Combination of Bottom-Up and Top-Down Characterization of Biologics Using a High Throughput Capable Workflow in Proteome Discoverer Software

Kai Scheffler,[1] Torsten Ueckert,[2] Carmen Paschke,[2] Bernard Delanghe[2]
Thermo Fisher Scientific, [1]Dreieich, [2]Bremen, Germany

Poster Note 64482

Overview

Purpose: Demonstrate the benefits of Thermo Scientific™ Proteome Discoverer™ 2.0 software for combining a top-down and bottom up protein analysis workflow for in-depth protein characterization.

Methods: A Sequest HT database search with precursor area calculation was combined with a ProSightPD and Byonic search.

Results: A 100% sequence coverage was obtained by combining bottom-up and top-down experiments.

Introduction

In the development as well as the production phase of biologics such as monoclonal antibodies the verification of the correct amino acid sequence as well as the assessment of the presence of the correct type and relative amount of glycosylation are essential steps and only a few out of many analytical methods applied. Most commonly, enzymatic digests of the proteins are analyzed in a peptide mapping type experiment. In addition, top-down experiments provide additional information regarding mass, amino acid and further modifications. Whereas previously data analysis had to be run through different software packages, Proteome Discoverer software now provides new nodes that enable the analysis of both bottom-up and top-down experiments in one workflow on one platform.

Methods

Sample Preparation

For bottom-up experiments, the monoclonal antibody Rituximab was denatured for 30 min in 7 M Urea and 50 mM Tris HCL at pH 8.00. The sample was reduced with 5 mM DTT for 30 min at 37°C, and alkylated by addition of 10 mM IAA for 30 min at room temperature. The reaction was quenched by addition of 10 mM DTT. Thermo Scientific™ Pierce™ Trypsin Protease (MS Grade) was added and digestion allowed to proceed over night at 37°C. Digests were stopped by addition of TFA to approximately pH 3.00.
The sample for the top-down characterization of the reduced antibody was produced by offline buffer exchange against 50mM ammonium acetate followed by incubation in 10 mM TCEP for 1 h at 60°C directly prior to analysis.

Liquid Chromatography

Thermo Scientific™ Vanquish™ UHPLC system with a 2.1 x 250 mm i.d. Thermo Scientific™ Acclaim™ 120 C18 column and gradients of water and ACN with 0.1% formic acid each were used to separate the peptide mixture. For the gradient 2-45% ACN were ramped for 30min at a flow rate of 0.4 mL/min.

Mass Spectrometry

Data acquisition was performed on the Thermo Scientific™ Q Exactive™ Plus and Thermo Scientific™ Q Exactive™ HF mass spectrometers, both of which were equipped with the protein mode feature. Bottom-up experiments were performed using a HESI source; top-down experiments were performed by direct infusion with a Thermo Scientific™ Nanospray Flex™ ion source with the static nanospray needle setup.

Data Analysis

Data analysis was performed using Proteome Discoverer 2.0 software with the Sequest HT search algorithm and ProsightPD [1-4] and Byonic [5] nodes enabled.

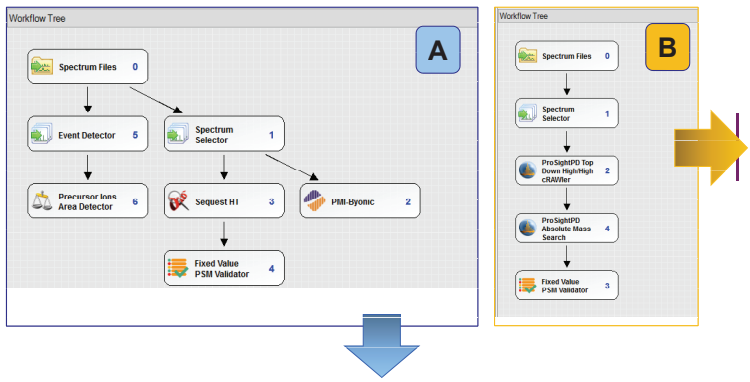
Results

In this study we have evaluated the new options in the Proteome Discoverer 2.0 software to include workflow nodes for Byonic and ProSightPC in a combined bottom-up and top-down analysis of a monoclonal antibody. The bottom-up workflow (Figure 1 A) consists of the combination of the major modules: the database search using Sequest HT and Byonic as well as the precursor ion quantification for the assessment of relative abundances of peptides.

