Bruker Daltonics

Technical Note # TN-40

Top-down Characterization of Intact Proteins by Ultra-high-Resolution maXis-ETD

The identification of proteins is usually achieved by the well established bottom-up approach where individual or crude mixtures of proteins are enzymatically digested at the whole. MS and MS/MS data information of the resulting peptides is then used for probability searches in sequence databases to reveal the protein identity. This approach typically covers 20-60% of the amino acid sequence for the protein hits. If a full protein characterization is required, like for the analysis of biopharmaceuticals [1], protein homologues or highly conserved proteins with various functional post-translational modifications, alternative or additional methods like top-down sequencing are required. For isolated or a lower number of protein mixtures, top-down can as well be a much faster and direct way to get maximum information on the sequence. Top-down sequencing starts typically by determining the intact protein mass, then fragmentation of the intact proteins is performed, resulting in product ions that allows to pinpoint the protein sequence as well as potential modifications or mutation in the their localization and structure [2].
The maXis is able to determine molecular weights of large intact proteins down to the ppm level, which in many cases is sufficient to get a first confirmation of the correct primary structure including glycosylations patterns. With the introduction of ETD (Electron Transfer Dissociation), direct MS/MS on intact proteins is possible for readout and confirmation of protein sequences including modifications. The principle and typical data is presented in this technical note.

Introduction

Detailed protein characterization is increasingly used in proteomics where researchers focus on a certain biologically active subset of proteins instead of looking at entire cell lysates. In many cases, protein candidates or protein patterns have been found in biomarker discovery experiments already and have to be identified at a later stage of the project. Doing this on the intact protein level has a lot of advantages like picking the right compound from the mixture and looking at a full characterization including modifications on the isolated species. Another relevant application is the regulatory approval for biopharmaceuticals, such as monoclonal antibodies [3]. Here, the complete characterization and comparison with reference standards is required since small process differences may significantly alter the final biologic product which may adversely affect the safety and efficacy of the drug.
Mass spectrometry combined with liquid chromatography facilitates the rapid characterization of proteins. Measuring the intact mass of a protein quickly confirms if the protein

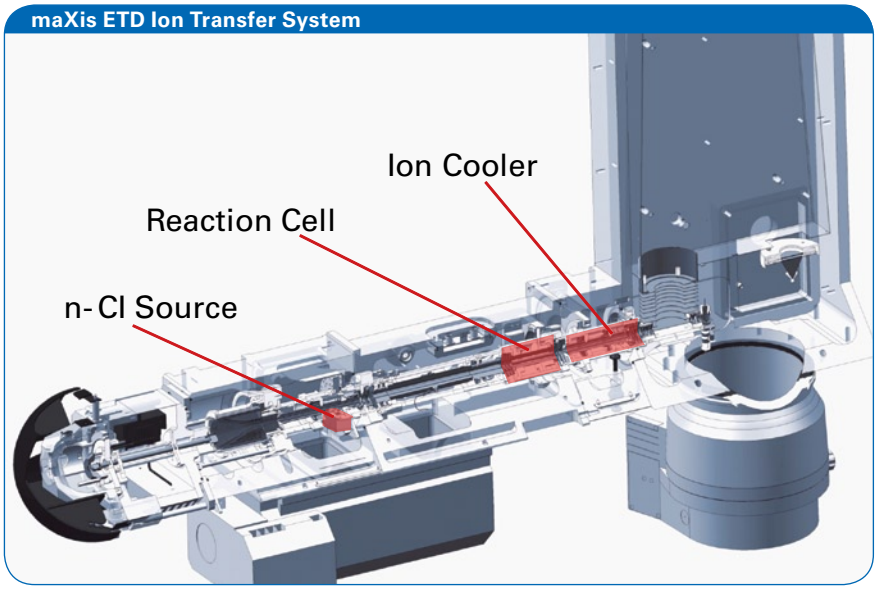


Figure 1: Innovative new technical developments on the maXis ETD for enhanced analytical performance.

