Development of Mass Spectrometry-Based Methods for Quality Assessment of Recombinant Proteins

Rosa Viner[1], Nan Liu[2], Seema Sharma[1], Sergei Snovida[3], Krishna Vattern[3], David Horn[1]
[1]Thermo Fisher Scientific, San Jose, CA
[2]Thermo Fisher Scientific, South San Francisco, CA
[3]Thermo Fisher Scientific, Rockford, IL

Poster Note 64761

ABSTRACT

Purpose: Development of protein product quality control method using mass spectrometric intact/top-down and bottom-up analysis.

Methods: For bottom up experiments, proteins were analyzed by nanoLC/MS. Intact and fragment samples were analyzed by standard flow LC- MS using a Thermo Scientific™ MabPac ™RP (2.1x50mm, 4 µm) column and a Thermo Scientific™ Orbitrap Fusion Lumos™ mass spectrometer. For top-down MS/MS experiments, ETD, EThcD, CID or HCD MS^2 fragmentation were used at a resolution of 120K@ *m/z* 200.
Results: We identified and quantified different forms of studied proteins by both bottom-up and top down approaches. The quantitative results produced by intact protein analysis strategy were more consistent. In general, the intact/top-down method outperformed the bottom up approach in these studies in terms of both efficiency and accuracy. Overall, a hybrid approach that combines the advantages of both bottom up and intact/top-down methods would be the most optimal for this type of experiment.

INTRODUCTION

Mass spectrometry has gained widespread acceptance for characterization of proteins during quality control and the regulatory approval processes. Improvements in mass spectrometer sensitivity, resolution and mass accuracy have enabled more detailed and confident analysis of larger biomolecules for confirmation of amino acid sequence, assessment of sequence variants and characterization of post-translational modifications. Traditionally, bottom-up mass spectrometry is typically used for such applications[1]. In this study, we demonstrate the suitability of a combined approach of intact MS and top down MS/MS analysis for comprehensive analysis of small footprint (SF) ligase performance in ProteinSEQ™ kits and characterization of expressed biotherapeutics. Our approach allows for rapid and accurate QC of protein products.

MATERIALS AND METHODS

Sample Preparation

SF ligase from ProteinSEQ CHO HCP quantitation kit , Granulocyte-colony stimulating factor (GCSF) expressed in *E.coli* or produced via 1 step in vitro translation kit (IVT) were acquired from Thermo Fisher Scientific. For bottom up experiments, following reduction and alkylation, proteins were digested using MS-sequencing grade LysC, chymotrypsin and trypsin for 4 hours at 37C. For intact/top down analysis, all proteins were desalted using an Amicon® centrifugal filter unit (10 kDa, EMD Millipore) prior to LC/MS experiments.

Liquid Chromatography

A Thermo Scientific™ EASY-nLC™ 1000 UPLC system and Thermo Scientific™ EASY-Spray™ source with 50 cm EASY-Spray column were used to separate peptides with a 25% acetonitrile gradient over 90 min at a flow rate of 300 nl/min for total run time of 120 min.

Intact proteins were analyzed on Thermo Scientific™ Dionex™ UltiMate™ 3000 RSLC system using a MabPac RP (2.1x50mm,4µm, heated at 80˚C) column.

For ligase intact mass analysis separation, gradient elution was performed from 10–25% over 1 min, from 25–40% over 15 min and from 40-95% over 1 min with ACN in 0.1% formic acid at a flow rate of 300 uL/min for a total run time of 20 min.

For top-down analysis, proteins were separated using an UltiMate 3000 RSLCnano system in microflow mode and a ProSwift™ RP-4H monolithic capillary column (200 um x 25 cm) with a 1 ul/injection. Gradient elution was performed from 10–25% over 5 min, from 25–40% over 5 min and from 40-70% over 3 min with ACN in 0.1% formic acid at a flow rate of 10 uL/min.