The top-down workflow (Figure 1 B) includes the ProSight High/High Crawler that performs the spectra deconvolution based on high resolution mass spectra combined with the ProSight absolute mass search.

For the complete characterization of biologics it is essential to confirm the expected sequence aiming at 100% sequence coverage and in addition the monitoring of relative abundances of glycoforms as well as the levels of modifications such as oxidation, deamidation and truncation of the heavy chains' C-terminal lysine residue. Figure 2A shows the base peak chromatogram and the separation achieved from the antibody digest sample with a separation time of 20min. Figure 2B is showing the XIC of the oxonium ion indicative for the presence of glycopeptides that are shown in Figures 2C and 2D on the Full MS level highlighting the presence of the glycopeptides in its various isoforms. The obtained sequence coverage from Sequest HT and Byonic are shown in Figures 4 and 5 for the light and heavy chain respectively. Figure 7 lists the identified glycoforms of the heavy chains' glycopeptide in two forms, identified with and without a missed cleavage.

FIGURE 1. Proteome Discoverer 2.0 software workflows for the data analysis of bottom-up (A) and top-down (B) spectra. The bottom-up workflow consist of nodes for protein ID using the Sequest HT search algorithm, the precursor ion area detector for relative quantitation and the new Byonic node. The top-down workflow contains the new ProSight TopDown High/High Crawler for spectra deconvolution combined with the Prosight absolute mass search node.



Thermo SCIENTIFIC

FIGURE 2. Base Peak Chromatogram of the monoclonal antibody digest (top trace). The peptide mixture was separated over a 20min gradient. The XIC of m/z 204.0685 shows where the glycopeptides elute (middle trace). A zoom into the averaged Full MS spectra shows the variety and relative abundances of the various glycoforms and different charge states of the glycopeptide EEQYNSTYR.

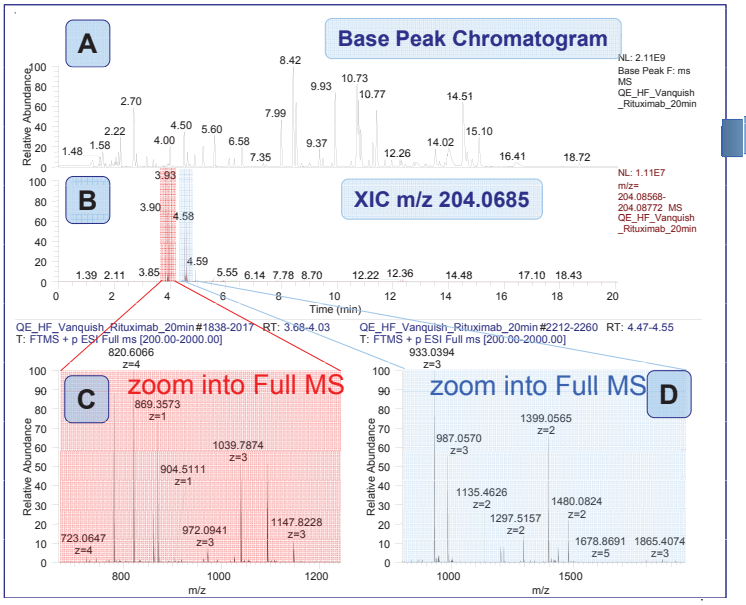


FIGURE 3. Top-down data of the reduced monoclonal antibody sample showing the Full MS spectrum (A), the MS/MS HCD spectrum of a single charge state of the antibodies' light chain (B). Trace C shows a zoom into the top-down spectrum highlighting a series of fragments of charge state z=6 and a number of low abundant higher charge states z=10,12. Trace D highlights the deconvoluted portion of the top-down spectrum shown in trace C highlighting the partial sequence aa81-83 DAA of the light chain.

