Journal of Visualized Experiments

# The Application of Open Searching-based Approaches for the Identification of Acinetobacter baumannii O-linked Glycopeptides

--Manuscript Draft--

| | |
|---|---|
| **Article Type:** | Invited Methods Article - JoVE Produced Video |
| **Manuscript Number:** | JoVE63242R3 |
| **Full Title:** | The Application of Open Searching-based Approaches for the Identification of Acinetobacter baumannii O-linked Glycopeptides |
| **Corresponding Author:** | Nichollas Scott<br>: The University of Melbourne Department of Microbiology and Immunology Melbourne, Victoria AUSTRALIA |
| **Corresponding Author's Institution:** | : The University of Melbourne Department of Microbiology and Immunology |
| **Corresponding Author E-Mail:** | nichollas.scott@unimelb.edu.au |
| **Order of Authors:** | Jessica Lewis |
| | Pauline Coulon |
| | Taylor McDaniels |
| | Nichollas Scott |
| **Additional Information:** | |
| **Question** | **Response** |
| Please specify the section of the submitted manuscript. | Biochemistry |
| Please indicate whether this article will be Standard Access or Open Access. | Standard Access ($1400) |
| Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations. | Melbourne, Victoria, Australia |
| Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below: | I agree to the Author License Agreement |
| Please confirm that you have read and agree to the terms and conditions of the video release that applies below: | I agree to the Video Release |
| Please provide any comments to the journal here. | |

**TITLE:**

# The Application of Open Searching-based Approaches for the Identification of *Acinetobacter baumannii* O-linked Glycopeptides

**AUTHORS AND AFFILIATIONS:**

Jessica M. Lewis[1*], Pauline M. L. Coulon[1*], Taylor A. McDaniels[1*], Nichollas E. Scott[1]

[1]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Parkville, Victoria, Australia

*These authors contributed equally

**Email addresses of co-authors:**

Jessica M. Lewis (jessica.lewis1@unimelb.edu.au)
Pauline M. L. Coulon (pauline.coulon@unimelb.edu.au)
Taylor A. McDaniels (tmcdaniels@student.unimelb.edu.au)

**Corresponding author:**

Nichollas E. Scott (nichollas.scott@unimelb.edu.au)

**KEYWORDS:**

Glycosylation, *Acinetobacter baumannii*, Open searching, Posttranslational modifications, Proteomics.

**SUMMARY:**

Open searching enables the identification of glycopeptides decorated with previously unknown glycan compositions. Within this article, a streamlined approach for undertaking open searching and subsequent glycan-focused glycopeptide searches are presented for bacterial samples using *Acinetobacter baumannii* as a model.

**ABSTRACT:**

Protein glycosylation is increasingly recognized as a common modification within bacterial organisms, contributing to prokaryotic physiology and optimal infectivity of pathogenic species. Due to this, there is increasing interest in characterizing bacterial glycosylation and a need for high-throughput analytical tools to identify these events. Although bottom-up proteomics readily enables the generation of rich glycopeptide data, the breadth and diversity of glycans observed in prokaryotic species make the identification of bacterial glycosylation events extremely challenging.

Traditionally, the manual determination of glycan compositions required within bacterial proteomic datasets made this a largely bespoke analysis restricted to field-specific experts. Recently, open searching-based approaches have emerged as a powerful alternative for the identification of unknown modifications. By analyzing the frequency of unique modifications observed on peptide sequences, open searching techniques allow the identification of common glycans attached to peptides within complex samples. This article presents a streamlined workflow for the interpretation/analysis of glycoproteomic data, demonstrating how open

searching techniques can be used to identify bacterial glycopeptides without prior knowledge of the glycan compositions.

Using this approach, glycopeptides within samples can rapidly be identified to understand glycosylation differences. Using *Acinetobacter baumannii* as a model, these approaches enable the comparison of glycan compositions between strains and the identification of novel glycoproteins. Taken together, this work demonstrates the versatility of open database-searching techniques for the identification of bacterial glycosylation, making the characterization of these highly diverse glycoproteomes easier than ever before.

**INTRODUCTION:**

Protein glycosylation, the process of attaching carbohydrates to protein molecules, is one of the most common posttranslational modifications (PTMs) in nature[1,2]. Across all domains of life, a range of complex machinery has evolved dedicated to the generation of glycoproteins that impact a myriad of cellular functions[1,3-5]. While protein glycosylation occurs on a range of amino acids[6,7], *N*-linked and *O*-linked glycosylation events are two dominant forms observed in nature. *N*-linked glycosylation involves the attachment of glycans to a nitrogen atom of asparagine (Asn) residues, while in *O*-linked glycosylation, the glycans are attached to an oxygen atom of serine (Ser), threonine (Thr), or tyrosine (Tyr) residues[7]. Despite the similarities in residues targeted by glycosylation systems, the differences within the glycans attached to proteins result in glycosylation being the most chemically diverse class of PTMs found in nature.

While eukaryotic glycosylation systems possess glycan diversity, these systems are typically restricted in the number of unique carbohydrates utilized. The resulting diversity stems from how these carbohydrates are arranged into glycans[8-12]. In contrast, bacterial and archaeal species possess virtually unlimited glycan diversity due to the sheer array of unique sugars generated within these systems[2,10,13-17]. These differences in the glycan diversity observed across domains of life represent a significant analytical challenge for the characterization and identification of glycosylation events. For eukaryotic glycosylation, the ability to anticipate glycan compositions has facilitated the growing interest in glycobiology; yet, the same is not true for bacterial glycosylation, which is still largely restricted to study by specialized laboratories. As the accessibility of mass spectrometry (MS) instrumentation has increased in the biosciences, MS-based approaches are now the primary method for glycoproteomic analysis.

MS has emerged as the quintessential tool for the characterization of glycoproteins, with both top-down and bottom-up approaches now commonly used to characterize glycoproteins[6]. While top-down proteomics is used to assess global glycosylation patterns of specific proteins[18,19], bottom-up approaches are used to enable the glycan-specific characterization of glycopeptides, even from complex mixtures[6,20-23]. For the analysis of glycopeptides, the generation of informative fragmentation information is essential for the characterization of glycosylation events[24,25]. A range of fragmentation approaches is now routinely accessible on instruments, including resonance ion trap-based collision-induced dissociation (IT-CID), beam-type collision-induced dissociation (CID), and electron transfer dissociation (ETD). Each approach possesses different strengths and weaknesses for glycopeptide analysis[25,26], with significant progress over

the last decade in applying these fragmentation approaches to analyze glycosylation[6,20]. However, for bacterial glycosylation analysis, the critical limitation has not been the ability to fragment glycopeptides but rather the inability to predict the potential glycan compositions within samples. Within these systems, the unknown nature of diverse bacterial glycans limits the identification of glycopeptides even with glycosylation-focused searching tools now commonplace for the analysis of eukaryotic glycopeptides, such as O-Pair[27], GlycopeptideGraphMS[28], and GlycReSoft[29]. To overcome this issue, an alternative searching approach is required, with the use of open searching tools emerging as a powerful approach for the study of bacterial glycosylation [30].

Open searching, also known as blind or wildcard searching, allows the identification of peptides with unknown or unexpected PTMs[21,30-32]. Open searches utilize a variety of computational techniques, including curated modification searches, multistep database searches, or wide-mass tolerant searching[33-37]. Although open searching has great potential, its use has typically been hindered by the significant increase in analysis times and loss in sensitivity of the detection of unmodified peptides compared to restrictive searches[31,32]. The decrease in the detection of unmodified peptide-spectral matches (PSMs) is a result of the increased false-positive PSM rates associated with these techniques, which requires increased stringent filtering to maintain the desired false discovery rates (FDRs)[33-37]. Recently, several tools have become available that significantly improve the accessibility of open searching, including Byonic[31,38], Open-pFind[39], ANN-SoLo[40], and MSFragger[21,41]. These tools enable the robust identification of glycosylation events by significantly reducing analysis times and implementing approaches to handle heterogeneous glycan compositions.

This article presents a streamlined method for the identification of bacterial glycopeptides by open searching, using the gram-negative nosocomial pathogen, *Acinetobacter baumannii*, as a model. *A. baumannii* possesses a conserved *O*-linked glycosylation system responsible for the modification of multiple protein substrates, known as the PglL protein glycosylation system[42-44]. While similar proteins are targeted for glycosylation between strains, the PglL glycosylation system is highly variable due to the biosynthesis of the glycan used for protein glycosylation being derived from the capsule locus (known as the K-locus)[44-46]. This results in diverse glycans (also known as a K-unit), derived from single or limited polymerized K-units, being added to protein substrates[30,44,46]. Within this work, the use of the open searching tool, MSfragger, within the software FragPipe, is used to identify glycans across *A. baumannii* strains. By combining open searching and manual curation, "glycan-focused searches" can be undertaken to further improve the identification of bacterial glycopeptides. Together, this multistep identification approach enables the identification of novel glycopeptides without extensive experience in the characterization of novel glycosylation events.

**PROTOCOL:**

NOTE: The preparation and analysis of bacterial glycopeptide samples can be divided into four sections (**Figure 1**). While many reviews and technique articles have focused on sample preparation and the enrichment of glycopeptides, the key focus of this article is the analysis and

characterization of glycopeptides with unknown glycan compositions, using the open searching capabilities of MSFragger. For this study, the glycosylation was assessed within three sequenced strains of *A. baumannii* (**Table 1**). Proteome FASTA databases of each of these strains are accessible via Uniprot. Refer to **Table 2** for the composition of buffers used in this protocol.

## 1. Preparation of protein samples for proteomic analysis

### 1.1. Isolation of proteome samples of interest

1.1.1. If using whole cells, ensure that the cells have been washed with a phosphate-buffered saline solution (PBS) to remove potential protein contaminants present in the medium. Snap-freeze whole cells after washing and store them at -80 °C until required.

1.1.2. If fractionated samples are used (such as membrane preparations), ensure that the reagents used will not interfere with downstream liquid chromatography MS (LC-MS) analysis[50].

1.1.3. If detergents, such as sodium dodecyl-sulfate (SDS), Triton X-100, NP-40, or lauroylsarcosine, have been used, remove these detergents using acetone precipitations, SP3 sample preparation methods[51], or commercial proteomic clean-up columns such as S-traps[52]. Alternatively, substitute incompatible detergents with an MS-compatible or removable detergent such as sodium deoxycholate (SDC) or octyl glucopyranoside.

1.1.4. Ensure all plasticware and glassware to be used for sample preparation have not been autoclaved.

NOTE: Autoclaved glassware and plastics are typically heavily contaminated with small molecular weight compounds, such as polymers, which are readily detected within the MS.