Mass Spectrometry

Samples were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer in intact protein mode using 2 mTorr ion-routing multipole (IRM) pressure. MS/MS spectra were acquired using Orbitrap CID and EThcD MS^2 fragmentation modes with Top 3-5 DDA methods. OT MS^1 data was acquired at resolution settings of 15–120K at *m/z* 200 and $OTMS^2$ at a resolution of 120K at *m/z* 200. Precursor ion isolation was performed with the mass selecting quadrupole and the isolation window was set to 3 m/z. The AGC target value was set to 5e5 for both MS^1 and MS^2; maximum injection times of 100 ms x 5 µscans for MS^1 and 200-250 ms x 5 µscans for MS^2 were used. For bottom-up analysis, MS data were acquired on an Orbitrap Fusion Tribrid mass spectrometer using MS^2 data-dependent acquisition mode. Maximum total cycle time was limited to 3 seconds. The most intense precursors selected from the FT MS^1 full scan (resolution 120,000 FWHM @ *m/z* 200) were quadrupole-isolated and fragmented by CID or HCD and detected in the dual-pressure linear ion trap (IT) or Orbitrap detector. The AGC target value was set to 4e5 for MS^1 and 5e4 (OT) or 1e4 (IT) for MS^2; maximum injection times of 100 ms for MS^1 and 35 or 60 ms for MS^2 were used.

Data Analysis

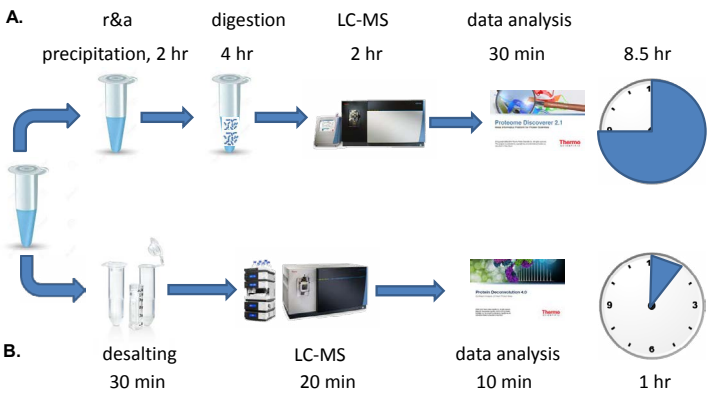
Intact protein spectra were deconvoluted with ReSpect™ (for 15K resolution) or Xtract (for 120K resolution) using the sliding window deconvolution algorithm in Thermo Fisher Scientific™ Protein Deconvolution™ 4.0 software. The top down data were analyzed with Thermo Fisher Scientific™ ProSightPC™ 4.0 and Thermo Fisher Scientific™ Proteome Discoverer™ 2.1 (utilizing the ProSightPD™ node) software packages. All searches were performed against custom databases. Final results were filtered using an E-value cutoff of 1 $x10^{-5}$ and search engine rank 1.

The bottom up data were analyzed using Thermo Scientific™ Proteome Discoverer™ 2.1 software using the SEQUEST® HT or Byonic™ search engines. Data were searched against a UniProt Ecoli or custom databases with a 1% FDR threshold.

RESULTS

Figure 1 shows the overall workflow for the protein product quality control method using MS intact/top-down and bottom-up (peptide mapping) analysis that was evaluated in this study.

Figure 1. Mass spectrometry-based workflows for quality assessment of recombinant proteins: A- bottom up; B- intact/top down.



