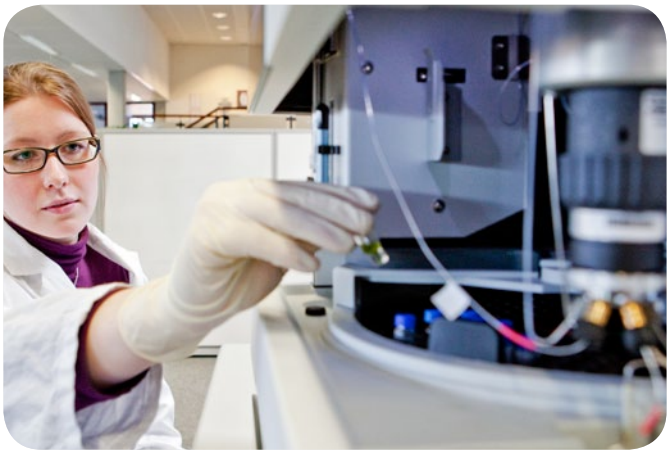

Application Note # ET-36

N-Terminal Top-Down Protein Sequencing by ETD-UHR-QTOF Mass Spectrometry

Abstract

MS based Top-Down protein sequencing techniques are becoming increasingly important mainly because relevant sequence information is obtained very quickly and accurately. This meets the demand of growing areas like identity conformation of biopharmaceuticals and their side products or metabolites.

In this study, we applied a general analysis strategy for Top-Down sequencing based on Electron Transfer Dissociation (ETD) mass spectrometry in order to obtain sequence information on 4 different proteins.

Authors

Christian Albers
Bruker Daltonik GmbH, Bremen, Germany

Keywords	Instrumentation and Software
Intact Protein	maXis 4G ETD
Top-Down Sequencing	BioTools
ETD	

Introduction

Detailed protein characterisation is increasingly used in different research fields like proteomics or biopharmaceutical characterisation. In the proteomics area researchers want to focus on a biologically interesting subset of proteins in more detail to gain more insight into protein candidates from biomarker discovery projects. In the biopharmaceutical area it is important to fully characterize a protein for regulatory approval or later stages in the processing pipeline like stability testing or impurity detection and identification.

In connection with these general protein characterisation tasks we participated in the 2012 study organized by the ABRF Protein Sequence Research Group (PSRG) [1]. This studies objective was to determine the capacity of various technology platforms for N-terminal sequencing of proteins - even in more complicated cases such as the presence of ragged ends, N-terminal modifications or near terminal fusion sites. Therefore, the ABRF 2012 PSRG study comprised of 3 pure protein samples the identities of which were provided to the participants whereas the N-terminal details were not revealed: one sample for establishing the technique (BSA), a fusion protein with a blocked N-terminus (recombinant Protein A) and a mixture of two proteins with different N-terminal variants (Endostatin).

MALDI-TOF on Bruker MALDI-TOF instruments is well established for these applications and the software well developed [1]. However, with a new generation of ultra-high resolving QTOF instruments with dramatically increased ETD capabilities another mass spectrometry option for Top-Down Sequencing (TDS) is available. We used the samples provided by the PSRG to evaluate the current ETD-UHR-QTOFs TDS capabilities.

Fig. 1 shows the schematic of the setup of the ion transfer system of a maXis 4G UHR-TOF with ETD option. A negative chemical ionization (nCI) source is integrated into the standard ion transfer system for generating reagent anions and injecting them onto the ion pathway. This arrangement of the nCI source and transfer octapole is similar to the setup of the amaZonTM ETD ion trap and has proven its robustness and ease-of-use for a long time.

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-C-α bond cleavages leading to c- and z-type fragments of the proteins or peptides. This non-ergodic process occurs randomly along the amino acid backbone. Posttranslational 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.

We describe here the N-Terminal sequencing of four different proteins in the mass range from approx. 20-70 kDa by maXis 4G ETD including two different processing strategies based on Brukers BioTools™ software for identification of intact proteins (Fig. 2a and 2b).

maXis ETD system

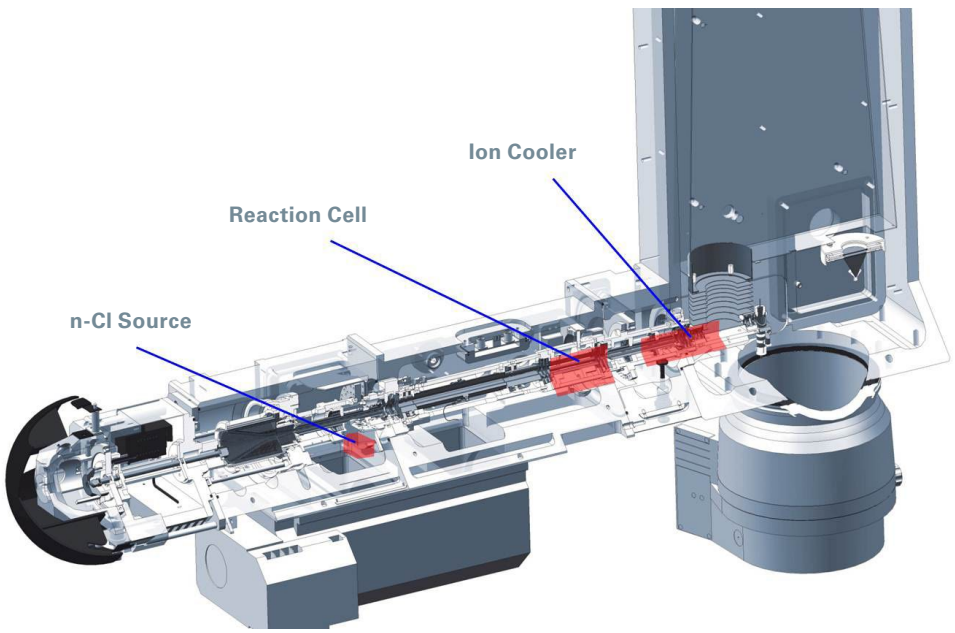

Figure 1: Innovative technical developments on the maXis 4G ETD system for enhanced analytical performance; the nCI source is integrated into the standard ion transfer system.

Experimental

Following proteins were provided by the ABRF PSRG: Bovine Serum Albumin (BSA) as test protein to establish and practice the preferred technology (approx.1 mg), Protein A, a fusion protein with blocked N-terminus (3 vials with 100 pmol each), and Endostatin with two different N-terminal variants as mixture (3 vials with 100 pmol each).

All samples were dissolved in a mixture of water/acetonitrile/2-propanol/formic acid to an approximate concentration of 2 pmol/µl and were analyzed without further sample preparation. Measurements were carried out with a maXis 4G ETD system (Bruker Daltonik, Bremen, Germany) coupled with a Triversa Nanomate (Advion, Ithaca, NY, USA) as offline nanospray experiments.

For each sample, optimization of MS transfer settings was performed first and MS spectra were acquired for each sample. These data were used to determine the intact masses of the provided proteins and to decide on the best m/z value for ETD fragmentation from the charge state distribution of each protein. After selecting a specific charge state as precursor for the Top-Down experiment the ETD settings (analyte accumulation time, reagent accumulation time, and extended reaction time) were optimized to obtain a good fragmentation pattern with a high yield of MS^2 fragment ions. Depending on the intensity of each precursor ion different acquisition times were chosen to obtain sufficient ion statistics for good sequence coverage.

The obtained deconvoluted peak list from each ETD experiment was opened in BioTools 3.2 software (Bruker) and processed either with a TDS database search strategy utilizing a standard Mascot 2.3 search engine (Matrix Science; Fig.2a) or using MS-BLAST for protein identification if the Mascot database search wasn't successful (Fig. 2b) due to unexpected sequence variations (N-term modification or in case of fusion protein).

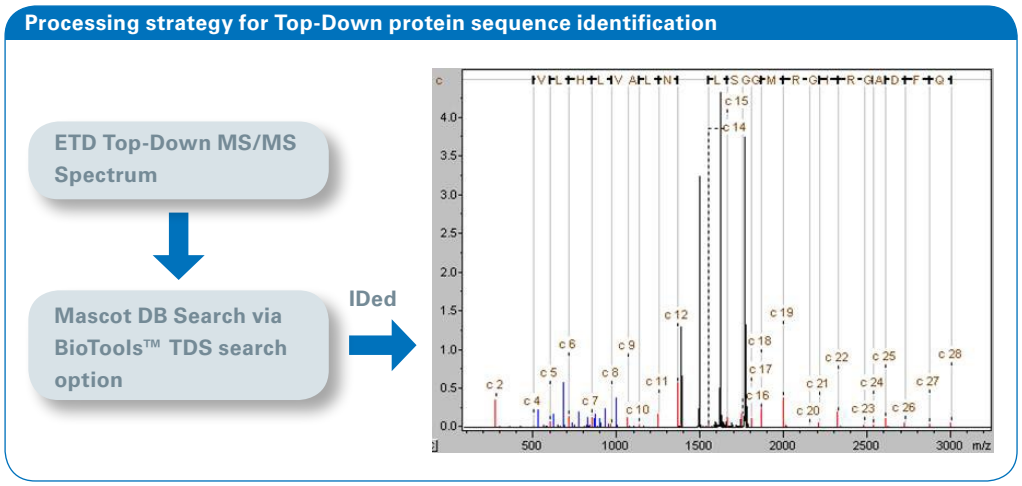

Figure 2a: Processing strategy for Top-Down protein sequence identification based on a standard Mascot server, and BioTools™ with TDS search option. The resulting sequence annotation as obtained in BioTools™ is shown.

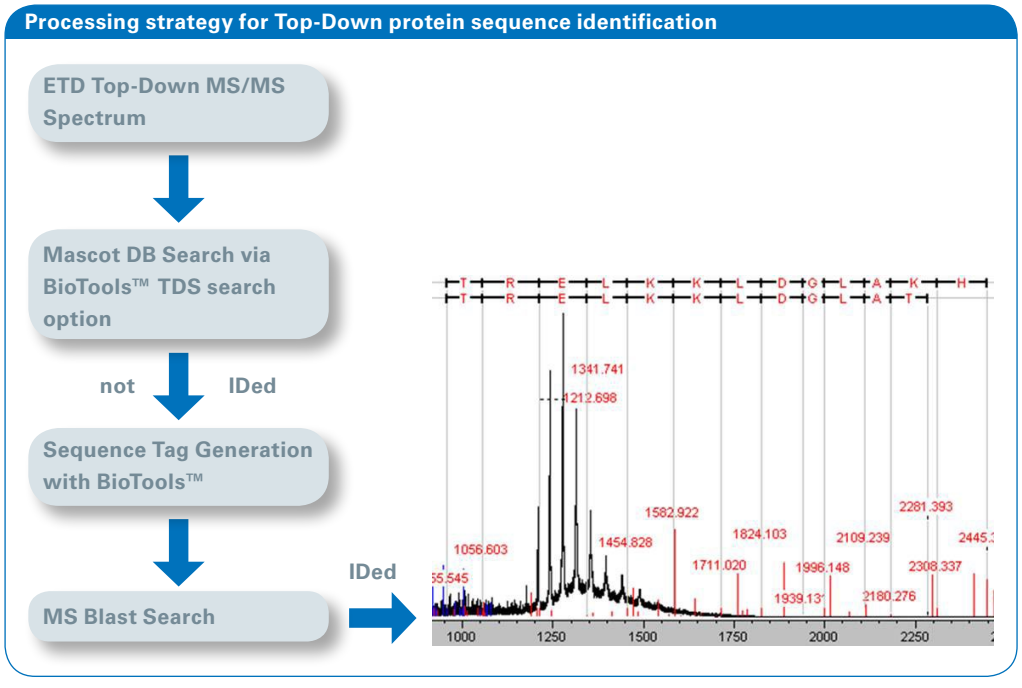

Figure 2b: Processing strategy for Top-Down protein sequence identification for samples not identified with standard database search; BioTools™ generates a sequence tag which is used for a MS BLAST search. The resulting sequence annotation as obtained in BioTools™ is shown.

Results

All samples could be analyzed with offline nanospray ESI MS without further sample cleaning. For the Endostatin sample it was necessary to include In-Source CID (ISCID) energy to obtain a good MS/MS spectrum with good signal intensities. This is quite surprising for such a small protein as ISCID energy is usually only necessary for proteins with a MW ≥ 80 kDa in order to decluster the analyte ions from solvent adducts. Possible reason for this behavior is the four metal binding sites in this protein which may support adduct formation leading to heterogeneous spectra. This is a nice example how challenging each individual protein can be for protein sequencing or characterization respectively.

The fragmentation data from intact BSA in Fig. 3 demonstrates the ability of the Bruker maXis 4G ETD ESI-UHR-TOF instrument to deliver comprehensive and information rich Top-Down fragmentation spectra. This complex spectrum with overlapping peak patterns from the huge number of obtained fragment ions is processed with a dedicated software algorithm (SNAP II™) to determine mono-isotopic masses of all fragment ions. The generated peak list is used for a Mascot database search based on a BioTools virtual precursor approach. This approach makes use of some abundant fragment ions with a molecular weight (MW) ≤16 kDa from the peak list and allows a database search with a standard Mascot server without MW restriction of measured intact proteins.

The database search result shows N-terminal sequencing for BSA until amino acid at position 51 with an excellent mass accuracy of fragment ions (Fig.4).

The virtual precursor strategy was also successful for the two Endostatin variants (Fig.5a). N-terminal sequencing was performed for both variants until the first disulfide bridge of the protein, sequencing of first 28 and 32 amino acids resp. was possible. The gaps in the annotated sequence (Fig.5 b) are proline gaps which can be explained by the N-C-α bond cleavages of the ETD fragmentation process. The N-C-α bond is part of the ring structure of proline. Therefore, c-ions are not detectable in a fragment spectrum at the position of proline in a protein sequence thus representing XP dimer sequences. For this sample intact high precision mass measurement delivered important information about this protein in addition to the TDS analysis (Fig. 6). The determined intact protein mass confirmed that the C-terminal Lysine was missing in both endostatin variants with an accuracy of 3 ppm.

For Protein A the database search wasn't successful as this fusion protein doesn't exist in common databases. Therefore, a different sequence analysis strategy was employed to determine the N-term sequence of this fusion protein (Fig.7). BioTools was used to calculate different sequence tags from the peak list. These sequence tags

Spectra section

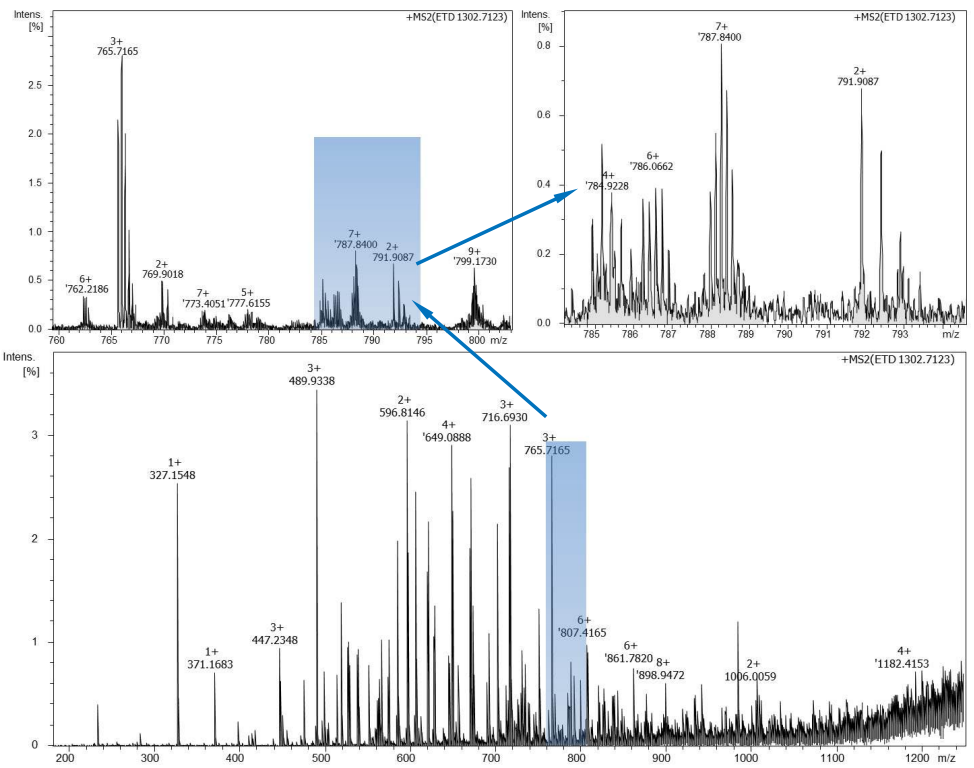

Figure 3: Spectra sections of the entire ETD-MS/MS fragment spectrum of intact BSA (approx. 67 kDa). All observed fragment ions are clearly isotopically resolved.

were used for a MS-BLAST [2] search and two different protein sequences were identified: S. aureus Protein A and E. coli α-glucuronidase. The nature of the protein A sample containing a fusion protein was thus confirmed, the identity of the fusion partners identified and the fusion site localized by sequencing across the fusion region. The first 30 amino acid residues were assigned, the first 18 residues originating from α-glucuronidase. In addition, the methylation of the N-terminal methionine residue was correctly detected from the ETD-TDS data. Overall, the maXis 4G ETD UHR-QTOF system showed excellence performance for the analytical task of this study. All complete results of the study can be accessed at the ABRF-PSRG website [1].

Zoomed ETD-MS/MS spectrum

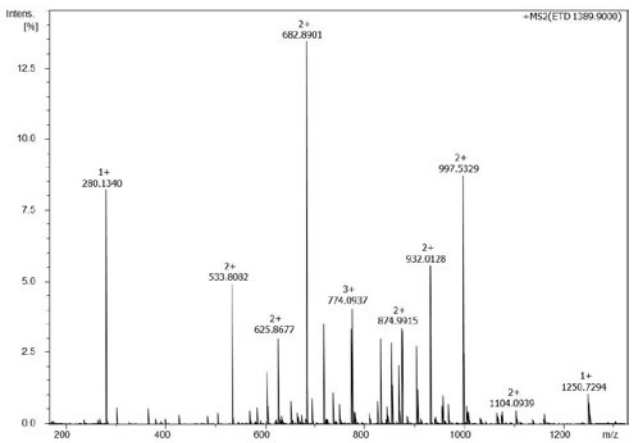

Figure 5a: Zoomed ETD-MS/MS spectrum from the Endostatin mixture. The spectrum shows the Endostatin form with the lower molecular weight.

Annotated ETD-MS/MS spectra in BioTools™

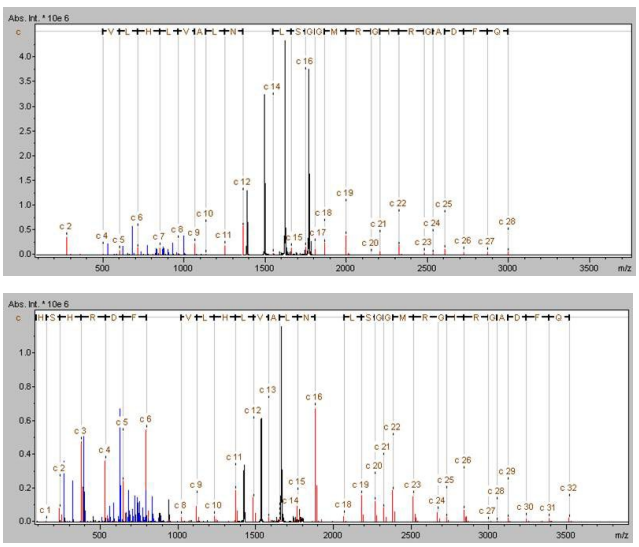

Figure 5b: Annotated ETD-MS/MS spectra in Biotools™ after Mascot database searches of the two analyzed forms from the Endostatin mixture. The spectrum at the top shows the Endostatin form with the lower molecular weight whereas the spectrum at the bottom shows the form with the higher molecular weight. The gaps in the two annotated sequences are caused by proline. Due to the ring structure of this amino acid including the N-C-α bond no ETD fragment ion can be detected for this position.

Mascot search result

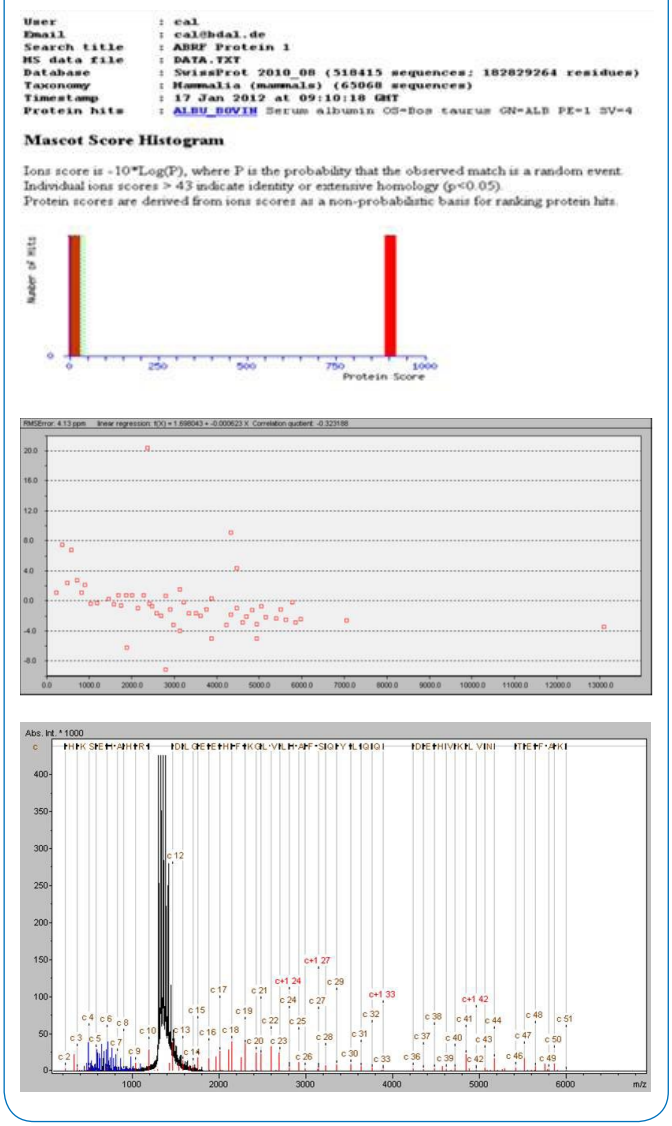

Figure 4: Mascot search result from database search with virtual precursors strategy of obtained ETD spectrum from intact BSA (top). Annotation within BioTools™ (bottom) shows fragment ions for almost complete N-Terminal sequence coverage up to position 51. In the middle the error distribution of all assigned fragment ions nicely shows the excellent mass accuracy of the maXis.

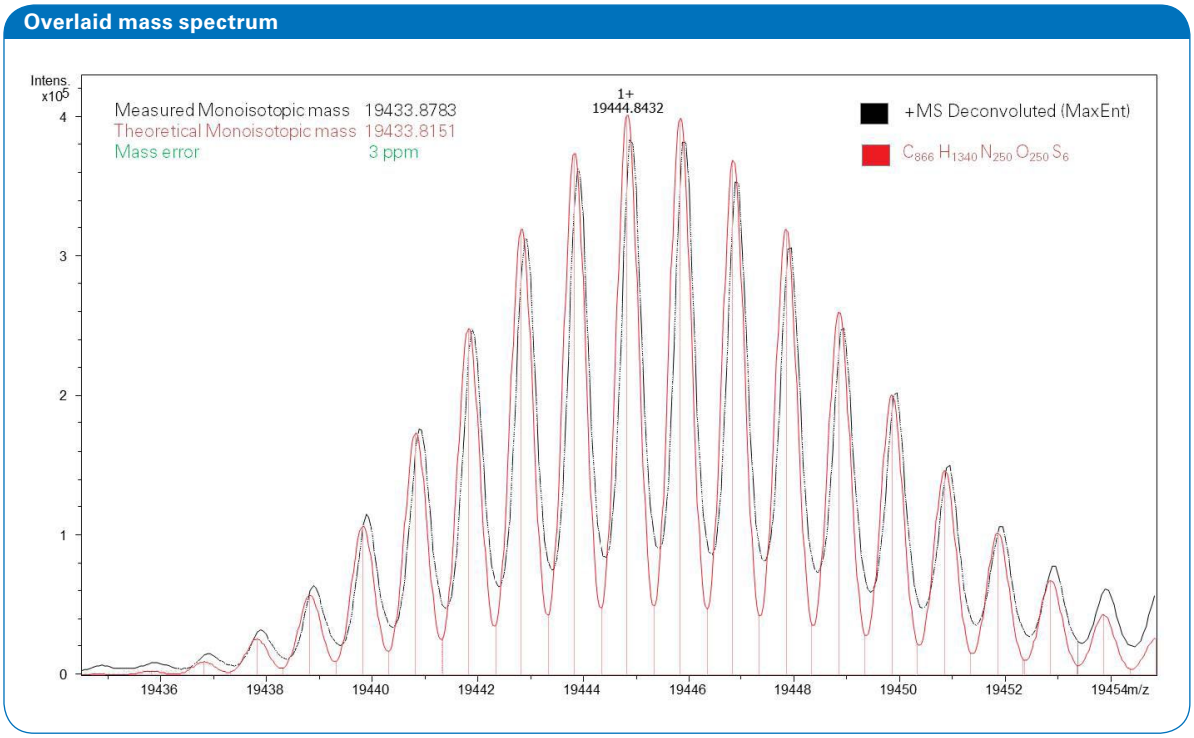

Figure 6: Overlaid measured intact mass spectrum (black) and simulated mass spectrum (red) without C-terminal Lysine of the protein with the lower molecular weight form from the Endostatin mixture. This spectrum shows the excellent mass accuracy the maXis offers for intact protein analysis and how the intact mass information complements the fragmentation data for a full protein characterization (same occurs for the second Endostatin protein; data not shown).

Conclusion

Top-Down protein analysis allows for identification and detailed 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 comprehensive characterization of proteins and artifacts introduced by tryptic digestion can be excluded.

The combination of high mass resolution, mass accuracy and high ETD efficiency makes the maXis 4G ETD UHR-QTOF system well suited for Top-Down protein investigations like bio-therapeutic drugs or isolated proteins from biomarker studies. Examples shown here cover a mass range from approx. 20 kDa up to 70 kDa. Together with its unique capabilities for intact protein mass measurements and excellent software support for fast and reliable evaluation strategies the maXis ETD system delivers unique performance characteristics for Top-Down experiments.

Acknowledgments

ABRF SPRG kindly provided samples for this study.

Workflow for identification and N-terminal sequencing of analyzed fusion protein

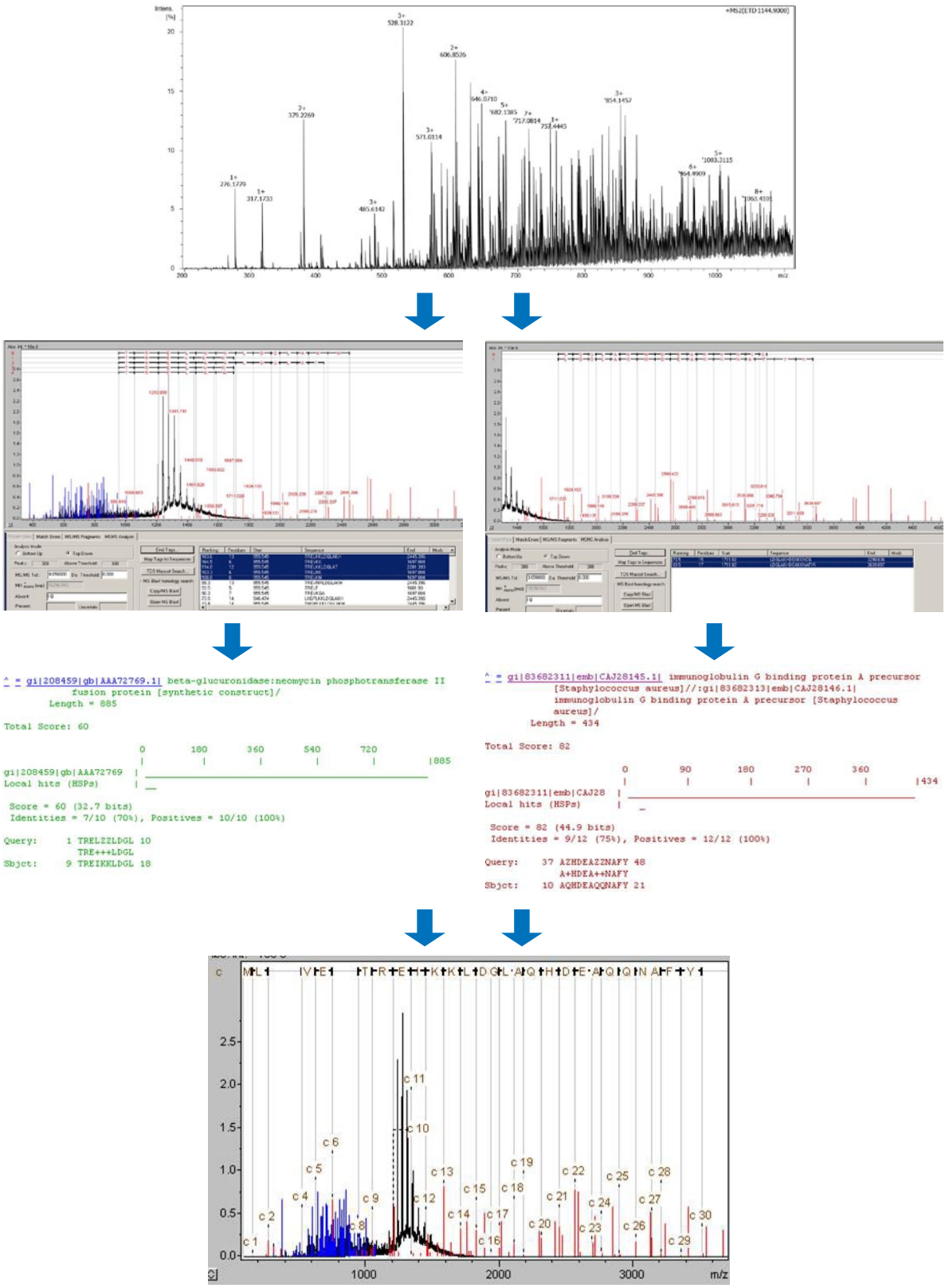

Figure 7: Workflow for identification and N-Terminal Sequencing of analyzed fusion protein (not Ided via Mascot database search). Top: Zoomed ETD Fragment spectrum of Protein A sample. Middle-1: Two sequence tags generated in different parts of the deconvoluted fragment spectrum. Middle-2: Results of MS-BLAST search for the two generated sequence tags leading to the identification of two different proteins; these two parts could be combined; a sequencing across the fusion region was possible. Bottom: Annotated Protein A across fusion region in BioTools™, gaps in the annotation are caused by proline gaps.

References

[1] http://www.abrf.org/ResearchGroups/EdmanSequencing/EPosters/PSRG2012_poster(FINAL).ppt

[2] Gish, W. (1996-2006) http://blast.wustl.edu

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

Bruker Daltonics is continually improving its products and reserves the right to change specifications without notice. © Bruker Daltonics 11-2012, ET-36, #704173

Bruker Daltonik GmbH

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

www.bruker.com

Bruker Daltonics Inc.

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

Fremont, CA · USA
Phone +1 (510) 683-4300
Fax +1 (510) 490-6586
ms-sales@bdal.com