### 1.2. Solubilization of whole-cell samples

1.2.1. Resuspend ~10 mg of washed, snap-frozen cells in 200 µL of freshly prepared sodium deoxycholate lysis buffer (SDC lysis buffer: 4% SDC in 100 mM Tris, pH 8.5).

NOTE: Proteinase inhibitors can be added to the SDC lysis buffer to limit protein degradation.

1.2.2. Boil the samples for 10 min at 95 °C with shaking (2000 rpm on a thermomixer), and then leave on ice for 10 min. Repeat this process twice to ensure efficient lysis and solubilization of the samples.

NOTE: Samples can be stored long-term at this point at -80 °C. If stored, resolubilize by heating at 95 °C before further processing.

### 1.3. Quantify sample protein concentrations using a bicinchoninic acid (BCA) protein assay[53].

Store samples on ice while undertaking quantification to limit protein degradation.

NOTE: For total proteome analysis, 20–100 µg of protein is more than sufficient for nano LC-MS, which typically requires less than 2 µg of protein digest per analysis. The preparation of excess peptide allows for replicate analysis or further fractionation if deep proteomic coverage is required. For glycopeptide enrichment-based analysis using hydrophilic interaction chromatography (HILIC), 100–500 µg of protein is required.

1.4. Reduce and alkylate samples.

1.4.1. Add 1/10$^{th}$ the volume of 10x reduction/alkylation buffer (100 mM Tris 2-carboxyethyl phosphine hydrochloride; 400 mM 2-chloroacetamide in 1 M Tris, pH 8.5) to samples for a final concentration of 1x and incubate samples in the dark for 30 min at 45 °C with shaking at 1,500 rpm.

NOTE: Check the pH of the 10x reduction/alkylation buffer to ensure a pH of approximately 7.0–8.0 before adding to the samples, as a lower pH will cause the SDC to precipitate.

1.5. Briefly spin down the samples and add the proteases trypsin/Lys-C (~10 µL, resuspended in 100 mM Tris, pH 8.5) for a final protease:protein ratio of 1:100. Incubate the digests overnight at 37 °C with shaking at 1,500 rpm (up to 18 h). To ensure complete protein digestion, use a Trypsin/Lys-C protease:protein ratio of 1:50 to 1:200.

1.6. Quench digests by adding 1.25 volumes of 100% isopropanol to the samples. Vortex the samples for 1 min to mix and briefly spin them down.

NOTE: Samples can be stored at -20 °C to be further processed later.

1.7. Acidify the samples by adding 0.115 volumes of 10% trifluoroacetic acid (TFA; final concentration of ~1% TFA), vortex the samples, and briefly spin them down.

**2. Processing of proteome samples**

2.1. Peptide clean-up of proteome samples

2.1.1. Prepare one styrenedivinylbenzene-reverse-phase sulfonate (SDB-RPS) Stop-and-go-extraction (Stage) Tip for each sample as previously described[54].

2.1.1.1. Empirically, excise three SDB-RPS discs from a 47 mm$^2$ SDB-RPS membrane using a blunt needle (14 G) for binding to 50 µg of peptide. For larger peptide amounts, increase the number of discs accordingly.

2.1.2. Prior to using SDB-RPS Stage Tips, prepare the tips by sequentially adding at least ten bed volumes of the following buffers and either spinning the buffer by centrifugation (25 °C, 3 min, 500 × *g*) or by gently applying pressure using a syringe.

2.1.2.1. Wet the tips with 150 µL of 100% acetonitrile.

2.1.2.2. Wash the tips with 150 µL of 30% methanol, 1% TFA in 18.2 MΩ $H_2O$.

2.1.2.3. Equilibrate the tips with 150 µL of 90% isopropanol, 1% TFA balanced with 18.2 MΩ $H_2O$.

2.1.3. Load the samples (containing 50% isopropanol, 1% TFA) onto the SDB-RPS Stage Tips by centrifugation (25 °C, 3 min, 500 × *g*) or by gently applying pressure using a syringe.

2.1.4. Wash the SDB-RPS Stage Tips with the following buffers by centrifugation (25 °C, 3 min, 500 × *g*) or by gently applying pressure using a syringe.

NOTE: Additional washes or alternative buffers can be used to remove non-peptide contaminants, such as the use of ethyl acetate instead of isopropanol[55].

2.1.4.1. Wash the tips with 150 µL of 90% isopropanol, 1% TFA.

2.1.4.2. Wash the tips with 150 µL of 1% TFA in 18.2 MΩ $H_2O$.

2.1.5. Elute the peptides from the SDB-RPS Stage Tips with 150 µL of 5% ammonium hydroxide in 80% acetonitrile by centrifugation or by gently applying pressure using a syringe. Collect the samples in individual tubes.

NOTE: Prepare 5% ammonium hydroxide in 80% acetonitrile in a plastic container immediately prior to use within a fume hood.

2.1.6. Dry the eluted peptides by vacuum centrifugation at 25 °C.

NOTE: If undertaking HILIC enrichment, 1–10% of the peptide eluate can be removed at this point, dried, and used as total proteome input control.

2.2. Enrichment of glycopeptide samples

2.2.1. Prepare Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC) Stage Tips as previously described[54,56].

2.2.1.1. Briefly, excise one $C_8$ disc from a 47 $mm^2$ $C_8$ membrane using a blunt needle (14 G) and pack the disc into a P200 tip to create a frit. Add approximately 5 mm of ZIC-HILIC material, resuspended in 50% acetonitrile, 50% 18.2 MΩ $H_2O$, onto the frit by gently applying pressure using a syringe.

2.2.2. Prior to using ZIC-HILIC Stage Tips, condition the resin by sequentially adding the following

buffers and gently applying pressure using a syringe.

NOTE: To ensure the integrity of the pseudo-water layer on the surface of the ZIC-HILIC resin (required to enrich glycopeptides), the resin must always remain wet. When washing the resin, always leave ~10 µL of solvent above the resin and ensure the washes/samples are pipetted directly into this residual solvent.

2.2.2.1. Equilibrate the resin with 20 bed volumes (200 µL) of ZIC-HILIC elution buffer (0.1% TFA in 18.2 MΩ $H_2O$).

2.2.2.2. Wash the resin with 20 bed volumes (200 µL) of ZIC-HILIC preparation buffer (95% acetonitrile in 18.2 MΩ $H_2O$).

2.2.2.3. Wash the resin with 20 bed volumes (200 µL) of ZIC-HILIC loading/wash buffer (80% acetonitrile, 1% TFA balanced with 18.2 MΩ $H_2O$).

2.2.3. Resuspend the dried digested samples (from step 2.1.6) in ZIC-HILIC loading/wash buffer to a final concentration of 4 µg/µL (e.g., for 200 µg of peptide, resuspend in 50 µL of ZIC-HILIC loading/wash buffer). Vortex briefly for 1 min to ensure the samples are resuspended, and spin down for 1 min at 2,000 × *g* at 25 °C.

2.2.4. Load the resuspended peptide sample onto a conditioned ZIC-HILIC column.

2.2.4.1. Wash three times with 20 bed volumes (200 µL) of ZIC-HILIC loading/wash buffer (for 60 bed volume washes total) by gently applying pressure using a syringe.

2.2.5. Elute glycopeptides with 20 bed volumes (200 µL) of ZIC-HILIC elution buffer into a 1.5 mL tube by gently applying pressure using a syringe and then dry the eluate by vacuum centrifugation at 25 °C.

## 3. LC-MS of proteome/glycopeptide-enriched samples

3.1. Resuspend the samples in Buffer A* (2% acetonitrile, 0.1% TFA) to a final concentration of 1 µg/µL (for example, for 50 µg of peptide, resuspend in 50 µL of Buffer A*).

3.2. Load the samples onto an HPLC/UPLC coupled to an MS to enable the separation and identification of glycopeptides.

NOTE: The column parameters, including inner diameter, length, flow rates, type of chromatography resin, and required peptide injection amounts, should be optimized for the analytical setup and gradient length to be used; for an example of how to undertake optimization of analytical setups, see [57].

3.3. Monitor the collection of the resulting MS data ensuring the data is being collected with the

desired parameters.

NOTE: For compositional analysis, CID fragmentation is sufficient. Due to the addition of glycans to glycopeptides, glycopeptide ions are typically observed with a higher *m/z* and lower charge density than unglycosylated peptides; a mass range from 400 to 2,000 *m/z* should be analyzed.

3.4. Fragment selected ions using CID, ensuring the collection of low *m/z* fragment ions that contain oxonium ions important for the characterization of glycans.

NOTE: Fragmentation of glycopeptides using CID is influenced by both the peptide and glycan sequences, as well as the energy applied during fragmentation[25,58,59]. While a range of different collision energies can be used, an optimal strategy for fragmenting glycopeptides is the use of stepped collision energies combining the use of multiple collision energies[25,59,60].

3.5. Use alternative fragmentation methods, if available, such as ETD for site localization or IT-CID, to aid in the determination of glycan compositions.

NOTE: Neither of these fragmentation approaches are essential for compositional analysis, yet can be collected to enable further interrogation of glycopeptides of interest.

**4. Analysis of proteome/glycopeptide-enriched samples**

4.1. Prefiltering data files to enable searching in FragPipe

4.1.1. If ETD or IT-CID scans have been acquired within datasets, filter these scan events from the datafiles using MSConvert[61] prior to searching with FragPipe.

NOTE: For the open searching parameters outlined below, only beam-type CID data are required.

4.2. Performing open searches in FragPipe

4.2.1. Open FragPipe and click the **Workflow** tab **|** the **workflow** pulldown menu; select the **Open** search option, and click **Add files** to import the data files to be searched into FragPipe (**Figure 2A**).

4.2.2. Click the **Database** tab **| Download** to launch the **download manager**, allowing proteome databases to be downloaded from Uniprot using a Uniprot accession number. Click the **Add decoys and contaminants** option within the **download manager** to incorporate decoy and contaminant proteins into this database.

4.2.3. For more stringent FDR thresholds, click the **Validation** tab and modify the **Filter and report** value from **0.01** to the **required FDR**.

NOTE: The default FragPipe settings will ensure a 1% FDR at the protein level.

4.2.4. Click the **MSfragger** tab. Within the **Peak matching** box, increase the **Precursor mass tolerance** from the default 500 Da to **2,000 Da** to identify large modifications (**Figure 2A**).

4.2.5. Click the **Run** tab and define the location of the outputs of FragPipe. Click the **Run** button to begin the search.

4.3. Using the PSMs identified across datasets (contained within the psm.tsv outputs from FragPipe), identify potential glycans by plotting the frequency of observed delta masses within datasets (**Figure 3**). Create delta mass plots for data searched with Byonic using R scripts[62] accessible via the PRIDE accession PXD018587[30] or PXD027820 (this study) for data searched with MSfragger.

