

[0147] Certain embodiments also provide for an isolated nucleic acid encoding the ComP protein, ComP glycosylation tag fragment, and/or ComP fusion protein described anywhere herein. In certain embodiments, an isolated nucleic acid referred to herein is a vector or is contained within a vector. In certain embodiments, an isolated nucleic acid referred to herein is inserted and/or has been incorporated into a heterologous genome or a heterologous region of a genome.

[0148] Disclosed herein is a pneumococcal bioconjugate vaccine containing a conventional vaccine carrier. Certain embodiments comprise the use of a ComP fragment as a glycosylation tag (aka "glycotag"). In certain embodiments, the glycosylation tag can be added to the C-terminus and/or N-terminus of a carrier protein. For example, in certain embodiments, the glycosylation tag is added to the C-terminus of the conventional carrier protein *Pseudomonas aeruginosa* Exotoxin A (EPA). It has been demonstrated that in certain embodiments, the glycosylation tag/carrier fusion protein can be paired with the CPS8 polysaccharide and use of PglS, generating a carrier protein-CPS8 bioconjugate, a first of its kind pneumococcal

bioconjugate vaccine. For example, in certain embodiments, an EPA fusion can be paired with the CPS8 polysaccharide and use of PglS, generating an EPA-CPS8 bioconjugate. It was demonstrated that the EPA-CPS8 bioconjugate vaccine elicited high IgG titers specific to serotype 8 specific that were protective as determined via bactericidal killing. Importantly, vaccination with as little as 100 ng of polysaccharide in the EPA-CPS8 bioconjugate was able to provide protection. Thus, certain embodiments provide for a CPS8 pneumococcal bioconjugate vaccine.

[0149] It is contemplated that a conjugate vaccine (such as the EPA vaccine construct) can comprise additional/multiple sites of glycosylation to increase the glycan to protein ratio as well as expand upon the number of serotypes in order to develop a comprehensive pneumococcal bioconjugate vaccine.

[0150] In certain embodiments, a bioconjugate or glycosylated fusion protein disclosed herein is a conjugate vaccine that can be administered to a subject for the prevention and/or treatment of an infection and/or disease. In certain embodiments, the conjugate vaccine is a prophylaxis that can be used, e.g., to immunize a subject against an infection and/or disease. In certain embodiments, the bioconjugate is associated with (such as in a therapeutic composition) and/or administered with an adjuvant. Certain embodiments provide for a composition (such as a therapeutic composition) comprising a conjugate vaccine described herein and an adjuvant. In certain embodiments, when the conjugate vaccine is administered to a subject, it induces an immune response. In certain embodiments, the immune response elicits long term memory (memory B and T cells). In certain embodiments, the immune is an antibody response. In certain embodiments, the antibody response is a serotype-specific antibody response. In certain embodiments, the antibody response is an IgG or IgM response. In certain embodiments where the antibody response is an IgG response, the IgG response is an IgG1 response. Further, in certain embodiments, the conjugate vaccine generates immunological memory in a subject administered the vaccine.

[0151] Certain embodiments also provide for producing a vaccine against an infection and/or disease. In certain embodiments a method comprises isolating a bioconjugate or fusion protein disclosed herein (conjugate vaccine) and combining the conjugate vaccine with an adjuvant. In certain embodiments, the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature. In certain embodiments, the vaccine is a conjugate vaccine against pneumococcal infection. In certain embodiments, the disease is pneumonia. In

certain embodiments, the infection is a systemic infection and/or an infection of the blood. In certain embodiments, the subject is a mammal. For example, in certain embodiments, a pig or a human.

[0152] Importantly, the aspects disclosed herein are not limited to pneumococcal polysaccharides, but in fact, have vast applicability for generating bioconjugate vaccines for many important human and animal pathogens that are incompatible with PglB and PglL. Notable examples include the human pathogens *Klebsiella pneumoniae* and Group B *Streptococcus* as well as the swine pathogen *S. suis*, all immensely relevant pathogens with no licensed vaccines available.

[0153] Provided herein are methods of inducing a host immune response against a pathogen. In certain embodiments, the pathogen is a bacterial pathogen. In certain embodiments, the host is immunized against the pathogen. In certain embodiments, the method comprises administering to a subject in need of the immune response an effective amount of a ComP conjugate vaccine, glycosylated fusion protein, or any other therapeutic/immunogenic composition disclosed herein. Certain embodiments provide a conjugate vaccine, glycosylated fusion protein, or other therapeutic/immunogenic composition disclosed herein for use in inducing a host immune response against a bacterial pathogen and immunization against the bacterial pathogen. Examples of immune responses include but are not limited to an innate response, an adaptive response, a humoral response, an antibody response, cell mediated response, a B cell response, a T cell response, cytokine upregulation or downregulation, immune system cross-talk, and a combination of two or more of said immune responses. In certain embodiments, the immune response is an antibody response. In certain embodiments, the immune response is an innate response, a humoral response, an antibody response, a T cell response, or a combination of two or more of said immune responses.

[0154] Also provided herein are methods of preventing or treating a bacterial disease and/or infection in a subject comprising administering to a subject in need thereof a conjugate vaccine, a fusion protein, or a composition disclosed herein. In certain embodiments, the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature. In certain embodiments, the disease is pneumonia. In certain embodiments, the infection is a systemic infection and/or an infection of the blood. In certain embodiments disclosed herein, the subject is a vertebrate. In certain embodiments the subject is a mammal such as a dog, cat,

cow, horse, pig, mouse, rat, rabbit, sheep, goat, guinea pig, monkey, ape, etc. And, for example, in certain embodiments the mammal is a human.

[0155] In any of the embodiments of administration disclose herein, the composition is administered via intramuscular injection, intradermal injection, intraperitoneal injection, subcutaneous injection, intravenous injection, oral administration, mucosal administration, intranasal administration, or pulmonary administration.

EXAMPLES

Example 1. Determination that cysteine residues flanking the site of glycosylation in ComP contribute to ComP stability and glycosylation.

[0156] Previously, it was demonstrated that the ComP protein from *Acinetobacter baylyi* ADP1 ($ComP_{ADP1}$) and *A. soli* strain 110264 ($ComP_{110264}$) are glycosylated at a homologous serine residue located at position 84 or 82, respectively, by the *O*-linking OTase PglS (Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891). Specifically, it was shown that the S84A point mutant of $ComP_{ADP1}$ and the S82A point mutant of $ComP_{110264}$ were not able to be glycosylated with the serotype 8 pneumococcal capsular polysaccharide by PglS. To further analyze the role of other amino acids important for PglS-dependent glycosylation, a series of point mutants were generated to alter the conserved cysteine residues flanking the site of glycosylation located at positions 75 and 95 of $ComP_{ADP1}$ (**Figure 1A**).

[0157] A series of point mutants was first generated replacing cysteine 75 with either alanine or glycine as these mutants would block the formation of a disulfide bond that may be formed between cysteines 75 and 95. The point mutants were then introduced into *E. coli* SDB1 co-expressing the *C. jejuni* heptasaccharide biosynthetic gene cluster and $PglS_{ADP1}$. As seen in **Figure 1B**, mutation of cysteine 75 significantly reduced the expression of the ComP protein mutants compared to wildtype (WT) ComP. In particularly, the C75A and C75G mutants displayed very low levels of protein expression and it appeared to exist only as a low molecular weight unglycosylated form. Next, a second series of point mutants was generated replacing cysteine 95 with either alanine, glycine or serine and then again introduced these mutant ComP constructs into *E. coli* SDB1 co-expressing the *C. jejuni* heptasaccharide biosynthetic gene cluster and $PglS_{ADP1}$. As seen in **Figure 1B**, ComP mutants were unable to be detected with either alanine, glycine, or serine in replace of cysteine 95. Last, a series of double point mutations

was generated consisting of all the permutations of cysteine 75 with alanine or glycine and cysteine 95 with alanine, glycine, or serine. As seen in **Figure 1B**, ComP double mutants were unable to be detected. Based on the lack of detectable expression for the different ComP point mutant variants, it is likely that the cysteine residues located at positions 75 and 95 form a disulfide bond flanking the site of glycosylation at serine 84. Blocking the formation of this disulfide bond appears detrimental to protein stability and protein glycosylation.

Example 2. A short PglS-dependent O-linking recognition motif is determined via a reductive cloning strategy.

[0158] It was demonstrated that a translational fusion containing $ComP_{110264}$ lacking the first 28 amino acids (herein referred to as $ComP\Delta28_{110264}$) fused at the C-terminus of a genetically inactivated variant of the exotoxin A protein from *Pseudomonas aeruginosa* (EPA) was efficiently glycosylated by PglS with multiple pneumococcal and *K. pneumoniae* capsular polysaccharides (Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891; Feldman MF, *et al.* (2019) A promising bioconjugate vaccine against hypervirulent Klebsiella pneumoniae. *Proc Natl Acad Sci U S A*). In order to shorten and define the minimal recognition site required for PglS dependent glycosylation, a reductive cloning strategy was pursued whereby fragments of $ComP_{110264}$ were translationally fused to the C-terminus of the EPA protein in between a glycine-glycine-glycine-serine (GGGS) linker and a hexahistidine tag (**Figure 2**). Specifically, multiple constructs were generated containing either a 25, 30, 35, 40, or 45 amino acid fragment of $ComP_{110264}$. Each fragment, irrespective of size, was shifted by one amino acid towards the stop codon of $ComP_{110264}$. As an example, the first construct contained a 25 amino acid fragment of $ComP_{110264}$ spanning residues 67 to 91, the second construct contained a 25 amino acid fragment of $ComP_{110264}$ spanning residues 68 to 92, the third construct contained a 25 amino acid fragment of $ComP_{110264}$ spanning residues 69 to 93 and so on. All fragments contained serine 82, the site of PglS glycosylation. EPA fusion constructs were then introduced into *E. coli* SDB1 co-expressing $PglS_{ADP1}$ and the pneumococcal CPS8.

[0159] As can be seen in **Figure 3**, three constructs containing a short 25 amino acid fragment of $ComP_{110264}$ were found to be glycosylated by $PglS_{ADP1}$ with the pneumococcal CPS8 as determined by a decreased electrophoretic mobility and the presence of multiple glycoforms (observed as a modal, ladder-like distribution above the unglycosylated protein) when analyzed via western blot. As a positive control, the EPA fusion containing the $ComP\Delta28_{110264}$ fragment

was included as this protein has previously been established to be glycosylated by $PglS_{APD1}$ with the CPS8. It is noteworthy that the only three 25 amino acid constructs found to be glycosylated: C1 (SEQ ID NO: 32); D1 (SEQ ID NO: 33); and E1 (SEQ ID NO: 34), all contained the conserved cysteine residues predicted to form a disulfide bond flanking the site of $ComP_{110264}$ glycosylation (serine 82).

[0160] As can be seen in **Figure 3** and **Figure 4**, seven constructs containing a 30 amino acid fragment of $ComP_{110264}$ were found to be glycosylated by $PglS_{ADP1}$ with the pneumococcal CPS8: E2 (SEQ ID NO: 41); F2 (SEQ ID NO: 42); G2 (SEQ ID NO: 43); H2 (SEQ ID NO: 44); A3 (SEQ ID NO: 45); B3 (SEQ ID NO: 46); C3 (SEQ ID NO: 47). All seven constructs contained the conserved cysteine residues predicted to form a disulfide bond flanking the site of $ComP_{110264}$ glycosylation.

[0161] As can be seen in **Figure 5** and **Figure 6**, nine constructs containing a 35 amino acid fragment of $ComP_{110264}$ were found to be glycosylated by $PglS_{ADP1}$ with the pneumococcal CPS8: D4 (SEQ ID NO: 55); E4 (SEQ ID NO: 56); F4 (SEQ ID NO: 57); G4 (SEQ ID NO: 58); A5 (SEQ ID NO: 59); B5 (SEQ ID NO: 60); D5 (SEQ ID NO: 61); E5 (SEQ ID NO: 62); F5 (SEQ ID NO: 63). All nine contained the conserved cysteine residues predicted to form a disulfide bond flanking the site of $ComP_{110264}$ glycosylation.

[0162] As can be seen in **Figure 6** and **Figure 7**, eight constructs containing a 40 amino acid fragment of $ComP_{110264}$ were found to be glycosylated by $PglS_{ADP1}$ with the pneumococcal CPS8: H6 (SEQ ID NO: 72); B7 (SEQ ID NO: 73); C7 (SEQ ID NO: 74); D7 (SEQ ID NO: 75); E7 (SEQ ID NO: 76); F7 (SEQ ID NO: 77); A8 (SEQ ID NO: 78); B8 (SEQ ID NO: 79). All eight contained the conserved cysteine residues predicted to form a disulfide bond flanking the site of $ComP_{110264}$ glycosylation.

[0163] As seen in **Figure 8**, **Figure 9**, and **Figure 10**, ten constructs containing a 45 amino acid fragment of $ComP_{110264}$ were found to be glycosylated by $PglS_{ADP1}$ with the pneumococcal CPS8: A10 (SEQ ID NO: 92); B10 (SEQ ID NO: 93); C10 (SEQ ID NO: 94); D10 (SEQ ID NO: 95); F10 (SEQ ID NO: 96); G10 (SEQ ID NO: 97); H10 (SEQ ID NO: 98); A11 (SEQ ID NO: 99); B11 (SEQ ID NO: 100); C11 (SEQ ID NO: 101). Again, all ten contained the conserved cysteine residues predicted to form a disulfide bond flanking the site of $ComP_{110264}$ glycosylation.

[0164] Based on the data presented in **Figure 3**, **Figure 4**, **Figure 5**, **Figure 6**, **Figure 7**, **Figure 8**, **Figure 9**, and **Figure 10**, the cysteine residues located at position 71 and 93 are necessary for glycosylation by $PglS_{ADP1}$ when translationally fused to the C-terminus of the EPA

carrier protein. In addition, the methionine residue located in position 104 appears to block $PglS_{ADP1}$ glycosylation when a part of the C-terminal glycosylation tag. This is particularly evidenced by the fact that constructs G5 (SEQ ID NO: 64), C8 (SEQ ID NO: 80), and D11 (SEQ ID NO: 102) each contain the required cysteines at position 71 and 93, but did not display any signs of glycosylation. While G5 (SEQ ID NO: 64), C8 (SEQ ID NO: 80), and D11 (SEQ ID NO: 102) contain fragments of $ComP_{110264}$ of 35, 40, and 45 amino acids in length, respectively, each fragment terminates with the methionine at position 104, demonstrating that this residue is sufficient to block glycosylation when included in the C-terminal glycotag. Moreover, all constructs containing methionine 104 in addition to the cysteines in position 71 and 93 did not display any sign of glycosylation (G5 (SEQ ID NO: 64), H5 (SEQ ID NO: 65), C8 (SEQ ID NO: 80), D8 (SEQ ID NO: 81), E8 (SEQ ID NO: 82), F8 (SEQ ID NO: 83), G8 (SEQ ID NO: 84), H8 (SEQ ID NO: 85), A9 (SEQ ID NO: 86), D11 (SEQ ID NO: 102), E11 (SEQ ID NO: 103), F11 (SEQ ID NO: 104), H11 (SEQ ID NO: 105), A12 (SEQ ID NO: 106), B12 (SEQ ID NO: 107), C12 (SEQ ID NO: 108), D12 (SEQ ID NO: 109), E12 (SEQ ID NO: 110), F12 (SEQ ID NO: 111), G12 (SEQ ID NO: 112). **Table 1** provides a summary of all $ComP_{110264}$ fragments tested for their ability to serve as *O*-linking glycosylation recognition motifs by $PglS_{ADP1}$.

Table 1.

ID	SEQ ID NO:	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation observed
A1	30	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASA$_{91}$	-
B1	31	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	-
C1	32	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	+
D1	33	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	+
E1	34	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	+
F1	35	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	-
G1	36	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSD$_{97}$	-
H1	37	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	-
A2	38	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	-
C2	39	$_{62}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASA$_{91}$	-
D2	40	$_{63}$TSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	-
E2	41	$_{64}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	+
F2	42	$_{65}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	+
G2	43	$_{66}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	+
H2	44	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	+
A3	45	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSD$_{97}$	+
B3	46	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	+
C3	47	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	+
E3	48	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	-
F3	49	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	-
G3	50	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	-
H3	51	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	-
A4	52	$_{76}$QIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{105}$	-
B4	53	$_{57}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASA$_{91}$	-
C4	54	$_{58}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	-
D4	55	$_{59}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	+
E4	56	$_{60}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	+
F4	57	$_{61}$GGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	+
G4	58	$_{62}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	+
A5	59	$_{64}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	+
B5	60	$_{65}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	+
D5	61	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	+
E5	62	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	+
F5	63	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	+
G5	64	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	-
H5	65	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{105}$	-
A6	66	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTD$_{106}$	-
B6	67	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDK$_{107}$	-
C6	68	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKA$_{108}$	-
D6	69	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	-
F6	70	$_{52}$TVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASA$_{91}$	-
G6	71	$_{53}$VSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	-
H6	72	$_{54}$SENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	+
B7	73	$_{56}$NIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	+
C7	74	$_{57}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	+

ID	SEQ ID NO:	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation observed
D7	75	$_{58}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSD$_{97}$	+
E7	76	$_{59}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	+
F7	77	$_{60}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	+
A8	78	$_{63}$TSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	+
B8	79	$_{64}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	+
C8	80	$_{65}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	-
D8	81	$_{66}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{105}$	-
E8	82	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTD$_{106}$	-
F8	83	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDK$_{107}$	-
G8	84	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKA$_{108}$	-
H8	85	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	-
A9	86	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKG$_{110}$	-
B9	87	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGV$_{111}$	-
C9	88	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVS$_{112}$	-
D9	89	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSI$_{113}$	-
E9	90	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIK$_{114}$	-
H9	91	$_{48}$AMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	-
A10	92	$_{49}$MKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	+
B10	93	$_{50}$KATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	+
C10	94	$_{51}$ATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	+
D10	95	$_{52}$TVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	+
F10	96	$_{54}$SENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	+
G10	97	$_{55}$ENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	+
H10	98	$_{56}$NIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI$_{100}$	+
A11	99	$_{57}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	+
B11	100	$_{58}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	+
C11	101	$_{59}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	+
D11	102	$_{60}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	-
E11	103	$_{61}$GGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{105}$	-
F11	104	$_{62}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTD$_{106}$	-
H11	105	$_{64}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKA$_{108}$	-
A12	106	$_{65}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	-
B12	107	$_{66}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKG$_{110}$	-
C12	108	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGV$_{111}$	-
D12	109	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVS$_{112}$	-
E12	110	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSI$_{113}$	-
F12	111	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIK$_{114}$	-
G12	112	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIKL$_{115}$	-
H12	113	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIKLT$_{116}$	-

[0165] *O*-linking glycosylation recognition motifs can be glycosylated by $PglS_{ADP1}$ when translationally fused N-terminally, in tandem at the N- or C-terminus, or simultaneously at the N- and C-terminus. Based on the data presented above, the D5 (SEQ ID NO: 61) fragment of $ComP_{110264}$ was selected for follow up experiments whereby the D5 (SEQ ID NO: 61) fragment or a derivative thereof (D5') was translationally fused to N-terminus and C-terminus in different combinations as outlined in **Figure 12A** and **Figure 12B**. As a positive control, the EPA fusion containing the $ComP\Delta28_{110264}$ fragment was included as this protein has previously been established to be glycosylated by $PglS_{APD1}$ with the CPS8. EPA fusion constructs were then introduced into *E. coli* SDB1 co-expressing the pneumococcal CPS8 in the presence of absence of $PglS_{ADP1}$. As seen in **Figure 12C**, all EPA-ComP110264 fusion constructs were glycosylated with the pneumococcal CPS8 indicating that ComP110264 *O*-linking glycosylation recognition motifs can be translationally fused in multiple combinations at the N-terminus or C-terminus and still be glycosylated by $PglS_{ADP1}$.

Example 3. A tandem, C-terminally fused double $ComP\Delta28_{110264}$ glycosylation tag is glycosylated by $PglS_{ADP1}$. EPA fusion constructs were built containing a tandem, C-terminally fused double $ComP\Delta28_{110264}$ glycosylation tag. The $ComP\Delta28_{110264}$ glycosylation tags were separated by either a glycine-glycine-glycine-glycine-serine (GGGS) linker (SEQ ID NO: 23) or by a proline-alanine-proline-alanine-proline (PAPAP) linker (SEQ ID NO: 25). Both constructs contained a hexahistidine tag to aid downstream purification. As a positive control, the EPA fusion containing the $ComP\Delta28_{110264}$ fragment was included as this protein has previously been established to be glycosylated by $PglS_{APD1}$ with the CPS8. The double tag EPA fusion constructs were then introduced into *E. coli* SDB1 co-expressing the pneumococcal CPS8 in the presence of absence of $PglS_{ADP1}$. As can be seen in **Figure 13**, EPA variant 7 and EPA variant 8 were both glycosylated the pneumococcal CPS8 when $PglS_{ADP1}$ was present. Moreover, the glycosylation appeared as very high molecular weight with immunoreactivity approaching the 250 kDa marker.

*****

[0166] The present disclosure is not to be limited in scope by the specific aspects described or preceding Examples which are intended as single illustrations of individual aspects of the disclosure, and any compositions or methods which are functionally equivalent are within the scope of this disclosure. Indeed, various modifications of the disclosure in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing

description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

[0167] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

REFERENCES

1. O'Brien, K. L. et al. Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. Lancet 374, 893-902, doi:10.1016/S0140-6736(09)61204-6 (2009).

2. Pneumococcal conjugate vaccine for childhood immunization--WHO position paper. Wkly Epidemiol Rec 82, 93-104 (2007).

3. Prevention, C. f. D. C. a. Pneumococcal Vaccination, <https://www.cdc.gov/vaccines/vpd/pneumo/index.html>

4. Pace, D. Glycoconjugate vaccines. Expert Opin Biol Ther 13, 11-33, doi:10.1517/14712598.2012.725718 (2013).

5. Vella, M. & Pace, D. Glycoconjugate vaccines: an update. Expert Opin Biol Ther 15, 529-546, doi:10.1517/14712598.2015.993375 (2015).

6. Pollard, A. J., Perrett, K. P. & Beverley, P. C. Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines. Nat Rev Immunol 9, 213-220, doi:10.1038/nri2494 (2009).

7. Avci, F. Y., Li, X., Tsuji, M. & Kasper, D. L. A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. Nat Med 17, 1602-1609, doi:10.1038/nm.2535 (2011).

8. Package Insert - Prevnar 13 - FDA, <https://www.fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM201669.pdf>

9. Prevention, C. f. D. C. a. Vaccines for Children Program (VFC), <https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html> (2018).

10. Pfizer Inc. 2017 Financial Report, <https://www.sec.gov/Archives/edgar/data/78003/000007800318000027/pfe-exhibit13x12312017x10k.htm> (2018).

11. Frasch, C. E. Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. Vaccine 27, 6468-6470, doi:10.1016/j.vaccine.2009.06.013 (2009).

12. Huttner, A. & Gambillara, V. The development and early clinical testing of the ExPEC4V conjugate vaccine against uropathogenic Escherichia coli. Clin Microbiol Infect, doi:10.1016/j.cmi.2018.05.009 (2018).

13. Huttner, A. et al. Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. Lancet Infect Dis 17, 528-537, doi:10.1016/S1473-3099(17)30108-1 (2017).

14. Riddle, M. S. et al. Safety and Immunogenicity of a Candidate Bioconjugate Vaccine against Shigella flexneri 2a Administered to Healthy Adults: a Single-Blind, Randomized Phase I Study. Clin Vaccine Immunol 23, 908-917, doi:10.1128/CVI.00224-16 (2016).

15. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta 1473, 4-8 (1999).

16. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol 8, 765-778, doi:10.1038/nrmicro2383 (2010).

17. Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L. & Feldman, M. F. Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. Mol Microbiol 89, 14-28, doi:10.1111/mmi.12265 (2013).

18. Ciocchini, A. E. et al. A bacterial engineered glycoprotein as a novel antigen for diagnosis of bovine brucellosis. Vet Microbiol 172, 455-465, doi:10.1016/j.vetmic.2014.04.014 (2014).

19. Garcia-Quintanilla, F., Iwashkiw, J. A., Price, N. L., Stratilo, C. & Feldman, M. F. Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. Front Microbiol 5, 381, doi:10.3389/fmicb.2014.00381 (2014).

20. Iwashkiw, J. A. et al. Exploiting the Campylobacter jejuni protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis. Microb Cell Fact 11, 13, doi:10.1186/1475-2859-11-13 (2012).

21. Wacker, M. et al. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proc Natl Acad Sci U S A 103, 7088-7093, doi:10.1073/pnas.0509207103 (2006).

22. Feldman, M. F. et al. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A 102, 3016-3021, doi:10.1073/pnas.0500044102 (2005).

23. Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S. & Feldman, M. F. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J Bacteriol 189, 8088-8098, doi:10.1128/JB.01318-07 (2007).

24. Geno, K. A. et al. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev 28, 871-899, doi:10.1128/CMR.00024-15 (2015).

25. Ihssen, J. et al. Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering. Open Biol 5, 140227, doi:10.1098/rsob.140227 (2015).

26. Pan, C. et al. Biosynthesis of Conjugate Vaccines Using an O-Linked Glycosylation System. MBio 7, e00443-00416, doi:10.1128/mBio.00443-16 (2016).

27. Wacker, M. et al. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science 298, 1790-1793, doi:10.1126/science.298.5599.1790 (2002).

28. Vik, A. et al. Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae. Proc Natl Acad Sci U S A 106, 4447-4452, doi:10.1073/pnas.0809504106 (2009).

29. Harding, C. M. et al. Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. Mol Microbiol 96, 1023-1041, doi:10.1111/mmi.12986 (2015).

30. Pan, Y. J. et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of Klebsiella spp. Sci Rep 5, 15573, doi:10.1038/srep15573 (2015).

31. Kovach, M. E. et al. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene 166, 175-176 (1995).

32. Wu, K. M. et al. Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol 191, 4492-4501, doi:10.1128/JB.00315-09 (2009).

33. Lery, L. M. et al. Comparative analysis of Klebsiella pneumoniae genomes identifies a phospholipase D family protein as a novel virulence factor. BMC Biol 12, 41, doi:10.1186/1741-7007-12-41 (2014).

34. Dykxhoorn, D. M., St Pierre, R. & Linn, T. A set of compatible tac promoter expression vectors. Gene 177, 133-136 (1996).

35. Arakawa, Y. et al. Biosynthesis of Klebsiella K2 capsular polysaccharide in Escherichia coli HB101 requires the functions of rmpA and the chromosomal cps gene cluster of the virulent strain Klebsiella pneumoniae Chedid (O1:K2). Infect Immun 59, 2043-2050 (1991).

36. Yeh, K. M. et al. Capsular serotype K1 or K2, rather than magA and rmpA, is a major virulence determinant for Klebsiella pneumoniae liver abscess in Singapore and Taiwan. J Clin Microbiol 45, 466-471, doi:10.1128/JCM.01150-06 (2007).

37. Kowarik, M. et al. Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25, 1957-1966, doi:10.1038/sj.emboj.7601087 (2006).

38. Comer, J. E., Marshall, M. A., Blanch, V. J., Deal, C. D. & Castric, P. Identification of the Pseudomonas aeruginosa 1244 pilin glycosylation site. Infect Immun 70, 2837-2845 (2002).

39. Scott, N. E. et al. Diversity within the O-linked protein glycosylation systems of acinetobacter species. Mol Cell Proteomics 13, 2354-2370, doi:10.1074/mcp.M114.038315 (2014).

40. Schwarz, F. et al. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol 6, 264-266, doi:10.1038/nchembio.314 (2010).

41. Porstendorfer, D., Drotschmann, U. & Averhoff, B. A novel competence gene, comP, is essential for natural transformation of Acinetobacter sp. strain BD413. Appl Environ Microbiol 63, 4150-4157 (1997).

42. Giltner, C. L., Nguyen, Y. & Burrows, L. L. Type IV pilin proteins: versatile molecular modules. Microbiol Mol Biol Rev 76, 740-772, doi:10.1128/MMBR.00035-12 (2012).

43. Pelicic, V. Type IV pili: e pluribus unum? Mol Microbiol 68, 827-837, doi:10.1111/j.1365-2958.2008.06197.x (2008).

44. Malik, A. Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of E. coli. 3 Biotech 6, 44, doi:10.1007/s13205-016-0397-7 (2016).

45. Ravenscroft, N. et al. Purification and characterization of a Shigella conjugate vaccine, produced by glycoengineering Escherichia coli. Glycobiology 26, 51-62, doi:10.1093/glycob/cwv077 (2016).

46. Schulz, B. L. et al. Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. PLoS One 8, e62768, doi:10.1371/journal.pone.0062768 (2013).

47. Castric, P. pilO, a gene required for glycosylation of Pseudomonas aeruginosa 1244 pilin. Microbiology 141 (Pt 5), 1247-1254, doi:10.1099/13500872-141-5-1247 (1995).

48. Power, P. M. et al. Genetic characterization of pilin glycosylation and phase variation in Neisseria meningitidis. Mol Microbiol 49, 833-847 (2003).

49. Stimson, E. et al. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. Mol Microbiol 17, 1201-1214 (1995).

50. Ishihama, Y., Rappsilber, J. & Mann, M. Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. J Proteome Res 5, 988-994, doi:10.1021/pr050385q (2006).

51. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols 2, 1896-1906, doi:10.1038/nprot.2007.261 (2007).

52. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom 11, 601, doi:10.1002/bms.1200111109 (1984).

53. Haurat, M. F. et al. Selective sorting of cargo proteins into bacterial membrane vesicles. J Biol Chem 286, 1269-1276, doi:10.1074/jbc.M110.185744 (2011).

54. Price, N. L. et al. Glycoengineered Outer Membrane Vesicles: A Novel Platform for Bacterial Vaccines. Sci Rep 6, 24931, doi:10.1038/srep24931 (2016).

55. Kay, E. J., Yates, L. E., Terra, V. S., Cuccui, J. & Wren, B. W. Recombinant expression of Streptococcus pneumoniae capsular polysaccharides in Escherichia coli. Open Biol 6, 150243, doi:10.1098/rsob.150243 (2016).

56. Szymanski CM, Yao R, Ewing CP, Trust TJ, & Guerry P (1999) Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol Microbiol* 32(5):1022-1030.

