Author/s:
Veith, PDD;Gorasia, DGG;Reynolds, ECC

Title:
Characterization of the O-Glycoproteome of Flavobacterium johnsoniae

Date:
2023-06

Citation:
Veith, P. D. D., Gorasia, D. G. G. & Reynolds, E. C. C. (2023). Characterization of the O-Glycoproteome of Flavobacterium johnsoniae. Journal of Bacteriology, 205 (6), https://doi.org/10.1128/jb.00093-23.

# Characterization of the O-Glycoproteome of *Flavobacterium johnsoniae*

Veith PD[#], Gorasia DG, Reynolds EC[#]

Oral Health Cooperative Research Centre, Melbourne Dental School, Bio21 Institute, The University of Melbourne, Parkville, Victoria 3010, Australia.

[#]Correspondence: Paul Veith (pdv@unimelb.edu.au); Eric Reynolds (e.reynolds@unimelb.edu.au)

**Keywords**
*Flavobacterium johnsoniae*, O-glycosylation, Glycoproteins, Proteome, Glycans, Mass spectrometry

## Abstract

*Flavobacterium johnsoniae* is a free-living member of the Bacteroidota phylum found in soil and water. It is frequently used as a model species for studying a type of gliding motility dependent on the type IX secretion system (T9SS). O-glycosylation has been reported in several Bacteroidota species and the O-glycosylation of S-layer proteins in *Tannerella forsythia* was shown to be important for certain virulence features. In this study we characterised the O-glycoproteome of *F. johnsoniae* and identified 325 O-glycosylation sites within 226 glycoproteins. The structure of the major glycan was found to be a hexasaccharide with the sequence Hex–(Me-dHex)–Me-HexA–Pent–HexA–Me-HexNAcA. Bioinformatic localisation of the glycoproteins predicted 68 inner membrane proteins, 60 periplasmic proteins, 26 outer membrane proteins, 57 lipoproteins and 9 proteins secreted by the T9SS. The glycosylated sites were predominantly located in the periplasm where they are postulated to be beneficial for protein folding/stability. Six proteins associated with gliding motility or the T9SS were demonstrated to be O-glycosylated.

## Importance

*Flavobacterium johnsoniae* is a Gram-negative bacterium found in soil and water. It is frequently used as a model species for studying gliding motility and the type IX secretion system (T9SS). In this study we characterised the O-glycoproteome of *F. johnsoniae* and identified 325 O-glycosylation sites within 226 glycoproteins. The glycosylated domains were mainly localised to the periplasm. The function of O-glycosylation is likely related to protein folding and stability, and therefore the significance of finding the glycosylation sites has relevance for studies involving expression of the proteins. Six proteins associated with gliding motility or the T9SS were demonstrated to be O-glycosylated which may impact the structure and function of these components.

## Introduction

Bacteria within the *Bacteroidota* phylum (formerly *Bacteroidetes*) are known or predicted to possess an O-glycosylation system (1), which was first described in *Bacteroides fragilis* (2). The O-glycosylation motif originally described for *B. fragilis* is D(S,T)(A,L,V,I,M,T), with glycosylation occurring on the Ser or Thr. A search within proteins predicted to be localised outside of the cytoplasm of 266 different *Bacteroidota* species found that around 50% typically contained the motif. For the larger genomes, this commonly equated to around 1000 predicted O-glycosylated proteins (1). Recent glycoproteomic studies have demonstrated that these predictions are realistic, with 257 putative glycosylation sites identified in 145 *Porphyromonas gingivalis* glycoproteins (out of 322 predicted) and 312 putative glycosylation sites in 145 *Tannerella. forsythia* glycoproteins (out of 557 predicted) (3, 4). These studies extended the O-glycosylation motif to D(S/T)(A/I/L/V/M/T/S/C/G/F).

The structure of the O-glycans have been elucidated in part or in whole for *B. fragilis* (5, 6), *Elizabethkingia meningoseptica* (formerly *Flavobacterium meningosepticum*)(7), two strains of *T. forsythia* (6, 8, 9), *P. gingivalis* (4), and *Flavobacterium columnare* (10) (**Fig. 1**). The O-glycans comprise between 6-10 sugars with the inner part (closest to the protein) exhibiting a degree of conservation across the phylum. An antibody raised to the first two sugars of the *B. fragilis* glycan was apparently able to cross-react with glycoproteins expressed by numerous species across the phylum (1) which is now somewhat surprising since these two sugars are not identical (**Fig. 1**). The biosynthesis of the inner glycans and the oligosaccharyl transferase (ligase) required to link the glycan to the protein is unknown, whereas the biosynthesis of the outer glycans is at least partially documented for *T. forsythia*, *B. fragilis* and *P. gingivalis* (1, 4, 6, 9).

The function of O-glycosylation in *Bacteroidota* is not well understood. The presence of such large numbers of glycoproteins of many different structural and enzymatic classes suggests that the major function of glycosylation is something basic, like promoting folding or stability as originally suggested, and supported by the finding that the protein BF2494 expressed with its glycosylation sites mutated was produced at only low levels (2). Within this system, the most highly glycosylated proteins identified to date are the S-layer proteins of *T. forsythia* with 19 O-glycosylation sites identified in TfsB and 11 O-glycosylation sites identified in TfsA (3). O-glycosylation of these proteins affects the virulence properties of the bacterium including biofilm formation and the elicited immune response (11). In contrast, O-glycosylation in *P. gingivalis* was shown to be biased against

surface proteins, potentially limiting the function of O-glycosylation in that species to protein folding/stability (4).

*F. johnsoniae* is a model organism for T9SS-associated gliding motility (12-14) however O-glycosylation has not yet been reported. In this study we identify 325 O-glycosylation sites within 226 glycoproteins and show that like *P. gingivalis*, O-glycosylation is primarily targeted to protein domains located in the periplasm.

## Methods

### Bacterial culture

Wild-type *Flavobacterium johnsoniae* UW101 was grown to late exponential phase as described previously (15). Briefly, the *F. johnsoniae* cells were grown in CYE medium (10 g/L casitone, 5 g/L yeast extract, 8 mM $MgSO_4$, 10 mM Tris, pH 7.6). Cultures were grown overnight at 30ºC with shaking at 225 rpm.

### Cell Fractionation

Cells (500 mL) were harvested at 10,000 *g* for 30 min and washed once in 50 mM phosphate-buffered saline, PH 7.5 (PBS). The culture fluid was also retained. The washed cells were resuspended in 45 mL PBS and a 30 mL portion was lysed on ice using a sonication probe (model CPX 750, Cole Parmer) fitted with a 6.5 mm tapered microtip. The amplitude was set to 40% and the pulser to 1 s on, 2 s off for a total of 45 min. The membrane fraction was separated from the soluble fraction by centrifugation at 50,000 *g* for 30 min. The supernatant was collected as the soluble fraction, and the pellet was washed once in PBS and retained as the membrane fraction.

