

# Glycoproteomics analysis to examine the role of chlamydial protease-like activity factor

Christa Feasley<sup>1</sup>, Fred Zinnel<sup>2</sup>, Julian Saba<sup>3</sup>, Stuart McCorrister<sup>4</sup>, Garrett Westmacott<sup>4</sup> and Grant McClarty<sup>4</sup>, Chris Grant<sup>4</sup>

<sup>1</sup>Thermo Fisher Scientific, West Palm Beach, FL, USA; <sup>2</sup>Thermo Fisher Scientific, Somerset, NJ, USA; <sup>3</sup>Thermo Fisher Scientific, Mississauga, ON, Canada; <sup>4</sup>National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada

## OVERVIEW

**Purpose:** Examination of the role the chlamydial protease-like activity factor targets has in altering protein expression and relative abundance of post translational modification by monitoring global glycopeptide changes.

**Methods:** HeLa 229 cells were infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain, enriched for glycopeptides and analyzed on a Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer. Thermo Scientific™ Proteome Discoverer™ software with Byonic node was used for data analysis.

**Results:** Overall we identified over 5000 unique intact glycopeptides translating into over 600 unique glycoproteins. Number of unique glycopeptides/glycoforms with levels greater than two-fold change were observed upon infection with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient and -deficient strains.

## INTRODUCTION

Recently proteomics studies have been performed to examine the role of chlamydial protease-like activity factor (CPAF), a secreted virulence factor and its role in immune evasion.[1] However, experiments thus far have ignored the presence of glycopeptides that are prevalent in the sample, as large scale intact glycopeptide analysis remains challenging by mass spectrometry due to the complexities associated with the glycopeptide structure. Not only must one sequence the peptide backbone, but also site localization and glycan composition are required for intact glycopeptide analysis. The challenge is further compounded owing to the fact that traditional fragmentation techniques are not ideal for glycopeptide sequencing. Here, we undertake qualitative and quantitative intact glycopeptide analysis to examine the role of CPAF targets and derive additional insights from glycoproteomics experiments.

## MATERIALS AND METHODS

HeLa 229 cells were infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain. 150 ug of proteins were extracted from the cell lysates in triplicates from each infection, digested with trypsin and labeled with Thermo Scientific™ Tandem Mass Tag™ 6-plex Reagents (TMT) per manufacturer's protocol. Proteins from HeLa 229 cells infected with the sufficient strain was labeled with 127 and 130 label, deficient strain with 128 and 131 label and the control HeLa 229 were labeled with 126 and 129 label. Upon labeling samples were enriched for glycopeptides with strong anion exchange (SAX) column.[2] The enriched glycopeptides were analyzed using a Thermo Scientific™ EASY-nLC™ 1000 with a Thermo Scientific™ EASY-Spray™ C18 PepMap column (2um, 100A, 75umx50 cm) on an Orbitrap Fusion Lumos mass spectrometer. [2] Glycopeptide MS2 data was collected by DDA using the following filters: the monoisotopic precursor selection, intensity threshold of 2e5, dynamic exclusion of 30 sec, and ions selected were fragmented both with HCD and ETD. Data analysis were performed using Proteome Discoverer 2.1 software using the Byonic node (Protein Metrics). Proteomics experiment were also performed for the sample. Figure 1 shows the schematic representation of the workflow undertaken.