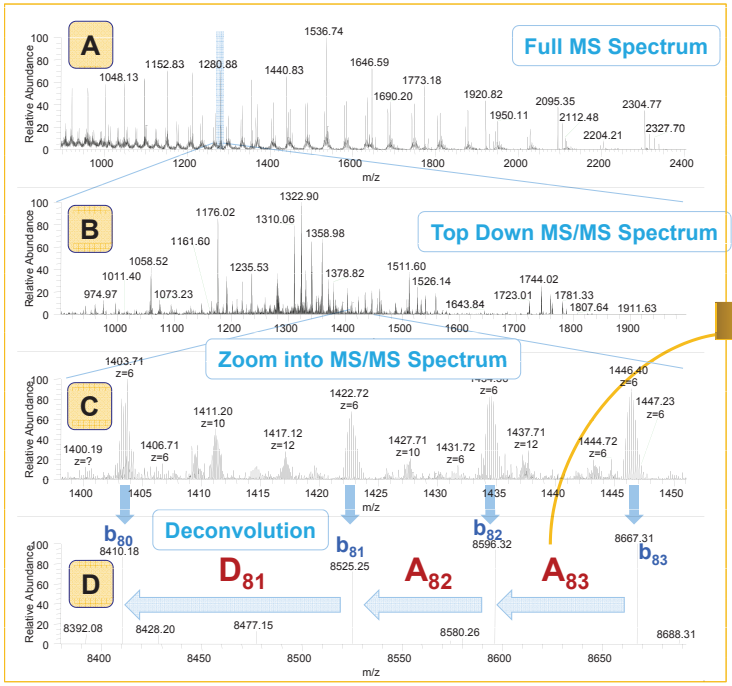


FIGURE 4: The sequence coverage obtained from the bottom-up approach for the antibodies light chain is 98.6%. Only the tripeptide EAK was missed due to the search algorithms' capabilities to identify peptides with at least 4 amino acids.

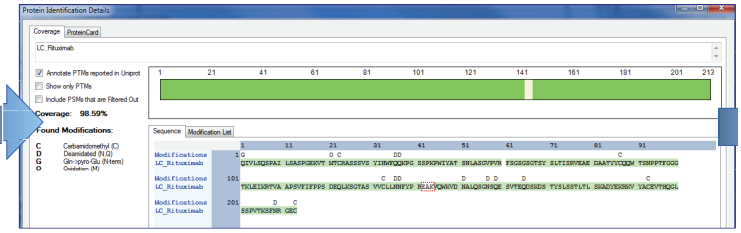


FIGURE 5. The sequence coverage obtained from the bottom-up approach for the antibodies heavy chain is 100%.

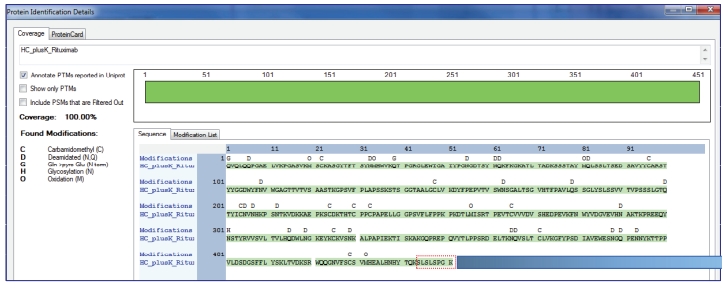


FIGURE 6: Sequence coverage of the antibodies' light chain based on the top down spectrum. Blue lines with kinks towards the left/right indicate the assignment of corresponding b/y-type fragment ions.