57. Iwashkiw JA, Vozza NF, Kinsella RL, & Feldman MF (2013) Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol* 89(1):14-28.

58. Wacker M, *et al.* (2002) N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* 298(5599):1790-1793.

59. Nothaft H & Szymanski CM (2010) Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* 8(11):765-778.

60. Schaffer C & Messner P (2017) Emerging facets of prokaryotic glycosylation. *FEMS Microbiol Rev* 41(1):49-91.

61. Comstock LE & Kasper DL (2006) Bacterial glycans: key mediators of diverse host immune responses. *Cell* 126(5):847-850.

62. De Gregorio E & Rappuoli R (2014) From empiricism to rational design: a personal perspective of the evolution of vaccine development. *Nat Rev Immunol* 14(7):505-514.

63. Berti F & Adamo R (2018) Antimicrobial glycoconjugate vaccines: an overview of classic and modern approaches for protein modification. *Chem Soc Rev* 47(24):9015-9025.

64. Frasch CE (2009) Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. *Vaccine* 27(46):6468-6470.

65. Terra VS, *et al.* (2012) Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design. *J Med Microbiol* 61(Pt 7):919-926.

66. Rappuoli R, De Gregorio E, & Costantino P (2019) On the mechanisms of conjugate vaccines. *Proc Natl Acad Sci U S A* 116(1):14-16.

67. Harding CM, *et al.* (2019) A platform for glycoengineering a polyvalent pneumococcal bioconjugate vaccine using E. coli as a host. *Nat Commun* 10(1):891.

68. Porstendorfer D, Gohl O, Mayer F, & Averhoff B (2000) ComP, a pilin-like protein essential for natural competence in Acinetobacter sp. Strain BD413: regulation, modification, and cellular localization. *J Bacteriol* 182(13):3673-3680.

69. Schulz BL, *et al.* (2013) Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. *PLoS One* 8(5):e62768.

70. Harding CM, *et al.* (2015) Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol Microbiol* 96(5):1023-1041.

71. Iwashkiw JA, *et al.* (2012) Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathog* 8(6):e1002758.

72. Geno KA, *et al.* (2015) Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clin Microbiol Rev* 28(3):871-899.

73. Carboni F, *et al.* (2017) Structure of a protective epitope of group B Streptococcus type III capsular polysaccharide. *Proc Natl Acad Sci U S A* 114(19):5017-5022.

74. Pan YJ, *et al.* (2015) Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of Klebsiella spp. *Sci Rep* 5:15573.

75. Feldman MF, *et al.* (2019) A promising bioconjugate vaccine against hypervirulent Klebsiella pneumoniae. *Proc Natl Acad Sci U S A*.

CLAIMS

What is claimed is:

1. A bioconjugate comprising an oligo- or polysaccharide covalently linked to a fusion protein:

wherein the fusion protein comprises a ComP protein (ComP) glycosylation tag;

wherein the ComP glycosylation tag comprises both a cysteine residue corresponding to the conserved cysteine residue at position 71 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1) and a cysteine residue corresponding to the conserved cysteine residue at position 93 of SEQ ID NO: 2 or both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1; and

wherein the fusion protein is glycosylated with the oligo- or polysaccharide on the ComP glycosylation tag at a serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 or position 84 of SEQ ID NO: 1.

2. The bioconjugate of Claim 1, wherein the ComP glycosylation tag does not comprise a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1);

optionally, wherein the fusion protein of the bioconjugate does not comprise, in relationship to the ComP glycosylation tag, a methionine residue at a position that would correspond to or correspond about to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

3. The bioconjugate of Claim 2, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

4. The bioconjugate of any one of Claims 1 to 3, wherein the ComP glycosylation tag has a length of between 18 and 50 amino acids in length; has a length of between 21 and 45 amino acids in length; or has a length of between 23 and 45 amino acids in length.

5. The bioconjugate of any one of Claims 1 to 4, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$);

optionally, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$) or SEQ ID NO: 8 ($ComP\Delta28_{110264}$);

optionally, wherein the ComP protein comprises SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$).

6. The bioconjugate of any one of Claims 1 to 4, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4

(Com$_{P50v1}$: PKD82822.1), SEQ ID NO: 5 (ComP$_{4466}$: SNX44537.1), or SEQ ID NO: 6 (ComP$_{SFC}$: OAL75955.1);

optionally, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 (ComP$_{ADP1}$: AAC45886.1) or SEQ ID NO: 2 (ComP$_{110264}$: ENV58402.1);

optionally, wherein the ComP protein comprises SEQ ID NO: 1 (ComP$_{ADP1}$: AAC45886.1), SEQ ID NO: 2 (ComP$_{110264}$: ENV58402.1), SEQ ID NO: 3 (ComP$_{GFJ-2}$: APV36638.1), SEQ ID NO: 4 (Com$_{P50v1}$: PKD82822.1), SEQ ID NO: 5 (ComP$_{4466}$: SNX44537.1), or SEQ ID NO: 6 (ComP$_{SFC}$: OAL75955.1).

7. The bioconjugate of Claim 1, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$X_1X_2GTX_5X_6X_7X_8X_9X_{10}X_{11}X_{12}$**C**$X_{14}GVX_{17}X_{18}IX_{20}X_{21}X_{22}$**AS**$X_{25}X_{26}TX_{28}NVX_{31}X_{32}AX_{34}$**C**

$X_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}X_{44}$ (SEQ ID NO: 27)

wherein:
X_1 is V, A, or no amino acid;
X_2 is A, G, T, or no amino acid;
X_5 is P, S, or Q;
X_6 is S, M, or I;
X_7 is T, P, or V;
X_8 is A, S, or T;
X_9 is G, N, S, or T;
X_{10} is N or no amino acid;
X_{11} is S, G, or A;
X_{12} is S or N;
X_{14} is V, T, or A;
X_{17} is Q, T, or E;
X_{18} is E, Q, or T;
X_{20} is S, N, A, or G;
X_{21} is S or no amino acid;
X_{22} is G or no amino acid;
X_{25} is N, S, or A;
X_{26} is A, S, or K;
X_{28} is T, S, or K;
X_{31} is A or E;
X_{32} is T or S;
X_{34} is T, Q, or A;

X_{36} is G, S, or T;
X_{37} is A, G, or D;
X_{38} is S, L, or A;
X_{39} is S, G, D, or T;
X_{40} is A, V, or G;
X_{41} is G, I, or V;
X_{42} is Q, T, or I;
X_{43} is I, V, T, or L; and
X_{44} is I, T, or V;

or a fragment of thereof, wherein the fragment of the ComP glycosylation tag comprises the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27;

or a variant of the amino acid consensus sequence of SEQ ID NO: 27 or the fragment thereof, having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27.

8. The bioconjugate of Claim 7, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$X_1X_2GTX_5X_6X_7X_8X_9X_{10}X_{11}X_{12}$**C**$X_{14}GVX_{17}X_{18}IX_{20}X_{21}X_{22}A$**S**$X_{25}X_{26}TX_{28}NVX_{31}X_{32}AX_{34}$**C**

$X_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}X_{44}$ (SEQ ID NO: 27)

wherein:
X_1 is V, A, or no amino acid;
X_2 is A, G, T, or no amino acid;
X_5 is P, S, or Q;
X_6 is S, M, or I;
X_7 is T, P, or V;
X_8 is A, S, or T;
X_9 is G, N, S, or T;
X_{10} is N or no amino acid;
X_{11} is S, G, or A;
X_{12} is S or N;
X_{14} is V, T, or A;
X_{17} is Q, T, or E;

X_{18} is E, Q, or T;
X_{20} is S, N, A, or G;
X_{21} is S or no amino acid;
X_{22} is G or no amino acid;
X_{25} is N, S, or A;
X_{26} is A, S, or K;
X_{28} is T, S, or K;
X_{31} is A or E;
X_{32} is T or S;
X_{34} is T, Q, or A;
X_{36} is G, S, or T;
X_{37} is A, G, or D;
X_{38} is S, L, or A;
X_{39} is S, G, D, or T;
X_{40} is A, V, or G;
X_{41} is G, I, or V;
X_{42} is Q, T, or I;
X_{43} is I, V, T, or L; and
X_{44} is I, T, or V;

or a fragment of thereof, wherein the fragment of the ComP glycosylation tag comprises the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27.

9. The bioconjugate of Claim 8, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

10. The bioconjugate of Claim 9, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 44 of SEQ ID NO: 27.

11. The bioconjugate of any one of Claims 8 to 10, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

12. The bioconjugate of Claim 1, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$CX_2GVX_5X_6IX_8X_9X_{10}ASX_{13}X_{14}TX_{16}NVX_{19}X_{20}AX_{22}C$ (SEQ ID NO: 28)

wherein:

X_2 is V, T, or A, optionally V;
X_5 is Q, T, or E, optionally Q;
X_6 is E, Q, or T;
X_8 is S, N, A, or G;
X_9 is S or no amino acid;
X_{10} is G or no amino acid;
X_{13} is N, S, or A, optionally N;
X_{14} is A, S, or K, optionally A;
X_{16} is T, S, or K;
X_{19} is A or E, optionally A;
X_{20} is T or S, optionally T; or
X_{22} is T, Q, or A, optionally T,

or a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains the cysteine residue at position 1 of SEQ ID NO: 28, the cysteine residue at position 23 of SEQ ID NO: 28, and the serine residue at position 12 of SEQ ID NO: 28.

13. The bioconjugate of Claim 12, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1)

14. The bioconjugate of Claim 13, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

15. The bioconjugate of any one of Claims 12 to 14, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

16. The bioconjugate of Claim 1, wherein the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11],

or a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1; and

wherein the variant maintains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1.

17. The bioconjugate of Claim 16, wherein the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID

NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11].

18. The bioconjugate of Claim 16 or 17, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

19. The bioconjugate of claim 18, wherein amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

20. The bioconjugate of any one of Claims 16 to 19, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

21. The bioconjugate of Claim 1, wherein the ComP glycosylation tag consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ

ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11].

22. The bioconjugate of any one of Claims 1 to 21, wherein the oligo- or polysaccharide is produced by a bacteria from the genus *Streptococcus*;

optionally, wherein the polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide.

23. The bioconjugate of Claim 22, wherein the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b.

24. The bioconjugate of any one of Claims 1 to 21, wherein the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*;

optionally, wherein the polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varricola*, *Klebsiella michinganenis*, or *Klebsiella oxytoca* capsular polysaccharide.

25. The bioconjugate of any one of Claims 1 to 21, wherein the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide.

26. The bioconjugate of Claim 25, wherein the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*.

27. The bioconjugate of any one of Claims 1 to 26, wherein the oligo- or polysaccharide comprises a glucose at its reducing end.

28. The bioconjugate of any one of Claims 1 to 27, wherein the bioconjugate is produced *in vivo*;

optionally, in a bacterial cell.

29. The bioconjugate of any one of Claims 1 to 28, wherein the fusion protein comprises a carrier protein selected from the group consisting of diphtheria toxoid CRM197, tetanus toxoid, *Pseudomonas aeruginosa* Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, and *Haemophilus influenza* protein D, or a fragment thereof.

30. The bioconjugate of Claim 29, wherein the ComP glycosylation tag is located at the N-terminal end of the fusion protein, at the C-terminal end of the fusion protein, and/or internally within the fusion protein;

optionally, wherein the carrier protein or fragment thereof is linked to the glycosylation tag via an amino acid linker.

31. The bioconjugate of any one of Claims 1 to 30, wherein the fusion protein comprises two or more, three or more, four or more, five or more, six or more, eight or more, ten or more, fifteen or more, or twenty or more ComP glycosylation tags.

32. The bioconjugate of any one of Claims 1 to 30, wherein the fusion protein comprises any of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 20 to any of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, or 25 ComP glycosylation tags.

33. The bioconjugate of Claim 31 or 32, wherein the ComP glycosylation tags are identical.

34. The bioconjugate of Claim 31 or 32, wherein at least two of the ComP glycosylation tags differ from each other,

optionally, wherein at least three, at least four, or at least five of the ComP glycosylation tags all differ from each other,

optionally, wherein none of the ComP glycosylation tags are the same.

35. The bioconjugate of any one of Claims 1 to 34, wherein the bioconjugate is a conjugate vaccine;

optionally, wherein the conjugate vaccine is a vaccine against *Streptococcus pneumoniae* serotype 8.

36. The bioconjugate of Claim 35, wherein when the conjugate vaccine induces an immune response when administered to a subject.

37. The bioconjugate of Claim 35 or 36, wherein the immune response elicits long term memory (memory B and T cells), is an antibody response, and is optionally a serotype-specific antibody response.

38. The bioconjugate of claim 37, wherein the antibody response is an IgG or IgM response.

39. The bioconjugate of claim 38, wherein the antibody response is an IgG response; optionally an IgG1 response.

40. The bioconjugate of any one of claims 35 to 39, wherein the conjugate vaccine generates immunological memory in a subject administered the vaccine.

41. A ComP glycosylation tag comprising an isolated fragment of a ComP protein, wherein the fragment comprises a serine residue corresponding to the conserved serine residue at position 84 in SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and both a cysteine residue corresponding to the conserved cysteine residue at position 71 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1) and a cysteine residue corresponding to the conserved cysteine residue at position 93 of SEQ ID NO: 2 or both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1.

42. The ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag does not comprise a methionine residue corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

43. The ComP glycosylation tag of Claim 42, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

44. The ComP glycosylation tag of any one of Claims 41 to 43, wherein the ComP glycosylation tag has a length of between 18 and 50 amino acids in length; has a length of between 21 and 45 amino acids in length; or has a length of between 23 and 45 amino acids in length.

45. The ComP glycosylation tag of any one of Claims 41 to 44, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$);

optionally, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$) or SEQ ID NO: 8 ($ComP\Delta28_{110264}$);

optionally, wherein the ComP protein comprises SEQ ID NO: 7 ($ComP\Delta28_{ADP1}$), SEQ ID NO: 8 ($ComP\Delta28_{110264}$), SEQ ID NO: 9 ($ComP\Delta28_{GFJ-2}$), SEQ ID NO: 10 ($ComP\Delta28_{P50v1}$), SEQ ID NO: 11 ($ComP\Delta28_{4466}$), or SEQ ID NO: 12 ($ComP\Delta28_{SFC}$).

46. The ComP glycosylation tag of any one of Claims 41 to 44, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4 (Com_{P50v1}: PKD82822.1), SEQ ID NO: 5 ($ComP_{4466}$: SNX44537.1), or SEQ ID NO: 6 ($ComP_{SFC}$: OAL75955.1);

optionally, wherein the ComP protein comprises an amino acid sequence that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) or SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1);

optionally, wherein the ComP protein comprises SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1), SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1), SEQ ID NO: 3 ($ComP_{GFJ-2}$: APV36638.1), SEQ ID NO: 4 (Com_{P50v1}: PKD82822.1), SEQ ID NO: 5 ($ComP_{4466}$: SNX44537.1), or SEQ ID NO: 6 ($ComP_{SFC}$: OAL75955.1).

47. The ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$X_1X_2GTX_5X_6X_7X_8X_9X_{10}X_{11}X_{12}\underline{C}X_{14}GVX_{17}X_{18}IX_{20}X_{21}X_{22}ASX_{25}X_{26}TX_{28}NVX_{31}X_{32}AX_{34}\underline{C}$
$X_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}X_{44}$ (SEQ ID NO: 27)

wherein:
X_1 is V, A, or no amino acid;
X_2 is A, G, T, or no amino acid;
X_5 is P, S, or Q;
X_6 is S, M, or I;
X_7 is T, P, or V;
X_8 is A, S, or T;
X_9 is G, N, S, or T;
X_{10} is N or no amino acid;
X_{11} is S, G, or A;
X_{12} is S or N;
X_{14} is V, T, or A;
X_{17} is Q, T, or E;
X_{18} is E, Q, or T;
X_{20} is S, N, A, or G;
X_{21} is S or no amino acid;
X_{22} is G or no amino acid;
X_{25} is N, S, or A;
X_{26} is A, S, or K;
X_{28} is T, S, or K;
X_{31} is A or E;
X_{32} is T or S;
X_{34} is T, Q, or A;
X_{36} is G, S, or T;
X_{37} is A, G, or D;
X_{38} is S, L, or A;
X_{39} is S, G, D, or T;

X_{40} is A, V, or G;
X_{41} is G, I, or V;
X_{42} is Q, T, or I;
X_{43} is I, V, T, or L; and
X_{44} is I, T, or V;

or a fragment of thereof, wherein the fragment of the ComP glycosylation tag comprises the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27;

or a variant of the amino acid consensus sequence of SEQ ID NO: 27 or the fragment thereof, having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27.

48. The ComP glycosylation tag of Claim 47, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$X_1X_2GTX_5X_6X_7X_8X_9X_{10}X_{11}X_{12}$**C**$X_{14}GVX_{17}X_{18}IX_{20}X_{21}X_{22}A$**S**$X_{25}X_{26}TX_{28}NVX_{31}X_{32}AX_{34}$**C**
$X_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}X_{44}$ (SEQ ID NO: 27)

wherein:
X_1 is V, A, or no amino acid;
X_2 is A, G, T, or no amino acid;
X_5 is P, S, or Q;
X_6 is S, M, or I;
X_7 is T, P, or V;
X_8 is A, S, or T;
X_9 is G, N, S, or T;
X_{10} is N or no amino acid;
X_{11} is S, G, or A;
X_{12} is S or N;
X_{14} is V, T, or A;
X_{17} is Q, T, or E;
X_{18} is E, Q, or T;
X_{20} is S, N, A, or G;
X_{21} is S or no amino acid;
X_{22} is G or no amino acid;

X_{25} is N, S, or A;
X_{26} is A, S, or K;
X_{28} is T, S, or K;
X_{31} is A or E;
X_{32} is T or S;
X_{34} is T, Q, or A;
X_{36} is G, S, or T;
X_{37} is A, G, or D;
X_{38} is S, L, or A;
X_{39} is S, G, D, or T;
X_{40} is A, V, or G;
X_{41} is G, I, or V;
X_{42} is Q, T, or I;
X_{43} is I, V, T, or L; and
X_{44} is I, T, or V;

or a fragment of thereof, wherein the fragment of the ComP glycosylation tag comprises the cysteine residue at position 13 of SEQ ID NO: 27, the cysteine residue at position 35 of SEQ ID NO: 27, and the serine residue at position 24 of SEQ ID NO: 27.

49. The ComP glycosylation tag of Claim 48, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

50. The ComP glycosylation tag of Claim 49, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

51. The ComP glycosylation tag of any one of Claims 48 to 50, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

52. The ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag comprises or consists of the amino acid consensus sequence of:

$CX_2GVX_5X_6IX_8X_9X_{10}ASX_{13}X_{14}TX_{16}NVX_{19}X_{20}AX_{22}C$ (SEQ ID NO: 28)

wherein:

X_2 is V, T, or A, optionally V;
X_5 is Q, T, or E, optionally Q;
X_6 is E, Q, or T;
X_8 is S, N, A, or G;
X_9 is S or no amino acid;
X_{10} is G or no amino acid;
X_{13} is N, S, or A, optionally N;
X_{14} is A, S, or K, optionally A;
X_{16} is T, S, or K;
X_{19} is A or E, optionally A;
X_{20} is T or S, optionally T; or
X_{22} is T, Q, or A, optionally T,

or a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains the cysteine residue at position 13 of SEQ ID NO: 28, the cysteine residue at position 35 of SEQ ID NO: 28, and the serine residue at position 24 of SEQ ID NO: 28.

53. The ComP glycosylation tag of Claim 52, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1)

54. The ComP glycosylation tag of Claim 53, wherein the amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

55. The ComP glycosylation tag of any one of Claims 52 to 54, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

56. The ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11],

or a variant thereof having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions,

wherein the variant maintains both a cysteine residue corresponding to the conserved cysteine residue at position 75 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1) and a cysteine residue corresponding to the conserved cysteine residue at position 95 of SEQ ID NO: 1; and

wherein the variant maintains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1.

57. The ComP glycosylation tag of Claim 56, wherein the ComP glycosylation tag comprises or consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32

[C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43 [G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11].

58. The ComP glycosylation tag of Claim 56 or 57, wherein the ComP glycosylation tag does not comprise a methionine residue in a position corresponding to the conserved methionine residue at position 104 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

59. The ComP glycosylation tag of claim 58, wherein amino acid sequence of the ComP glycosylation tag does not extend in the C-terminus direction beyond the amino acid residue corresponding to position 103 of SEQ ID NO: 2 ($ComP_{110264}$: ENV58402.1).

60. The ComP glycosylation tag of any one of Claims 56 to 59, wherein the ComP glycosylation tag is not more than 25, 30, 40, 45, or 50 amino acids in length.

61. The ComP glycosylation tag of Claim 41, wherein the ComP glycosylation tag consists of an amino acid sequence selected from the group consisting of: SEQ ID NO: 32 [C1]; SEQ ID NO: 33 [D1]; SEQ ID NO: 34 [E1]; SEQ ID NO: 41 [E2]; SEQ ID NO: 42 [F2]; SEQ ID NO: 43

[G2]; SEQ ID NO: 44 [H2]; SEQ ID NO: 45 [A3]; SEQ ID NO: 46 [B3]; SEQ ID NO: 47 [C3]; SEQ ID NO: 55 [D4]; SEQ ID NO: 56 [E4]; SEQ ID NO: 57 [F4]; SEQ ID NO: 58 [G4]; SEQ ID NO: 59 [A5]; SEQ ID NO: 60 [B5]; SEQ ID NO: 61 [D5]; SEQ ID NO: 62 [E5]; SEQ ID NO: 63 [F5]; SEQ ID NO: 72 [H6]; SEQ ID NO: 73 [B7]; SEQ ID NO: 74 [C7]; SEQ ID NO: 75 [D7]; SEQ ID NO: 76 [E7]; SEQ ID NO: 77 [F7]; SEQ ID NO: 78 [A8]; SEQ ID NO: 79 [B8]; SEQ ID NO: 92 [A10]; SEQ ID NO: 93 [B10]; SEQ ID NO: 94 [C10]; SEQ ID NO: 95 [D10]; SEQ ID NO: 96 [F10]; SEQ ID NO: 97 [G10]; SEQ ID NO: 98 [H10]; SEQ ID NO: 99 [A11]; SEQ ID NO: 100 [B11]; and SEQ ID NO: 101 [C11].

62. A fusion protein comprising the ComP glycosylation tag of any of Claims 41 to 61;

optionally, wherein the fusion protein is glycosylated at a serine residue on the glycosylation tag corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC45886.1).

63. The fusion protein of Claim 62, wherein the fusion protein is glycosylated with an oligo- or polysaccharide and wherein the oligo- or polysaccharide is produced by a bacteria from the genus *Streptococcus*;

optionally, wherein the polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide.

64. The fusion protein of Claim 63, wherein the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b.

65. The fusion protein of Claim 62, wherein the fusion protein is glycosylated with an oligo- or polysaccharide and wherein the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*;

optionally, wherein the polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varricola*, *Klebsiella michinganenis*, or *Klebsiella oxytoca* capsular polysaccharide.

66. The fusion protein of Claim 65, wherein the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide.

67. The fusion protein of Claim 66, wherein the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*.

68. The fusion protein of any one of Claims 62 to 67, wherein the oligo- or polysaccharide comprises a glucose at its reducing end.

69. The fusion protein of any one of Claims 62 to 68, wherein the fusion protein is produced *in vivo*;

optionally, in a bacterial cell.

70. The fusion protein of any one of Claims 62 to 69, wherein the fusion protein comprises a carrier protein selected from the group consisting of diphtheria toxoid CRM197, tetanus toxoid, *Pseudomonas aeruginosa* Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, and *Haemophilus influenza* protein D, or a fragment thereof.

71. The fusion protein of Claim 70, wherein the ComP glycosylation tag is located at the N-terminal end of the fusion protein, at the C-terminal end of the fusion protein, and/or internally within the fusion protein;

optionally, wherein the carrier protein or fragment thereof is linked to the glycosylation tag via an amino acid linker.

72. The fusion protein of any one of Claims 62 to 71, wherein the fusion protein comprises two or more, three or more, four or more, five or more, six or more, eight or more, ten or more, fifteen or more, or twenty or more ComP glycosylation tags.

73. The fusion protein of any one of Claims 62 to 71, wherein the fusion protein comprises any of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 20 to any of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, or 25 ComP glycosylation tags.

74. The fusion protein of Claim 72 or 73, wherein the ComP glycosylation tags are identical.

75. The fusion protein of Claim 72 or 73, wherein at least two of the ComP glycosylation tags differ from each other,

optionally, wherein at least three, at least four, or at least five of the ComP glycosylation tags all differ from each other,

optionally, wherein none of the ComP glycosylation tags are the same.

76. A method of *in vivo* conjugation of an oligo- or polysaccharide to an acceptor polypeptide, the method comprising covalently linking the oligo- or polysaccharide to the

acceptor polypeptide with a PglS oligosaccharyltransferase (OTase), wherein the acceptor polypeptide comprises the ComP glycosylation tag of any one of Claims 41 to 61;

optionally, wherein the ComP glycosylation tag is linked to a heterologous carrier protein.

77. The method of Claim 76, wherein the PglS OTase is $PglS_{110264}$, $PglS_{ADP1}$, $PglS_{GFJ-2}$, $PglS_{50v1}$, $PglS_{4466}$, or $PglS_{SFC}$.

78. The method of Claim 76 or 77, wherein the oligo- or polysaccharide is linked to the ComP glycosylation tag at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 ($ComP_{ADP1}$: AAC4588631).

79. The method of any one of Claims 76 to 78, wherein the *in vivo* conjugation occurs in a host cell.

80. The method of Claim 79, wherein the host cell is a bacterial cell.

81. The method of Claim 80, wherein the bacterial host cell is *E. coli*.

82. The method of any one of Claims 79 to 81 comprising culturing a host cell that comprises: (a) a genetic cluster encoding for the proteins required to synthesize the oligo- or polysaccharide; (b) a PglS OTase; and (3) the acceptor polypeptide.

83. The method of any one of Claims 76 to 82, wherein production of the oligo- or polysaccharide is enhanced by the *K. pneumoniae* transcriptional activator rmpA (*K. pneumoniae*

NTUH K-2044) or a homolog of the *K. pneumoniae* transcriptional activator rmpA (*K. pneumoniae* NTUH K-2044).

84. The method of any one of Claims 76 to 83, wherein the method produces a conjugate vaccine.

85. A host cell comprising (a) a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide; (b) a PglS OTase; and (3) an acceptor polypeptide comprising the ComP glycosylation tag of any one of Claims 41 to 61.

86. The host cell of Claim 85, wherein the acceptor polypeptide is a fusion protein.

87. The host cell of Claim 85 or Claim 86, wherein the host cell comprises a nucleic acid encoding the PglS OTase.

88. The host cell of any one of Claims 85 to 87, wherein the host cell comprises a nucleic acid encoding the acceptor polypeptide.

89. An isolated nucleic acid encoding the ComP glycosylation tag of any one of Claims 41 to 61 and/or the fusion protein of any one of Claims 62 to 75.

90. The isolated nucleic acid of Claim 78, wherein the nucleic acid is a vector.

91. A host cell comprising the isolated nucleic acid of Claim 78 or 79.

92. A composition comprising the conjugate vaccine of any one of Claims 35 to 40 or the fusion protein of any one of Claims 62 to 75, and an adjuvant.

93. A method of inducing a host immune response against a bacterial pathogen, the method comprising administering to a subject in need of the immune response an effective amount of the conjugate vaccine of any one of Claims 35 to 40, the fusion protein of any one of Claims 62 to 75, or the composition of Claim 92.

94. The method of Claim 93, wherein the immune response is an antibody response.

95. The method of claim 94, wherein the immune response is selected from the group consisting of an innate response, an adaptive response, a humoral response, an antibody response, cell mediated response, a B cell response, a T cell response, cytokine upregulation or downregulation, immune system cross-talk, and a combination of two or more of said immune responses.

96. The method of claim 95, wherein the immune response is selected from the group consisting of an innate response, a humoral response, an antibody response, a T cell response, and a combination of two or more of said immune responses.

97. A method of preventing or treating a bacterial disease and/or infection in a subject comprising administering to a subject in need thereof the conjugate vaccine of any one of Claims 35 to 40, the fusion protein of any one of Claims 62 to 75, or the composition of Claim 92.

98. The method of Claim 97, wherein the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature.

99. The method of Claim 97, wherein the disease is pneumonia.

100. The method of Claim 98, wherein the infection is a systemic infection and/or an infection of the blood.

101. The method of any one of Claims 97 to 100, wherein the subject is a human.

102. The method of any one of claims 97 to 101, wherein the composition is administered via intramuscular injection, intradermal injection, intraperitoneal injection, subcutaneous injection, intravenous injection, oral administration, mucosal administration, intranasal administration, or pulmonary administration.

103. A method of producing a pneumococcal conjugate vaccine against pneumococcal infection, the method comprising:

(a) isolating the bioconjugate of any one of Claims 1 to 40 or a glycosylated fusion protein of any one of Claims 62 to 75; and

(b) combining the isolated conjugate vaccine or isolated glycosylated fusion protein with an adjuvant.

A.

$ComP_{ADP1}$ fragment: $_{75}$CVGVQEISASNATTNVATATC$_{95}$ (SEQ ID NO: 26)

B.