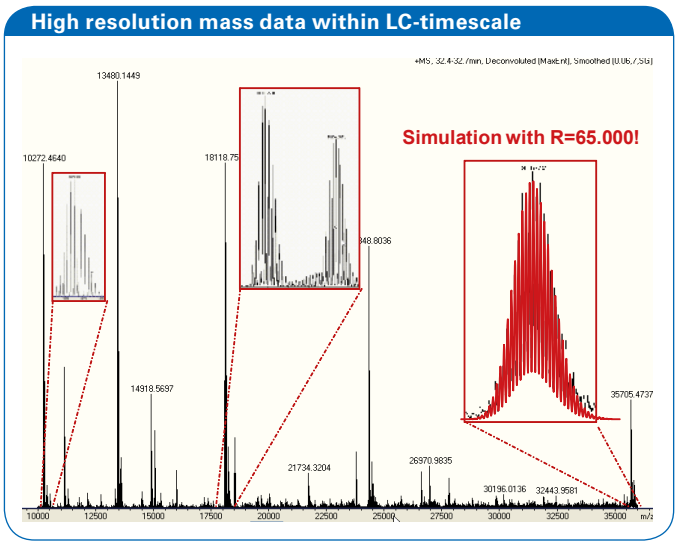


Figure 2: Maximum Entropy deconvoluted mass spectrum of an LC/MS run of intact *E- coli* proteins at the retention time of ca. 32 min. All co-eluting proteins in the mass range of 10 – 36 kDa are isotopically resolved. The experimental signal at 36 kDa (right inset) is in perfect agreement with the theoretical isotope pattern of this compound at a simulated resolution of 65,000.

drug has been correctly manufactured and purified and provides information about heterogeneity and any potential contaminants. Additional top-down sequencing by ETD (Electron Transfer Dissociation) can give further, quick insight into sequence details and modifications.
In ETD, the fragmentation of peptides or intact proteins is induced by converting the positively charged protonated ions into radical cations by an electron transfer from radical anions. The resulting analyte cation radicals are unstable and typically undergo rapid N-Cα bond cleavages leading to c- and z-type fragments of the proteins or peptide. This non-ergodic process occurs randomly along the amino acid backbone. Post-translational modifications are not affected but remain on their respective amino acid and can thus be observed in the ETD-MS/MS spectra by significant mass differences between adjacent fragment peaks. Moreover, due to the statistical, rapid N-Cα bond cleavage, also large intact proteins and antibodies can be dissociated and sequenced.
The Bruker maXis™ UHR-TOF LC-MS, provides ultra high resolution, exceptional mass accuracy (confident sub ppm) data at full sensitivity. The combination of high mass accuracy and resolution allows rapid detection of modifications within MS spectra, whilst the addition of ETD fragmentation enables direct insight into the protein sequence for further detailed characterization. This very unique combination of features makes the maXis the ideal instrument for any kind of top-down analysis in protein investigations.

Experimental

IgG was obtained from Chinese Hamster Ovary (CHO) cells. Proteins were separated with a Zorbax SBC8, Rapid Resolution Cartridge (2.1 x 30 mm, 3.5 µm) within 15 minutes and directly analyzed by the maXis. IgG was reduced and alkylated to release the heavy (HC) and light chain (LC).
LC-MS: Intact, heavy and light chain IgG proteins were separated with a Zorbax SBC8, Rapid Resolution Cartridge (2.1 x 30 mm, 3.5 µm) within 15 minutes and directly analyzed by the maXis™.
Bacterial lysis supernatant (*E. coli*) was separated with a Dionex Acclaim C4 150X1 mm column at a flow rate of 180µl/min @ 30 °C. Gradient 1: 0 to 50 % B at 17,5 min (A: 1% HCOOH in H_2O ; B: 0,8 %HCOOH in CH_3CN. Gradient 2 : 0 to 50 % B at 40 min (A: 1% HCOOH in H_2O ; B: 0,8 % HCOOH in CH_3CN. 1µl from a dilution by 10 was injected.

IgG1: assignment of the major glycosylation isoforms

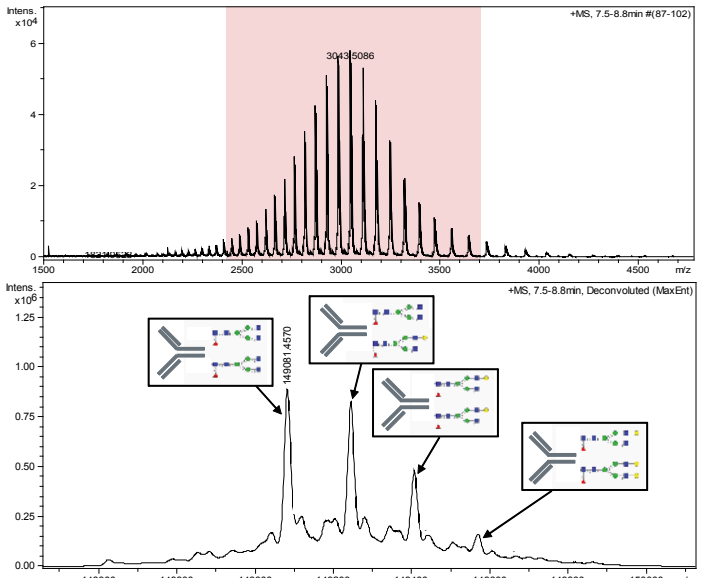


Figure 3: The mass spectrum before and after charge deconvolution applying the Maximum Entropy algorithm. The mass deviation between measured and theoretical mass is better than 2 ppm. The high resolving power of maXis rapidly identified heterogeneous glycosylation patterns; the major glycosylated species have been annotated.