Thermo SCIENTIFIC

LC-MS QC analysis of SF ligase

Small footprint ligase (Figure 2B) is a member of DNA ligase family - a specific type of enzyme that facilitates the joining of DNA strands together by catalyzing the formation of a phosphodiester bond[2] (Figure 2C). These ligases are widely used in protein sequencing kits. Five lots of SF ligase have been tested during the product development of ProteinSEQ CHO HCP assay kit (Figure 2A). All other lots of SF ligase except for one passed the product requirements for intra-assay precision (Figure 2D). Interestingly, all other QC manufacturing tests were positive for all lots. Standard QC methods did not reveal the reason for low performance of the outlying SF ligase lot. To identify the root cause using MS tools, we first performed a standard bottom-up/peptide mapping analysis (Figure 3). Multiple differences in quantities of different forms of ligase and numbers of host cell proteins were observed (Table 1). Complete sequence coverage was obtained for SF ligase from each lot and Lys 27 was identified as main site of adenylation (Figure 3C). Adenylation (addition of AMP) of a lysine(K_{27}) residue in the active center of the enzyme is required for ligase activity and the enzyme provided in the kit is pre-charged with ATP. We identified both forms of ligase with and without adenylate using the bottom-up approach, but the quantitative results produced by this strategy were inconsistent due to multiple peptide forms that span the modification site between the two intact proteoforms (Figure 3B, Table 1). The main reason for presence of multiple forms of *N* or *C*- ragged peptides containing Lys27 is that the modification interferes with the activity of the chosen proteolytic enzyme (LysC). Similar findings were reported previously for quantification of other PTM-containing peptides[3]. To quantify both forms using bottom up approach, we either combined all peptides containing the Lys 27 site or used only the 6-35 peptide (Table 1). Significant changes in the ratio between the pre-charged/non charged forms were observed for those two approaches as shown in Table 1 with the single peptide strategy being more accurate. Overall, based solely on the bottom up analysis results, we were unable to pinpoint reason for poor performance of lot 5 vs other lots. On the other hand, LC-MS analysis of intact ligase from different lots clearly identified different amounts of adenylated vs unmodified enzyme as the root cause for compromised performance of lot 5 (Table 1, Figures 4-5). The ratio of adenylated to unmodified enzyme was ~2:1 in all working lots and 0.7:1 in the low precision lot.

Figure 2. ProteinSEQ CHO HCP assay kit (A): SF Ligase (B); reaction mechanism (C); high variation of precision of lot 5 (D)

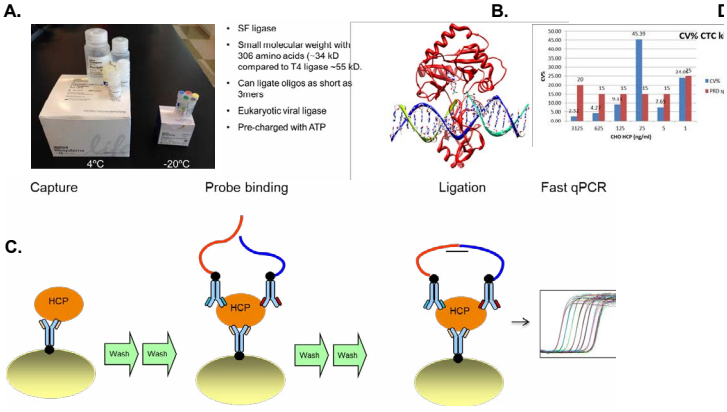
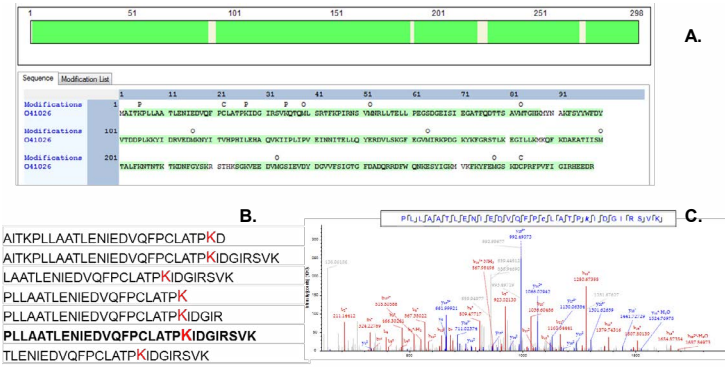


Figure 3. Bottom up analysis of SF ligase: A. Sequence coverage; B. Identified Lys27 peptides; C. Lys 27 adenylation site identification