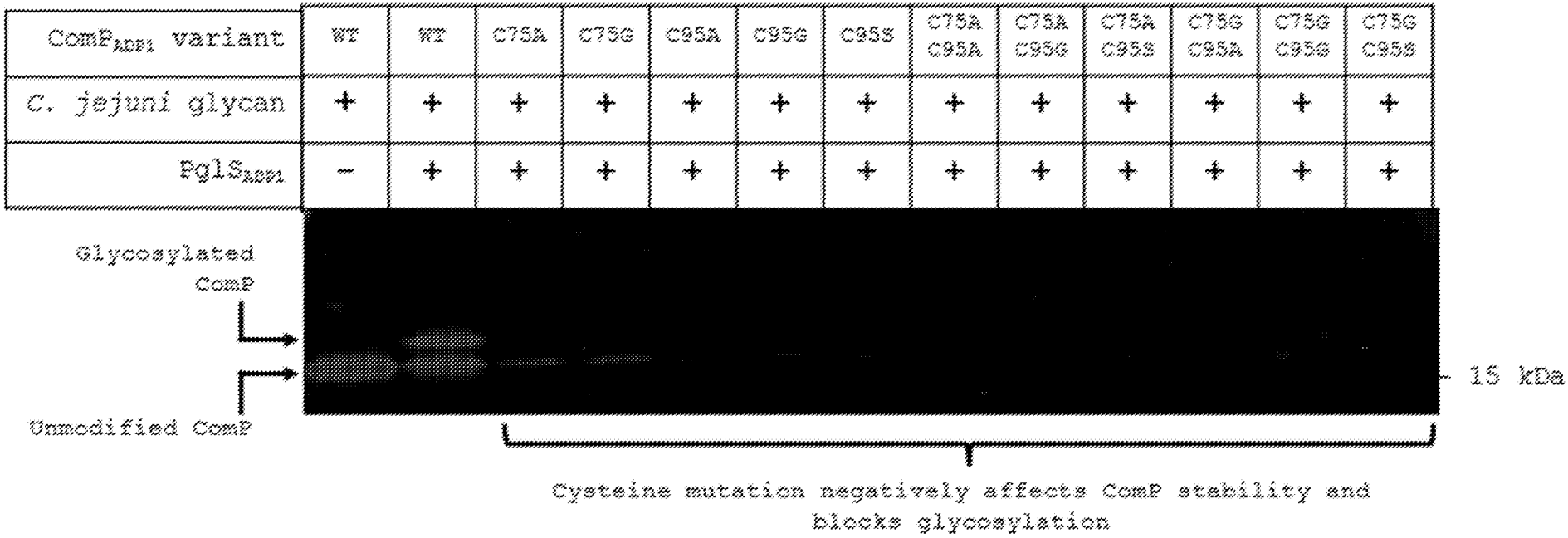


Figure 1A,B

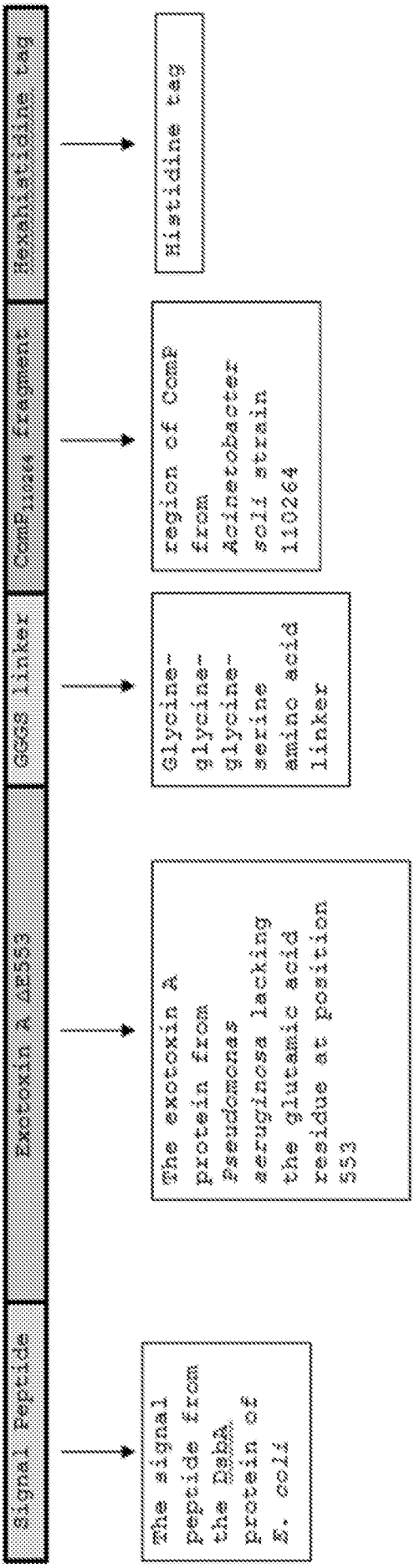
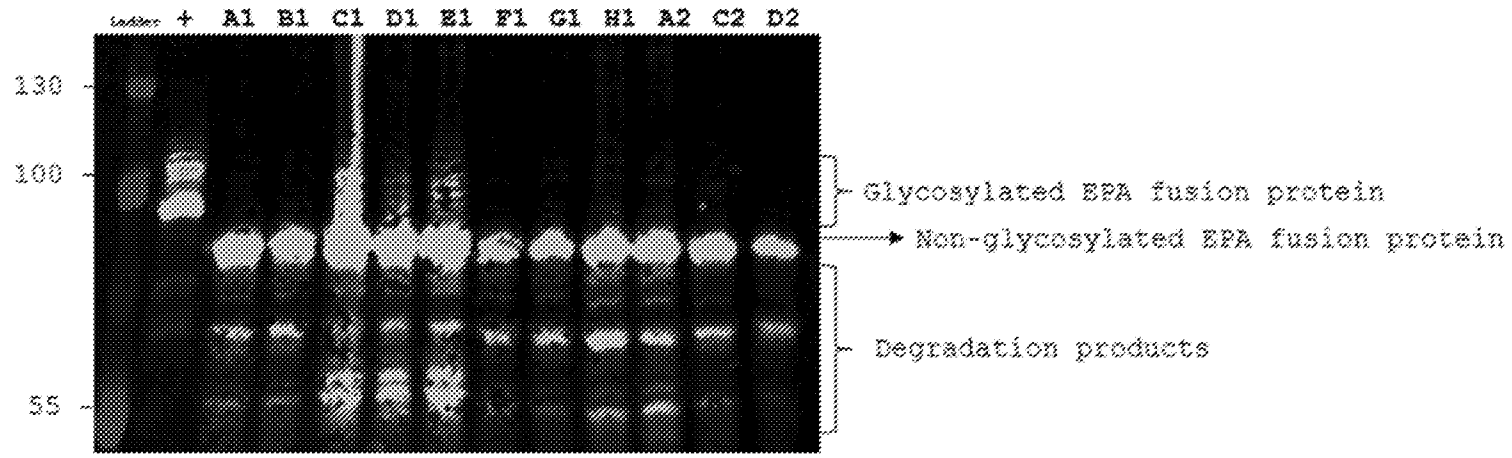

Signal Peptide
Exotoxin A (E553
GGGS linker
Complase fragment
Hexahistidine tag
The signal peptide from the DsbA protein of E. coli
The exotoxin A protein from Pseudomonas aeruginosa lacking the glutamic acid residue at position 553
glycine-glycine-glycine-serine amino acid linker
region of Comp from Acinetobacter soli strain 110264
Histidine tag

Figure 2

A.



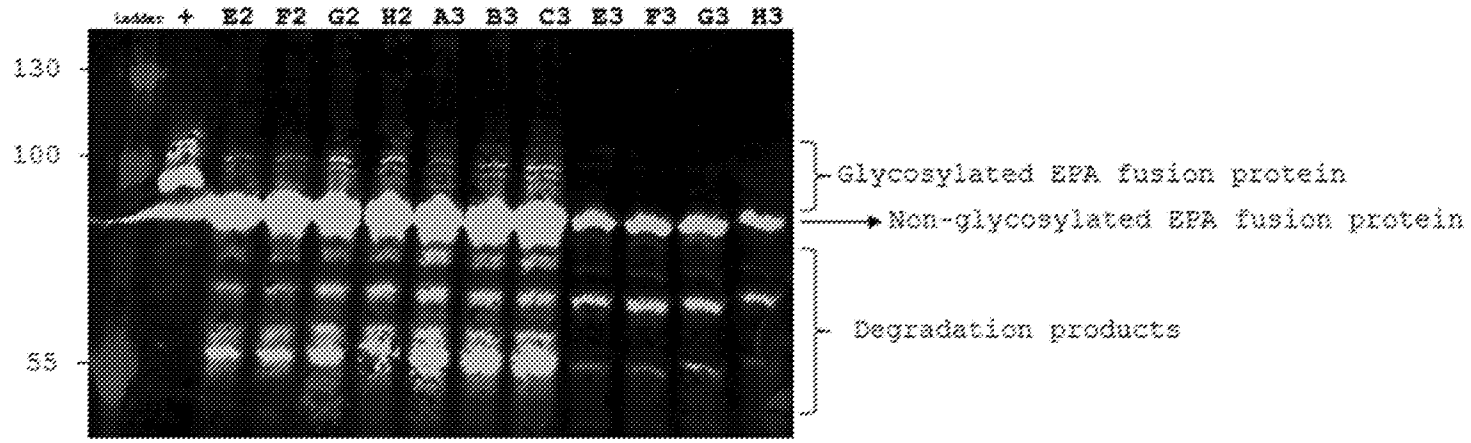
1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

B.

Lane ID	ComP$_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCS DSDGVITVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{148}$	Yes
A1	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASA$_{93}$	No
B1	$_{68}$SGNCTGVTQIASGA**S**AATTNVASAQ$_{94}$	No
C1	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQC$_{95}$	Yes
D1	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCS$_{96}$	Yes
E1	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSD$_{97}$	Yes
F1	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDS$_{98}$	No
G1	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSD$_{99}$	No
H1	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDG$_{100}$	No
A2	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGV$_{101}$	No
C2	$_{67}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASA$_{93}$	No
D2	$_{68}$TSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{94}$	No

Figure 3A,B

A.



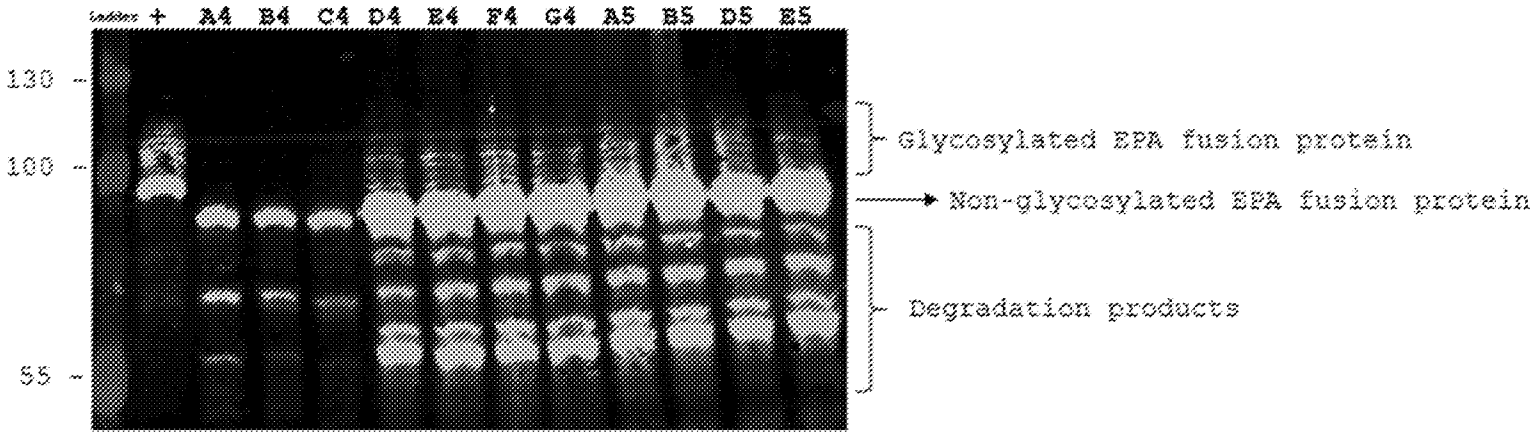
1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

B.

Lane ID	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{20}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{148}$	Yes
E2	$_{44}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{83}$	Yes
F2	$_{45}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{84}$	Yes
G2	$_{46}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{85}$	Yes
H2	$_{47}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{86}$	Yes
A3	$_{48}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSD$_{87}$	Yes
B3	$_{49}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{88}$	Yes
C3	$_{50}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{89}$	Yes
E3	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	No
F3	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	No
G3	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	No
H3	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	No

Figure 4A,B

A.



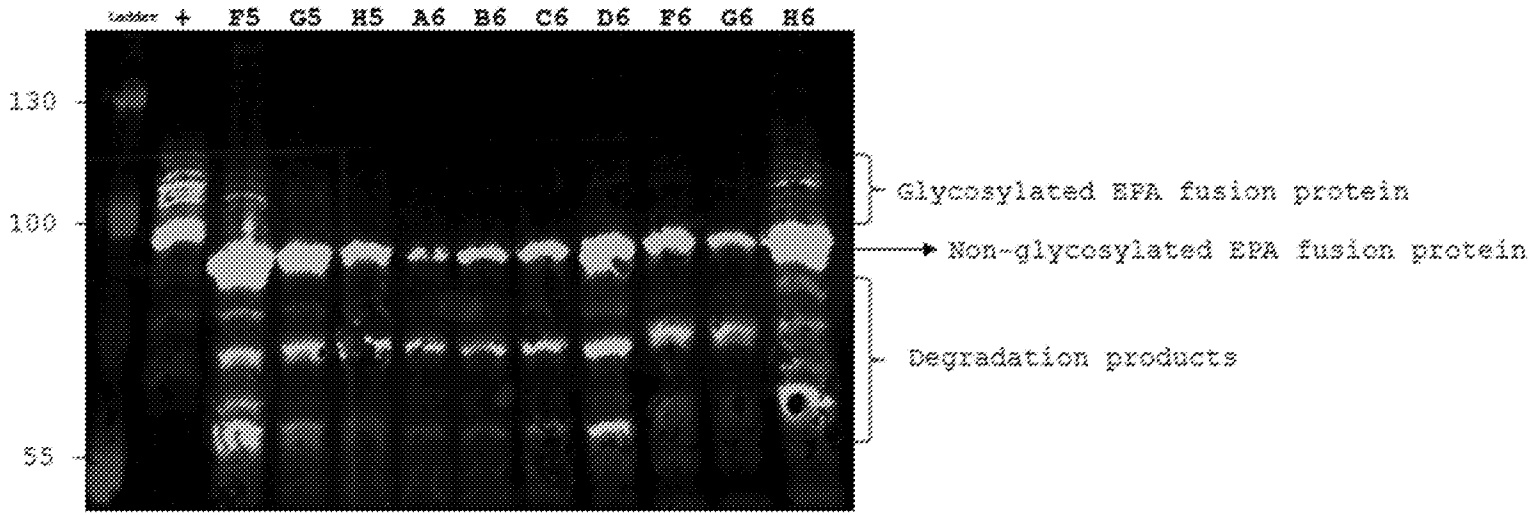
1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

B.

Lane ID	ComP$_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{20}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{115}$	Yes
A4	$_{26}$QIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{108}$	No
B4	$_{37}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASA$_{91}$	No
C4	$_{38}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	No
D4	$_{39}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	Yes
E4	$_{40}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	Yes
F4	$_{41}$GGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	Yes
G4	$_{42}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	Yes
A5	$_{44}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	Yes
B5	$_{45}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	Yes
D5	$_{47}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	Yes
E5	$_{48}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	Yes

Figure 5A,B

A.

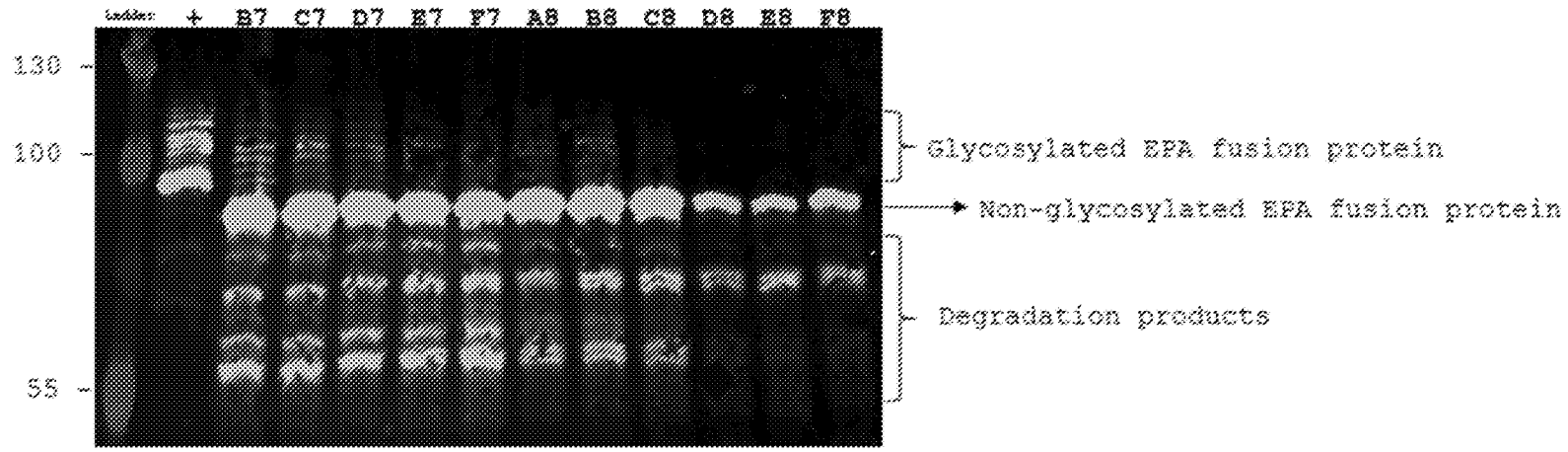


B.

Lane ID	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMFSSGNCTGVTQIASGASAATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{145}$	Yes
F5	$_{69}$GNCTGVTQIASGASAATTNVASAQCSDSDGVITVT$_{103}$	Yes
G5	$_{70}$NCTGVTQIASGASAATTNVASAQCSDSDGVITVTM$_{104}$	No
H5	$_{71}$CTGVTQIASGASAATTNVASAQCSDSDGVITVTMT$_{105}$	No
A6	$_{72}$TGVTQIASGASAATTNVASAQCSDSDGVITVTMTD$_{106}$	No
B6	$_{73}$GVTQIASGASAATTNVASAQCSDSDGVITVTMTDK$_{107}$	No
C6	$_{74}$VTQIASGASAATTNVASAQCSDSDGVITVTMTDKA$_{108}$	No
D6	$_{75}$TQIASGASAATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	No
F6	$_{87}$TVSENIMNAGGTSMFSSGNCTGVTQIASGASAATTNVASA$_{121}$	No
G6	$_{88}$VSENIMNAGGTSMFSSGNCTGVTQIASGASAATTNVASAQ$_{122}$	No
H6	$_{89}$SENIMNAGGTSMFSSGNCTGVTQIASGASAATTNVASAQC$_{123}$	Yes

Figure 6A,B

A.



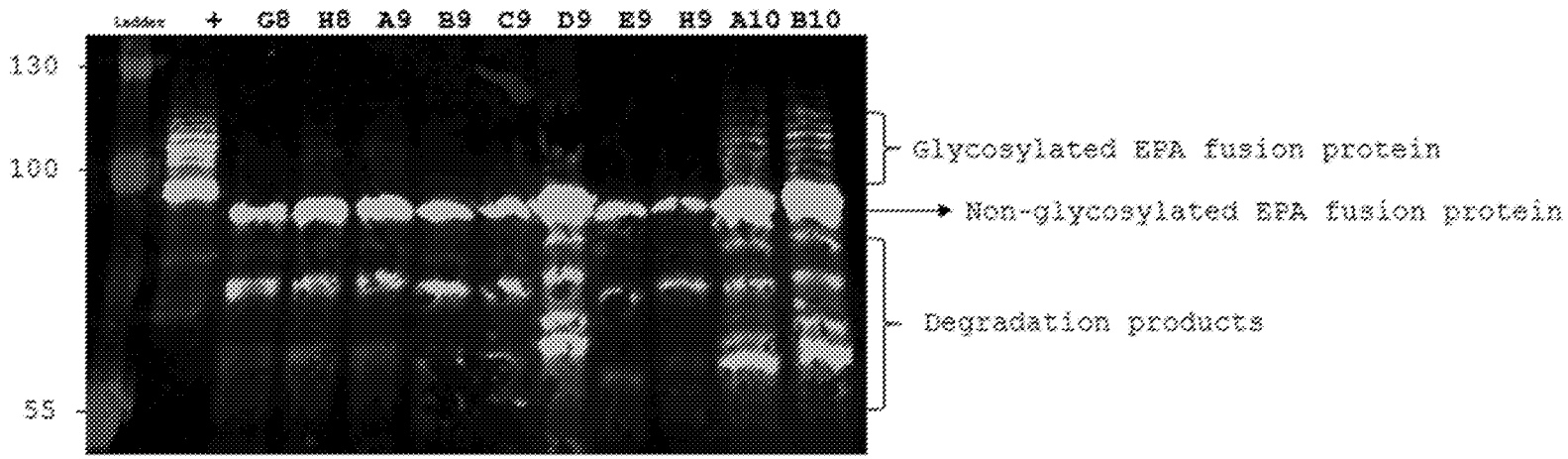
1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

B.

Lane ID	ComP$_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLITASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{145}$	Yes
B7	$_{56}$NIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{98}$	Yes
C7	$_{57}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{99}$	Yes
D7	$_{58}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSD$_{100}$	Yes
E7	$_{59}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{101}$	Yes
F7	$_{60}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{102}$	Yes
A8	$_{61}$TSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{103}$	Yes
B8	$_{62}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{104}$	Yes
C8	$_{63}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{105}$	No
D8	$_{64}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{106}$	No
E8	$_{65}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTD$_{107}$	No
F8	$_{66}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDK$_{108}$	No

Figure 7A,B

A.



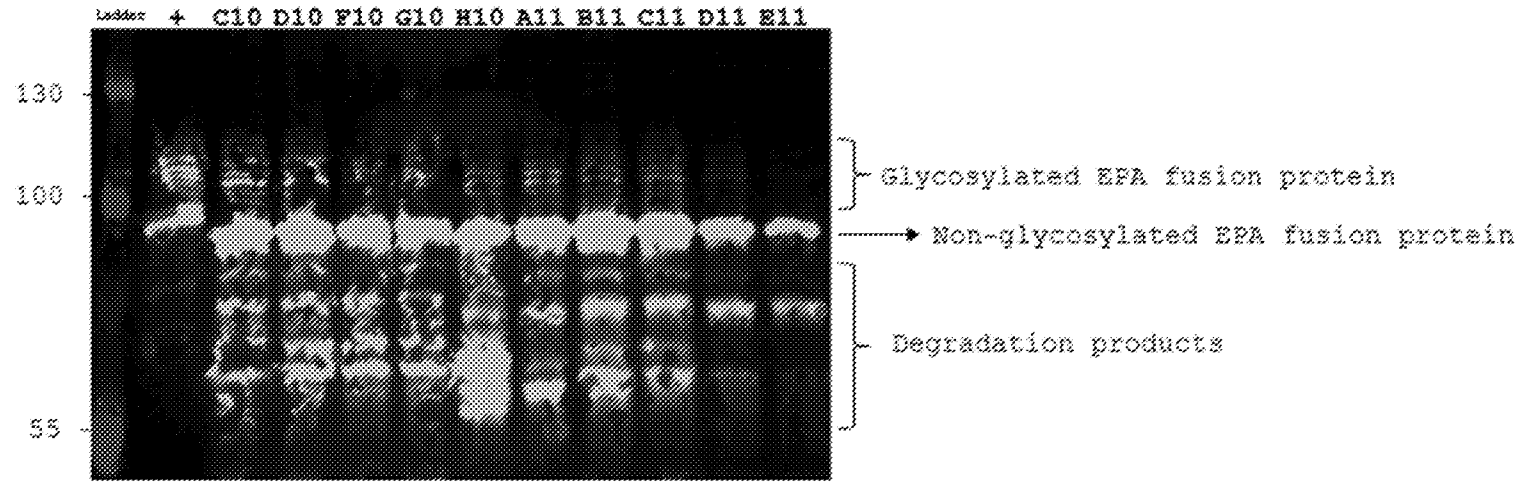
1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

B.

Lane ID	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{115}$	Yes
G8	$_{69}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKA$_{108}$	No
H8	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	No
A9	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKG$_{110}$	No
B9	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGV$_{111}$	No
C9	$_{73}$GVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVS$_{112}$	No
D9	$_{74}$VTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSI$_{113}$	No
E9	$_{75}$TQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIK$_{114}$	No
H9	$_{48}$AMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQ$_{92}$	No
A10	$_{49}$MKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQC$_{93}$	Yes
B10	$_{50}$KATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCS$_{94}$	Yes

Figure 8A,B

A.

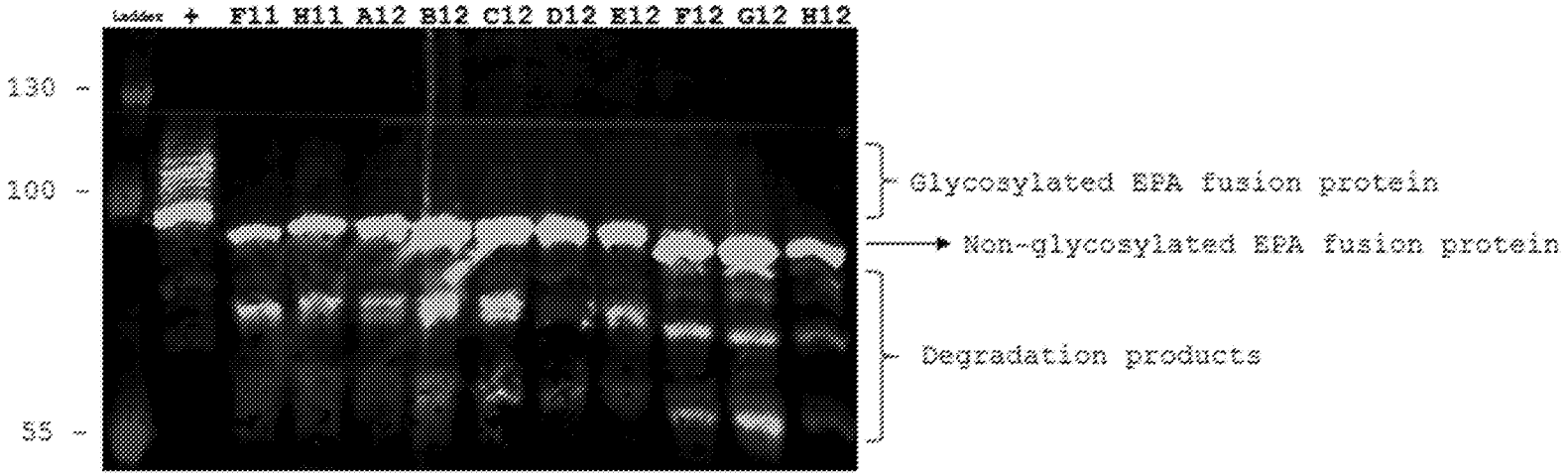


B.

Lane ID	$ComP_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSSCRGT$_{148}$	Yes
C10	$_{51}$ATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD$_{95}$	Yes
D10	$_{52}$TVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDS$_{96}$	Yes
F10	$_{54}$SENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDG$_{98}$	Yes
G10	$_{55}$ENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGV$_{99}$	Yes
H10	$_{56}$NIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI$_{100}$	Yes
A11	$_{57}$IMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVIT$_{101}$	Yes
B11	$_{58}$MNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITV$_{102}$	Yes
C11	$_{59}$NAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVT$_{103}$	Yes
D11	$_{60}$AGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTM$_{104}$	No
E11	$_{61}$GGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT$_{105}$	No

Figure 9A,B

A.

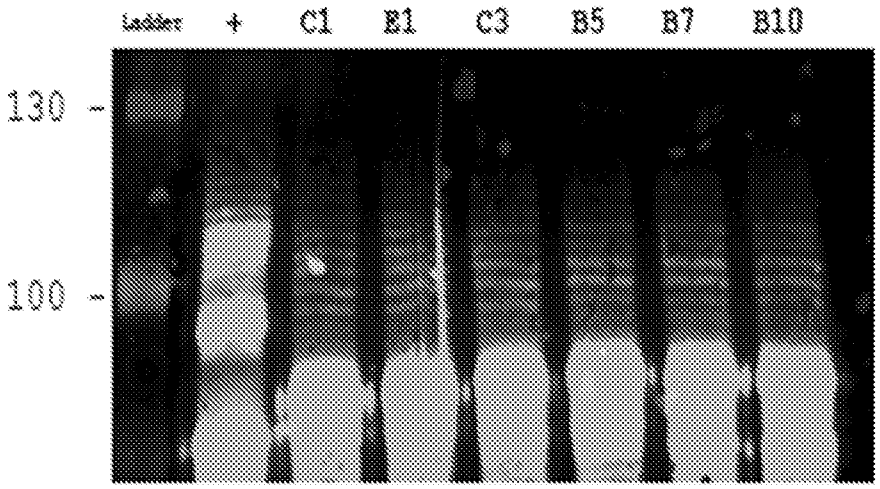


B.

Lane ID	ComP$_{110264}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{29}$AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGT$_{145}$	Yes
F11	$_{82}$GTSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTD$_{106}$	No
H11	$_{84}$SMPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKA$_{108}$	No
A12	$_{85}$MPSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAK$_{109}$	No
B12	$_{86}$PSSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKG$_{110}$	No
C12	$_{87}$SSGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGV$_{111}$	No
D12	$_{88}$SGNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVS$_{112}$	No
E12	$_{89}$GNCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSI$_{113}$	No
F12	$_{70}$NCTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIK$_{114}$	No
G12	$_{71}$CTGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIKL$_{115}$	No
H12	$_{72}$TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMTDKAKGVSIKLT$_{116}$	No

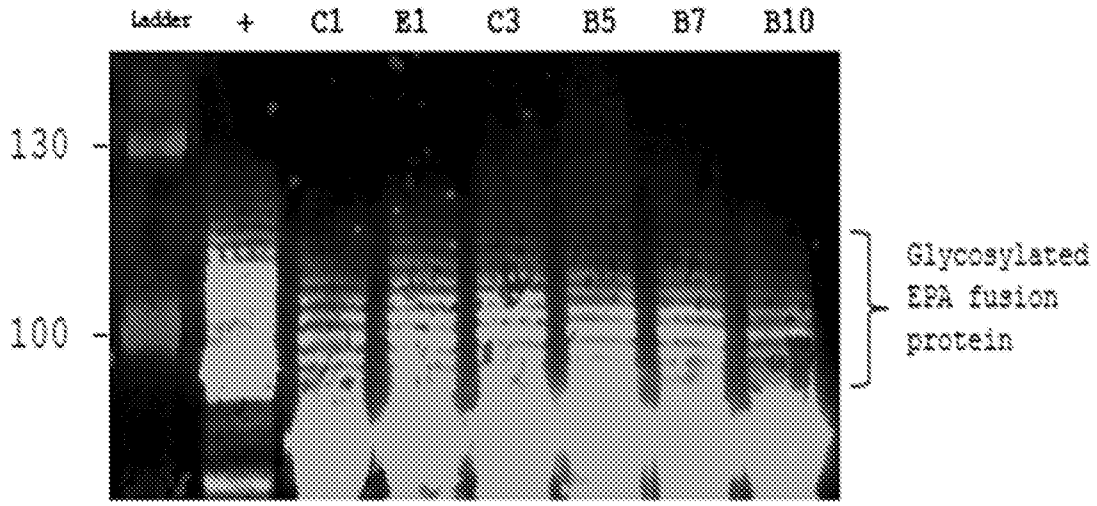
Figure 10A,B

A.