IgG heavy chain measurement

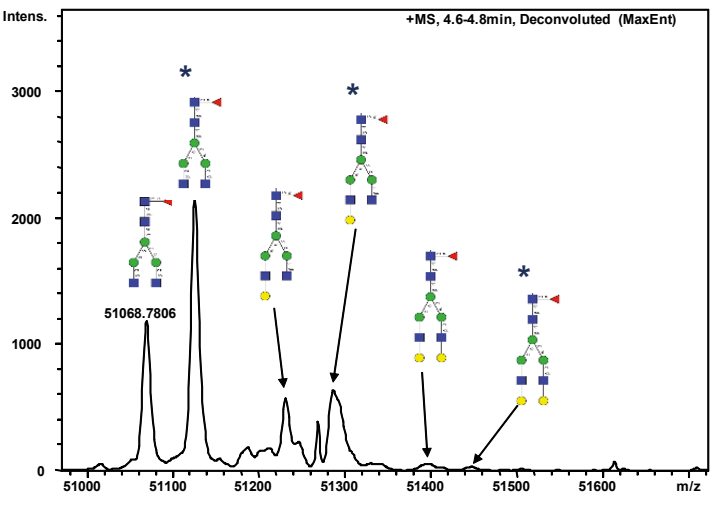


Figure 4: Maximum Entropy deconvoluted spectrum of the released heavy chain of the human IgG1. Taking into account average atomic weights from organic sources, the calculated mass of the mass labeled glycosylation form is 51,068.88 Da, which deviates less than 2ppm from the measured mass. Annotated peaks labeled with * (peak shifts of 57 Da) represent chemical artifacts due to overalkylation.

Ubiquitin and Serotransferrin were obtained from Sigma Aldrich. For ETD MS/MS all protein were prepared in 50:50 methanol/water with 0.1% formic acid and ionized by pneumatic electrospray at a rate of 3mL/min.

Fig. 1 shows the schematic of the setup of the ion transfer system. ETD on the maXis is performed by the addition of a negative chemical ionization source into the ion transfer system for generating reagent anions and injecting them onto the ion pathway. The arrangement of the nCI source and transfer octapole is similar to the setup of the amaZon™ ETD ion trap.

Protein precursor cations as well as reagent anions are discontinuously and consecutively accumulated in the reaction cell for performing the ETD reaction. Following the ETD experiment the product ions are extracted to the ion cooler. The ion cooler acts as an interface between the ETD reaction cell and the TOF analyzer. The ion cooler collects the ETD product ions and ejects them under conditions appropriate for the orthogonal acceleration into the TOF. The pressure in the cooling cell provides substantial collisional cooling to allow for the optimal energy ion beam exiting the ion cooler. The exit of the cooling cell is pulsed synchronously with the TOF analyzer at a kHz repetition rate. After the ions have been extracted into the ion cooler the ETD reaction cell is ready for the next ETD experiment. While the next ETD experiment takes place product ions are extracted from the cooling cell.

Results

Fig. 2 demonstrates the ability of the maXis instrument to generate high-resolution, high accuracy mass data of intact proteins even within an LC-timescale. The deconvoluted spectrum shows several proteins from E. coli which were online LC-separated. Even under those time-constraints, full isotopic resolution at a level of R = 65,000 is achieved up to the molecular weight of ca. 36 kDa – unbeaten within mass spectrometry, except FT-ICR MS. Fig. 3 demonstrates the rapid assignment of the major glycosylation isoforms of human IgG1. Maximum Entropy deconvolution of the data yields a mass for the major glycosylation form, which fits with the expected average mass range based on a calculation of atomic weights from organic sources by 2 ppm. Further glycosylation isoforms are assigned by the characteristic mass distances of 162 Da revealing additional galactose units.

The released heavy chain of the intact human IgG was measured in Fig. 4. The experimental mass was found to be in excellent agreement within 2 ppm to the expected one.