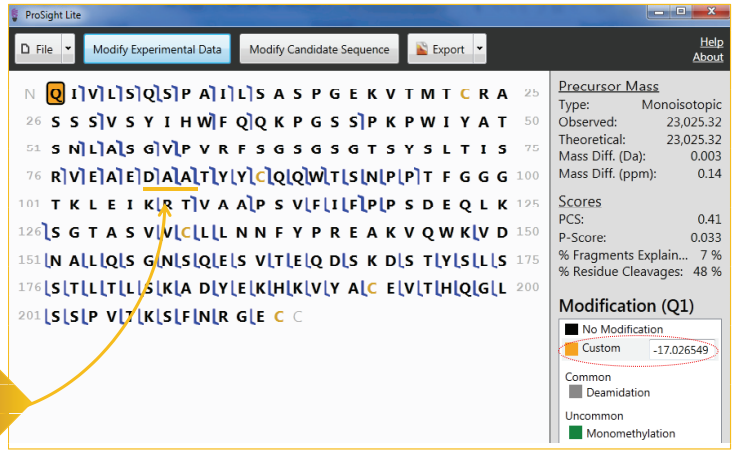


FIGURE 7: List of the identified glycoforms of the antibodies' glycopeptide EEQYNSTYR and TKPREEQYNSTYR using the Byonic search node.

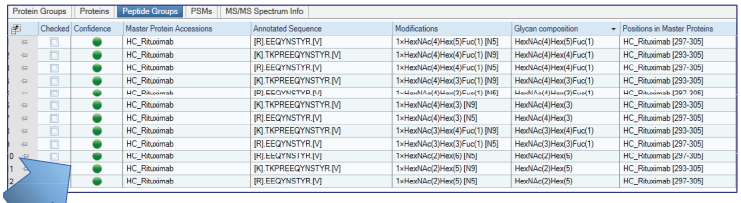
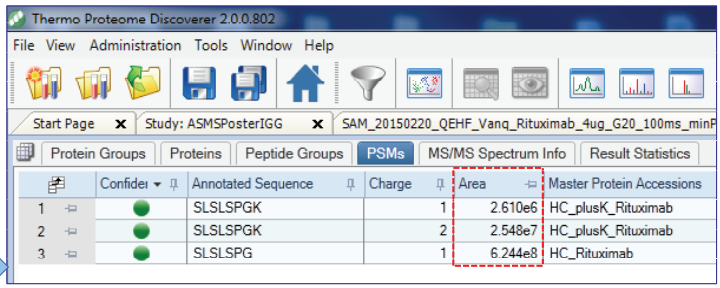


FIGURE 8: This table shows the list of identified modifications based on the included search criteria for the bottom-up database search:Gln->pyroGlu, deamidation and oxidation.

Description	Gln->pyro-Glu(Q) Count	Gln->pyro-Glu(Q) Positions
	Deamidated(N; Q) Count	Deamidated(N; Q) Positions
	Oxidation(M) Count	Oxidation(M) Positions
HC_plusK_Rituximab	2	Gln->pyro-Glu [N-Term; N-Term]
	22	Deamidated [Q6; N33; N55; N61; Q62; Q82; N163; Q179; Q200; N205; N207; N212; N290; Q315; N319; N329; N365; Q366; N388; Q390; N393]
	5	Oxidation [M20; M34; M81; M256; M432]
LC_Rituximab	1	Gln->pyro-Glu [N-Term]
	9	Deamidated [Q36; Q37; N136; N137; N151; N157; Q159; Q165; N209]
	1	Oxidation [M31]
HC_Rituximab	2	Gln->pyro-Glu [N-Term; N-Term]
	20	Deamidated [N33; N55; N61; Q62; Q82; N163; Q179; Q200; N205; N207; N212; N290; Q315; N319; N329; N365; Q366; N388; Q390; N393]
	5	Oxidation [M20; M34; M81; M256; M432]

FIGURE 9 The truncation of the C-terminal lysine residue of the heavy chain is quite commonly observed. Using the precursor ion quantification in the bottom-up workflow an assessment of relative quantitation of the peptide with and without the C-terminal lysine can be performed including different charge states. The ratio of the truncated to untruncated peptide based on the precursor are is 6.2e8:(2.6e6+2.5e7)=23 corresponding to ~4.3% of the untruncated form.