### SDS-PAGE and in-gel digestion

For the culture fluid sample, 15 mL was concentrated by ultrafiltration through a 10 kDa cut-off membrane (Amicon Ultra-15, Millipore) at 4 °C to a volume of approximately 1 mL, and then precipitated in 13% trichloroacetic acid (TCA). The TCA precipitated material was washed once with ice-cold acetone. For the soluble fraction, 10 µL was diluted to 1 mL with deionised water and precipitated with TCA and washed with acetone as described above. For the membrane sample, a portion equivalent to 30 µL of cell sonicate was used. All samples were separated by reducing SDS-PAGE and fractionated into 12 gel segments respectively as shown to cover the entire separation range (**Fig 2**). The segments were in-gel digested with trypsin as previously described (16) and extracted once with 1% aqueous TFA and once with 50% acetonitrile–0.1% aqueous TFA, both for 15 min in an ultrasonication bath. Extracts were combined, evaporated in a vacuum centrifuge and dissolved in 2% acetonitrile–0.1% aqueous TFA for MS analysis.

### Mass Spectrometry

LC-MS/MS experiments were conducted on a Dionex Ultimate 3000 UHPLC interfaced with an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) as previously described (17) with the following modifications. An LC gradient of 2-40% acetonitrile over 60 min was employed. A stepped FAIMS method was employed alternating between -25 V and -45 V. Specific glycan fragment ions (204.0867, 138.0545, 218.0659 and 232.0816 m/z) were used to trigger the additional CID, ETD and HCD scans (all in the orbitrap) with previously described scanning parameters (17).

**Peptide Identification**

Proteins and peptides were identified by searching against the *F. johnsoniae* UW101 sequence database (Uniprot proteome: UP000006694; Genome accession: CP000685) (18). All searches were conducted using trypsin with up to two missed cleavages allowed, peptide mass tolerance was set to 10 ppm, fragment mass tolerance to 0.04 Da, and a fixed modification of carbamidomethyl (C), and a variable modification of oxidation (M). Initially, the raw MS data was searched with Byonic™ v3.9.6 (Protein Metrics Inc) using the wildcard setting, set to 100-1500 Da (19). After identification of the various glycan additions, the data was searched with Mascot v2.6.2 (Matrix Science, UK) with an additional modification set for the major glycan at 1051.323 Da according to its calculated Δmass.

A putative glycopeptide identified by Byonic was included for the Δmass analysis if it contained the sequence of the glycosylation motif and exhibited a |Log Prob| score of greater than 1.0. For the Mascot searches, glycopeptides were considered identified and included in **Table S1** if they met the following criteria. Where the protein was identified with just one peptide (the putative glycopeptide), the Mascot score had to be at least 40 and the match was manually checked. For all other proteins, a glycosylation site was identified when > 1 peptide bearing the site was identified above threshold ($p<0.05$); or if one peptide, the Mascot score had to be at least 20 and was manually checked if <30. Manual checking involved finding signature low mass ions corresponding to HexNAcA-O-methyl at 232.082, strong matches to b or y ion series and peaks corresponding to glycan cleavages. Checking was also performed on proteins predicted to be cytoplasmic as these should not in theory be glycosylated.

The false discovery rate (FDR) of identifying peptides (PSMs) in the Mascot searches above the identity threshold (score ~15) was determined in decoy mode. The reported FDR values were 3.5%, 2.9% and 2.0% for the membrane, soluble and culture fluid samples respectively. No glycopeptides were identified for any of the samples in the decoy mode.

**Protein localization**

The localisation of proteins was predicted by bioinformatic approaches as follows. Proteins exhibiting a C-terminal signal for the T9SS were designated as “secreted”, and may be associated with the cell surface or secreted into the culture fluid. The proteins were found using published lists (20) and the finding of conserved CTD domains such as TIGR04183, TIGR04131 or NF033708 in the NCBI conserved domain database (www.ncbi.nlm.nih.gov/Structure/cdd). Proteins were localised to the IM and OM using DeepTMHMM (https://dtu.biolib.com/DeepTMHMM/) (21). Lipoproteins were predicted using Signal P 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) (22). Proteins were predicted to localize to the periplasm when signal peptides were identified using Signal P 6.0 or DeepTMHMM and the protein wasn't predicted to localize to a membrane or to be secreted across the outer membrane. Proteins that could not be confidently predicted to a locale were left as “uncertain”.

**Glycan nomenclature**

The nomenclature used for drawing and abbreviating sugars was taken from the Symbol Nomenclature for Glycans (SNFG) at https://www.ncbi.nlm.nih.gov/glycans/snfg.html (23).

**Protein Modelling**

Pre-computed protein models of BamA, LptD, SecDF, Signal peptidase I, and a lipoprotein (Acc: A5FM71) were obtained from the Alphafold protein structure database (alphafold.ebi.ac.uk) (24). The pdb models were downloaded and edited in ChimeraX version 1.5 (25) to show the location of glycosylation sites.

**Data availability**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD040877 and 10.6019/PXD040877.

**Results**

*F. johnsoniae* cultures were fractionated into culture fluid, membrane and soluble fractions and the proteins were separated by SDS-PAGE, digested with trypsin and analysed by LC-MS/MS as outlined (**Fig 2**). These analyses led to the identification of 325 O-glycosylation sites within 226 glycoproteins (**Table S1**).

To identify the major glycoforms that modify *F. johnsoniae* proteins, the data was first analysed using Byonic with the wildcard setting to identify peptides with any modification up to a Δmass of 1500 Da. In the initial analysis (not shown), a Δmass of approximately 1051 was the only prominent value, and the corresponding peptide sequences exhibited the Bacteroidota O-glycosylation motif consistent with the presence of one major glycoform. To increase the sensitivity of the analysis, only peptide sequences that included the motifs DSX or DTX were considered. The Δmass plot of these confirmed the prominence of the cluster around 1051 Da, and showed other potential glycoforms (**Fig 3**). As discussed previously (3, 4), the clusters are due to Byonic incorrectly assigning the monoisotopic peaks necessitating manual checking of the MS spectra to calculate the correct monoisotopic mass of the precursor and hence the correct Δmass values. This was performed on a representative number of spectra matching the common Δmass values starting with 981 (**Table 1**). Inspection of the data associated with a Δmass of 981 revealed that the peptides were not glycosylated, but rather the charge states of the precursor peaks were incorrectly assigned. The others were all due to glycosylation and are discussed below.