1° Antibody: rabbit polyclonal anti-exotoxin A antisera
2° Antibody: goat anti-rabbit IgG antisera

B.



1° Antibody: rabbit polyclonal anti-His antisera
2° Antibody: goat anti-rabbit IgG antisera

Figure 11A,B

C,

Lane ID	$ComP_{110Z64}$ fragment fused to C-terminus of EPA	Glycosylation Observed
+	$_{2}$AYTDYTVRSRVTEGLMTASAMKATVSENIMAGGTSMPSSGNCTGVTQIASGASAATTNVASAQCSDSDGVI TVTMTDKAKGVSIKLTPSFSSTGSVSMKCTYSSDKKYVPSEGRGT$_{145}$	Yes
C1	$_{6}$GNCTGVTQIASGASAATTNVASAQC$_{68}$	Yes
E1	$_{71}$CTGVTQIASGASAATTNVASAQCSD$_{85}$	Yes
C3	$_{79}$NCTGVTQIASGASAATTNVASAQCSDSDGV$_{99}$	Yes
B5	$_{6}$MPSSGNCTGVTQIASGASAATTNVASAQCSDSDGV$_{99}$	Yes
B7	$_{5}$NTTMNAGGTSMPSSGNCTGVTQIASGASAATTNVASAQCSD$_{95}$	Yes
B10	$_{5}$KATVSENTMNAGGTSMPSSGNCTGVTQIASGASAATTNVASAQCS$_{94}$	Yes

Figure 11C

A.

GlycoTag ID	ComP110264 Fragment
D5	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASA QCSDSDGVIT$_{101}$
D5'	$_{67}$SSGNCTGVTQIASGA**S**AATTNVASA QCSDSDG$_{93}$

B.

EPA Variant Legend

1 = EPA-GGGS-ComPΔ28$_{110264}$-His6X
2 = EPA-GGGS-D5-His6X
3 = EPA-GGGS-D5-D5'-His6X
4 = D5'-EPA-GGGS-His6X
5 = D5'-D5'-EPA-GGGS-His6X
6 = D5'-EPA-GGGS-D5-His6X

C.

EPA variant	1	1	2	2	3	3	4	4	5	5	6	6
CPS8	+	+	+	+	+	+	+	+	+	+	+	+
PglS$_{MDP1}$	-	+	-	+	-	+	-	+	-	+	-	+

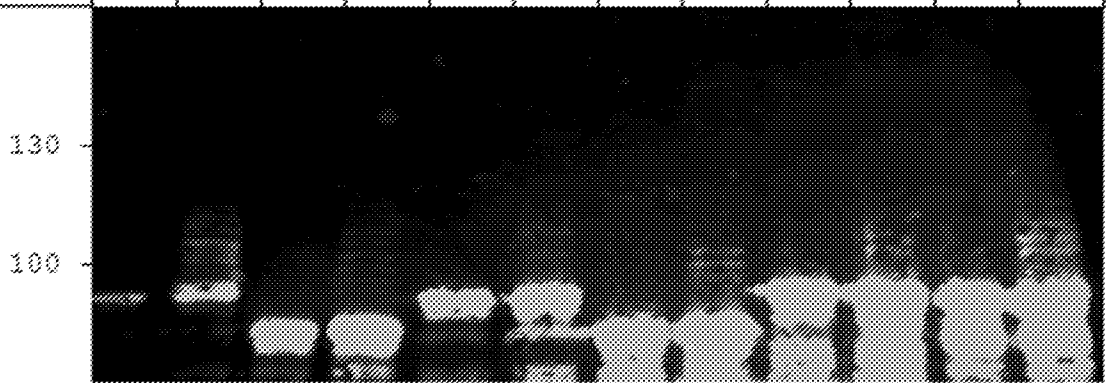
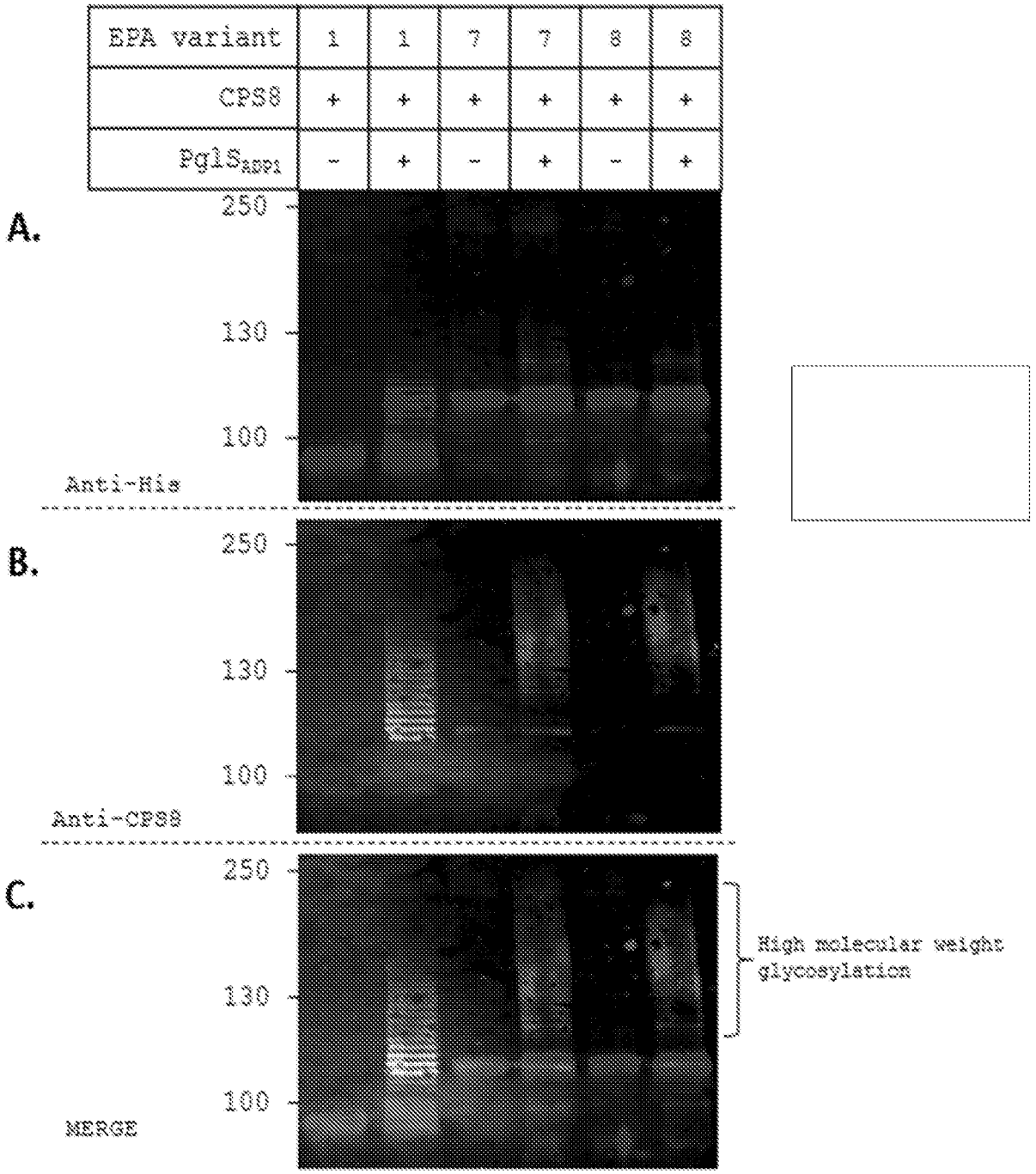


Figure 12A,B,C



EPA Variant Legend

1 = EPA-GGGS-ComPΔ28$_{110264}$-His6X

7 = EPA-GGGS-ComPΔ28$_{110236}$-GGGGS-ComPΔ28$_{110264}$-His6X

8 = EPA-GGGS-ComPΔ28$_{110264}$-PAPAP-ComPΔ28$_{110264}$-His6X

Figure 13A,B,C

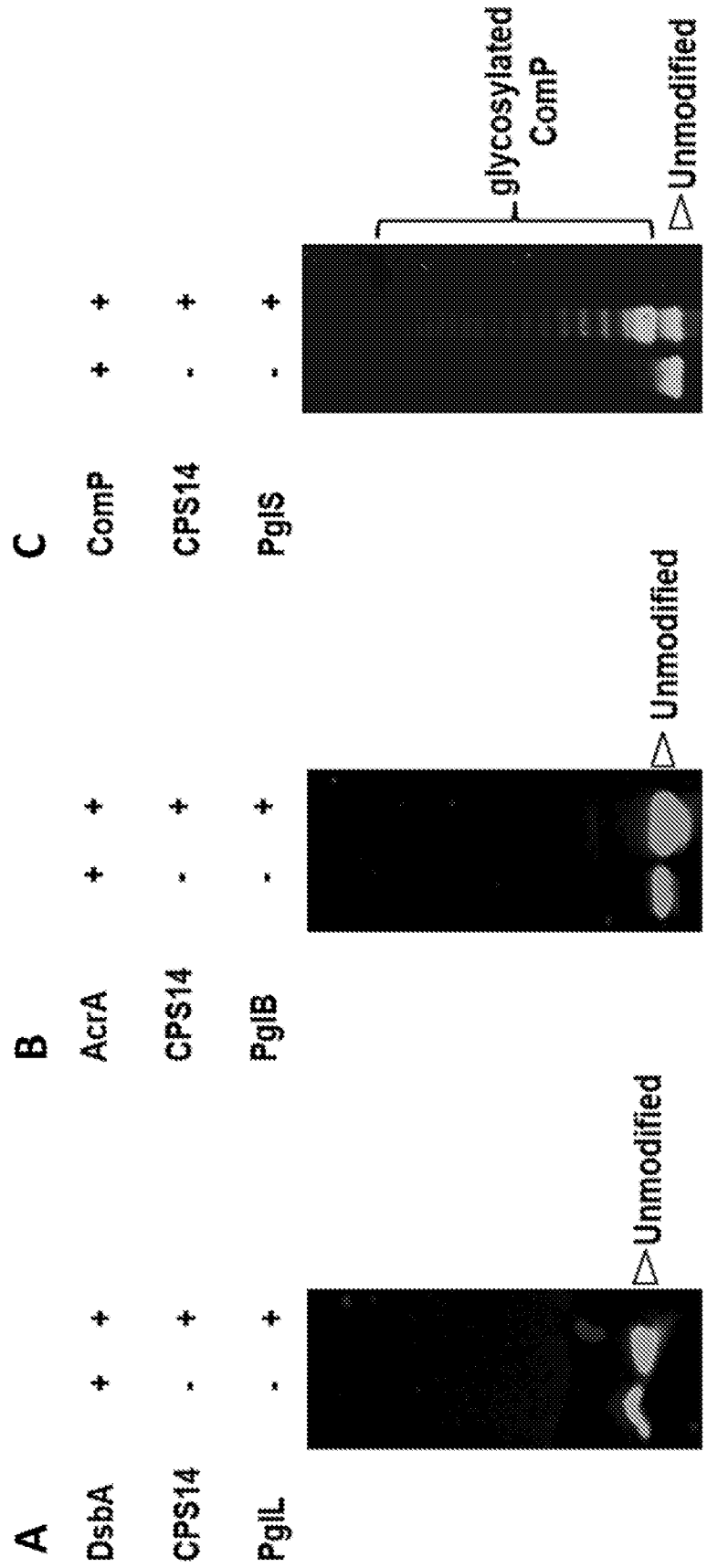


Figure 14A,B,C

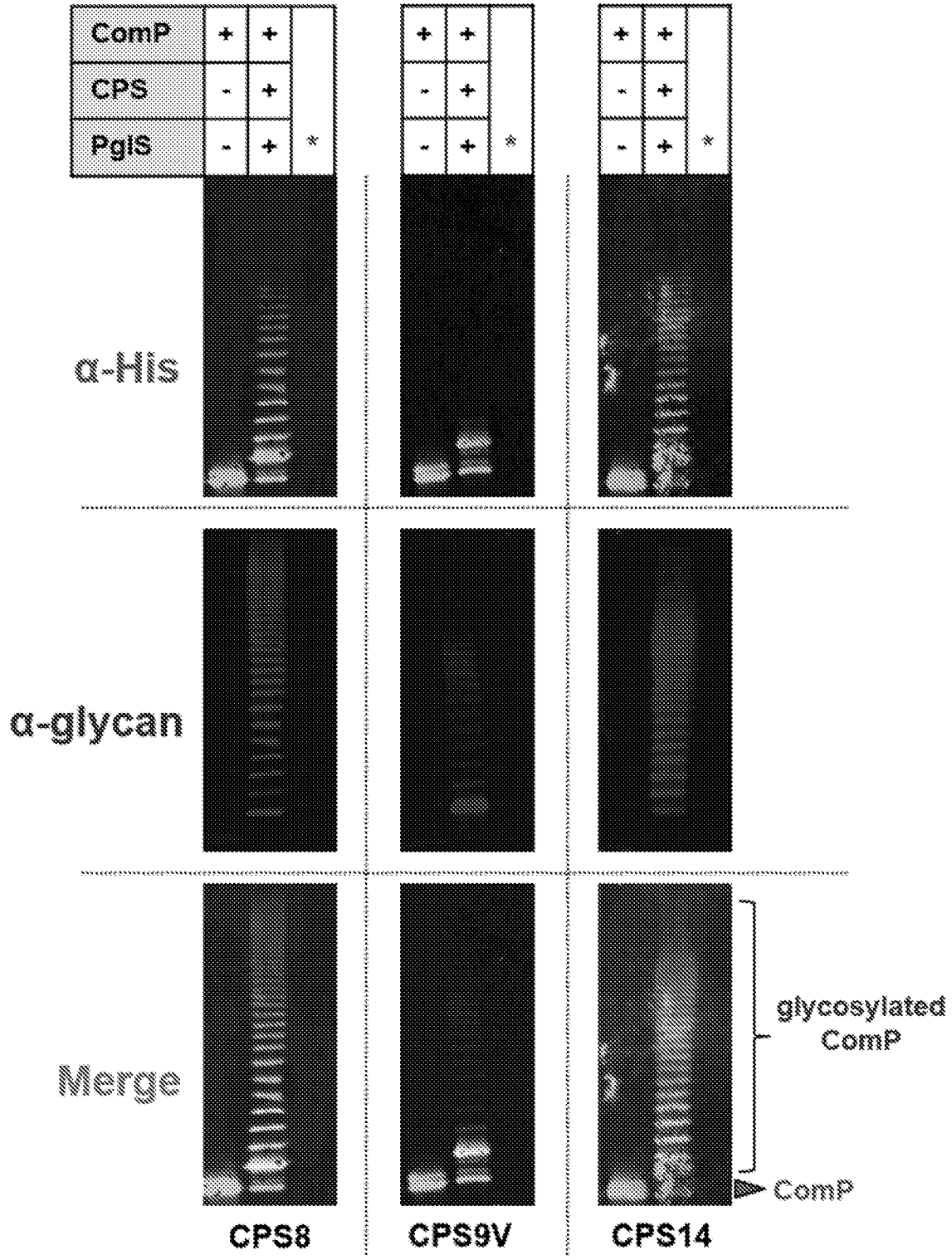

ComP
CPS
PgIS
+ + +
- + +
- + *
α-His
α-glycan
Merge
glycosylated ComP
ComP
CPS8
CPS9V
CPS14

Figure 15

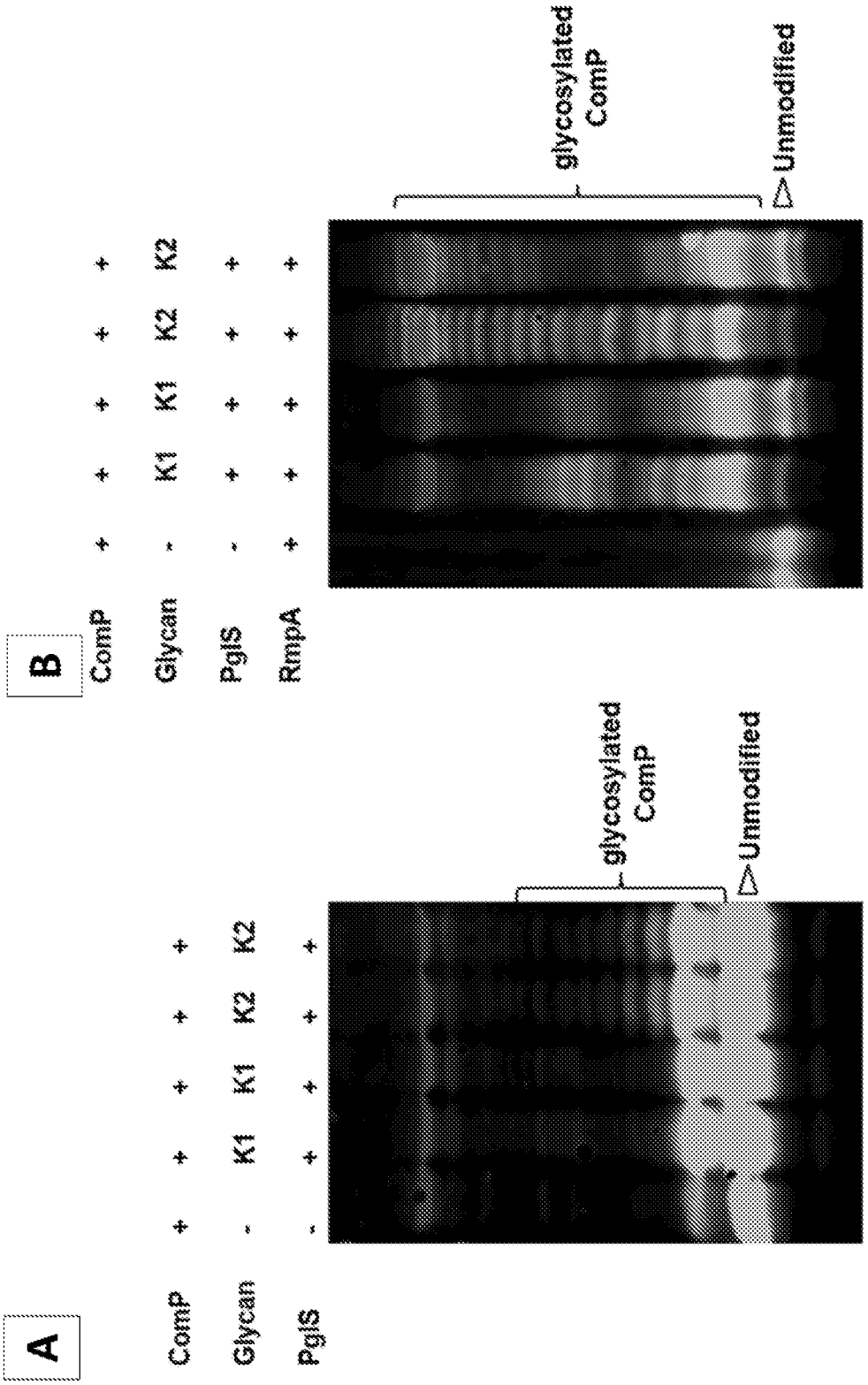

A
ComP + + + +
Glycan - K1 K2 K2
PglS - + + +
glycosylated ComP
Unmodified
B
ComP + + + +
Glycan - K1 K1 K2 K2
PglS - + + + +
RmpA + + + + +
glycosylated ComP
Unmodified

Figure 16A,B

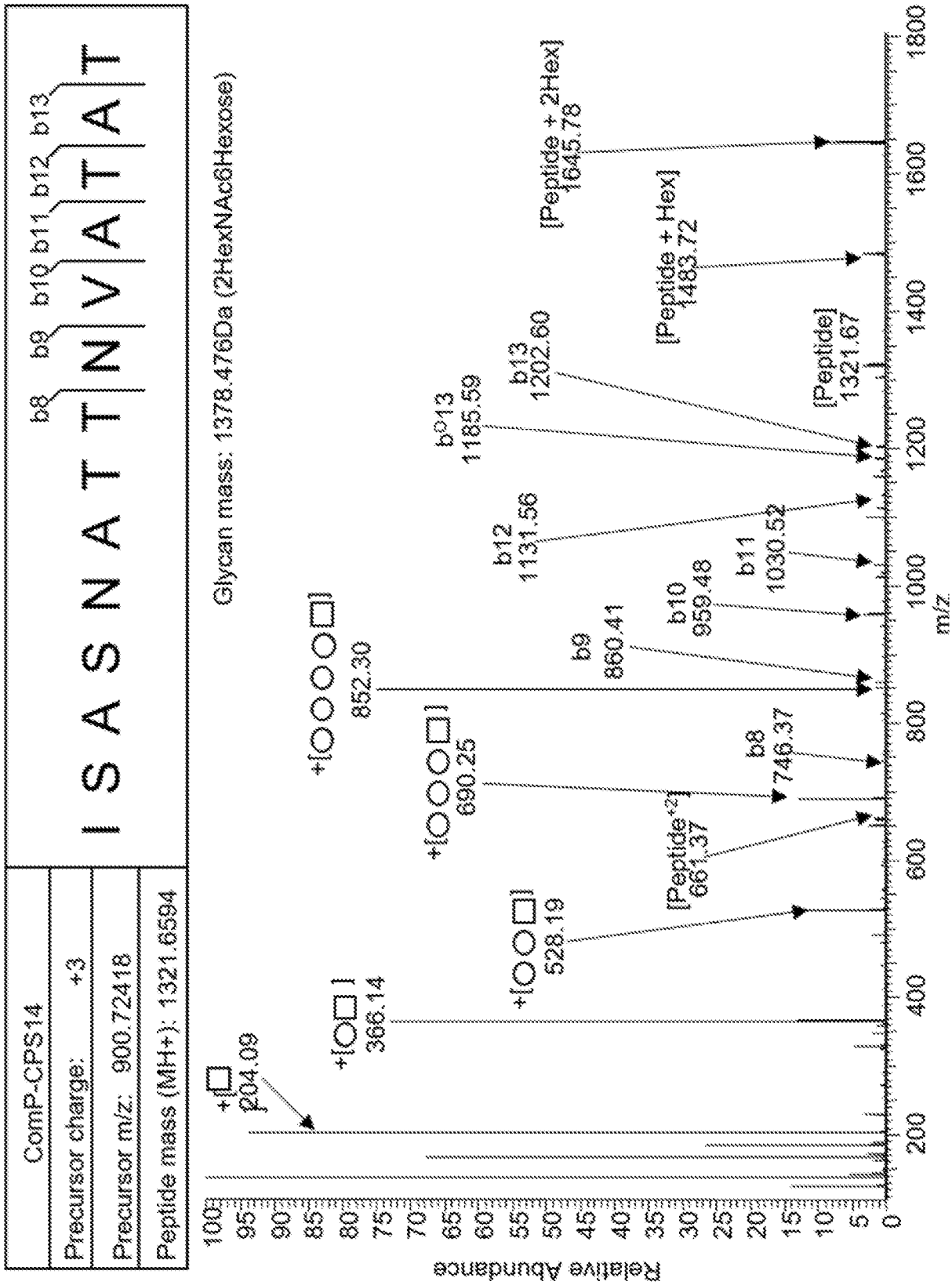

ComP-CPS14
Precursor charge: +3
Precursor m/z: 900.72418
Peptide mass (MH+): 1321.6594
I S A S N A T T N V A T A T
b8 b9 b10 b11 b12 b13
Glycan mass: 1378.476Da (2HexNAc6Hexose)
Relative Abundance
m/z
+[□ 204.09
+[O□] 366.14
+[O□□□□] 852.30
+[O□□□□] 690.25
+[O□□□] 528.19
[Peptide+2H]+2 661.37
b8 746.37
b9 860.41
b10 959.48
b11 1030.52
b12 1131.56
b°13 1185.59
b13 1202.60
[Peptide] 1321.67
[Peptide + Hex] 1483.72
[Peptide + 2Hex] 1645.78

Figure 17A

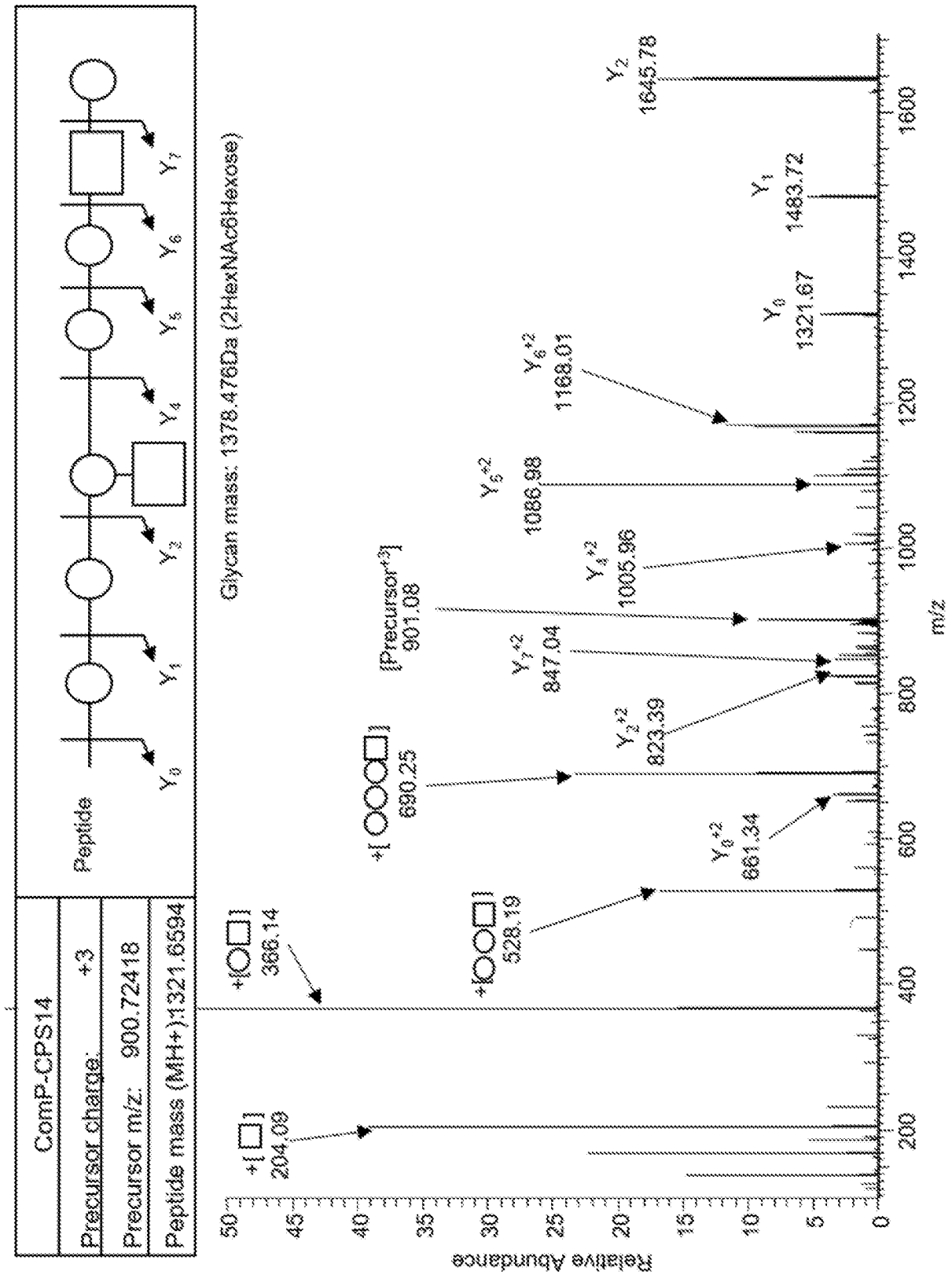

ComP-CPS14
Precursor charge: +3
Precursor m/z: 900.72418
Peptide mass (MH+):1321.6594
Peptide
Glycan mass: 1378.476Da (2HexNAc6Hexose)
Relative Abundance
m/z
+[ ] 204.09
+[ ] 366.14
+[ ] 528.19
+[ ] 690.25
[Precursor+3] 901.08
Y0 1321.67
Y1 1483.72
Y2 1645.78
Y0+2 661.34
Y2+2 823.39
Y4+2 1005.96
Y6+2 1168.01
Y7+2 847.04
Y5+2 1086.98