NOTE: Minimal postprocessing of the open searching results is undertaken within these scripts, as the main purpose of these scripts is to aid in the visualization of delta mass profiles. Importantly, the observation of abundant delta masses alone is not proof that a modification is a potential glycan, as assigning delta masses as glycans requires further analysis of the corresponding MS2 events.

4.4. To enable the characterization of glycopeptides within samples, focus on high confidence delta mass identifications, corresponding to assignments with high hyperscores.

4.4.1. To aid in assessing glycopeptide spectra, use peptide annotation tools, such as the Interactive Peptide Spectral Annotator[63], which enables the assignment of peptide-associated ions within spectra, allowing the manual identification of the glycan-associated ions (**Figure 4**).

NOTE: Within the datasets presented here, hyperscores of >30 are considered high scoring, as these correspond to scores within the top 50% of all identified glycopeptides (**Figure 5**).

4.5. With high-confidence glycopeptides assigned, identify commonly observed glycan-associated ions (**Figure 4**) to improve the identification of glycopeptides.

NOTE: By incorporating glycan-associated ions within searches, known as glycan-focused searches, the quality of glycopeptide assignments can be improved.

4.5.1. Click the **MSfragger** tab of FragPipe, incorporate the determined delta masses of the observed glycans into the **Variable modifications** and **Mass Offsets** sections. Add these masses by typing values into the **Variable modifications** and **Mass Offsets** sections with individual masses separated with a **/**. Add the glycan-associated fragment masses of these glycans into the **Glyco/Labile mods** section of MSFragger.

NOTE: **Figure 2B** outlines the key information required for a glycan-focused search of the *A. baumannii* strain AB307-0294.

4.6. Upload all MS data associated with proteomic studies to centralized Proteomic repositories such as the PRIDE or MASSIVE repositories.

NOTE: All data associated with this study have been deposited into the PRIDE proteomic repository and can be accessed via the PRIDE accession: PXD027820 using the Username: reviewer_pxd027820@ebi.ac.uk and password: erNH3WbT.

**REPRESENTATIVE RESULTS:**

To illustrate the utility of open searching for bacterial glycopeptide analysis, the chemical diversity of *O*-linked glycans within three strains of *A. baumannii*—AB307-0294, ACICU, and D1279779—was assessed. The *O*-linked glycoproteomes are highly variable between *A. baumannii* strains as the glycans used for glycosylation are derived from the highly variable capsule loci[44-46]. This chemical diversity makes *A. baumannii* an ideal model system for open searching studies. While the glycoproteomes of the three strains have not previously been assessed, the capsule structures of two of these strains, AB307-0294[64] and ACICU[65], are known, with the capsule of D1279779 yet to be elucidated.

Consistent with the capsule of AB307-0294, open searching of AB307-0294 revealed two dominant delta masses, corresponding to 648.25 Da and 692.28 Da (**Figure 3A**). These masses match the capsule structures previously determined by Russo et al., -β-D-QuiNAc4NR-α-GlcNAc6OAc-α-D-GalNAcA, where the QuiNAc4NR residue corresponds to 2,4-diamino-2,4,6-trideoxy-glucose, modified with either 3-OH-butyrate or an acetyl group[64]. Similarly, the ACICU capsule is known to be composed of a tetrasaccharide K-unit, previously identified as Pse5Ac7Ac-β-D-Glc*p*-β-D-Gal*p*-β-D-Gal*p*-NAc-, corresponding to a predicted mass of 843.31 Da[65]. This capsule structure is consistent with the most frequently observed delta mass within the ACICU data (**Figure 3B**). Finally, open searching analysis of the strain D1279779 revealed the presence of multiple delta masses consistent with 203.08, 794.31, and 1588.62 Da (**Figure 3C**). While the mass 203.08 is consistent with that of a single HexNAc residue, 794.31 and 1588.62 correspond to unknown modifications. It should be noted that the mass 1588.62 is exactly double that of 794.31, suggesting these peptides are decorated with glycan K-unit dimers, as has been previously observed within other *Acinetobacter* glycoproteomes[30,44,46].

Delta mass plots provide a quick and effective way to assess the frequency of observed modifications within samples. However, for complex modifications, such as glycans, the presence of a delta mass alone does not provide the information to define the exact composition of a glycan. To aid in the determination of glycan compositions, the resulting MS/MS data of PSMs assigned to specific delta masses can be used. By focusing on high-confidence glycopeptide assignments (in MSFragger corresponding to PSMs with high hyperscores), manual analysis can be used to further characterize the glycans utilized by a strain for protein glycosylation. It should be noted that high-confidence peptide assignments typically contain robust glycan information that allows the determination of glycan compositions based on Y ions. This information can be used to identify the corresponding fragments where the glycans have remained attached to peptides, as well as to B and oxonium-related ions useful for the assignment of glycans[66]. Detailed guidelines on how to assign glycan compositions have been previously outlined, and readers are

recommended to consult Harvey et al.[67]. Using tools such as the Interactive Peptide Spectral Annotator[63], peptide-associated ions can be easily identified, allowing users to determine the identity of the glycans attached to peptides.

Using the Interactive Peptide Spectral Annotator, glycopeptides identified across these three *A. baumannii* samples were assessed, revealing potential glycan-associated ions. Consider the MS/MS spectra of the AB307-0294 glycopeptide SAGDQAASDIATATDNASAK, decorated with the glycans 648.25 and 692.28 Da (**Figure 4A,B**); under similar fragmentation conditions, the glycan-related ions are highly similar, yet shift in mass by 44.02 Da. Analysis of the glycan ions within these PSMs enables confirmation that the delta masses observed for these glycopeptides correspond to trisaccharides containing four different carbohydrates: dHexNAcNAc (228 Da), dHexNAcNBu (272 Da), HexNAcA (217 Da), and HexNAc (203 Da). It should be noted that when assigning glycan fragments, multiple monosaccharides are prone to losing water when fragmented using higher energy C-trap dissociation. This can be seen in the glycan of ACICU, where the monosaccharide Pse5Ac7Ac (316 Da) leads to the generation of two prominent glycan-related fragments: $MH^+$ 299.12 and 317.13, the difference in mass corresponding to the mass of water (18.01 Da) (**Figure 4C**).

Additionally, when analyzing bacterial glycans, it is common to observe unexpected monosaccharide masses corresponding to sugars not observed within eukaryotes. An example of this can be seen within the glycopeptide PSMs of D1279779, where analysis of the most frequent delta mass 794.31 Da reveals the presence of an unusual monosaccharide dHexNAc (187 Da, observed as an $MH^+$ of 188.09, **Figure 4D**), which has been previously observed within *A. baumannii O*-linked glycans[44]. While high mass accuracy measurements and the characteristic behavior of glycans as labile modifications, which are lost from the peptide backbone, can aid in the determination of potential chemical compositions of glycans, it is advisable to minimize the overinterpretation of MS data. If stereochemistry of specific sugars or the linkage information of a glycan is unknown, it is best practice to use structurally agonistic assignments, including referring to monosaccharides by their specific classes. An example of this is referring to 203.08 Da residues as *N*-acetylhexosamines (HexNAc) and avoiding the assignment of linkage types. If required, it is recommended that additional techniques are utilized to define the exact chemical identities of an unknown modification, such as the use of Nuclear Magnetic Resonance (NMR).

While open searching approaches enable the rapid identification of modified peptides, it is important to note that this searching approach can overlook potential glycopeptides within datasets. Within open searches, glycopeptides can fail to be identified for several reasons, including insufficient peptide fragmentation information or penalization of the assigned peptide due to the presence of multiple unassigned ions within the MS2 event. The latter is especially true for bacterial glycopeptide PSMs, as these spectra can contain glycan-associated fragments not considered by default open searching parameters. As unmatched features adversely impact the scoring of PSMs, minimizing unmatched ions within MS/MS spectra enables the identification of glycopeptides that were initially excluded due to the FDR-controlled minimal score thresholds. To overcome this limitation, the masses defined from the open searches (**Figure 3**) and the manually defined glycan-associated ions (**Figure 4**) can be incorporated into search parameters

to improve the identification of glycopeptides. It is important to note that these glycan-associated ions can be identified manually based on *de novo* determination of the glycan, prior knowledge of common oxonium ions[44,68], or the use of tools that can capture reoccurring ions within spectra, such as the SPectral Immonium Ion Detection tool[69,76].

To demonstrate this, the masses of atypical glycan fragments identified through open searching analysis were incorporated into the search parameters in MSFragger, and glycan-focused searching undertaken (the glycan-focused settings for strain AB307-0294 are shown in **Figure 2B**). It should be noted that when multiple glycans are being searched together, care should be taken in the selection of Y ion and diagnostic masses to ensure they accurately reflect the fragment ions associated with each delta mass. While this may not always be possible for combinations of Y and diagnostic masses that are mutually exclusive, such as the fragments of the 648.25 and 692.28 glycans of AB307-0294 (**Figure 4A,B**), glycan-focused searching is still possible, albeit this may impact the robustness of the search results. Glycan-focused searches lead to a notable increase in the total number of glycopeptide PSMs identified across all three *A. baumannii* strains, corresponding to a 37% increase in ACICU, a 117% increase in AB307-0294, and a 363% increase in D1279779 (**Figure 5A–C**). For individual MS events, the inclusion of glycan-specific information typically results in an increase within the observed hyperscore, albeit this increase is highly glycan-dependent (**Figure 5D,E**). Thus, by refining the search parameters using the masses of glycans revealed through open searching and subsequently incorporating manually curated glycan fragment information into targeted searches, a more detailed analysis of glycopeptides within samples can be obtained.

**FIGURE AND TABLE LEGENDS:**

**Figure 1: Overview of the key steps in the preparation and analysis of bacterial glycopeptides.** The successful identification of bacterial glycopeptides is dependent on the generation of high-quality samples containing glycopeptides. Glycopeptide-containing samples are then analyzed by LC-MS, and subsequently, open searching approaches are used to identify potential glycopeptides based on unique delta masses. Using manual curation, these delta masses can be identified as glycans. Finally, to improve the identification of glycopeptides, glycan information can be incorporated into glycan-focused database searches. Abbreviations: LC-MS = liquid chromatography coupled to mass spectrometry; OD = optical density; SDB-RPS = styrenedivinylbenzene-reverse-phase sulfonate; ZIC-HILIC = Zwitterionic Hydrophilic Interaction Liquid Chromatography; PSMs = peptide-spectral matches; CID = collision-induced dissociation; ETD = electron transfer dissociation.