Peptides glycosylated with an accurate Δmass of 1135.340 (**Table 1**) together with the corresponding CID spectra matched to the same glycan published for *F. columnare* (**Fig 1C**, **Fig 4A**). This glycan has two O-acetyl groups on the pentose sugar. Relative to this glycoform, the Δmass of 1093.330 has just one O-acetyl group on the pentose (star) as indicated by the smaller m/z values starting with the $Y_3$ ion (**Fig 4B**). Similarly, the major glycan of Δmass 1051.319 is missing both O-acetyl groups and now the difference between the $Y_2$ and $Y_3$ ions is 132 Da which corresponds to an unmodified pentose residue (**Fig 4C**). In **Fig 4** and **Fig 5**, the ions depicted are doubly charged unless otherwise shown, and hence differences in m/z values need to be doubled to calculate the mass difference.

The accurate Δmass of 1034.291 corresponded to peptides modified with the major glycoform (1051 Da) but with their N-terminal Gln converted to pyroglutamate (pGlu, **Table 1**). The Δmass of 1037.304 corresponds to a missing methyl group (**Fig 5A**). As compared to the major glycan modification of the same peptide (**Fig 5B**), the absence of a methyl (14 Da) can be seen in the $Y_2$, $Y_3$ and $Y_4$ ions whereas the $Y_{1b}$ ions are the same in both spectra indicating that the methyl group is

missing from the deoxyhexose residue (dHex, triangle). Interestingly, the Δmass of 1065.299 is 13.979 Da higher than the major glycoform and corresponds to an additional O but minus two H (**Table 1**). This difference was localised to the first two sugars and may correspond to an oxidation, either an additional oxidation of the uronic acid (diamond) or an oxidation of the hexose to form a uronic acid (**Fig 5C**). Due to the very low abundance of all these additional glycoforms, only the major glycoform of Δmass 1051.319 was used for mascot searching to construct the main data table (**Table S1**).

The O-glycosylation motif found for *F. johnsoniae* was compared to our earlier findings for *P. gingivalis* and *T. forsythia* (**Fig 6**). Overall, the motif preferences were similar across the three species, however *F. johnsoniae* had a relatively strong preference for amino acid residues DTI and DTT and a relatively weak preference for DSA, DSV and DTA. DSN was uniquely observed in *F. johnsoniae* while DTC, DTG and DTM were absent.

The 226 identified glycoproteins were localised using a range of bioinformatic tools, and found to be associated predominantly with the inner membrane (IM), periplasm or were lipoproteins (**Fig 7**, **Table S2**). Only 9 were predicted to be T9SS cargo proteins, and these exhibited only a total of 10 identified glycosylation sites. Of the 29 predicted Omps, those with the most identified sites included BamA and TamA-like proteins required for Omp insertion into the OM (26) and LptD required for LPS insertion into the OM (27). Each of these proteins are known to include large periplasmic domains, and it is these that harbour the major sites for glycosylation (**Supp Figure -A & B**). Since O-glycosylation in the Bacteroidetes and other Gram-negative bacteria is known to occur in the periplasm, IM proteins can only be glycosylated on the periplasmic side. Many of the identified IM proteins are also known to have large periplasmic domains, and it is these domains that appear to be glycosylated. Examples of such domains with multiple glycosylation sites include models of SecDF and signal peptidase I (**Supp Figure -C&D**). The protein identified with the most glycosylated sites was a predicted lipoprotein with five β-sandwich domains. All glycosylation sites were located in loop regions or at the end of β-strands (**Supp Figure E**).

## Discussion

In this study, we have identified numerous O-glycosylated peptides and proteins in *F. johnsoniae*, and found that the major glycan is similar to that observed previously in the related species, *F. columnare* (10). In that study, the structure of the hexasaccharide glycan was elucidated from a small number of related protein sequences exhibiting the motif DSA. Helpfully, the structure of the glycan

was fully described, however the prevalence of this glycan was not determined. For *F. johnsoniae*, we found a glycan consistent with the *F. columnare* hexasaccharide except that with only MS data we could not confirm the isomeric forms of the sugars. The major glycan that we observed however was lacking the two acetyl groups on the pentose. It is uncertain whether this is a real difference between the two species, perhaps driven by differing levels of activity of the O-acetyltransferases, or is a difference due to the very small number of glycopeptides analysed for *F. columnare.* It should also be noted that O-glycan structures can vary between strains (8), and therefore other strains of *F. johnsoniae* need to be analysed to determine whether the glycan identified here is representative of the species.

As could be expected, the flavobacterial O-glycans are most similar to the glycan of the closely related *E. meningoseptica* (Fig 1), with five sugars in common, albeit with different methylation sites. The first three sugars comprising a hexose, a deoxyhexose and a uronic acid are the best conserved across the five species (Fig 1). An antibody raised to the "core glycan" composed of the first two sugars of the *B. fragilis* O-glycan (Man-2MeRha) cross-reacted with glycoproteins produced across the Bacteroidota phylum (1). However, the purity of this "core glycan" was not shown and conceivably contained minor glycans containing additional sugars. In particular, in the published studies of the O-glycan biosynthesis, the transferase responsible for transferring the third sugar (hexuronate) has not been determined (1, 4, 9), and therefore it is possible that this sugar is also targeted by the core glycan antibody.

We identified 325 O-glycosylation sites within 226 glycoproteins, more than the number we previously identified in *P. gingivalis* (257 sites & 145 proteins (4)) and *T. forsythia* (312 sites & 145 proteins (3)). This reflects the much greater number of predicted glycoproteins in *F. johnsoniae* (1041 proteins) based on the presence of the original O-glycosylation motif in the protein sequences (1). The glycoproteins were predicted to localize mainly to the IM, periplasm and to lipoproteins that are most likely located in the periplasm and attached to either membrane. Fewer glycoproteins were localised to the OM, and the most prominent of these exhibited large periplasmic domains. This localisation distribution was similar to what we observed in *P. gingivalis* and is consistent with glycosylation playing a role in the folding or stabilisation of proteins or protein domains within the periplasmic compartment (4).

Of special interest are the glycoproteins associated with gliding motility and the T9SS. Six of these were found to be glycosylated, namely GldH, GldJ, GldK, GldN, SprE and one of the PorE

homologs, Fjoh_2275 (A5FHL7) (**Table S2**). Further studies are required to determine if their glycosylation impacts the efficiency of gliding or secretion.

**Acknowledgements**

We thank Susan Veith for her research assistance, and the staff of the Mass Spectrometry and Proteomics Facility at the Bio21 Institute, University of Melbourne, Australia for the acquisition of Orbitrap LC-MS/MS data and related technical support. This work was supported by the Australian National Health and Medical Research Council grant ID 1123866, the Australian Research Council grant DP200100914 and the Australian Government Department of Industry, Innovation and Science Grant ID 20080108.