In addition, the high-definition data provided by the maXis also allows unambiguous assignment of chemical artefacts which were generated during sample preparation. In the example shown here, spectral peaks that did not match the masses of expected protein sequence/glycosylation structures (denoted with *, peaks shifted by 57Da) could be assigned to unwanted, over-alkylated by-products originating from the alkylation step.

Even large non-covalent protein complexes can be investigated with the maXis. In Fig. 5, the interaction of an IgG with a smaller protein at 25 kDa is observed. The maXis shows an excellent ion transmission in the m/z range between 1,000 and 7,000 m/z where the charge state distributions of the individual compounds and the complex are detected. In the deconvoluted spectrum, the intact non-covalent complex at a molecular weight of 172 kDa is observed, derived even from overlapping signals in the raw data spectrum originating from the IgG and the protein complex.

Analysis of a non-covalent complex

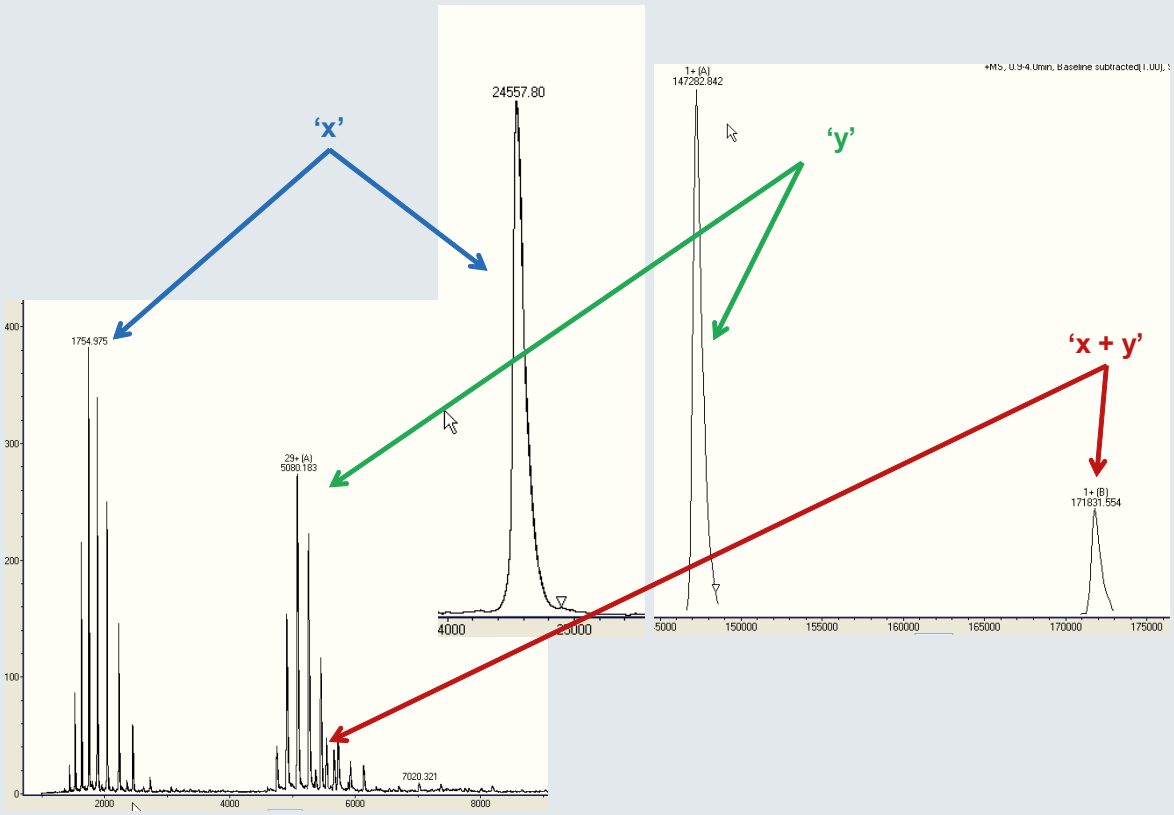


Figure 5: Example of a study on a non-covalent complex formed by a monoclonal antibody and a protein of 25 kDa molecular weight. Protein signals appear in the wide m/z range of 1,000 – 7,000 (left), requiring a mass spectrometer with a high transmission in that range. After deconvolution (right) of the raw data, signals of both individual components as well as the entire complex at 172 kDa are observed. Samples are courtesy of CIPF (Centre d'Immunologie Pierre Fabre).