**Figure 2: Overview of how to enable open searching and glycan-focused searches in FragPipe.** (**A**) Open searching can be enabled by loading the **Open** Workflow within the **Workflow** tab of FragPipe. To enable the identification of glycans greater than 500 Da in size, increase the Precursor mass tolerance window to 2,000 Da. (**B**) Glycan-focused searches for atypical glycans require the entry of glycan information in the **MSFragger** tab to define the expected masses of the modifications (in the **Variable modifications** and **Mass Offsets** sections), the Y ions associated with these glycans (in the **Y Ion Masses** list of the **Glyco/Labile** section), and the low-mass glycan

fragment ions (into the **Diagnostic Fragment Masses** list of the **Glyco/Labile** section).

**Figure 3: Delta mass plots of open searching results for the three strains of *Acinetobacter baumannii* (AB307-0294, ACICU, and D1279779).** Consistent with the diversity of the capsule loci, each *A. baumannii* strain possesses a unique delta mass profile with prominent modifications of greater than 140 Da in size, denoted by the observed delta masses. (**A**) Within the AB307-0294 delta mass plot, the masses 648.25 and 692.28 correspond to modifications consistent with -β-D-QuiNAc4NR-α-GlcNAc6OAc-α-D-GalNAcA-, where the QuiNAc4NR is modified with either a 3-OH-butyrate or acetyl group[64]. (**B**) Within the ACICU delta mass plot, the masses 203.08 and 843.31 correspond to modifications consistent with HexNAc and Pse5Ac7Ac-β-D-Glc*p*-β-D-Gal*p*-β-D-Gal*p*-NAc, respectively[65]. (**C**) Within the D1279779 delta mass plot, the masses 203.08, 794.31, and 1588.62 correspond to HexNAc, HexNAc-217-187-187, and a dimer of HexNAc-217-187-187. Delta masses manually assessed and confirmed in **Figure 4** to be glycans denoted by *.

**Figure 4: MS/MS characterization of delta masses observed across the three strains of *Acinetobacter baumannii.*** (**A, B**) AB307-0294, (**C**) ACICU, and (**D**) D1279779. Interactive Peptide Spectral Annotator-assisted annotated spectra of representative glycopeptides. For each spectrum, the peptide fragment ions are denoted in blue and red, while the manually assigned glycan-associated ions are denoted in green. Glycans have been annotated using the Symbol Nomenclature for Glycans with HexNAc denoted as a square, Hex as a circle, and atypical monosaccharides denoted as trapezoids with the mass of the glycan denoted in the trapezoids. Abbreviation: MS/MS = tandem mass spectrometry.

**Figure 5: Comparison of glycopeptides identified across strains of *Acinetobacter baumannii* using open vs glycan-focused searching.** (**A–C**) Comparison of hyperscores and number of PSM assignments within D1279779, ACICU, and AB307-0294 using open and glycan-focused searching. (**D–F**) Comparison of individual MS2 events assigned to the same peptide sequence using open and glycan-focused searches for D1279779, ACICU, and AB307-0294. Abbreviation: PSM = peptide-spectral match.

**Table 1. Strain and Uniprot Proteome IDs used within this study.**

**Table 2: Composition of buffers used in this study.**

**DISCUSSION:**

Open searching is an effective and systematic method for the identification of unknown modifications. While the identification of unknown glycans within bacterial proteome samples has traditionally been a time-consuming and technically specialized undertaking, the recent developments of tools such as MSfragger[21,41] and Byonic[31,38] now enable the quick and effective identification of delta masses for further characterization as potential glycans attached to peptides. Open searching of both standard and glycopeptide-enriched proteome samples can identify potential glycopeptides (**Figure 1**). The commonality and abundance of glycopeptides within the proteome of many bacterial systems make the direct detection of these modifications

possible without glycopeptide enrichment[6]. Although glycopeptide enrichment approaches typically still outperform nonenrichment-based methods for many glycosylation systems[70], it is notable that not all glycans are equally amenable to enrichment[71,72]. This fact should be considered when determining the most appropriate approach for identifying glycosylation events using open searching for a given biological sample.

While open searching effectively identifies glycans observed at a high frequency within a sample, it remains ineffective at identifying glycosylation events that rarely occur within datasets. In practical terms, for a unique delta mass of a glycan to be identified, at least 10 PSMs with an identical delta mass must be present within samples. This minimum frequency allows a potential glycan to be distinguishable above background delta mass assignments (**Figure 3**). While this criterion is typically satisfied by general glycosylation systems targeting multiple substrates for glycosylation, systems where a single protein is targeted for glycosylation may be less amenable to this approach. As the speed and sensitivity of MS instruments improve, it is likely that this will increasingly enable the assessment of lower abundance glycopeptide populations, which will, in turn, enhance the effectiveness of open searching approaches, provided sufficient quality spectra can be generated for rare/low abundant glycoforms.

While MS enables the effective identification of novel glycans, it is important to note that this insight alone rarely provides a complete structural characterization of glycopeptides/glycans, with tools such as NMR spectroscopy still being critical for complete glycan characterization. As observed here, MS provided complementary confirmation of K-units previously determined using NMR from isolated capsules of *A. baumannii* ACICU and AB307-0294[64,65]. However, for the major glycan identified in D1279779, only information about the putative classes of monosaccharides within this glycan was assignable. MS provided no information about the linkage types or the stereochemistry of these sugar units (**Figure 4**). As new MS fragmentation methods become more widely available, such as ultraviolet photodissociation[73,74], more detailed structural characterization of glycopeptides may be possible. However, care should be taken when inferring linkages or stereochemistry information with current instrumentation. In addition to these considerations, it should be noted that questions have been raised recently about the appropriateness of FDR-based controls for open searching[75]. Thus, while open searching is a powerful approach, these considerations highlight the need for care to be taken in the interpretation of glycan assignments based on open searching information alone.

While this demonstrates that open searching is an effective approach for the unsupervised identification of glycans used for bacterial glycosylation, open searching alone provides access to only a subset of all glycopeptides contained within samples. Due to the labile nature of glycans, glycopeptide PSMs typically contain a diverse range of glycan-associated fragments that, if not accounted for, will adversely impact the scoring of PSMs. To overcome this issue and allow more in-depth identification of glycopeptides, the glycan information obtained through open searching can subsequently be used to inform glycan-focused searches (**Figure 2B**). In this way, open searching, coupled with the manual determination of glycan-associated fragments, can be used to improve the quality and quantity of glycopeptides assigned within samples (**Figure 5**). Although the refinement of glycopeptide search parameters is undertaken manually with current

versions of tools, such as MSfragger, it is likely that as the field matures, these steps will be done in an automated manner. Studies have already demonstrated computational approaches to identify repeating low molecular weight ions associated with specific PTMs[69], as well as strategies to identify oxonium-ions and repeating Y-ion masses associated with unknown glycopeptides[76]. Thus, it is likely that these steps will be incorporated into new iterations of tools such as FragPipe, making it even easier to identify previously unknown glycosylation events within bacterial and eukaryotic samples.

**ACKNOWLEDGMENTS:**

N.E.S is supported by an Australian Research Council Future Fellowship (FT200100270) and an ARC Discovery Project Grant (DP210100362). We thank the Melbourne Mass Spectrometry and Proteomics Facility of The Bio21 Molecular Science and Biotechnology Institute for access to MS instrumentation.

**DISCLOSURES:**

The authors have no conflicts of interest.

**REFERENCES:**

1 Varki, A. et al. (Eds) Essentials of glycobiology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2015).

2 Schaffer, C., Messner, P. Emerging facets of prokaryotic glycosylation. *FEMS Microbiology Reviews.* **41** (1), 49–91 (2017).

3 Abu-Qarn, M., Eichler, J., Sharon, N. Not just for Eukarya anymore: Protein glycosylation in bacteria and archaea. *Current Opinion in Structural Biology.* **18** (5), 544–550 (2008).

4 Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., Guerry, P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology.* **32** (5), 1022–1030 (1999).

5 Koomey, M. O-linked protein glycosylation in bacteria: snapshots and current perspectives. *Current Opinion in Structural Biology.* **56**, 198–203 (2019).

6 Riley, N. M., Bertozzi, C. R., Pitteri, S. J. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Molecular & Cellular Proteomics: MCP.* **20**, 100029 (2020).

7 Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology.* **12** (4), 43R–56R (2002).

8 Hu, H., Khatri, K., Klein, J., Leymarie, N., Zaia, J. A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconjugate Journal.* **33** (3), 285–296 (2016).

9 Moremen, K. W., Tiemeyer, M., Nairn, A. V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nature Reviews Molecular Cell Biology.* **13** (7), 448–462 (2012).

10 Bohm, M. et al. Glycosciences.DB: An annotated data collection linking glycomics and proteomics data (2018 update). *Nucleic Acids Research.* **47** (D1), D1195–D1201 (2019).

11 Kawano, S., Hashimoto, K., Miyama, T., Goto, S., Kanehisa, M. Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. *Bioinformatics.* **21** (21), 3976–3982 (2005).

12 McDonald, A. G., Tipton, K. F., Davey, G. P. A knowledge-based system for display and prediction of *O*-glycosylation network behaviour in response to enzyme knockouts. *PLOS Computational Biology.* **12** (4), e1004844 (2016).
13 Eichler, J., Koomey, M. Sweet new roles for protein glycosylation in prokaryotes. *Trends in Microbiology.* **25** (8), 662–672 (2017).
14 Scott, N. E. et al. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the *N*-linked glycoproteome of *Campylobacter jejuni*. *Molecular & Cellular Proteomics: MCP.***10** (2), M000031–MCP000201 (2011).
15 Russo, T. A. et al. The K1 capsular polysaccharide of *Acinetobacter baumannii* strain 307-0294 is a major virulence factor. *Infection and Immunity.* **78** (9), 3993–4000 (2010).
16 Szymanski, C. M., Wren, B. W. Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology.* **3** (3), 225–237 (2005).
17 Imperiali, B. Bacterial carbohydrate diversity - a brave new world. *Current Opinion in Chemical Biology.* **53**, 1–8 (2019).
18 Čaval, T., Buettner, A., Haberger, M., Reusch, D., Heck, A. J. R. Discrepancies between high-resolution native and glycopeptide-centric mass spectrometric approaches: A case study into the glycosylation of erythropoietin variants. *Journal of the American Society for Mass Spectrometry.* **32** (8), 2099–2104 (2021).
19 Čaval, T., Heck, A. J. R., Reiding, K. R. Meta-heterogeneity: Evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Molecular & Cellular Proteomics: MCP.* **20**, 100010 (2020).
20 Thomas, D. R., Scott, N. E. Glycoproteomics: growing up fast. *Current Opinion in Structural Biology.* **68**, 18–25 (2020).
21 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., Nesvizhskii, A. I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods.* **14** (5), 513–520 (2017).
22 Cao, W. et al. Recent advances in software tools for more generic and precise intact glycopeptide analysis. *Molecular & Cellular Proteomics: MCP.* doi:10.1074/mcp.R120.002090 (2021).
23 Yu, A. et al. Advances in mass spectrometry-based glycoproteomics. *Electrophoresis.* **39** (24), 3104–3122 (2018).
24 Reiding, K. R., Bondt, A., Franc, V., Heck, A. J. R. The benefits of hybrid fragmentation methods for glycoproteomics. *Trends in Analytical Chemistry.* **108**, 260–268 (2018).
25 Riley, N. M., Malaker, S. A., Driessen, M., Bertozzi, C. R. Optimal dissociation methods differ for *N*- and *O*-glycopeptides. *Journal of Proteome Research.* **19** (8), 3286–3301 (2020).
26 Mao, Y. et al. Systematic evaluation of fragmentation methods for unlabeled and isobaric mass tag-labeled *O*-glycopeptides. *Analytical Chemistry.* **93** (32), 11167–11175 (2021).
27 Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R., Smith, L. M. O-Pair Search with MetaMorpheus for O-glycopeptide characterization. *Nature Methods.* **17** (11), 1133–1138 (2020).