## References


1. Coyne MJ, Fletcher CM, Chatzidaki-Livanis M, Posch G, Schaffer C, Comstock LE. 2013. Phylum-wide general protein O-glycosylation system of the Bacteroidetes. Mol Microbiol 88:772-83.
2. Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE. 2009. A general O-glycosylation system important to the physiology of a major human intestinal symbiont. Cell 137:321-31.
3. Veith PD, Scott NE, Reynolds EC. 2021. Characterization of the O-Glycoproteome of *Tannerella forsythia*. mSphere 6:e0064921.
4. Veith PD, Shoji M, Scott NE, Reynolds EC. 2022. Characterization of the O-Glycoproteome of *Porphyromonas gingivalis*. Microbiol Spectrum 10:e0150221.
5. Posch G, Pabst M, Neumann L, Coyne MJ, Altmann F, Messner P, Comstock LE, Schaffer C. 2013. "Cross-glycosylation" of proteins in Bacteroidales species. Glycobiology 23:568-77.
6. Tomek MB, Janesch B, Braun ML, Taschner M, Figl R, Grunwald-Gruber C, Coyne MJ, Blaukopf M, Altmann F, Kosma P, Kahlig H, Comstock LE, Schaffer C. 2021. A Combination of Structural, Genetic, Phenotypic and Enzymatic Analyses Reveals the Importance of a Predicted Fucosyltransferase to Protein O-Glycosylation in the Bacteroidetes. Biomolecules 11.
7. Reinhold BB, Hauer CR, Plummer TH, Reinhold VN. 1995. Detailed structural analysis of a novel, specific O-linked glycan from the prokaryote *Flavobacterium meningosepticum*. J Biol Chem 270:13197-203.
8. Friedrich V, Janesch B, Windwarder M, Maresch D, Braun ML, Megson ZA, Vinogradov E, Goneau MF, Sharma A, Altmann F, Messner P, Schoenhofen IC, Schaffer C. 2017. *Tannerella forsythia* strains display different cell-surface nonulosonic acids: biosynthetic pathway characterization and first insight into biological implications. Glycobiology 27:342-357.
9. Tomek MB, Maresch D, Windwarder M, Friedrich V, Janesch B, Fuchs K, Neumann L, Nimeth I, Zwickl NF, Dohm JC, Everest-Dass A, Kolarich D, Himmelbauer H, Altmann F, Schaffer C. 2018. A General Protein O-Glycosylation Gene Cluster Encodes the Species-Specific Glycan of the Oral Pathogen *Tannerella forsythia*: O-Glycan Biosynthesis and Immunological Implications. Front Microbiol 9:2008.
10. Vinogradov E, Perry MB, Kay WW. 2003. The structure of the glycopeptides from the fish pathogen *Flavobacterium columnare*. Carbohydr Res 338:2653-8.
11. Bloch S, Tomek MB, Friedrich V, Messner P, Schaffer C. 2019. Nonulosonic acids contribute to the pathogenicity of the oral bacterium *Tannerella forsythia*. Interface Focus 9:20180064.
12. McBride MJ. 2019. Bacteroidetes Gliding Motility and the Type IX Secretion System. Microbiol Spectr 7.
13. Nakane D, Sato K, Wada H, McBride MJ, Nakayama K. 2013. Helical flow of surface protein required for bacterial gliding motility. Proc Natl Acad Sci U S A 110:11145-50.
14. Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, McBride MJ, Rhodes RG, Nakayama K. 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. Proc Natl Acad Sci U S A 107:276-81.
15. McBride MJ, Kempf MJ. 1996. Development of techniques for the genetic manipulation of the gliding bacterium *Cytophaga johnsonae*. J Bacteriol 178:583-90.
16. Gorasia DG, Veith PD, Chen D, Seers CA, Mitchell HA, Chen YY, Glew MD, Dashper SG, Reynolds EC. 2015. *Porphyromonas gingivalis* Type IX Secretion Substrates Are Cleaved and Modified by a Sortase-Like Mechanism. PLoS Pathog 11:e1005152.
17. Ahmad Izaham AR, Ang CS, Nie S, Bird LE, Williamson NA, Scott NE. 2021. What are we missing by using hydrophilic enrichment? Improving bacterial glycoproteome coverage using total proteome and FAIMS analyses. J Proteome Res 20:599-612.
18. McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, Goltsman E, Wang W, Xu J, Hunnicutt DW, Staroscik AM, Hoover TR, Cheng YQ, Stein JL. 2009. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. Appl Environ Microbiol 75:6864-75.
19. Roushan A, Wilson GM, Kletter D, Sen KI, Tang W, Kil YJ, Carlson E, Bern M. 2020. Peak filtering, peak annotation, and wildcard search for glycoproteomics. Mol Cell Proteomics 20:100011.

20. Kulkarni SS, Johnston JJ, Zhu Y, Hying ZT, McBride MJ. 2019. The Carboxy-Terminal Region of *Flavobacterium johnsoniae* SprB Facilitates Its Secretion by the Type IX Secretion System and Propulsion by the Gliding Motility Machinery. J Bacteriol 201.
21. Hallgren J, Tsirigos KD, Damgaard Pedersen M, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. BioRxiv doi:https://doi.org/10.1101/2022.04.08.487609.
22. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023-1025.
23. Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lutteke T, O'Boyle N, Packer NH, Stanley P, Toukach P, Varki A, Woods RJ, Group SD. 2019. Updates to the symbol nomenclature for glycans guidelines. Glycobiology 29:620-624.
24. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Zidek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439-D444.
25. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 30:70-82.
26. Stubenrauch CJ, Lithgow T. 2019. The TAM: A Translocation and Assembly Module of the beta-Barrel Assembly Machinery in Bacterial Outer Membranes. EcoSal Plus 8.
27. Botos I, Noinaj N, Buchanan SK. 2017. Insertion of proteins and lipopolysaccharide into the bacterial outer membrane. Philos Trans R Soc Lond B Biol Sci 372.

**Table 1.** Assignment of Δmass values and clusters to glycoforms

| Δmass (Byonic)[a] | Δmass (manual)[b] | Difference[c] | Assignment | Note |
|---|---|---|---|---|
| 1051 cluster | 1051.319 | 0 | - | Major glycoform |
| 981 | Not calculated | - | - | *Not a glycoform |
| 1033 | 1034.291 | -17.029 | $-NH_3$ | pGlu at N-terminus of peptide |
| 1036 | 1037.304 | -14.015 | $-CH_2$ | loss of methyl on dHex |
| 1065 cluster | 1065.299 | +13.979 | $+O-H_2$ | Formation of uronic acid? |
| 1077 cluster | 1077.335 | +26.016 | $+C_2H_2$ | Artefactual modification at peptide N-terminus |
| 1093 cluster | 1093.330 | +42.011 | $+ C_2H_2O$ | O-acetylation |
| 1135 cluster | 1135.340 | +84.020 | $+ 2C_2H_2O$ | O-acetylation x 2 |

[a]Each Δmass value in this column corresponds to Fig 3. Where the value is labelled as a "cluster", it corresponds to a shaded region in Fig 3 represented by multiple Δmass values.