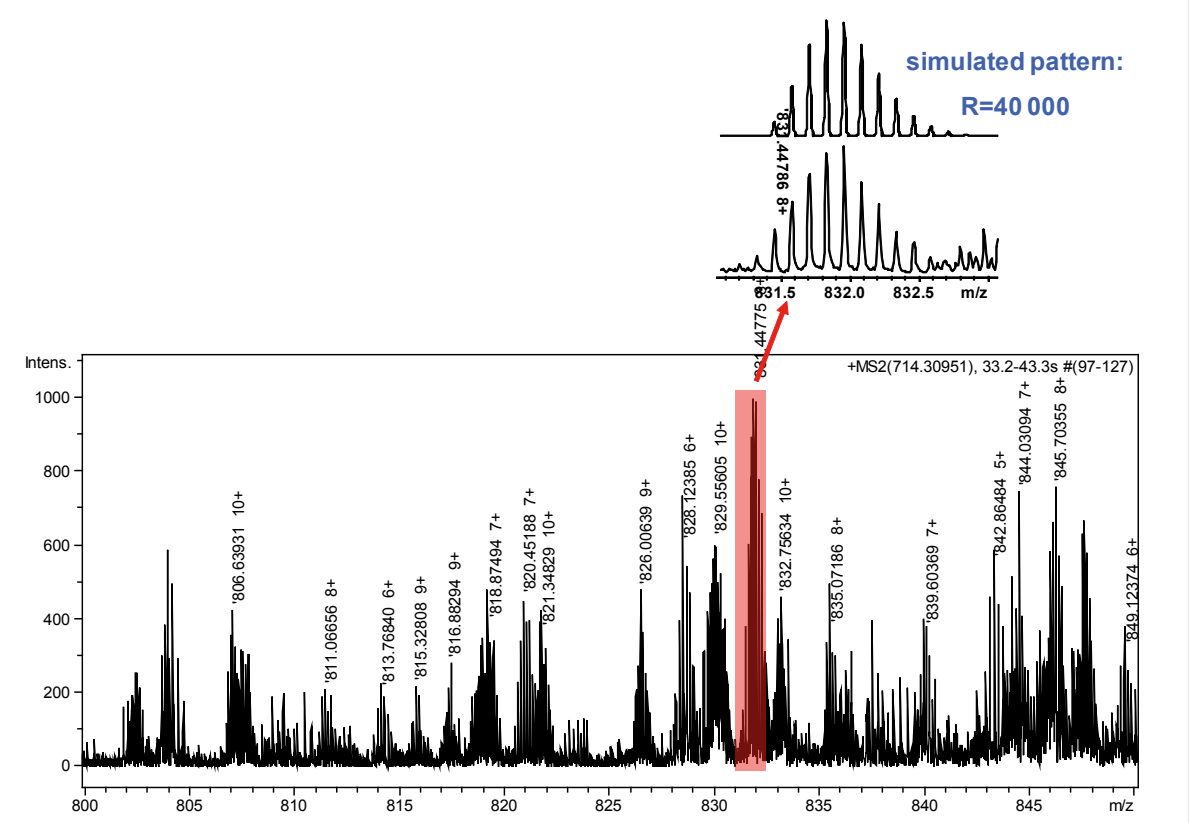
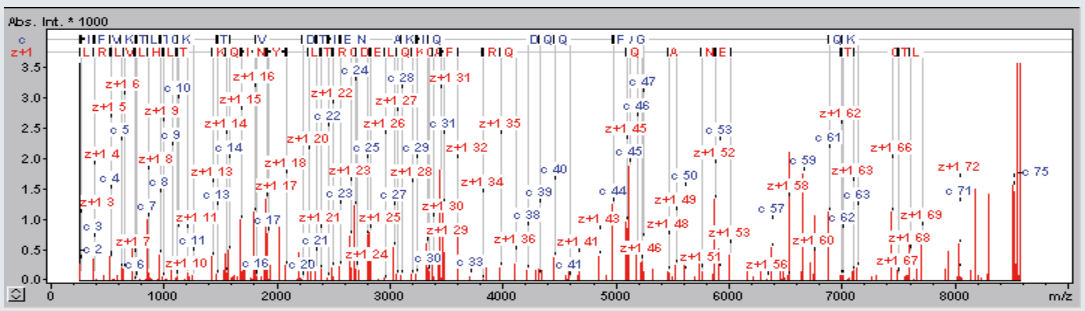
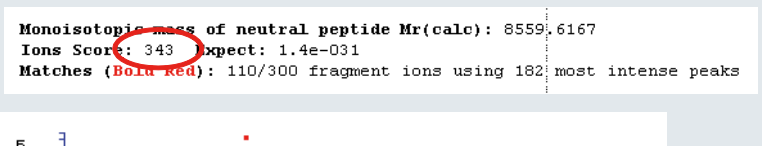


Figure 6a: Zoom of the entire ETD-MS/MS spectrum of intact Ubiquitin (MW 8565 Da). All observed fragment ions with charge states between +6 and +10 are clearly isotopically resolved. The comparison of the experimental signal at m/z 831 with the simulated isotope pattern of this fragment (top) shows that the instrument has a mass resolution of the fragments in the range of R = 40,000.





