

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2019287659 B2**

(54) Title  
**Glycosylated ComP pilin variants, methods of making and uses thereof**

(51) International Patent Classification(s)  
**A61K 38/02** (2006.01)                      **C07K 19/00** (2006.01)  
**A61K 39/02** (2006.01)                      **C12N 9/10** (2006.01)  
**A61P 31/04** (2006.01)                      **C12P 21/00** (2006.01)  
**C07K 14/195** (2006.01)

(21) Application No: **2019287659**                      (22) Date of Filing: **2019.06.14**

(87) WIPO No: **WO19/241672**

(30) Priority Data

(31) Number	(32) Date	(33) Country
<b>62/685,970</b>	<b>2018.06.16</b>	<b>US</b>
<b>62/783,971</b>	<b>2018.12.21</b>	<b>US</b>

(43) Publication Date: **2019.12.19**

(44) Accepted Journal Date: **2025.02.20**

(71) Applicant(s)  
**VaxNewMo LLC**

(72) Inventor(s)  
**HARDING, Christian; FELDMAN, Mario**

(74) Agent / Attorney  
**RNB IP PTY LTD, PO Box 9530, Deakin, ACT, 2600, AU**

(56) Related Art  
**WO 2016/134485 A1**



(51) International Patent Classification:

A61K 38/02 (2006.01) C07K 19/00 (2006.01)  
A61K 39/02 (2006.01) C12N 9/10 (2006.01)  
A61P 31/04 (2006.01) C12P 21/00 (2006.01)  
C07K 14/195 (2006.01)

(21) International Application Number:

PCT/US2019/037251

(22) International Filing Date:

14 June 2019 (14.06.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/685,970 16 June 2018 (16.06.2018) US  
62/783,971 21 December 2018 (21.12.2018) US

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

(72) Inventors: HARDING, Christian; c/o VaxNewMo LLC, 4340 Duncan Ave., Suite 202, St. Louis, MO 63110 (US).  
FELDMAN, Mario; c/o VaxNewMo LLC, 4340 Duncan Ave., Suite 202, St. Louis, MO 63110 (US).

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

(54) Title: GLYCOSYLATED COMP PILIN VARIANTS, METHODS OF MAKING AND USES THEREOF

(57) Abstract: Provided herein are glycosylated ComP proteins, fragments and fusion proteins thereof, and methods of making, for example, for use in the production of conjugate vaccines. Also provided herein are conjugate vaccines against diseases including bacterial diseases.

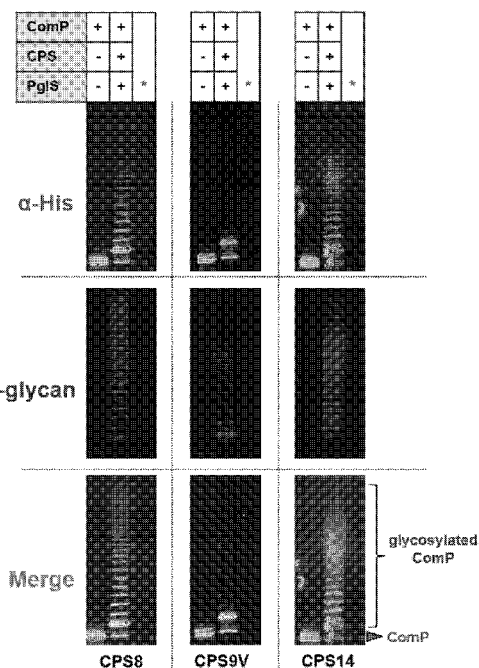


Figure 2

WO 2019/241672 A3

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

**(88) Date of publication of the international search report:**  
06 February 2020 (06.02.2020)

- 1 -

## GLYCOSYLATED COMPILIN VARIANTS, METHODS OF MAKING AND USES THEREOF

Inventors: Christian Harding  
Mario Feldman

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This international application claims the benefit of U.S. Provisional Application 62/685,970, filed on June 16, 2018 and U.S. Provisional Application 62/783,971, filed on December 21, 2018, both of which are incorporated herein in their entireties.

[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.

### GOVERNMENT FUNDING STATEMENT

[0003] This invention was made with government support under the R41 AI131742 grant awarded by the National Institute for Allergy and Infectious Disease (NIAID). The government has certain rights in the invention.

### BACKGROUND

[0004] *Streptococcus pneumoniae* (*S. pneumoniae*) is a leading cause of pneumonia globally, particularly, afflicting children five years of age or younger (O'Brien, K.L. et al. *Lancet* 374, 893-902 (2009)). Estimates indicate that ~1.5 million people die each year as a result of *S. pneumoniae* infection, almost one million of those deaths are among children (Pneumococcal conjugate vaccine for childhood immunization—WHO position paper. *Wkly Epidemiol Rec* 82, 93-104 (2007)). The recommended prophylactic treatments comprise multiple commercially licensed vaccines (Prevention, C.f.D.C.a. Pneumococcal Vaccination, on the world wide web at [cdc.gov/vaccines/vpd/pneumo/index.html](http://cdc.gov/vaccines/vpd/pneumo/index.html)). PNEUMOVAX 23®, a 23-valent polysaccharide vaccine, is used in elderly populations as polysaccharide vaccines usually act as T cell independent antigens, do not elicit high avidity IgG responses or B cell memory, and are not effective in children (Pace, D. *Expert Opin Biol Ther* 13, 11-33 (2013); Vella, M. & Pace, D. *Expert Opin Biol Ther* 15, 529-546 (2015)). On the other hand, pneumococcal conjugate

- 2 -

vaccines, comprised of pneumococcal capsular polysaccharide covalently attached to a carrier protein, have been shown to be effective and generate immunological memory across all age groups due to their ability to act as T-cell dependent antigens (Pollard, A.J., Perrett, K.P. & Beverley, P.C. *Nat Rev Immunol* 9, 213-220 (2009); Avci, F.Y., Li, X., Tsuji, M. & Kasper, D.L. *Nat Med* 17, 1602-1609 (2011)).

**[0005]** Three pneumococcal conjugate vaccines have been commercially licensed since the year 2000: PREVNAR®; SYNFLORIX; and PREVNAR 13®, the most broadly protecting pneumococcal conjugate vaccine, is comprised of 13 protein-polysaccharide conjugates consisting of pneumococcal serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19A, 19F, and 23F, each individually linked to the genetically inactivated diphtheria toxoid CRM<sub>197</sub> (Package Insert - PREVNAR 13® - FDA, on the world wide web at [fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM201669.pdf](http://fda.gov/downloads/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM201669.pdf)). Although highly protective in a three dose primary schedule, Prevnar 13 is one of the most expensive vaccines produced due to its complex manufacturing process resulting in a price tag of ~\$600 US dollars for primary and booster immunizations (Prevention, C.f.D.C.a. Vaccines for Children Program (VFC), on the world wide web at [cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html](http://cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html) (2018)). In fact, PREVNAR 13® has been Pfizer's leading revenue generating product from 2015-2017 with total revenues exceeding 17.5 billion U.S. dollars (Pfizer Inc. 2017 Financial Report, on the world wide web at [sec.gov/Archives/edgar/data/78003/000007800318000027/pfe-exhibit13x12312017x10k.htm](http://sec.gov/Archives/edgar/data/78003/000007800318000027/pfe-exhibit13x12312017x10k.htm) (2018)). Although pneumococcal conjugate vaccines, namely PREVNAR 13® and SYNFLORIX, have significantly reduced the burden of invasive pneumococcal disease, variations in global serotype distributions as well as serotype replacement and displacement events necessitate the introduction of a broader PCV providing additional protection to vulnerable patient populations.

**[0006]** Currently licensed pneumococcal conjugate vaccines are synthesized chemically, which is a tedious process plagued with technical challenges, low yields, and batch-to-batch variations; highlighting the need for improved conjugate vaccine synthetic methodologies (Frasch, C.E. *Vaccine* 27, 6468-6470 (2009)). Over the last 15 years, *in vivo* conjugation using bacterial protein glycosylation systems has emerged as a feasible alternative to chemical conjugations, with multiple bioconjugate vaccine candidates now in various stages of development and clinical trials (Huttner, A. & Gambillara, V. *Clin Microbiol Infect* (2018));

Huttner, A. et al. *Lancet Infect Dis* 17, 528-537 (2017); Riddle, M.S. et al. *Clin Vaccine Immunol* 23, 908-917 (2016)). Protein glycosylation is a ubiquitous post-translational modification in which carbohydrates, also known as sugars or glycans, are covalently linked to proteins (aka polypeptides) (Apweiler, R., Hermjakob, H. & Sharon, N. *Biochim Biophys Acta* 1473, 4-8 (1999)). In bacteria, glycans are commonly bound to proteins via *N*- or *O*-linkages on asparagine and serine/threonine residues respectively (Nothhaft, H. & Szymanski, C.M. *Nat Rev Microbiol* 8, 765-778 (2010)). Several pathways for bacterial glycosylation have been characterized, and one of the best described is the oligosaccharyltransferase (OTase)-dependent glycosylation pathway in Gram negative bacteria (Iwashkiw, J.A., Vozza, N.F., Kinsella, R.L. & Feldman, M. F. *Mol Microbiol* 89 (2013)). In this system, generally a lipid-linked oligosaccharide is assembled sequentially at the cytoplasmic leaflet of the inner membrane, flipped to the periplasmic leaflet, and then transferred to acceptor proteins by either *N*- or *O*-OTases depending on the site of glycan attachment generating a variety of glycoproteins (Iwashkiw, J.A., Vozza, N.F., Kinsella, R.L. & Feldman, M.F. *Mol Microbiol* 89, 14-28 (2013)).

**[0007]** Glycoproteins have been recombinantly synthesized in *Escherichia coli* (*E. coli*) for use as vaccines and/or diagnostics by co-expressing three components: (1) a genetic cluster encoding for the proteins required to synthesize a glycan of interest; (2) an OTase; and (3) an acceptor protein (Ciocchini, A.E. et al. *Vet Microbiol* 172, 455-465 (2014); Garcia-Quintanilla, F., Iwashkiw, J.A., Price, N.L., Stratilo, C. & Feldman, M.F. *Front Microbiol* 5, 381 (2014); Iwashkiw, J.A. et al. *Microb Cell Fact* 11, 13 (2012)). One drawback of this process is the apparent substrate specificity of the known OTases, which has been suggested to be regulated by the reducing end sugar (Wacker, M. et al. *Proc Natl Acad Sci USA* 103, 7088-7093 (2006)) (*i.e.*, the first monosaccharide in the growing polysaccharide chain). Although OTases are able to transfer many different oligo- and polysaccharide structures, some sugars are not efficiently conjugated by the known OTases to acceptor proteins. Therefore, characterizing novel OTases is paramount for developing the next generation of conjugate vaccines.

**[0008]** OTases currently used for commercially synthesizing glycoconjugates are the *Campylobacter jejuni* *N*-OTase PglB and the *Neisseria meningitidis* *O*-OTase PglL, both of which exhibit a great deal of promiscuity towards glycan substrates (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)). However, neither enzyme has been experimentally demonstrated to conjugate *Streptococcus pneumoniae* capsular

- 4 -

polysaccharides (CPSs) containing glucose as the reducing end sugar to proteins (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)), to proteins. More than 90 CPS serotypes have been characterized for pneumococcus, each possessing a structurally distinct capsular polysaccharide structure; however, more than 70% of *S. pneumoniae* CPSs contain glucose as the reducing end sugar (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015)). Therefore, to complement and/or replace existing manufacturing pipelines in order to more rapidly generate the next generation of pneumococcal conjugate vaccine, novel methods of pneumococcal vaccine synthesis are needed.

## SUMMARY

[0009] Provide for herein is a bioconjugate, for example certain aspects and features of which are described in this paragraph, comprising an oligo- or polysaccharide covalently linked to a fusion protein, wherein the fusion protein comprises a ComP protein (ComP) or a glycosylation tag fragment thereof. In certain aspects, the fusion protein is glycosylated with the oligo- or polysaccharide on the ComP protein or glycosylation tag fragment thereof at a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, 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 $\Delta$ 28<sub>ADP1</sub>), and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the ComP protein comprises SEQ ID NO: 7 (ComP $\Delta$ 28<sub>ADP1</sub>), SEQ ID NO: 8 (ComP $\Delta$ 28<sub>110264</sub>), SEQ ID NO: 9 (ComP $\Delta$ 28<sub>GFJ-2</sub>), SEQ ID NO: 10 (ComP $\Delta$ 28<sub>P50v1</sub>), SEQ ID NO: 11 (ComP $\Delta$ 28<sub>4466</sub>), or SEQ ID NO: 12 (ComP $\Delta$ 28<sub>SFC</sub>). In certain aspects, 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<sub>ADP1</sub>: AAC45886.1), and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the ComP protein comprises SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1), SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1), SEQ ID NO: 3 (ComP<sub>GFJ-2</sub>: APV36638.1), SEQ ID NO: 4 (ComP<sub>50v1</sub>: PKD82822.1), SEQ ID NO: 5 (ComP<sub>4466</sub>: SNX44537.1), or SEQ ID NO: 6 (ComP<sub>SFC</sub>: OAL75955.1). In certain aspects, the glycosylation tag fragment of the ComP protein is a ComP $\Delta$ 28 polypeptide lacking amino acid residues corresponding to amino acid residues 1 to 28 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In

- 5 -

certain aspects, the ComPA28 polypeptide is selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12. In certain aspects, the glycosylation tag fragment of the ComP protein comprises a region corresponding to the region of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) comprising the serine residue at position 82 of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) flanked by a disulfide bond connecting the alpha beta loop to the beta strand region. In certain aspects, the glycosylation tag fragment of the ComP protein comprises an amino acid sequence that is at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to an amino acid selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag fragment of the ComP protein comprises an amino acid sequence selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag fragment of the ComP protein comprises the amino acid consensus sequence of SEQ ID NO: 37, or a fragment of at least 5, 10, 15, 20, 30, 35, or 40 consecutive amino acids thereof, wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag fragment of the ComP protein comprises the amino acid consensus sequence of SEQ ID NO: 38 or 45, or a fragment of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive amino acids thereof, wherein said glycosylation tag fragment comprises the

- 6 -

serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the oligo- or polysaccharide is produced by a bacteria from the genus *Streptococcus*. In certain aspects, the polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide. In certain aspects, the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b. In certain aspects, the capsular polysaccharide is CPS8. In certain aspects, the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*. In certain aspects, the oligo- or polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varriicola*, *Klebsiella michinganensis*, or *Klebsiella oxytoca* capsular polysaccharide. In certain aspects, the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide. In certain aspects, the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*. In certain aspects, the oligo- or polysaccharide comprises a glucose at its reducing end. In certain aspects, the bioconjugate is produced *in vivo*. In certain aspects, the bioconjugate is produced in a bacterial cell. In certain aspects, 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. In certain aspects, the ComP protein or glycosylation tag fragment thereof 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. In certain aspects, the carrier protein or fragment thereof is linked to the ComP protein or glycosylation tag fragment thereof via an amino acid linker. In certain aspects, 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 glycosylation tag fragments of a ComP protein. In certain aspects, 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 glycosylation tag fragments of a ComP protein. In certain aspects, the bioconjugate is a conjugate vaccine. In certain aspects, the conjugate vaccine is a vaccine against *Streptococcus pneumoniae* serotype 8. In certain aspects, the conjugate vaccine is administered to a subject, it induces an immune response. In certain aspects, the immune response elicits long term memory (memory B and T cells), is an antibody response. In certain aspects, the antibody response is a serotype-specific antibody response. In certain aspects,

- 7 -

the antibody response is an IgG or IgM response. In certain aspects, the antibody response is an IgG response. In certain aspects, the IgG response is an IgG1 response. In certain aspects, the conjugate vaccine generates immunological memory in a subject administered the vaccine.

**[0010]** Provided herein is a ComP glycosylation tag, for example certain aspects and features of which are described in this paragraph, 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<sub>ADP1</sub>: AAC45886.1). In certain aspects, the fragment comprises at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, or 40 amino acids of the ComP protein. In certain aspects, the ComP protein is a ComP protein as described above. In certain aspects, the glycosylation tag is attached to a heterologous carrier protein. In certain aspects, the heterologous carrier protein is 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. In certain aspects, the glycosylation tag is a ComPΔ28 polypeptide lacking amino acid residues corresponding to amino acid residues 1 to 28 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the ComPΔ28 polypeptide is selected from the group consisting of SEQ ID NO: 7 (ComPΔ28<sub>ADP1</sub>), SEQ ID NO: 8 (ComPΔ28<sub>110264</sub>), SEQ ID NO: 9 (ComPΔ28<sub>GFJ-2</sub>), SEQ ID NO: 10 (ComPΔ28<sub>P50v1</sub>), SEQ ID NO: 11 (ComPΔ28<sub>4466</sub>), and SEQ ID NO: 12 (ComPΔ28<sub>SFC</sub>). In certain aspects, the glycosylation tag comprises a region corresponding to the region of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) comprising the serine residue at position 82 of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) flanked by a disulfide bond connecting the alpha beta loop to the beta strand region. In certain aspects, the glycosylation tag comprises an amino acid sequence that is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to an amino acid selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), wherein said glycosylation tag comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag comprises an amino acid sequence selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID

- 8 -

NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag comprises the amino acid consensus sequence of SEQ ID NO: 37, or a fragment of at least 5, 10, 15, 20, 30, 35, or 40 consecutive amino acids thereof, wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag comprises the amino acid consensus sequence of SEQ ID NO: 38 or 45, or a fragment of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive amino acids thereof, wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag comprises a ComP protein amino acid sequence that corresponds to any of amino acid residues 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, or 82 of SEQ ID NO: 1 (ComPADP1: AAC45886.1) to any of amino acid residues 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, or 147 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag comprises an amino acid sequence comprising any of amino acid residues 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, or 82 of SEQ ID NO: 1 (ComPADP1: AAC45886.1) to any of amino acid residues 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106,

- 9 -

107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, or 147 of SEQ ID NO: 1 (Comp<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag is not more than 124, 120, 115, 110, 100, 90, 80, 75, 70, 60, 50, 40, 30, 25, 20, 15, 10, or 5 amino acids in length. In certain aspects, the glycosylation tag is covalently linked to an oligo- or polysaccharide at a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (Comp<sub>ADP1</sub>: AAC45886.1). In certain aspects, the oligo- or polysaccharide is produced by a bacteria from the genus *Streptococcus*. In certain aspects, the polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide. In certain aspects, the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b. In certain aspects, the capsular polysaccharide is CPS8. In certain aspects, the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*. In certain aspects, the oligo- or polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varriicola*, *Klebsiella michinganensis*, or *Klebsiella oxytoca* capsular polysaccharide. In certain aspects, the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide. In certain aspects, the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*. In certain aspects, the oligo- or polysaccharide comprises a glucose at its reducing end.

**[0011]** Provided for herein is a fusion protein, for example certain aspects and features of which are described in this paragraph, comprising a ComP glycosylation tag as described above. In certain aspects, 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<sub>ADP1</sub>: AAC45886.1). In certain aspects, 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. In certain aspects, the fusion protein comprises an amino acid linker sequence.

**[0012]** Provided herein is a method of *in vivo* conjugation of an oligo- or polysaccharide to an acceptor polypeptide, for example certain aspects and features of which are described in this paragraph, the method comprising covalently linking the oligo- or polysaccharide to the acceptor polypeptide with a PglS oligosaccharyltransferase (OTase), wherein the acceptor polypeptide comprises a ComP protein or a glycosylation tag fragment thereof. In certain aspects, the ComP protein or glycosylation tag fragment thereof is linked to a heterologous carrier protein. In certain

- 10 -

aspects, the PglS OTase is PglS<sub>110264</sub>, PglS<sub>ADP1</sub>, PglS<sub>GFJ-2</sub>, PglS<sub>50v1</sub>, PglS<sub>4466</sub>, or PglS<sub>SFC</sub>. In certain aspects, the ComP protein is ComP<sub>110264</sub>, ComP<sub>ADP1</sub>, ComP<sub>GFJ-2</sub>, ComP<sub>50v1</sub>, ComP<sub>4466</sub>, or ComP<sub>SFC</sub>. In certain aspects, the PglS OTase is PglS<sub>ADP1</sub>. In certain aspects, the PglS OTase is PglS<sub>ADP1</sub> but the ComP protein is not ComP<sub>ADP1</sub>. In certain aspects, the PglS OTase is PglS<sub>ADP1</sub> and the ComP protein is ComP<sub>110264</sub>. In certain aspects, the oligo- or polysaccharide is linked to the ComP protein or glycosylation tag fragment thereof at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the *in vivo* conjugation occurs in a host cell. In certain aspects, the host cell is a bacterial cell. In certain aspects, the host cell is *E.coli*. In certain aspects, 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. In certain aspects, 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 aspects, the carrier protein is 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. In certain aspects, the method produces a conjugate vaccine.

**[0013]** Provided for herein is a host cell, for example certain aspects and features of which are described in this paragraph, 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 a ComP protein or a glycosylation tag fragment thereof. In certain aspects, the acceptor polypeptide is a fusion protein. In certain aspects, the host cell comprises a nucleic acid encoding the PglS OTase. In certain aspects, the host cell comprises a nucleic acid encoding the acceptor polypeptide.

**[0014]** Provided for herein is an isolated nucleic acid, for example certain aspects and features of which are described in this paragraph, encoding a Comp glycosylation tag described above and/or a fusion protein described above. In certain aspects, the nucleic acid is a vector. Also provided for is a host cell comprising the isolated nucleic acid.

**[0015]** Provided for herein is a composition comprising a conjugate vaccine described above or a fusion protein described above, and an adjuvant.

[0016] Provided for herein is a method of inducing a host immune response against a bacterial pathogen, for example certain aspects and features of which are described in this paragraph, the method comprising administering to a subject in need of the immune response an effective amount of a conjugate vaccine described above, the fusion protein described above, or the composition described above. In certain aspects, the immune response is an antibody response. In certain aspects, 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. In certain aspects, 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.

[0017] Provided for herein is a method of preventing or treating a bacterial disease and/or infection in a subject, for example certain aspects and features of which are described in this paragraph, comprising administering to a subject in need thereof a conjugate vaccine described above, a fusion protein described above, or a composition described above. In certain aspects, the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature. In certain aspects, the disease is pneumonia. In certain aspects, the infection is a systemic infection and/or an infection of the blood. In certain aspects, the subject is a human. In certain aspects, the composition is administered via intramuscular injection, intradermal injection, intraperitoneal injection, subcutaneous injection, intravenous injection, oral administration, mucosal administration, intranasal administration, or pulmonary administration.

[0018] Provided for herein is a method of producing a pneumococcal conjugate vaccine against pneumococcal infection, the method comprising: (a) isolating a bioconjugate described above or a glycosylated fusion protein described above; and (b) combining the isolated conjugate vaccine or isolated glycosylated fusion protein with an adjuvant.

## **BRIEF DESCRIPTION OF THE DRAWINGS**

[0019] **Figure 1A-C.** **Figure 1A, 1B, and 1C** show that PglS (**1C**), but not PglB (**1B**) or PglL (**1A**), 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.

[0020] **Figure 2.** **Figure 2** shows that PglS from *A. baylyi* ADP1 (PglS<sub>ADP1</sub>) can transfer multiple pneumococcal capsular polysaccharides to ComP from *A. baylyi* ADP1 (ComP<sub>ADP1</sub>). Western blot analysis on purified ComP<sub>ADP1</sub> variants probing for hexa-histidine tagged ComP<sub>ADP1</sub> 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<sub>ADP1</sub> was glycosylated with the correct pneumococcal polysaccharide. The asterisk indicates samples that were treated with proteinase K for 2 hours.

[0021] **Figure 3A,B.** **Figure 3A and 3B** show that PglS<sub>ADP1</sub> can transfer the K1 and K2 capsular polysaccharides of *K. pneumoniae* to ComP<sub>ADP1</sub>. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged ComP<sub>ADP1</sub> variants and RNA polymerase. RNA polymerase was used as a loading control.

[0022] **Figure 4A.** **Figure 4A** shows mass spectrometry of CPS14-ComP<sub>ADP1</sub> identified a single glycosylated peptide. ISASNATTNVATAT (SEQ ID NO: 22).

[0023] **Figure 4B.** **Figure 4B** shows mass spectrometry of CPS14-ComP<sub>ADP1</sub> identified a single glycosylated peptide.

[0024] **Figure 5.** **Figure 5** shows Serine 84 of ComP<sub>ADP1</sub> is the site of PglS dependent glycosylation. Western blot analysis on *E. coli* whole cell lysates probing for hexa-histidine tagged ComP<sub>ADP1</sub> variants and the *Campylobacter jejuni* heptasaccharide. The ComP[S84A]<sub>ADP1</sub> variant was expressed; however, was not glycosylated as indicated by the absence of any reactive bands probing with the anti-hR6 heptasaccharide antisera.

[0025] **Figure 6.** **Figure 6** lists ComP ortholog amino acid sequences. The site of predicted glycosylation is **bolded**, flanked by a predicted disulfide bond (underlined) linking the predicted alpha beta loop to the beta strand region.

[0026] **Figure 7.** **Figure 7** shows that PglS<sub>ADP1</sub>, but not PglS<sub>110264</sub>, efficiently glycosylates both its cognate ComP<sub>ADP1</sub> as well as ComP<sub>110264</sub> 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.

[0027] **Figure 8.** **Figure 8** shows that PglS<sub>ADP1</sub> efficiently glycosylates DsbA-ComP $\Delta$ 28<sub>110264</sub> fusions but not DsbA-ComP $\Delta$ 28<sub>ADP1</sub> 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 $\Delta$ 28<sub>110264</sub> or ComP $\Delta$ 28<sub>ADP1</sub>.

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.

**[0028] Figure 9.** **Figure 9** shows that PglS<sub>ADP1</sub> efficiently glycosylates MBP-ComP $\Delta$ 28<sub>110264</sub> fusions but not MBP-ComP $\Delta$ 28<sub>ADP1</sub> 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 $\Delta$ 28<sub>110264</sub> or ComP $\Delta$ 28<sub>ADP1</sub>. 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.

**[0029] Figure 10.** **Figure 10** PglS<sub>ADP1</sub>, but not PglS<sub>110264</sub>, efficiently glycosylates EPA-GGGS-ComP $\Delta$ 28<sub>110264</sub> 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<sub>110264</sub> variant.

**[0030] Figure 11.** **Figure 11** shows amino acid sequences of representative ComP $\Delta$ 28<sub>110264</sub> fusion proteins.

**[0031] Figure 12A-C.** **Figure 12A, 12B, and 12C** show that a monovalent CPS14-ComP<sub>ADP1</sub> bioconjugate vaccine induces serotype specific IgG antibodies.

**[0032] Figure 13.** **Figure 13** shows that a trivalent bioconjugate vaccine against serotypes 8, 9V, and 14 induces serotype specific IgG titers at comparable levels to Prevnar 13.

**[0033] Figure 14.** **Figure 14** 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<sub>ADP1</sub>: AAC45886.1) have been removed. The site of predicted glycosylation is **bolded**, flanked by a predicted disulfide bond (underlined) linking the predicted alpha beta loop to the beta strand region.

**[0034] Figure 15.** **Figure 15** 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<sub>ADP1</sub>: AAC45886.1).

**[0035] Figure 16.** **Figure 16** 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.

**[0036] Figure 17.** Figure 17 shows higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested ComP glycosylated with the *C. jejuni* heptasaccharide (ComP-Glycan<sub>Cj</sub>). GluC digested ComP-Glycan<sub>Cj</sub> 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.

**[0037] Figure 18.** Figure 18 shows higher energy collisional dissociation (HCD) fragmentation spectra of GluC digested ComP glycosylated with the *C. jejuni* heptasaccharide (ComP-Glycan<sub>Cj</sub>). GluC digested ComP-Glycan<sub>Cj</sub> 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.

**[0038] Figure 19A-I.** Figure 19A-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.

**[0039] Figure 20A,B.** Figure 20A,B 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 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.

**[0040] Figure 21A-F.** **Figure 21A-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 CSP8 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.

**[0041] Figure 22A-F.** **Figure 22A-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 were 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.

**[0042] Figure 22G-L.** **Figure 22G-L** shows 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 biconjugate vaccine. Sera were 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.

**[0043] Figure 23A,B.** **Figure 23A,B** 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.

**[0044] Figure 24A,B.** **Figure 24A,B** 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.

**[0045] Figure 24C.** **Figure 24C** 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<sup>numeric</sup>” 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.

**[0046] Figure 24D.** **Figure 24D** provides a zoomed in view of the varying EPA-CPS8 glycoforms from **Figure 24C**.

**[0047] Figure 25A,B.** **Figure 25A,B** 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.

**[0048] Figure 26A,B,C.** **Figure 26A,B,C** 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<sub>ADP1</sub> and the homologous serines 79 and 82 of ComP<sub>110264</sub> 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.

**[0049] Figure 27A,B,C.** **Figure 27A,B,C** 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<sup>0</sup>” denotes the non-glycosylated EPA fusion and “g<sup>n</sup>” denotes the EPA fusion glycosylated with different sized K1 or K2 repeating subunits as depicted in panel B or C, respectively.

**[0050] Figure 28A,B.** **Figure 28A,B** 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<sup>numeric</sup>” 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

**[0051]** 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.

**[0052]** 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.

**[0053]** 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.

**[0054]** 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).

**[0055]** 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.

**[0056]** 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. 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.

**[0057]** 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.

**[0058]** As used herein, the term "non-naturally occurring" condition, substance, polypeptide, polynucleotide, composition, entity, organism, individual, and/or any combination thereof, or any grammatical variants thereof, is a conditional term that explicitly excludes, but only excludes, those forms 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."

**[0059]** As used herein, the term "bioconjugate" is a molecule comprising a peptide, oligopeptide, polypeptide, etc. covalently linked to a sugar (saccharide, oligosaccharide, polysaccharide, etc.).

**[0060]** As used herein, an "oligo- or polysaccharide" refers to a carbohydrate consisting of more than one monosaccharide unit bonded together.

**[0061]** As used herein, the term "lipid-linked oligo- or polysaccharide" refers to any isoprenoid moiety linked by a pyrophosphate to an oligo- or polysaccharide.

**[0062]** As used herein, the term "fusion protein" refers to a polypeptide comprising two or more amino acid sequences that are covalently linked in an arrangement that they would not naturally occur. Such amino acid sequences can include sequences from heterologous proteins, from different regions of the same protein, and/or repeated sequences from the same protein.