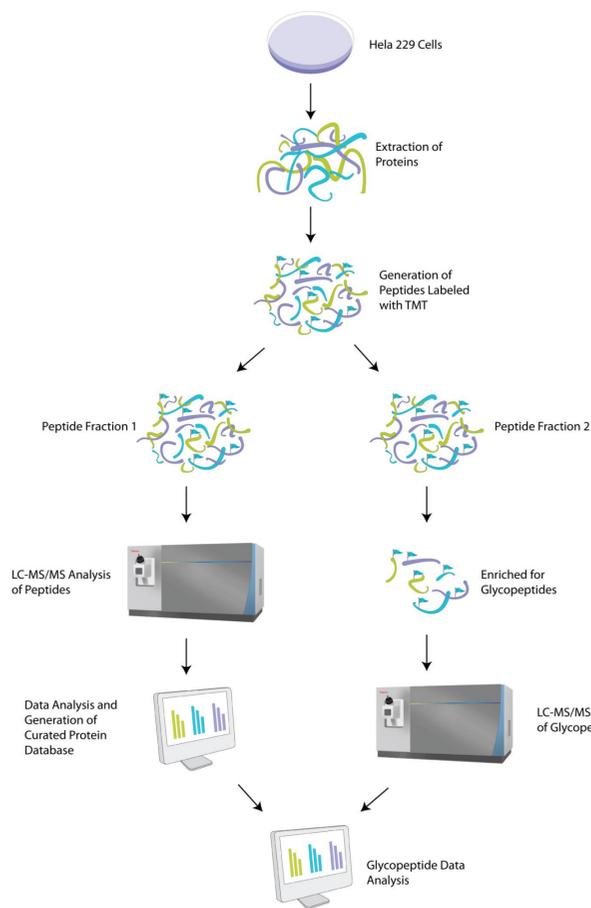
## RESULTS

In a typical proteomics experiment ~40-50% of MS/MS spectra are identified. In recent years researchers have focused on developing numerous software algorithms to sequence the remainder of the spectra with the hypothesis that these are indeed identifiable spectra. The main problem with the hypothesis is that the unidentified spectra are considered to be produced from conventional peptides, ignoring the possibility that spectra could be the result of post-translationally modified peptides. We have observed, depending upon sample type, that some of these unidentified spectra (2-20%) can be the result of glycopeptides. Figure 2 shows the extracted ion chromatogram (XIC) for *m/z* 204.087 from our proteomics LC-MS/MS experiment. The 204.087 ion is extracted from the MS/MS spectra and indicative HexNAc oxonium ion. The presence of this fragment ion in the MS/MS spectrum confirms that the spectra is produced by a glycopeptide. The MS/MS spectrum in the experiment is acquired at a resolution of 60000 at *m/z* 200, which is more than sufficient resolution to be confident that this ion is coming from HexNAc and not something else. Further examination of MS/MS spectra confirms additional glycan oxonium ions. In fact Figure 2 shows that the entire run is full of glycopeptides. The abundance of glycopeptides present in the sample is further illustrated by zooming into narrower windows in the chromatogram. Figure 2 shows 10 and 5 minute retention time window highlighting the magnitude of the glycopeptides present. Unfortunately, conventional fragmentation techniques that are present in commercial mass spectrometers are not ideal for glycopeptide sequencing. Further exacerbating the issue is that these glycopeptides are present in low abundance which makes acquiring quality MS/MS spectra very difficult, even when suitable fragmentations are used. The last panel in Figure 2 shows an example of MS/MS spectrum acquired during this proteomics experiment. Due to insufficient fragmentation information it was impossible to identify this spectrum. So in proteomics experiments like these, glycopeptides are ignored in the overall identification process.

Currently identifying glycopeptides in the presence of non-glycopeptides is challenging due to the limited dynamic range of commercial mass spectrometers. This makes it very difficult to detect all glycoforms. Furthermore, the low abundance of detected glycopeptides makes it very hard to generate quality MS/MS data. In order to resolve these challenges we undertook glycoproteomics