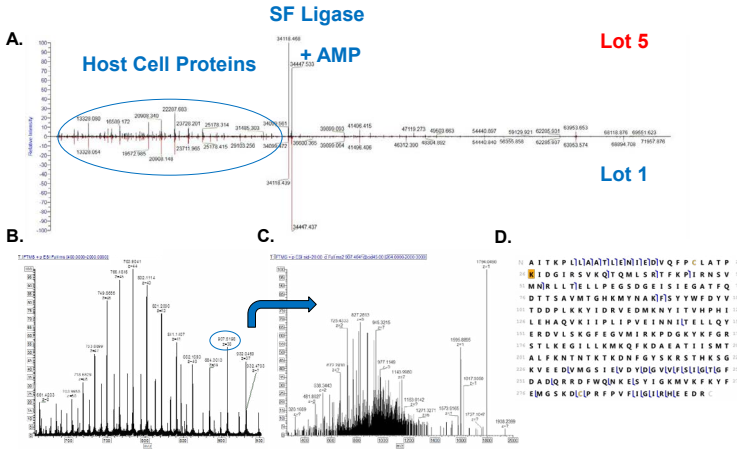


LC/MS intact analysis of ligase samples was performed using low (15K at *m/z* 200) or high (120K at *m/z* 200) resolution and produced identical quantification results (Figures 4A,B). Host cell protein profiles were similar between lots (Figure 4 A) and confirmed that indeed, diminished performance of lot 5 was due to different ratios of ligase forms. OT CID fragmentation in DDA mode was done to confirm protein identity (Figure 4C,D). As only one type of fragmentation (CID) and a top 5 DDA method was employed, the exact site of the PTM between Lys_{27} and Lys_{35} cannot be pinpointed, but the modified form of SF ligase was identified with high confidence.

Table 1. LC/MS QC analysis of SF ligase from 5 lots of ProteinSEQ CHO HCP assay kit. Lot 5* is the lot with compromised performance.

	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5*
# *E.coli* proteins	65	58	34	35	44
Peptide 6-35: AMP charged form /native form	3.11	2.56	0.75	1.76	1.12
All peptides with K27: AMP charged form /native form	1.74	1.2	0.44	0.5	1.17
Intact mass: AMP charged form /native form	2.17	2.22	1.96	2.11	0.69

Figure 4. Intact/top down analysis of SF ligase: A. Lot 1 vs Lot 5 intact mass measurements at 15K @ m/z 200 using the sliding window deconvolution method with ReSpect; B. High resolution average spectrum of Lot 1 ligase; C. OT CID spectrum of *m/z* 907(+38) precursor; D. Ligase adenylated form identification using ProSightPD 1.1 node in Proteome Discoverer 2.1



Overall, the intact/top-down QC protein analysis was simple and more straightforward than the bottom-up approach, which demonstrated strong dependence on proteolytic enzyme activity, PTM site location and required extensive data analysis using multiple software. Additionally, the intact/top-down method outperformed bottom up approach for such studies in terms of both efficiency (it only requires ~1 hr vs more than 8hr (Figure 1)) and quantification accuracy (Table 1).

Sequence Confirmation analysis of a recombinant Granulocyte-Colony Stimulating Factor (GCSF)

rGCSF[4], an 18.8 kDa protein used either as protein therapeutic drug or in immunoassays, was expressed in *E.coli* and MS analysis was performed as part of product validation. Complete protein sequence coverage was obtained from combined chymotrypsin and trypsin peptide mapping experiments, including identification of *N* and *C*-terminal peptides (Figure 5A). Surprisingly, we observed a significantly reduced abundance of the expected *N*-terminal peptide compared to the other peptides (Figure 5B). LC/MS intact mass measurement results confirmed this observation as the main detected proteoform was 19035.9 Da or 380 Da heaver than theoretical mass of 18655.64 with 2 S-S bonds (Figure 6A,B). Targeted LC/MS top down analysis of main proteoform using CID, ETD, EThcD fragmentations in combination with ProSightPC data processing allowed us to pinpoint the *N*-terminus as the site of modification (Figure 6 C,D). Additionally, we identified deamidation of Q_{20} and localized both disulfide bonds. Based on these results, the modification at the *N*-terminus has a mass of 379.2 Da and can be either a remnant of the expression tag or a chemical modification. Multiple additional proteoforms, including the unmodified primary structure of the N-terminus were identified (Figure 6E).

Figure 5. Bottom up analysis of rGCSF: A. Combined chemotrypsin and trypsin digests sequence coverage; B. Identification and TopN quantification of rGCSF chemotrypsin peptides. *N*-terminal unmodified peptide is highlighted.

