

## Technical Note # TN-36

# Automated Acquisition of MALDI-ISD Spectra for the N- and C-terminal Sequence Determination of Intact Proteins

### Abstract

For over 10 years, MALDI-Top-Down Sequencing (MALDI-TDS) of intact proteins has proved particularly useful for the determination of N- and C-terminal protein sequences. This technology note describes a procedure using the recently introduced 1,5-diaminonaphthalene (DAN) matrix and the new ultrafleXtreme™ MALDI-TOF/TOF that automates the entire analysis – from sample preparation and data acquisition to N- and C-terminal sequence assignments.

### Introduction

MALDI-TDS is a powerful method for characterization of purified proteins without prior digestion [1]. The In-Source-Decay (ISD) fragmentation occurs rapidly during the MALDI process in the plume, and provides long, contiguous c- and z+2-ion sequence tags accompanied by a- and y-ions. This enables extended N- and C-terminal protein sequences to be derived from single spectra. Recent improvements in MALDI-TOF instrument performance [2] and sample preparation [3] enable generation of highly resolved spectra with ~ 10 ppm accuracy and monoisotopically resolved fragments up to approx. 8 kDa. Additional developments in dedicated top-down analysis software are facilitating and accelerating

data interpretation. Here we describe the latest methods for making MALDI-TDS an even more powerful tool for protein characterization.

### Materials & Methods

Test proteins were recombinant Protein A (rPA) [Repligen, New York], RNase B, hen egg white lysozyme (HEL), and carbonic anhydrase (CA) [all Sigma-Aldrich]. BSA [GERBU, Gaiberg, Germany] was used as calibrant of the ISD spectra. TCEP (Tris(2-Carboxyethyl) phosphine hydrochloride) [Sigma-Aldrich] was used for reduction of HEL and RNase B (5 mM final concentration TCEP in 0.1 % TFA, 1 h at 50 °C).

1,5-diaminonaphthalene (DAN) [ACROS Organics, Belgium] was used as the MALDI matrix without further purification. All spectra were acquired on an ultrafleXtreme with smartbeam™ II solid state 1kHz laser, FlashDetector™ and PAN™ broadband resolution.

### Sample preparation

A saturated colorless to pale violet solution of DAN was prepared fresh in 50 % acetonitrile/ 50 % water containing 0.1 % TFA. Speedy sample preparation is required to prevent in-solution oxidation of DAN (indicated by the solution turning dark violet) . Test proteins were dissolved

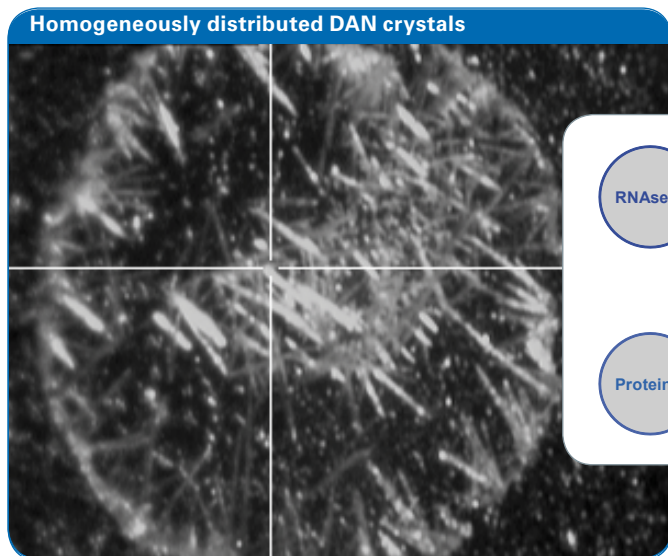


Fig. 1: Video image of the DAN matrix preparation on an 800 μm AnchorChip spot. Homogeneous crystallization provides ISD signals across the entire covered area and is a precondition for automated spectra acquisition.

at 5– 50 pmol/μl in water/0.1 % TFA and mixed at a ratio of 1:2 with the matrix solution. 0.5 μl aliquots of the protein/matrix solutions were spotted onto an 800 μm AnchorChip sample holder (MTP AnchorChip™ 800/384 T F) and air-dried at ambient temperature. The spots were examined to ensure that they consisted of homogeneous needle-like crystals evenly distributed over the target spot (Fig. 1). If crystallization was heterogeneous, 0.5 μl of pure matrix solution was added to the spot. This typically led to homogeneous re-crystallization.

Four analytical samples were prepared around the central calibrant BSA for near-neighbor calibration (Fig. 2).

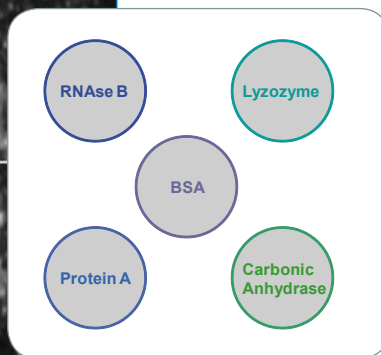


Fig. 2: Test proteins were grouped around a central calibrant spot.