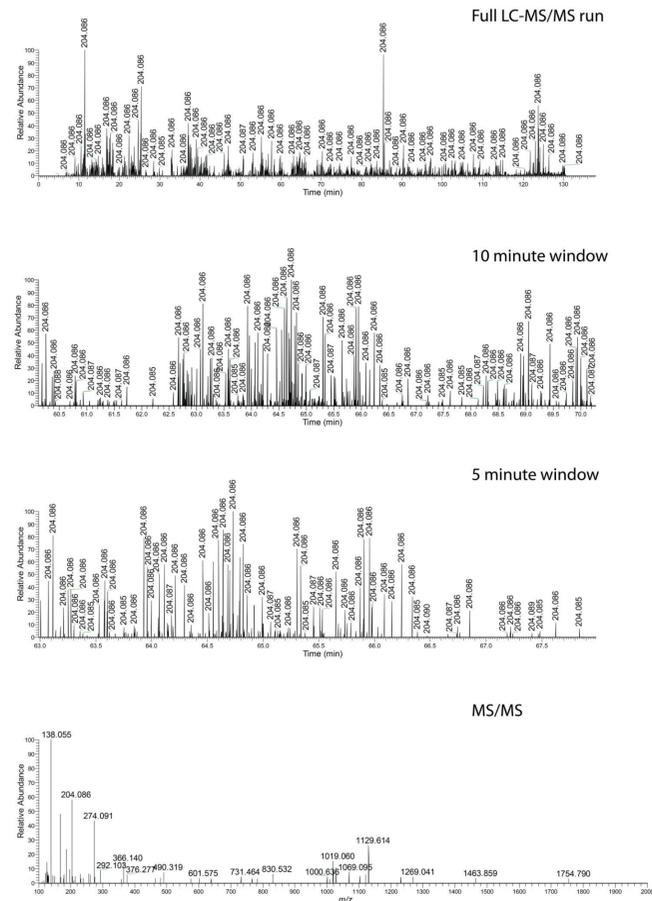
experiments. Glycopeptides were enriched with SAX and targeted with an Orbitrap Fusion Lumos MS using HCD and ETD for sequencing. Figure 3 highlights the advantage of enrichment. 10 and 1 minute window comparison are shown between the proteomics and glycoproteomics LC-MS/MS runs. The glycoproteomics runs shows far more glycopeptides and much higher abundance. This enables the mass spectrometer to target the glycopeptides much more efficiently.

Figure 1. Schematic representation of the workflow.



The glycoproteomics experiments also showed the presence of phosphorylated glycopeptides. This is the presence of phosphorylation on the glycan rather than the peptide. Where mannose 6-phosphate (Man-6-P) occurs on high mannose glycans. Figure 4 shows the XIC for 243.026, which is the fragment ion generated from phosphorylated mannose. Since MS/MS is acquired at a resolution of 60000 at *m/z* 200 we can clearly distinguish it from sulfated species as those would have a fragment mass of 243.162. The presence of additional glycan oxonium ions in the MS/MS spectrum further confirms that it is indeed a glycopeptide. Example of this is shown in the bottom spectrum in Figure 4. Approximately 7% of the glycopeptides that we identified in these experiments were phosphorylated glycopeptides.

Figure 2. LC-MS/MS of tryptically digested HeLa 229 cells infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain. Top chromatogram shows the XIC of 204.087 (indicative of HexNAc) for the full run. Subsequent XICs show 10 and 1 minute windows respectively. Bottom spectra is a representative HCD glycopeptide spectrum.



The identification of LC-MS/MS glycopeptide data can be challenging. To minimize false positives, we performed both proteomics and glycoproteomics experiment (Figure 1). Total lysate samples were initially searched from the standard DDA proteomics LC-MS/MS run to identify the proteins present in the sample using conventional database search approach. The initial protein identifications allow for creation of a custom database that we can use to search in a more targeted glycoproteomics identification search. We use unmodified peptides from these survey runs to create a curated database of proteins that can also be glycoproteins will be present. The smaller curated protein database is used to search the intact glycopeptide data. This strategy not only minimizes the false positives but also speeds up the search time.

Figure 3. LC-MS/MS of tryptically digested proteomics and glycoproteomics HeLa 229 cells infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain. Top chromatogram shows the XIC of 204.087 for the proteomics run and the bottom for the glycoproteomics run. 10 and 1 minute windows shown respectively.