**[0063]** As used herein, the term “identity” or “identical” e.g., “percent (%) identity” or “percent (%) identical” to an amino acid sequence or to a nucleotide sequence disclosed herein refers to a relationship between two or more nucleotide sequences or between two or more amino acid sequences. When a position in one sequence is occupied by the same nucleic acid base or amino acid in the corresponding position of the comparator sequence, the sequences are said to be “identical” at that position. The percentage “sequence identity” is calculated by determining the number of positions at which the identical nucleic acid base or amino acid occurs in both sequences to yield the number of “identical” positions. The number of “identical” positions is then divided by the total number of positions in the comparison window and multiplied by 100 to yield the percentage of “sequence identity.” Percentage of “sequence identity” is determined by comparing two optimally aligned sequences over a comparison window of the entire length of a reference sequence. In order to optimally align sequences for comparison, the portion of a nucleotide or amino acid sequence in the comparison window can comprise additions or deletions termed gaps while the reference sequence is kept constant. An optimal alignment is that alignment which, even with gaps, produces the greatest possible number of “identical” positions between the reference and comparator sequences. Percentage “sequence identity” between two sequences can be determined using, e.g., the program “BLAST” which is available from the National Center for Biotechnology Information, and which program incorporates the programs BLASTN (for nucleotide sequence comparison) and BLASTP (for amino acid sequence comparison), which programs are based on the algorithm of Karlin and Altschul (*Proc. Natl. Acad. Sci. USA* 90(12):5873-5877, 1993).

**[0064]** 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.

While a polypeptide can be derived from a natural biological source or produced by recombinant technology, is not necessarily translated from a designated nucleic acid sequence. It can be generated in any manner, including by chemical synthesis.

**[0065]** 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.

**[0066]** By an “isolated” polypeptide or a fragment, variant, or derivative thereof or the like 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. An isolated polypeptide or fragment, variant, or derivative thereof or the like can be associated, bound, etc., with a cofactor. Likewise, a purified or purified and isolated polypeptide or fragment, variant, or derivative thereof or the like can be associated, bound, etc., with a cofactor.

**[0067]** Other polypeptides disclosed herein are fragments, derivatives, analogs, or variants of the polypeptides disclosed herein, and any combination thereof. The terms “fragment,” “variant,” “derivative” and “analog” when referring to polypeptide subunit or multimeric protein as disclosed herein can include any polypeptide or protein that retain at least some of the activities of the complete polypeptide or protein (for example retain the ability to be glycosylated), but which is structurally different. Fragments of polypeptides include, for example, proteolytic fragments, as well as deletion fragments. Variants include fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants can occur spontaneously or be intentionally constructed. Intentionally constructed variants can be produced using art-known mutagenesis techniques. Variant polypeptides can comprise conservative or non-conservative amino acid substitutions, deletions or additions (one of ordinary skill in the art would understand that a “conservative amino acid substitution” is not the same as a “conserved amino acid residue/position”). Variant polypeptides can also be referred to herein as “polypeptide analogs.” Derivatives are variants of polypeptides that have been altered so as to exhibit additional features not found on the native polypeptide. Examples include fusion proteins. As used herein a “derivative” also refers to a subject polypeptide having one or more amino acids chemically derivatized by reaction of a functional

side group. Also included as "derivatives" are those peptides that contain one or more standard or synthetic amino acid derivatives of the twenty standard amino acids. For example, 4-hydroxyproline can be substituted for proline; 5-hydroxylysine can be substituted for lysine; 3-methylhistidine can be substituted for histidine; homoserine can be substituted for serine; and ornithine can be substituted for lysine.

**[0068]** As used herein, a "single amino acid substitution" means replacing an amino acid residue in a polypeptide sequence with a different amino acid residue (such as replacing the native residue in a wild-type sequence with a non-native amino acid), unless otherwise specified. Also encompassed by the disclosure are a "single amino acid deletion" and/or a "single amino acid insertion."

**[0069]** A "conservative amino acid substitution" is one in which one amino acid is replaced with another amino acid having a similar side chain. Families of amino acids having similar side chains have been defined in the art, including basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). For example, substitution of a phenylalanine for a tyrosine is a conservative substitution. Methods of identifying nucleotide and amino acid conservative substitutions which do not eliminate protein activity or functionality are well-known in the art (see, *e.g.*, Brummell et al. *Biochem.* 32: 1180-1 187 (1993); Kobayashi et al. *Protein Eng.* 12(10):879-884 (1999); Burks et al. *Proc. Natl. Acad. Sci. USA* 94:412-417 (1997)).

**[0070]** 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.

**[0071]** As used herein, a "coding region" is a portion of nucleic acid comprising codons translated into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is not translated into an amino acid, it can be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, and the like, are not part of a coding region. Two or more coding regions can be present in a single polynucleotide construct, *e.g.*, on a single vector, or in separate polynucleotide constructs, *e.g.*, on separate (different) vectors. Furthermore, any vector can contain a single coding region, or can comprise two or more coding regions, *e.g.*, a single vector can separately encode a selection marker gene and a gene of interest. In addition, a vector, polynucleotide, or nucleic acid can encode heterologous coding regions, either fused or unfused to a nucleic acid encoding a polypeptide subunit or fusion protein as provided herein. Heterologous coding regions include without limitation specialized elements or motifs, such as a secretory signal peptide or a heterologous functional domain.

**[0072]** In certain aspects, the polynucleotide or nucleic acid is DNA. In the case of DNA, a polynucleotide comprising a nucleic acid that encodes a polypeptide normally can include a promoter and/or other transcription or translation regulatory elements operably associated with one or more coding regions. An operable association or linkage can be when a coding region for a gene product, *e.g.*, a polypeptide, can be associated with one or more regulatory sequences in such a way as to place expression of the gene product under the influence or control of the regulatory sequence(s). Two DNA fragments (such as a polypeptide coding region and a promoter associated therewith) can be "operably associated" or "operably linked" if induction of promoter function results in the transcription of mRNA encoding the desired gene product and if the nature of the linkage between the two DNA fragments does not interfere with the ability of the expression regulatory sequences to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be operably associated with a nucleic acid encoding a polypeptide if the promoter was capable of effecting transcription of that nucleic acid. The promoter can be a cell-specific promoter that directs

substantial transcription of the DNA only in predetermined cells. Other transcription regulatory elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can be operably associated with the polynucleotide to direct cell-specific transcription.

**[0073]** A variety of transcription regulatory regions are known to those skilled in the art. These include, without limitation, transcription regulatory regions that function in vertebrate cells, such as, but not limited to, promoter and enhancer segments from cytomegaloviruses (the immediate early promoter, in conjunction with intron-A), simian virus 40 (the early promoter), and retroviruses (such as Rous sarcoma virus). Other transcription regulatory regions include those derived from vertebrate genes such as actin, heat shock protein, bovine growth hormone and rabbit  $\beta$ -globin, as well as other sequences capable of controlling gene expression in eukaryotic cells. Additional suitable transcription regulatory regions include tissue-specific promoters and enhancers.

**[0074]** Similarly, a variety of translation regulatory elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, and elements derived from picornaviruses (particularly an internal ribosome entry site, or IRES, also referred to as a CITE sequence).

**[0075]** In other aspects, a polynucleotide can be RNA, for example, in the form of messenger RNA (mRNA).

**[0076]** Polynucleotide and nucleic acid coding regions can be associated with additional coding regions that encode secretory or signal peptides, which direct the secretion of a polypeptide encoded by a polynucleotide as disclosed herein.

**[0077]** 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. Illustrative types of vectors include plasmids, phages, viruses and retroviruses.

**[0078]** 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 prokaryotic cell (e.g., bacterial) or a eukaryotic cell (e.g., mammalian).

**[0079]** 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.

**[0080]** 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.

**[0081] Overview.** 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.

**[0082]** 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.

**[0083]** 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.

**[0084]** The oligosaccharyltransferase PglS – previously referred to as PglL by Schulz *et al.* (PMID23658772) and PglL<sub>ComP</sub> 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)).

**[0085]** ComP is orthologous to type IV pilin proteins, like PilA from *Pseudomonas aeruginosa* and Pile from *Neisseria meningitidis*, 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. meningitidis*. Comprehensive glycopeptide analysis, however, revealed this serine and the surrounding residues were not the site of glycosylation in ComP. Here it is disclosed that PglS glycosylates ComP at a single serine residue located at position corresponding to the conserved serine at position 84 of ComP<sub>ADP1</sub>: AAC4588631 (SEQ ID NO: 1) (also corresponding to the conserved serine at position 82 of ComP<sub>110264</sub>: 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.

**[0086] 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. meningitidis*, 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 1A** and **Figure 1B**, both acceptor proteins were expressed; however, no evidence for CPS14 glycosylation to either acceptor protein was observed.

**[0087]** *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<sup>28</sup>. 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 1C**, 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 1B**). 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.

**[0088]** 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](http://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 2**, 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.

**[0089]** 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.

**[0090]** **PglS can transfer capsular polysaccharides of *Klebsiella* to ComP.** *Klebsiella pneumoniae* (*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 3B**). 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. pneumoniae*. Increased efficiency of conjugation was observed with co-expression of the transcriptional activator *rmpA* from *K. pneumoniae*.

**[0091]** 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. varriicola*, *K. michiganensis*, and *K. oxytoca*.

**[0092]** 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<sub>ADP1</sub>. *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 4A** and **Figure 4B**, 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 HexNA<sub>2</sub>Hexose<sub>6</sub> (**Figure 4B**). Additional glycopeptides were also observed decorated with extended glycans corresponding to up to four tetrasaccharide repeat units (**Figure 16**).

**[0093]** 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 5**, 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 17 and Figure 18**). As a negative control, a catalytically inactive PglS mutant (H324A) was generated, that when co-expressed with the *C. jejuni* heptasaccharide 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.

**[0094] 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<sub>ADP1</sub>) 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<sub>ADP1</sub> 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<sub>ADP1</sub> against all proteins from bacteria in the *Acinetobacter* genus. Expectedly, many *Acinetobacter* type IVa pilin orthologs, including ComP<sub>ADP1</sub>, 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 6**) were identified based on the presence of the conserved serine at position 84 relative to ComP<sub>ADP1</sub> 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.

**[0095]** 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.

**[0096]** In certain aspects disclosed herein, 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 (ComPADP1: AAC45886.1), SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1), SEQ ID NO: 3 (ComP<sub>GFJ-2</sub>: APV36638.1), SEQ ID NO: 4 (ComP<sub>P50v1</sub>: PKD82822.1), SEQ ID NO: 5 (ComP<sub>4466</sub>: SNX44537.1), or SEQ ID NO: 6 (ComP<sub>SFC</sub>: OAL75955.1), and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects disclosed herein, 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 (ComPADP1: AAC45886.1), and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain of these aspects, the ComP protein comprises a region corresponding to the region of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) comprising the serine residue at position 82 of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1) flanked by a disulfide bond connecting the alpha beta loop to the beta strand region. In certain aspects, the ComP protein comprises the consensus sequence of SEQ ID NO: 37, SEQ ID NO: 38, or SEQ ID NO: 45 (Table 3 below). In certain aspects, the ComP protein comprises a region having the amino acid sequence of ADP1 VGVQEISASNATTNVATAT (SEQ ID NO: 39), 110264 TGVQTIASGASAATTNVAQAQ (SEQ ID NO: 40), GFJ-2 VGVQEINASSSTSNVATAT (SEQ ID NO: 41), SFC AGVETIGASNKTKNVESAA (SEQ ID NO: 42), P50v1 VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and 4466 NGVISASATTNVASSA (SEQ ID NO: 44), or variant of SEQ ID NO: 39, 40, 41, 42, 43, or 44 having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or

- 34 -

deletions, wherein the variant sequence maintains the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, amino acid substitutions are conservative amino acid substitutions. It is also evident to one of ordinary skill in the art that in any aspect disclosed herein, the ComP protein is capable of being glycosylated on the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1).

**[0097] 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<sub>110264</sub>) is 71% identical at the amino acid level when compared to the ComPADP1. However, consistent with the features above, ComP<sub>110264</sub> 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 ComPADP1. Moreover, a PglS ortholog can be found immediately downstream of ComP<sub>110264</sub>. To determine whether PglS from *A. baylyi* ADP1 (PglS<sub>ADP1</sub>) could glycosylate ComP<sub>110264</sub>, PglS<sub>ADP1</sub> was cloned into pACT3 and ComP<sub>110264</sub> 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<sub>110264</sub>) with ComPADP1. As seen in **Figure 7**, PglS<sub>110264</sub> minimally glycosylated its cognate acceptor pilin ComP<sub>110264</sub> as indicated by higher molecular weight ComP pilin variants when compared to whole cell lysates lacking PglS<sub>110264</sub>. Based on western blot analysis, PglS<sub>110264</sub> appears to not glycosylate ComPADP1. On the other hand, PglS<sub>ADP1</sub> efficiently glycosylated both ComPADP1 and ComP<sub>110264</sub> as indicated by the robust increase of His-reactive signals of increasing electrophoretic mobility. Collectively, PglS<sub>ADP1</sub> 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.

**[0098] 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<sub>ADP1</sub> and ComP<sub>110264</sub> 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*<sub>110264</sub> or *pglS*<sub>ADP1</sub>. As seen in **Figure 8**, *E. coli* cells expressing either DsbA-AAA-ComP $\Delta$ 28<sub>110264</sub> or DsbA-GGGS-ComP $\Delta$ 28<sub>110264</sub> in combination with PglS<sub>ADP1</sub> 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 $\Delta$ 28<sub>ADP1</sub> 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 9**, *E. coli* cells expressing either MBP-AAA-ComP $\Delta$ 28<sub>110264</sub> or MBP-GGGS-ComP $\Delta$ 28<sub>110264</sub> in combination with PglS<sub>ADP1</sub> demonstrated detectable levels of glycosylation as indicated by the modal distribution of anti-His reactive signals; whereas, fusions with ComP $\Delta$ 28<sub>ADP1</sub> 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 ComP $\Delta$ 28<sub>110264</sub> 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<sub>ADP1</sub> in both whole cell extracts and in periplasmic extracts. As seen in **Figure 10**, EPA-GGGS-ComP $\Delta$ 28<sub>110264</sub> constructs were found to be glycosylated in both the whole cell extract and periplasmic extracts of cells co-expressing the CPS8 glycan and PglS<sub>ADP1</sub> 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<sub>110264</sub>. Collectively, PglS<sub>ADP1</sub> 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 11**.

**[0099] 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 12A**, 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 12B**, IgG1 titers appeared to be higher than the other subtypes.

**[0100]** 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 13**). The monovalent immunization against serotype 14 also showed significant induction of serotype specific IgG titers, which were similar to the preliminary immunization (**Figure 12** and **Figure 13**). 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 13**).

**[0101] Bioconjugates.** Provide herein are bioconjugates comprising an oligo- or polysaccharide covalently linked to a fusion protein. In certain aspects, the fusion protein comprises a ComP protein (ComP). In certain aspects, the fusion protein comprises a glycosylation tag or a glycosylation tag fragment of a ComP protein (as described in detail elsewhere herein).

**[0102]** As described herein, it has been discovered that ComP is glycosylated on a serine (S) residue. This serine residue is conserved in ComP and corresponds to position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). This serinre residue also corresponds to position 82 of SEQ ID NO: 2 (ComP110264: ENV58402.1) (**Figure 26A, B, and C**). Thus, in certain aspects, a fusion protein is glycosylated with an oligo- or polysaccharide on a ComP protein or glycosylation tag

fragment thereof at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 (Comp<sub>ADP1</sub>: AAC45886.1). **Figure 15** 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<sub>ADP1</sub>: AAC45886.1), which is conserved across the ComP sequences.

**[0103]** 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.

**[0104]** 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<sub>ADP1</sub>: AAC45886.1) and contains a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (Comp<sub>ADP1</sub>: 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 $\Delta$ 28<sub>ADP1</sub>), SEQ ID NO: 8 (Comp $\Delta$ 28<sub>110264</sub>), SEQ ID NO: 9 (Comp $\Delta$ 28<sub>GFJ-2</sub>), SEQ ID NO: 10 (Comp $\Delta$ 28<sub>P50v1</sub>), SEQ ID NO: 11 (Comp $\Delta$ 28<sub>4466</sub>), or SEQ ID NO: 12 (Comp $\Delta$ 28<sub>SFC</sub>) 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<sub>ADP1</sub>: 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 $\Delta$ 28<sub>ADP1</sub>) 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<sub>ADP1</sub>: AAC45886.1). In certain aspects, the ComP protein comprises SEQ ID NO: 7 (ComP $\Delta$ 28<sub>ADP1</sub>), SEQ ID NO: 8 (ComP $\Delta$ 28<sub>110264</sub>), SEQ ID NO: 9 (ComP $\Delta$ 28<sub>GFJ-2</sub>), SEQ ID NO: 10 (ComP $\Delta$ 28<sub>P50v1</sub>), SEQ ID NO: 11 (ComP $\Delta$ 28<sub>4466</sub>), or SEQ ID NO: 12 (ComP $\Delta$ 28<sub>SFC</sub>). In certain aspects, the ComP protein is SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1), SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1), SEQ ID NO: 3 (ComP<sub>GFJ-2</sub>: APV36638.1), SEQ ID NO: 4 (ComP<sub>50v1</sub>: PKD82822.1), SEQ ID NO: 5 (ComP<sub>4466</sub>: SNX44537.1), or SEQ ID NO: 6 (ComP<sub>SFC</sub>: OAL75955.1).

**[0105]** In certain aspects, the oligo- or polysaccharide is produced by a bacteria or a mammalian cell. In certain aspects, the bacteria is a Gram negative bacteria. In certain aspects, the bacteria is from the genus *Streptococcus*. In certain aspects, the bacteria is from the genus *Klebsiella*. In certain aspects, the oligo- or polysaccharide is a *S. pneumoniae*, *S. agalactiae*, or *S. suis* capsular polysaccharide, for example, wherein the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b of *S. pneumoniae*. In certain aspects, the oligo- or polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varriicola*, *Klebsiella michinganensis*, or *Klebsiella oxytoca* capsular polysaccharide. In certain aspects, the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide. For example, in certain aspects, the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*.

**[0106]** In certain aspects, the oligo- or polysaccharide comprises a glucose (Glc) at its reducing end, the significance of which is discussed elsewhere herein.

**[0107]** 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 is incorporated herein by reference.

**[0108]** As discussed elsewhere herein, in certain applications, it may be advantageous to form a fusion protein with a carrier protein or fragment thereof. In certain application, the carrier protein is one recognized in the art as useful in producing conjugate vaccines. In certain aspects, 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 aspects, 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 aspects, the carrier protein or fragment thereof is linked to the ComP protein or glycosylation tag fragment thereof via an amino acid linker, for example (GGGS)<sub>n</sub> (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 aspects, 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 glycosylation tag fragments of a ComP protein. In certain aspects, 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 glycosylation tag fragments of a ComP protein. In certain aspects, multiple glycosylation tag fragments are arranged in tandem to one another in the fusion protein. In certain aspects, multiple glycosylation tag fragments are arranged apart from one another in the fusion protein, for example separated by sequences of carrier protein. In certain aspects, the glycosylation tag fragment(s) can be, for example, located at the N-terminal end of the carrier protein and/or fusion protein. In certain aspects, the glycosylation tag fragment(s) can be, for example, located at the C-terminal end of the carrier protein and/or fusion protein. In certain aspects, the glycosylation tag fragment(s) can be located internally within the carrier protein and/or fusions protein, for example, wherein a glycosylation tag fragment is located between multiple carrier proteins in a fusion protein. In certain aspects, the multiple carrier proteins can be identical in type or different in type.

**[0109] Glycosylation tag fragment.** In certain applications, such as any of the aspects described herein, there may be advantages to using less than a whole length ComP protein, such as by removing the leader sequences or using an even smaller fragment of ComP that can still be glycosylated. Because the glycosylation site of ComP is disclosed herein as the serine at residue 84 of SEQ ID NO: 1, or a corresponding serine residue in other ComP sequences, fragments of ComP proteins can be identified comprising the ComP glycosylation site. As used herein, a fragment of a ComP protein that comprises a serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1, and that can be glycosylated by a PglS OTase

- 40 -

when incorporated into a fusion protein, is referred to herein as a glycosylation tag or glycosylation tag fragment of a ComP protein.

**[0110]** In certain aspects, a ComP glycosylation tag comprises an isolated fragment of ComP, wherein the fragment comprises at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, or 50 amino acids of a ComP protein and comprises a serine residue corresponding to serine residue 84 in SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, a glycosylation tag comprises a ComP protein amino acid sequence that corresponds to any of amino acid residues 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, or 82 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1) to any of amino acid residues 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, or 147 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, a glycosylation tag comprises a ComP protein amino acid sequence comprising any of amino acid residues 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, or 82 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1) to any of amino acid residues 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, or 147 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). In certain aspects, the glycosylation tag is not more than 124, 120, 119, 118, 117, 116, 115, 100, 90, 80, 75, 70, 60, 50, 40, 30, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, or 5 amino acids in length.

**[0111]** In certain aspects, the glycosylation tag fragment of a ComP protein is a ComP $\Delta$ 28 polypeptide lacking amino acid residues corresponding to amino acid residues 1 to 28 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1). For example, representative examples of ComP $\Delta$ 28 polypeptides include, but are not limited to, SEQ ID NOs: 7-12. In certain aspects, the glycosylation tag fragment of the ComP protein comprises a region corresponding to the region of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1) comprising the serine residue at position 84 flanked by a disulfide bond connecting predicted the alpha beta loop to the beta strand region. For

- 41 -

example, representative examples of a region corresponding to the region of SEQ ID NO: 1 (ComPADP1: AAC45886.1) comprising the serine residue at position 84 flanked by a disulfide bond connecting the predicted alpha beta loop to the beta strand region include, but are not limited to: ADP1 VGVQEISASNATTNVATAT (SEQ ID NO: 39), 110264 TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), GFJ-2 VGVQEINASSSTSNVATAT (SEQ ID NO: 41), SFC AGVETIGASNKTKNVESAA (SEQ ID NO: 42), P50v1 VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and 4466 NGVISASATTNVASSA (SEQ ID NO: 44).

**[0112]** In certain aspects disclosed herein, the glycosylation tag fragment of ComP comprises an amino acid sequence that is at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to an amino acid selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), wherein the glycosylation tag comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag fragment of ComP comprises an amino acid sequence selected from the group consisting of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), VGVQTIAASNATKNVATAT (SEQ ID NO: 43), and NGVISASATTNVASSA (SEQ ID NO: 44), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, amino acid substitutions are conservative amino acid substitutions. In certain aspects, the glycosylation tag fragment of ComP comprises an amino acid sequence that is at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to VGVQEISASNATTNVATAT (SEQ ID NO: 39), wherein the glycosylation tag comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, the glycosylation tag fragment of ComP comprises the amino acid sequence VGVQEISASNATTNVATAT (SEQ ID NO: 39), 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 serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, amino acid substitutions are conservative amino acid substitutions.

**[0113]** In certain aspects, a glycosylation tag fragment of a ComP protein comprises an amino acid sequence of at least 5, 10, 15, 20, 30, 35, or 40 consecutive amino acids of the amino acid consensus sequence of SEQ ID NO: 37 (Table 3 below), wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, a glycosylation tag fragment of a ComP protein comprises the amino acid consensus sequence of SEQ ID NO: 37, or variant of SEQ ID NO: 37 having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions, wherein the variant sequence maintains the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, amino acid substitutions are conservative amino acid substitutions.

**[0114]** In certain aspects, a glycosylation tag fragment of a ComP protein comprises an amino acid sequence of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive amino acids of the amino acid consensus sequence of SEQ ID NO: 38 or 45 (Table 3 below), wherein said glycosylation tag fragment comprises the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, a glycosylation tag fragment of a ComP protein comprises the amino acid consensus sequence of SEQ ID NO: 38 or 45, or variant of SEQ ID NO: 38 or 45 having one, two, three, four, five, six, or seven amino acid substitutions, additions, and/or deletions, wherein the variant sequence maintains the serine residue corresponding to the conserved serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1). In certain aspects, amino acid substitutions are conservative amino acid substitutions.

**[0115]** In certain aspects, the glycosylation tag is attached to a heterologous protein such as a carrier protein. Thus, certain aspects provide for a fusion protein comprising a ComP glycosylation tag disclosed herein. In certain aspects, the fusion protein comprises a carrier protein, representative examples of which include but are not limited to diphtheria toxoid CRM197, tetanus toxoid, *Pseudomonas aeruginosa* Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, *Haemophilus influenzae* protein D, or a fragment thereof. In certain aspects, the fusion protein comprises a linker sequence as disclosed elsewhere herein. In certain

aspects, the fusion protein is glycosylated and further in certain aspects, the fusion protein is glycosylated on a glycosylation tag region at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 (ComP<sub>ADP1</sub>: AAC45886.1).

**[0116] Conjugate vaccines.** Disclosed herein is a pneumococcal bioconjugate vaccine containing a conventional vaccine carrier. Certain aspects comprise the use of a ComP fragment as a glycosylation tag (aka “glycotag”). In certain aspects, the glycotag can be added to the C-terminus and/or N-terminus of a carrier protein. For example, in certain aspects, the glycotag is added to the C-terminus of the conventional carrier protein *Pseudomonas aeruginosa* Exotoxin A (EPA). It has been demonstrated that in certain aspects, the glycotag/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 aspects, 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 aspects provide for a CPS8 pneumococcal bioconjugate vaccine.

**[0117]** 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.

**[0118]** In certain aspects, 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 aspects, the conjugate vaccine is a prophylaxis that can be used, e.g., to immunize a subject against an infection and/or disease. In certain aspects, the bioconjugate is associated with (such as in a therapeutic composition) and/or administered with an adjuvant. Certain aspects provide for a composition (such as a therapeutic composition) comprising a conjugate vaccine described herein and an adjuvant. In certain aspects, when the conjugate vaccine is administered to a subject, it induces an immune response. In certain aspects, the immune response elicits long term memory (memory B and T cells). In certain aspects, the immune is an antibody response. In certain aspects, the antibody response is a serotype-specific antibody response. In certain aspects, the antibody response is an IgG or IgM response. In certain

aspects where the antibody response is an IgG response, the IgG response is an IgG1 response. Further, in certain aspects, the conjugate vaccine generates immunological memory in a subject administered the vaccine.

**[0119]** Certain aspects provide for producing a vaccine against an infection and/or disease. In certain aspects the method comprises isolating a bioconjugate or fusion protein disclosed herein (conjugate vaccine) and combining the conjugate vaccine with an adjuvant. In certain aspects, the vaccine is a conjugate vaccine against pneumococcal infection. In certain aspects, the disease is pneumonia.

**[0120]** 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.

**[0121] Methods and reagents for *in vivo* glycosylation.** Disclosed herein are methods for the *in vivo* conjugation of an oligo- or polysaccharide to a polypeptide (*in vivo* glycosylation). In certain aspects, the method comprises covalently linking the oligo- or polysaccharide to the polypeptide with a PglS oligosaccharyltransferase (OTase) (described elsewhere herein). In certain aspects, the polypeptide comprises a ComP protein or a glycosylation tag fragment thereof. In certain aspects, the polypeptide comprises a ComP protein or a glycosylation tag fragment thereof linked to a heterologous polypeptide such as a carrier protein. Representative examples of PglS OTases include, but are not limited to PglS<sub>110264</sub>, PglS<sub>ADP1</sub>, PglS<sub>GFJ-2</sub>, PglS<sub>50v1</sub>, PglS<sub>4466</sub>, and PglS<sub>SFC</sub>. ComP proteins are described in detail elsewhere and representative examples include, but are not limited to ComP<sub>110264</sub>, ComP<sub>ADP1</sub>, ComP<sub>GFJ-2</sub>, ComP<sub>50v1</sub>, ComP<sub>4466</sub>, and ComP<sub>SFC</sub>. It will be recognized that while a PglS OTase from an organism would naturally glycosylate the ComP protein from that organism (e.g., PglS<sub>110264</sub> glycosylates ComP<sub>110264</sub>) in certain aspects, a PglS from one organism glycosylates a ComP from a different organism (e.g., PglS<sub>ADP1</sub> glycosylates ComP<sub>110264</sub>). For example, in certain aspects, the PglS OTase is PglS<sub>ADP1</sub>. In certain aspects, where the PglS OTase is PglS<sub>ADP1</sub>, the ComP protein glycosylated is not ComP<sub>ADP1</sub>. For example, in certain aspects where the PglS OTase is PglS<sub>ADP1</sub>, the ComP protein is ComP<sub>110264</sub>. 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.

[0122] In certain aspects 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<sub>ADP1</sub>: AAC45886.1).

[0123] In certain aspects disclosed herein, the *in vivo* glycosylation occurs in a host cell. In certain aspects, 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 aspects, the host cell is a bacterial cell, for example, *E. coli*.

[0124] In certain aspects, 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 aspects, 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 aspects, 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 aspects, the method further comprises expressing and/or providing such a transcriptional activator in the host cell along with the other components.

[0125] In certain aspects, the carrier protein linked to the ComP protein or a glycosylation tag fragment thereof is, for example, diphtheria toxoid CRM197, tetanus toxoid, *Pseudomonas aeruginosa* Exotoxin A (EPA), tetanus toxin C fragment, cholera toxin B subunit, *Haemophilus influenzae* protein D, or a fragment thereof.

[0126] In certain aspects, the method produces a conjugate vaccine as described herein.

[0127] Certain aspects 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 aspects, 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 aspects, the

acceptor polypeptide is a fusion protein. In certain aspects, the host cell further comprises a transcriptional activator such as described above along with the other components.

**[0128]** In certain aspects, a host cell comprises an isolated nucleic acid encoding a PglS OTase. In certain aspects a host cell comprises an isolated nucleic acid encoding the ComP acceptor polypeptide. In certain aspects, a host cell comprises a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide. In certain aspects, 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 certain 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.

**[0129]** Certain aspects provide for an isolated nucleic acid encoding the ComP protein, ComP glycosylation tag fragment, and/or ComP fusion protein described anywhere herein. In certain aspects, an isolated nucleic acid referred to herein is a vector or is contained within a vector. In certain aspects, 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.

**[0130] Administration.** Provided herein are methods of inducing a host immune response against a pathogen. In certain aspects, the pathogen is a bacterial pathogen. In certain aspects, the host is immunized against the pathogen. In certain aspects, 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 aspects 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 aspects, the immune response is an antibody response. In certain aspects, 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.

[0131] 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 aspects, the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature. In certain aspects, the disease is pneumonia. In certain aspects, the infection is a systemic infection and/or an infection of the blood. In certain aspects disclosed herein, the subject is a vertebrate. In certain aspects 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 aspects the mammal is a human.

[0132] In any of the aspects of administration disclosed 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

[0133] **Bacterial strains, plasmids and growth conditions.** Strains and plasmids used in this work are listed in Table 1.