28 Choo, M. S., Wan, C., Rudd, P. M., Nguyen-Khuong, T. GlycopeptideGraphMS: Improved glycopeptide detection and identification by exploiting graph theoretical patterns in mass and retention time. *Analytical Chemistry.* **91** (11), 7236–7244 (2019).
29 Maxwell, E. et al. GlycReSoft: a software package for automated recognition of glycans from LC/MS data. *PLoS One.* **7** (9), e45474 (2012).
30 Ahmad Izaham, A. R., Scott, N. E. Open database searching enables the identification and comparison of bacterial glycoproteomes without defining glycan compositions prior to searching. *Molecular & Cellular Proteomics: MCP.* **19** (9), 1561–1574 (2020).
31 Bern, M., Kil, Y. J., Becker, C. Byonic: Advanced peptide and protein identification software. *Current Protocols in Bioinformatics.* **Chapter 13**, Unit 13.20 (2012).
32 Na, S., Bandeira, N., Paek, E. Fast multi-blind modification search through tandem mass spectrometry. *Molecular & Cellular Proteomics: MCP.* **11** (4), M111 010199 (2012).
33 Creasy, D. M., Cottrell, J. S. Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics.* **2** (10), 1426–1434 (2002).
34 Craig, R., Beavis, R. C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics.* **20** (9), 1466–1467 (2004).
35 Ahrné, E., Nikitin, F., Lisacek, F., Müller, M. QuickMod: A tool for open modification spectrum library searches. *Journal of Proteome Research.* **10** (7), 2913–2921 (2011).
36 Shortreed, M. R. et al. Global identification of protein post-translational modifications in a single-pass database search. *Journal of Proteome Research.* **14** (11), 4714–4720 (2015).
37 Chick, J. M. et al. A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature Biotechnology.* **33** (7), 743–749 (2015).
38 Roushan, A. et al. Peak filtering, peak annotation, and wildcard search for glycoproteomics. *Molecular & Cellular Proteomics: MCP.* **20**, 100011 (2020).
39 Chi, H. et al. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nature Biotechnology.* doi:10.1038/nbt.4236 (2018).
40 Bittremieux, W., Laukens, K., Noble, W. S. Extremely fast and accurate open modification spectral library searching of high-resolution mass spectra using feature hashing and graphics processing units. *Journal of Proteome Research.* **18** (10), 3792–3799 (2019).
41 Polasky, D. A., Yu, F., Teo, G. C., Nesvizhskii, A. I. Fast and comprehensive *N*- and *O*-glycoproteomics analysis with MSFragger-Glyco. *Nature Methods.* **17** (11), 1125–1132 (2020).
42 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. *Molecular Microbiology.* **96** (5), 1023–1041 (2015).
43 Iwashkiw, J. A. et al. Identification of a general *O*-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathogens.* **8** (6), e1002758 (2012).
44 Scott, N. E. et al. Diversity within the *O*-linked protein glycosylation systems of *Acinetobacter* species. *Molecular & Cellular Proteomics: MCP.* **13** (9), 2354–2370 (2014).
45 Kenyon, J. J., Hall, R. M. Variation in the complex carbohydrate biosynthesis loci of *Acinetobacter baumannii* genomes. *PLoS One.* **8** (4), e62160 (2013).
46 Lees-Miller, R. G. et al. A common pathway for *O*-linked protein-glycosylation and synthesis of capsule in *Acinetobacter baumannii*. *Molecular Microbiology.* **89** (5), 816–830 (2013).

47 Iacono, M. et al. Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group. *Antimicrobial Agents and Chemotherapy.* **52** (7), 2616–2625 (2008).
48 Hamidian, M. et al. Insights from the revised complete genome sequences of *Acinetobacter baumannii* strains AB307-0294 and ACICU belonging to global clones 1 and 2. *Microbial Genomics.* **5** (10), e000298 (2019).
49 Farrugia, D. N. et al. The complete genome and phenome of a community-acquired *Acinetobacter baumannii*. *PLoS One.* **8** (3), e58628 (2013).
50 Keller, B. O., Sui, J., Young, A. B., Whittal, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Analytica Chimica Acta.* **627** (1), 71–81 (2008).
51 Batth, T. S. et al. Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Molecular & Cellular Proteomics: MCP.* **18** (5), 1027–1035 (2019).
52 HaileMariam, M. et al. S-Trap, an ultrafast sample-preparation approach for shotgun proteomics. *Journal of Proteome Research.* **17** (9), 2917–2924 (2018).
53 Smith, P. K. et al. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry.* **150** (1), 76–85 (1985).
54 Rappsilber, J., Mann, M., Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols.* **2** (8), 1896–1906 (2007).
55 Harney, D. J. et al. Proteomic analysis of human plasma during intermittent fasting. *Journal of Proteome Research.* **18** (5), 2228–2240 (2019).
56 Scott, N. E. Characterizing glycoproteins by mass spectrometry in *Campylobacter jejuni*. *Methods in Molecular Biology.* **1512**, 211–232 (2017).
57 Bian, Y. et al. Robust microflow LC-MS/MS for proteome analysis: 38000 runs and counting. *Analytical Chemistry.* **93** (8), 3686–3690 (2021).
58 Diedrich, J. K., Pinto, A. F., Yates, J. R., 3rd. Energy dependence of HCD on peptide fragmentation: stepped collisional energy finds the sweet spot. *Journal of the American Society for Mass Spectrometry.* **24** (11), 1690–1699 (2013).
59 Liu, M. Q. et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nature Communications.* **8** (1), 438 (2017).
60 Hinneburg, H. et al. The art of destruction: Optimizing collision energies in quadrupole-time of flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. *Journal of the American Society for Mass Spectrometry.* **27** (3), 507–519 (2016).
61 Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology.* **30** (10), 918–920 (2012).
62 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, http://www.r-project.org/index.html (2020).
63 Brademan, D. R., Riley, N. M., Kwiecien, N. W., Coon, J. J. Interactive peptide spectral annotator: A versatile web-based tool for proteomic applications. *Molecular & Cellular Proteomics: MCP.* **18** (8 suppl 1), S193–S201 (2019).

64 Russo, T. A. et al. The K1 capsular polysaccharide from *Acinetobacter baumannii* is a potential therapeutic target via passive immunization. *Infection and Immunity.* **81** (3), 915–922 (2013).
65 Senchenkova, S. N. et al. Structure of the capsular polysaccharide of *Acinetobacter baumannii* ACICU containing di-*N*-acetylpseudaminic acid. *Carbohydrate Research.* **391**, 89–92 (2014).
66 Domon, B., Costello, C. E. A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates. *Glycoconjugate Journal.* **5** (4), 397-409 (1988).
67 Harvey, D. J., Dwek, R. A., Rudd, P. M. Determining the structure of glycan moieties by mass spectrometry. *Current Protocols in Protein Science.* **Chapter 12,** Unit 12.7 (2006).
68 Medzihradszky, K. F., Kaasik, K., Chalkley, R. J. Characterizing sialic acid variants at the glycopeptide level. *Analytical Chemistry.* **87** (5), 3064–3071 (2015).
69 Kelstrup, C. D., Frese, C., Heck, A. J., Olsen, J. V., Nielsen, M. L. Analytical utility of mass spectral binning in proteomic experiments by SPectral Immonium Ion Detection (SPIID). *Molecular & Cellular Proteomics:MCP.* **13** (8), 1914–1924 (2014).
70 Ahmad Izaham, A. R. et al. What are we missing by using hydrophilic enrichment? Improving bacterial glycoproteome coverage using total proteome and FAIMS analyses. *Journal of Proteome Research.* **20** (1), 599–612 (2021).
71 Neue, K., Mormann, M., Peter-Katalinic, J., Pohlentz, G. Elucidation of glycoprotein structures by unspecific proteolysis and direct nanoESI mass spectrometric analysis of ZIC-HILIC-enriched glycopeptides. *Journal of Proteome Research.* **10** (5), 2248–2260 (2011).
72 Mysling, S., Palmisano, G., Hojrup, P., Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Analytical Chemistry.* **82** (13), 5598–5609 (2010).
73 Escobar, E. E. et al. Precision mapping of *O*-Linked *N*-acetylglucosamine sites in proteins using ultraviolet photodissociation mass spectrometry. *Journal of the American Chemical Society.* **142** (26), 11569–11577 (2020).
74 Madsen, J. A. et al. Concurrent automated sequencing of the glycan and peptide portions of *O*-linked glycopeptide anions by ultraviolet photodissociation mass spectrometry. *Analytical Chemistry.* **85** (19), 9253–9261 (2013).
75 Lysiak, A., Fertin, G., Jean, G., Tessier, D. Evaluation of open search methods based on theoretical mass spectra comparison. *BMC Bioinformatics.* **22** (Suppl 2), 65 (2021).
76 Pabst, M. et al. A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *ISME Journal.* doi:10.1038/s41396-021-01073-y (2021).