[b]The Δmass values in this column were calculated accurately by manually checking the mass of the monoisotopic peaks.

[c]The difference is calculated relative to the major glycoform in the first row of mass 1051.319

**Figure Legends**

**Figure 1. Comparison of O-glycan structures amongst *Bacteroidota*.** O-glycans from A. *P. gingivalis* ATCC 33277 (4), B. *B. fragilis* (5, 6). C. *F. columnare* (10). D. *E. meningoseptica* (7) and E. *T. forsythia* (3, 6, 9). Note that striped sugar symbols are used to show the predicted isomeric form of the sugar – see legend.

**Figure 2. Overview of sample preparation.** *F. johnsoniae* cells were divided into membrane (Mem), soluble (Sol) and culture fluid (CF) fractions and separated by SDS-PAGE. Each gel lane was divided into 12 fractions for digestion with trypsin and LC-MS/MS analysis.

**Figure 3. Detection of glycoforms using Byonic wildcard.** Membrane, soluble and culture fluid samples (Fig 2) were analysed by LC-MS/MS and the data searched using Byonic in wildcard mode. The search results were binned into integer Δmass values, and then the frequency of each value was plotted. Only Δmass values with a frequency of at least 10 in any sample are shown. Shaded regions correspond to clusters of adjacent Δmass values - likely corresponding to the same glycoform (see main text). Some low intensity glycoforms exhibiting Δmass values between 1047 and 1049 appeared to have an accurate Δmass 2 Da lower than the major glycoform of 1051 Da, possibly due to double bond formation resulting in the loss of $H_2$.

**Figure 4. CID spectra of tryptic peptides modified with glycans varying in O-acetyl substitution.** Each CID spectrum corresponds to the peptide AAEKPAEGAAPTTDTTAK from Cytochrome C oxidase subunit II (A5FJE3) with Δmass 1135.34 Da (A), Δmass 1093.33 Da (B) and Δmass 1051.32 Da (C). All spectra are from the membrane sample (band 6 or 7) and the CID spectra were triggered by the presence of HexNAcA-O-methyl (m/z 232.082) in the corresponding HCD spectra. For all spectra, the labeled ions are $2^+$ unless shown otherwise. Sugar symbols follow the SNFG as described in Methods.

**Figure 5. CID spectra of tryptic peptides modified with minor glycans.** Each CID spectrum corresponds to the peptide GEQQLKDTLTIGK from PKD domain containing protein (A5FK25) with Δmass 1037.30

Da, missing one methyl group (A), Δmass 1051.32 Da, the major glycan (B) and Δmass 1065.299 Da, with oxidized sugar (C). All spectra are from the culture fluid sample (band 8) and the CID spectra were triggered by the presence of HexNAcA-O-methyl (m/z 232.082) in the corresponding HCD spectra. For all spectra, the labeled ions are $2^{+}$. Sugar symbols follow the SNFG as described in Methods.

**Figure 6. O-glycosylation motif comparison across species.** Frequency of observed O-glycosylation motifs in *F. johnsoniae* (FJ, n=310)*, P. gingivalis* (PG, n=254) and *T. forsythia* (TF, n=318). The total number (n) of observed motifs for each species is different to the number of reported sites identified since in some peptides there is more than one possible motif and hence it is uncertain which motif is glycosylated.

**Figure 7. Predicted localisation of identified glycoproteins.** The predicted localisation of proteins to the outer membrane (OM, 26), periplasm (PP, 60) and inner membrane (IM, 68) are plotted together with T9SS secreted proteins (S, 9), and lipoproteins (LP, 57) which can be associated with either membrane. The black wedge corresponds to 3 glycoproteins of uncertain localization.

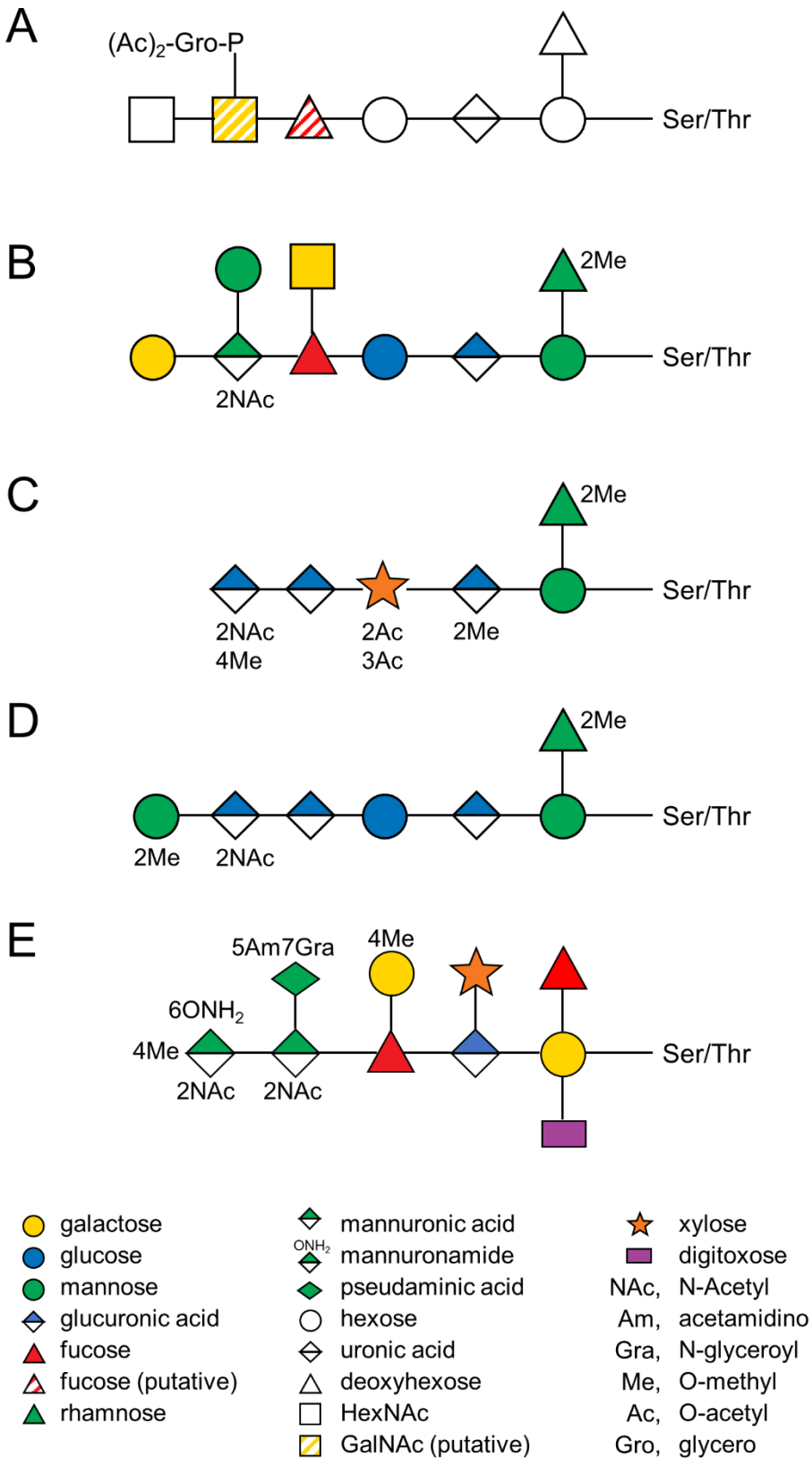
A
(Ac)2-Gro-P
Ser/Thr
B
2Me
Ser/Thr
2NAc
C
2Me
Ser/Thr
2NAc
4Me
2Ac
3Ac
2Me
D
2Me
Ser/Thr
2Me
2NAc
E
5Am7Gra
4Me
6ONH2
4Me
Ser/Thr
2NAc
2NAc
galactose
glucose
mannose
glucuronic acid
fucose
fucose (putative)
rhamnose
mannuronic acid
ONH2 mannuronamide
pseudaminic acid
hexose
uronic acid
deoxyhexose
HexNAc
GalNAc (putative)
xylose
digitoxose
NAc, N-Acetyl
Am, acetamidino
Gra, N-glyceroyl
Me, O-methyl
Ac, O-acetyl
Gro, glycero