**Table 1.** Strains and plasmids employed in this study.

<b>Strains/ Plasmids</b>	<b>Description</b>
<b>Strains</b>	
<i>E. coli</i> SDB1	W3110, <i>A waaL</i> ligase, <i>AwecA</i> glycosyltransferase
<i>E. coli</i> DH5 $\alpha$	General cloning strain
<i>S. pneumoniae</i> serotype 8, 9V, and 14	Wild type pneumococci strains expressing either the serotype 8, 9V, or 14 capsular polysaccharides
<i>K. pneumoniae</i> serotype K1 and K2	Wild type <i>K. pneumoniae</i> strains expressing either the serotype K1 or K2 capsular polysaccharides
<b>Plasmids</b>	
pEXT20	Cloning vector, Amp <sup>R</sup> , IPTG inducible

pACT3	Cloning vector, Cm <sup>R</sup> , IPTG inducible
pMN1	C-6X His-tagged ComP cloned in <i>Bam</i> HI and <i>Sal</i> I sites of pEXT20, Amp <sup>R</sup> , IPTG inducible
pMN2	Non-coding region and PglS cloned in <i>Sal</i> I and <i>Pst</i> I sites of pMN1, Amp <sup>R</sup> , IPTG inducible
pMN4	PglS[H324A] in pMN2 background
pMN8	Non-coding region and PglS cloned in <i>Sal</i> I and <i>Pst</i> I sites of pEXT20, Amp <sup>R</sup> , IPTG inducible
pMN9	ComP[S82A] mutant of pMN2
pMN10	ComP[S84A] mutant of pMN2
pMAF10	HA-tagged PglB cloned in pMLBAD, Tp <sup>R</sup> , Arabinose inducible
pAMF10	C-10× His-tagged NmpglL cloned into pEXT20, Amp <sup>R</sup> , IPTG inducible
piH18	C-6X His-tagged AcrA from <i>C. jejuni</i> cloned into pEXT21, Sp <sup>R</sup> , IPTG inducible
pAMF22	C-6X His-tagged <i>dsbA1</i> from <i>N. meningitidis</i> MC58 cloned into pMLBAD, Tp <sup>R</sup> Arabinose inducible
pACYCpglBmut	pACYC184-based plasmid encoding the <i>C. jejuni</i> pgl locus with mutations W458A and D459A in PglB. Cm <sup>R</sup> , IPTG inducible.
pNLP80	<i>S. pneumoniae</i> CPS14 cluster on pWSK129, Kan <sup>R</sup>
pB-8	<i>S. pneumoniae</i> CPS8 cluster on pBBR1MCS-3, Tc <sup>R</sup>
pWKS130-9V	<i>S. pneumoniae</i> CPS9v cluster on pWKS130, Kan <sup>R</sup>
pBBR1MCS-K1	<i>K. pneumoniae</i> K1 cluster in pBBR1MCS2
pBBR1MCS-K2	<i>K. pneumoniae</i> K2 cluster in pBBR1MCS2
pACT3- <i>rmpA</i>	<i>rmpA</i> cloned into pACT3
pEXT20-ComP-PglS <sub>110264</sub>	ComP and PglS from <i>A. soli</i> CIP 110264 cloned into pEXT20
pEXT20-ComP <sub>110264</sub>	ComP <sub>110264</sub> with a c-terminal hexa-his tag from <i>A. soli</i> CIP 110264 cloned into pEXT20
pACT3-PglS <sub>110264</sub>	PglS <sub>110264</sub> w from <i>A. soli</i> CIP 110264 cloned into pEXT20
pACT3-PglS <sub>ADP1</sub>	PglS <sub>ADP1</sub> w from <i>A. baylyi</i> ADP1 cloned into pEXT20
pACT3- <i>rmpA</i> - <i>pglS</i> <sub>ADP1</sub>	<i>rmpA</i> and <i>pglS</i> <sub>ADP1</sub> cloned into pACT3
pEXT20-DsbA-AAA-ComP $\Delta$ 23 <sub>110264</sub>	DsbA fused to a triple alanine peptide linking ComP $\Delta$ 23 <sub>110264</sub> with a c-terminal hex-his tag
pEXT20-DsbA-GGGS-ComP $\Delta$ 23 <sub>110264</sub>	DsbA fused to a gly-gly-gly-ser peptide linking ComP $\Delta$ 23 <sub>110264</sub> with a c-terminal hex-his tag
pEXT20-DsbA-AAA-ComP $\Delta$ 23 <sub>ADP1</sub>	DsbA fused to a triple alanine peptide linking ComP $\Delta$ 23 <sub>ADP1</sub> with a c-terminal hex-his tag

pEXT20-DsbA-GGGS-ComP $\Delta$ 23 <sub>ADP1</sub>	DsbA fused to a gly-gly-gly-ser peptide linking ComP $\Delta$ 23 <sub>ADP1</sub> with a c-terminal hex-his tag
pEXT20-MBP-AAA-ComP $\Delta$ 23 <sub>110264</sub>	MBP fused to a triple alanine peptide linking ComP $\Delta$ 23 <sub>110264</sub> with a c-terminal hex-his tag
pEXT20-MBP-GGGS-ComP $\Delta$ 23 <sub>110264</sub>	MBP fused to a gly-gly-gly-ser peptide linking ComP $\Delta$ 23 <sub>110264</sub> with a c-terminal hex-his tag
pEXT20-MBP-AAA-ComP $\Delta$ 23 <sub>ADP1</sub>	MBP fused to a triple alanine peptide linking ComP $\Delta$ 23 <sub>ADP1</sub> with a c-terminal hex-his tag
pEXT20-MBP-GGGS-ComP $\Delta$ 23 <sub>ADP1</sub>	MBP fused to a gly-gly-gly-ser peptide linking ComP $\Delta$ 23 <sub>ADP1</sub> with a c-terminal hex-his tag
pEXT20-EPA-GGGS-ComP $\Delta$ 23 <sub>110264</sub>	The DsbA signal peptide fused to EPA fused to a gly-gly-gly-ser peptide linking ComP $\Delta$ 23 <sub>110264</sub> with a c-terminal hex-his tag

**[0134]** Unless otherwise stated, *E. coli* strains were grown in Terrific Broth (TB) at 37°C overnight. *S. pneumoniae* strains were grown in brain heart infusion (BHI) broth or sheep blood agar plates at 37°C in 5% CO<sub>2</sub>. For plasmid selection the antibiotics were used at the following concentrations: ampicillin (100 µg/mL), tetracycline (20 µg/mL), chloramphenicol (12.5 µg/mL), kanamycin (20 µg/mL) and spectinomycin (80 µg/mL) were added as needed.

**[0135] Heterologous glycosylation in *E. coli*.** For all heterologous glycosylation experiments, the *E. coli* SDB1 cell line was used as it has previously been established as a suitable strain for glycoengineering. Electrocompetent *E. coli* SDB1 was prepared as described by Dower and colleagues. Cells were electroporated with plasmids encoding the glycan synthesis loci, acceptor proteins and OTases. Colonies were picked and grown at 37°C in TB with appropriate antibiotic selection and immediately induced with 0.05-0.1 mM IPTG or 0.2% arabinose as needed and left overnight at 37°C. Cultures requiring arabinose induction received a second dose of arabinose after 4 hours. Cell pellets were obtained at stationary phases and prepared for western blot analysis.

**[0136] Western blotting.** Cell lysates containing the equivalent of OD<sub>600</sub> = 0.1 units were loaded on 12.5% in-house prepared SDS-PAGE gels, which were then transferred to nitrocellulose membranes (Biorad). Western blotting was performed according to previously published protocols. Nitrocellulose membranes were then visualized using an Odyssey Infrared Imaging System (LiCor Biosciences, USA).

**[0137] Purification of proteins and glycoproteins.** C-terminally Hexa-histidine-tagged ComP and ComP bioconjugates were purified from *E. coli* total membrane preparations. Cells were grown overnight in 2 L of terrific broth at 37°C, washed with phosphate buffered saline (PBS) buffer, and resuspended in 60 mL of the same buffer. Cells were lysed by two rounds of cell disruption at approximately 20 kPSI using a French press (Aminco) followed by the addition of a protease inhibitor cocktail (Roche). Lysates were centrifuged twice for 30 minutes at 20,000 x g to pellet cell debris. Supernatants were ultra-centrifuged at 100,000 x g for 60 minutes to pellet total membranes. The pellets were resuspended in PBS buffer containing 0.5% n-dodecyl- $\beta$ -D-maltoside (DDM) and membrane proteins were solubilized by tumbling for 48 hours. An equal volume of PBS was added to the suspension to reduce detergent concentration to 0.25% and the suspension was ultra-centrifuged at 100,000 x g for 60 minutes. Solubilized membranes were filtered through 0.45 $\mu$ m and 0.22 $\mu$ m filters and loaded on a His-Trap HP column (GE Healthcare) fitted to an ÄKTA purifier (Amersham Biosciences, Sweden). The column was equilibrated with a PBS/DDM buffer containing 20 mM imidazole before loading the sample. Unbound proteins were removed by washing the column with seven column volumes of buffer containing 20mM and 30mM imidazole in PBS stepwise. To elute proteins bound to the column, a gradient elution with an incremental increase in imidazole concentration was used. The majority of unconjugated and conjugated ComP eluted between 180mM and 250mM imidazole. Imidazole was removed by an overnight round of dialysis followed by two 2-hour rounds through a 3.5 kDa dialysis membrane (Spectrum labs) in a 250mL dialysis buffer composed of PBS containing 0.25% w/v DDM. The final theoretical concentration of imidazole post dialysis was about 0.007 mM. Proteins were quantified using a DC kit (biorad) after which the samples were diluted to the appropriate concentrations for mouse immunizations.

**[0138] Murine model immunizations.** The immunogenicity of a CPS14-ComP bioconjugate in a murine vaccination model was evaluated. Two groups of mice (n=10) individually received 3  $\mu$ g of either unglycosylated ComP or CPS14-ComP bioconjugate. Mice were boosted on days 14 and 28, and sacrificed on day 49 for whole blood collection. Each vaccine was formulated based on total protein. Using an enzyme linked immunosorbent assay (ELISA) with a serotype 14 strain of *S. pneumoniae* adsorbed to each well, IgM and IgG responses to CPS14 were compared. As seen in **Figure 12A,B**, sera collected from mice vaccinated with a CPS14-ComP bioconjugate had an increased IgG response specific to CPS14 (**Figure 12A**) but not and increased IgM response (**Figure 12B**). Further, secondary HRP-tagged anti-IgG subtype antibodies to were

employed determine which of the IgG subtypes were present in CPS14-ComP vaccinated mice (**Figure 12C**). As seen in **Figure 12C**, the CPS14-specific IgG1 response was higher than the other subtypes, which is consistent with previous findings for pneumococcal conjugate vaccines (Wuorimaa et al. *J Infect Dis* 184, 1211-1215 (2001); Soininen, A., Seppala, I., Nieminen, T., Eskola, J. & Kayhty, H. *Vaccine* 17, 1889-1897 (1999)).

**[0139]** There are more than 90 serotypes of *S. pneumoniae* (Geno, K.A. et al. *Clin Microbiol Rev* 28, 871-899 (2015); Bentley, S.D., et al. *PLoS Genet* 2, e31 (2006)). Many increasingly prevalent serotypes, like serotypes 8, 22F, and 33F are not included in currently licensed vaccines (Pilishvili, T., et al. *J Infect Dis* 201, 32-41 (2010)). Therefore, versatility 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) was tested. The aforementioned CPSs all contain Glc as the reducing end sugar and are therefore not compatible with other commercially exploited conjugating enzymes. As seen in **Figure 21A-F**, western blot analyses of affinity purified proteins from whole cells co-expressing PglS, ComP, and either the CPS8 or CPS9V polysaccharides resulted in the generation CPS-specific ComP bioconjugates, respectively. Again, to confirm that the material purified was not contaminated with lipid-linked polysaccharides, samples were treated with proteinase K and observed a loss of signal when analyzed via western blotting, confirming that the bioconjugates were proteinaceous.

**[0140]** Next, a vaccination trial was performed to determine the immunogenicity of a trivalent CPS8-, CPS9V-, and CPS14-ComP pneumococcal bioconjugate vaccine (**Figure 22A-L**). Three control groups were included: one group receiving carrier protein alone (unglycosylated ComP); another group receiving a monovalent dose of the CPS14-ComP bioconjugate to account for IgG specificity when analyzing immune responses against other serotypes; and a third group receiving PREVNAR 13® as a positive control. All immunogen groups contained and equal mixture of Freund's adjuvant, including mice receiving PREVNAR 13®. Day 49 sera from each group were employed for ELISAs on plates coated with *S. pneumoniae* serotypes 8, 9V and 14. As mentioned above, serotypes 9V and 14 are included in PREVNAR 13® and elevated IgG response could be seen in PREVNAR 13® immunized mice against these two serotypes 49 days post vaccination (**Figure 13**). Mice receiving the monovalent CPS14-ComP bioconjugate also showed significant IgG increase specific to serotype 14 (**Figure 12** and **Figure 13**). Mice receiving the trivalent CPS8/CPS9V/CPS14-ComP bioconjugate also

had statistically significant increases in serotype specific IgG responses 49 days post vaccination (**Figure 13**).

**[0141]** Immunizations were conducted at the Southern Alberta Cancer Research Institute (SACRI) antibody services. For the CPS14 ComP monovalent immunizations, 4-6 weeks old female BALB/c mice were injected with 100  $\mu$ L of purified protein/glycoprotein (3  $\mu$ g total protein) with 50  $\mu$ L of Freund's adjuvant. Two groups of mice (n=10) were injected either unglycosylated ComP (placebo) or CPS-ComP conjugate. Sera from the mice were obtained before immunizations and 7, 21, 35 and 49 days post immunizations. Booster doses were given on days 14 and 28. The same procedure was followed for the trivalent immunization, except four groups of mice (n=10) were used for the four different immunization groups. These groups were injected with 100  $\mu$ L containing 3  $\mu$ g of unconjugated ComP (placebo) and Freund's adjuvant, 100  $\mu$ L containing 3  $\mu$ g of ComP-CPS14 conjugate and Freund's adjuvant, 100  $\mu$ L containing 9  $\mu$ g of a glycoprotein mixture (ComP-CPS8, ComP-CPS9V and ComP-CPS14) and Freund's adjuvant, or 100  $\mu$ L of a 1:3 diluted stock of PREVNAR 13<sup>®</sup> and Freund's adjuvant. CPS-ComP bioconjugates were formulated by total protein for this immunization.

**[0142]** Because Freund's adjuvant is not a suitable adjuvant for human clinical development, another immunization trial was performed with vaccines containing formulated with Imject Alum Adjuvant, a mild adjuvant containing a mixture of aluminum hydroxide and magnesium hydroxide. Vaccination cohorts included a buffer/adjuvant test group, a PREVNAR 13<sup>®</sup> test group, and a trivalent CPS8-/CPS9V-/CPS14-ComP bioconjugate test group. Groups of three mice were vaccinated on days 1, 14, and 28. Serum was collected on day 42 and used to determine effector functions via an opsonophagocytosis assay (OPA). Given the limited amounts of sera collected from individual mice, sera were tested for bactericidal activity against serotypes 8 and 14, as one serotype is included in PREVNAR 13<sup>®</sup> (serotype 14) and one is not (serotype 8). As seen in **Figure 23A and 23B**, serum from a representative mouse vaccinated with the trivalent CPS8-/CPS9V-/CPS14-ComP bioconjugate had increased bactericidal activity against *S. pneumoniae* serotype 14 strain when compared to sera from a mock vaccinated mouse. Importantly, that same bioconjugate vaccinated serum had high bactericidal activity against a *S. pneumoniae* serotype 8 strain, which was not observed for PREVNAR 13<sup>®</sup> vaccinated sera due to the absence of this conjugate in its formulation.

**[0143]** Another trivalent immunization experiment was conducted with groups of three 4-6 week old female BALB/c mice. Each immunization group was subcutaneously injected with 100

μL of a 1:1 immunogen (3 μg of protein of each of the trivalent bioconjugate or a 1:10 diluted stock of PREVNAR 13®) to Imject Alum Adjuvant. Mice were vaccinated on day 0, 14, and 28 and then sacrificed on day 42 for sera collection.

[0144] Another immunization experiment was conducted with groups of three 4-6 week old BALB/c mice (five female and five male per group). Mice were immunized subcutaneously with 100 μL of EPA (5 μg total protein), 100 μL of ComP-CPS8 (5 μg total polysaccharide), or 100 μL of EPA-CPS8 (0.1 μg total polysaccharide) on day 0, 14, and 28 and then sacrificed on day 42 for sera collection. Vaccines were formulated 1:1 with Imject Alum Adjuvant.

[0145] **Glycoengineering a pneumococcal bioconjugate with a conventional vaccine carrier.** Up to this point, the use of ComP from *A. baylyi* ADP1 has been exploited as a carrier protein for pneumococcal bioconjugate vaccine production. To increase the commercial applicability of this technology, however, a conventional vaccine carrier was sought to be compatible with the O-linked OTase. Chimeric fusion proteins were generated consisting of the ΔE553 variant of Exotoxin A from *Pseudomonas aeruginosa* (EPA) C-terminally fused to a ComP fragment lacking its first 28 amino acids (ComPΔ28). A ComP ortholog from *A. soli* strain 110264 was used as it was most efficiently glycosylated by PglS and also found to be glycosylated at the same conserved serine as ComP from *A. baylyi* ADP1. The EPA fusion was linked to ComPΔ28 with a glycine-glycine-glycine-serine (GGGS; SEQ ID NO: 23) linker and trafficked to the periplasm with a DsbA signal sequence.

[0146] Because current formulations of pneumococcal conjugate vaccines do not contain a conjugate for serotype 8, focus was placed on generating an EPA-CPS8 pneumococcal bioconjugate. The EPA fusion was introduced into SDB1 cells co-expressing PglS and CPS8, subsequently purified, and then probed for glycosylation. The EPA fusion was efficiently glycosylated with CPS8. Furthermore, mass spectrometry analysis of intact glycoproteins confirmed that the EPA fusion was repetitively modified with an increasing mass unit of 662 Da, which is the calculated mass of a single CPS8 subunit. The EPA fusion was found to be glycosylated with at least 11 CPS8 subunits by intact protein analysis; however, western blot and Coomassie analyses indicated that >15 subunits were able to be transferred.

[0147] Subsequently, a vaccination experiment was performed comparing the immunogenicity of an EPA-CPS8 pneumococcal bioconjugate to a ComP-CPS8 pneumococcal bioconjugate. Groups of 10 mice were either vaccinated with 5 μg of EPA alone (based on total protein), 5 μg of ComP-CPS8 (based on polysaccharide as determined by anthrone sulfuric acid),

- 54 -

or 100ng of EPA-CPS8 (based on polysaccharide as determined by mass spectrometry of intact EPA-CPS8). Mice were vaccinated on days 1, 14, and 28 with serum collected on day 42. All vaccines were formulated 1:1 with Imject Alum Adjuvant. ELISAs were subsequently performed to determine the IgG titers specific to CPS8. As seen in **Figure 25A**, mice vaccinated with either CompP-CPS8 or EPA-CPS8 has statistically significant increases in IgG titers specific to CPS8 when compared to EPA vaccinated mice. Additionally, the protective capacity of sera from vaccinated mice was determined using a murine adapted opsonophagocytosis assay (OPA) with whole blood leukocytes. As shown in **Figure 25B**, sera from vaccinated mice immunized with CompP-CPS8 displayed high levels of bactericidal killing ranging from 84- 50% with one mouse not displaying any killing activity. Moreover, sera from EPA-CPS8 vaccinated mice also displayed bactericidal ranging from 88%-10% with three mice displaying no killing activity. Expectedly, sera from EPA vaccine mice did not display killing activity.

**[0148] Enzyme linked immunosorbent assays (ELISAs).** *S. pneumoniae* strains grown overnight in BHI broth at 37°C in 5% CO<sub>2</sub> were washed in PBS and the optical density was adjusted to OD<sub>600</sub> =0.6 units. Cells were heat inactivated at 60°C for 2-4 hours followed by immobilization on high binding 96 well plates (Corning) by adding 50 µL/well. Plates were incubated on a tumbler overnight at 4°C. The following day, wells were washed three times with PBST (Phosphate buffered saline-tween) (100 µL/well) before blocking with 5% skimmed milk (250 µL/well) for 2 h. The wells were washed three times with PBST. Plates were incubated for an 1 hour at room temperature with mouse sera (100 µL/well) at a 1:500 dilution in 2.5% skimmed milk in PBST. For the positive control, commercial rabbit polyclonal antibodies against CPS were used (Statens serum institute). Negative control wells were treated with skimmed milk without any primary antibody. After incubation with the primary antibody, wells were washed three times with PBST followed by a one hour incubation with secondary HRP-conjugated antibodies (100 µL/well) diluted in 2.5% skimmed milk in PBST. After incubation, the wells were washed three times with PBST and 100 µL of the chromogenic substrate TMB (Cell Signaling Technology) was added to each well. Plates were incubated at room temperature for 5 minutes after which the absorbance at 650 nm was measured using a BioTek™ plate reader.

**[0149]** For IgG titer determinations, ELISA plates were coated with 100 µL of 1 x 10<sup>8</sup> CFU/mL of *S. pneumoniae* serotype 8 grown approximately to mid-log phase. Bacteria were washed twice in PBS and suspended in water prior to coating. ELISA plates were allowed to air dry in a biological hood for 24 hours. Fifty microliters of methanol were then added to each well

and allowed to air dry. Plates were stored in a re-sealable bag protected from the light until use. To perform the titration of mouse total IgG antibodies, day 42 sera was serially diluted (2-fold) in PBST and antibodies were detected using an anti-mouse, HRP-linked IgG (Cell Signaling Technology # 7076) diluted 1:4000. For mouse serum titrations, the reciprocal of the last serum dilution that resulted in an optical density at 450nm equal to or lower than 0.2 was considered the titer of that serum. For representation purposes, negative titers (less than or equal to the cutoff) were given an arbitrary titer value of 10. Inter-plate variations were controlled by including an internal reference positive control on each plate. This control was hyper-immune sera from a mouse previously immunized with the ComP-CPS8 bioconjugate vaccine. The ELISA reactions in TMB were stopped when an OD<sub>450nm</sub> of ~1 was obtained for the internal positive control.

**[0150] Site directed mutagenesis.** Site-directed mutagenesis was carried out to mutate the residues H325 in PglM and S82 and S84 of ComP as previously described (Fisher and Pei, 1997). Mutagenic primers were designed using Primer X, a web-based primer design program (<http://www.bioinformatics.org/primerx/>). Primers used are listed in Supplementary Table 1. PCR reactions were performed using *Pfu* polymerase and 2-10 ng of pMN2 as template. The PCR reaction consisted of an initial denaturation of 30 s at 95°C followed by 16 cycles of 30s at 95°C, 60s at 55°C, 360s at 68°C with no final extension. PCR reactions were DpnI digested for 2 hours to remove the template plasmid, then transformed into electrocompetent DH5α cells and grown on ampicillin for plasmid selection. Colonies were sequenced to confirm mutagenesis.

**[0151] Digestion of ComP-CPS14 conjugate.** Isolated ComP bands were processed as previously described with minor modification. Briefly, gel separated ComP bands were excised and destained in a 50:50 solution of 50 mM NH<sub>4</sub>HCO<sub>3</sub> : 100% ethanol for 20 minutes at room temperature with shaking at 750 rpm. Destained bands were then washed with 100% ethanol, vacuum-dried for 20 minutes and rehydrated in 10 mM DTT in 50mM NH<sub>4</sub>HCO<sub>3</sub>. Reduction was carried out for 60 minutes at 56°C with shaking. The reducing buffer was then removed and the gel bands washed twice in 100% ethanol for 10 minutes to ensure the removal of remaining DTT. Reduced ethanol washed samples were sequentially alkylated with 55 mM Iodoacetamide in 50mM NH<sub>4</sub>HCO<sub>3</sub> in the dark for 45 minutes at RT. Alkylated samples were then washed with 2 rounds of Milli-Q water and 100% ethanol then vacuum-dried. Alkylated samples were then rehydrated with 10 ng μl<sup>-1</sup> GluC (Promega, Madison WI) in 40 mM NH<sub>4</sub>HCO<sub>3</sub> at 4°C for 1 hr. Excess GluC was removed, gel pieces were covered in 40 mM NH<sub>4</sub>HCO<sub>3</sub> and incubated for 24 hours at 37°C. Peptides were concentrated and desalted using C<sub>18</sub> stage tips (Ishihama, Y.,

Rappsilber, J. & Mann, M. *J Proteome Res* 5, 988-994 (2006); Rappsilber, J., Mann, M. & Ishihama, Y. *Nature protocols* 2, 1896-1906 (2007)) and stored on tip at 4°C. Peptides were eluted in buffer B (0.5% acetic acid, 80% MeCN) and dried before analysis by LC-MS.

**[0152] Identification of glycopeptides using reversed phase LC-MS and HCD MS-MS.**

Purified peptides were re-suspended in Buffer A\* and separated using an in-house packaged 25 cm, 75 µm inner diameter, 360 µm outer diameter, 1.7µm 130Å CSH C<sub>18</sub> (Waters, Manchester, UK) reverse phase analytical column with an integrated HF etched nESI tip. Samples were loaded directly onto the column using an ACQUITY UPLC M-Class System (Waters) at 600 nl/min for 20 minutes with Buffer A (0.1% FA) and eluted at 300 nl/min using a gradient altering the concentration of Buffer B (99.9% ACN, 0.1% FA) from 2% to 32% B over 60 minutes, then from 32% to 40% B in the next 10 minutes, then increased to 80% B over 8 minutes period, held at 100% B for 2 minutes, and then dropped to 2% B for another 10 minutes. RP separated peptides were infused into a Q-EXACTIVE (Thermo Scientific) mass spectrometer and data acquired using data dependent acquisition. Two methods were used to identify putative glycopeptides. Method A aimed to enable robust peptide identification in which one full precursor scan (resolution 70,000; 350-1850 m/z, AGC target of  $1 \times 10^6$ ) was followed by 10 data-dependent HCD MS-MS events (resolution 35k AGC target of  $1 \times 10^5$  with a maximum injection time of 110 ms, NCE 26 with 25% stepping) with 90 seconds dynamic exclusion enabled. Method B aimed to enable more complete characterization of glycans within glycopeptides with one full precursor scan (resolution 70,000; 350-1850 m/z, AGC target of  $1 \times 10^6$ ) followed by 10 data-dependent HCD MS-MS events (resolution 35k AGC target of  $5 \times 10^5$  with a maximum injection time of 250 ms, NCE 13 with 25% stepping) with 90 seconds dynamic exclusion enabled.

**[0153] Database interrogation of identified glycopeptides.** Raw files were processed manually to identify potential glycopeptides based on the diagnostic oxonium 204.08 m/z ion. Putative glycopeptide derived scans were manually inspected and identified as possible GluC derived ComP glycopeptides based on the presence of an intense deglycosylated ComP derived peptide ion, matching within 10 ppm using the ExPASy FindPept tool (on the world wide web at [web.expasy.org/findpept/](http://web.expasy.org/findpept/)). To facilitate peptide assignments the resulting glycopeptides was manually annotated according to (Roepstorff, P. & Fohlman, J. *Biomed Mass Spectrom* 11, 601 (1984)) with the aid of the Protein Prospector tool MS-Product (on the world wide web at [prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct](http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct)).

**[0154] Intact Protein Analysis.** Intact analysis was performed using a 6520 Accurate mass Q-TOF mass spectrometer (Agilent, Santa Clara, CA). Protein samples were re-suspended in 2% acetonitrile, 0.1% TFA and immediately loaded onto a C5 Jupiter 5  $\mu\text{m}$  300Å 50mm \* 2.1 mm column (Phenomenex, Torrance, CA) Using an Agilent 1200. Samples were desalted by washing with buffer A (2% acetonitrile, 0.1% formic acid) for 4 minutes and then separated with a 12 min linear gradient from 2 to 100% buffer B (80% acetonitrile, 0.1% formic acid) at a flow rate of 0.200ml/min. MS1 Mass spectra were acquired at 1 Hz between a mass range of 300–3,000  $m/z$ . Intact mass analysis and deconvolution was performed using MassHunter B.06.00 (Agilent).

**[0155] Opsonophagocytosis Assay (OPA).** Assays were performed as previously described ( ; ) 47,48 and are briefly described below. Blood collection. Blood was collected by intracardiac puncture from naïve female mice (Charles River, Wilmington, MA), treated with sodium heparin, then diluted to obtain  $6.25 \times 10^6$  leukocytes/mL in RPMI 1640 supplemented with 5% heat-inactivated fetal bovine serum, 10 mM HEPES, 2 mM L-glutamine and 50  $\mu\text{M}$  2-mercaptoethanol. All reagents were from Gibco (Invitrogen, Burlington, ON, Canada). Bacterial suspension preparation. Isolated colonies on sheep blood agar plates of either *S. pneumoniae* serotypes 8 or 14 (Statens Serum Institut, Denmark) were inoculated in 5 ml of Todd-Hewitt Broth (THB) (Oxoid, Thermo Fisher Scientific, Nepean, Canada) and incubated for 16 hours at 37 °C with 5% CO<sub>2</sub>. Working cultures were prepared by transferring 0.1 mL of 16 h-cultures into 10 mL of THB, which was then incubated for 5 hours. Bacteria were washed 3 times and resuspended in PBS to obtain an OD<sub>600</sub> value of 0.6, which corresponds to  $1.5 \times 10^8$  and colony forming units (CFU)/mL and to  $3.5 \times 10^8$  CFU/mL for serotype 8 and serotype 14, respectively. Final bacterial suspensions were prepared in complete cell culture medium to obtain a concentration of  $6.25 \times 10^4$  CFU/mL. The number of CFU/mL in the final suspensions was determined by plating samples onto Todd-Hewitt Agar (THA). Opsonophagocytosis Assay. Diluted whole blood ( $5 \times 10^5$  total leukocytes) was mixed with  $5 \times 10^3$  CFU of *S. pneumoniae* serotype 8 or 14 (MOI of 0.01) and 5% (v/v) of serum from control (placebo) or vaccinated mice in a microtube to a final volume of 0.2 mL. Microtubes were incubated for 4 hours at 37 °C with 5% CO<sub>2</sub>, with shaking. After incubation, viable bacterial counts were performed on THA. Tubes with the addition of naïve mouse sera (5% v/v) or of commercial rabbit anti-*S. pneumoniae* types 8 or 14 serum (5% v/v) (Statens Serum Institut, Denmark), were used as negative and positive controls, respectively. The percent of bacteria killing was determined using the following

formula: percent bacteria killed =  $[1 - (\text{bacteria recovered from sample tubes} / \text{bacteria recovered from negative control tubes with naïve sera})] \times 100$ .

**Table 2.** Primers.

Primer	Sequence
igrF	ACTGGTCGACTAGTAGTACTATATGGCTTTAAA (SEQ ID NO: 25)
igrR	ACTGCTGCAGTTAATATTCTATTGAACAAAATTTAAC (SEQ ID NO: 26)
H325A F	GAGAATGGTTTACATACTCAGCGAATTTGTTCTTAGATTTAATG (SEQ ID NO: 27)
H325A R	CATTAAATCTAAGAACAAATTCGCTGAGTATGTAAACCATTCTC (SEQ ID NO: 28)
S82A F – ADP1	GGAGTCCAAGAAATTGCGGCAAGTAATGCCA (SEQ ID NO: 29)
S82A R – ADP1	GTGGCATTACTTGCCGCAATTTCTTGGACTCC (SEQ ID NO: 30)
S84A F- ADP1	CAAGAAATTTCAGCAGCGAATGCCACTACGAAC (SEQ ID NO: 31)
S84A R – ADP1	GTTCGTAGTGGCATTGCTGCTGAAATTTCTTG (SEQ ID NO: 32)
S82A 110254 F	ACAGATCGCGTCCGGCGCCGCGCAGCAGCGACAACAAATGTAGCGT (SEQ ID NO: 33)
S82A 110254 R	ACGCTACATTTGTTGTCGCTGCTGCGGCGCCGACGCGATCTGT (SEQ ID NO: 34)
S84A 110254 F	CGGGCGTCACACAGATCGCGGCCGGCGCCTCAGCAGCGACAACA (SEQ ID NO: 35)
S84A 110254 R	TGTTGTCGCTGCTGAGGCGCCGGCCGCGATCTGTGTGACGCCCG (SEQ ID NO: 36)

**Table 3.** Glycosylation Tag Sequences

X<sub>1</sub>X<sub>2</sub>GTX<sub>5</sub>X<sub>6</sub>X<sub>7</sub>X<sub>8</sub>X<sub>9</sub>X<sub>10</sub>X<sub>11</sub>X<sub>12</sub>CX<sub>14</sub>G<sup>V</sup>X<sub>17</sub>X<sub>18</sub>I<sup>X</sup>X<sub>20</sub>X<sub>21</sub>X<sub>22</sub>ASX<sub>25</sub>X<sub>26</sub>TX<sub>28</sub>NVX<sub>31</sub>X<sub>32</sub>AX<sub>34</sub>CX<sub>36</sub>X<sub>37</sub>  
 X<sub>38</sub>X<sub>39</sub>X<sub>40</sub>X<sub>41</sub>X<sub>42</sub>X<sub>43</sub>X<sub>44</sub>X<sub>45</sub>X<sub>46</sub> (SEQ ID NO: 37)

Wherein:

X<sub>1</sub> is V, A, or no amino acid;  
 X<sub>2</sub> is A, G, T, or no amino acid;  
 X<sub>5</sub> is P, S, or Q;  
 X<sub>6</sub> is S, M, or I;  
 X<sub>7</sub> is T, P, or V;  
 X<sub>8</sub> is A, S, or T;  
 X<sub>9</sub> is G, N, S, or T;  
 X<sub>10</sub> is N or no amino acid;  
 X<sub>11</sub> is S, G, or A;  
 X<sub>12</sub> is S or N;  
 X<sub>14</sub> is V, T, or A;  
 X<sub>17</sub> is Q, T, or E;  
 X<sub>18</sub> is E, Q, or T;  
 X<sub>20</sub> is S, N, A, or G;  
 X<sub>21</sub> is S or no amino acid;  
 X<sub>22</sub> is G or no amino acid;  
 X<sub>25</sub> is N, S, or A;  
 X<sub>26</sub> is A, S, or K;  
 X<sub>28</sub> is T, S, or K;  
 X<sub>31</sub> is A or E;  
 X<sub>32</sub> is T or S;  
 X<sub>34</sub> is T, Q, or A;  
 X<sub>36</sub> is G, S, or T;  
 X<sub>37</sub> is A, G, or D;  
 X<sub>38</sub> is S, L, or A;  
 X<sub>39</sub> is S, G, D, or T;  
 X<sub>40</sub> is A, V, or G;  
 X<sub>41</sub> is G, I, or V;  
 X<sub>42</sub> is Q, T, or I;  
 X<sub>43</sub> is I, V, T, or L;  
 X<sub>44</sub> is I, T, or V;  
 X<sub>45</sub> is M or no amino acid; and  
 X<sub>46</sub> is D or no amino acid.

- 60 -

CX<sub>14</sub>GVX<sub>17</sub>X<sub>18</sub>IX<sub>20</sub>X<sub>21</sub>X<sub>22</sub>ASX<sub>25</sub>X<sub>26</sub>TX<sub>28</sub>NVX<sub>31</sub>X<sub>32</sub>AX<sub>34</sub>C (SEQ ID NO: 45)X<sub>14</sub>GVX<sub>17</sub>X<sub>18</sub>IX<sub>20</sub>X<sub>21</sub>X<sub>22</sub>ASX<sub>25</sub>X<sub>26</sub>TX<sub>28</sub>NVX<sub>31</sub>X<sub>32</sub>AX<sub>34</sub> (SEQ ID NO: 38)

Wherein:

X<sub>14</sub> is V, T, or A, optionally V;  
X<sub>17</sub> is Q, T, or E, optionally Q;  
X<sub>18</sub> is E, Q, or T;  
X<sub>20</sub> is S, N, A, or G;  
X<sub>21</sub> is S or no amino acid;  
X<sub>22</sub> is G or no amino acid;  
X<sub>25</sub> is N, S, or A, optionally N;  
X<sub>26</sub> is A, S, or K, optionally A;  
X<sub>28</sub> is T, S, or K;  
X<sub>31</sub> is A or E, optionally A;  
X<sub>32</sub> is T or S, optionally T; or  
X<sub>34</sub> is T, Q, or A, optionally T.

\*\*\*\*\*

**[0156]** 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.

**[0157]** 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., Voza, 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).

CLAIMS

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 fragment that comprises an isolated fragment of a ComP protein,

(i) wherein the ComP glycosylation tag fragment comprises the amino acid consensus sequence of

X<sub>14</sub>G<sub>17</sub>V<sub>18</sub>X<sub>19</sub>I<sub>20</sub>X<sub>21</sub>X<sub>22</sub>A<sub>25</sub>X<sub>26</sub>T<sub>28</sub>N<sub>31</sub>X<sub>32</sub>A<sub>34</sub> (SEQ ID NO: 38),

wherein:

X<sub>14</sub> is V, T, or A;

X<sub>17</sub> is Q, T, or E;

X<sub>18</sub> is E, Q, or T;

X<sub>20</sub> is S, N, A, or G;

X<sub>21</sub> is S or no amino acid;

X<sub>22</sub> is G or no amino acid;

X<sub>25</sub> is N, S, or A;

X<sub>26</sub> is A, S, or K;

X<sub>28</sub> is T, S, or K;

X<sub>31</sub> is A or E;

X<sub>32</sub> is T or S; and

X<sub>34</sub> is T, Q, or A;

and/or,

(ii) wherein the ComP glycosylation tag fragment comprises the amino acid sequence of

VG<sub>1</sub>V<sub>2</sub>Q<sub>3</sub>E<sub>4</sub>I<sub>5</sub>S<sub>6</sub>A<sub>7</sub>S<sub>8</sub>N<sub>9</sub>A<sub>10</sub>T<sub>11</sub>T<sub>12</sub>N<sub>13</sub>V<sub>14</sub>A<sub>15</sub>T<sub>16</sub>A<sub>17</sub>T<sub>18</sub> (SEQ ID NO: 39),

TG<sub>1</sub>V<sub>2</sub>T<sub>3</sub>Q<sub>4</sub>I<sub>5</sub>A<sub>6</sub>S<sub>7</sub>G<sub>8</sub>A<sub>9</sub>S<sub>10</sub>A<sub>11</sub>T<sub>12</sub>T<sub>13</sub>N<sub>14</sub>V<sub>15</sub>A<sub>16</sub>S<sub>17</sub>A<sub>18</sub>Q<sub>19</sub> (SEQ ID NO: 40),

VG<sub>1</sub>V<sub>2</sub>Q<sub>3</sub>E<sub>4</sub>I<sub>5</sub>N<sub>6</sub>A<sub>7</sub>S<sub>8</sub>S<sub>9</sub>T<sub>10</sub>S<sub>11</sub>N<sub>12</sub>V<sub>13</sub>A<sub>14</sub>T<sub>15</sub>A<sub>16</sub>T<sub>17</sub>A<sub>18</sub>T<sub>19</sub> (SEQ ID NO: 41),

AG<sub>1</sub>V<sub>2</sub>E<sub>3</sub>T<sub>4</sub>I<sub>5</sub>G<sub>6</sub>A<sub>7</sub>S<sub>8</sub>N<sub>9</sub>K<sub>10</sub>T<sub>11</sub>K<sub>12</sub>N<sub>13</sub>V<sub>14</sub>E<sub>15</sub>S<sub>16</sub>A<sub>17</sub>A<sub>18</sub> (SEQ ID NO: 42), or

VG<sub>1</sub>V<sub>2</sub>Q<sub>3</sub>T<sub>4</sub>I<sub>5</sub>A<sub>6</sub>A<sub>7</sub>S<sub>8</sub>N<sub>9</sub>A<sub>10</sub>T<sub>11</sub>K<sub>12</sub>N<sub>13</sub>V<sub>14</sub>A<sub>15</sub>T<sub>16</sub>A<sub>17</sub>T<sub>18</sub>A<sub>19</sub> (SEQ ID NO: 43),

or a variant of any thereof having one or two amino acid substitutions, one or two amino acid additions, and/or one or two amino acid deletions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (ComP110264: ENV58402.1);

wherein the ComP glycosylation tag fragment is not more than 50 amino acids in length,  
and

wherein the fusion protein is glycosylated with the oligo- or polysaccharide on the ComP glycosylation tag fragment at a serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1).

2. The bioconjugate of claim 1,  
wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
VGVQEISASNATTNVATAT (SEQ ID NO: 39),  
TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
VGVQEINASSSTSNVATAT (SEQ ID NO: 41),  
AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or  
VGVQTIAASNATKNVATAT (SEQ ID NO: 43),  
or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (ComP<sub>110264</sub>: ENV58402.1); or  
wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
VGVQEISASNATTNVATAT (SEQ ID NO: 39),  
TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
VGVQEINASSSTSNVATAT (SEQ ID NO: 41),  
AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or  
VGVQTIAASNATKNVATAT (SEQ ID NO: 43).
3. The bioconjugate of claim 1,  
wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
VGVQEISASNATTNVATAT (SEQ ID NO: 39), or  
TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (COMP<sub>110264</sub>: ENV58402.1);  
or wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
VGVQEISASNATTNVATAT (SEQ ID NO: 39),

TGVTQIASGASAATTNVAQAQ (SEQ ID NO: 40).

4. The bioconjugate of any one of Claims 1 to 3, 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.
5. The bioconjugate of Claim 4, wherein the capsular polysaccharide is CPS14, CPS8, CPS9V, or CPS15b.
6. The bioconjugate of any one of Claims 1 to 3, wherein the oligo- or polysaccharide is produced by a bacteria from the genus *Klebsiella*; optionally wherein the oligo- or polysaccharide is a *Klebsiella pneumoniae*, *Klebsiella varriicola*, *Klebsiella michinganensis*, or *Klebsiella oxytoca* capsular polysaccharide.
7. The bioconjugate of any one of Claims 1 to 3, wherein the polysaccharide is a *Klebsiella pneumoniae* capsular polysaccharide, optionally wherein the polysaccharide is a serotype K1 or serotype K2 capsular polysaccharide of *Klebsiella pneumoniae*.
8. The bioconjugate of any one of Claims 1 to 7, wherein the oligo- or polysaccharide comprises a glucose at its reducing end.
9. The bioconjugate of any one of Claims 1 to 8, wherein the bioconjugate is produced *in vivo*; optionally in a bacterial cell.
10. The bioconjugate of any one of Claims 1 to 9, wherein the bioconjugate is a conjugate vaccine; optionally wherein the conjugate vaccine is a vaccine against *Streptococcus pneumoniae* serotype 8.

11. A ComP glycosylation tag fragment that comprises an isolated fragment of a ComP protein:

(i) wherein the ComP glycosylation tag fragment comprises the amino acid consensus sequence of

X<sub>14</sub>G<sub>17</sub>V<sub>18</sub>X<sub>18</sub>I<sub>20</sub>X<sub>21</sub>X<sub>22</sub>A<sub>25</sub>S<sub>26</sub>T<sub>28</sub>N<sub>31</sub>V<sub>32</sub>A<sub>34</sub> (SEQ ID NO: 38),

wherein:

X<sub>14</sub> is V, T, or A;

X<sub>17</sub> is Q, T, or E;

X<sub>18</sub> is E, Q, or T;

X<sub>20</sub> is S, N, A, or G;

X<sub>21</sub> is S or no amino acid;

X<sub>22</sub> is G or no amino acid;

X<sub>25</sub> is N, S, or A;

X<sub>26</sub> is A, S, or K;

X<sub>28</sub> is T, S, or K;

X<sub>31</sub> is A or E;

X<sub>32</sub> is T or S; and

X<sub>34</sub> is T, Q, or A;

and/or,

(ii) wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39),

TGVTQIASGASAATTNVAQAQ (SEQ ID NO: 40),

VGVQEINASSSTSNVATAT (SEQ ID NO: 41),

AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or

VGVQTIAASNATKNVATAT (SEQ ID NO: 43),

or a variant of any thereof having one or two amino acid substitutions, one or two amino acid additions, and/or one or two amino acid deletions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2

(ComP110264: ENV58402.1);

wherein the ComP glycosylation tag fragment is not more than 50 amino acids in length.

12. The ComP glycosylation tag fragment claim 11,

wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or VGVQTIAASNATKNVATAT (SEQ ID NO: 43),

or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (ComP110264: ENV58402.1); or

wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), VGVQEINASSSTSNVATAT (SEQ ID NO: 41), AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or VGVQTIAASNATKNVATAT (SEQ ID NO: 43).

13. The ComP glycosylation tag fragment claim 11, wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39), or TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40), or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (COMP110264: ENV58402.1);

or wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39), TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40).

14. The ComP glycosylation tag fragment of any one of Claims 11 to 13 attached to a heterologous carrier protein.

15. A fusion protein comprising a ComP glycosylation tag fragment:

(i) wherein the ComP glycosylation tag fragment comprises the amino acid consensus sequence of

X<sub>14</sub>GVX<sub>17</sub>X<sub>18</sub>IX<sub>20</sub>X<sub>21</sub>X<sub>22</sub>ASX<sub>25</sub>X<sub>26</sub>TX<sub>28</sub>NVX<sub>31</sub>X<sub>32</sub>AX<sub>34</sub> (SEQ ID NO: 38),

wherein:

X<sub>14</sub> is V, T, or A;

X<sub>17</sub> is Q, T, or E;

X<sub>18</sub> is E, Q, or T;

X<sub>20</sub> is S, N, A, or G;

X<sub>21</sub> is S or no amino acid;

X<sub>22</sub> is G or no amino acid;

X<sub>25</sub> is N, S, or A;

X<sub>26</sub> is A, S, or K;

X<sub>28</sub> is T, S, or K;

X<sub>31</sub> is A or E;

X<sub>32</sub> is T or S; and

X<sub>34</sub> is T, Q, or A;

and/or,

(ii) wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39),

TGVTQIASGASAATTNVAQAQ (SEQ ID NO: 40),

VGVQEINASSSTSNVATAT (SEQ ID NO: 41),

AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or

VGVQTIAASNATKNVATAT (SEQ ID NO: 43),

or a variant of any thereof having one or two amino acid substitutions, one or two amino acid additions, and/or one or two amino acid deletions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2

(ComP110264: ENV58402.1);

wherein the ComP glycosylation tag fragment is not more than 50 amino acids in length.

16. The fusion protein of claim 15,  
wherein the ComP glycosylation tag fragment comprises the amino acid sequence of VGVQEISASNATTNVATAT (SEQ ID NO: 39),

TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
 VGVQEINASSSTSNVATAT (SEQ ID NO: 41),  
 AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or  
 VGVQTIAASNATKNVATAT (SEQ ID NO: 43),

or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (ComP110264: ENV58402.1); or

wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
 VGVQEISASNATTNVATAT (SEQ ID NO: 39),  
 TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
 VGVQEINASSSTSNVATAT (SEQ ID NO: 41),  
 AGVETIGASNKTKNVESAA (SEQ ID NO: 42), or  
 VGVQTIAASNATKNVATAT (SEQ ID NO: 43).

17. The fusion protein of claim 15,  
 wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
 VGVQEISASNATTNVATAT (SEQ ID NO: 39), or  
 TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40),  
 or a variant of any thereof having one or two amino acid substitutions, wherein the variant maintains the serine residue corresponding to the conserved serine residue at position 82 of SEQ ID NO: 2 (COMP110264: ENV58402.1);  
 or wherein the ComP glycosylation tag fragment comprises the amino acid sequence of  
 VGVQEISASNATTNVATAT (SEQ ID NO: 39),  
 TGVTQIASGASAATTNVASAQ (SEQ ID NO: 40).

18. The fusion protein of any one of Claims 15 to 17 comprising 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.

19. 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 fragment of any one of claims 11 to 13,

wherein the oligo- or polysaccharide is linked to the ComP glycosylation tag fragment at a serine residue corresponding to the serine residue at position 84 of SEQ ID NO: 1 (ComPADP1: AAC45886.1), and

wherein the ComP protein or glycosylation tag fragment thereof is linked to a heterologous carrier protein.

20. The method of claim 19, wherein the *in vivo* conjugation occurs in a host cell.
21. The method of Claim 20, wherein the host cell is a bacterial cell, optionally, wherein the bacterial host cell is *E.coli*.
22. The method of Claim 20 or 21 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 (c) the acceptor polypeptide.
23. The method of any one of Claims 19 to 22, wherein the method produces a conjugate vaccine.
24. A host cell comprising (a) a genetic cluster encoding for the proteins required to synthesize an oligo- or polysaccharide; (b) a PglS OTase; and (c) an acceptor polypeptide comprising the ComP glycosylation tag fragment of any one of claims 11 to 13.
25. The host cell of Claim 24, wherein the acceptor polypeptide is a fusion protein.
26. The host cell of Claim 24 or 25, wherein the host cell comprises a nucleic acid encoding the PglS OTase, and/or wherein the host cell comprises a nucleic acid encoding the acceptor polypeptide.
27. An isolated nucleic acid encoding the ComP glycosylation tag fragment of any one of Claims 11 to 13 and/or the fusion protein of any one of Claims 15 to 18;

- optionally wherein the nucleic acid is a vector.
28. A host cell comprising the isolated nucleic acid of Claim 27.
  29. A composition comprising the conjugate vaccine of Claim 10 and an adjuvant.
  30. 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 Claim 10 or the composition of Claim 29.
  31. The method of claim 30, wherein the immune response is an antibody response.
  32. 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 Claim 10 or the composition of Claim 29.
  33. The method of Claim 32, wherein the infection is a localized or systemic infection of skin, soft tissue, blood, or an organ, or is auto-immune in nature.
  34. The method of Claim 32, wherein the disease is pneumonia.
  35. The method of Claim 32, wherein the infection is a systemic infection and/or an infection of the blood.
  36. The method of any one of Claims 32 to 35, wherein the subject is a human.
  37. A method of producing a pneumococcal conjugate vaccine against pneumococcal infection, the method comprising:
    - (a) isolating the bioconjugate of Claim 10; and
    - (b) combining the isolated conjugate vaccine or isolated glycosylated fusion protein with an adjuvant.