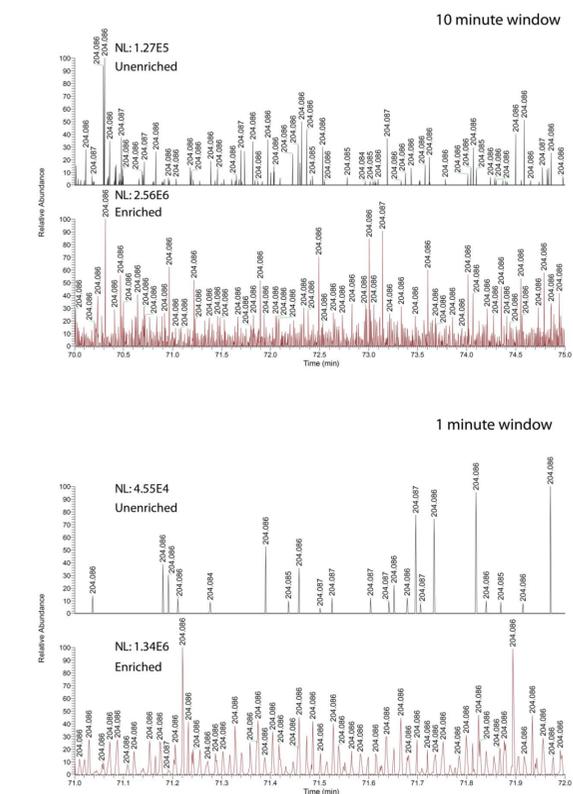
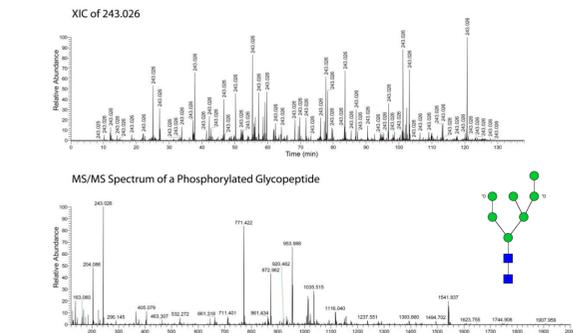
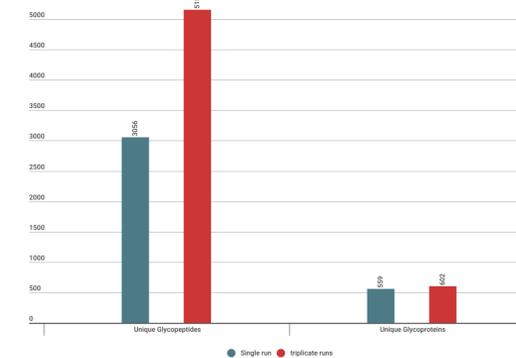


Figure 4. LC-MS/MS of tryptically digested glycoproteomics HeLa 229 cells infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain. Top chromatogram shows the XIC of 243.026, which is indicative of Man-6-P.



Overall we identified over 3000 unique intact glycopeptides in a single LC-MS/MS run and over 5000 unique intact glycoproteins in three runs, translating into over 600 unique glycoproteins. To our knowledge this is by far the largest number of intact glycopeptides reported in a single experiment without fractionation. In order to increase identification we have fractionated the sample and are currently doing experiments to further our insight.

Figure 5. Unique glycopeptides and glycoproteins identified in single and triplicate runs.



For quantitative glycoproteomics analysis using Proteome Discoverer, there were differences in glycoprotein and/or glycopeptide glycoform abundance levels between HeLa 229 cells, HeLa 229 cells infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain and RST17 CPAF-deficient strain. Those cells infected with *Chlamydia trachomatis* L2 RST5 CPAF-sufficient strain showed 23 unique glycopeptides/glycoforms with levels greater than two-fold change and cells infected with RST17 CPAF-deficient strain showed 81 unique glycopeptides/glycoforms with levels greater than two-fold change.

## CONCLUSIONS

- This study represents the first qualitative and quantitative glycoproteomics analysis to examine the role of chlamydial protease-like activity factor targets
- Over 5000 unique glycopeptides and 600 glycoproteins identified representing the largest number reported in a single study. Number of unique glycopeptides/glycoforms were observed to show change in expression upon infection
- Further functional and cellular pathways analysis with additional glycoproteomics experiments are planned for the future

## REFERENCES

- Patton MJ, McCorrister S, Grant C, Westmacott G, Fariss R, Hu P, Zhao K, Blake M, Whitmire B, Yang C, Caldwell HD, McClarty G. MBio. 2016 Sep 27;7(5). pii: e01427-16
- Totten SM, Feasley CL, Bermudez A, Pitteri SJ. J Proteome Res 2017 Mar 3; 16 (3)1249-1260.

## TRADEMARKS/LICENSING

© 2017 Thermo Fisher Scientific Inc. All rights reserved. Tandem Mass Tag and TMT are trademarks of Proteome Sciences plc. Byonic is a trademark of Protein Metrics. Swiss-Prot is a registered trademark of Institut Suisse de Bioinformatique. All other trademarks are the property of Thermo Fisher Scientific and its subsidiaries. Information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.