Figure 17B

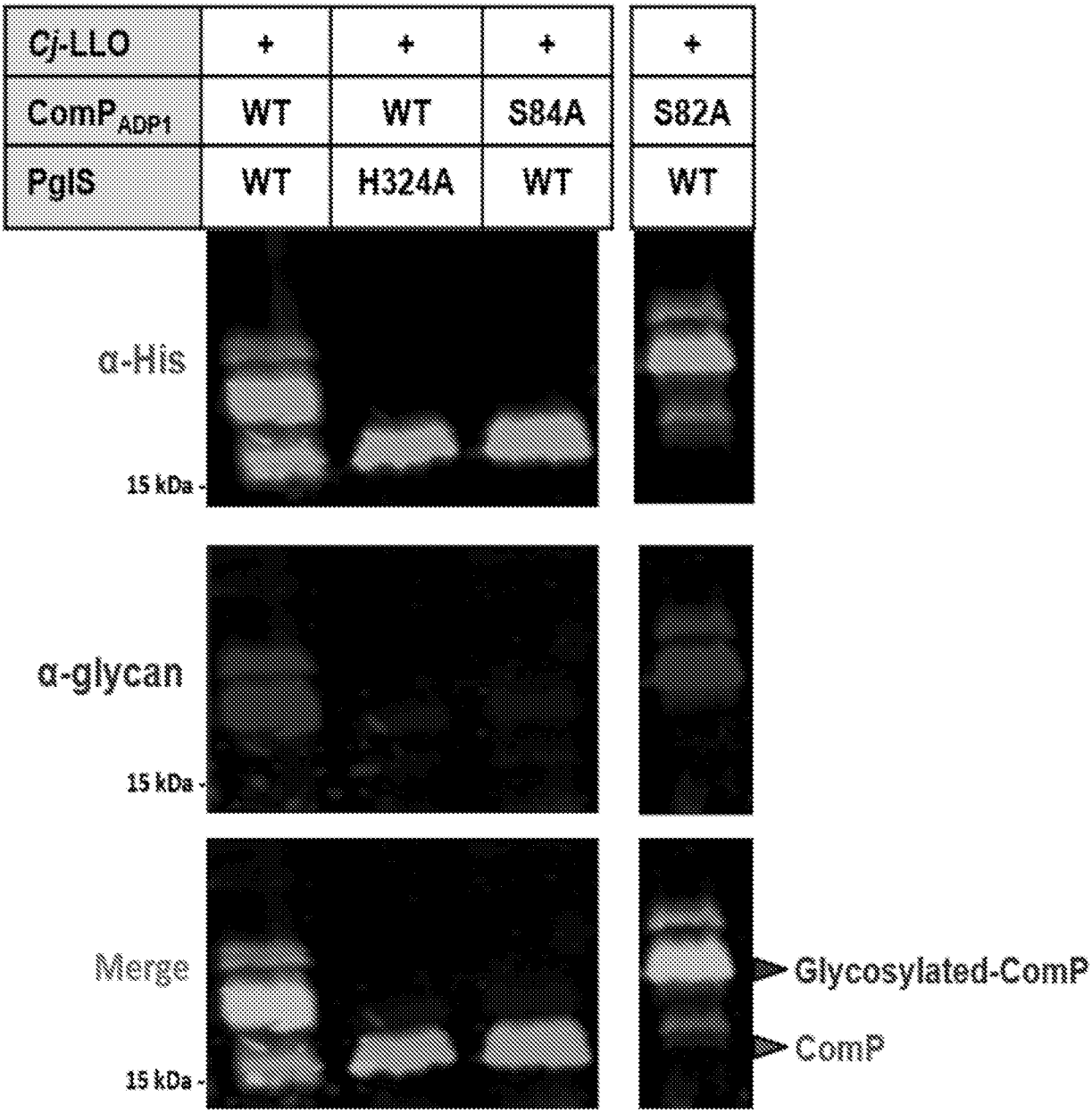

Cj-LLO
+
+
+
+
ComP ADP1
WT
WT
S84A
S82A
PglS
WT
H324A
WT
WT
α-His
15 kDa
α-glycan
15 kDa
Merge
15 kDa
Glycosylated-ComP
ComP

Figure 18

>AAC45886.1 ComP [*Acinetobacter* sp. ADP1]
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR
VSEGLTAASSMKTTVSENILNAGALVAGTPSTAG
SSCVGVQEISA**S**NATTNVATATCGASSAGQIIVTM
DTTKAKGANITLTPTYASGAVTWKCTTTSDKKYV
PSECRG (SEQ ID NO: 1)

>ENV58402.1 hypothetical protein F951_00736
[*Acinetobacter soli* CIP 110264]
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRSR
VTEGLTTASAMKATVSENIMNAGGTSMPSSGNC
TGVTQIASGA**S**AATTNVASAQCSDSDGVITVTMT
DKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYV
PSECRGT (SEQ ID NO: 2)

>APV36638.1 competence protein [*Acinetobacter soli* GFJ-2]
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR
VSEGLTTASAMKATVSENILSAGQIVTGTPSTANS
SCVGVQEINA**S**SSTSNVATATCSGLGVITVTMDS
TKAKGVNLTLTPTYTTSNAVTWKCTTTSDKKYVP
SECRN (SEQ ID NO: 3)

Figure 19

>PKD82822.1 competence protein [*Acinetobacter radioresistens* 50v1]
MNTQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR
VTEAVSTASSMKATVSENIMNAGGTQIPTSGNCV
GVQTIAA**S**NATKNVATATCTDSTGVIVVTTTPAAK
SVPLTLTPTYTGGNVKWACSTTANFKNYVPSEC
RS (SEQ ID NO: 4)

>SNX44537.1 type IV pilus assembly protein PilA [*Acinetobacter puyangensis* ANC 4466]
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR
VTEALTTASAMKATVSENIMSAGGTTIASSACNG
VISA**S**ATTNVASSACSGSGVISVTTTAAAKGIVLTL
TPKYTGGNVAWQCTTTSGDAQKYVPSECRTTS
(SEQ ID NO: 5)

>OAL75955.1 competence protein [*Acinetobacter* sp. SFC]
MNTQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAK
VTEAISTASAMKATVSENLMSAGGTSIVSTNANC
AGVETIGA**S**NKTKNVESAACTAATGVILVTTTAEA
KSVPLTLKPTYTGSNVQWKCGTTAAAFKYVPSE
CRNDSSGTGF (SEQ ID NO: 6)

Figure 19 Cont.

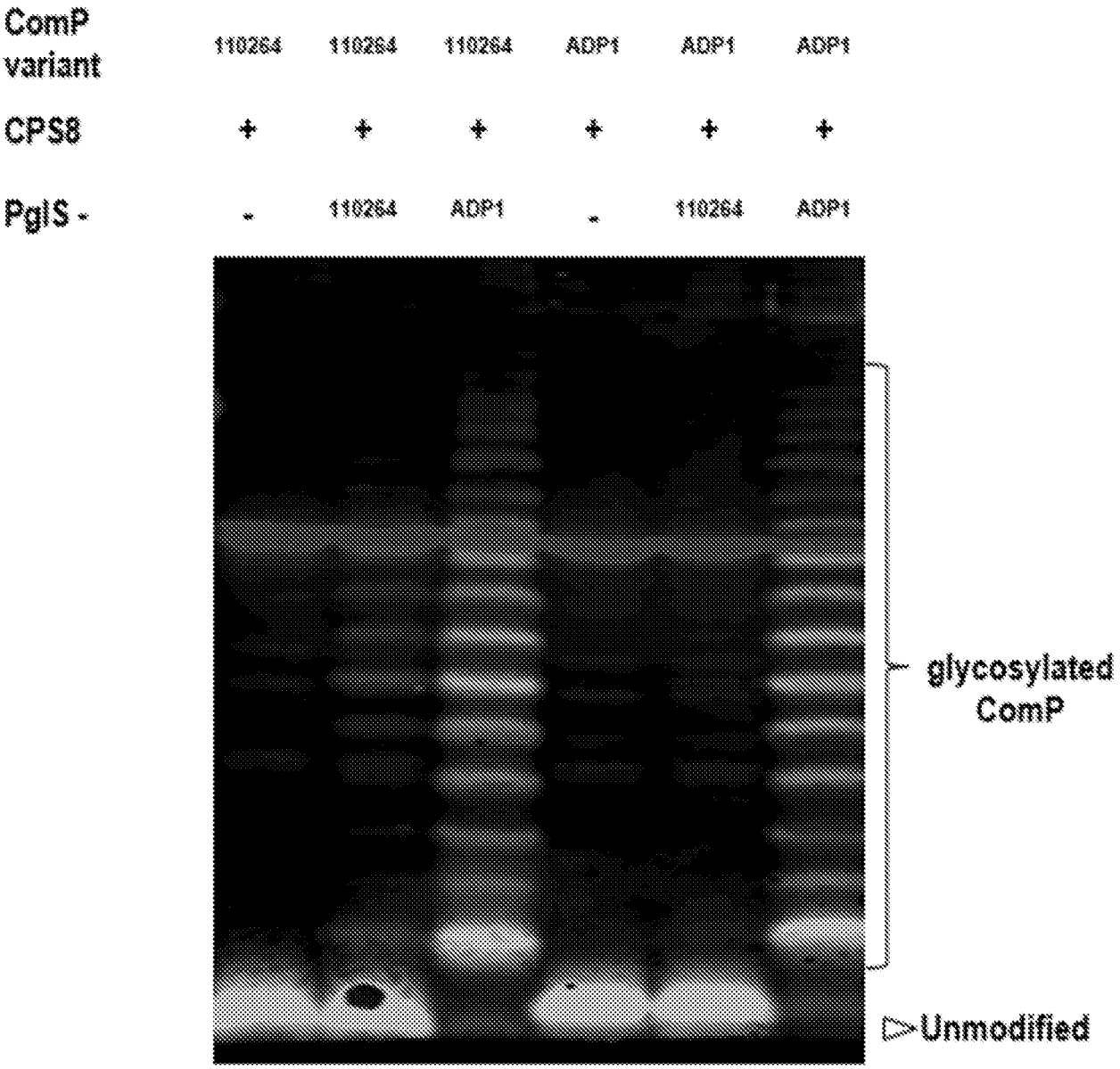

ComP variant
110264
110264
110264
ADP1
ADP1
ADP1
CPS8
+
+
+
+
+
+
PglS -
-
110264
ADP1
-
110264
ADP1
glycosylated ComP
Unmodified

Figure 20

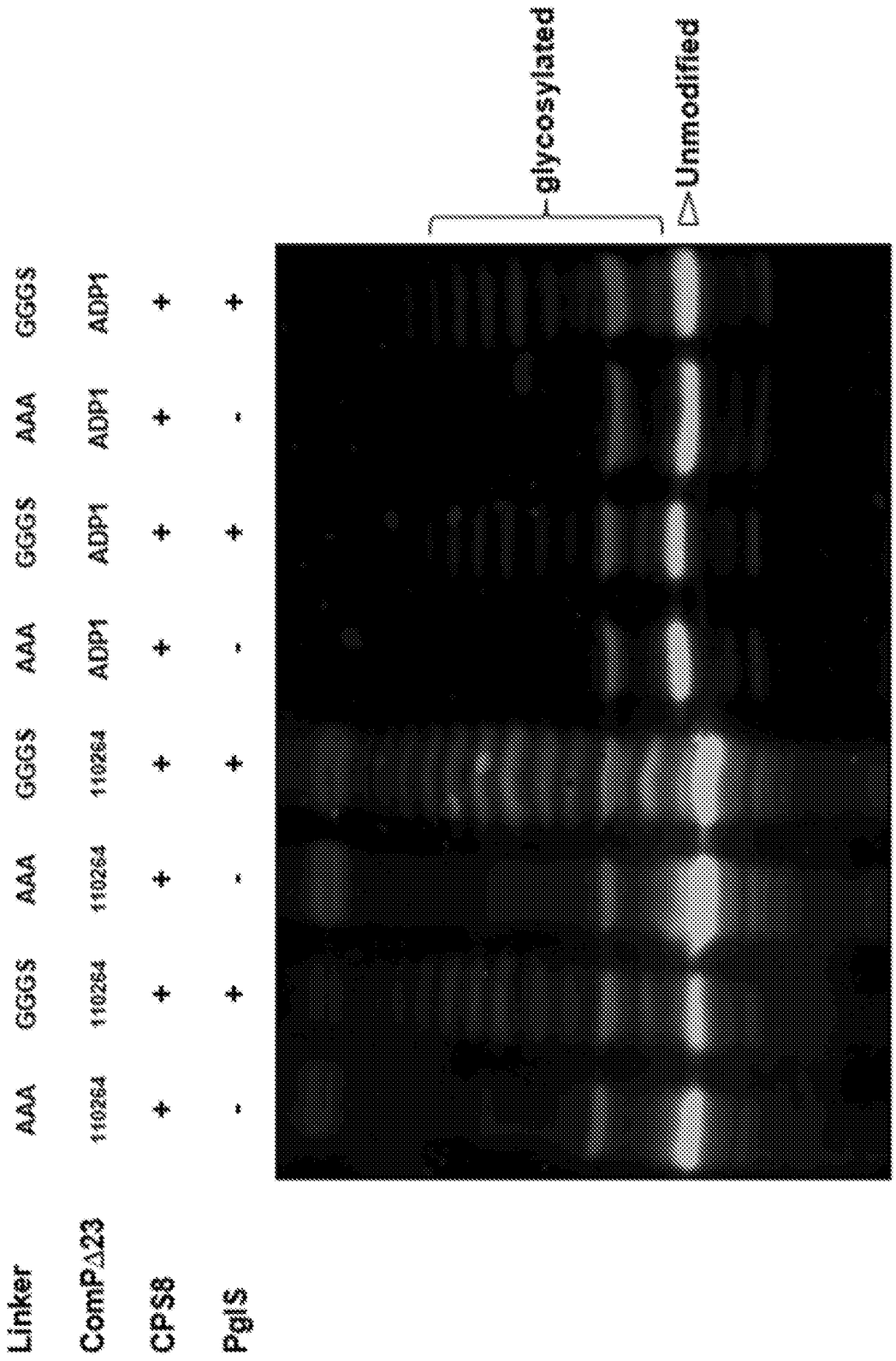

glycosylated
Unmodified
Linker
ComP∆23
CPS8
PglS
AAA
GGGS
AAA
GGGS
AAA
GGGS
AAA
GGGS
110264
110264
110264
110264
ADP1
ADP1
ADP1
ADP1
+
+
+
+
+
+
+
+
-
+
-
+
-
+
-
+

Figure 21

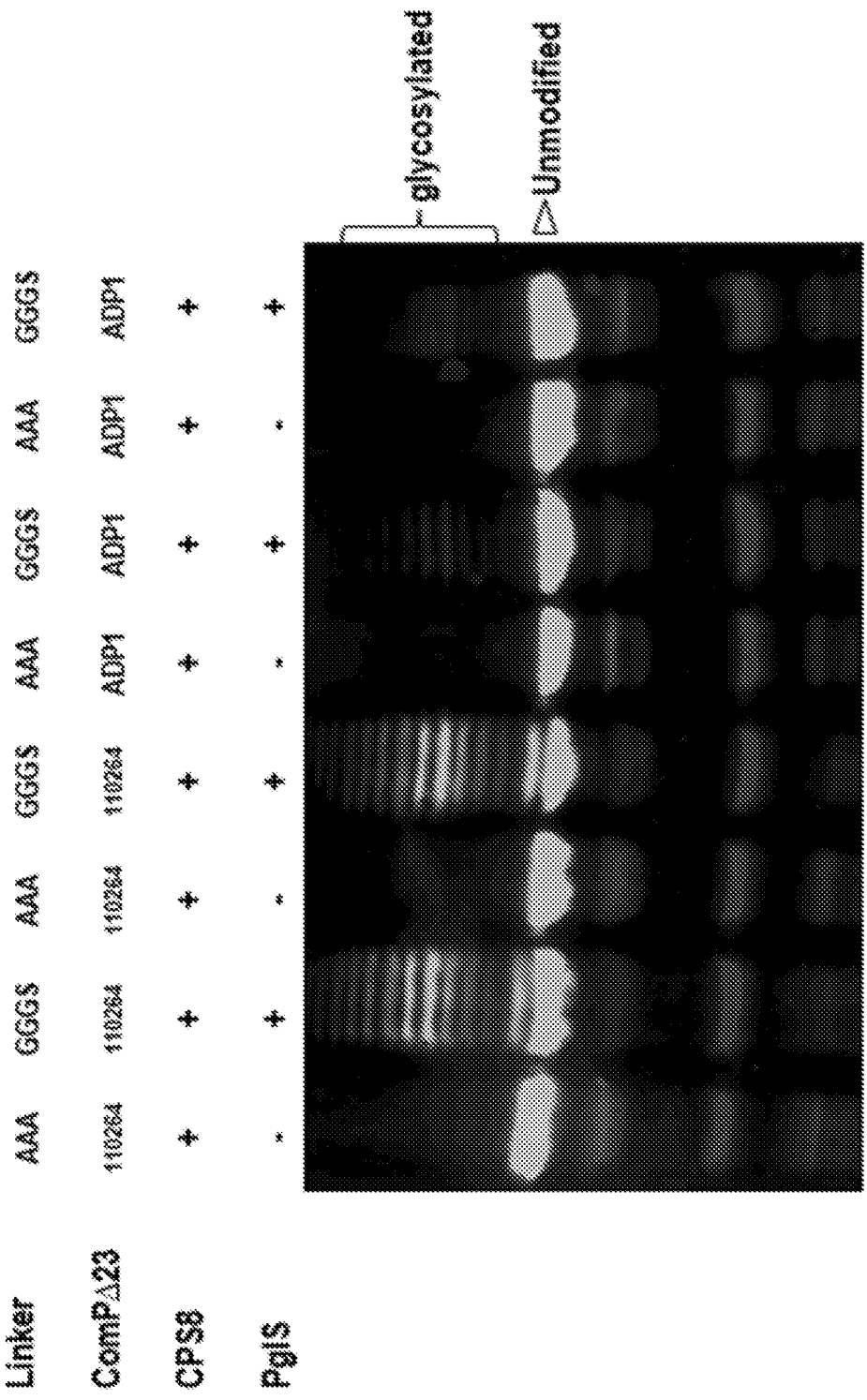

glycosylated
Unmodified
Linker
ComP∆23
CPS8
PglS
AAA
GGGS
AAA
GGGS
AAA
GGGS
AAA
GGGS
110264
110264
110264
110264
ADP1
ADP1
ADP1
ADP1
+
+
+
+
+
+
+
+
-
+
-
+
-
+
-
+

Figure 22

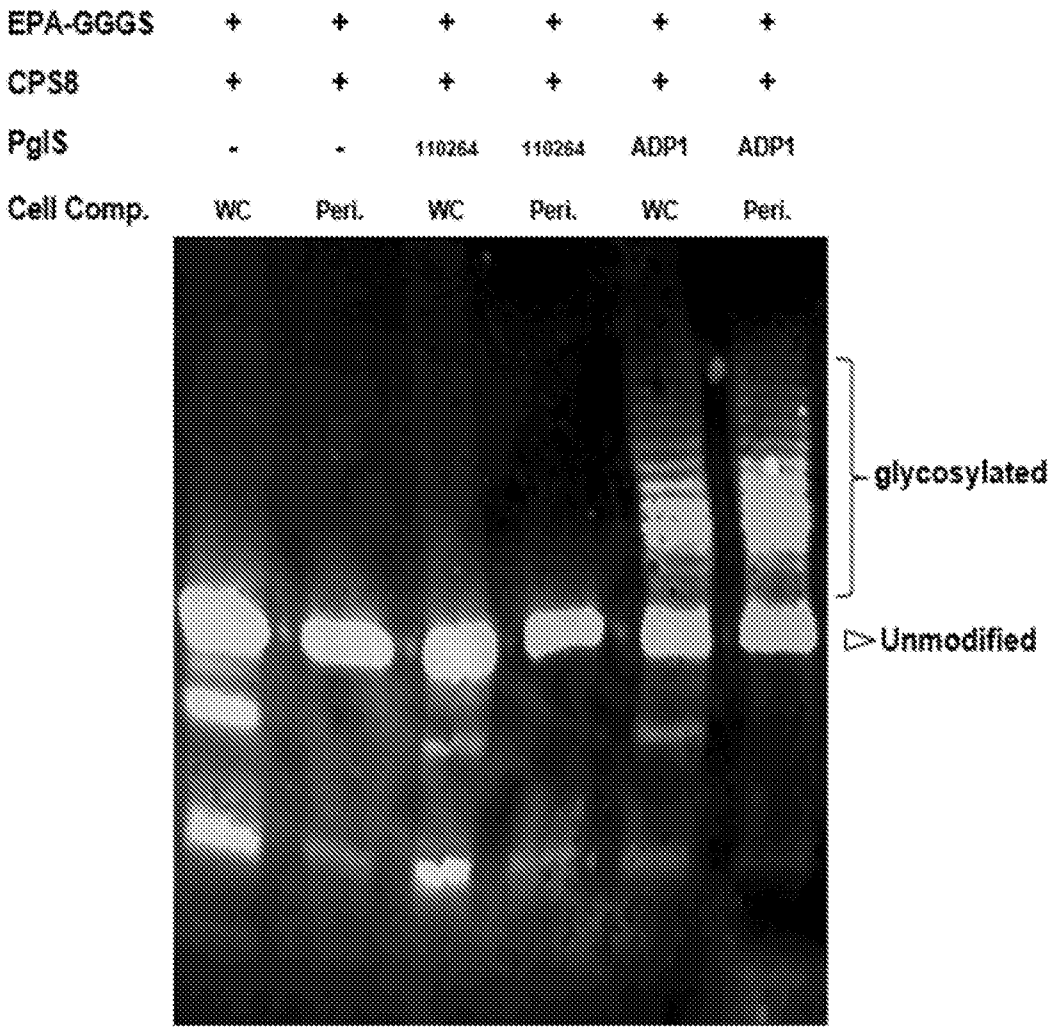

EPA-GGGS + + + + + +
CPS8 + + + + + +
PgIS - - 110284 110284 ADP1 ADP1
Cell Comp. WC Peri. WC Peri. WC Peri.
glycosylated
Unmodified

Figure 23

>DsbA-GGGS-ComPΔ28$_{110264}$-His
MKKIWLALAGLVLAFSASAAQYEDGKQYTTLEKPVA
GAPQVLEFFSFFCPHCYQFEEVLHISDNVKKKLPEGV
KMTKYHVNFMGGDLGKDLTQAWAVAMALGVEDKVT
VPLFEGVQKTQTIRSASDIRDVFINAGIKGEEYDAAW
NSFVVKSLVAQQEKAAADVQLRGVPAMFVNGKYQL
NPQGMDTSNMDVFVQQYADTVKYLSEKK*GGGS*AYT
DYTVRSRVTEGLTTASAMKATVSENIMNAGGTSMPS
SGNCTGVTQIASGASAATTNVASAQCSDSDGVITVTM
TDKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPS
ECRGTHHHHHH (SEQ ID NO: 18)

Legend
DsbA – thiol disulfide oxidoreductase protein (*dsbA* gene)
MBP – maltose binding protein (*malE* gene)
DsbA$_{SP}$ – signal seuqnce of DsbA (first 19 amino acids)
EPA – exotoxin A from *Pseudomonas aeruginosa*
AAA – triple alanine linker
GGGS – glycine-glycine-glycine-serine linker
ComPΔ28$_{110264}$ – ComP from *A. soli* CIP 110264 without the first 28 amino acids
His – hexa-histidine tag

Figure 24

>MBP-AAA-ComPΔ28$_{110264}$-His
MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWIN
GDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFP
QVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAF
QDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLP
NPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWP
LIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLV
DLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAW
SNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAA
SPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS
YEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAV
RTAVINAASGRQTVDEALKDAQTN*AAA*AYTDYTVRS
RVTEGLTTASAMKATVSENIMNAGGTSMPSSGNCTG
VTQIASGASAATTNVASAQCSDSDGVITVTMTDKAKG
VSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECRGTHH
HHHH (SEQ ID NO: 19)

Figure 24 Cont.

>DsbA$_{SP}$-EPA-GGGS-ComPΔ28$_{110264}$-His
MKKIWLALAGLVLAFSASAAEEAFDLWNECAKACVL
DLKDGVRSSRMSVDPAIADTNGQGVLHYSMVLEGGN
DALKLAIDNALSITSDGLTIRLEGGVEPNKPVRYSYTR
QARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSH
MSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTL
AISHAGVSVVMAQAQPRREKRWSEWASGKVLCLLD
PLDGVYNYLAQQRCNLDDTWEGKIYRVLAGNPAKHD
LDIKPTVISHRLHFPEGGSLAALTAHQACHLPLETFTR
HRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVD
QVIRNALASPGSGGDLGEAIREQPEQARLALTLAAAE
SERFVRQGTGNDEAGAASADVVSLTCPVAAGECAG
PADSGDALLERNYPTGAEFLGDGGDISFSTRGTQNW
TVERLLQAHRQLEERGYVFVGYHGTFLEAAQSIVFG
GVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDA
RGRIRNGALLRVYVPRSSLPGFYRTGLTLAAPEAAGE
VERLIGHPLPLRLDAITGPEEEGGRLTILGWPLAERTV
VIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQ
PGKPPREDLK*GGGS*AYTDYTVRSRVTEGLTTASAMK
ATVSENIMNAGGTSMPSSGNCTGVTQIASGASAATTN
VASAQCSDSDGVITVTMTDKAKGVSIKLTPSFSSTGS
VGWKCTTSSDKKYVPSECRGTHHHHHH (SEQ ID
NO: 20)

Figure 24 Cont.

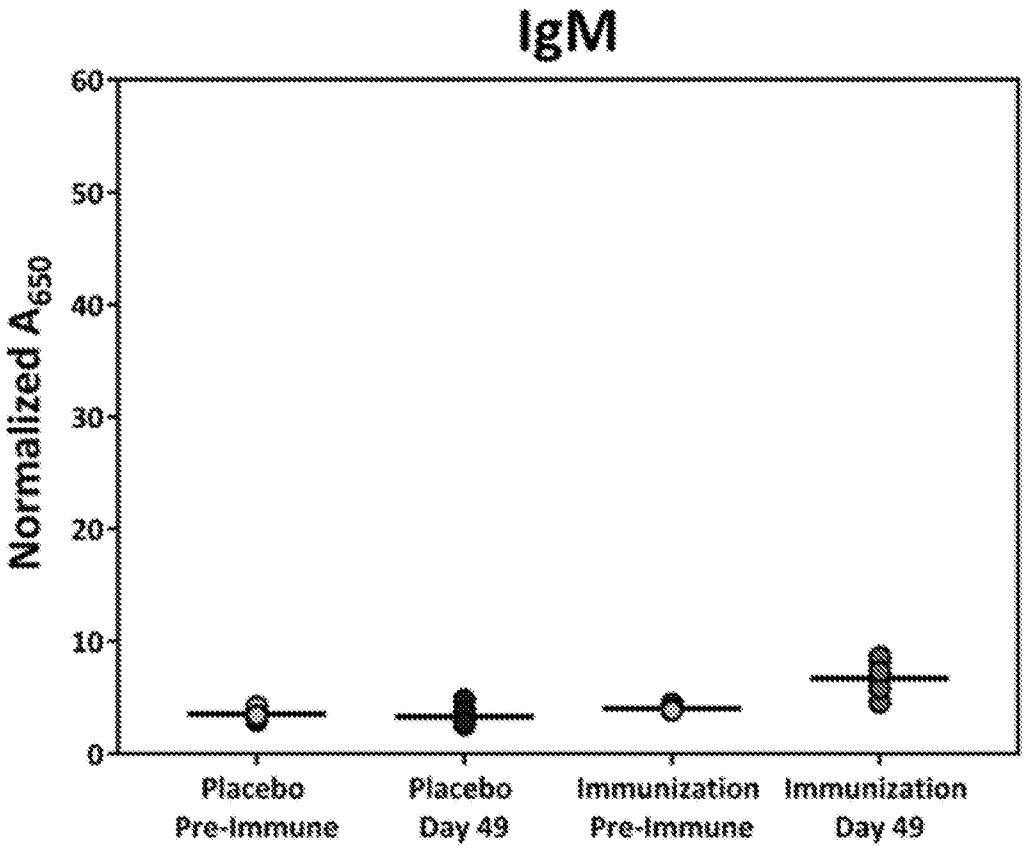

IgM
Normalized A_{650}
60
50
40
30
20
10
0
Placebo Pre-Immune
Placebo Day 49
Immunization Pre-Immune
Immunization Day 49

Figure 25A

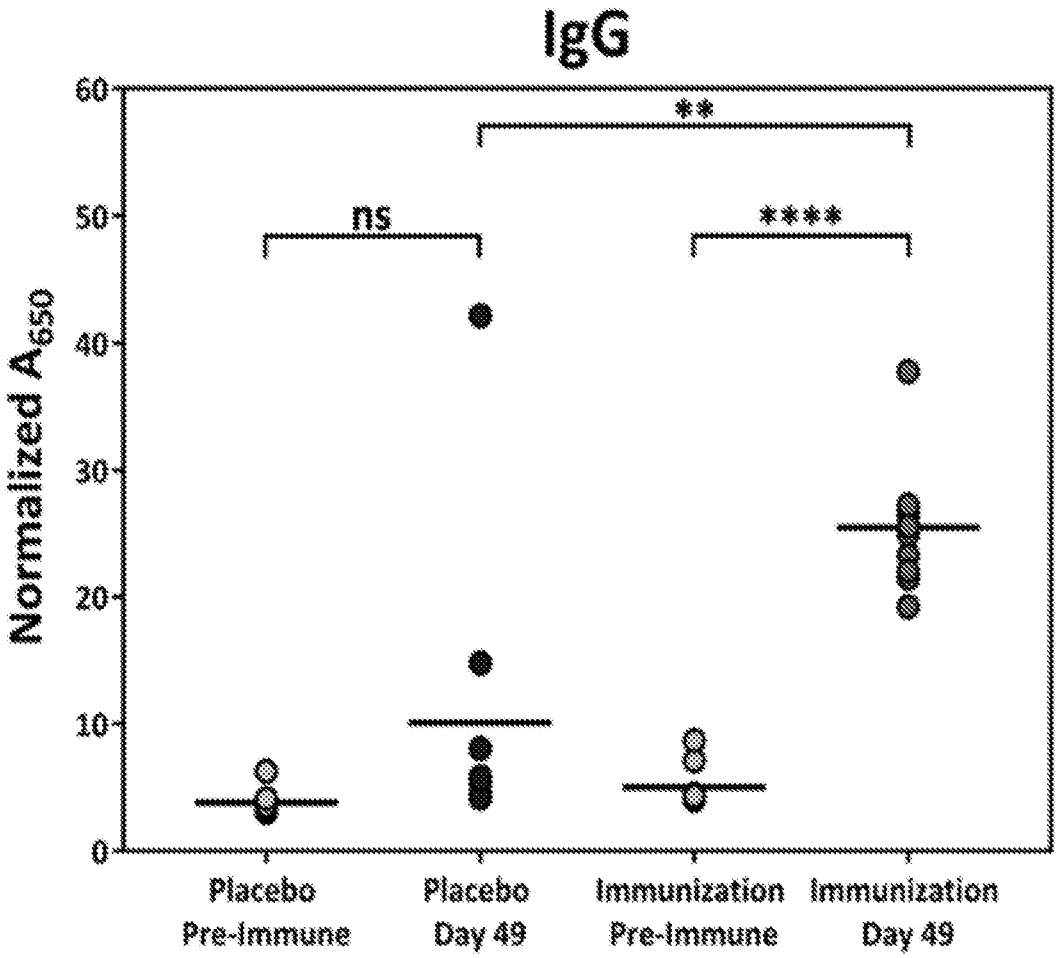

IgG
Normalized A_{650}
60
50
40
30
20
10
0
**
ns
****
Placebo Pre-Immune
Placebo Day 49
Immunization Pre-Immune
Immunization Day 49

Figure 25B

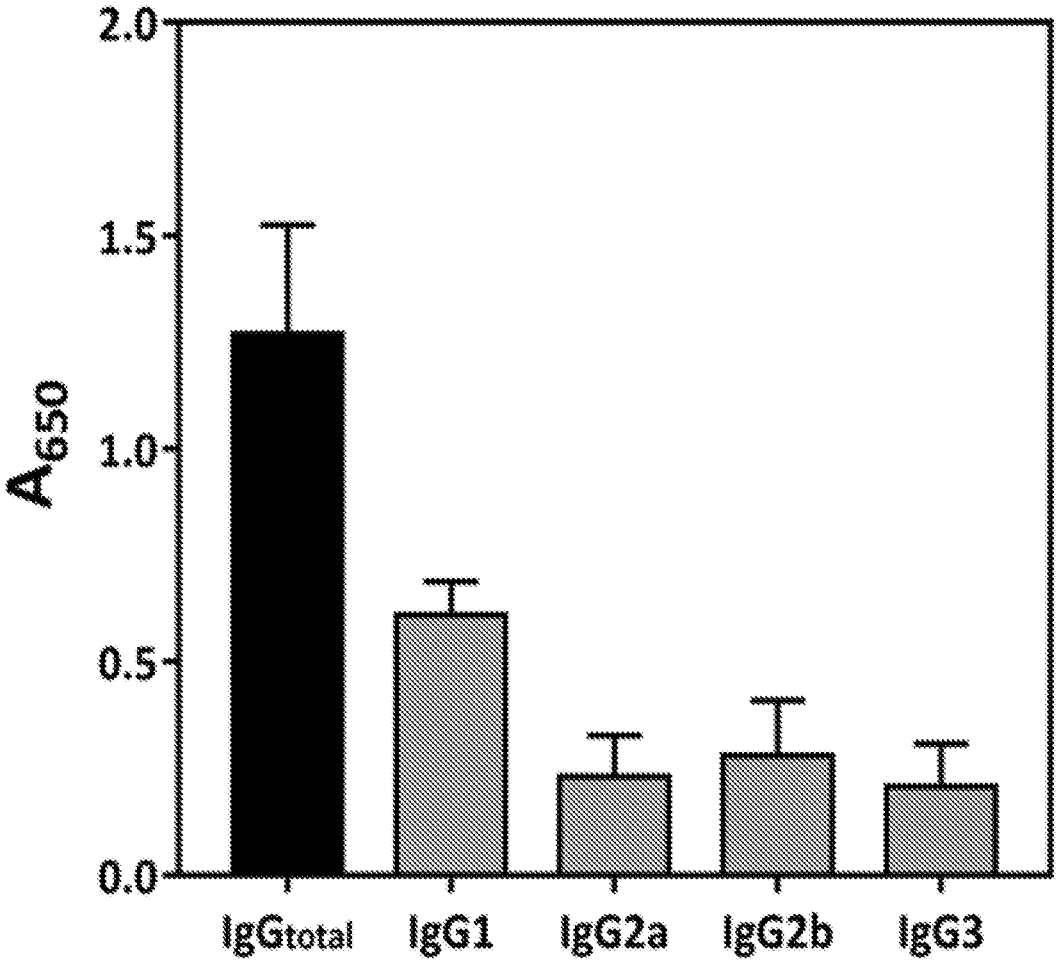


Figure 25C

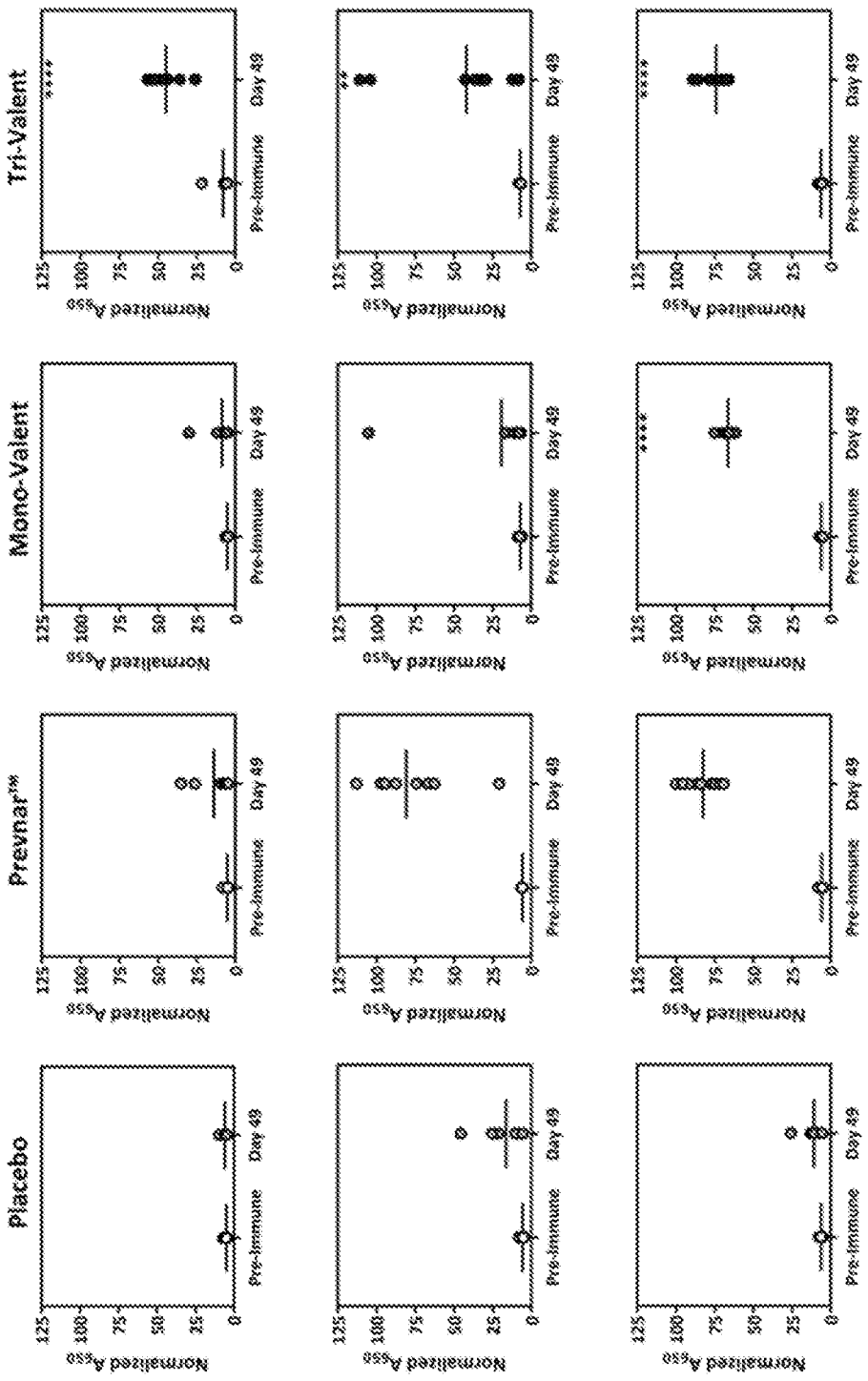

Placebo
Prevnar™
Mono-Valent
Tri-Valent
S. pneumoniae 8
S. pneumoniae 9v
S. pneumoniae 14
Normalized A_{450}
Pre-Immune
Day 49
****
**

Figure 26

>AAC45886.1 ComPΔ28 [*Acinetobacter* sp. ADP1]
AYTDYTVRARVSEGLTAASSMKTTVSENILNAGA
LVAGTPSTAGSSCVGVQEISA**S**NATTNVATATCG
ASSAGQIIVTMDTTKAKGANITLTPTYASGAVTWK
CTTTSDKKYVPSECRG (SEQ ID NO: 7)

>ENV58402.1 hypothetical protein F951_00736 Δ28 [*Acinetobacter soli* CIP 110264]
AYTDYTVRSRVTEGLTTASAMKATVSENIMNAGG
TSMPSSGNCTGVTQIASGA**S**AATTNVASAQCSD
SDGVITVTMTDKAKGVSIKLTPSFSSTGSVGWKC
TTSSDKKYVPSECRGT (SEQ ID NO: 8)

>APV36638.1 competence protein Δ28 [*Acinetobacter soli* GFJ-2]
AYTDYTVRARVSEGLTTASAMKATVSENILSAGQI
VTGTPSTANSSCVGVQEINA**S**SSTSNVATATCSG
LGVITVTMDSTKAKGVNLTLTPTYTTSNAVTWKC
TTTSDKKYVPSECRN (SEQ ID NO: 9)

Figure 27

>PKD82822.1 competence protein Δ28
[*Acinetobacter radioresistens* 50v1]
AYTDYTVRARVTEAVSTASSMKATVSENIMNAG
GTQIPTSGNCVGVQTIAA**S**NATKNVATATCTDST
GVIVVTTTPAAKSVPLTLTPTYTGGNVKWACSTT
ANFKNYVPSECRS (SEQ ID NO: 10)

>SNX44537.1 type IV pilus assembly protein PilA
Δ28 [*Acinetobacter puyangensis* ANC 4466]
AYTDYTVRARVTEALTTASAMKATVSENIMSAGG
TTIASSACNGVISA**S**ATTNVASSACSGSGVISVTT
TAAAKGIVLTLTPKYTGGNVAWQCTTTSGDAQKY
VPSECRTTS (SEQ ID NO: 11)

>OAL75955.1 competence protein Δ28
[*Acinetobacter* sp. SFC]
AYTDYTVRAKVTEAISTASAMKATVSENLMSAGG
TSIVSTNANCAGVETIGA**S**NKTKNVESAACTAAT
GVILVTTTAEAKSVPLTLKPTYTGSNVQWKCGTT
AAAFKYVPSECRNDSSGTGF (SEQ ID NO: 12)

Figure 27 Cont.

```
ADP1     CVGVQEIS--ASNATTNVATATC (SEQ ID NO: 13)
1102264  CTGVTQIASGASAATTNVASAQC (SEQ IS NO: 14)
GFJ_2    CVGVQEIN--ASSTSNVATATC (SEQ ID NO: 15)
SFC      CAGVETIG--ASNKTKNVESAAC (SEQ ID NO: 16)
P50v1    CVGVQTIA--ASNATKNVATATC (SEQ ID NO: 17)
```

Figure 28

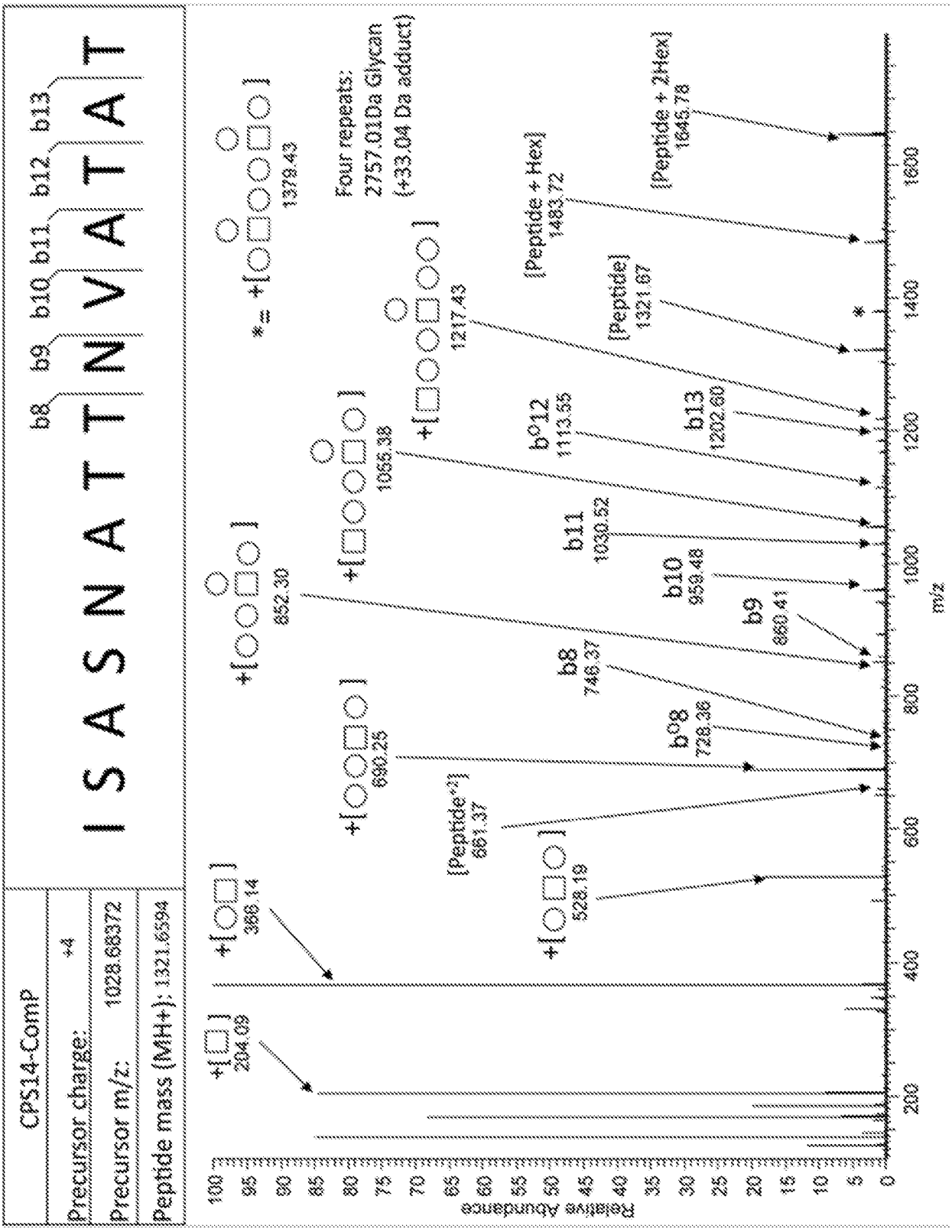

CPS14-ComP
Precursor charge: +4
Precursor m/z: 1028.68372
Peptide mass (MH+): 1321.6594
I S A S N A T T N V A T A T
b8 b9 b10 b11 b12 b13
Four repeats:
2757.01Da Glycan
(+33.04 Da adduct)
*= +[○□○○□○□]
1378.43
+[□]
204.09
+[○□]
366.14
+[○□○]
528.19
+[○□○□○]
690.25
+[○○□○□○]
852.30
+[□○○□○□○]
1055.38
+[□○○□○□○○]
1217.43
[Peptide+23]
661.37
b8
746.37
b°8
728.36
b9
860.41
b10
959.48
b11
1030.52
b°12
1113.55
b13
1202.60
[Peptide]
1321.67
[Peptide + Hex]
1483.72
[Peptide + 2Hex]
1645.78
*
Relative Abundance
m/z
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
200 400 600 800 1000 1200 1400 1600

Figure 29

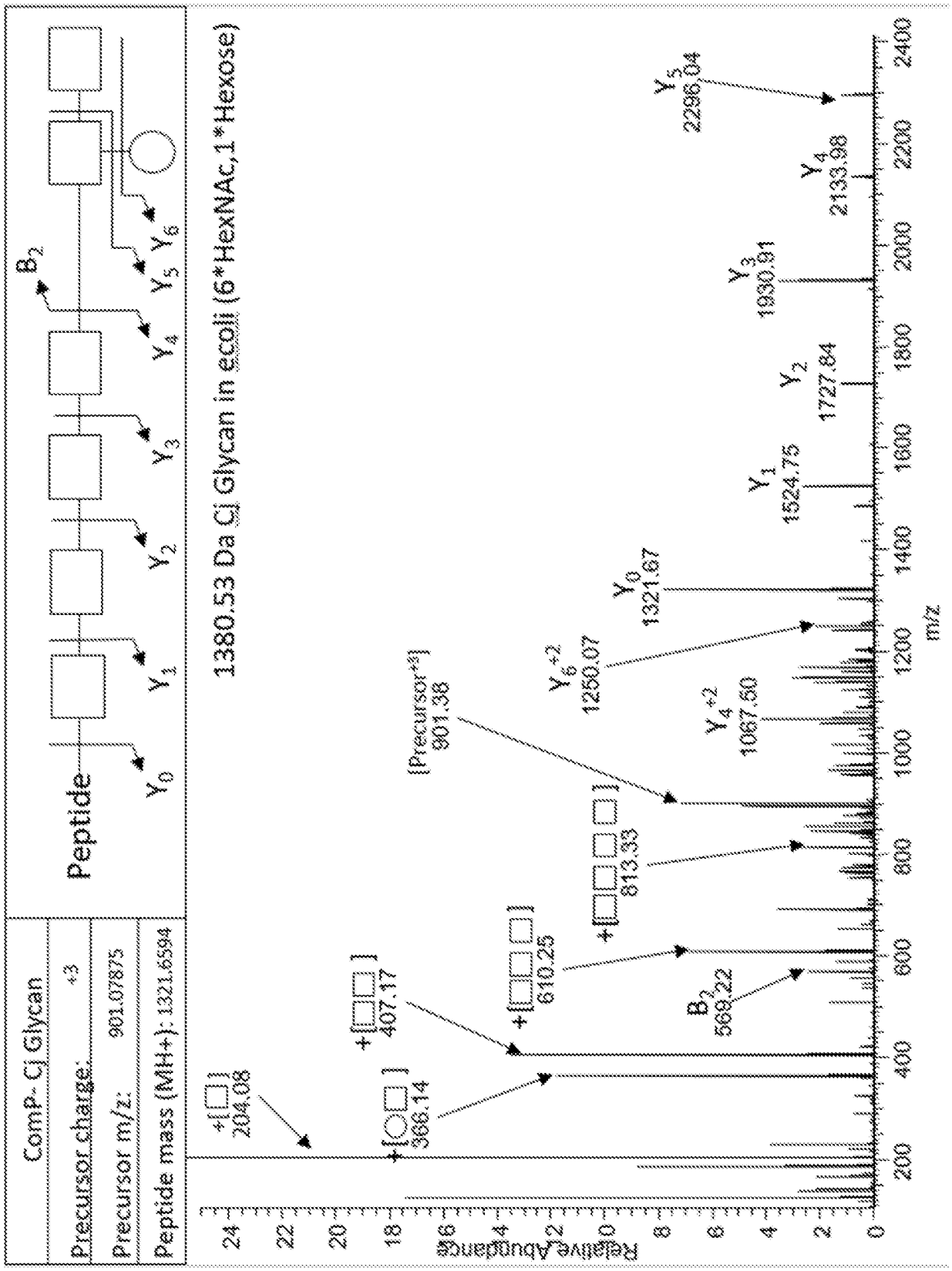

ComP- Cj Glycan
Precursor charge: +3
Precursor m/z: 901.07875
Peptide mass (MH+): 1321.6594
Peptide
Y0
Y1
Y2
Y3
Y4
Y5
Y6
B2
1380.53 Da Cj Glycan in ecoli (6*HexNAc,1*Hexose)
Relative Abundance
m/z
[Precursor+3] 901.38
Y0 1321.67
Y6+2 1250.07
Y1 1524.75
Y2 1727.84
Y3 1930.91
Y4 2133.98
Y5 2296.04
Y4+2 1067.50
B2 569.22
610.25
813.33
407.17
366.14
204.08

Figure 30

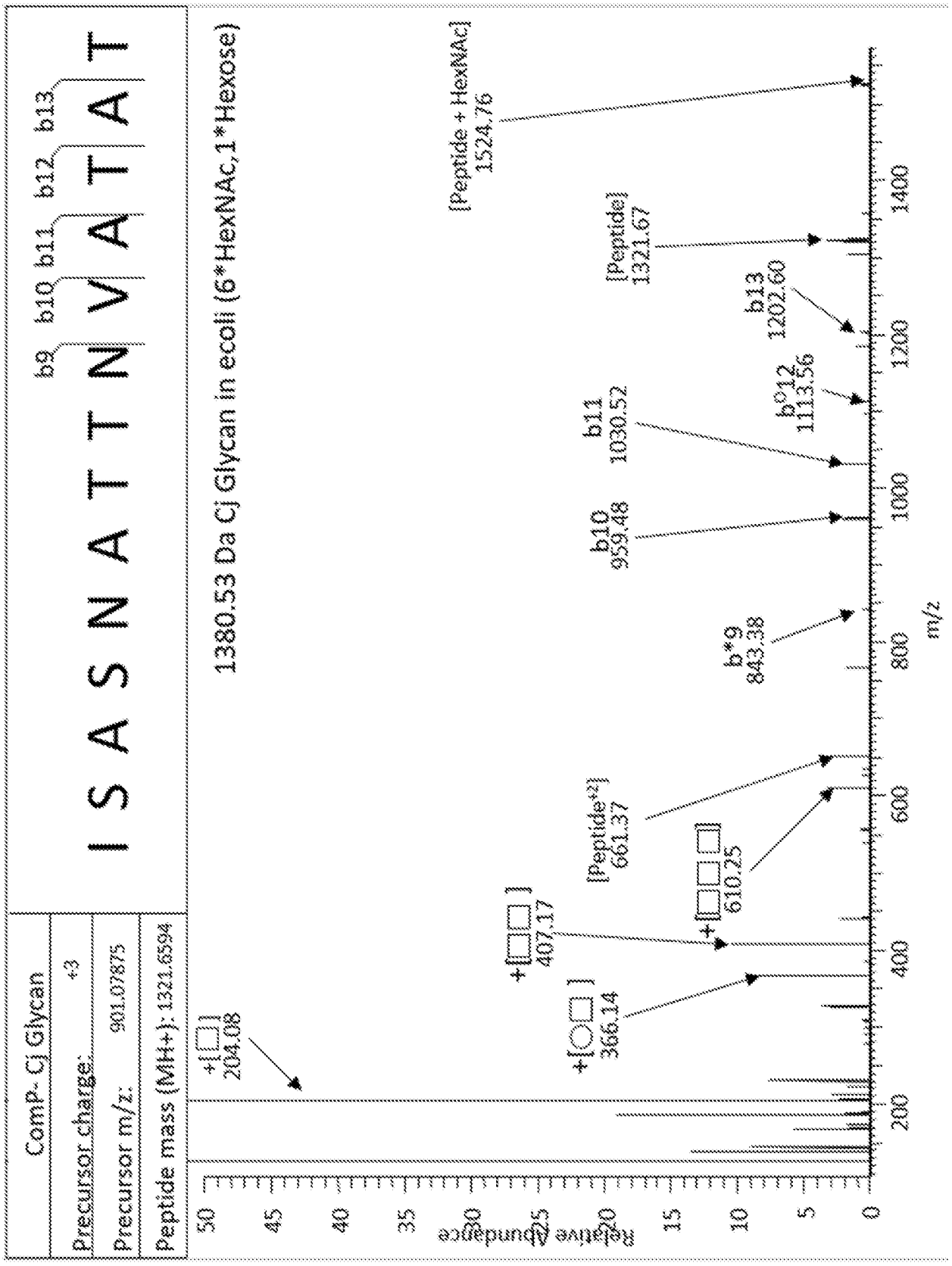

ComP- Cj Glycan
Precursor charge: +3
Precursor m/z: 901.07875
Peptide mass (MH+): 1321.6594
I S A S N A T T N V A T A T
b9 b10 b11 b12 b13
1380.53 Da Cj Glycan in ecoli (6*HexNAc,1*Hexose)
Relative Abundance
m/z
+[□] 204.08
+[□□] 407.17
+[○□] 366.14
+[□□□] 610.25
[Peptide*2] 661.37
b*9 843.38
b10 959.48
b11 1030.52
b°12 1113.56
b13 1202.60
[Peptide] 1321.67
[Peptide + HexNAc] 1524.76

Figure 31

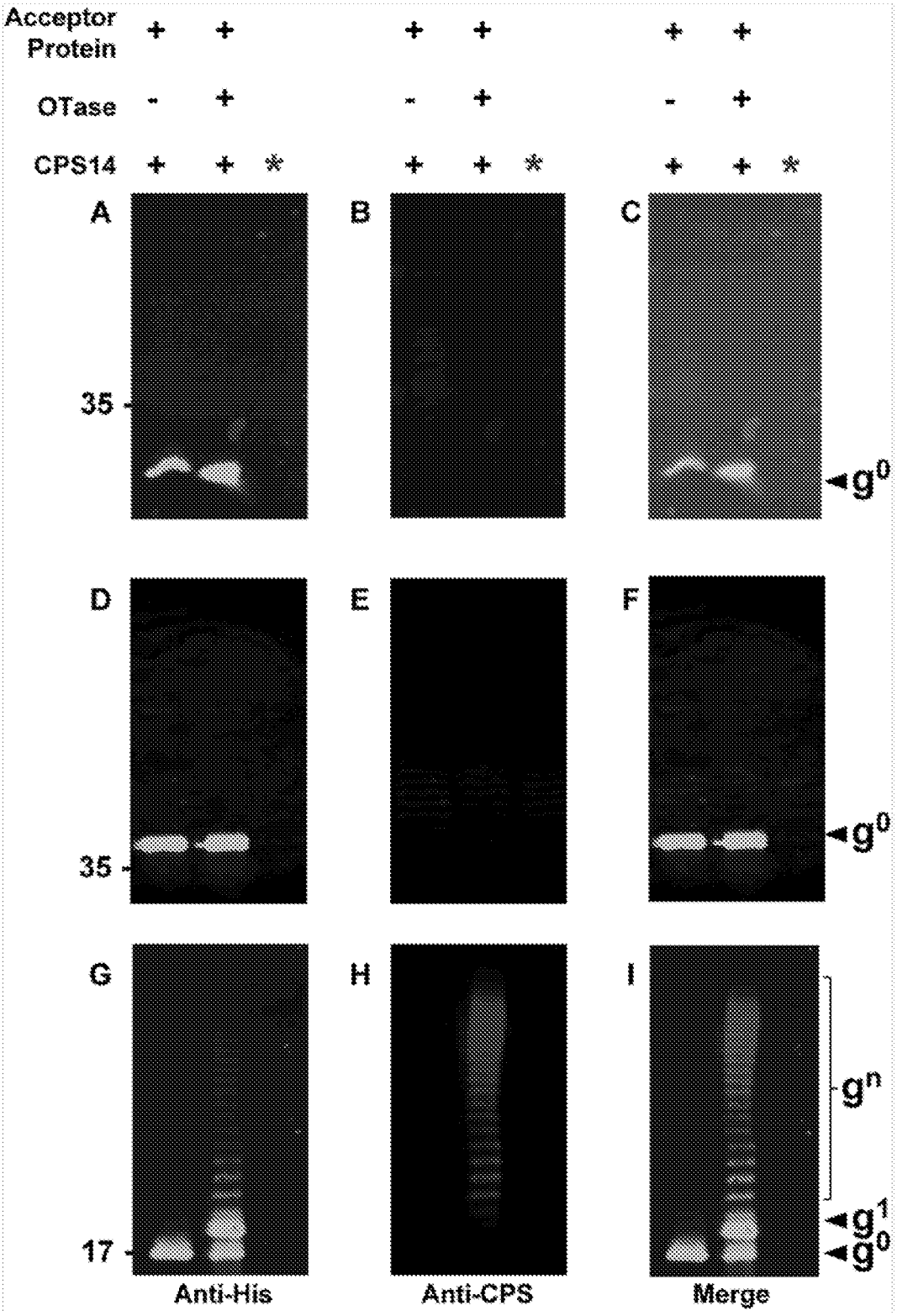

Acceptor Protein
OTase
CPS14
A
B
C
35
g^0
D
E
F
35
g^0
G
H
I
g^n
g^1
g^0
17
Anti-His
Anti-CPS
Merge

Figure 32A-I

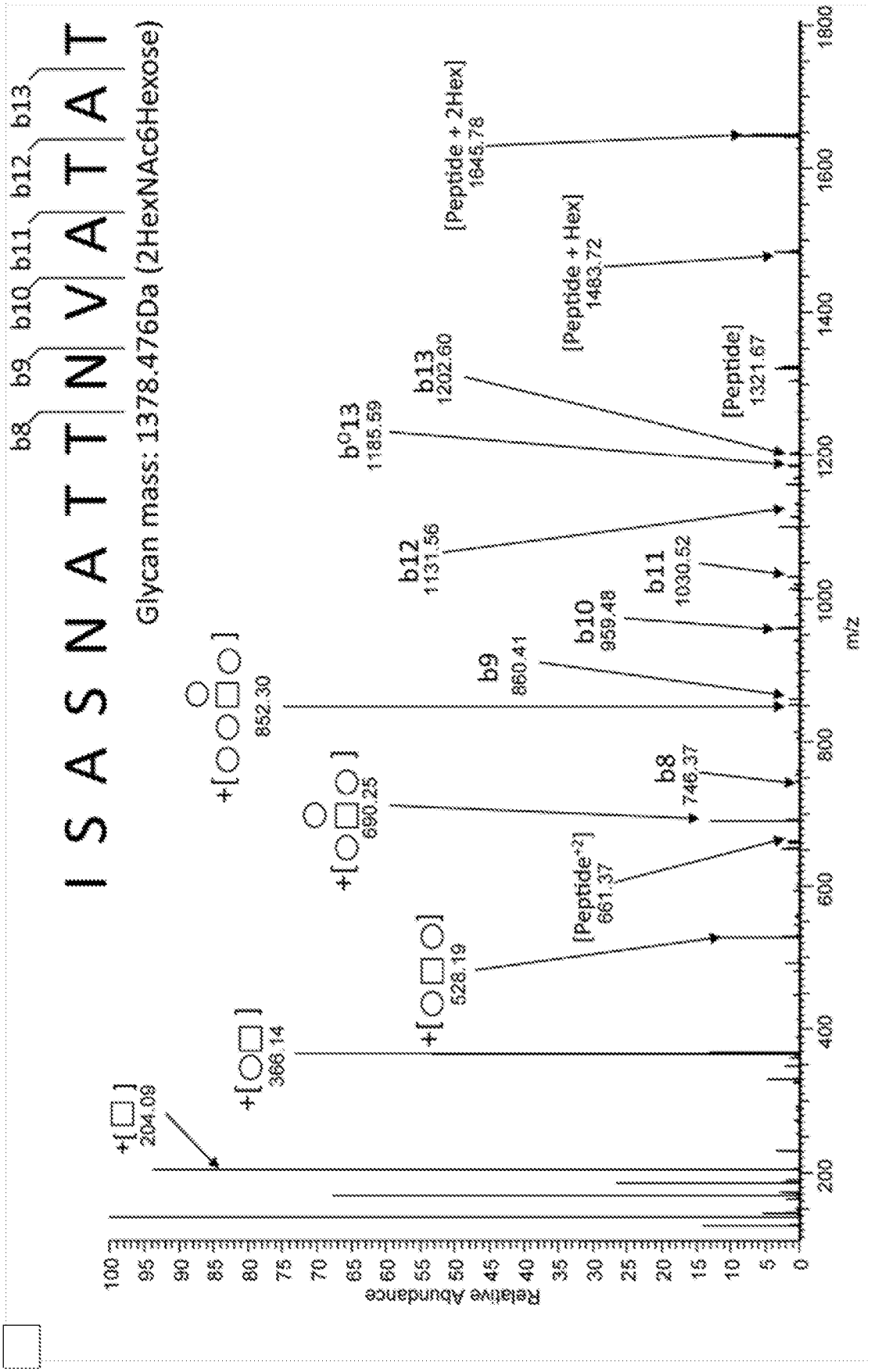

I S A S N A T T N V A T A T
b8 b9 b10 b11 b12 b13
Glycan mass: 1378.476Da (2HexNAc6Hexose)
+[□] 204.08
+[○□] 366.14
+[○□○] 528.19
+[○□○] 690.25
+[○○□○] 852.30
[Peptide+2] 661.37
b8 746.37
b9 860.41
b10 959.48
b11 1030.52
b12 1131.56
b°13 1185.59
b13 1202.60
[Peptide] 1321.67
[Peptide + Hex] 1483.72
[Peptide + 2Hex] 1645.78
Relative Abundance
m/z

Figure 33A

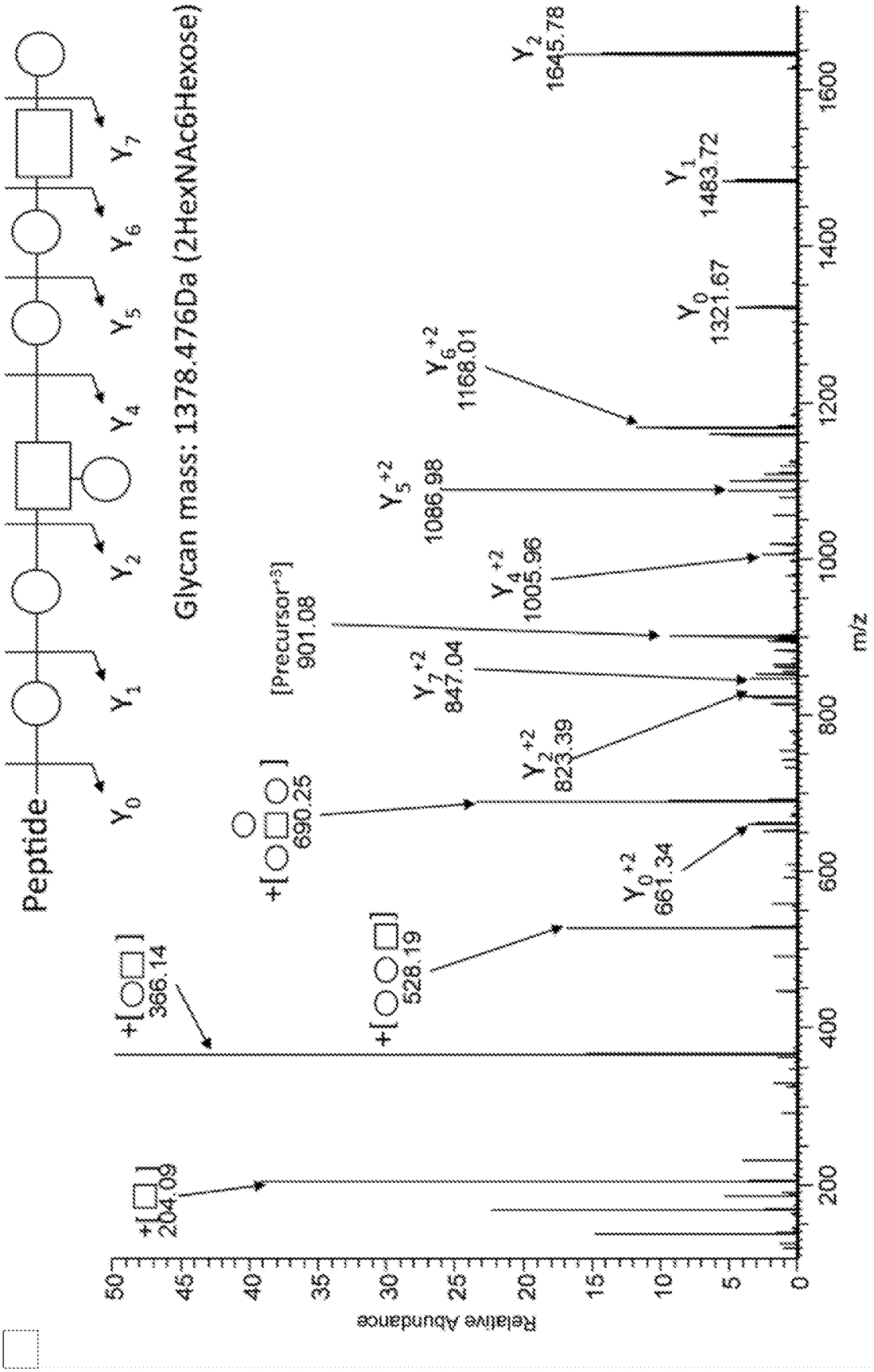

Peptide
Glycan mass: 1378.476Da (2HexNAc6Hexose)
Y0
Y1
Y2
Y4
Y5
Y6
Y7
[Precursor+3]
901.08
+[ ]
204.09
+[ ]
366.14
+[ ]
528.19
+[ ]
690.25
Y0+2
661.34
Y2+2
823.39
Y7+2
847.04
Y4+2
1005.96
Y5+2
1086.98
Y6+2
1168.01
Y0
1321.67
Y1
1483.72
Y2
1645.78
Relative Abundance
m/z

Figure 33B

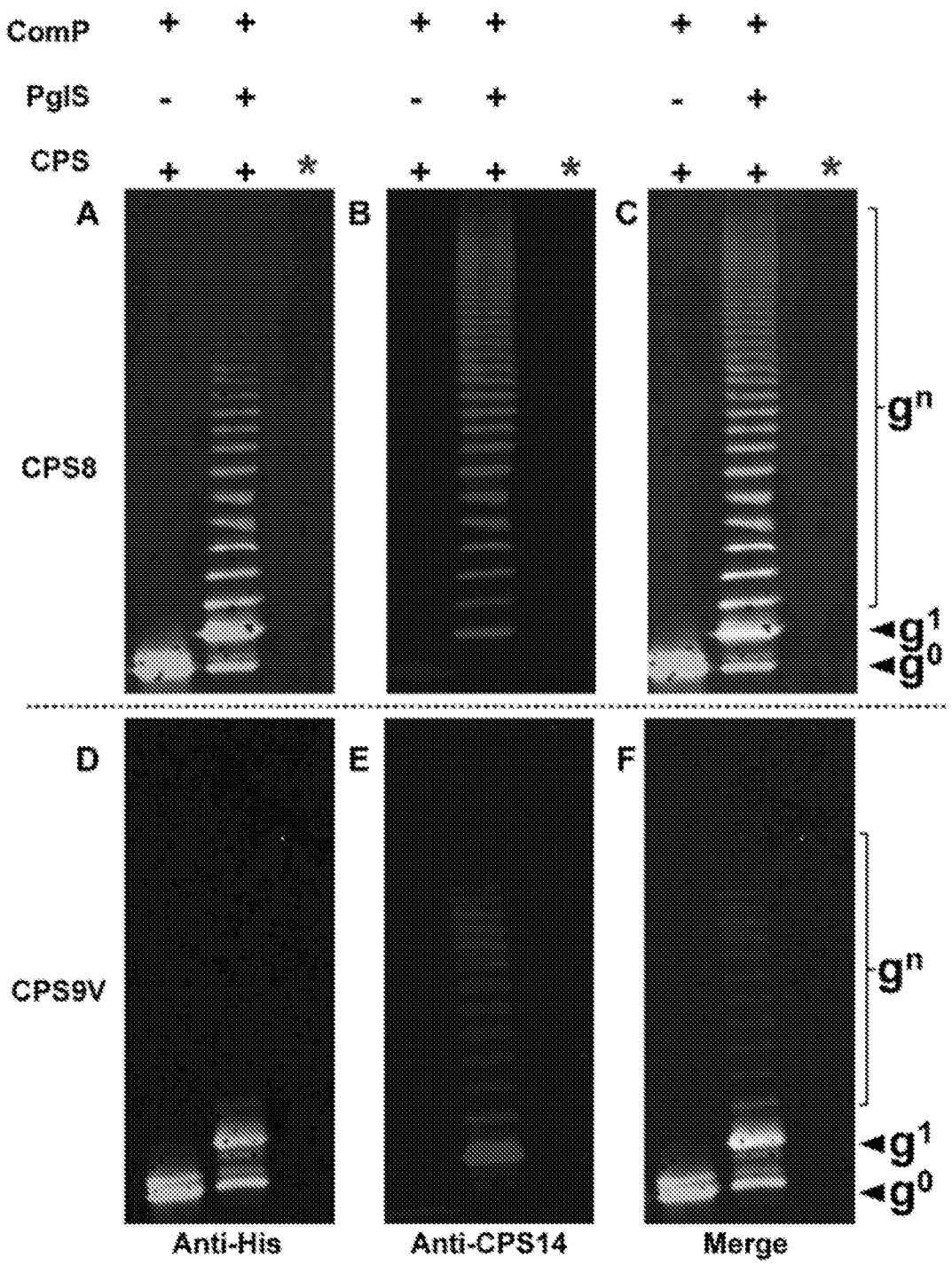

ComP
PglS
CPS
A
B
C
CPS8
g^n
g^1
g^0
D
E
F
CPS9V
g^n
g^1
g^0
Anti-His
Anti-CPS14
Merge

Figure 34A-F

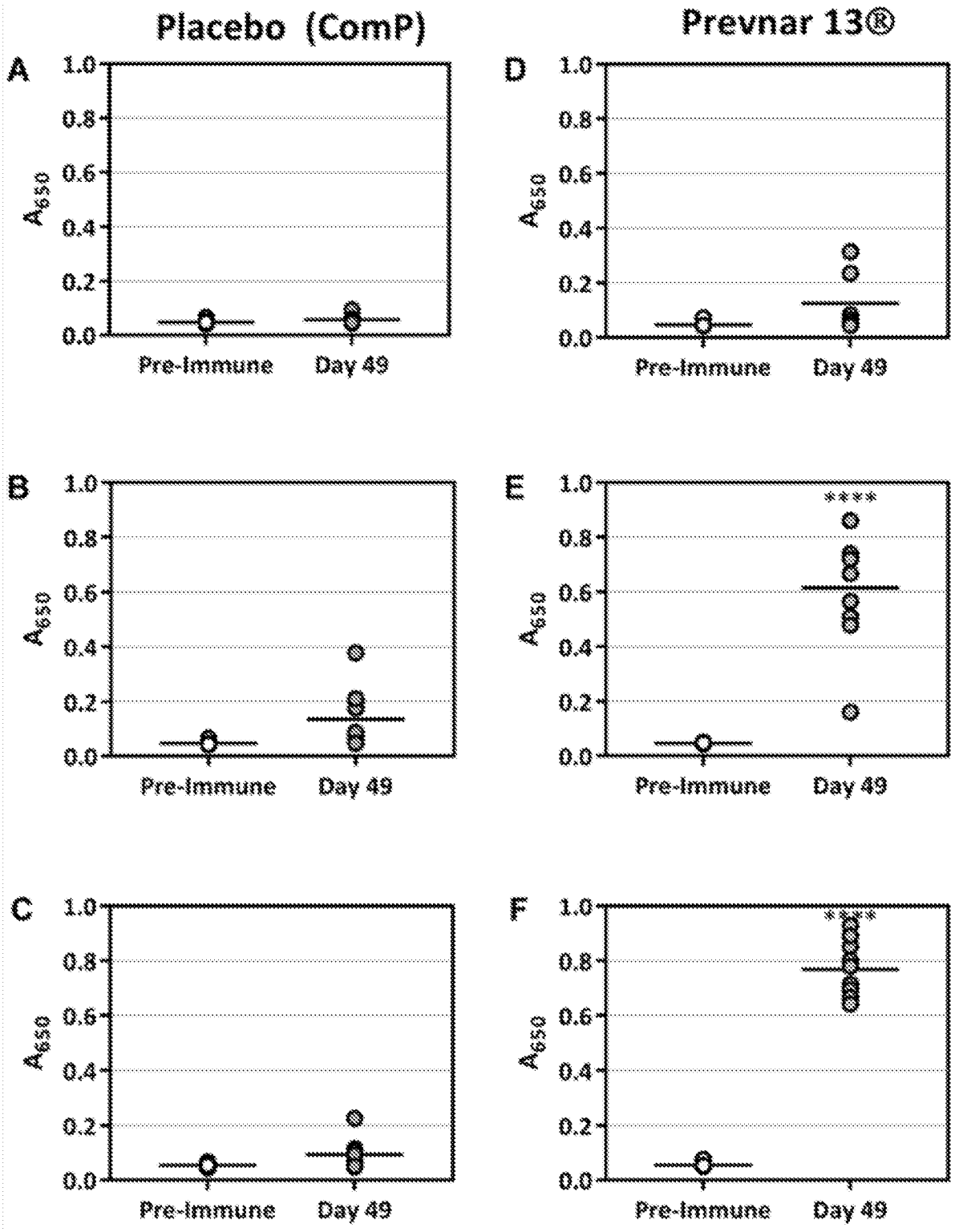

Placebo (ComP)
Prevnar 13®
A
B
C
D
E
F
A_{650}
1.0
0.8
0.6
0.4
0.2
0.0
Pre-Immune
Day 49
****

Figure 35A-F

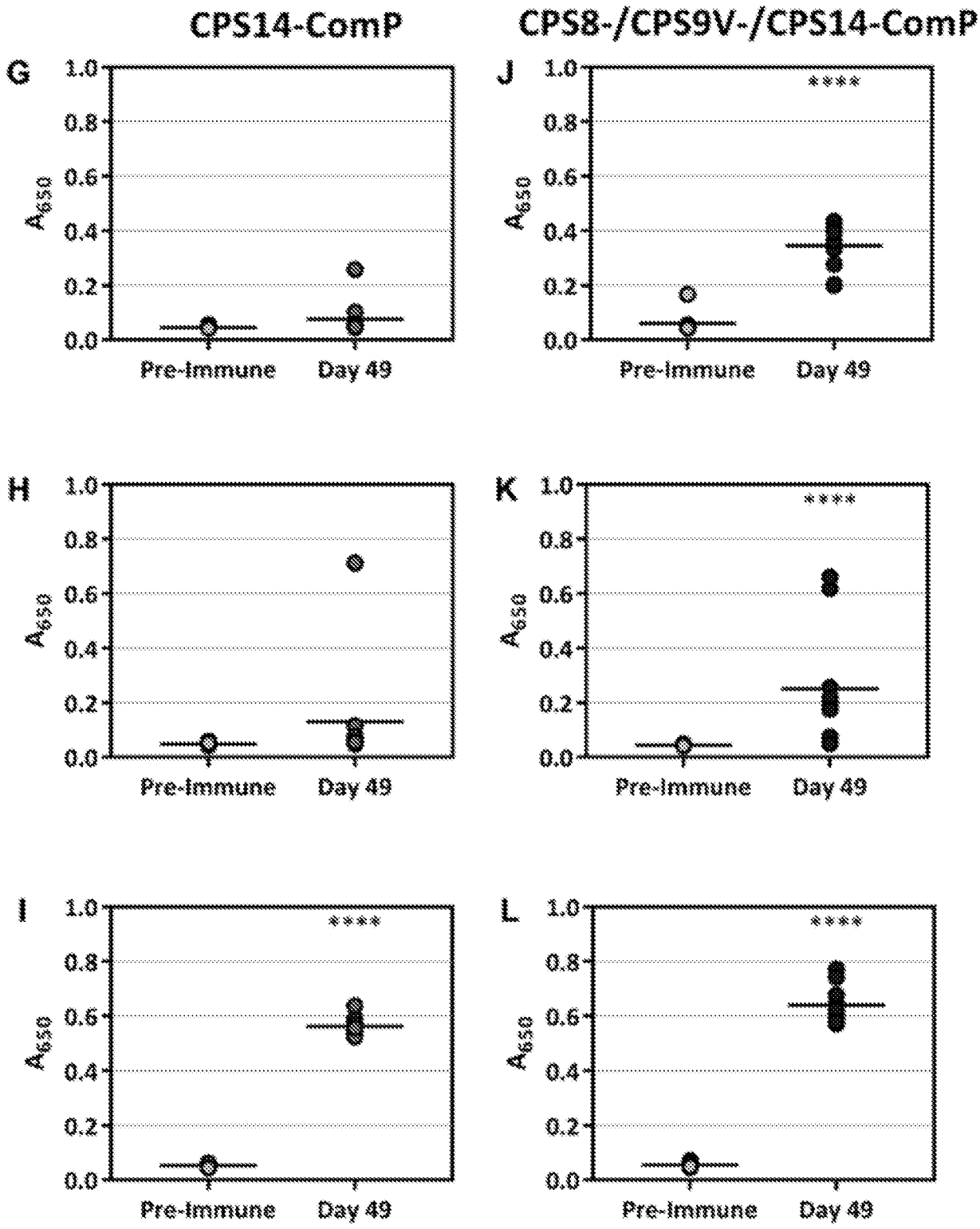

CPS14-ComP
CPS8-/CPS9V-/CPS14-ComP
G
H
I
J
K
L
A_{650}
1.0
0.8
0.6
0.4
0.2
0.0
Pre-Immune
Day 49
****

Figure 35G-L

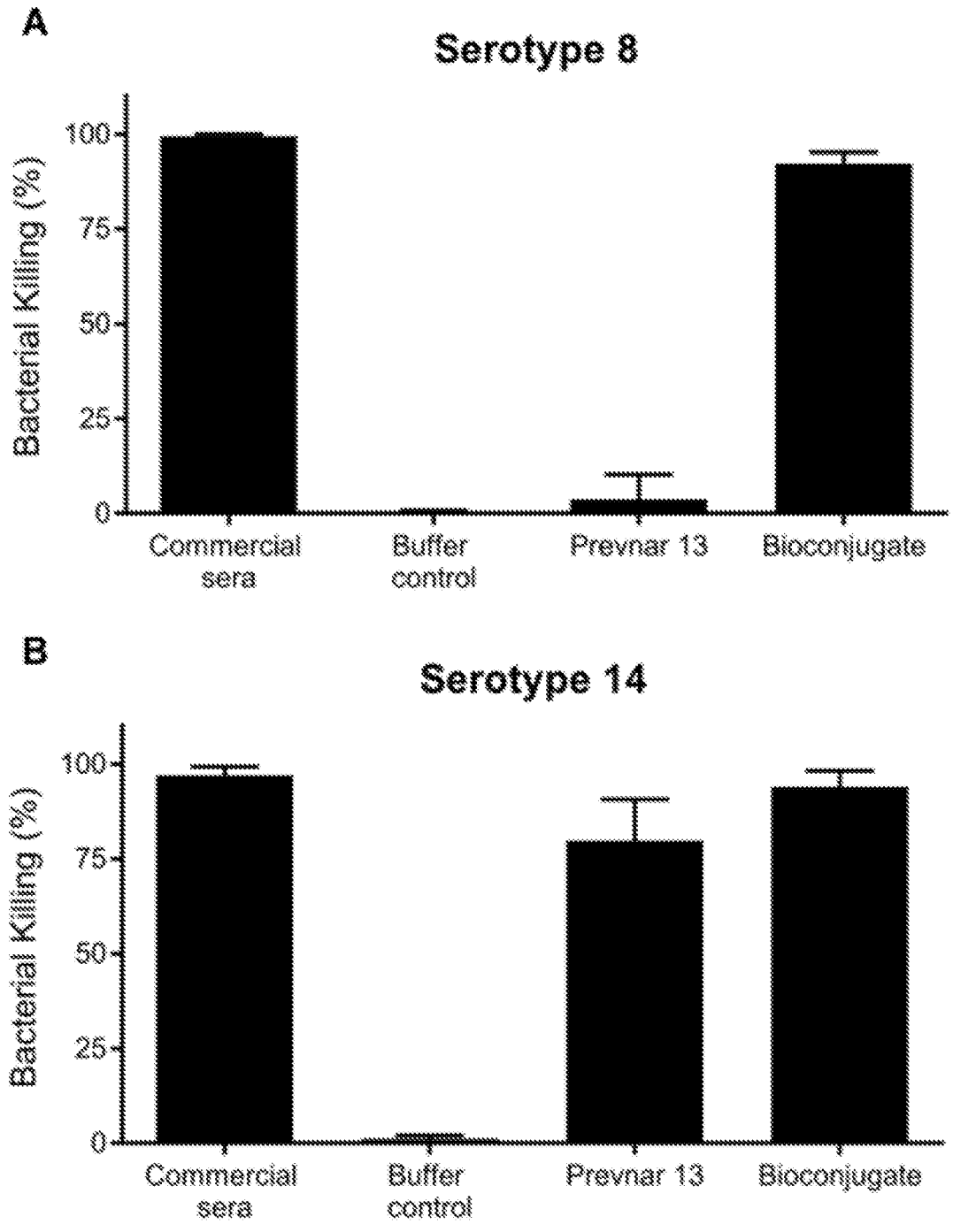

A
Serotype 8
Bacterial Killing (%)
100
75
50
25
0
Commercial sera
Buffer control
Prevnar 13
Bioconjugate
B
Serotype 14
Bacterial Killing (%)
100
75
50
25
0
Commercial sera
Buffer control
Prevnar 13
Bioconjugate

Figure 36A,B

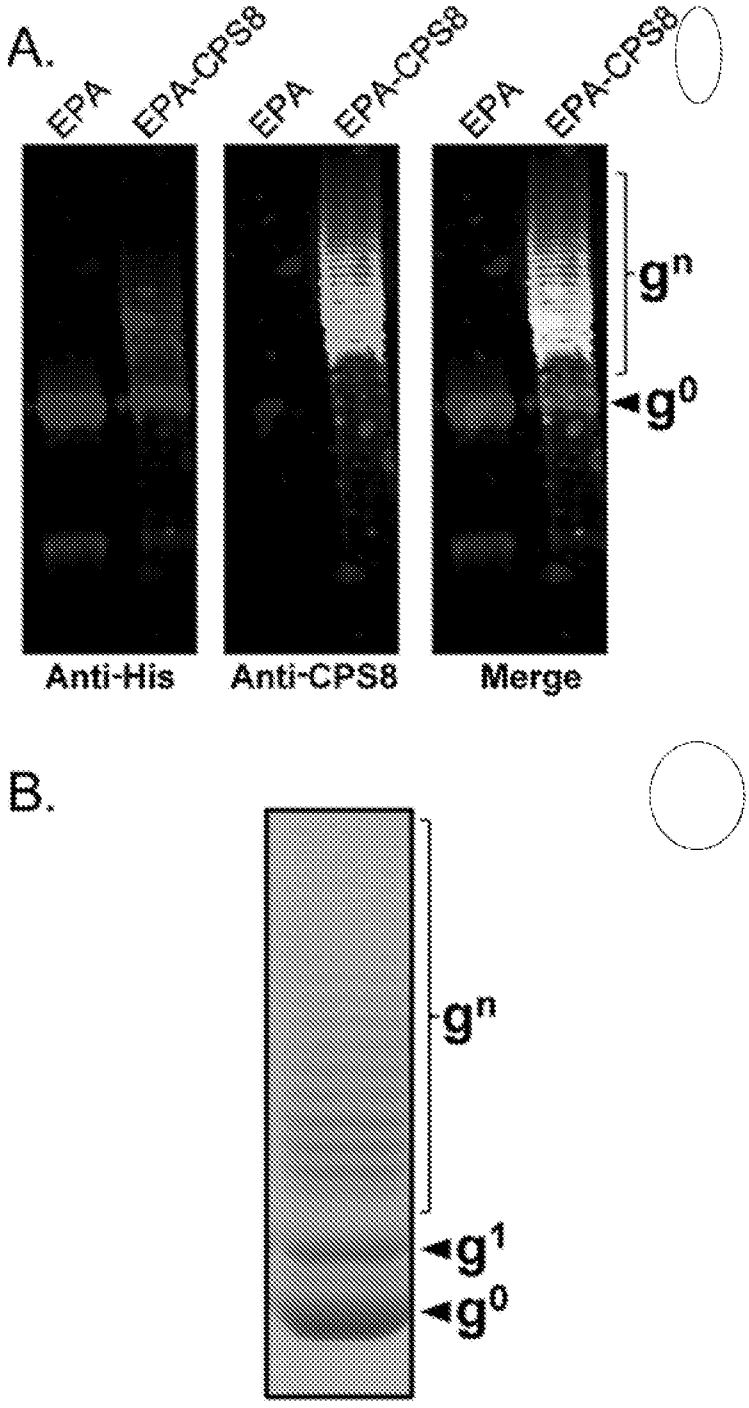

A.
EPA
EPA-CPS8
EPA
EPA-CPS8
EPA
EPA-CPS8
g^n
g^0
Anti-His
Anti-CPS8
Merge
B.
g^n
g^1
g^0

Figure 37A,B

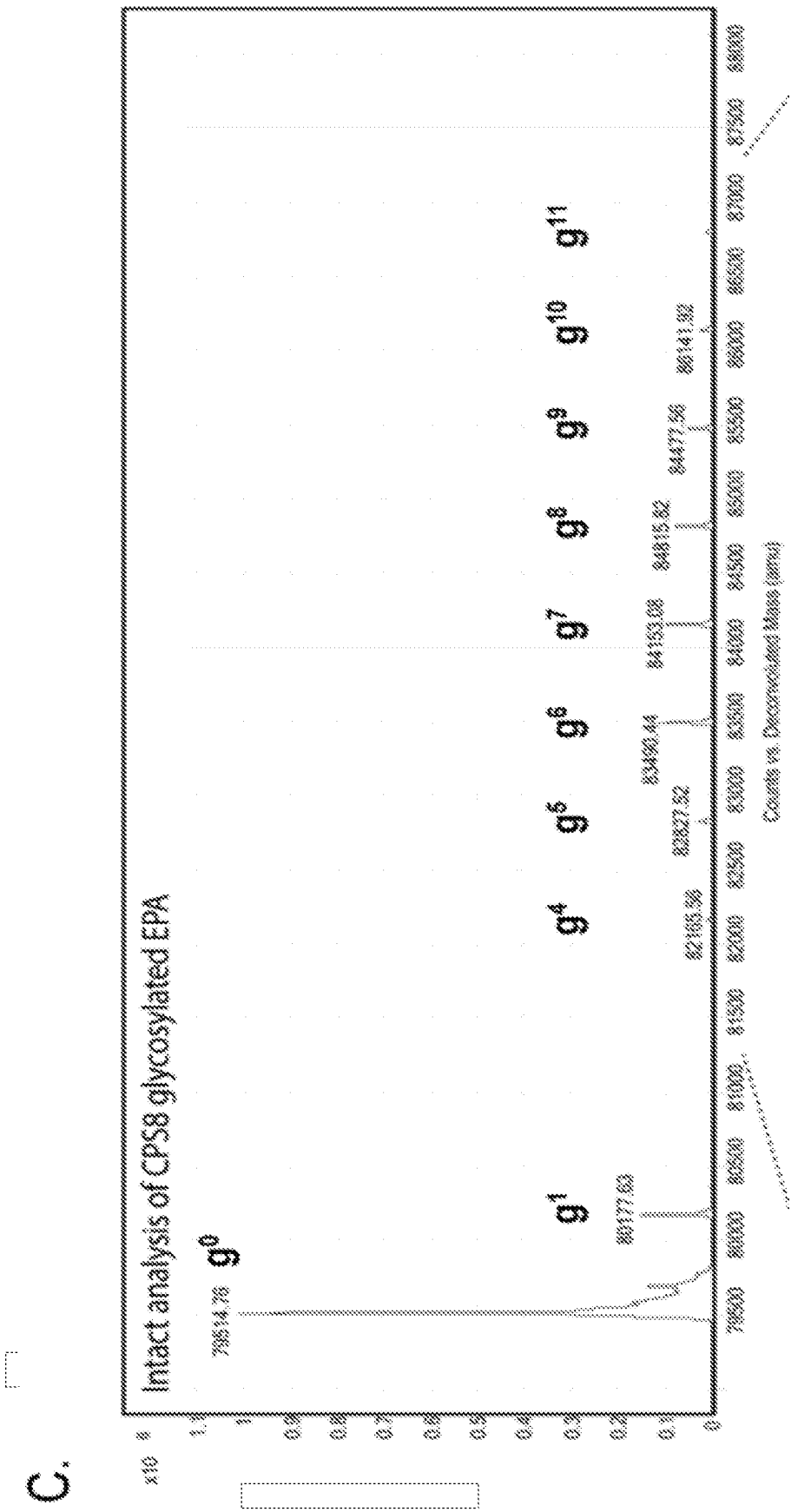

C:
Intact analysis of CPS8 glycosylated EPA
g0
g1
g4
g5
g6
g7
g8
g9
g10
g11
Counts vs. Deconvoluted Mass (amu)

Figure 37C

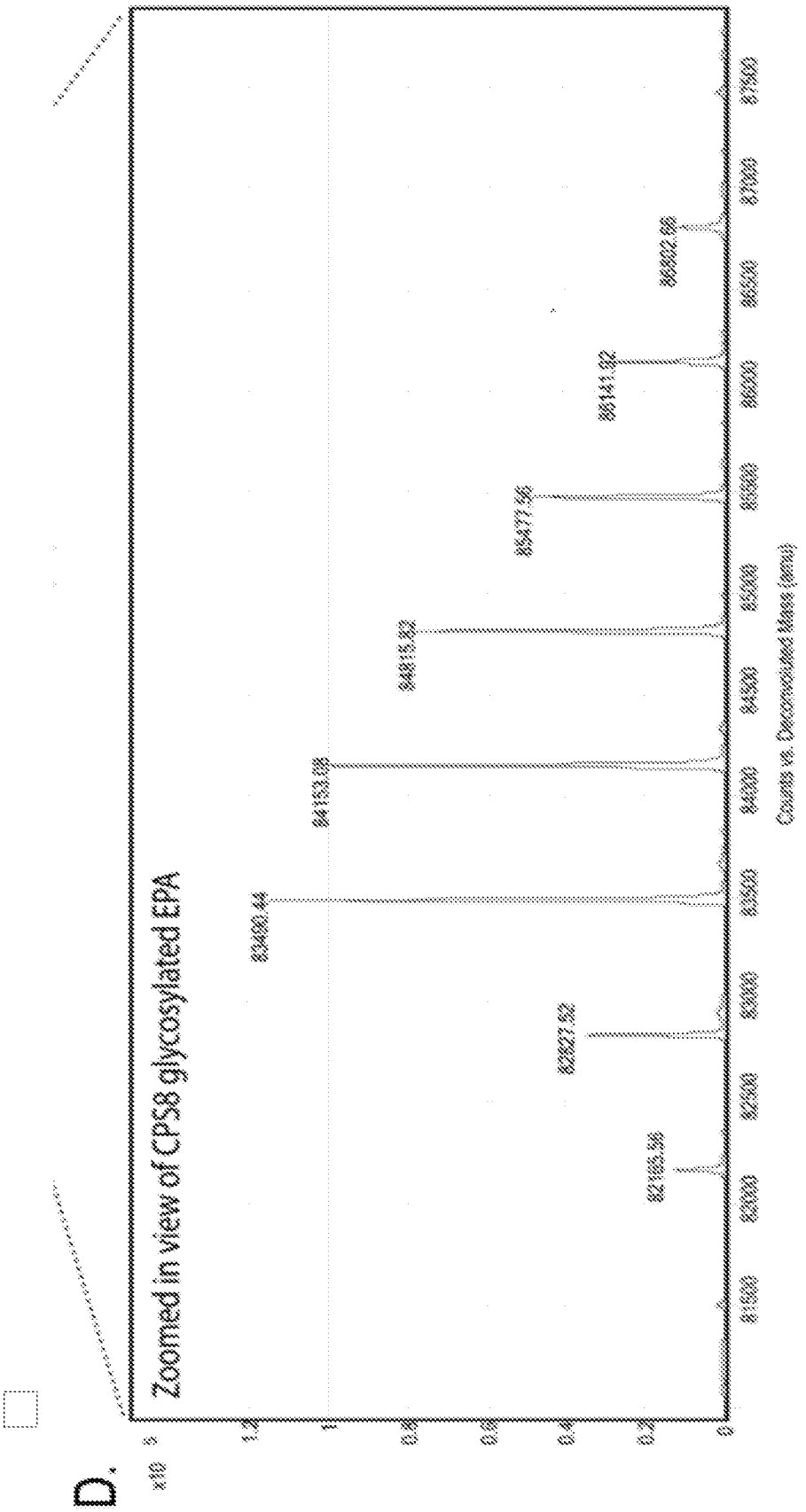

D.
Zoomed in view of CPS8 glycosylated EPA
x10
Counts vs. Deconvoluted Mass (amu)
82165.36
82821.52
83480.44
84153.68
84815.32
85477.56
86141.92
86802.68

Figure 37D

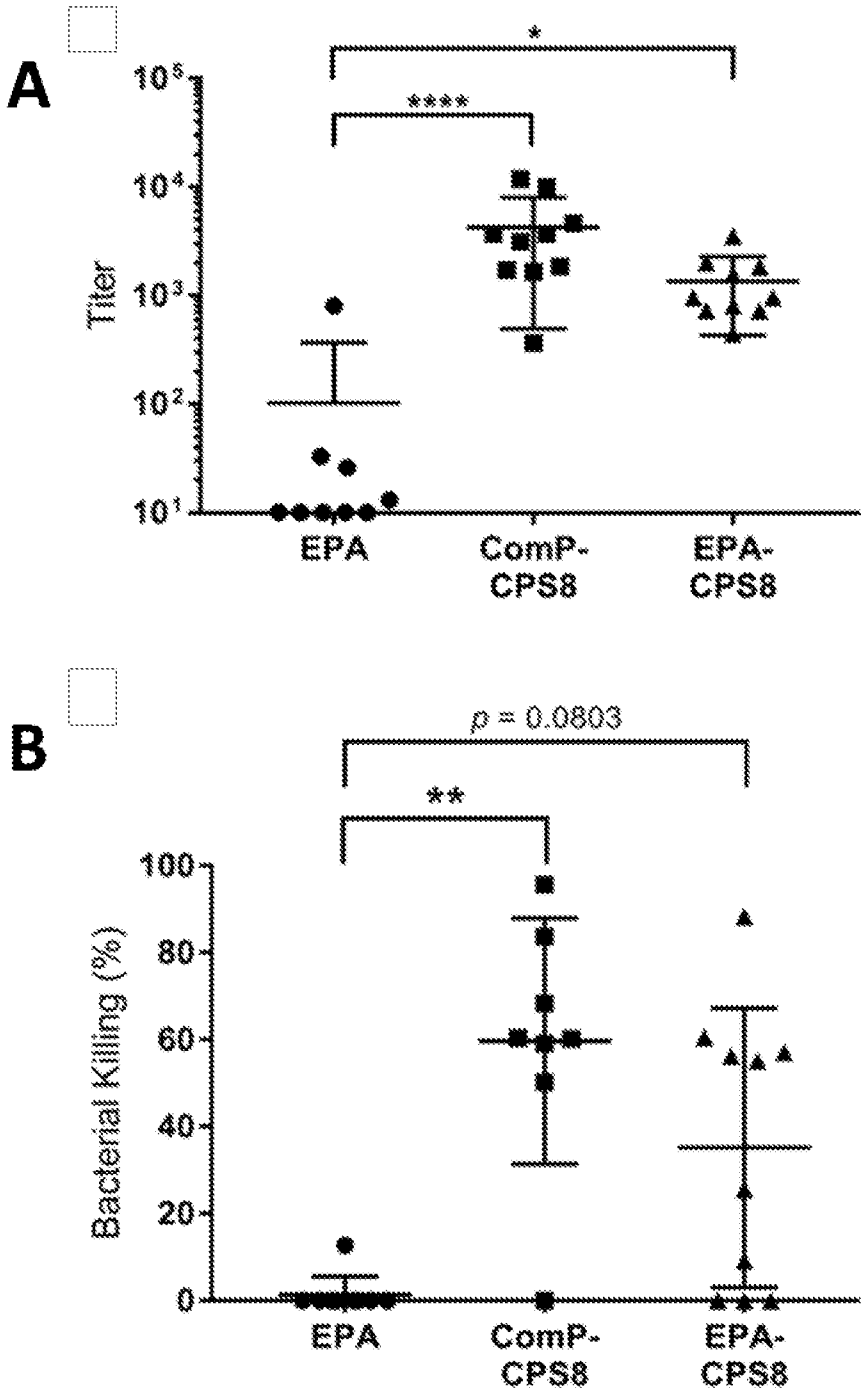

A
*
****
10^5
10^4
10^3
10^2
10^1
Titer
EPA
ComP-CPS8
EPA-CPS8
B
p = 0.0803
**
100
80
60
40
20
0
Bacterial Killing (%)
EPA
ComP-CPS8
EPA-CPS8

Figure 38A,B

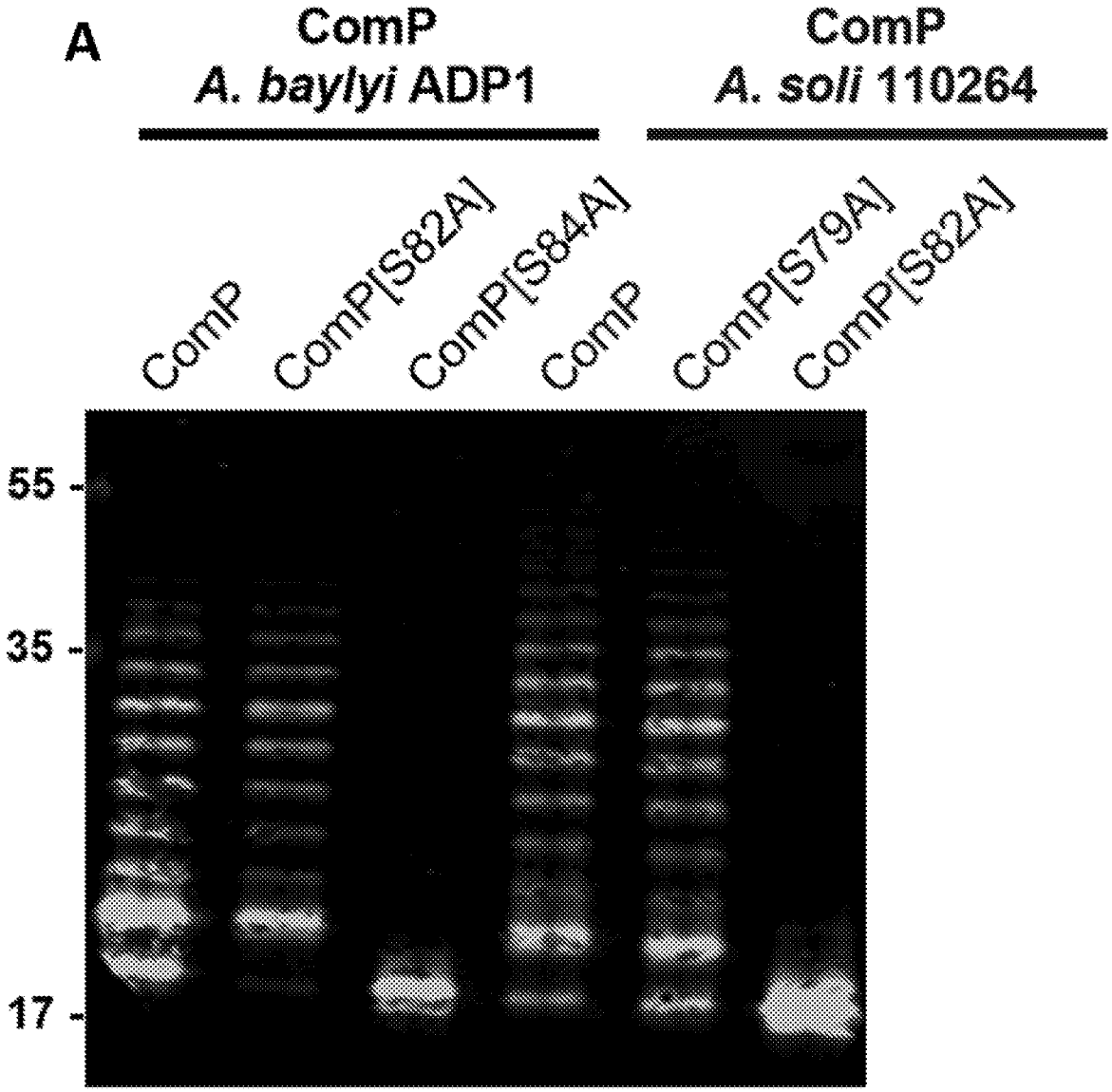

A
ComP
A. baylyi ADP1
ComP
A. soli 110264
ComP
ComP[S82A]
ComP[S84A]
ComP
ComP[S79A]
ComP[S82A]
55
35
17

Figure 39A

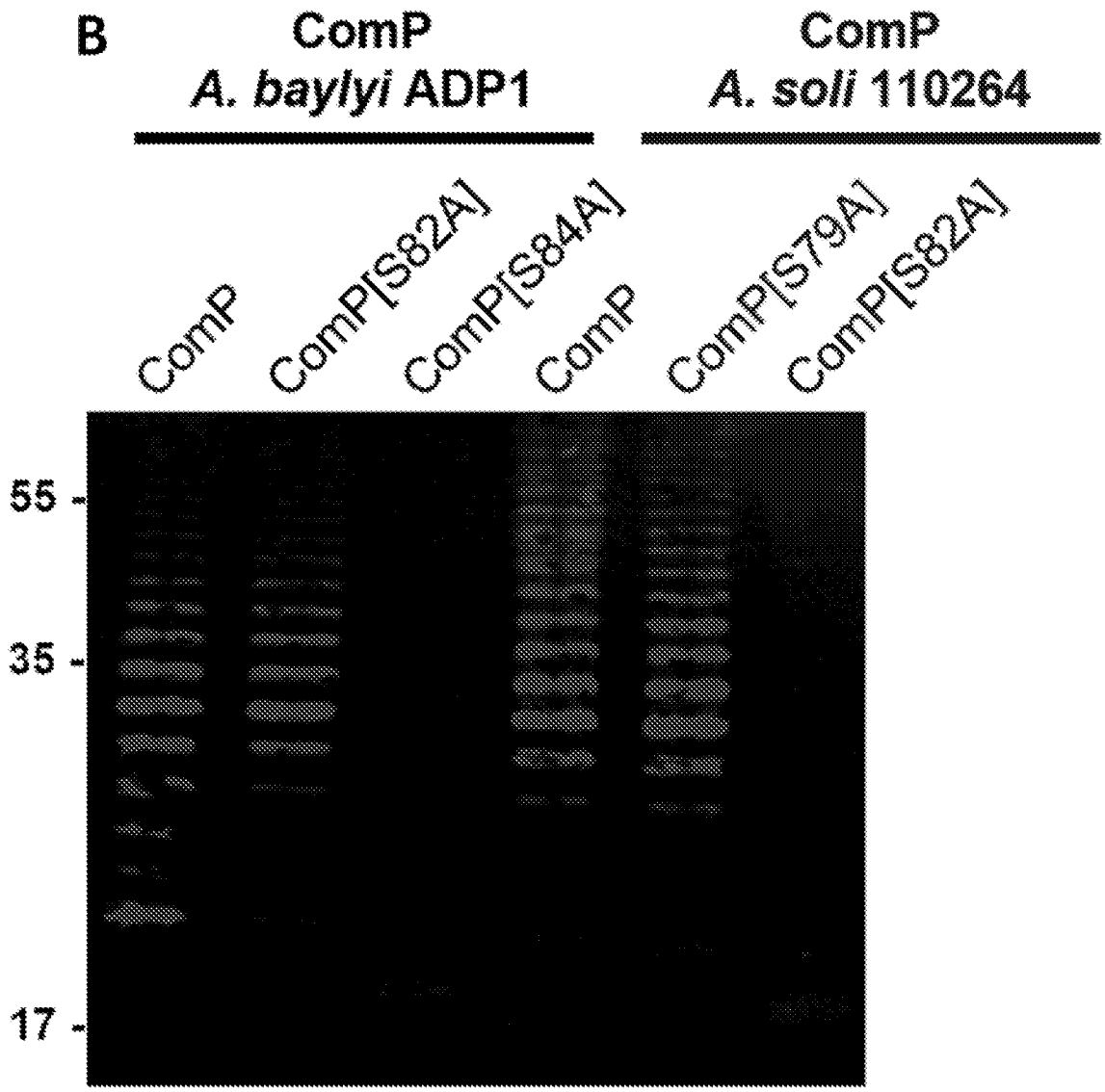

B
ComP
A. baylyi ADP1
ComP
A. soli 110264
ComP
ComP[S82A]
ComP[S84A]
ComP
ComP[S79A]
ComP[S82A]
55
35
17

Figure 39B

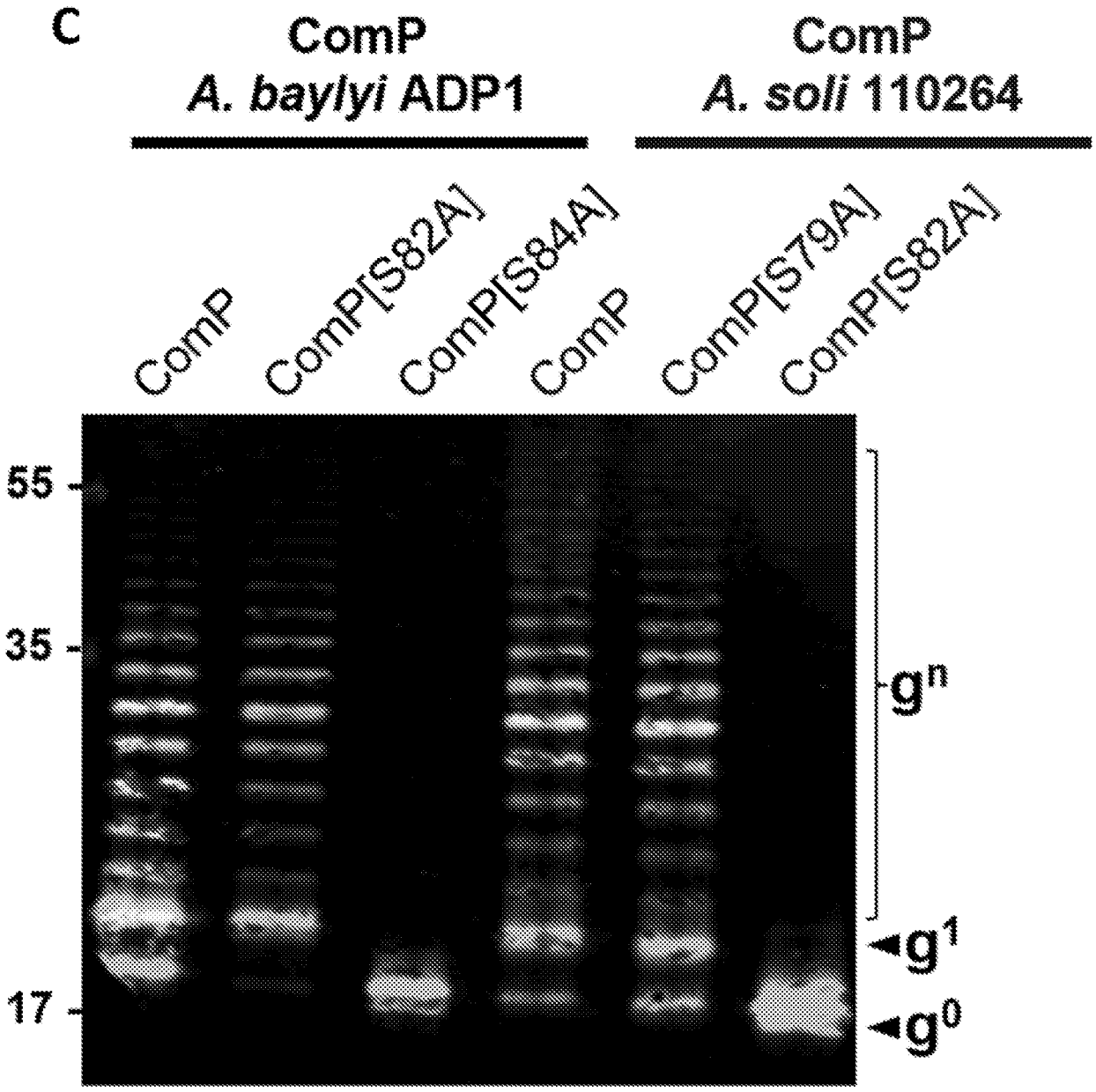

C
ComP
A. baylyi ADP1
ComP
A. soli 110264
ComP
ComP[S82A]
ComP[S84A]
ComP
ComP[S79A]
ComP[S82A]
55
35
17
gn
g1
g0

Figure 39C

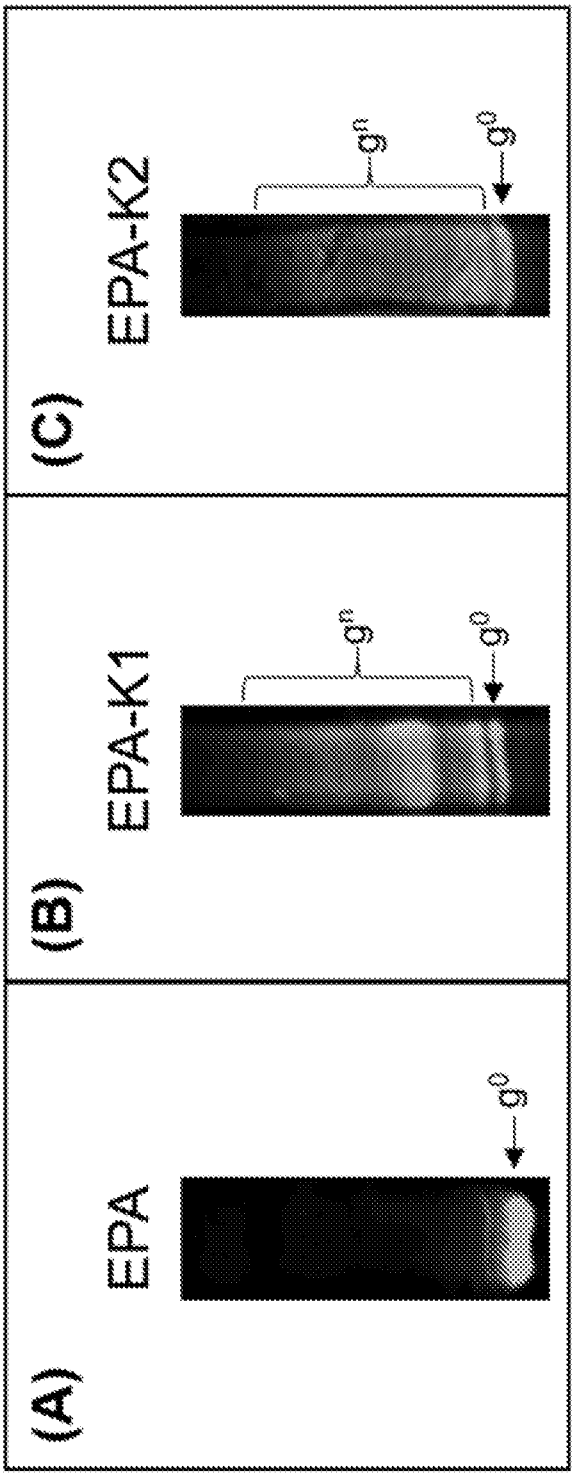


Figure 40A,B,C

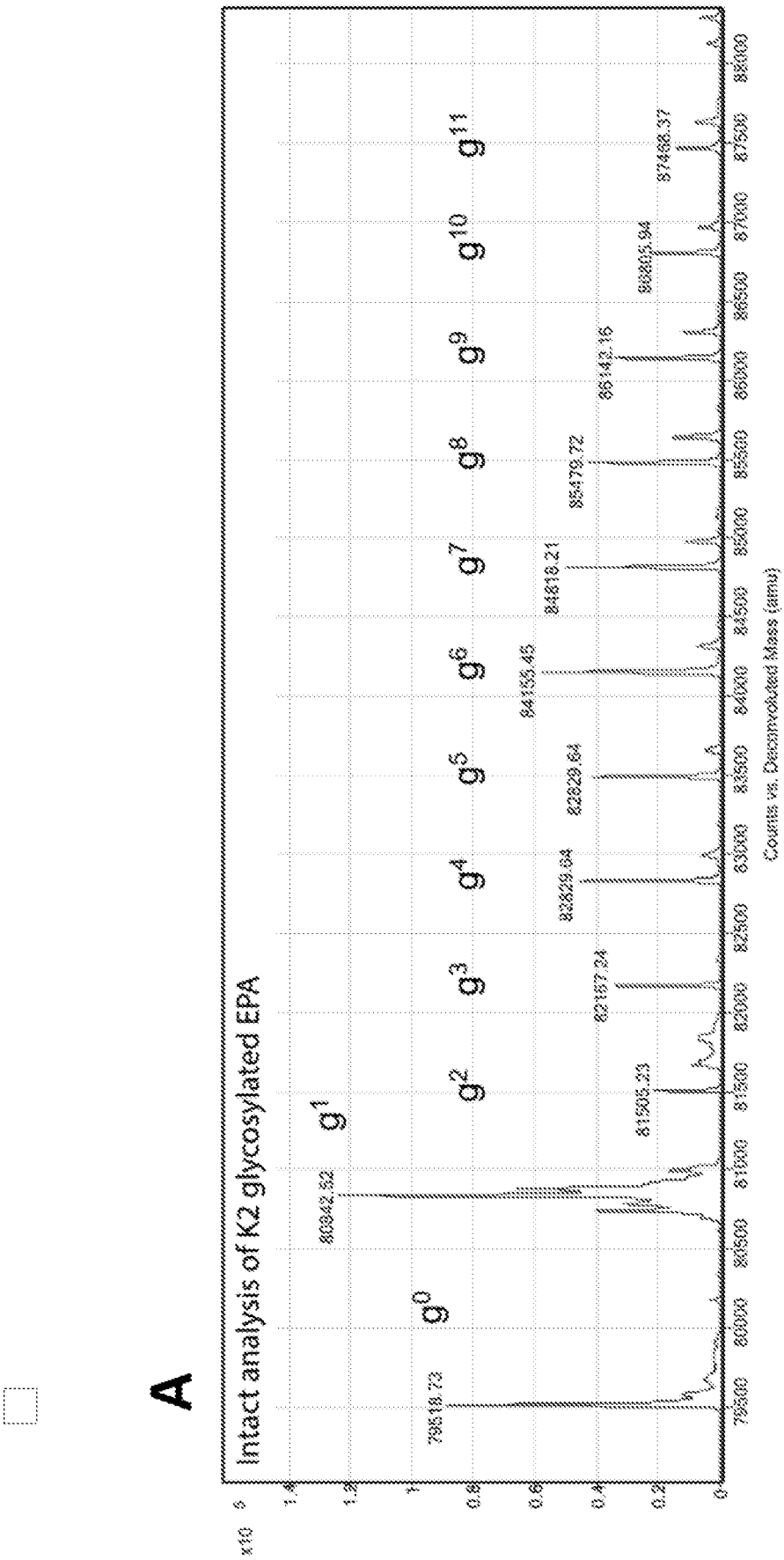

A
Intact analysis of K2 glycosylated EPA
x10 5
Counts vs. Deconvoluted Mass (amu)
79518.72
80342.52
81505.23
82187.24
82829.64
82829.64
84155.46
84816.21
85479.72
86143.16
86805.94
87468.37
g0
g1
g2
g3
g4
g5
g6
g7
g8
g9
g10
g11

Figure 41A

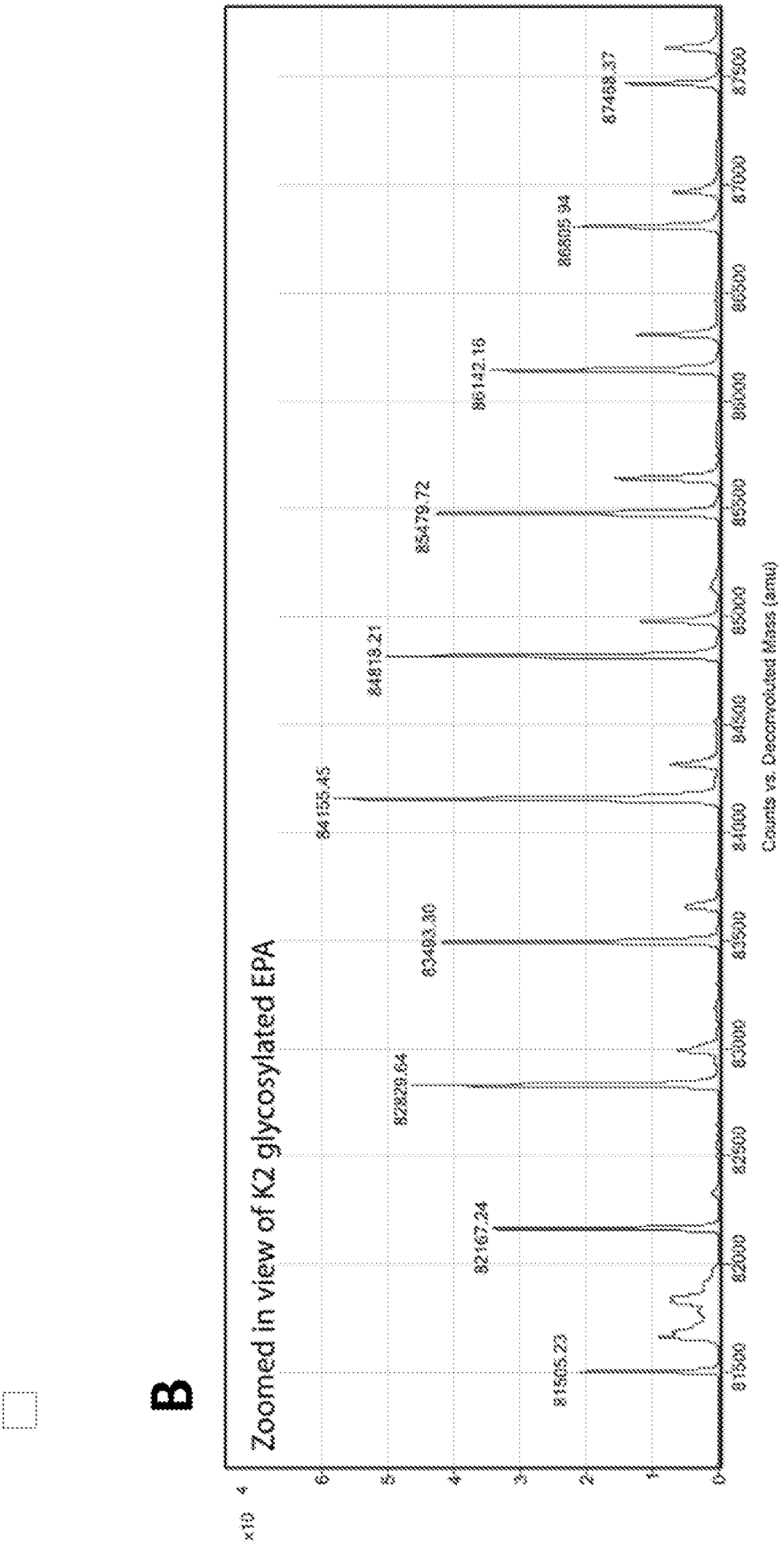

B
Zoomed in view of K2 glycosylated EPA
x10 4
Counts vs. Deconvoluted Mass (amu)
81505.23
82167.24
82928.64
83483.30
84155.45
84819.21
85479.72
86142.16
86855.94
87468.37

Figure 41B

A.

$ComP_{ADP1}$ fragment: $_{75}$CVGVQEISA**S**NATTNVATATC$_{95}$ (SEQ ID NO: 26)