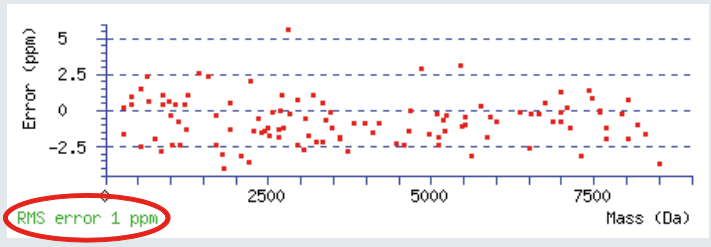


Figure 6b: Result of the Mascot database search on the ETD spectrum of intact Ubiquitin. The peak annotation within biotools (top) shows almost complete sequence coverage from the intact mass of 8565 Da down to ca. 200 Da. The Mascot score of 343 (middle) represents the highly confidential identification of Ubiquitin. The average mass error of all fragments is 1 ppm (bottom), i.e. the maXis shows similar mass accuracies for both MS and MS/MS spectra.

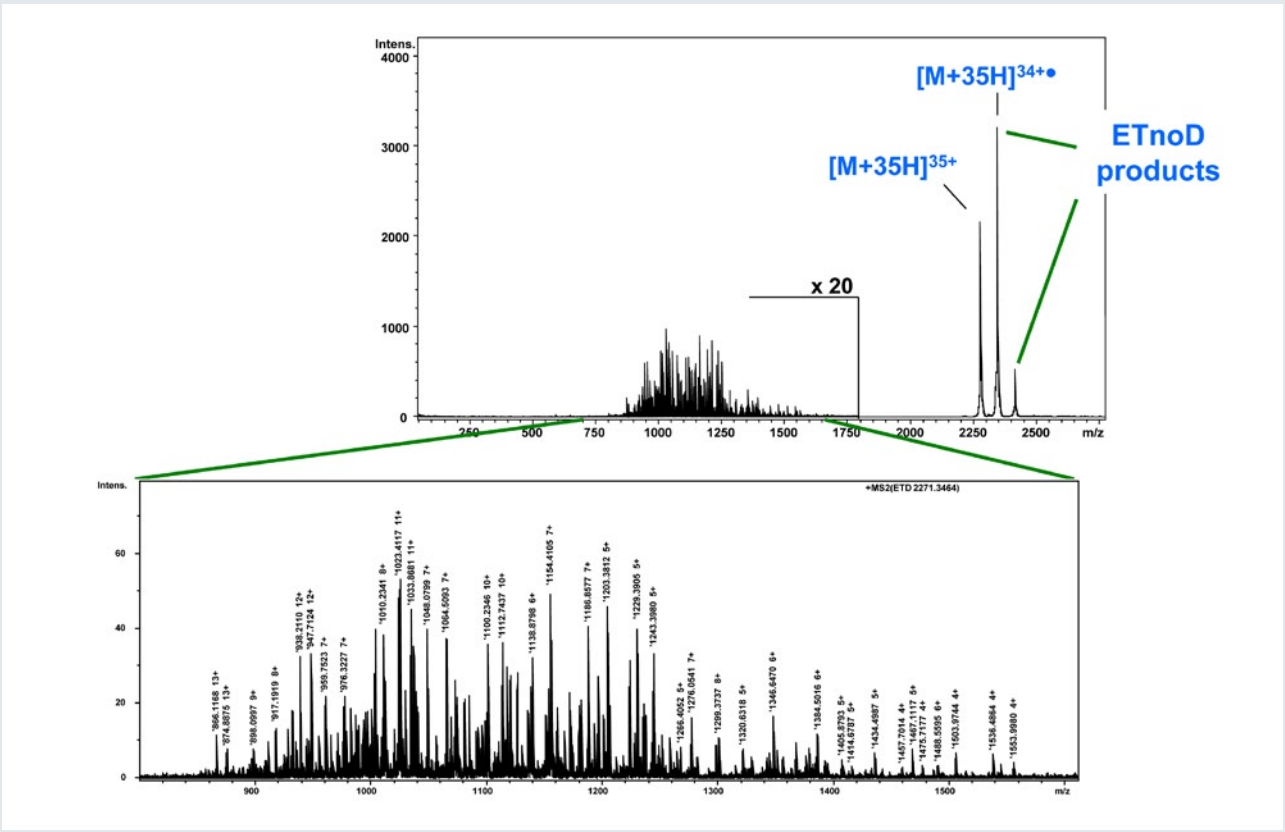


Figure 7a: Raw ETD-MS/MS spectrum of intact Serotransferrin (MW ca. 80 kDa). After isolation and