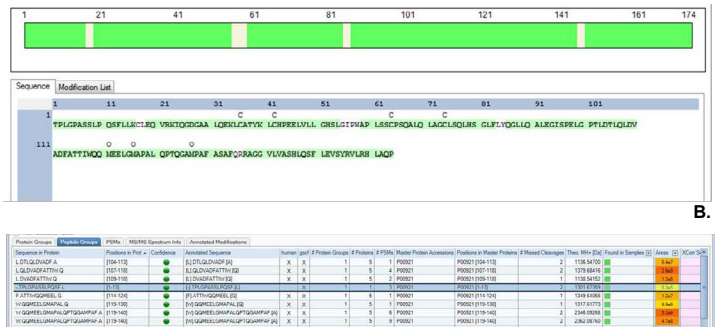
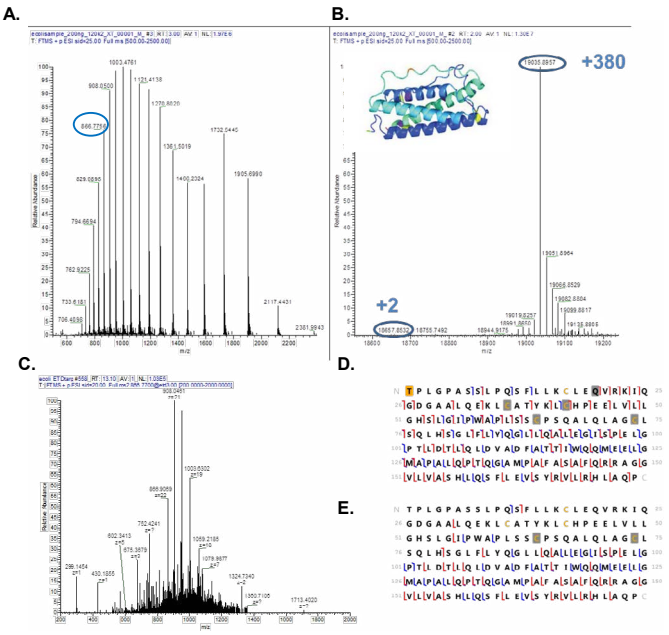


Figure 6. Intact/top-down analysis of rGCSF: Intact mass measurement at 120K @ *m/z* 200 (A) Xtract deconvolution spectrum with the crystal structure of GCSF inset (B); OT ETD spectrum of *m/z* 866 (+22) precursor (C); Combined ETD&CID fragment ion maps for proteoforms 19035.9 (D) and 18657.9 (E).



CONCLUSIONS

- The root cause of SF ligase poor performance in ProteinSEQ CHO HCP assay kit was due to the lower abundance of the ATP pre-charged protein as was clearly demonstrated only by intact protein mass spec analysis.
- Peptide level quantification of ligase proteoforms demonstrated strong dependence on PTM site location and digestion conditions and thus was less consistent compared to the intact/top-down approach.
- Overall, the intact/top-down method outperformed bottom-up approach for QC analysis of recombinant proteins in terms of both efficiency and accuracy.
- A hybrid approach that combines the positive aspects of both bottom up (PTM site identification) and intact/top-down (quantification) methods would be the most optimal for such type of experiments.

REFERENCES

1. Zhang, Z., Pan, H., Chen, X. *Mass Spectrom. Rev,* **2009**, 28, 147-176.
2. Shuman, S. J Biol.Chemistry, **2009**, 284, 17365-17369.
3. Leymarie, N, et *al*, *Mol Cell Proteomics*, **2013** Oct; 12(10): 2935–2951
4. Souza, L.M et *al*, *Science*, **1986**, 232, 61.

www.thermofisher.com

Thermo Fisher Scientific, Sunnyvale, CA USA is ISO 9001Certified.

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Brazil +55 11 2730 3006
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 6 6885 1213
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
Taiwan +886 2 8751 6655
UK/Ireland +44 1442 233555
USA +1 800 532 4752

PN64798-EN 0616S

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand