



Application Note # MT-90

MALDI-TDS: A Coherent MALDI Top-Down-Sequencing Approach Applied to the ABRF-Protein Research Group Study 2008

In the ABRF-PRG study 2008 [*] the ability to characterize recombinant proteins was challenged. The study asked for analytical capabilities typically required to QC recombinant proteins. We used a Top-Down-LC-MALDI-reISD strategy to ID the protein and to assign C-terminal heterogeneity of RAGE_HUMAN derived proteins

Introduction

The aim of the study was to survey the useful approaches for qualitative protein analysis in a comparative approach on recombinant proteins. The application resembles closely the analytical task to detect and elucidate possible differences in batches of recombinant proteins, being a standard QC type of question.

We used a Top-Down-LC-MALDI-reISD strategy to initially LC-separate the proteins contained in each sample as intact, undigested molecules, determined the respective molecular weights and identified the protein species by Top-Down Sequencing using a MALDI-TOF/TOF instrument. This MALDI-TDS approach was the only native Top-Down strategy employed in this study, yielding the detection, identification and full characterization of all protein expression products on which the study was focussing.

Sample

Two study samples (A and B) were supplied by ABRF PRG. These consisted of separate preparations of related human proteins that were expressed in *E. coli* and affinity purified.

Methods

Sample preparation

Both samples were reduced and alkylated, then LC separated using an ML-CAP-PM monolithic column with a 200 μm inner diameter and a PS-DVB (Dionex) stationary phase. The flow rate was set at 2 $\mu\text{l}/\text{min}$. Only a few picomols of each sample were loaded and separated across a progressive gradient of 20 – 40 % solvent B in 45 min.

Solvent A = 0.1 % TFA in water

Solvent B = Acetonitrile/ 0.1 % TFA

Bruker AnchorChip™ targets (2 mm diameter) were manually supplied with a thin layer of sinapinic acid. Post-column MALDI preparation was carried out using a Bruker PROTEINEER fc™ LC-MALDI fraction collector. 30 sec fractions were collected on the target spots combined with a sheath flow of sinapinic acid (1 μl of saturated solution in 30 % ACN/ 0.1 % TFA in water per fraction).

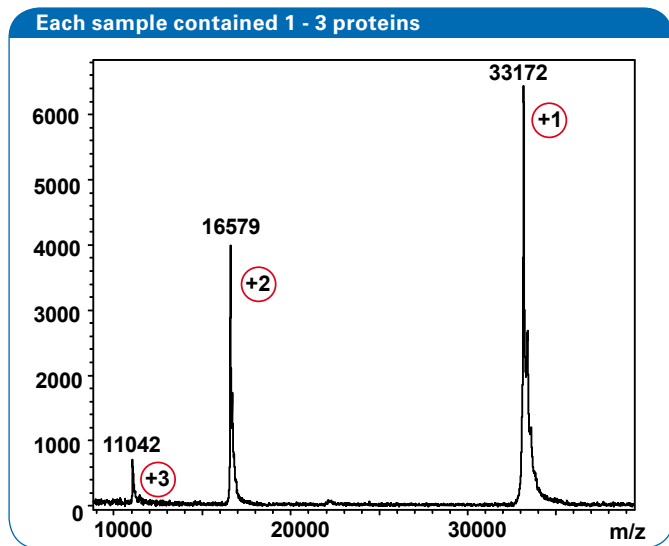


Fig. 1: Sample A: Linear MALDI-TOF spectrum of a single protein.

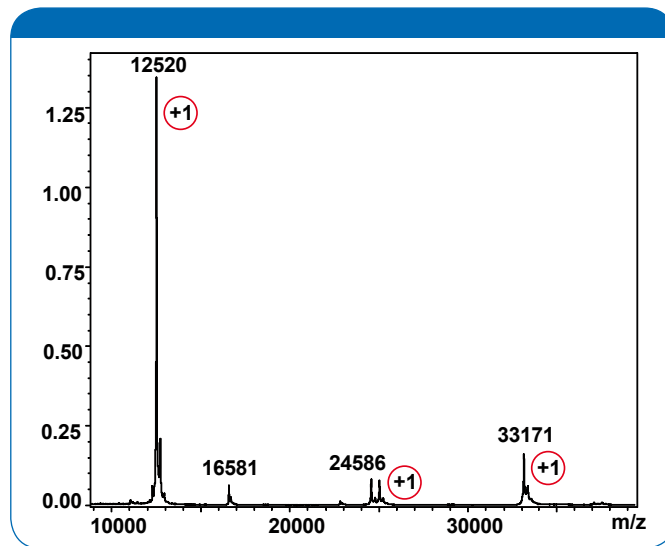


Fig. 2: Sample B: Linear MALDI-TOF spectrum shows three proteins in a mixture.

Spectra acquisition

All spectra were acquired on a Bruker ultraflex III MALDI TOF/TOF™ mass spectrometer. Intact protein molecular weights were obtained in linear mode, in-source-decay spectra in reflector mode (relSD) essentially as described in [1] and all bottom-up MS/MS spectra were acquired without the use of collision gas as described in [2]. The WARP-LC™ 1.1 software tracked the fractions throughout the analytical procedure and was used for LC-data analysis.

Data analysis

Monoisotopic peaks were annotated in the acquired spectra in Bruker Compass™ 1.2 using the SNAP algorithm. All protein and peptide identifications were triggered in BioTools™ 3.1 (Bruker) using Mascot™ 2.2 (Matrix Science) as search engine. Sequence tags and all Top-Down related sequence analyses were generated in BioTools using a mixture of automated calculations and interactive sequence assignments.

Results

Determination of Sample Composition

An analysis of both samples in their native form provided molecular weights of the intact proteins. The results showed an immediate correlation between the two samples and provided evidence that each sample contained 1-3 proteins. Figure 1 and Figure 2 are the MALDI-TOF outputs from this first analysis. Sample A shows the clear spectrum of a single protein (charge states 1 to 3), sample B consists of at least 3 different sized protein species. The molecular masses of the proteins were also used to confirm suggested protein structures later in the analysis process.

Automated LC-MALDI-MS analysis was used to generate in depth comparative views of Sample A and Sample B (Figure 3) and to separate the protein mixture for downstream sequence analysis. The 33.2 kDa protein can be seen to be present in both samples. However, Sample B has additional strong peaks at 24.6 kDa and 12.5 kDa, confirming the presence of 3 proteins in Sample B.

Top-Down Sequencing and Protein Assignments

Reflector in-source decay (relSD) spectra of the intact species were then collected after LC-separation. RelSD produces very regular fragmentation of intact protein

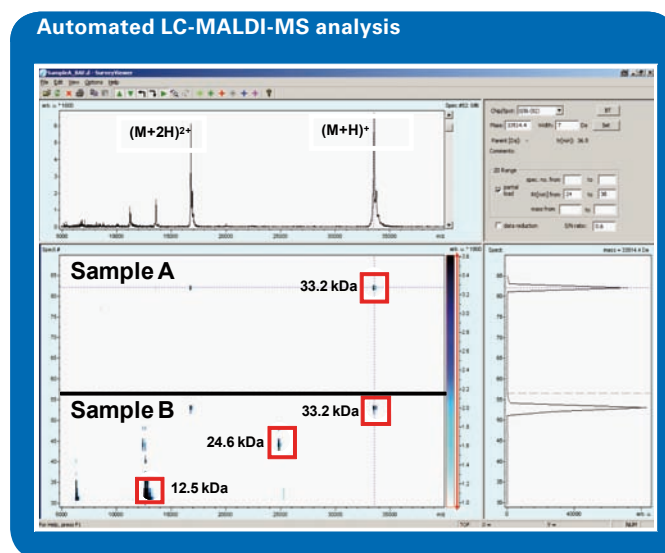


Fig. 3: Top-Down LC-MALDI analysis of Sample A and Sample B (RT vs. m/z), Sample A containing a single 33.2 kDa protein, Sample B containing additional strong peaks at 24.6 kDa and 12.5 kDa. Red boxes indicate the detected proteins that were sequenced with relSD.

molecules [1,2], permitting direct sequence analysis of undigested proteins. This approach of MALDI Top-Down Sequencing is dubbed "MALDI-TDS". The peak list was processed through Bruker's BioTools 3.1 software where c-ion sequence tags were generated (Figure 4) and used for protein identification either using Sequence Queries (in Figure 5) or homology searches in the MS-BLAST (EMBL/Harvard) software. A Mascot Sequence Query returned a high probability match (Protein Score 119) with the RAGE_HUMAN (Q15109) advanced glycosylation

end product-specific precursor (Figure 6). The entire sequence was then matched the BioTools sequence editor with the experimental sequences derived from the peptide fragments. The signal peptide was identified by the software (Figure 7) and removed from the experimental sequence, confirming the N terminus as GSHMAQNITARIGE and BioTools successfully annotated the matched ISD fragments and the spectrum (Figure 8), allowing the N- and C-Termini to be confirmed on the intact Sample A protein. The annotation process was repeated for the 24.6 kDa and the 12.5 kDa fragments included in Sample B. These were found to have the same N-Terminus as the intact protein but with different C-Terminal truncations (Figure 9, Figure 10 and Figure 11). The assignments were guided by the intact protein molecular weights and confirmed by the fully assigned reISD spectra.

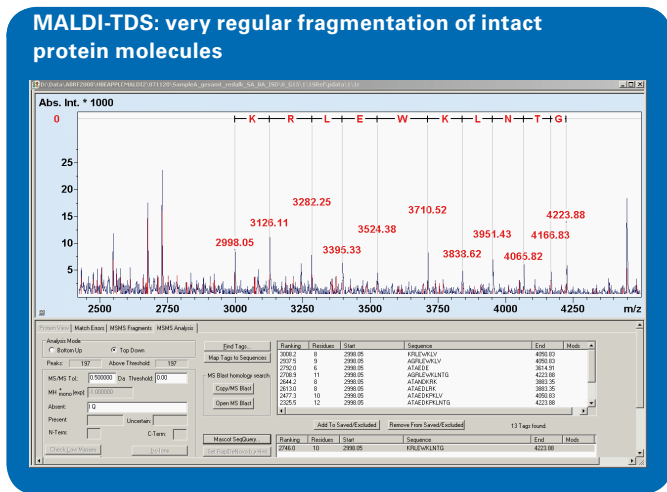


Fig. 4: Automated sequence tag generation in BioTools from the reISD spectrum from Sample A.

100 % Sequence coverage confirmed by bottom-up-LC-MALDI

At this point, the identity of each protein species had been confirmed and all N- and C-termini confidently assigned. Although ISD only acquires fragment ions approx. covering residues 10-60 from each end of the sequence, the correctness of the assignment of the core part of the sequence can be deduced from the match of the calculated molecular weights based of the established sequence with the intact protein mass spectra. All 3 protein forms provided matches of the calculated vs. experimental molecular weights within better than 5 Da. This suggested a correct structural assignment of all proteins without any additional modifications such as phosphorylation. A bottom-up analysis of Sample A was conducted purely to confirm the 100 % structural assignments of the Top-Down data. Trypsin would normally be a favored protease for this purpose, cleaving the peptide chain at the carboxyl side of lysine and arginine. However the initial Top-Down analysis indicated that this would result in peptides and amino acids which were either too small or too large for effective MS/MS work. Endoproteinase Glu-C was therefore selected as a more effective protease for this purpose. It hydrolyzes peptide bonds at the carboxyl side of Glu and Asp and typically provides a high sequence coverage in LC-MALDI-MS/MS. The peptides resulting from the Glu-C digest of Sample A were run through LC-MALDI MS and MS/MS and searched with Mascot without any enzyme restriction, providing 100 % coverage of the RAGE_HUMAN protein (Figure 12) including the central region of the protein. The N-terminal fusion peptide did not provide a match as it is not present in the native sequence listed in Swiss-Prot. However, from the molecular weight deduced from the suggested protein sequence it was easy to find the peptide in the LC-MS/MS data set, assign the putative structure to the respective MS/MS spectrum and verify its proper match (Figure 13).

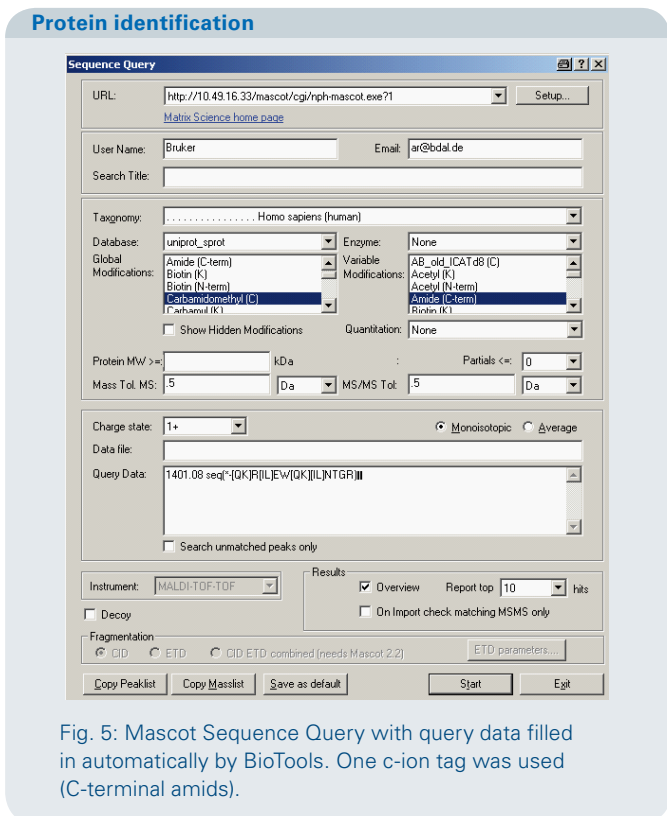


Fig. 5: Mascot Sequence Query with query data filled in automatically by BioTools. One c-ion tag was used (C-terminal amids).

Sample B fragments have the same N-Terminus but different C-Terminal truncations

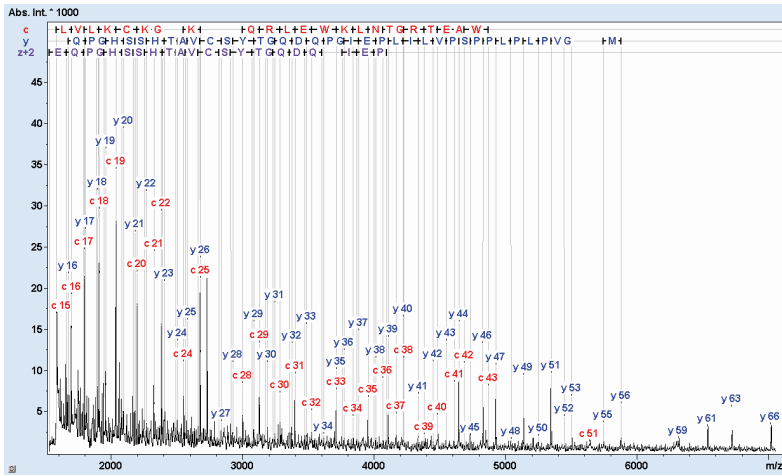


Fig. 9: Sample B: 35.2 kDa analysis.

Sample B fragments have the same N-Terminus but different C-Terminal truncations

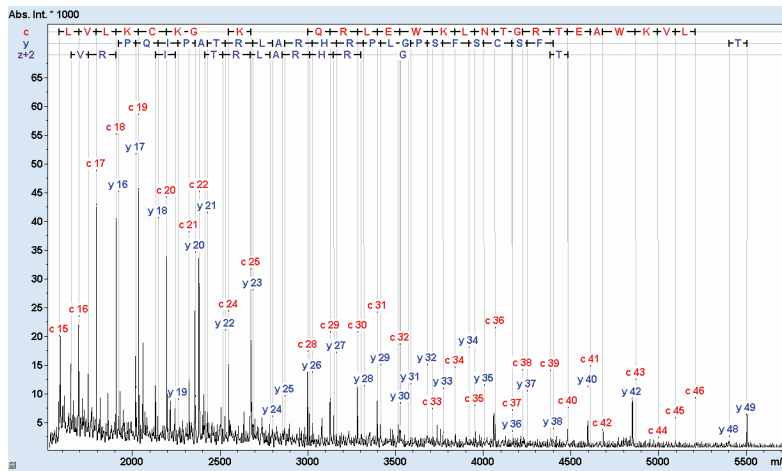


Fig. 10: Sample B: 24.6 kDa analysis.

Sample B fragments have the same N-Terminus but different C-Terminal truncations

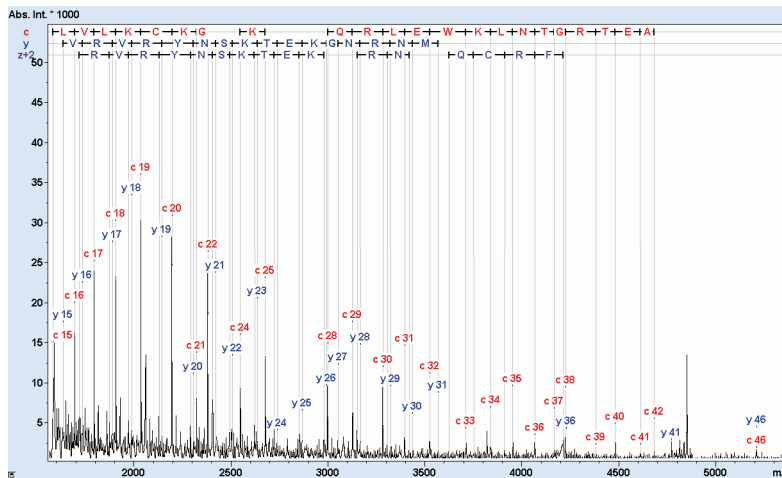


Fig. 11: Sample B: 12.5 kDa analysis.

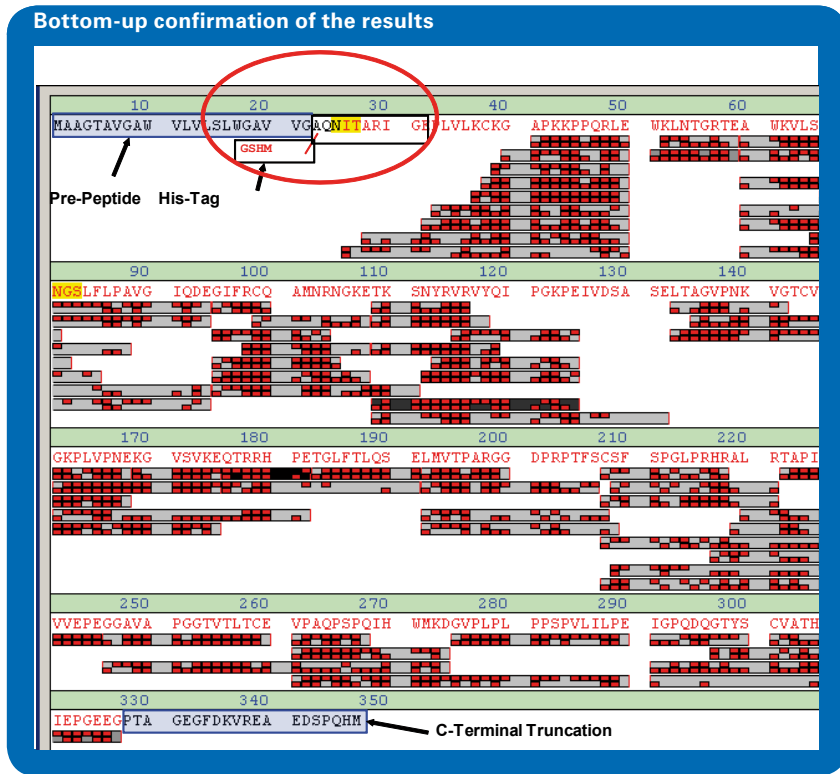


Fig. 12: LC-MALDI result of Glu-C digest of Sample A. Peptide matches (gray bars) as well as fragment matches (red bricks), top rows: N-terminal fragments, bottom rows: C-terminal y-ions.

Keywords
MALDI-TDS
Top-Down Sequencing
Protein identification
N-terminus assignment
C-terminus assignment

Instrumentation & Software
ultraflex III TOF/TOF
PROTEINEER fc
AnchorChip
WARP-LC
BioTools

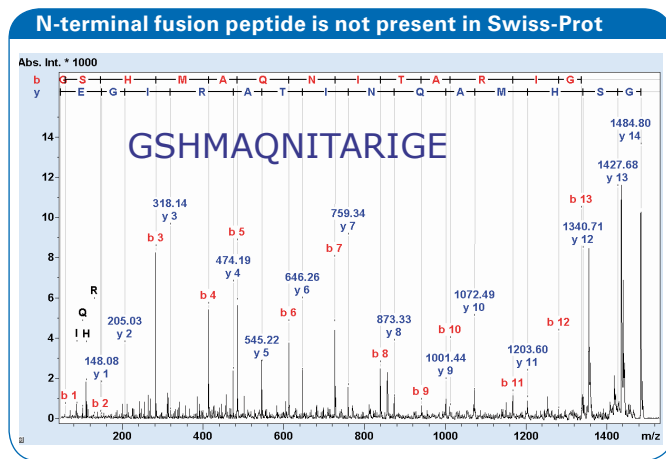


Fig. 13: Sample A: N-Terminus validation by MS/MS.

Authors

A. Resemann, M. Kay, S. Brand, D. Suckau
Bruker Daltonik GmbH, Bremen, Germany

References

- [1] Detlev Suckau and Anja Resemann (2003) T³-Sequencing: Targeted Characterization of the N- and C-Termini of Undigested Proteins by Mass Spectrometry Anal. Chem., 75, 5817-5824.
 - [2] Detlev Suckau, Anja Resemann, Martin Schuerenberg, Peter Hufnagel, Jochen Franzen, Armin Holle (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics Anal Bioanal Chem 376 : 952-965.
 - [3] <http://www.abrf.org/ResearchGroups/Proteomics/Studies/27960.pdf>
 - [4] http://www.abrf.org/ResearchGroups/Proteomics/Studies/PRG08_poster.pdf
- [*] The ABRF PRG group prepared and provided the sample to all members and vendors, but did not participate in the vendor's study and does not endorse any specific manufacturer, instrument or strategy

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

Mascot is a trademark of Matrix Science Ltd.

www.bdal.com • Bruker Daltonik GmbH

Bremen · Germany
Phone +49 (421) 2205-0
Fax +49 (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