Figure 1



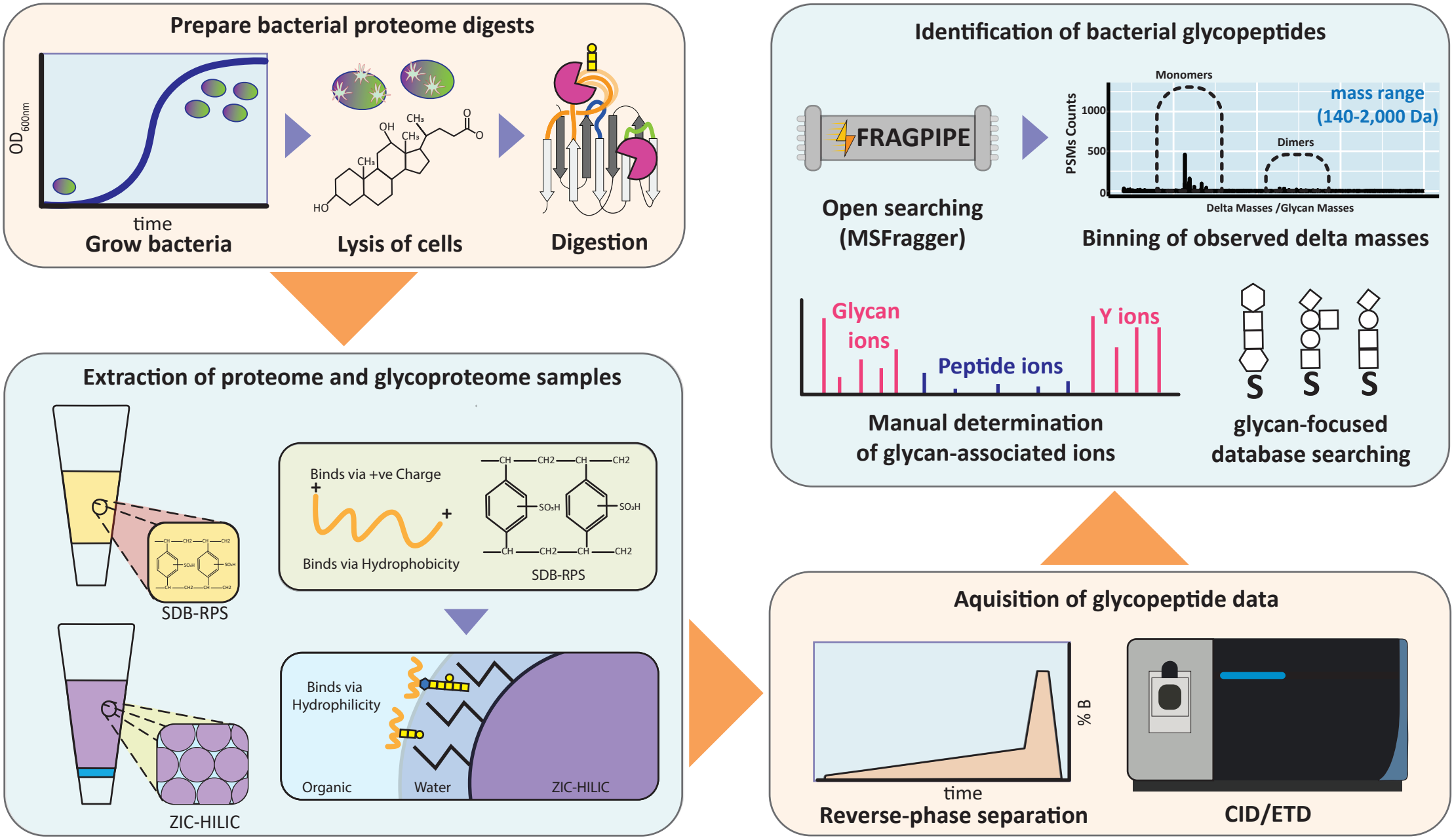
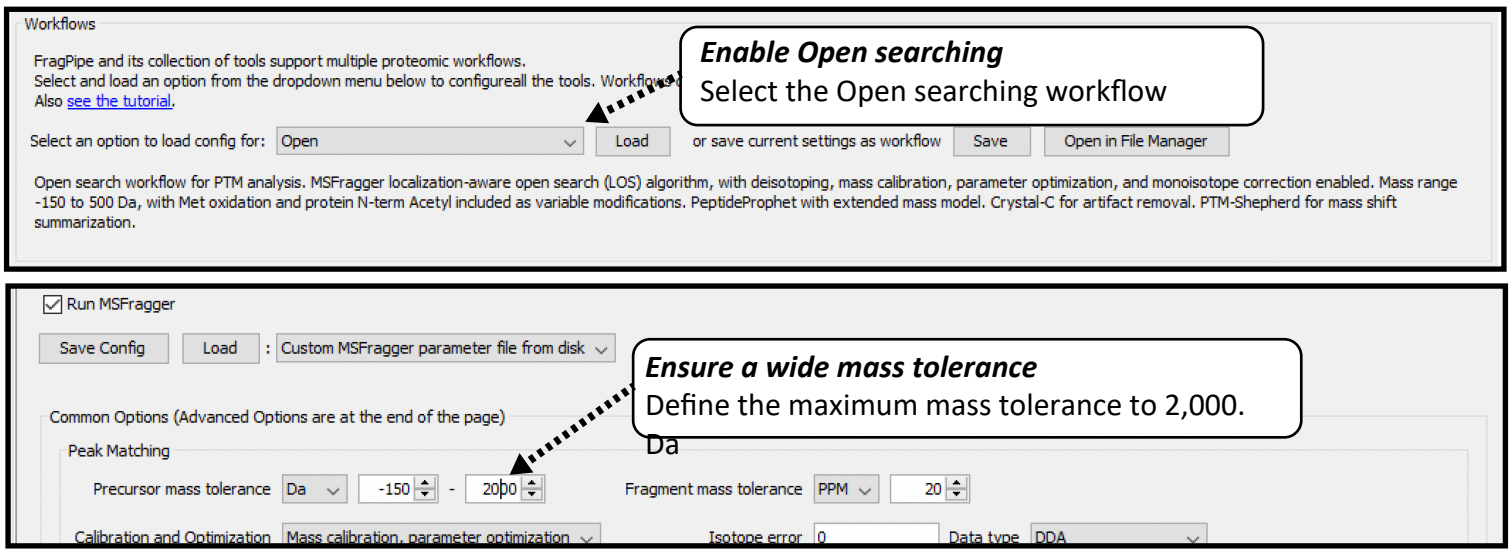