Results continued

Figure 8 lists the identified modifications found in the bottom up approach: oxidation, deamidation and the conversion of the N-termini of both the light and heavy chain from Gln to pyro-Glu. Figure 9 highlights the relative quantitation of the C-terminal peptide of the heavy chain with and without the terminal lysine. For the full length peptide both charge states +1 and +2 were considered, the truncated peptide was detected only in charge state 1 due to the lack of the proton-capturing lysine. The areas obtained for the two versions of the peptide allow the determination of the ratio truncated:untruncated peptide to be ~23, corresponding to ~4.3% abundance of the untruncated form.

Figure 3 shows the Full MS spectrum of the reduced antibody with the overlapping charge envelopes for the light and heavy chains. For top-down analysis and fragmentation a single charge state was isolated with a 10Da isolation window in order to get a fragment ion spectrum of either light chain (Figure 3B) or the heavy chain (not shown) was obtained. When top-down analyses are performed involving chromatographic separation the isolation window can be widened up to several hundreds of Da in order to co-isolate and fragment several charge states of either species in parallel. Figure 3C highlights the number and density of fragment ions observed, and in particular fragment ions in higher charges states due to the use of the protein mode available on the Q Exactive mass spectrometer. Figure 3D shows the deconvoluted spectrum of the zoomed MS/MS spectrum highlighting the fragment ion series b_{80}-b_{83} supporting the amino acid sequence "DAA". The assignment of all detected fragment ions to the sequence of the light chain is highlighted in Figure 6. The isoform with Gln->pyro-Glu (D mass -17.0264 Da) was selected so this modification was added to the sequence. Overall 48% of all residue cleavages were confirmed with the top-down approach This resulted in a combined sequence coverage of 100 %.

Conclusion

Here we demonstrate the advantages of the possibilities to implement nodes for Byoinc and ProSightPC into Proteome Discoverer software for a high-throughput capable analysis of biologics. In the example detailed on this poster results were obtained for:

- The sequence coverage for the light (100%) and heavy chain (100%).
- The identification of various glycoforms of the heavy chains' glycopeptide.
- Various modifications of light and heavy chain, in particular oxidation, deamidation and Gln->pyroGluc conversion.
- The assessment of relative ratio of the truncation of the terminal lysine residue on the heavy chain.
- The light chain top-down data confirmed the sequence and supported the bottom-up analysis with 48% obtained residue cleavages.

References

1. LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Res. 2004, 32: W340–W345.
2. Durbin KR, Tran JC, Zamdborg L, Sweet SM, Catherman AD, Lee JE, Li M, Kellie JF, Kelleher NL. Intact mass detection, interpretation, and visualization to automate Top-Down proteomics on a large scale. Proteomics. 2010 Oct;10(20):3589-97.
3. Leduc RD, Kelleher NL. Using ProSight PTM and related tools for targeted protein identification and characterization with high mass accuracy tandem MS data. Curr Protoc Bioinformatics. 2007 Sep;Chapter 13:Unit 13.6.
4. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, EarlyBP, Bluhm EJ, Babai S, Kelleher NL. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. Nucleic Acids Res. 2007 Jul;35 :W701-6.
5. Byonic: advanced peptide and protein identification software. Curr Protoc Bioinformatics. 2012 Dec; Chapter 13:Unit13.20. Bern M, Kil YJ, Becker C http://www.ncbi.nlm.nih.gov/pubmed/23255153

6. Acknowledgements

We would like to thank Martin Samonig for providing the LC-MS data file of the antibody digest.

www.thermoscientific.com

Thermo Fisher Scientific, San Jose, CA USA is ISO 13485 Certified.

ISO 13485

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
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 45 453 9100
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
Spain +34 914 845 965
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
UK +44 1442 233555
USA +1 800 532 4752

Thermo SCIENTIFIC
A Thermo Fisher Scientific Brand

PN64482-EN 0615S