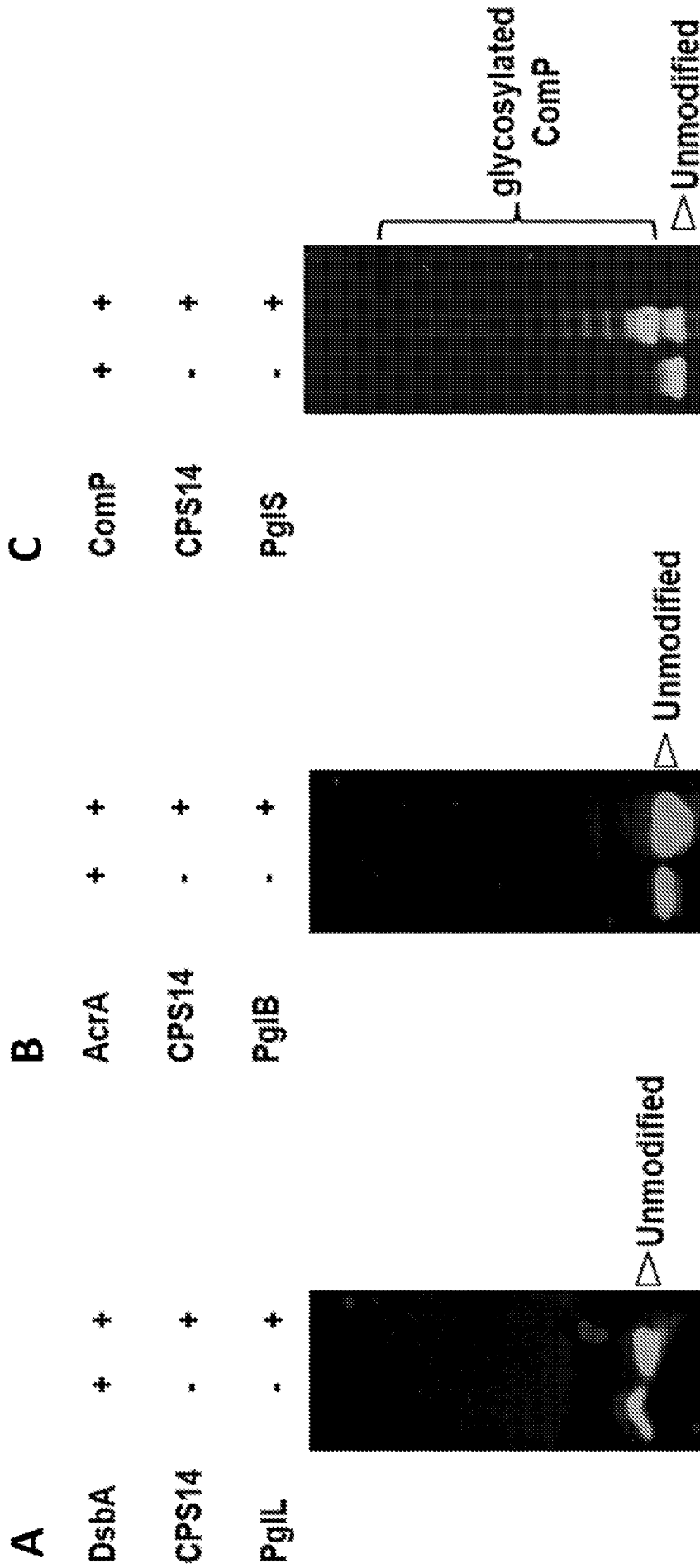


Figure 1A-C

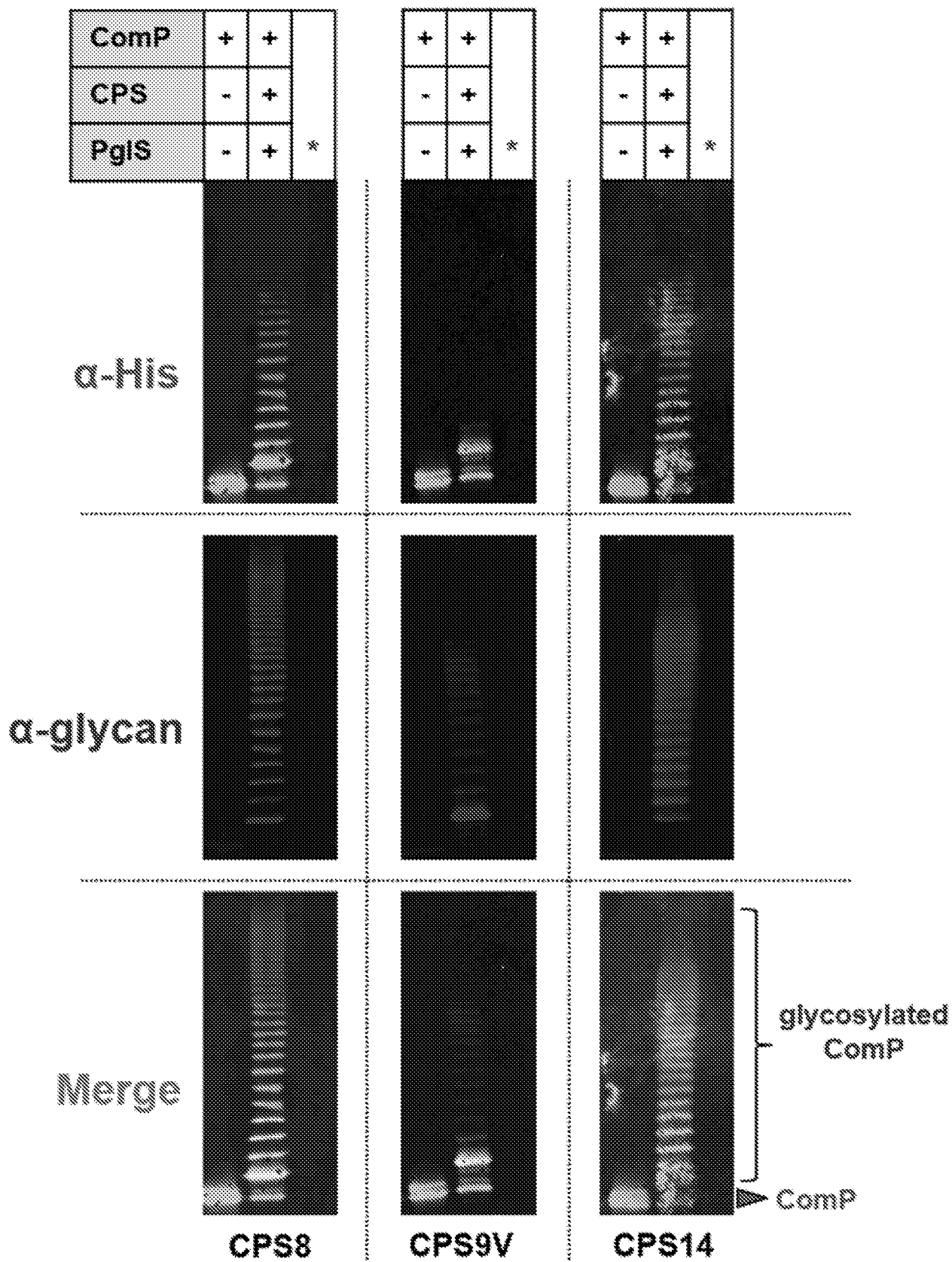


Figure 2

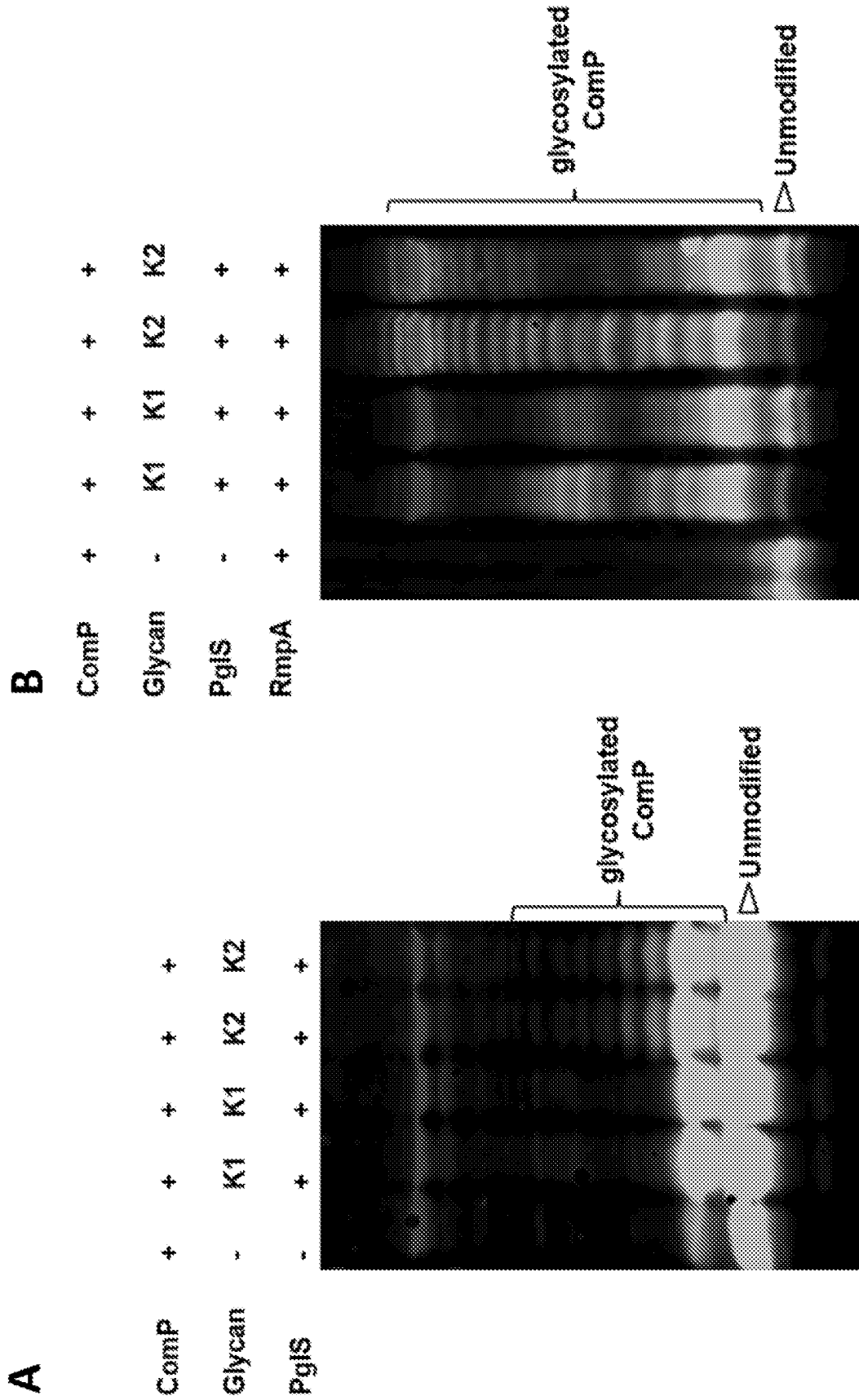


Figure 3A,B

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

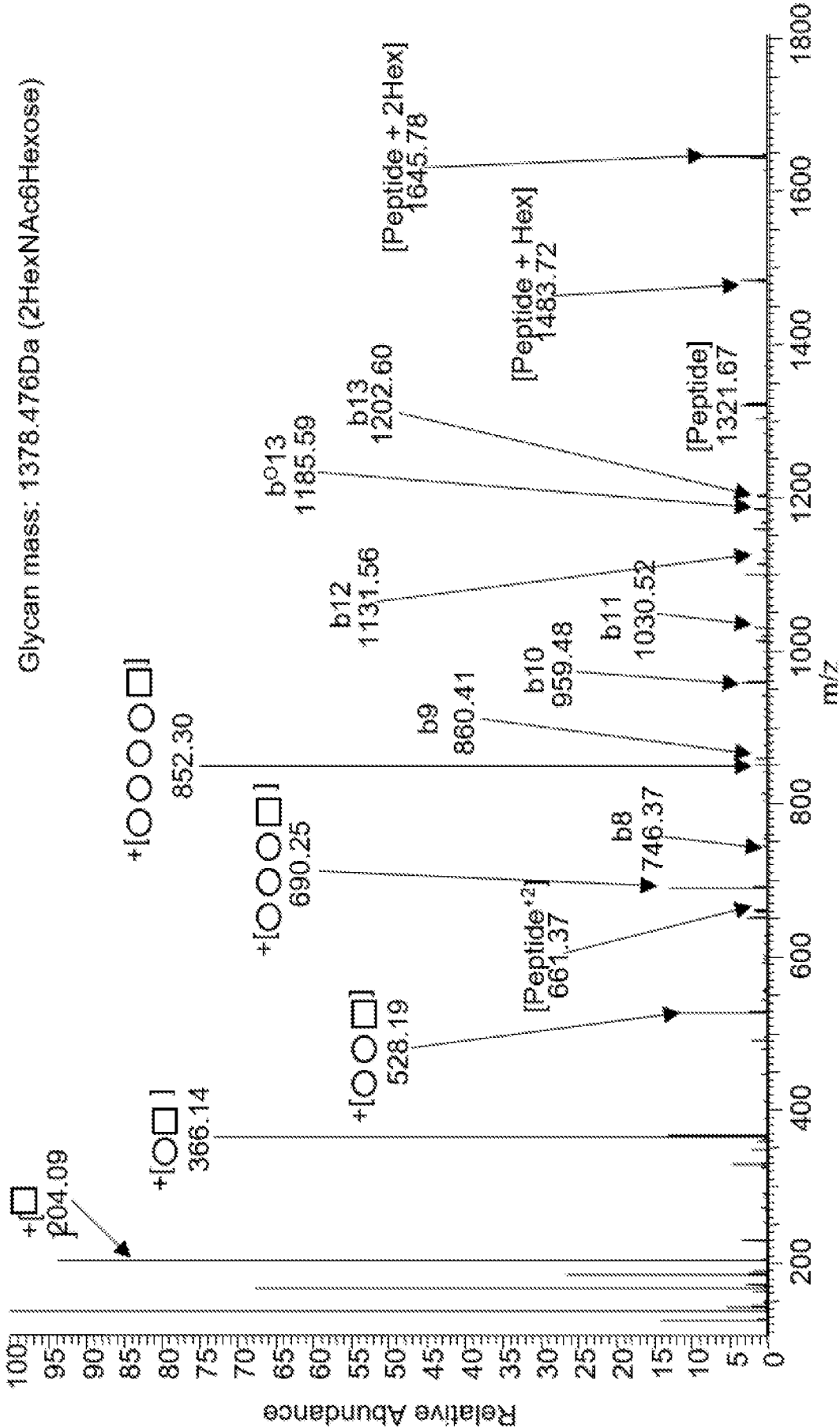


Figure 4A

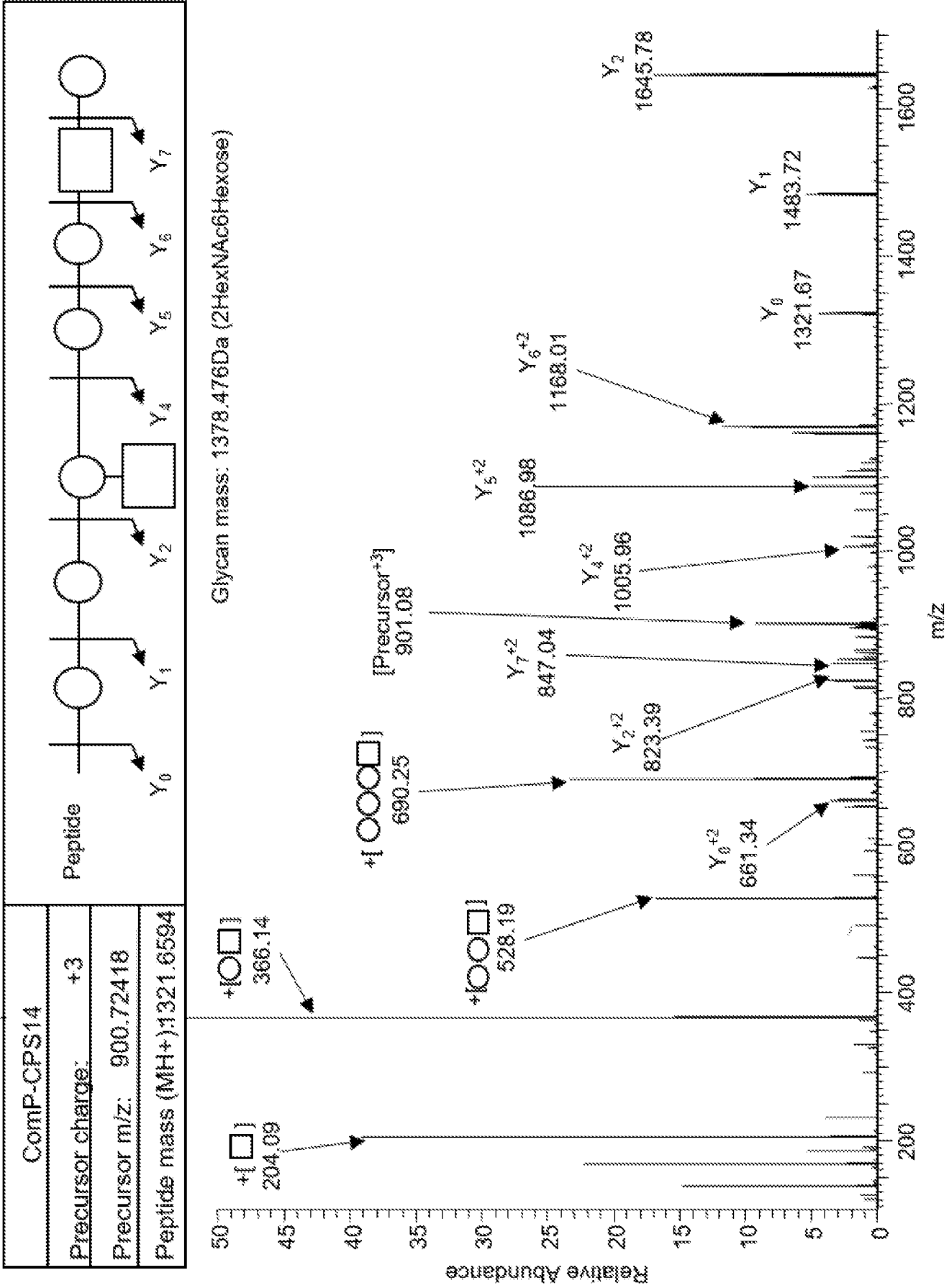


Figure 4B

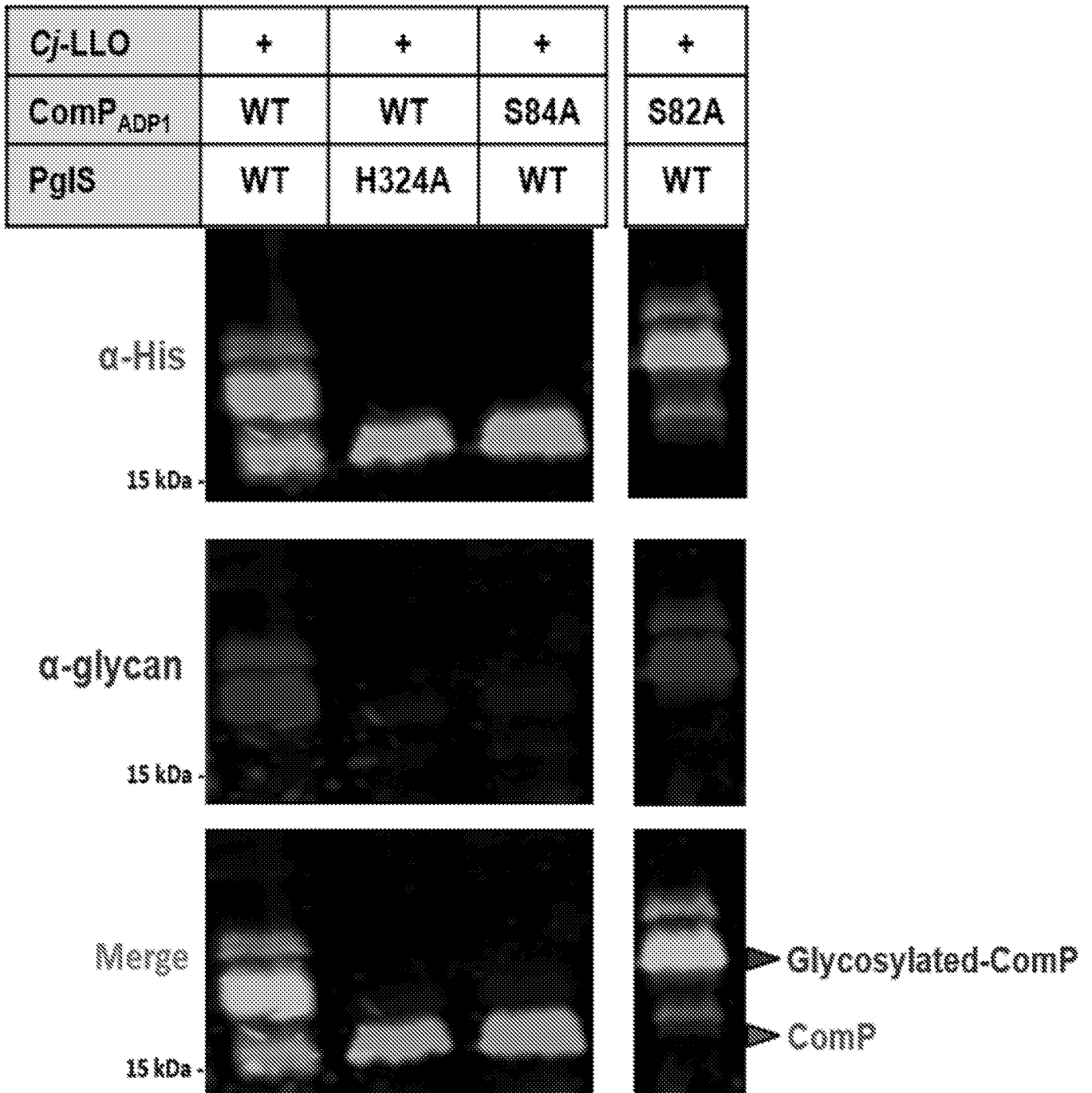


Figure 5

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

>ENV58402.1 hypothetical protein F951\_00736  
[*Acinetobacter soli* CIP 110264]  
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRSR  
VTEGLTTASAMKATVSENIMNAGGTSMPSGNC  
TGVTQIASGASAATTNVAQAQCSSDSDGVITVTMT  
DKAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYV  
PSECRGT (SEQ ID NO: 2)

>APV36638.1 competence protein [*Acinetobacter*  
*soli* GFJ-2]  
MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR  
VSEGLTTASAMKATVSENILSAGQIVTGTPSTANS  
SCVGVQEINASSSTSNVATATCSGLGVITVTMDS  
TKAKGVNLTLTPTYTTSSNAV TWKCTTTSDKKYVP  
SECRN (SEQ ID NO: 3)

Figure 6

>PKD82822.1 competence protein [*Acinetobacter radioresistens* 50v1]

MNTQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR  
VTEAVSTASSMKATVSENIMNAGGTQIPTSGNCV  
GVQTIAASNATKNVATATCTDSTGVIVVTTTPAAK  
SVPLTLPTYTGGNVKWACSTTANFKNYVPSEC  
RS (SEQ ID NO: 4)

>SNX44537.1 type IV pilus assembly protein Pila  
[*Acinetobacter puyangensis* ANC 4466]

MNAQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAR  
VTEALTTASAMKATVSENIMSAGGTTIASSACNG  
VISASATTNVASSACSGSGVISVTTTAAAKGIVLTL  
TPKYTGGNVAWQCTTTSGDAQKYVPSECRRTS  
(SEQ ID NO: 5)

>OAL75955.1 competence protein [*Acinetobacter*  
sp. SFC]

MNTQKGFTLIELMIVIAIIGILAAIAIPAYTDYTVRAK  
VTEAISTASAMKATVSENLMMSAGGTSIVSTNANC  
AGVETIGASNKTKNVESAACTAATGVILVTTTAEA  
KSVPLTLKPTYTGSNVQWKCGTTAAAFKYVPSE  
CRNDSSGTGF (SEQ ID NO: 6)

**Figure 6 Cont.**

<b>ComP variant</b>	110264	110264	110264	ADP1	ADP1	ADP1
<b>CPS8</b>	+	+	+	+	+	+
<b>PglS -</b>	-	110264	ADP1	-	110264	ADP1

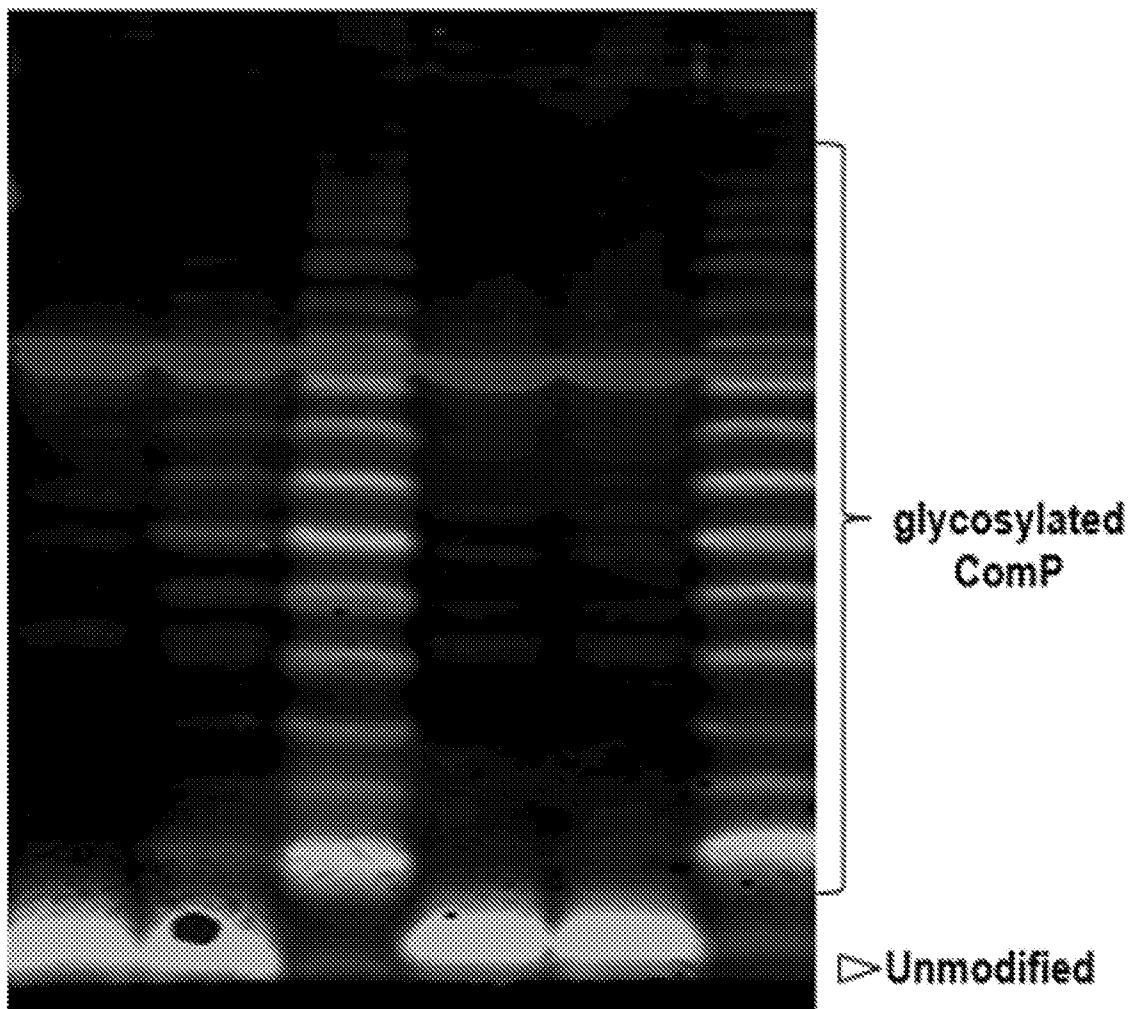


Figure 7

Linker	AAA	GGGS	AAA	GGGS	AAA	GGGS	AAA	GGGS	ADP1
CompP $\Delta$ 23	110264	110264	110264	110264	ADP1	ADP1	ADP1	ADP1	ADP1
CPS8	+	+	+	+	+	+	+	+	+
PgIS	-	+	-	+	-	+	-	+	+

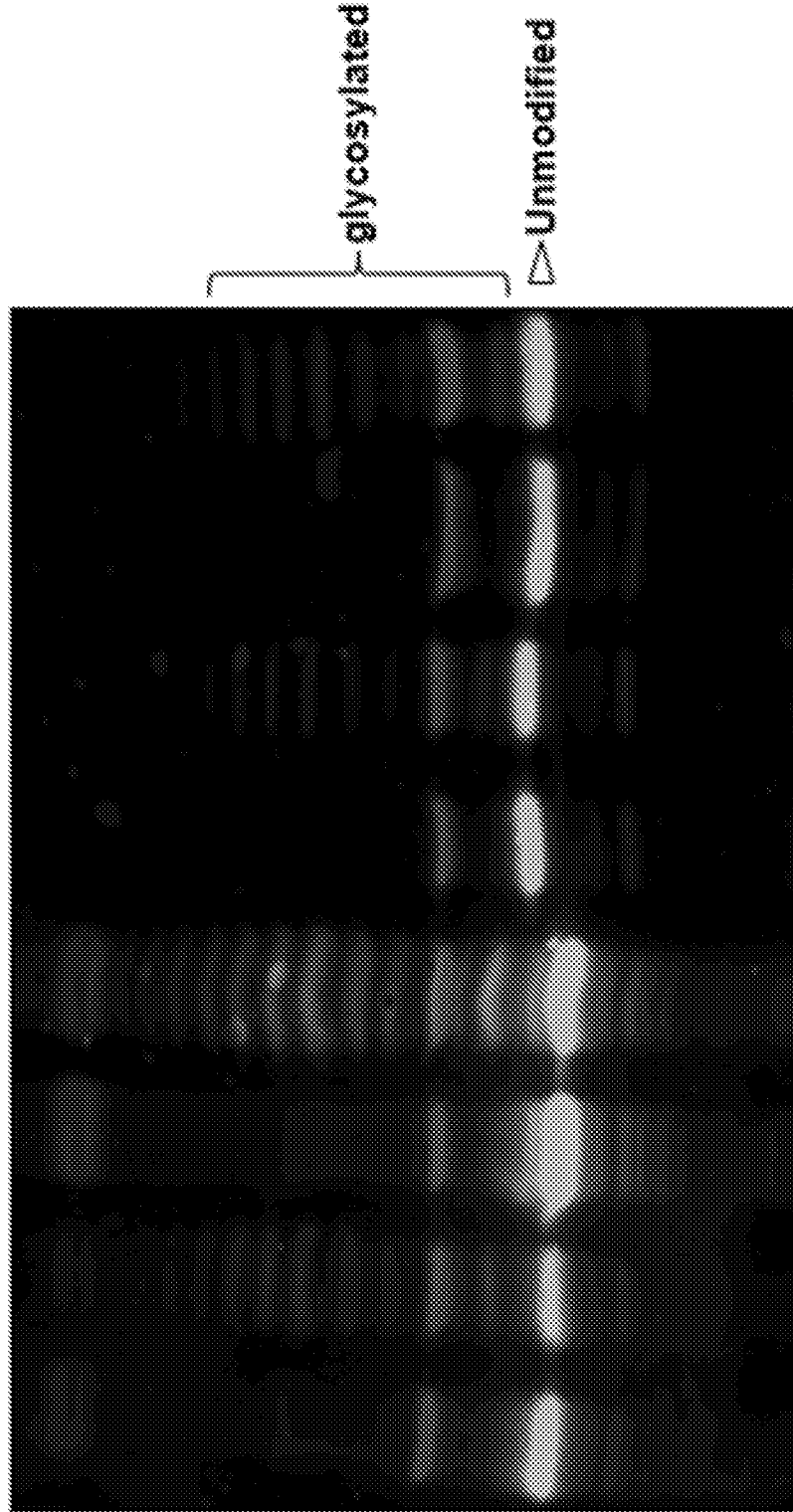


Figure 8

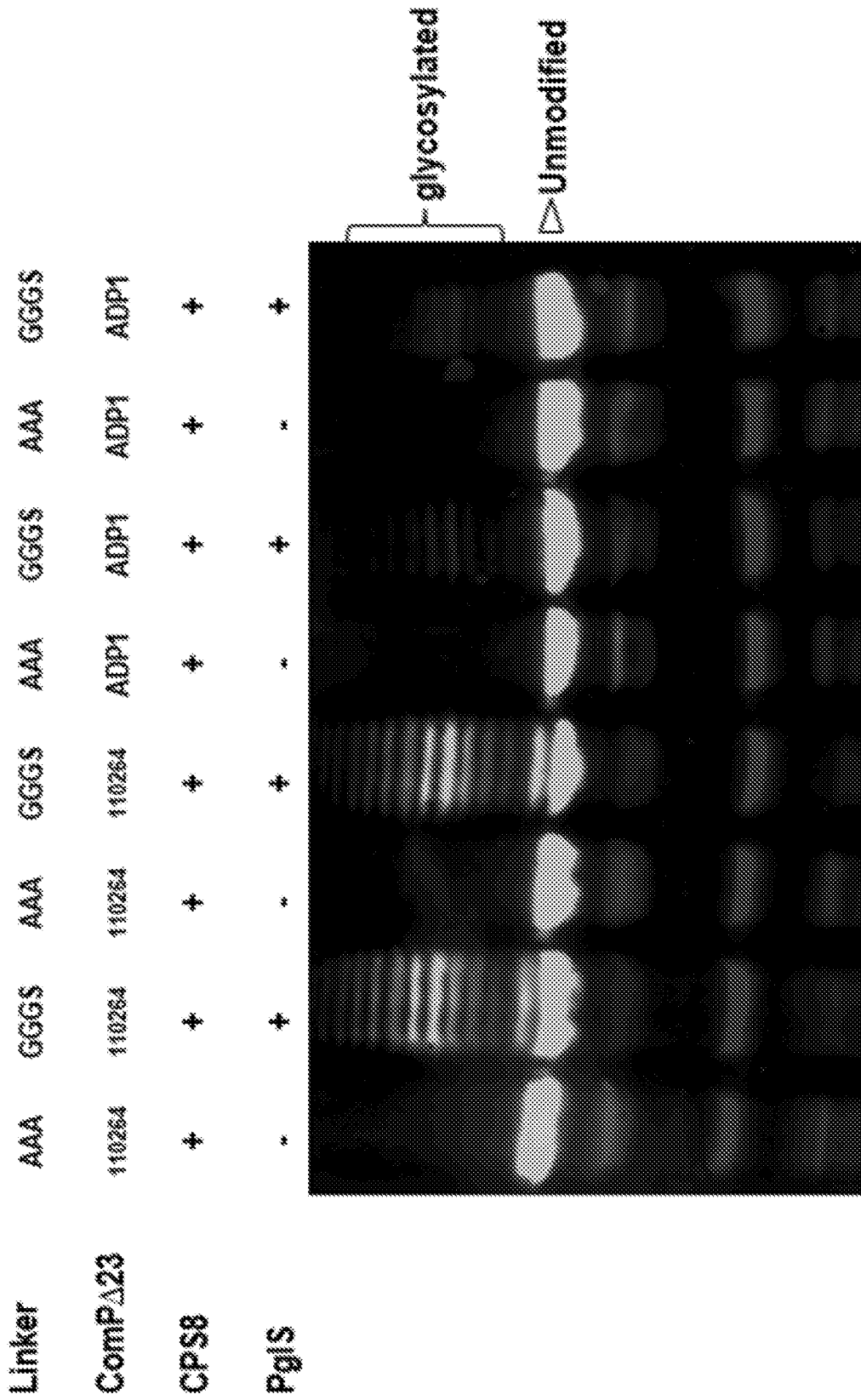


Figure 9

EPA-GGS	+	+	+	+	+	+
CPS8	+	+	+	+	+	+
PgIS	-	-	110264	110264	ADP1	ADP1
Cell Comp.	WC	Peri.	WC	Peri.	WC	Peri.

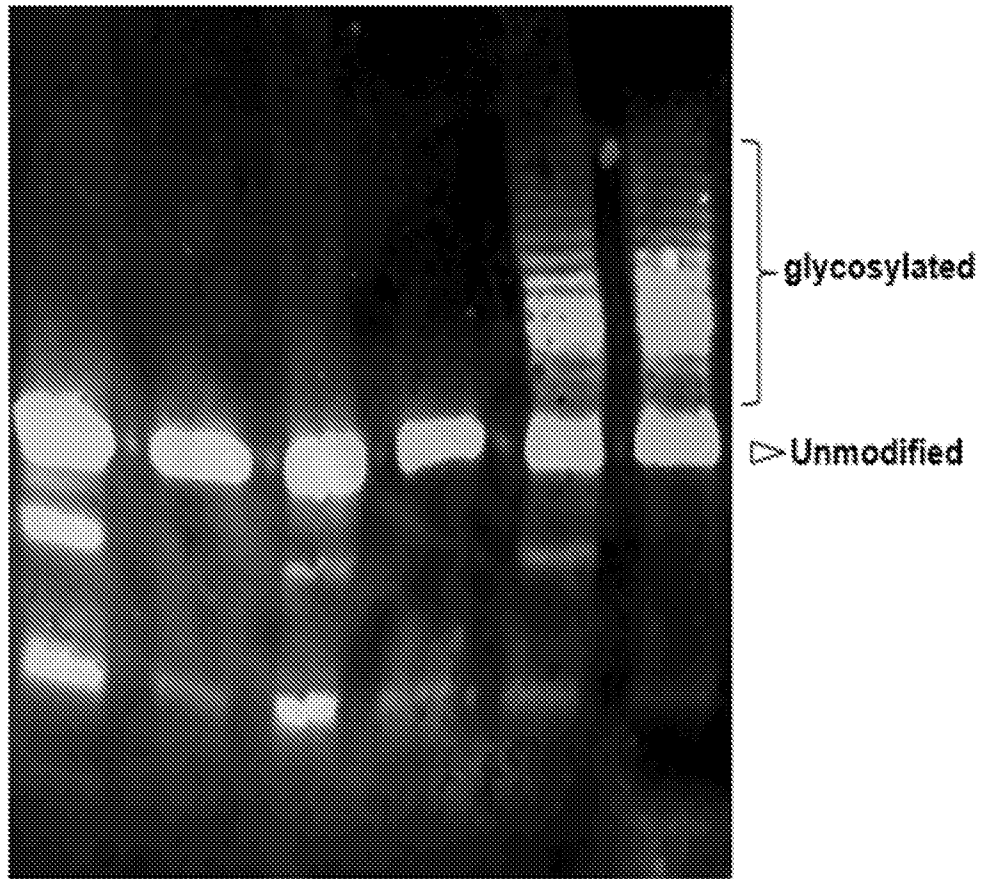


Figure 10



>DsbA-GGGS-ComP $\Delta$ 28<sub>110264</sub>-His  
 MKKIWLALAGLVLA FSASAAQYEDGKQYTTLEKPVA  
 GAPQVLEFFSFFCPHCYQFEEVLHISDNVKKKLPEGV  
 KMTKYHVNFMGDDLKDLTQAWAVAMALGVEDKVT  
 VPLFEGVQKTQTIRSASDIRDVFINAGIKGEEYDAAWN  
 SFVVKSLVAQQEKAAADVQLRGVPAMFVNGKYQLNP  
 QGMDTSNMDVFVQQYADTVKYLSEKKGGGSAYTDY  
 TVRSRVTEGLTTASAMKATVSENIMNAGGTSMPSG  
 NCTGVTQIASGASAATTNVASAQCSDSDGVITVTMTD  
 KAKGVSIKLTPSFSSTGSVGWKCTTSSDKKYVPSECR  
 GTHHHHHH (SEQ ID NO: 18)

### Legend

DsbA – thiol disulfide oxidoreductase protein (*dsbA* gene)

MBP – maltose binding protein (*malE* gene)

DsbA<sub>SP</sub> – signal sequence of DsbA (first 19 amino acids)

EPA – exotoxin A from *Pseudomonas aeruginosa*

AAA – triple alanine linker

GGGS – glycine-glycine-glycine-serine linker

ComP $\Delta$ 28<sub>110264</sub> – ComP from *A. soli* CIP 110264 without the first 28 amino acids

His – hexa-histidine tag

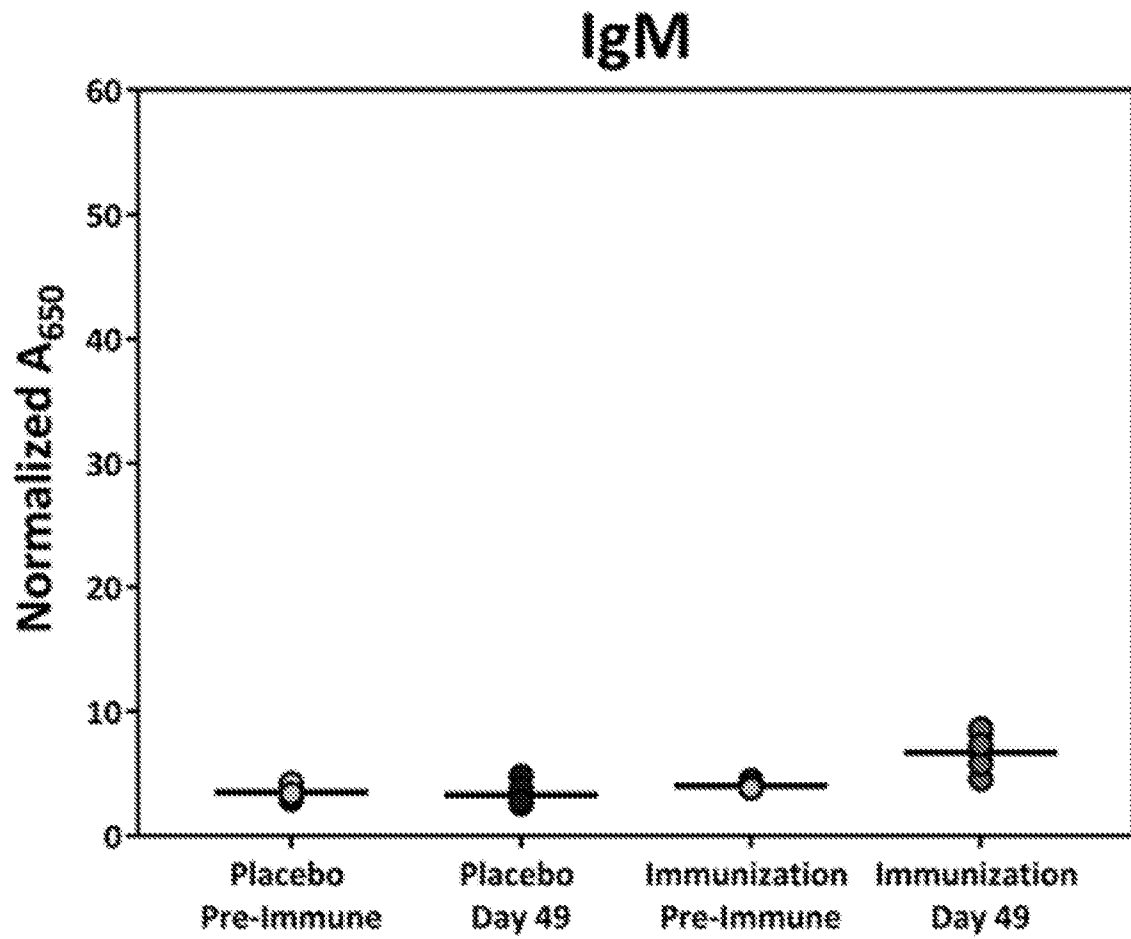
Figure 11

>MBP-AAA-ComP $\Delta$ 28<sub>110264</sub>-His  
MKIKTGARILALSALTTMMFSASALAKIEEGKLVWING  
DKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQV  
AATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQD  
KLYPFTWDAVRYNGKLIAYPIAVEALS LIYNKDLLPNPP  
KTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAA  
DGGYAFKYENGGYDIKDVGVNDAGAKAGLTFLVDLIK  
NKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDT  
SKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKE  
LAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEEL  
AKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVIN  
AASGRQTVDEALKDAQTNAAAAYTDYTVRSRVTEGL  
TTASAMKATVSENIMNAGGTSMPSGNCCTGVTQIASG  
ASAATTNVASAQCSDSDGVITVTMTDKAKGVSIKLTPS  
FSSTGSGWKCTTSSDKKYVPSECRGTHHHHHH  
(SEQ ID NO: 19)

Figure 11 Cont.

>DsbA<sub>sp</sub>-EPA-GGGS-ComP $\Delta$ 28<sub>110264</sub>-His  
MKKIWLALAGLVLAFSASAAEEAFDLWNECAKACVL  
DLKDGVRSSRMSVDPAIADTNGQGVLHYSMVLEGGN  
DALKLAIDNALSITSDGLTIRLEGGVEPNKPVRYSTR  
QARGSWSLNWLVPIGHEKPSNIKVFIHELNAGNQLSH  
MSPIYTIEMGDELLAKLARDATFFVRAHESNEMQPTL  
AISHAGVSVVMAQAQPRREKRWSEWASGKVLCLLD  
PLDGVYNYLAQQRCNLDDTWEGKIYRVLAGNPAKHD  
LDIKPTVISHRLHFPEGGSLAALTAHQACHLPLETFTR  
HRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVD  
QVIRNALASPGSGDLGEAIREQPEQARLALTAAAE  
SERFVRQGTGNDEAGAASADVSLTCPVAAGECAGP  
ADSGDALLERNYPTGAEFLGDGGDISFSTRGTQNWT  
VERLLQAHRQLEERGYVFGYHGTFLAAQSIVFGGV  
RARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARG  
RIRNGALLRVYVPRSSLPGFYRTGLTLAAPEAAGEVE  
RLIGHPLPLRLDAITGPEEEGGRLTILGWPLAERTVVIP  
SAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGK  
PPREDLKGGGSSAYTDYTVRSRVTEGLTTASAMKATVS  
ENIMNAGGTSMPSSGNCTGVTQIASGASAATTNVA  
QCSDSDGVITVTMTDKAKGVSIKLTPSFSSTGSVGWK  
CTTSSDKKYVPSECRGTHHHHHH (SEQ ID NO: 20)

Figure 11 Cont.

**Figure 12A**

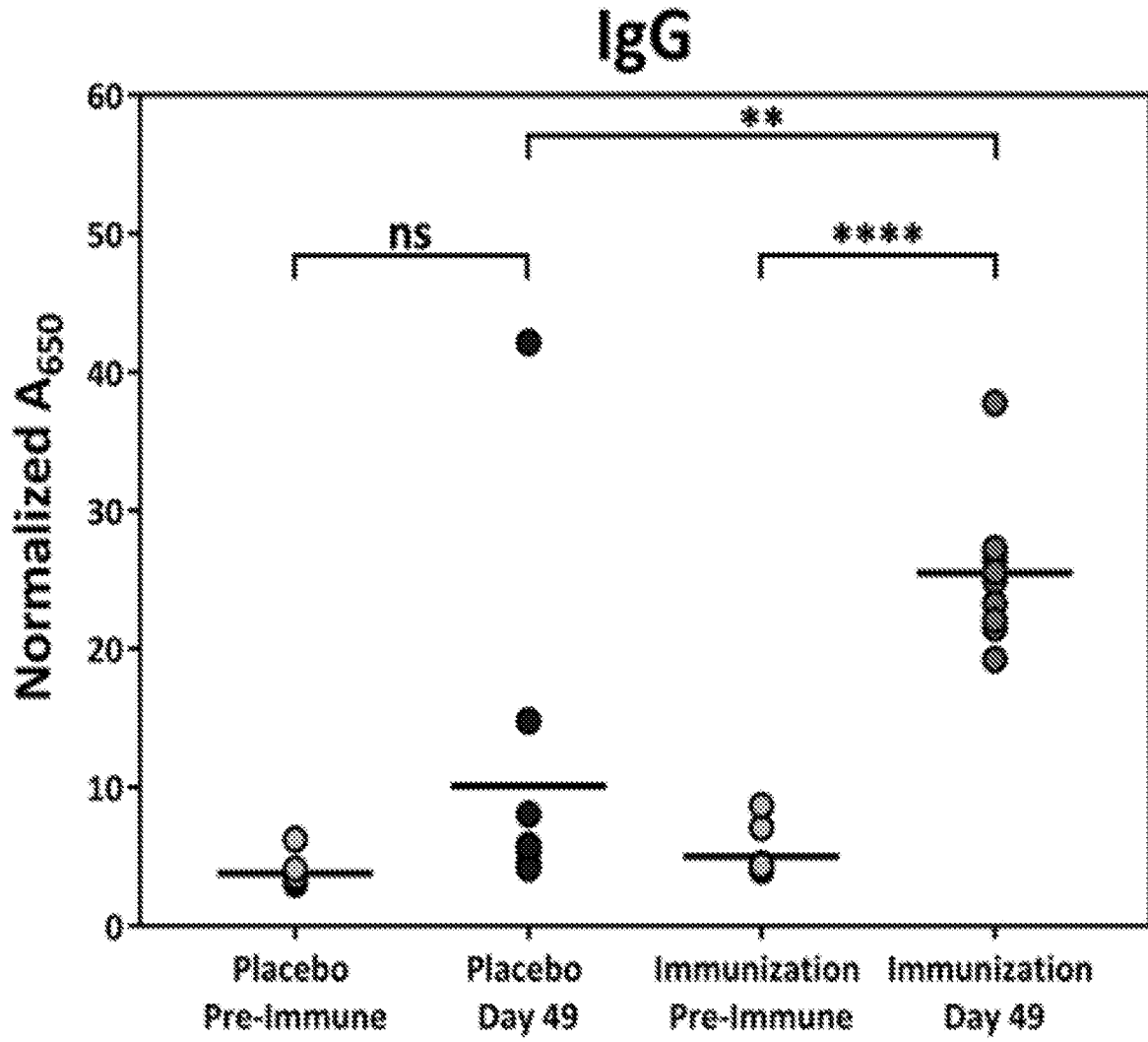


Figure 12B

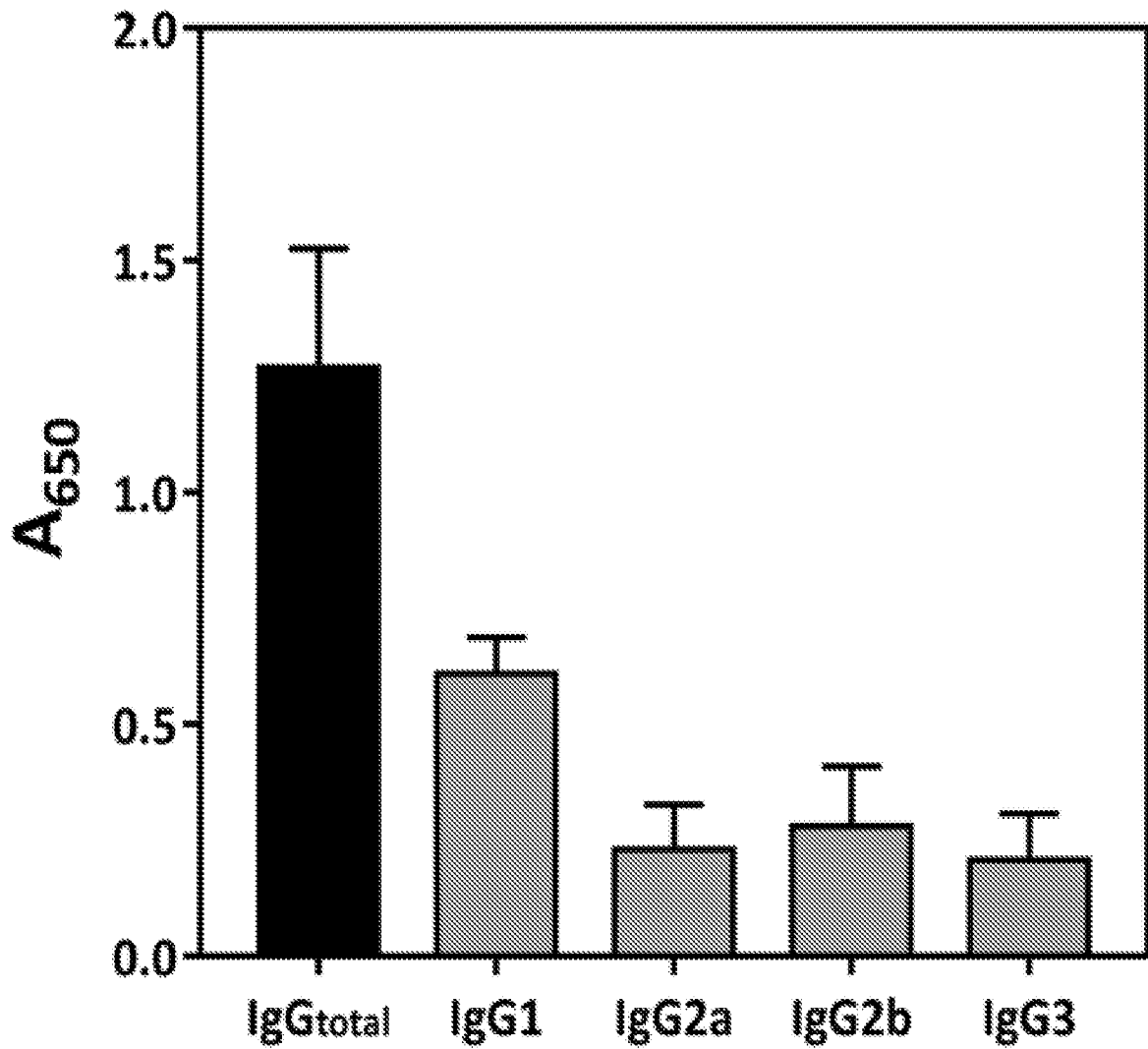
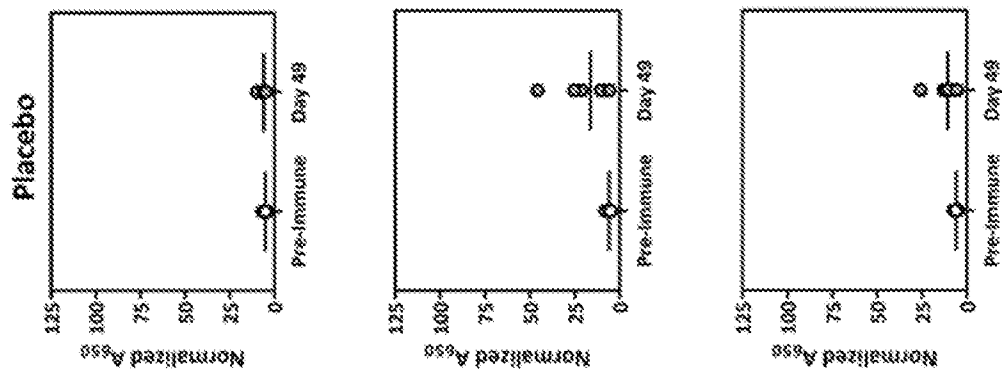
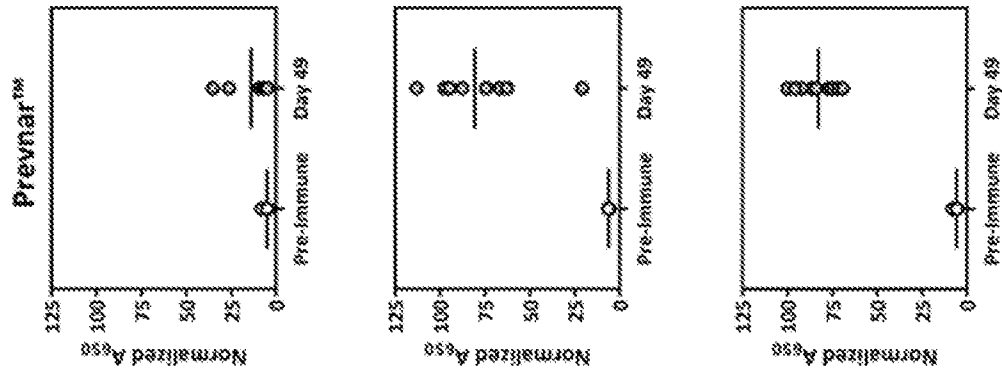
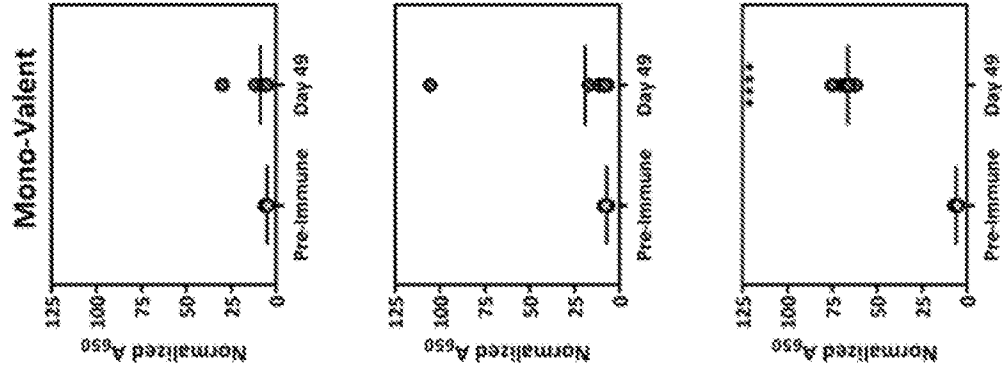
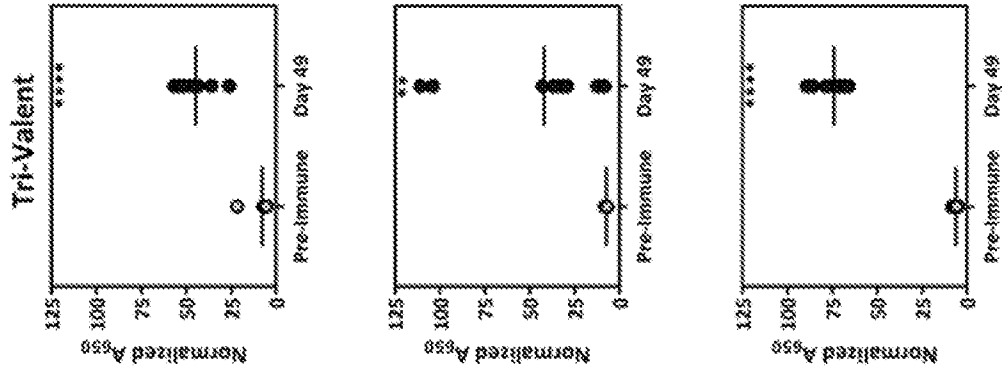
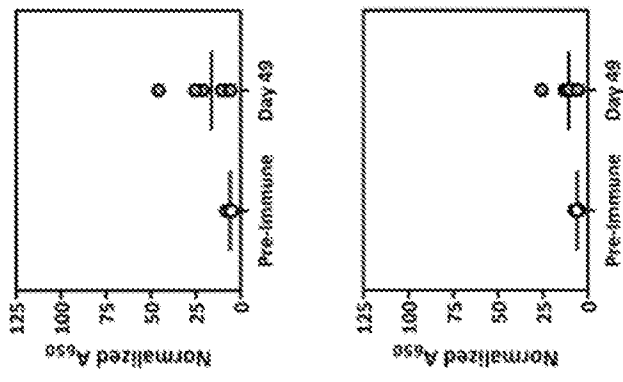
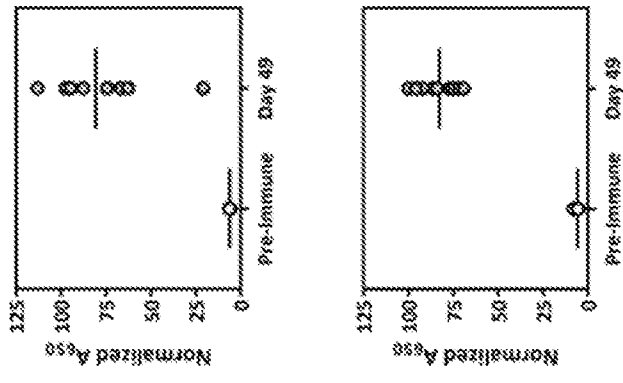
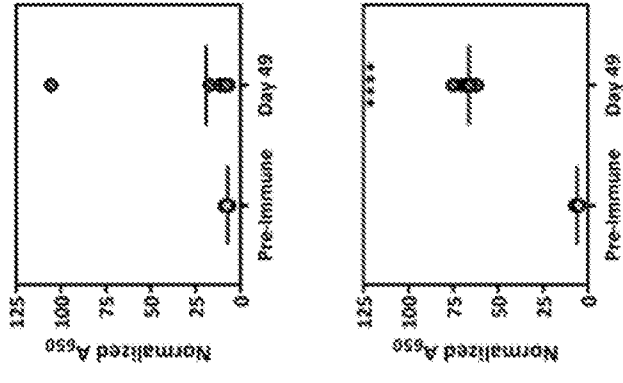
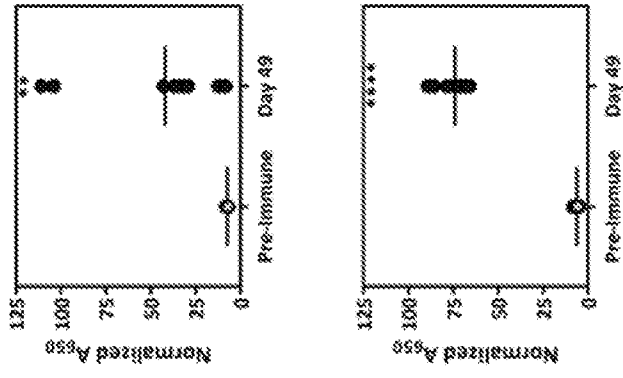


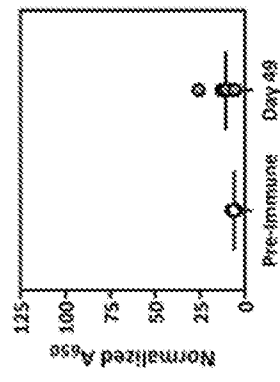
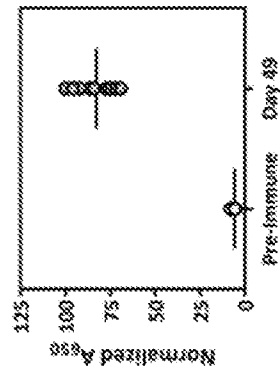
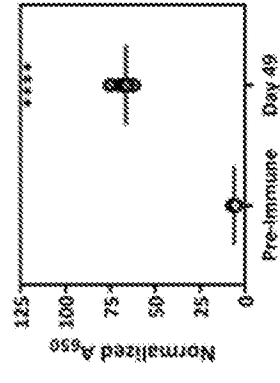
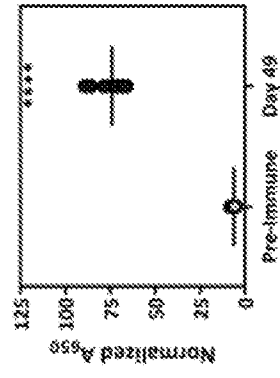
Figure 12C



*S. pneumoniae* 8



*S. pneumoniae* 9v



*S. pneumoniae* 14

Figure 13

>AAC45886.1 ComP $\Delta$ 28 [*Acinetobacter* sp. ADP1]  
AYTDYTVR~~ARV~~SEGLTAASSMKTTVSENILNAGA  
LVAGTPSTAGSSCVGQEI~~SAS~~**S**NATTNVATATCG  
ASSAGQIIVTMDTTKAKGANITLTPTYASGAVTWK  
CTTSDKKYVPSECRG (SEQ ID NO: 7)

>ENV58402.1 hypothetical protein F951\_00736  
 $\Delta$ 28 [*Acinetobacter soli* CIP 110264]  
AYTDYTVR~~SRV~~TEGLTTASAMKATVSENIMNAGG  
TSMPSGGNCTGVTQIASGAS**A**AATTNVASAQCSD  
SDGVITVTMTDKAKGVSILTPSFSSTGSVGWKC  
TTSSDKKYVPSECRGT (SEQ ID NO: 8)

>APV36638.1 competence protein  $\Delta$ 28  
[*Acinetobacter soli* GFJ-2]  
AYTDYTVR~~ARV~~SEGLTTASAMKATVSENILSAGQI  
VTGTPSTANSSCVGQEI**NS**SSTSNVATATCSG  
LGVITVTMDSTKAKGVNLTLTPTYTTSSNAVTWKC  
TTTSDKKYVPSECRN (SEQ ID NO: 9)

**Figure 14**

>PKD82822.1 competence protein  $\Delta$ 28  
[*Acinetobacter radioresistens* 50v1]  
AYTDYTVRARVTEAVSTASSMKATVSENIMNAGG  
TQIPTSGNCVGVQTIAASNATKNVATATCTDSTG  
VIVVTTTPAAKSVPLTLTPTYTGGNVKWACSTTA  
NFKNYVPSECRS (SEQ ID NO: 10)

>SNX44537.1 type IV pilus assembly protein Pila  
 $\Delta$ 28 [*Acinetobacter puyangensis* ANC 4466]  
AYTDYTVRARVTEALTTASAMKATVSENIMSAGG  
TTIASSACNGVISASATTNVASSACSGSGVISVTT  
TAAAKGIVLTLTPKYTGGNVAWQCTTTSGDAQKY  
VPSECRRTS (SEQ ID NO: 11)

>OAL75955.1 competence protein  $\Delta$ 28  
[*Acinetobacter* sp. SFC]  
AYTDYTVRAKVTEAISTASAMKATVSENLMMSAGG  
TSIVSTNANCAGVETIGASSNKTKNVESAACTAAT  
GVILVTTTAEAKSVPLTLKPTYTGSNVQWKCGTT  
AAAFKYVPSECRNDSSGTGF (SEQ ID NO: 12)

**Figure 14 Cont.**

ADP1	CVGVQEIIS--AS	SNATTNVATATC	(SEQ	ID	NO:	13)
110264	CTGVTQIASGAS	AATTNVASAQC	(SEQ	IS	NO:	14)
GFJ_2	CVGVQEIIN--AS	SSSTSNVATATC	(SEQ	ID	NO:	15)
SFC	CAGVETIG--AS	NKTKNVEAAC	(SEQ	ID	NO:	16)
P50v1	CVGVQETIA--AS	NATKNVATATC	(SEQ	ID	NO:	17)

Figure 15

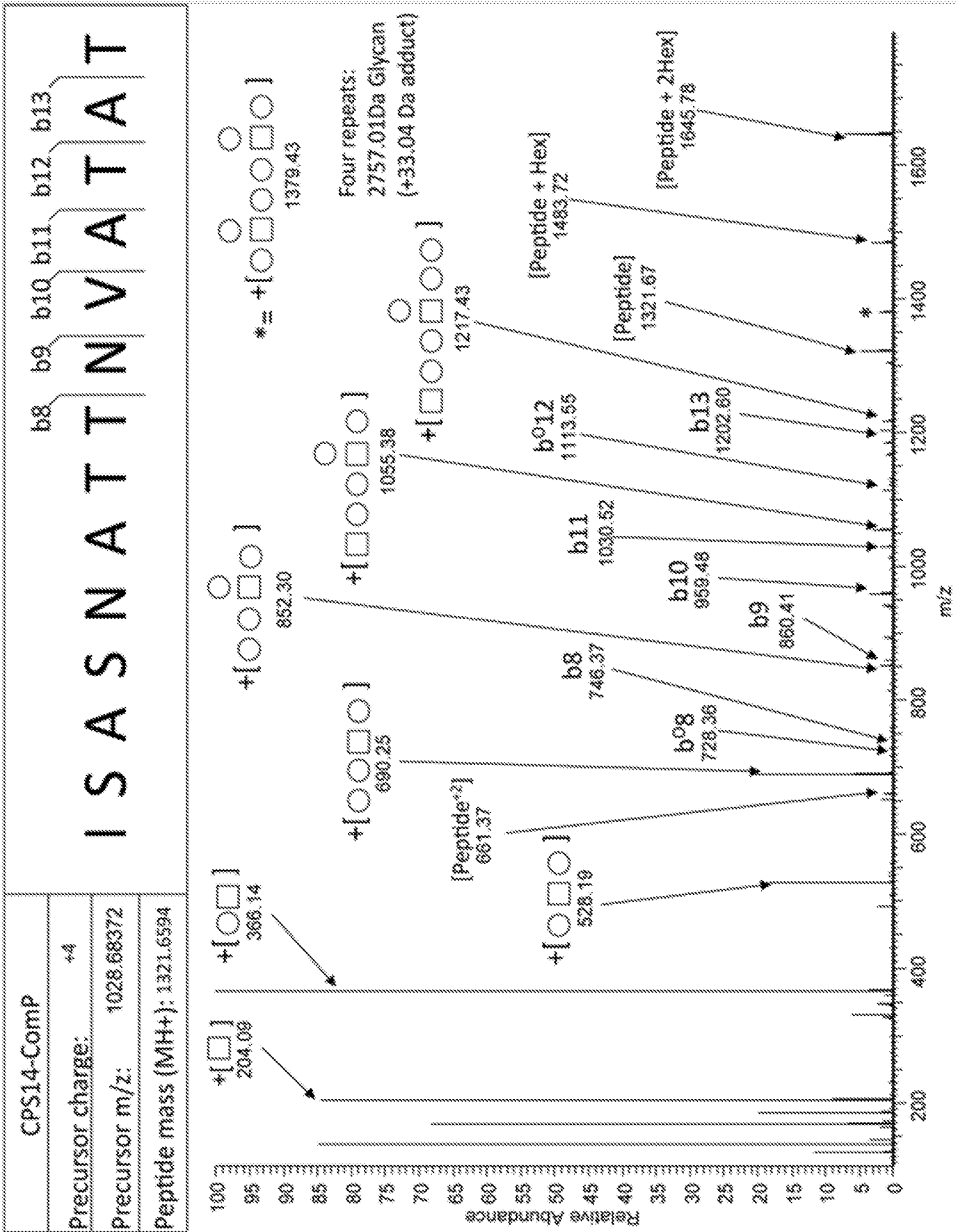


Figure 16

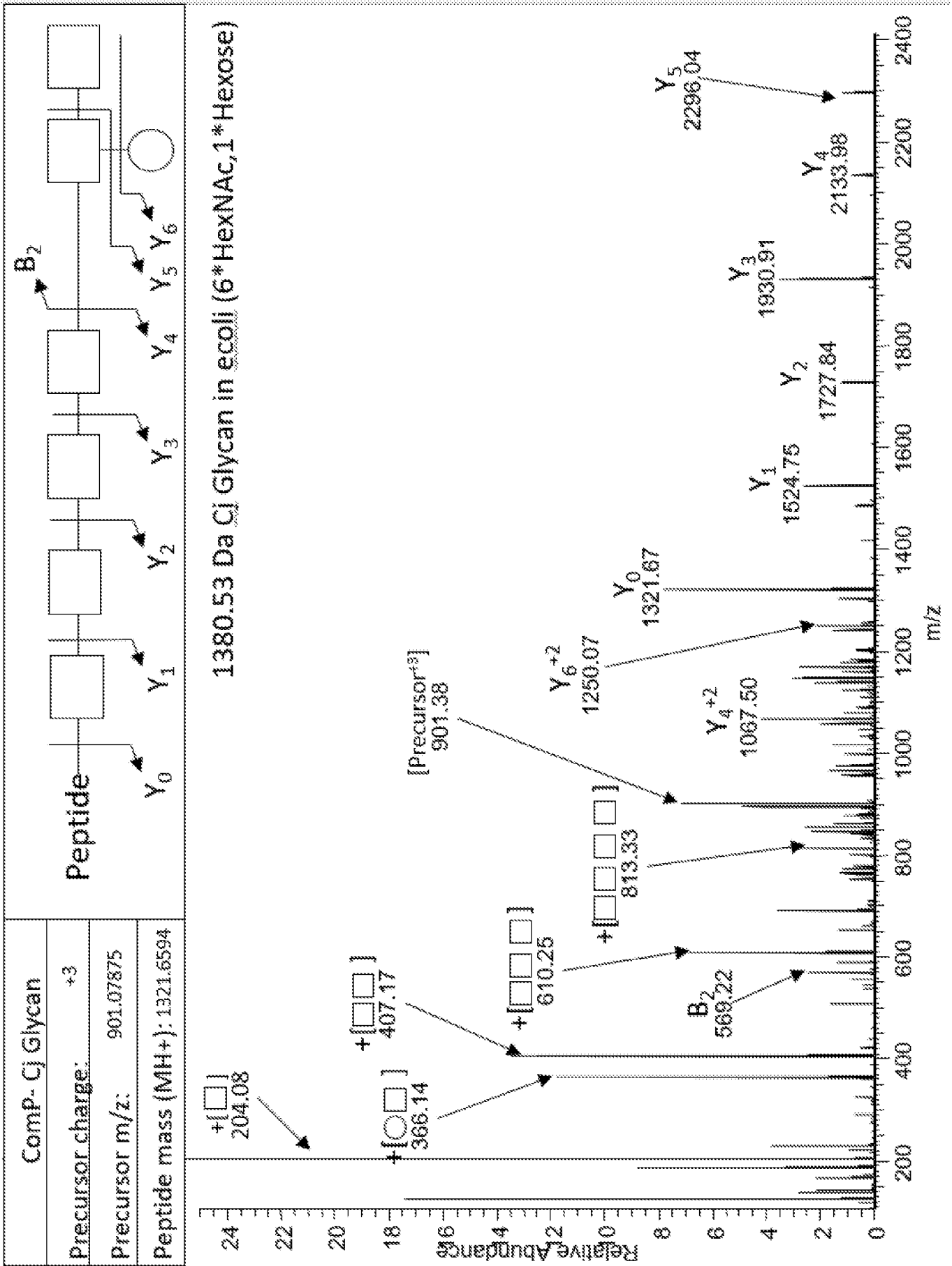
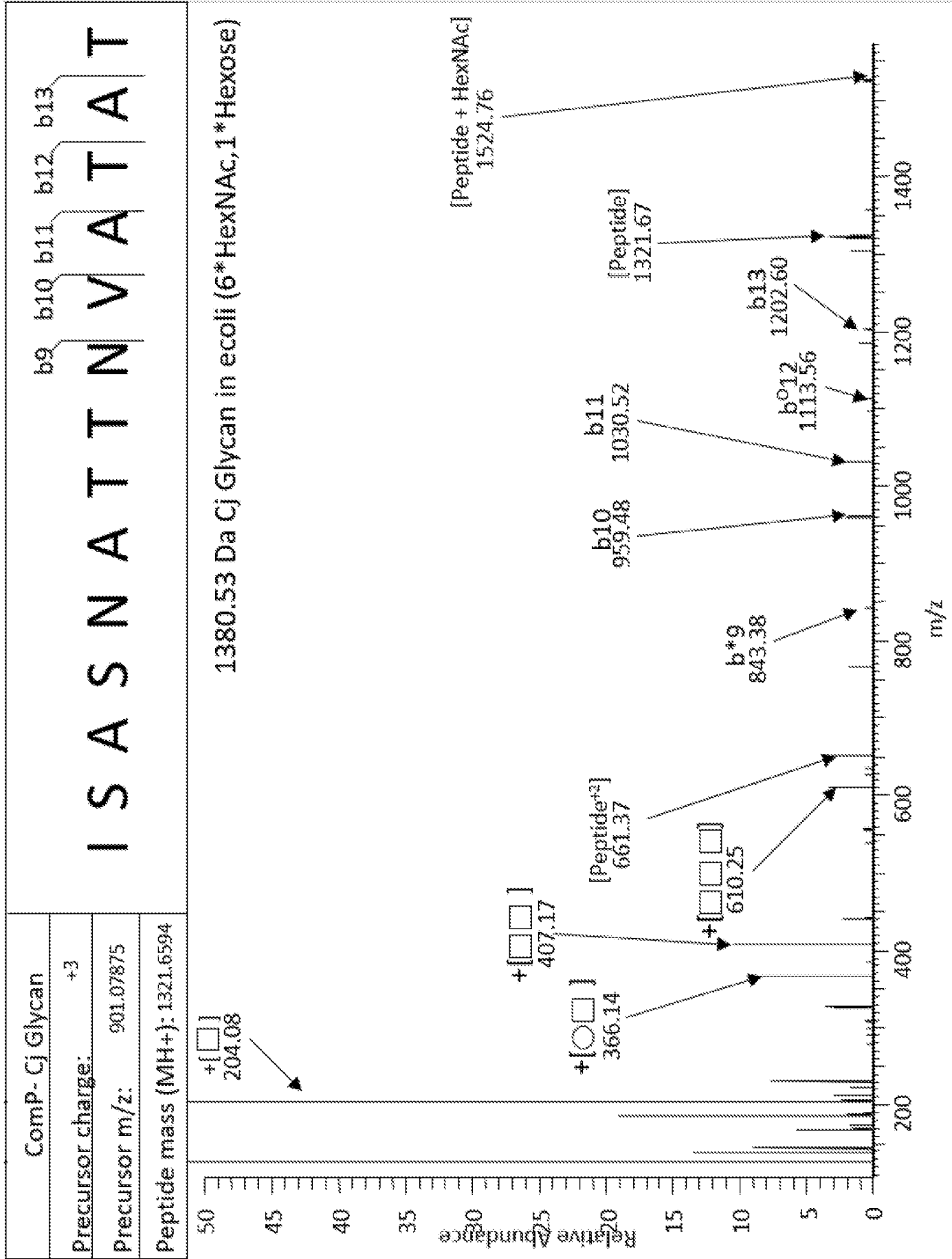


Figure 17



**Figure 18**

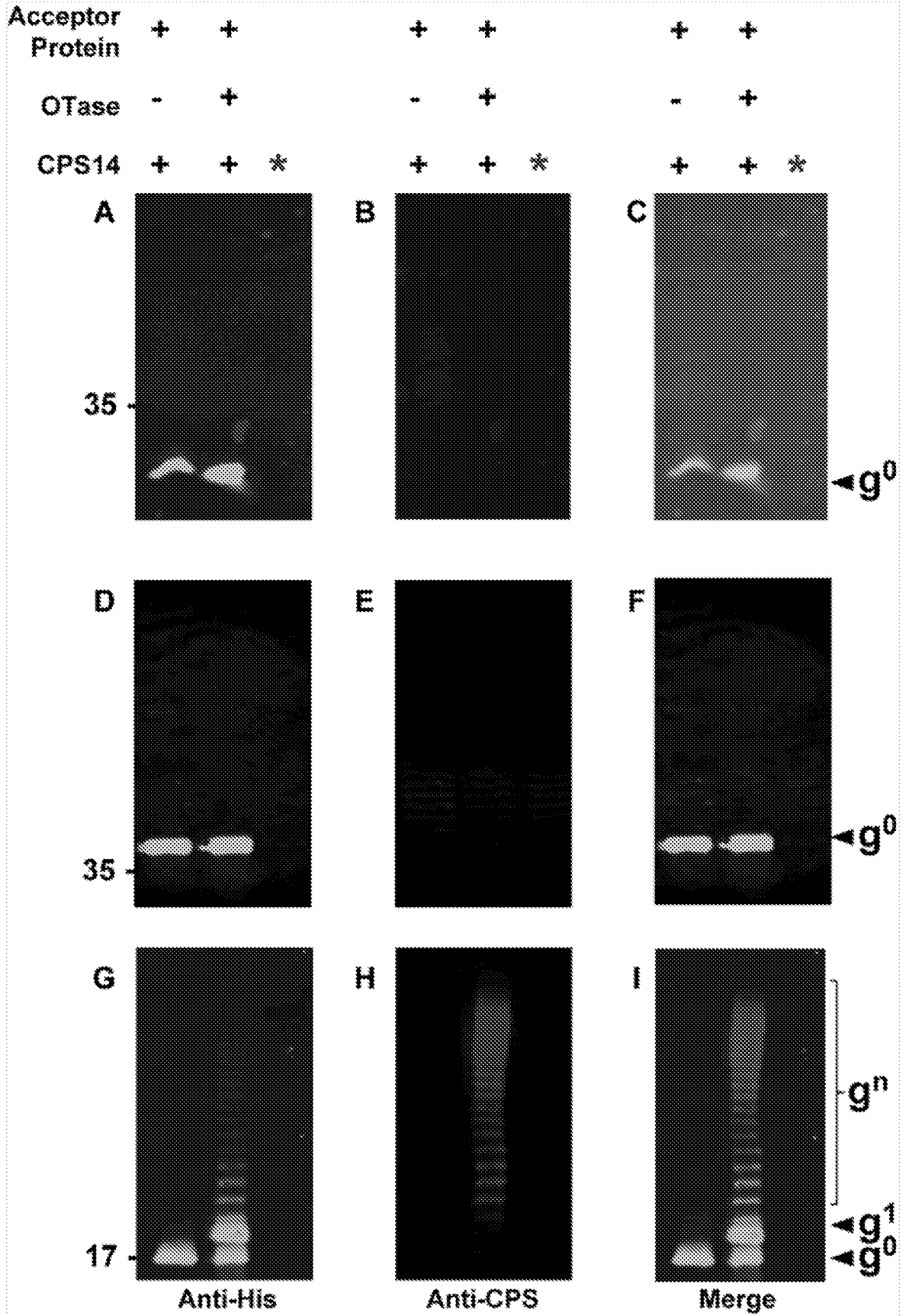


Figure 19A-I

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)

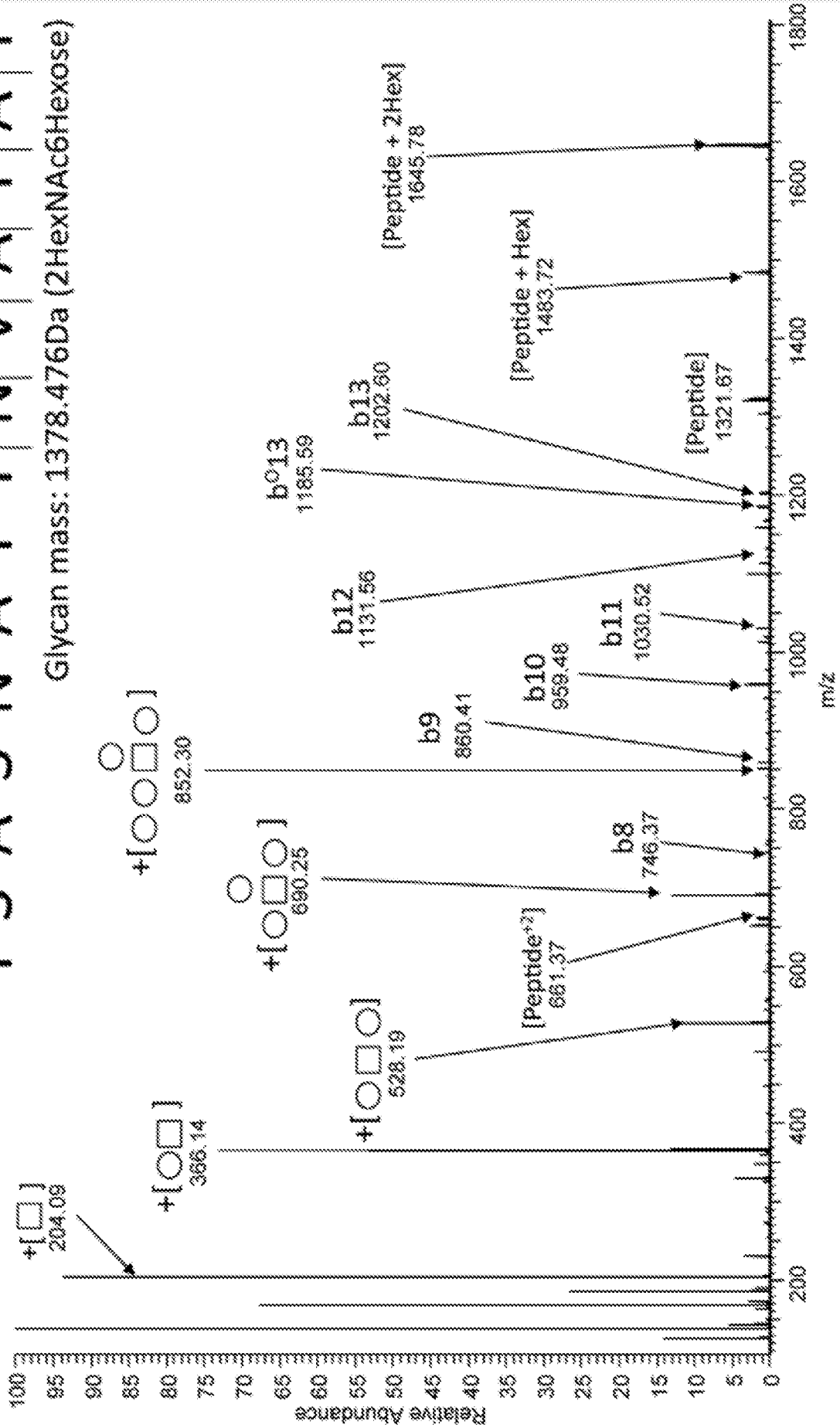


Figure 20A

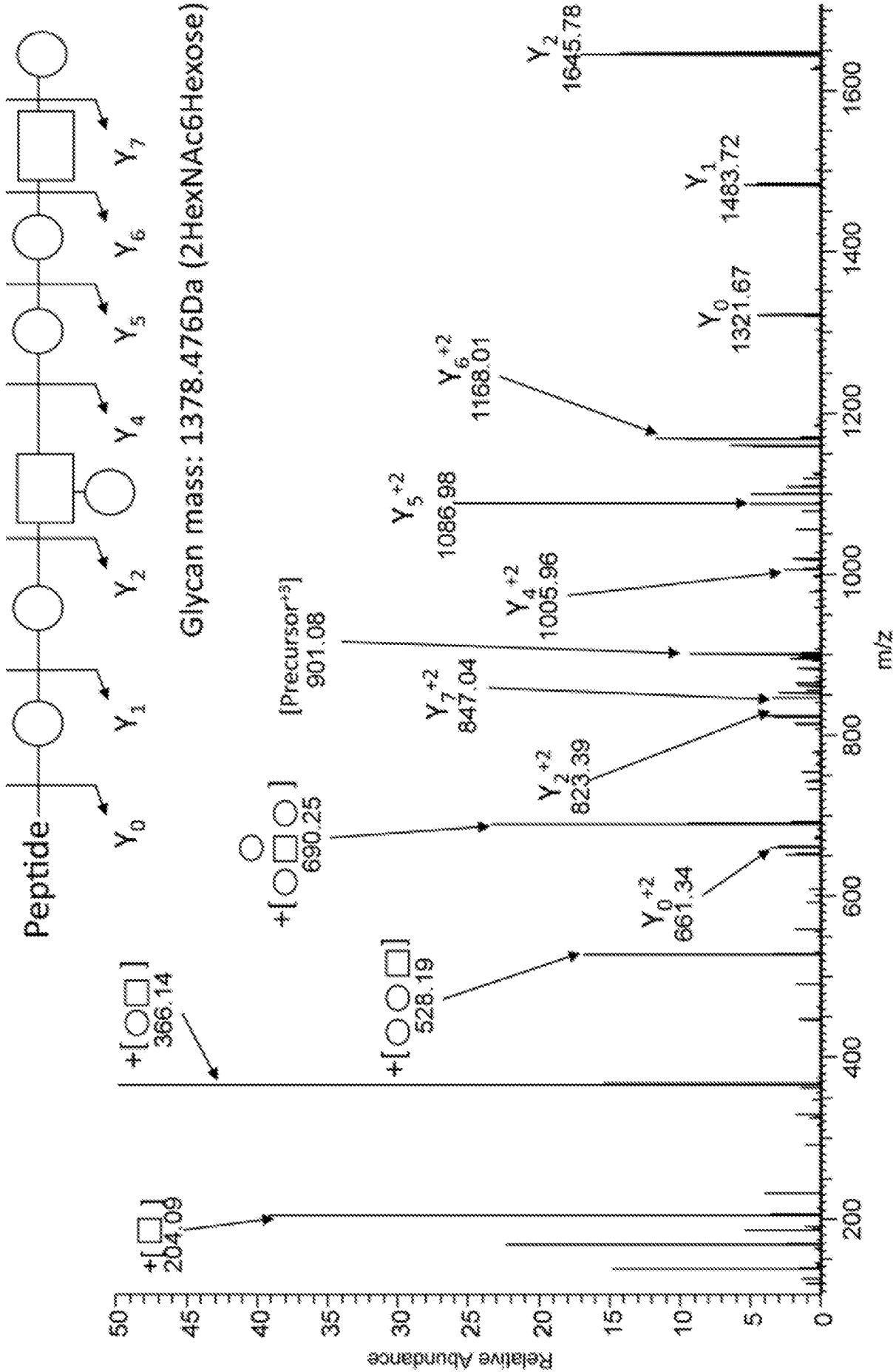


Figure 20B

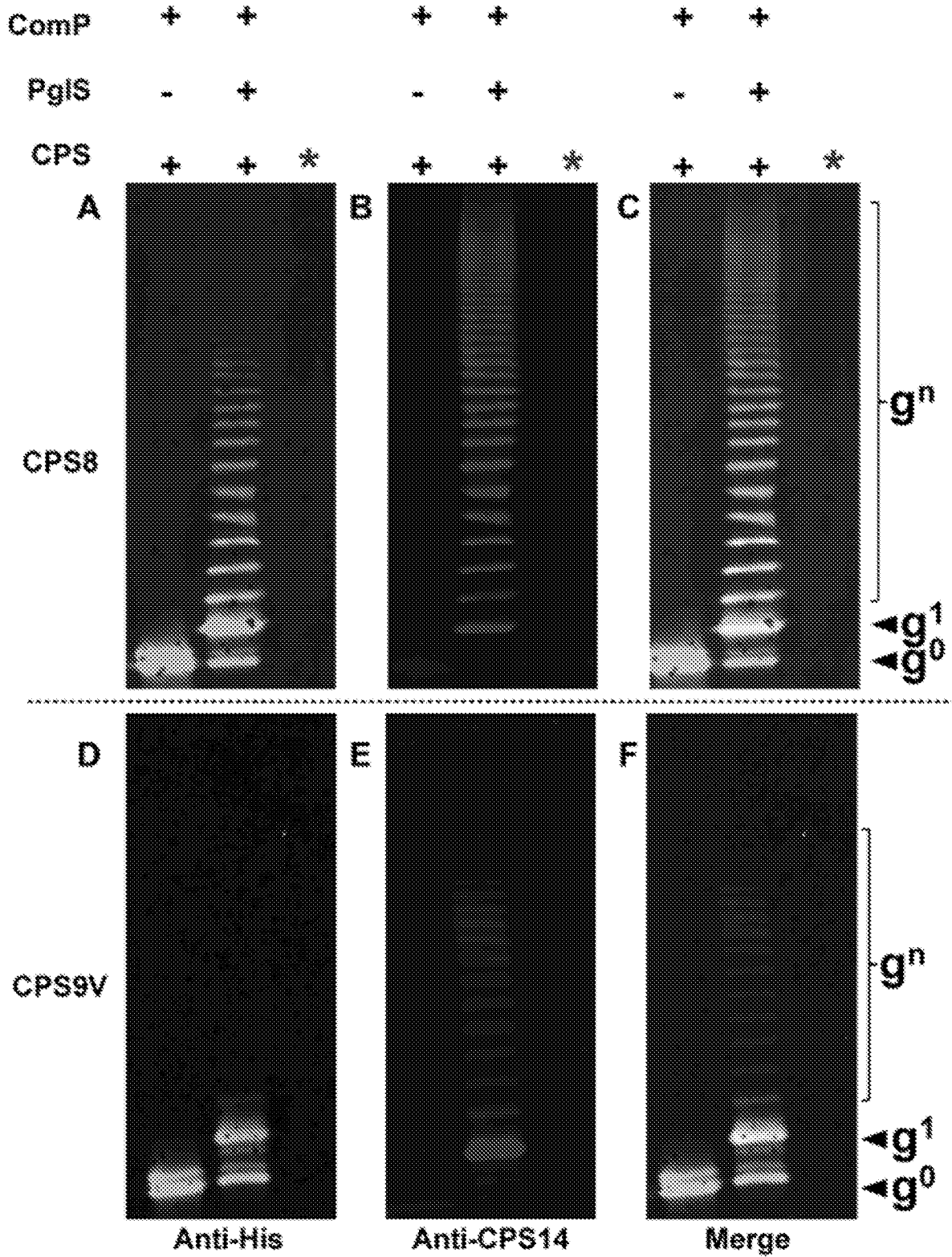


Figure 21A-F

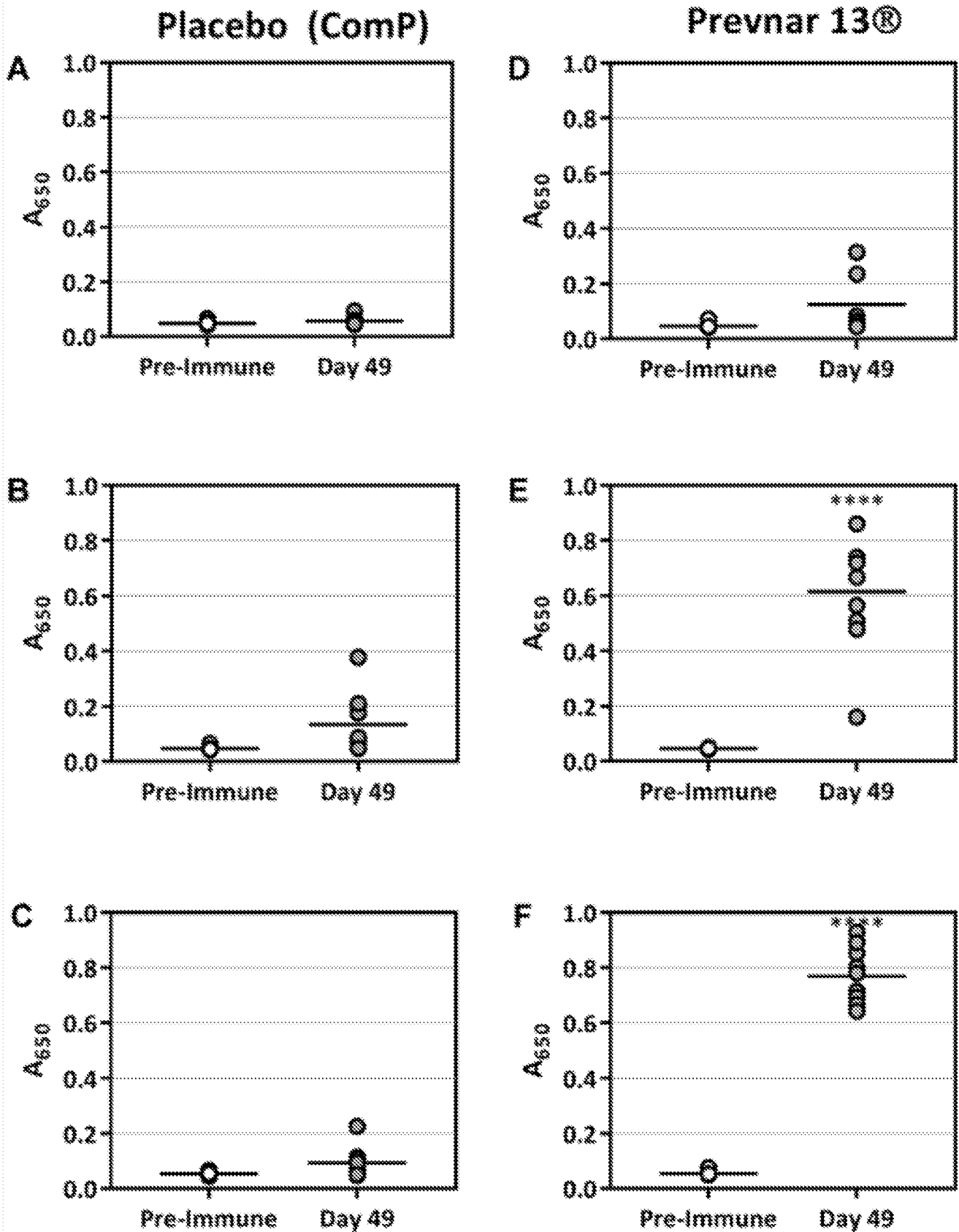


Figure 22A-F

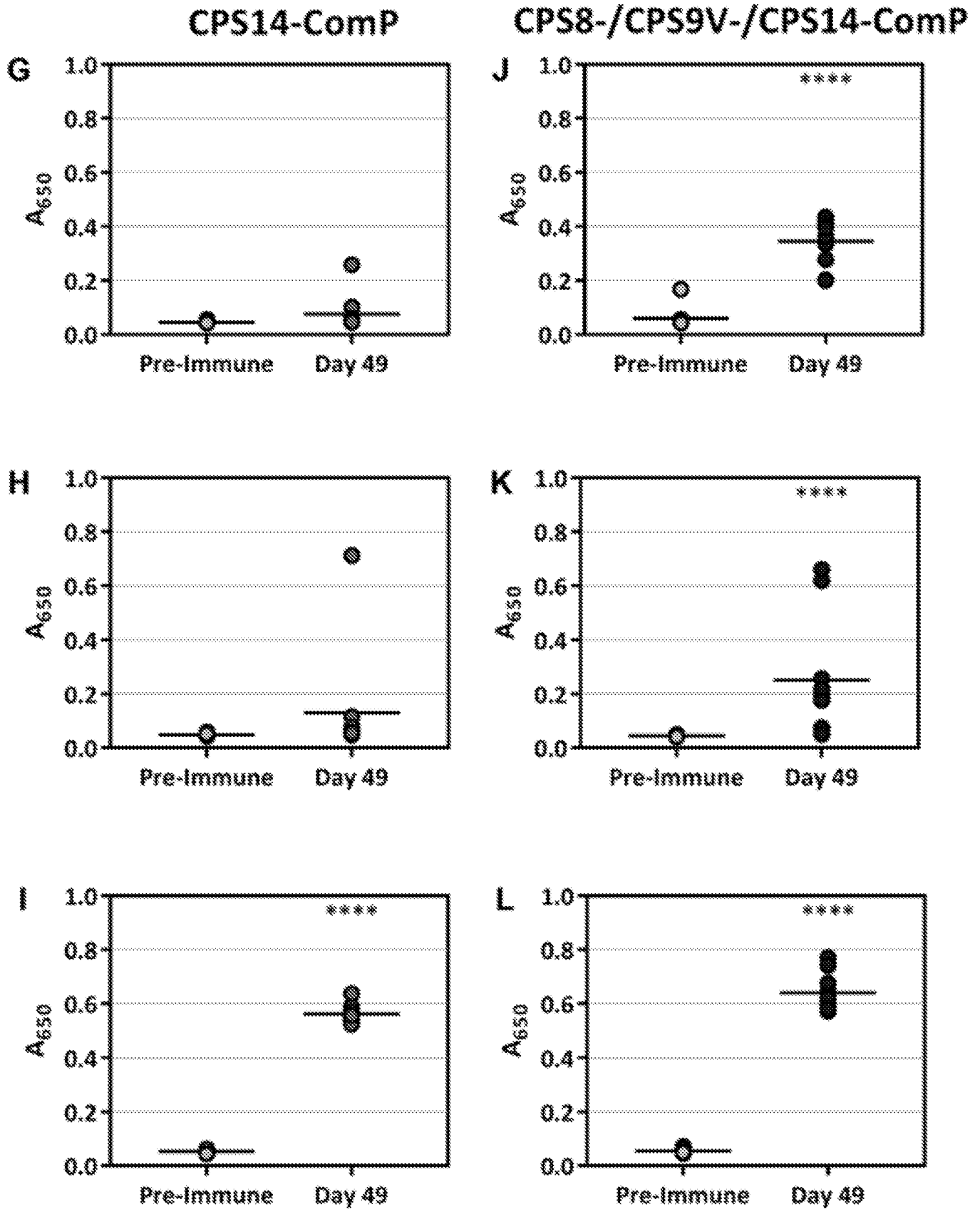
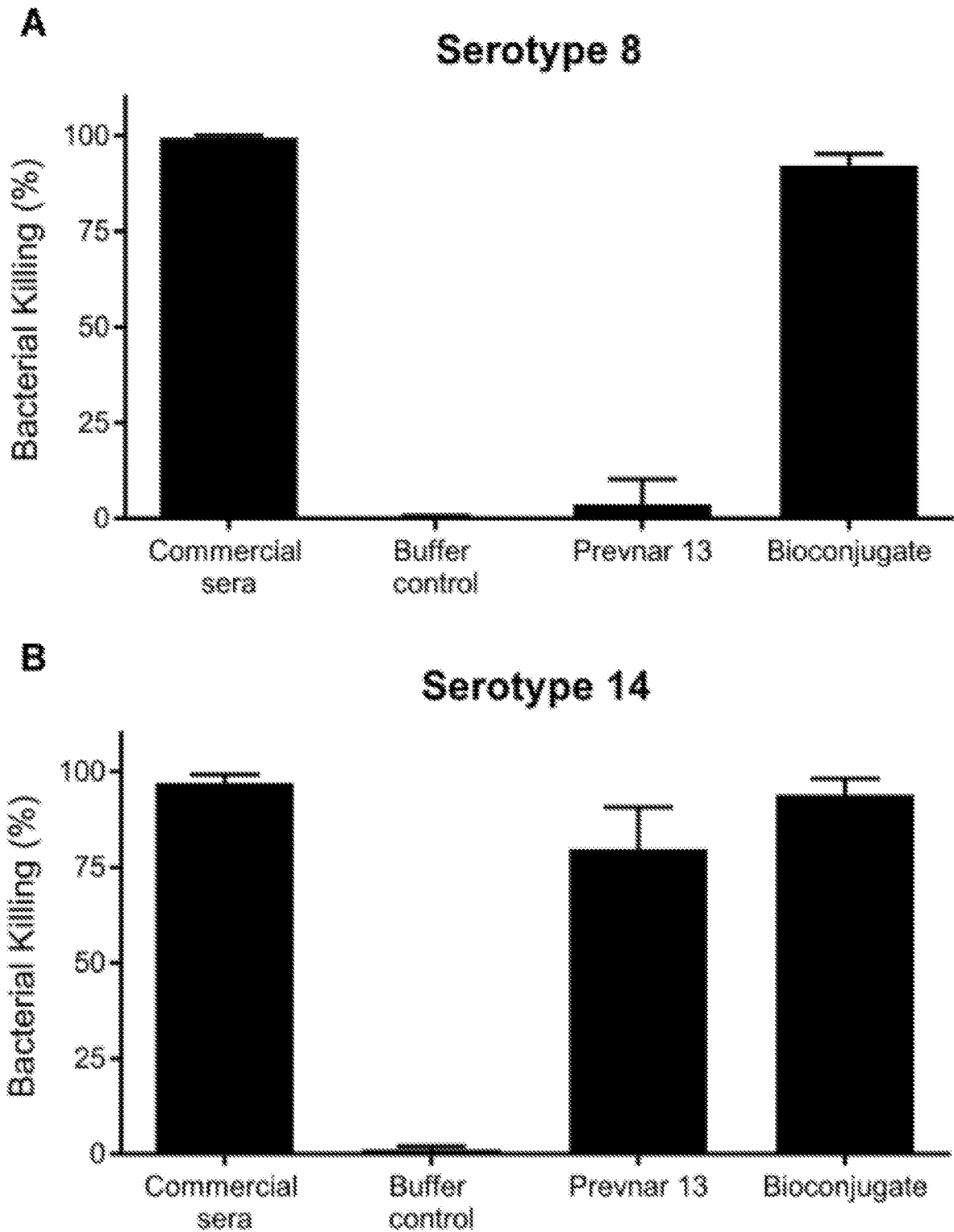


Figure 22G-L

**Figure 23A,B**

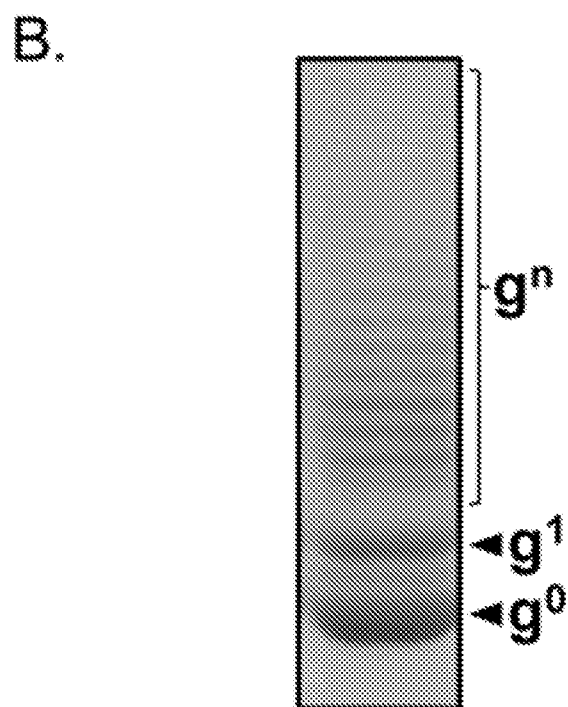
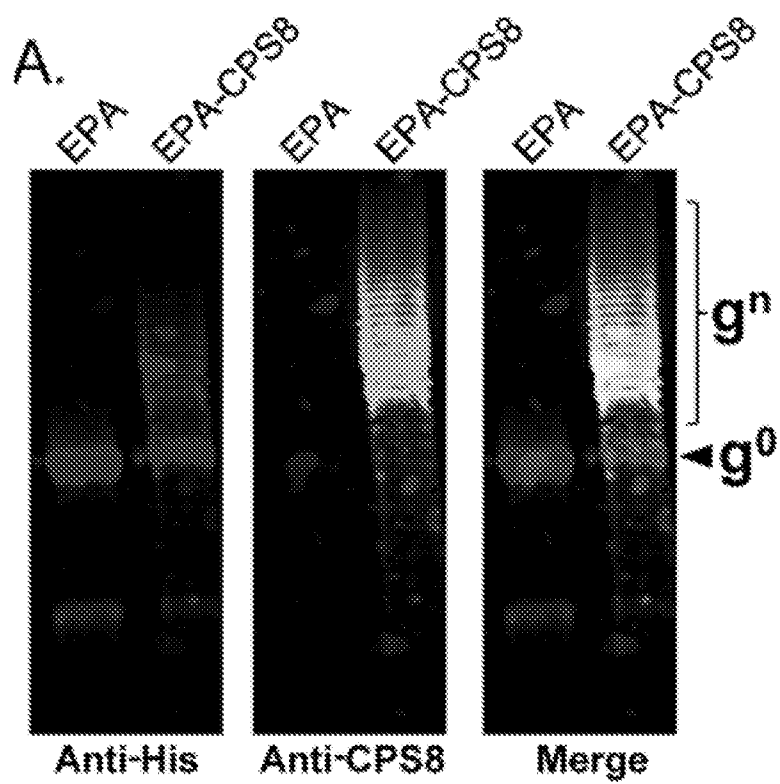


Figure 24A,B

C.

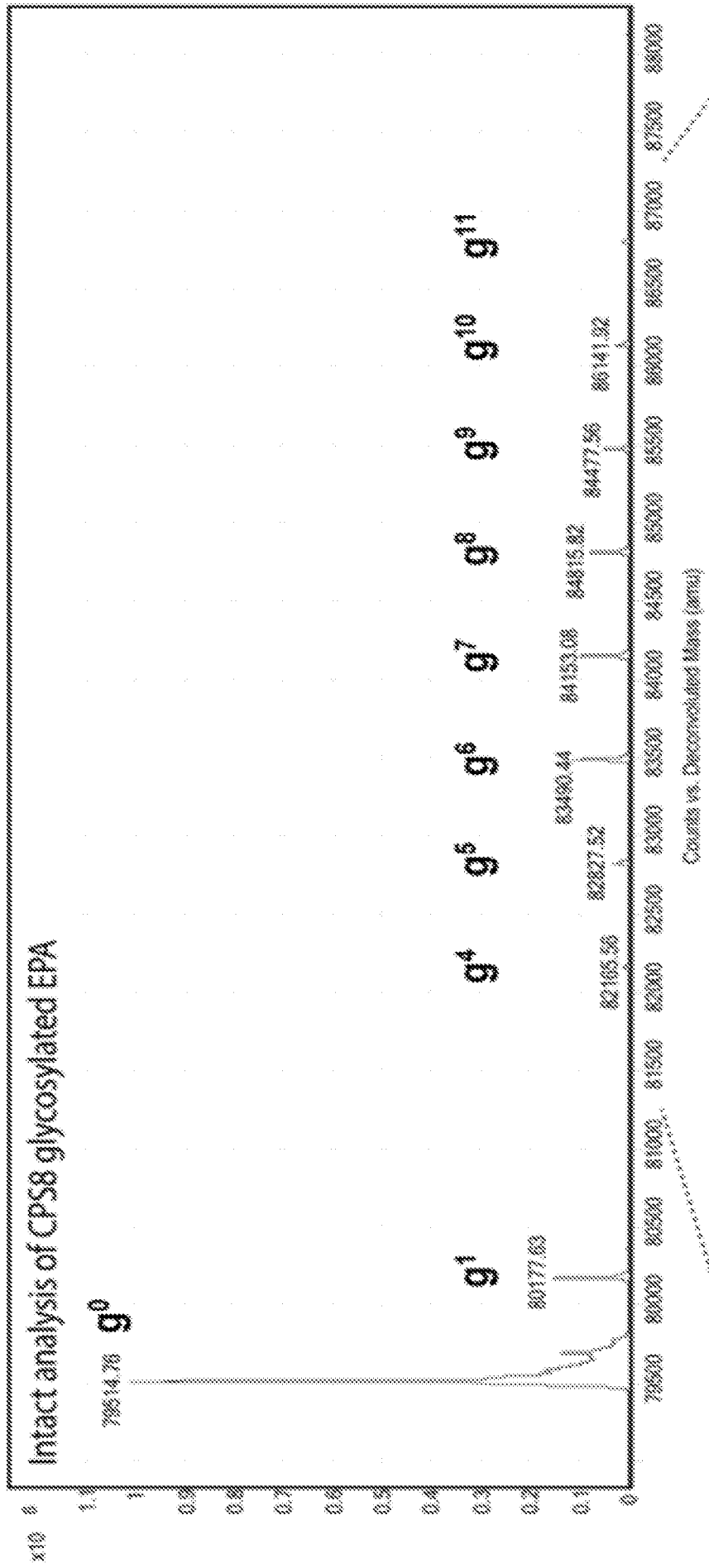


Figure 24C

D.

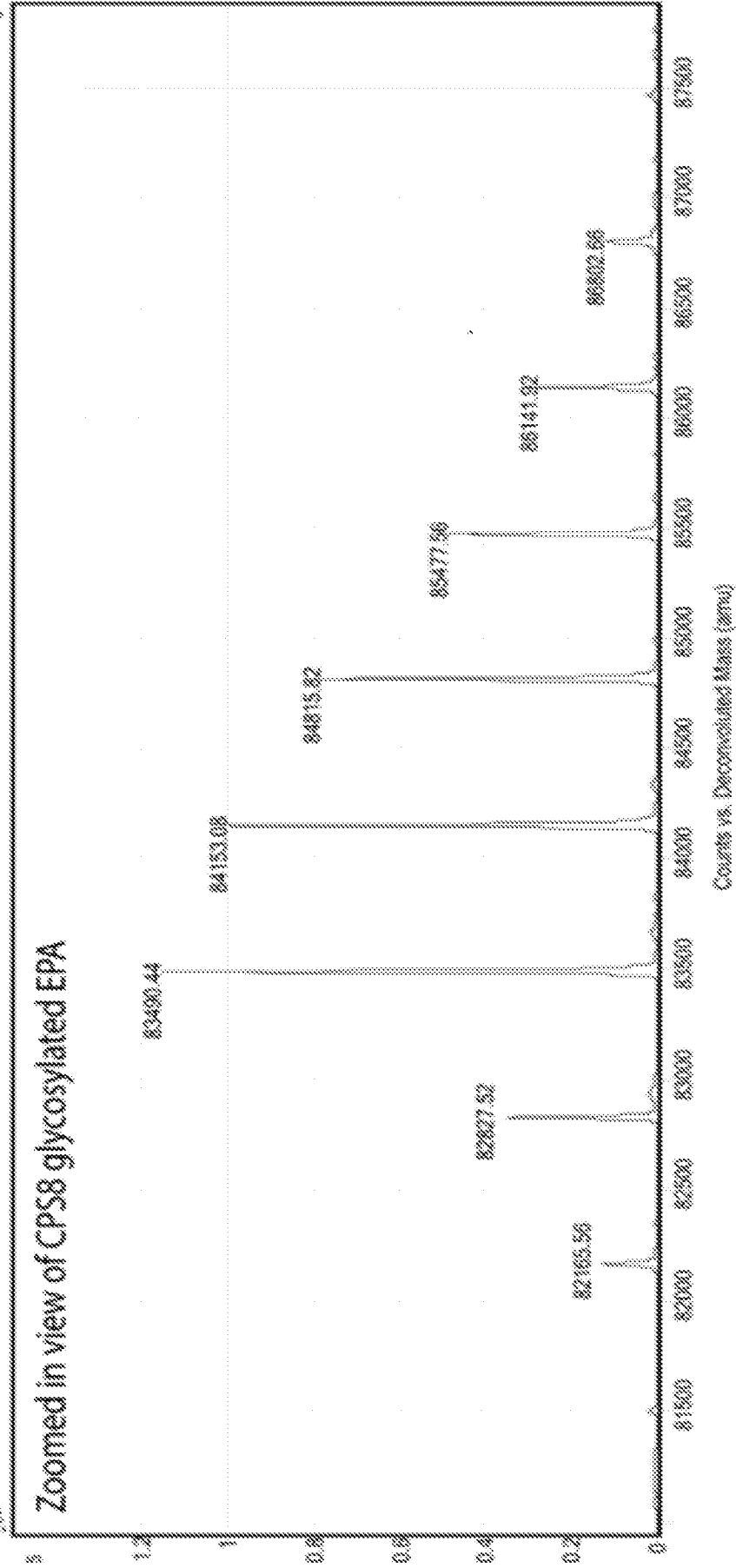


Figure 24D



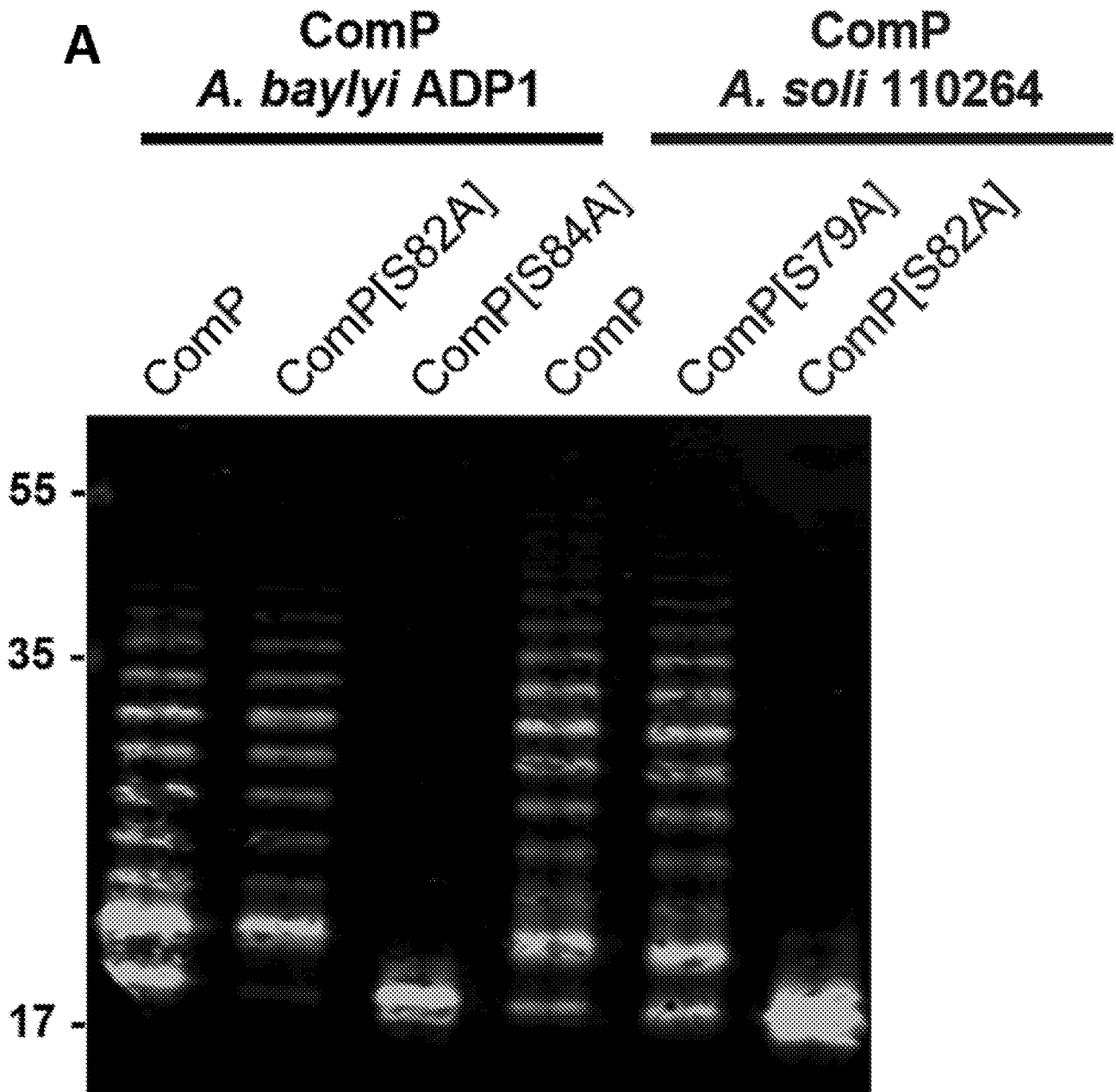


Figure 26A

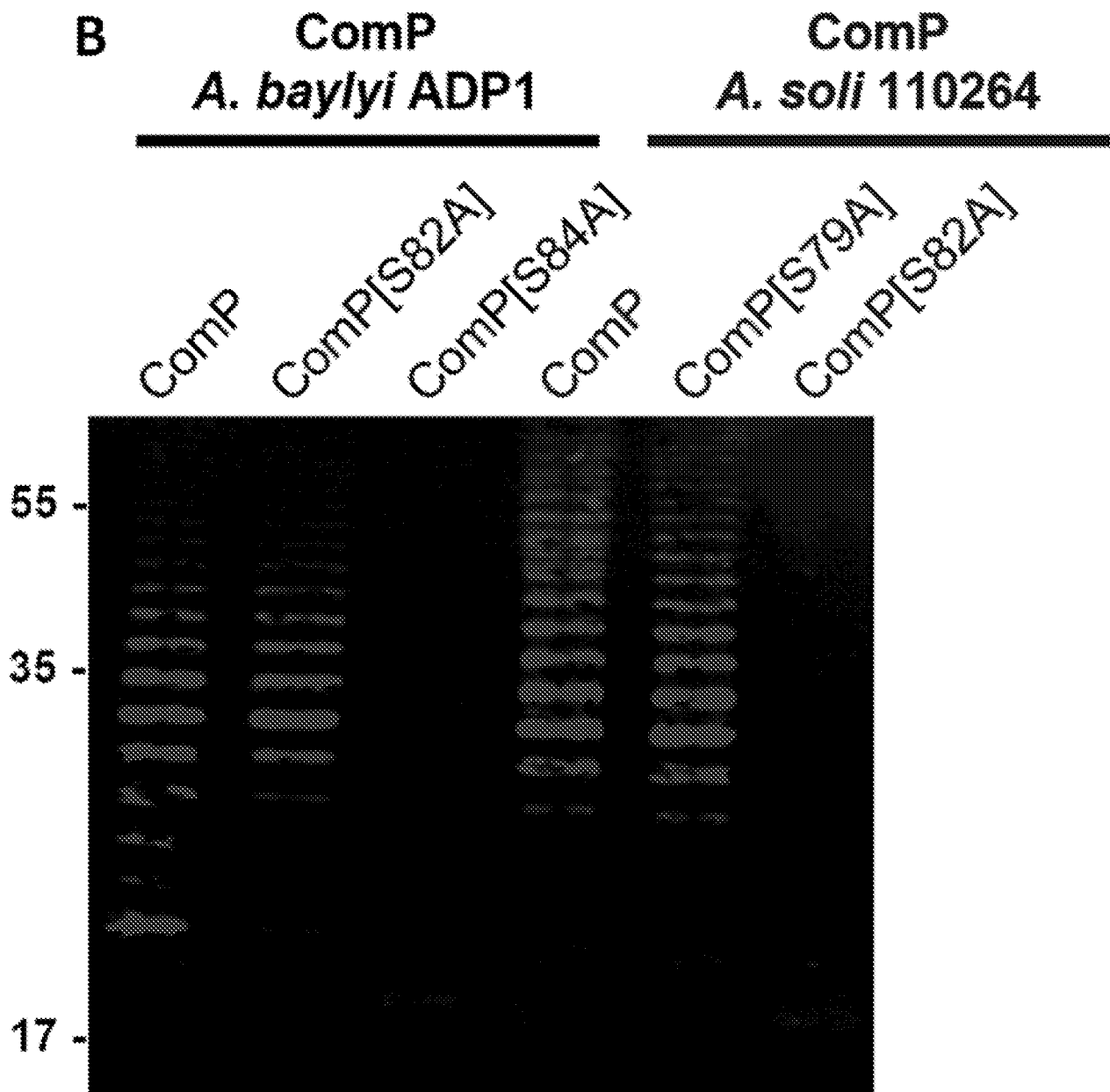


Figure 26B

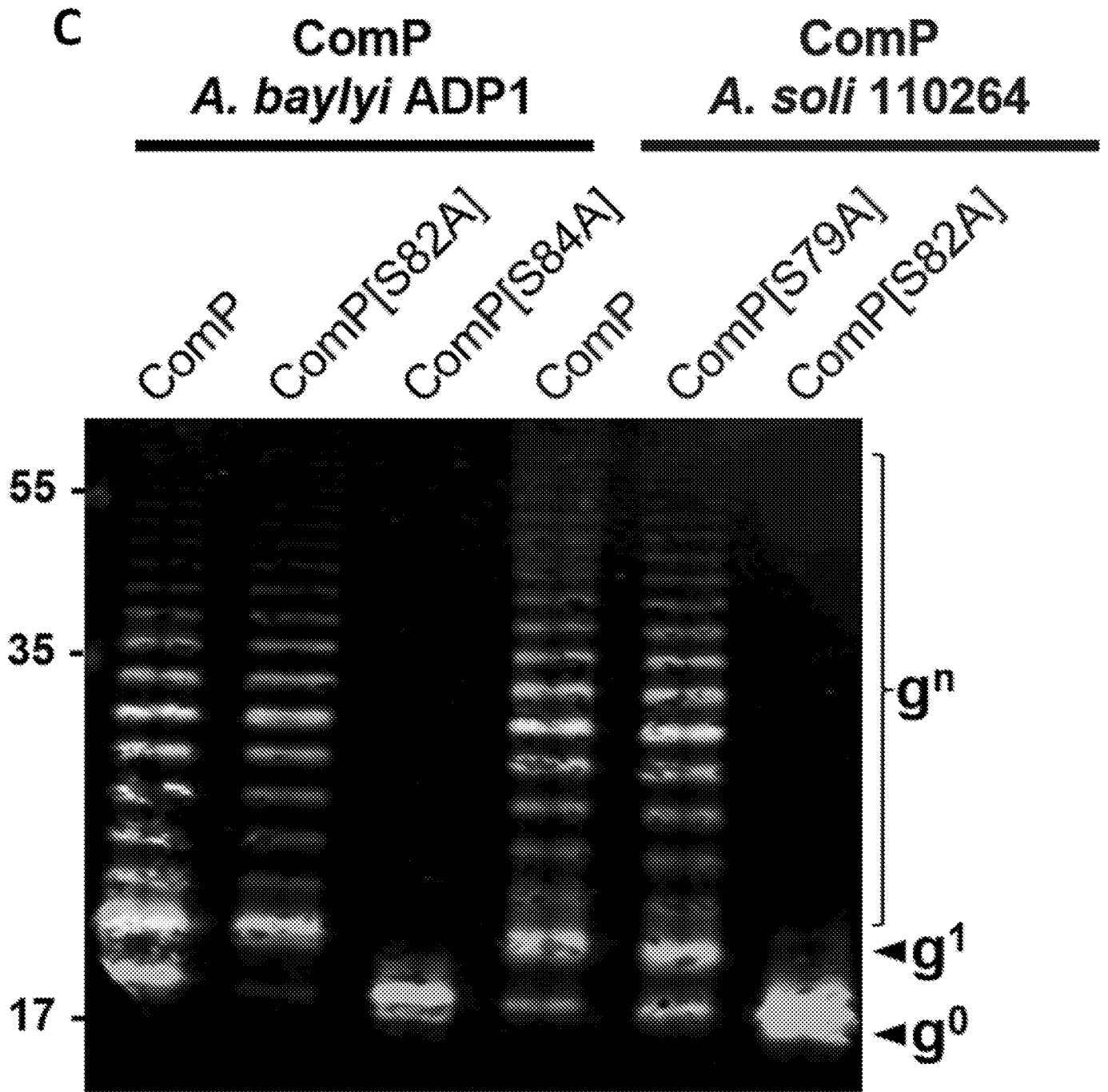


Figure 26C

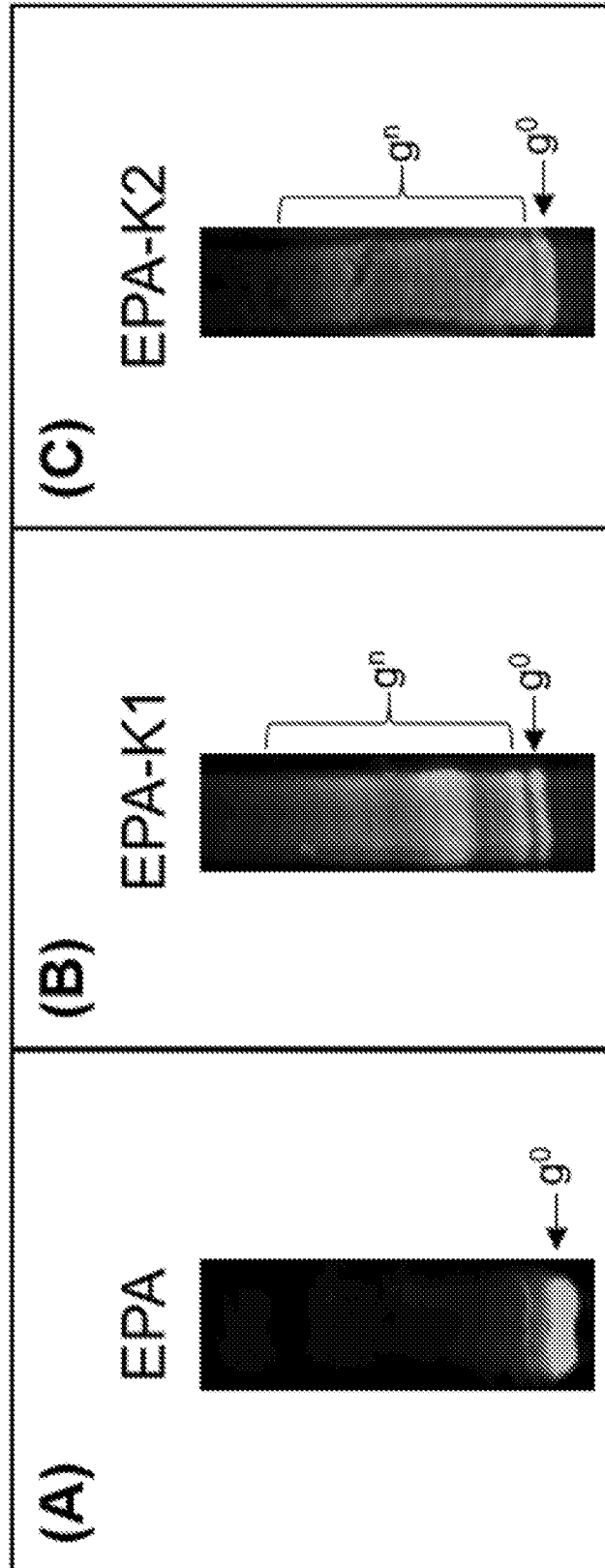


Figure 27A,B,C

A

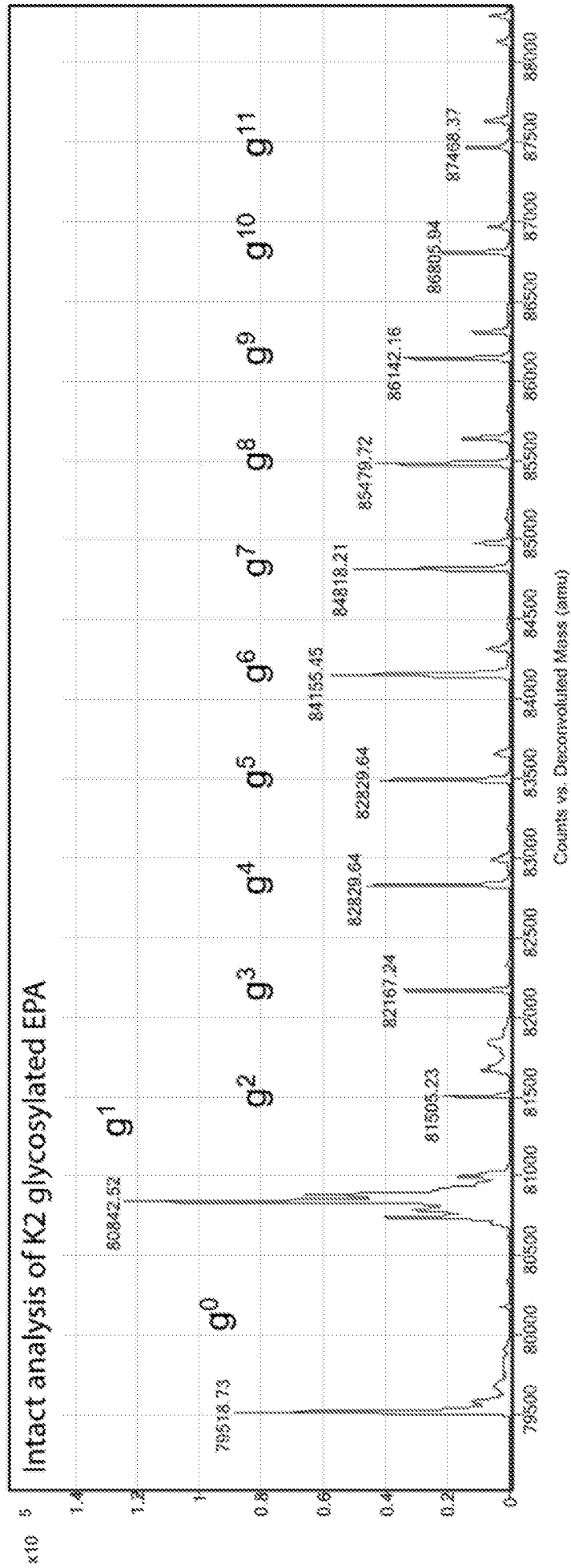
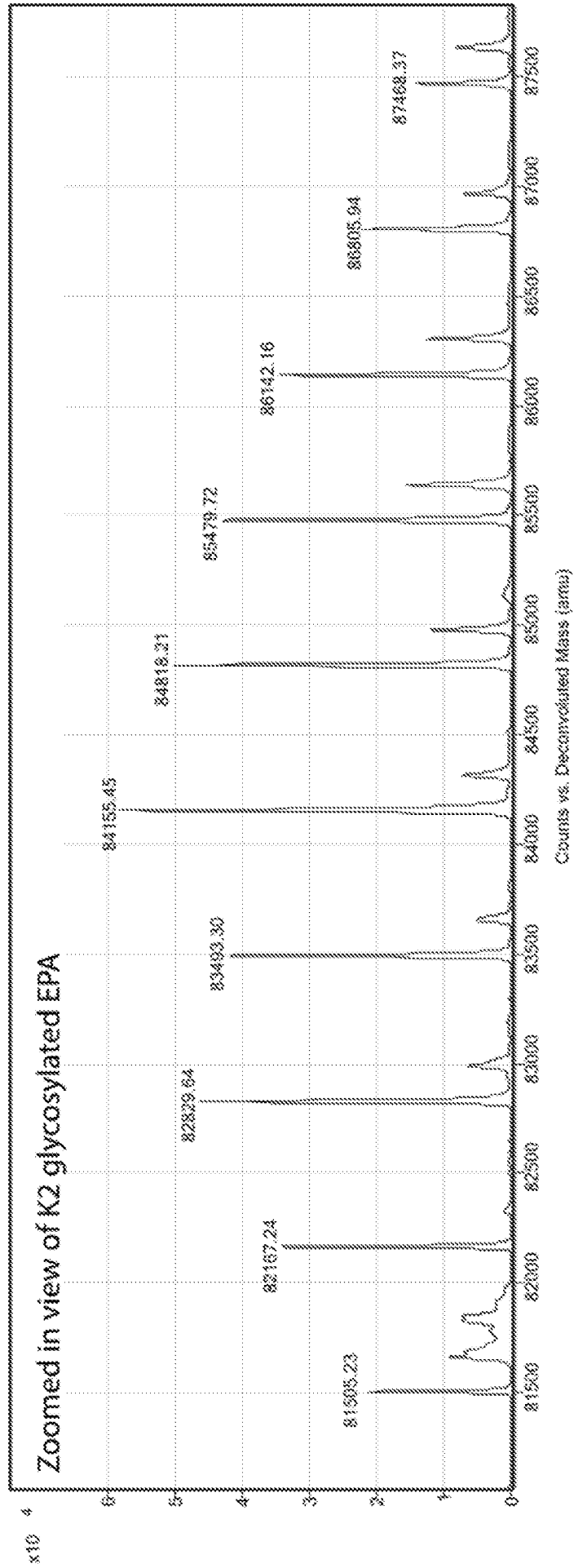


Figure 28A

**B**



**Figure 28B**