fragmentation of the +35 charged ions at m/z 2271, multiple fragment ions appear in the m/z range 750 – 1750 with charge states of +4 to +13.

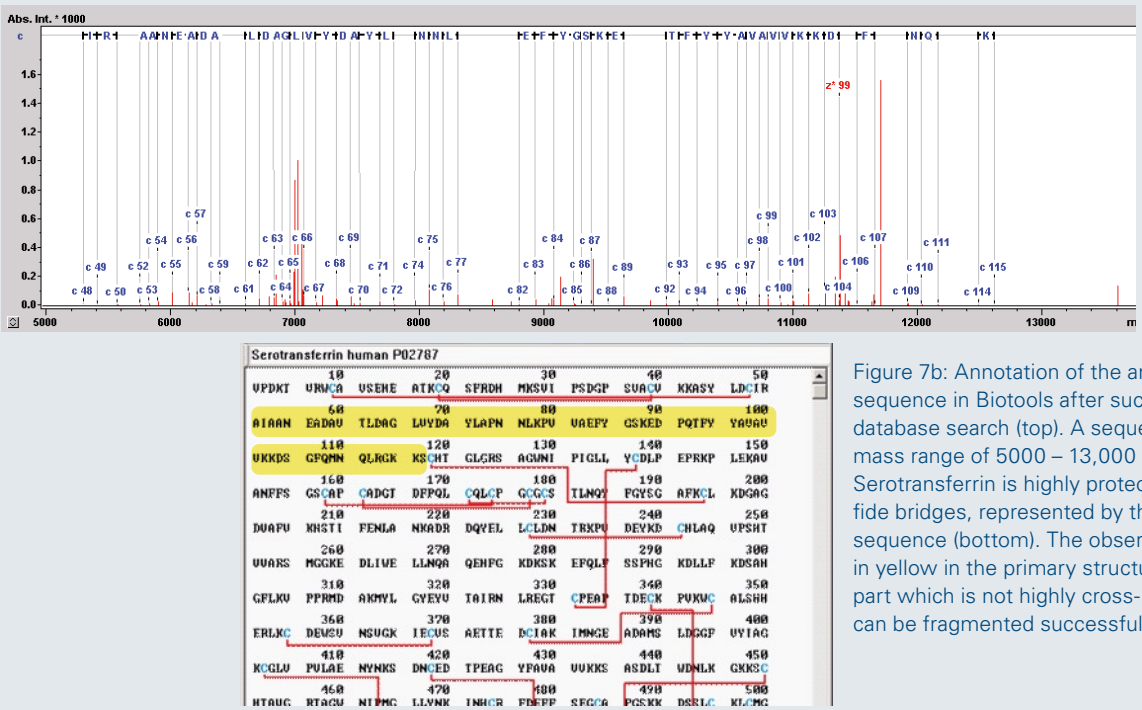


Figure 7b: Annotation of the amino acid sequence in Biotools after successful Mascot database search (top). A sequence tag in the mass range of 5000 – 13,000 Da is observed. Serotransferrin is highly protected by 19 disulfide bridges, represented by the red lines in the sequence (bottom). The observed tag is marked in yellow in the primary structure and is the only part which is not highly cross-linked: therefore it can be fragmented successfully.