Figure 2

Click here to access/download;Figure;Figure2.pdf

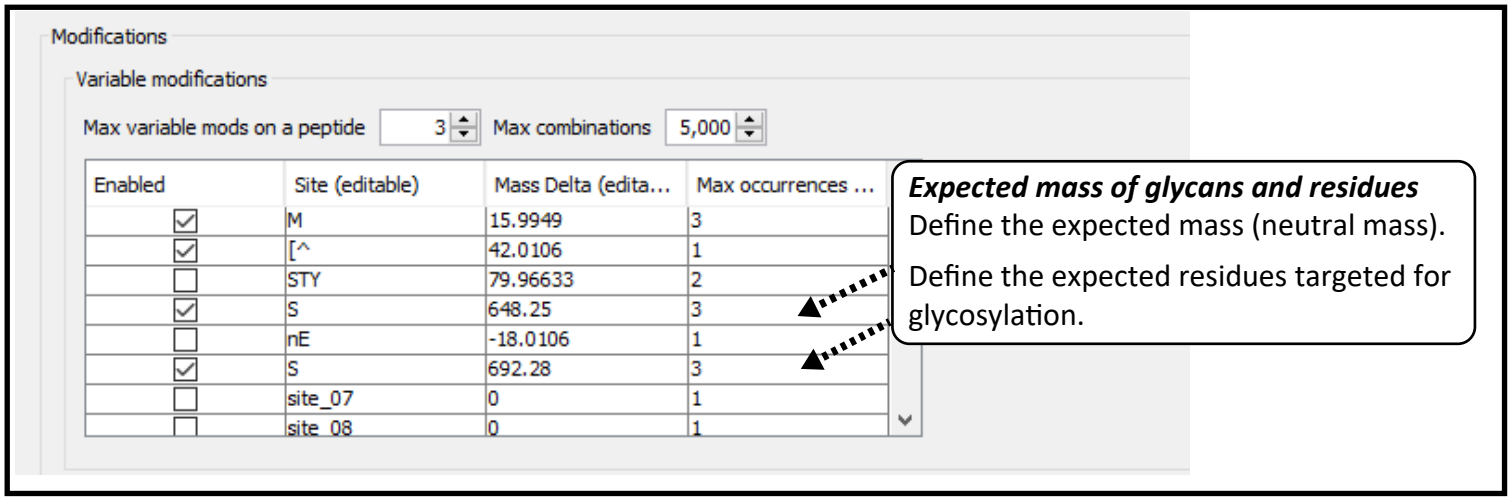
# A) *Open searching*







# B) *Glycan-focused search*



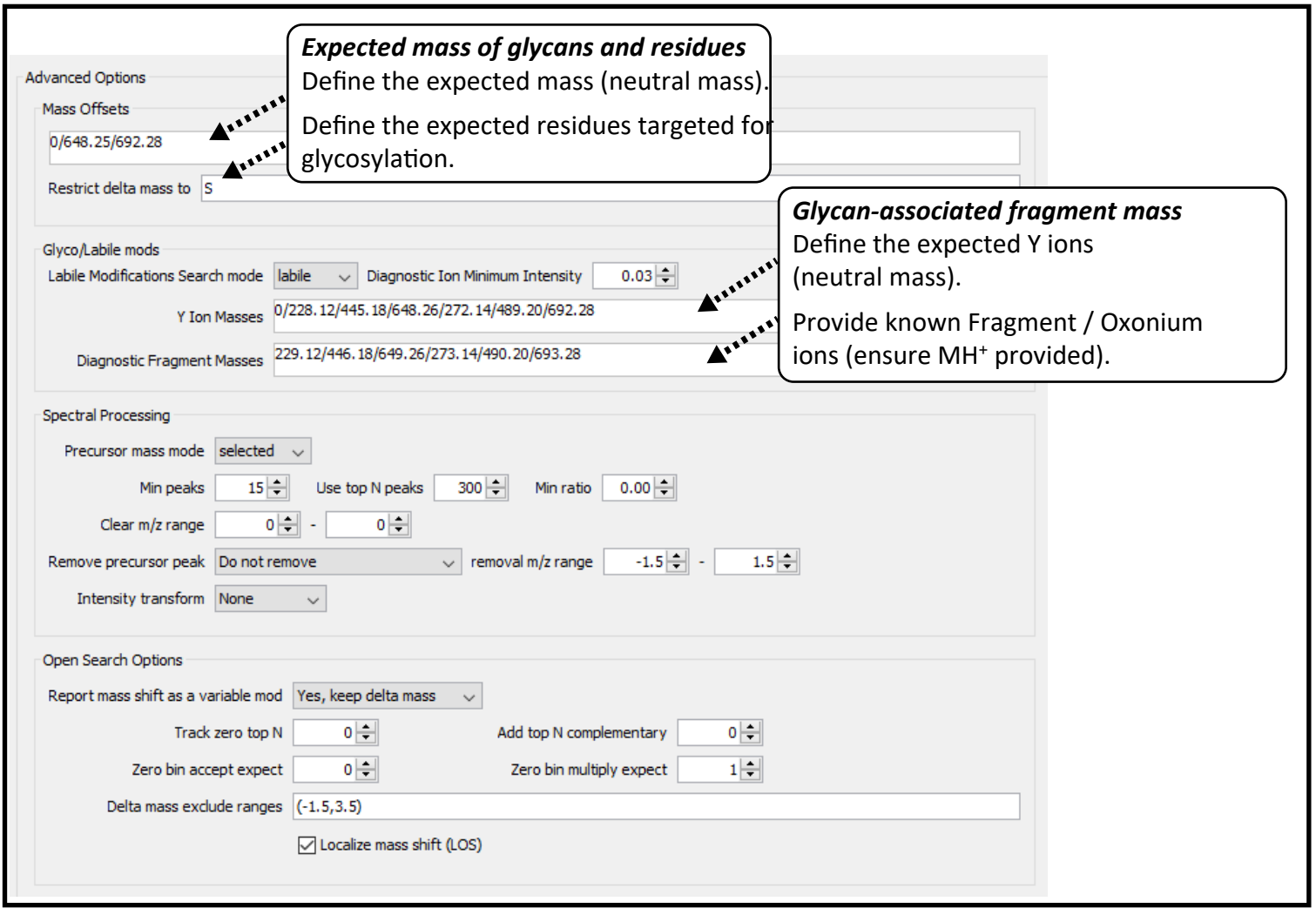



Figure 3



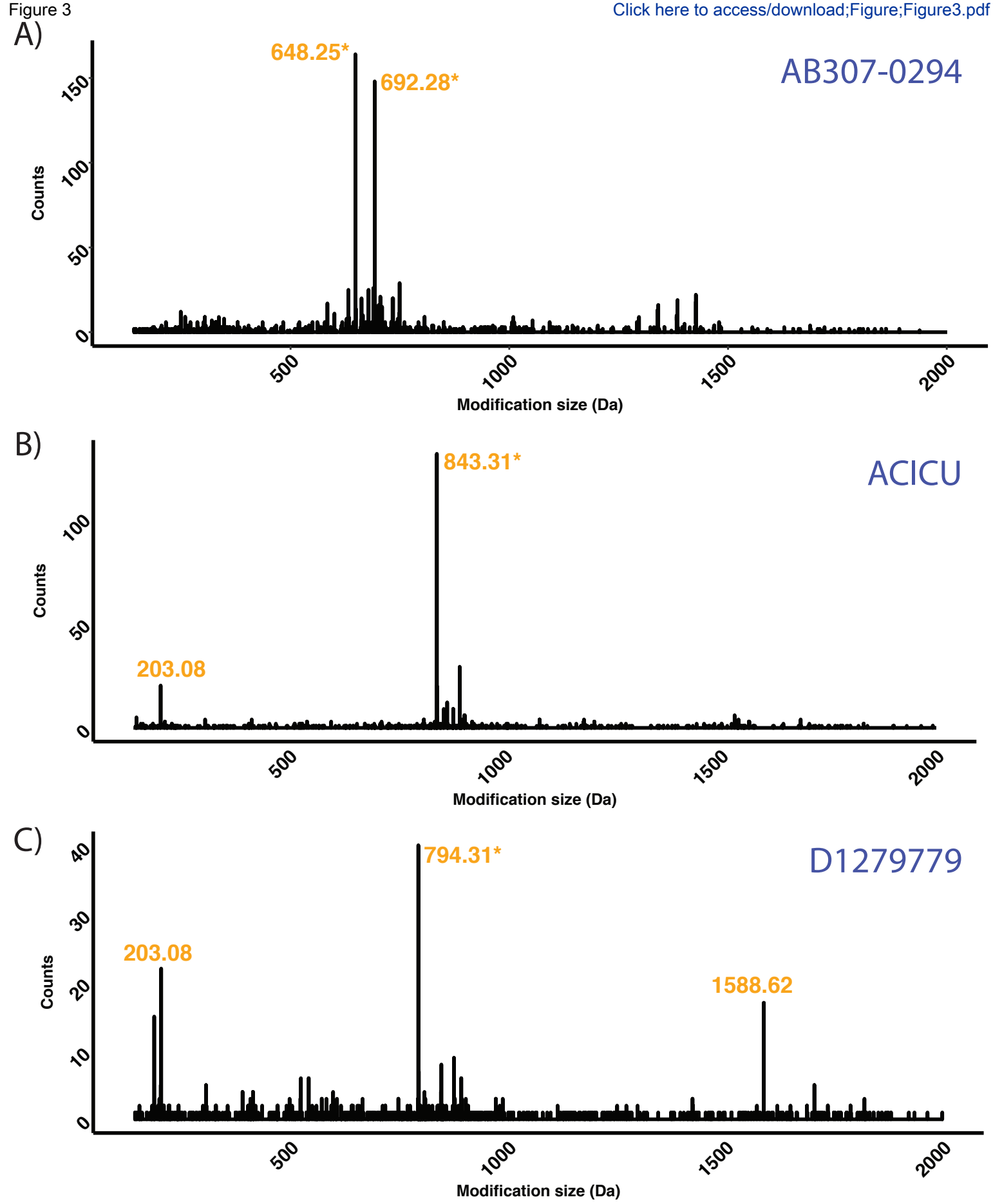


Figure 4

Click here to access/download;Figure;Figure4.pdf

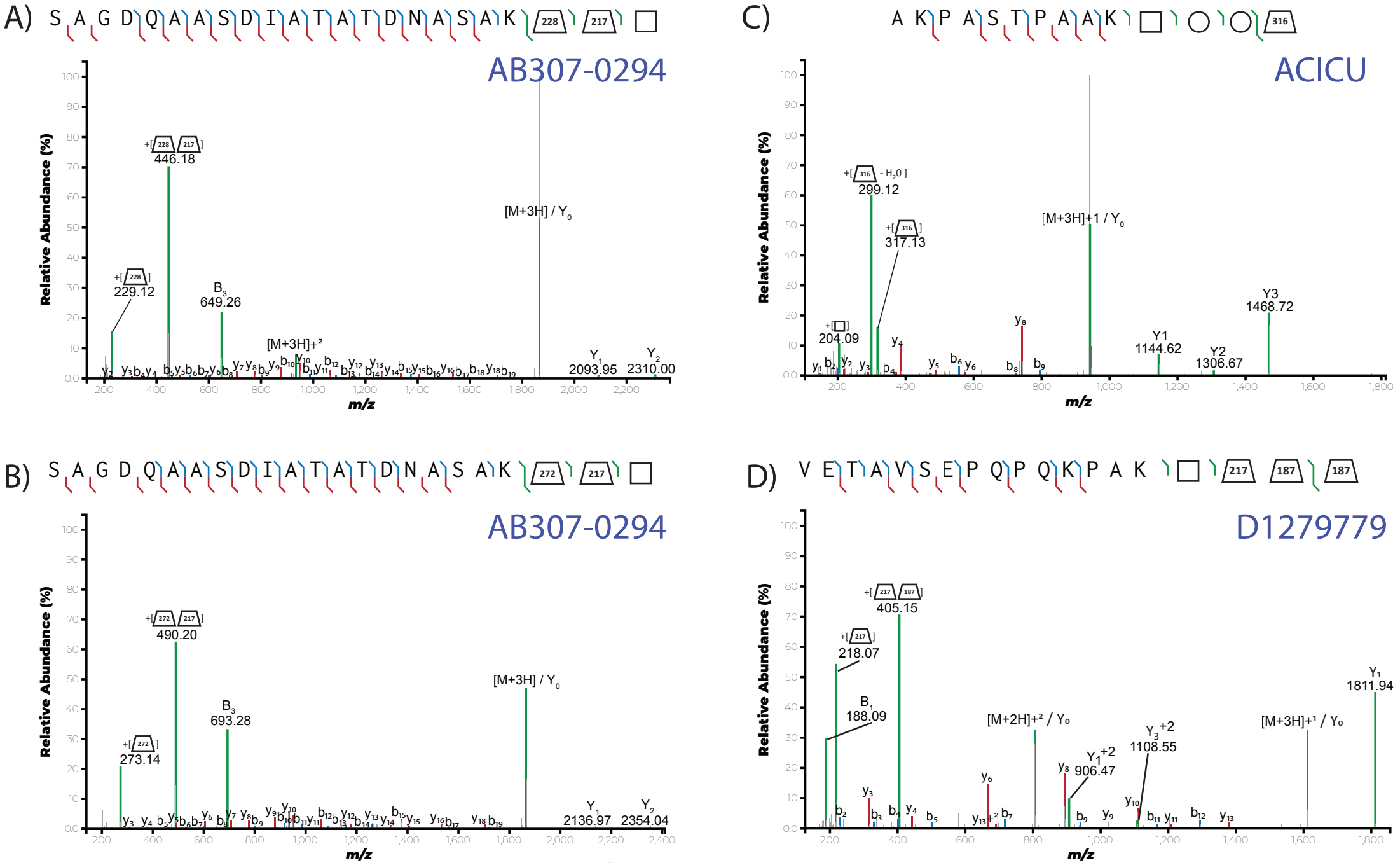


Figure 5



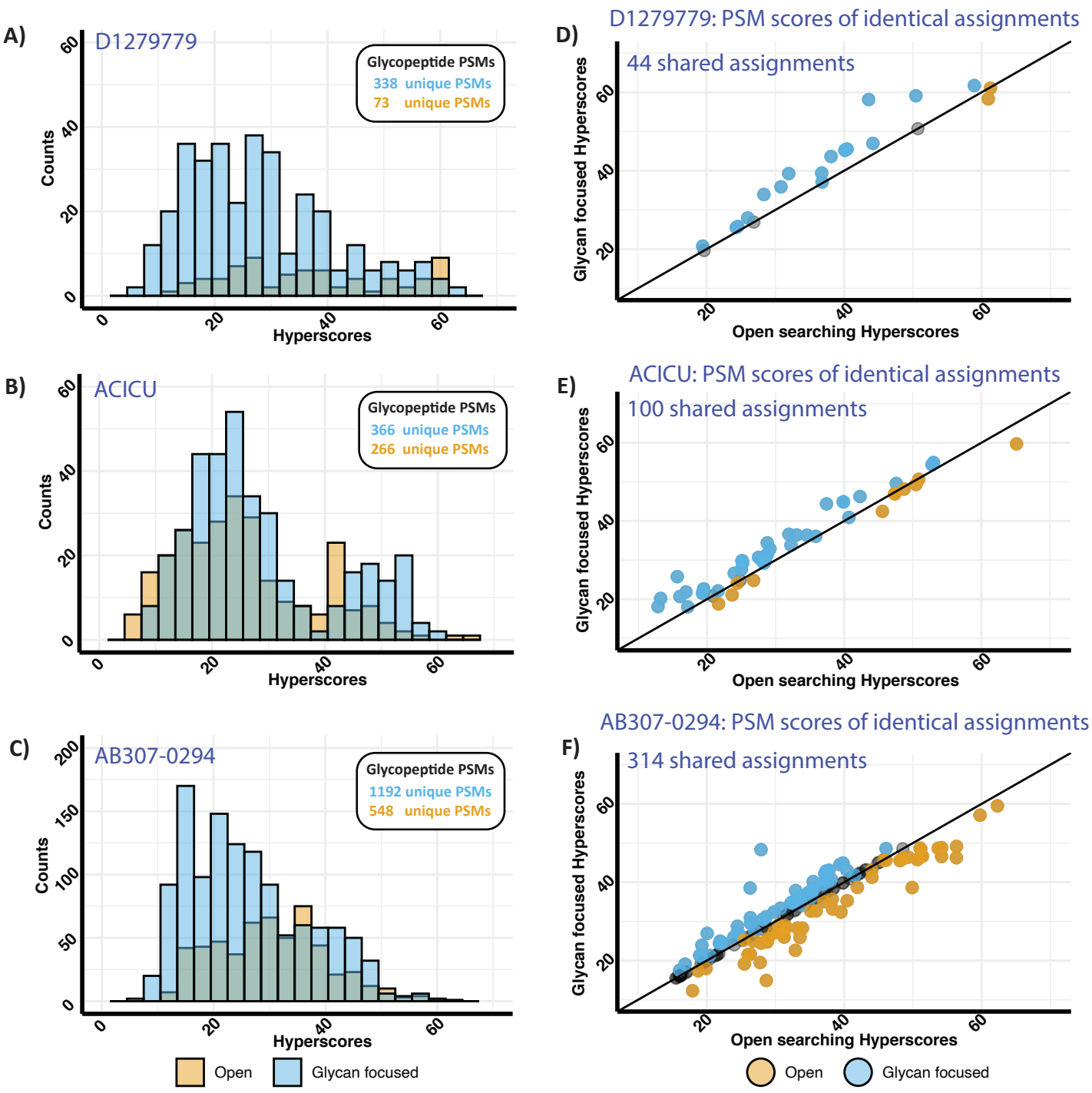


Table 1



| ***A. baumannii* Strains** | **Uniprot Proteome IDs** |
|---|---|
| AB307-0294 [15] | UP000006924 |
| ACICU [47,48] | UP000008839 |
| D1279779 [49] | UP000011860 |

Table 2

Click here to access/download;Table;JoVE_Table_of_Buffer_compositions_20

| Buffer Name | Buffer composition / pH |
|---|---|
| Buffer A* | 2% acetonitrile in 18.2 MΩ H2O, 0.1% trifluroacetic acid / pH 1.0 |
| PBS | 137 mM NaCl, 2.7 mM KCl, 8 mM $Na_2HPO_4$, and 2 mM $KH_2PO_4$ / pH 7.4 |
| Reduction/alkylation buffer | 100 mM Tris 2-carboxyethyl phosphine hydrochloride, 400 mM 2-Chloroacetamide in 1 M Tris / pH 7.5 |
| SDB-RPS Stage Tip elution buffer | 5% ammonium hydroxide in 80% acetonitrile / pH 11 |
| SDB-RPS Stage Tip equlibration buffer | 30% methanol, 1% trifluroacetic acid, in 18.2 MΩ $H_2O$ / pH 1.0 |
| SDB-RPS Stage Tip equlibration/wash buffer | 90% isopropanol, 1% trifluroacetic acid / pH 1.0 |
| SDB-RPS Stage Tip wash buffer | 1% trifluroacetic acid in 18.2 MΩ $H_2O$ / pH 1.0 |
| SDC lysis buffer | 4% SDC in 100 mM Tris / pH 8.5 |
| ZIC-HILIC elution buffer | 0.1% trifluroacetic acid in 18.2 MΩ $H_2O$ /pH 1.0 |
| ZIC-HILIC loading/wash buffer | 80% acetonitrile, 1% trifluroacetic acid in 18.2 MΩ $H_2O$ / pH 1.0 |
| ZIC-HILIC preparation buffer | 95% acetonitrile in 18.2 MΩ $H_2O$ / pH 7.0 |

Doctor Nichollas Scott
ARC Future Fellow and Group leader
Department of Microbiology and Immunology, University of Melbourne
The Peter Doherty Institute for Infection and Immunity
Victoria 3000, Australia
Nichollas.scott@unimelb.edu.au

20th of October 2021

Dear JoVE Editors,

Please find enclosed our revision JoVE63242R1 submission entitled "**The application of open searching based approaches for the identification of *Acinetobacter baumannii* O-linked glycopeptides**" by Lewis *et al.* for consideration for publication within JoVE.

We were delighted with the very supportive comments from the reviewers and that they were satisfied with our corrections to the original manuscript. We feel their feedback has been instrumental in improving the clarity and completeness of the manuscript.

To aid in the revision of this manuscript we have provided a detailed point by point reply to all questions/comments raised by the reviewers. In addition, we have also provided a marked version of the manuscript where all alterations associated with the initial resubmission are highlighted in red while the new corrections associated with this revision are in dark red.

Again, we thank you for considering our work for publication in JoVE and we look forward to hearing from you soon.

Sincerely

Nichollas E. Scott
**ARC Future Fellow and Group Leader**
**The Department of Microbiology and Immunology**
**University of Melbourne, Australia**

**Reviews comments: (Responses comments denoted in dark red)**

**Reviewer #1:**

Manuscript Summary:

The authors have greatly improved the manuscript, especially the description of how glycans are identified from the delta mass results, and what the limitations of this approach are. Only minor comments remain.

We thank the reviewer for their supportive comments and have addressed the minor concerns listed below.

Major Concerns:

None

Minor Concerns:

1. Line 29: "breath" should be "breadth"

We thank the reviewer for highlighting this error, it has now been corrected.

2. Figure 2: I would still recommend that within MSFragger-Glyco, the mixing of different glycans and their respective Y-ion masses and diagnostic fragment masses should be avoided. As mentioned before, the selection of parameters as presented by the authors in Figure 2 includes the search for fragment ions like 229 and 446 Da for all defined mass offsets (648 and 692 Da) instead of only for the specific 648 Da mass offset. Currently, MS2 spectra with a mass offset of 692 Da, containing ions corresponding to 229 and 446 Da will receive an increased score, although these fragment ions are not indicative for the correct identification of the 692 Da glycan. Therefore, mixing different glycans and their corresponding fragment ions (as diagnostic ions) will affect the robustness of the search results.

The reviewer is correct that listing multiple mass offsets which are observed mutually exclusively could impact searches, yet this is largely unavoidable in the current version of MSFragger-Glyco making this a limitation of the software. As we agree this should at least be highlighted to users we have added the following statement to the results section:

“It should be noted when multiple glycans are being searched together care should be taken in the selection of Y ion and diagnostic masses to ensure they accurately reflect the fragment ions associated with each delta mass. While this may not always be possible for combinations of Y and diagnostic masses which are mutually exclusive, such as the fragments of the 648.25 and 692.28 glycans of AB307-0294 (Figure 4A and B), glycan-focused searching is still possible albeit this may impact the robustness of the search results.” [see line 425 to 431]

3. In lines 397 - 406, the authors describe very well that high mass accuracy MS can be used to obtain information about the composition of glycans, but over interpretation in regard to glycan structures should be avoided, since it requires other techniques like NMR. Yet, the authors refer

to the identification of "glycan structures" based on MS data throughout the manuscript. Using a different terminology (e.g. glycan composition, or monosaccharide composition) could be helpful to avoid confusion.

We thank the reviewer for their suggestion and agree the use of the term "glycan structures" is not ideal and have changed this to “glycan composition” throughout the manuscript.

Furthermore, in their response, the authors explained the appropriate use of structurally agnostic assignments ("The overinterpretation we aimed to highlight to be avoided was the assignment of structural information including glycan linkages, the stereochemistry of the specific sugars being assigned […]. Within this document we have chosen to use structural agonistic assignments such as referring to monosaccharides."). However, in line 386/387 the authors still refer to "6-dHexNAc4NAc" and "6-dHexNAc4NBu" which are claimed to be assigned based on the glycan ions of the corresponding PSMs. Removing the structural information by just naming the monosaccharides without information about the position of the modifications would seem more appropriate here.

We thank the reviewer for highlighting this oversight, we have now modified these assignments to "dHexNAcNAc" and "dHexNAcNBu" removing the assigned position information.

4. It seems like A. baumannii glycans do not contain pentose, so a filtering of delta masses for >140 Da is appropriate. But it might be worth noting that pentose is a common monosaccharide used for protein glycosylation in other species, so filtering e.g. for >130 Da would be suggested as a more general approach.

While the reviewer is correct as the script for the creation of the delta mass figures are provided and can be modify to any delta mass threshold desired, we feel highlighting this point is unnecessary.

**Reviewer #2:**

Manuscript summary:

The authors sufficiently addressed all raised concerns, and the referee suggests only a few additional changes in order to improve accuracy and clarity of the manuscript. Overall, Lewis and co-workers did a great job in constructing a user guideline for an open source tool, which will be of great value for the (microbial) glycoproteomics community.

We thank the reviewer for their supportive comments and have now addressed the minor concerns listed below

Major suggestions

-> Table 1: The "UP…X" numbers are "Uniprot Proteome IDs" rather than "Uniprot Accession numbers".

We thank the reviewer for highlighting this oversight and have made this correction.

-> Line 161: "less than 2 µg of peptide per analysis", should be rather "than 2 µg of protein digest per analysis"

We thank the reviewer for highlighting this oversight and have made this correction.

-> Line 417-425 (original comment/response see below, italic): The SPectral Immonium Ion Detection tool performs binning and supports the ID of stable amino acid modifications (not novel sugar oxonium ions). The extension of this approach is described in reference 76, which supports the de novo identification of (novel) sugar oxonium ions. For clarity/completeness possibly add here also reference 76 (ISME J, 2021).

The reference has been added and we thank the reviewer for highlighting this oversight.

Additional minor suggestions

-> Abstract, line 44: " of these unusual glycoproteomes easier than ever before." In fact, those (diverse) glycoproteomes are the majority (considering their biomass on earth). Maybe rather "highly diverse glycoproteomes".

We thank the reviewer for their suggestion and have modified the sentence accordingly.

-> Line 116: "the identification of atypical glycopeptides". Bacterial glycans are only "atypical" if compared to mammalian glycosylation. Possibly the authors wanted to say "novel"?

We thank the reviewer for their suggestion and have modified the sentence accordingly.