Figure 1

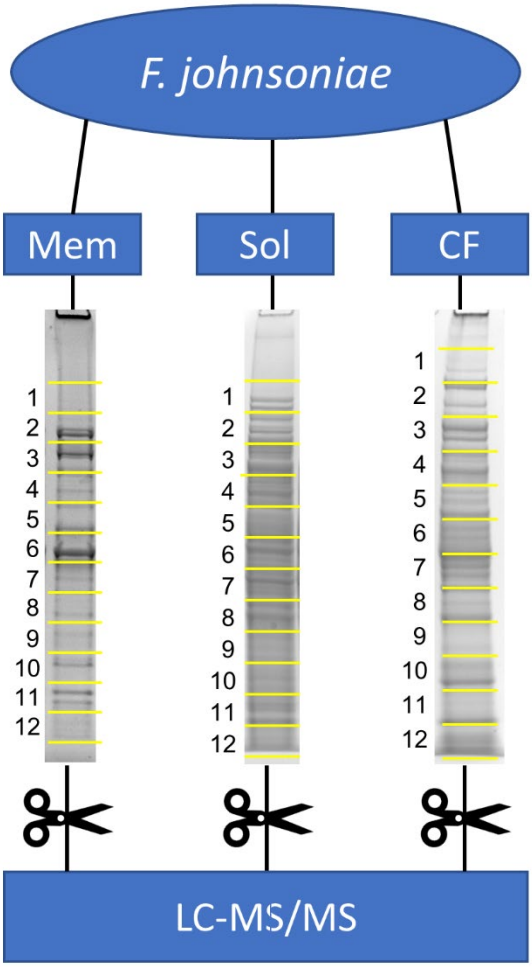


Figure 2

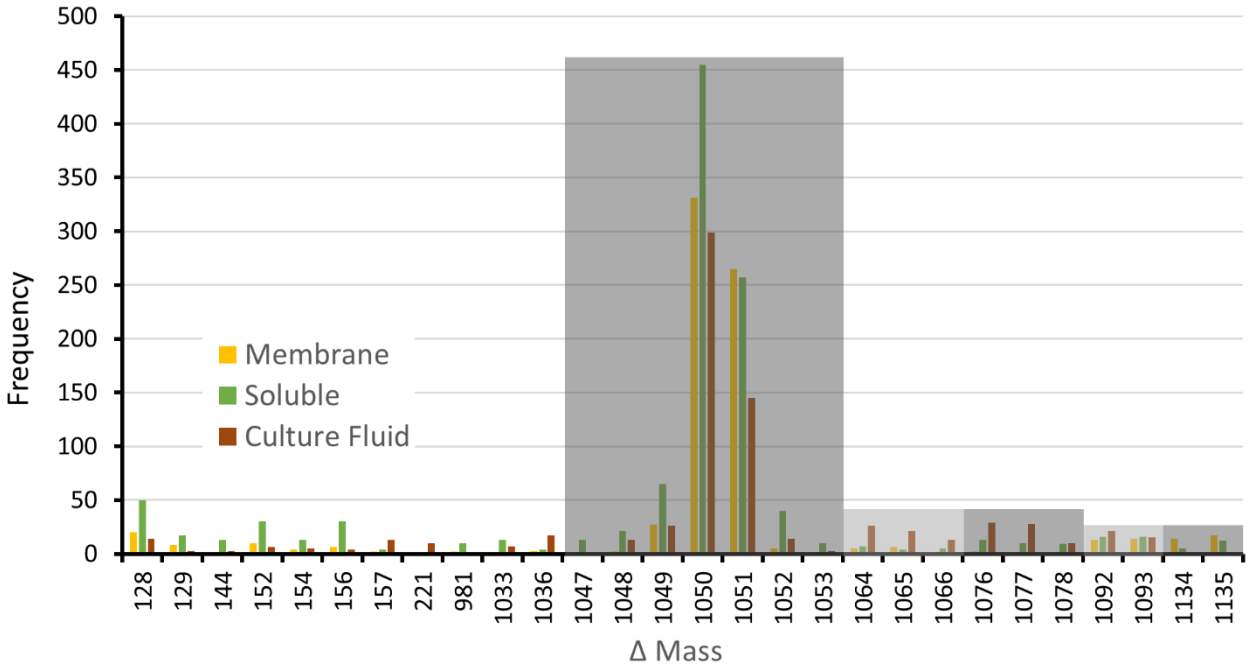


Figure 3

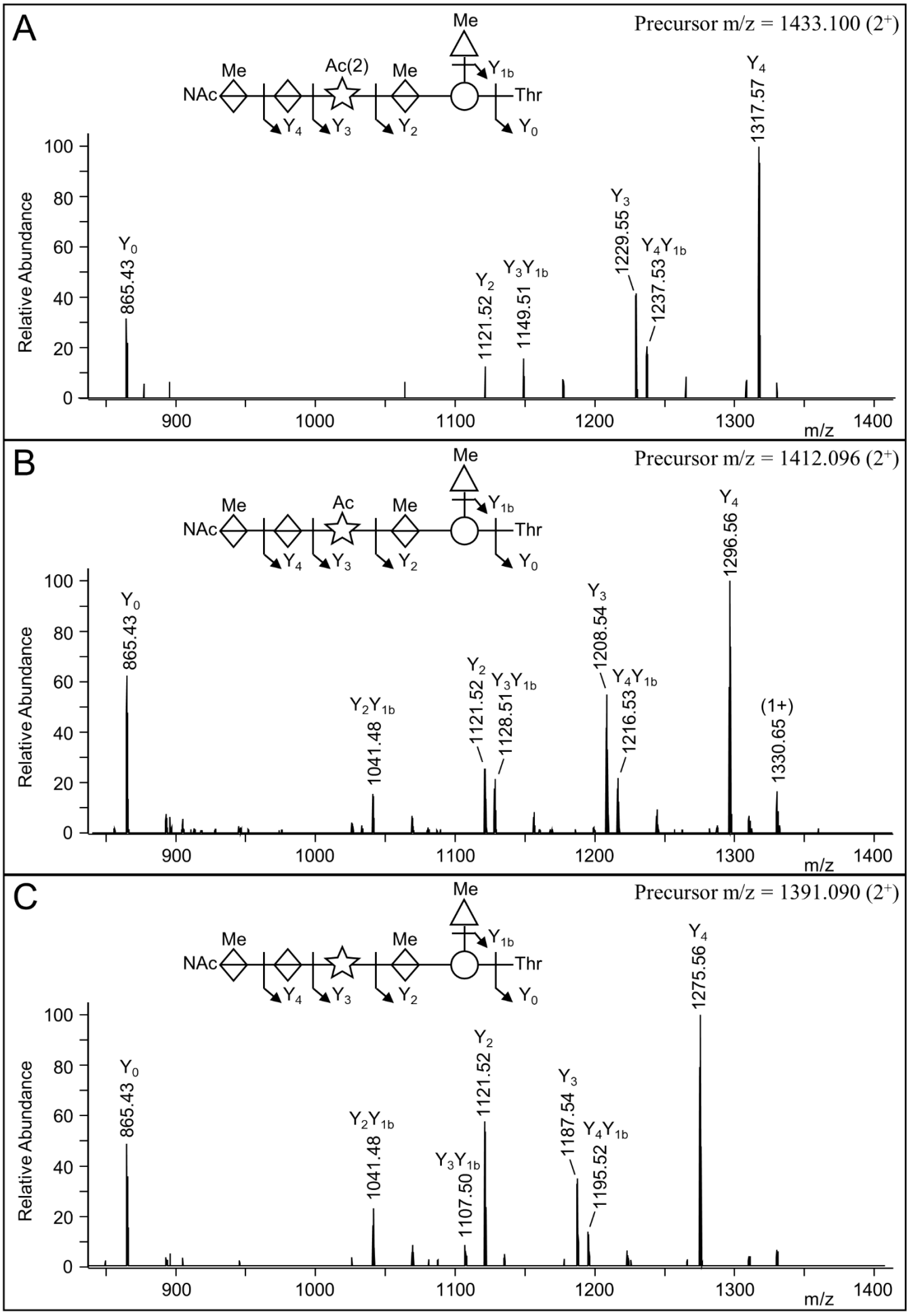
A
Precursor m/z = 1433.100 (2+)
Me
Ac(2)
Me
Me
NAc
Thr
Y4
Y3
Y2
Y0
Y1b
Y4
1317.57
Y3
1229.55
Y4Y1b
1237.53
Y3Y1b
1149.51
Y2
1121.52
Y0
865.43
Relative Abundance
m/z
B
Precursor m/z = 1412.096 (2+)
Ac
Y4
1296.56
Y3
1208.54
Y4Y1b
1216.53
Y2
1121.52
Y3Y1b
1128.51
Y2Y1b
1041.48
Y0
865.43
(1+)
1330.65
C
Precursor m/z = 1391.090 (2+)
Y4
1275.56
Y2
1121.52
Y3
1187.54
Y4Y1b
1195.52
Y3Y1b
1107.50
Y2Y1b
1041.48
Y0
865.43