Aside of the pure high mass accuracy molecular weight assignments, direct analysis of the primary structure of the investigated proteins is of major interest. Here, not only best ion transmission for large intact protein ions into the mass analyzer is important, but also the efficiency of the ETD process which should lead to maximum sequence coverage in the resulting MS/MS spectrum. Fig. 6a demonstrates the excellent performance of the maXis with regard to ETD of intact proteins. The +12 precursor ion of Ubiquitin (MW 8565 Da) at m/z 714 was isolated in the quadrupole and fragmented in the reaction cell of the maXis. Shown is part of the full ETD spectrum with various highly charged fragment ions between +6 and +10 charge states. All are isotopically resolved, as can be seen in the inset: the +9 charged ions at m/z 831 has a mass resolution of ca. 40,000 as compared with its simulated pattern which has been assigned to the c59 fragment of ubiquitin by the Mascot database search. In Fig. 6b, the result of the subsequent Mascot database search can be seen. Almost the entire amino acid sequence can be revealed (coverage ca. 90%), fragments are observed from the molecular weight of the intact protein at 8565 Da all the way down to the mass range of di- und tripeptides. Ubiquitin was identified with a very confident Mascot score of 343. Remarkably, the mass errors of all fragments, along the entire mass range up to 8.5 kDa, are within ± 2.5 ppm, the averaged (RMS) error is 1 ppm. The high mass accuracy of the maXis for both MS and MS/MS results, at a simultaneously high sensitivity, is truly unique and provides highly confident protein identification and sequence information. This combination of high mass resolution, high mass accuracy and excellent ETD efficiency in general makes the maXis the ideal instrument for top-down protein analysis.

In the next example (Fig. 7a), a truly large intact protein was analyzed by maXis-ETD, Serotransferrin, a glycoprotein from human liver with a molecular weight of ca. 80 kDa. Here, the high isolation m/z range of the maXis quadrupole allowed the isolation and fragmentation of the +35 precursor ion at m/z 2271. While a large portion of the product ions are possibly just radical, still cross-linked or charge-reduced species (m/z range 2200 – 2600), a considerable number of fragments occur in the m/z range 750 – 1750. As visible in the zoomed inset in Fig. 7a, charge states of those fully resolved fragments vary between +4 and +13. Fig. 7b shows the amino acid sequence annotation of the ETD MS/MS spectrum in biotools™: the sequence tag between the amino acid positions A51 and S117 is revealed. Serotransferrin is highly inter-connected by disulfide bridges which were not reduced prior to the analysis. The visible tag covers the sequence between C48 and C118: that is the only region within Serotransferrin which is not highly cross-linked by disulfide bridges. For a full characterization, the protein must be reduced and as well analyzed by bottom-up in addition, but this example demonstrates nevertheless the high and unique capabilities of maXis-ETD for the direct top-down analysis of proteins in the range 50 – 100 kDa.

Conclusions

Top-down protein analysis allows for the identification and characterization of intact proteins without the need for prior enzymatic digestion. Due to the analysis starting from the entire protein sequence, the top-down approach is ideally suited for the comprehensive characterization of proteins intended to the used as biotherapeutics or the identification of biomarkers discovered by other mass spectrometry techniques like MALDI Imaging or quantitative proteomic studies.

ETD coupled to the maXis UHR-TOF LC-MS provides unbeatable, unique technical performance. The combination of high mass resolution, mass accuracy and high ETD efficiency makes it the ideal tool for top-down protein investigations on intact proteins like biotherapeutic drugs, isolated proteins from biomarker studies or even protein mixtures in an LC-MS/MS setup. Examples shown here cover a mass range of 10 kDa up to 80 kDa for MS/MS analysis. Fragments masses as well as molecular weights of the intact proteins up to 172 kDa, including glycosylation patterns, can be determined within 2 ppm, thus providing data with extremely high confidence for sequence confirmation.

References

[1] Bruker Daltonics Application Note ET-18:
Rapid Quality Control of Biopharmaceutical products

[2] Bruker Technical Note TN-37:
Towards 100% sequence coverage in protein QC: In-depth characterization of monoclonal antibodies using the ProteinS cape database software

[3] Bruker Daltonics Application Note ET-17/MT-99:
Characterization of the N-glycosylation Pattern of Antibodies by ESI - and MALDI mass spectrometry

Authors

Ralf Hartmer, Carsten Stoermer, Laura Main, Dirk Wunderlich, Arnd Ingendoh, Christian Albers, Romano Hebeler, Alexander Harder, Pierre-Olivier Schmit
Bruker Daltonik, Bremen, Germany

Keywords
Biopharma Quality Control
Complete characterization
Top-down protein analysis
maXis
ETD
recombinant proteins

Instrumentation & Software
maXis

For research use only. Not for use in diagnostic procedures.

www.bdal.com

Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (0)421-2205-0
Fax +49 (0)421-2205-103
sales@bdal.de

Bruker Daltonics Inc.

Billerica, MA · USA
Phone +1 (978) 663-3660
Fax +1 (978) 667-5993
ms-sales@bdal.com