Figure 4

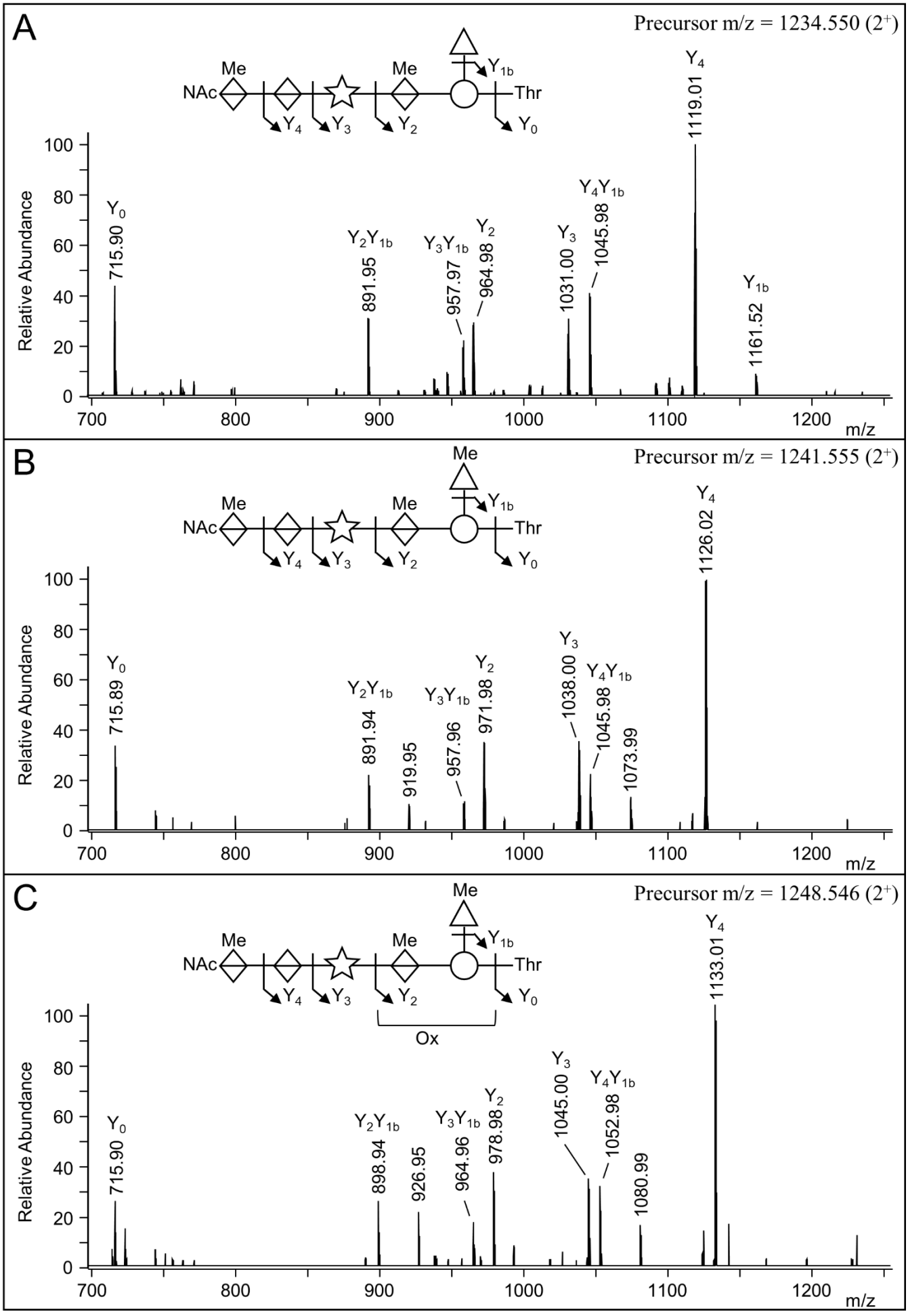
A
Precursor m/z = 1234.550 (2+)
Me
NAc
Me
Thr
Y1b
Y4
Y3
Y2
Y0
Relative Abundance
m/z
Y0 715.90
Y2Y1b 891.95
Y3Y1b 957.97
Y2 964.98
Y3 1031.00
Y4Y1b 1045.98
Y4 1119.01
Y1b 1161.52
B
Precursor m/z = 1241.555 (2+)
Y0 715.89
Y2Y1b 891.94
919.95
Y3Y1b 957.96
Y2 971.98
Y3 1038.00
Y4Y1b 1045.98
1073.99
Y4 1126.02
C
Precursor m/z = 1248.546 (2+)
Ox
Y0 715.90
Y2Y1b 898.94
926.95
Y3Y1b 964.96
Y2 978.98
Y3 1045.00
Y4Y1b 1052.98
1080.99
Y4 1133.01

Figure 5

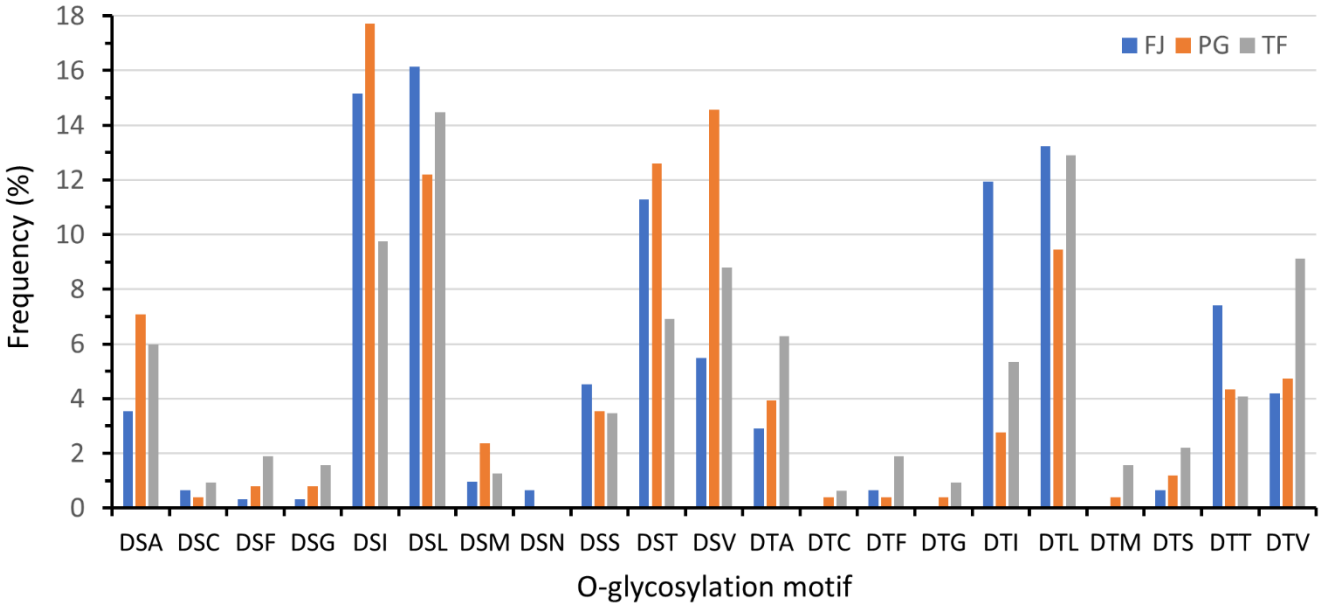


Figure 6

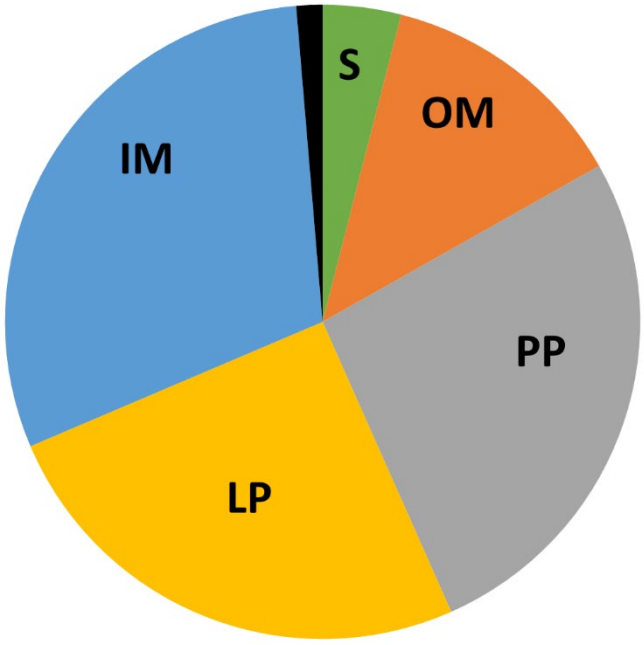


Figure 7