## Spectra acquisition and data analysis

DAN spots typically appeared as small needle-like crystals that evenly covered the hydrophilic anchors of the MALDI sample plate (Fig. 1). For automated sample analysis, 5000 laser shots were accumulated from each spot by moving the sample holder randomly during 1 kHz acquisition. An automated acquisition sequence of 10 sets of 4 intact standard proteins and 1 intact BSA calibrant was generated. Fully automated acquisition and processing was performed on an ultrafleXtreme using a flexControl method optimized for ISD. Acquired spectra were automatically processed with Compass™ 1.3 software using the SNAP algorithm for monoisotopic peak annotation. During this procedure, the reflector ISD (relISD) spectrum of BSA was internally calibrated using the mass control list "BSA\_ISD mono.mcl", which contains theoretical c-ion masses from 1192.62 to 5784.6 Da. Average mass errors of ~ 6 ppm were achieved over the whole ISD mass range (Fig. 3). This calibration was automatically applied to the sample spectra.

All top-down related sequence analyses were carried out with BioTools™ 3.2 software. Database searches were performed using standard Mascot 2.2 with an additional MALDI-ISD instrument definition [4] on mascot server/configuration editor/instruments. All the methods used in this approach, together with a detailed description of where to save the methods and how to use them in an automated run, are available from [www.bdal.com/downloads](http://www.bdal.com/downloads). After login, navigate to "MALDI Compass" and select "MALDI-TDS".

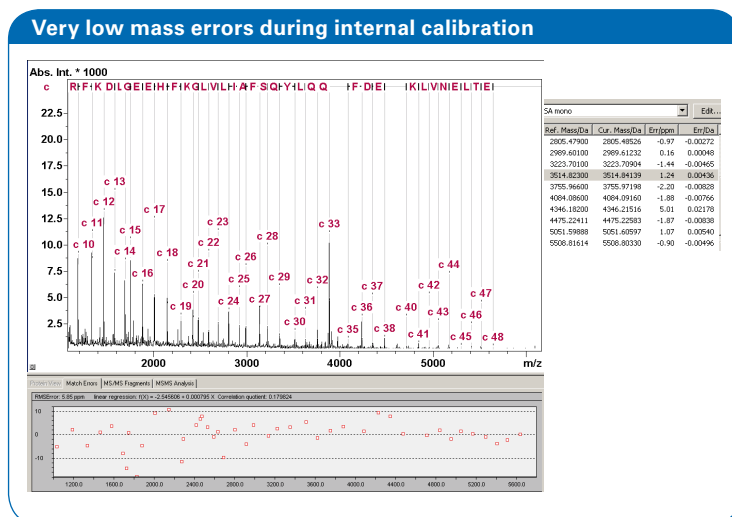


Fig. 3: MALDI-ISD spectrum of BSA in DAN and results of internal calibration. A) Matching c-ions are annotated in BioTools. B) Mass error map C) File BSA\_mono.SPL showing the theoretical masses of the c-ions used for calibration and the internal calibration errors on the right. D) Inset showing that isotopic resolution was obtained across the full mass range.

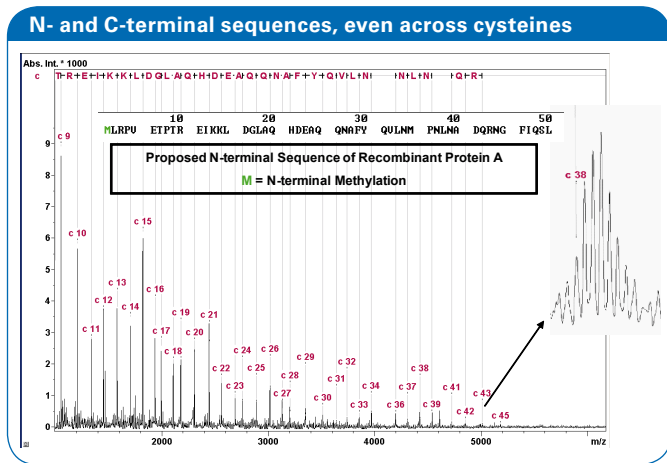


Fig. 4: MALDI-MS/MS spectrum of hen egg white lysozyme (HEL) in DAN. Due to DAN's reducing properties, sequences extend beyond cysteines that form disulfide bridges in native HEL.

## Data quality

ISD fragment intensities obtained using conventional MALDI matrices (e.g. SA, DHB and sDHB) are relatively low. In contrast, ISD fragments from DAN preparations are isotopically well-resolved and yield high intensities. Signals from intact proteins are barely detectable due to DAN's strong tendency to transfer hydrogen radicals, which induce ISD. The DAN matrix mainly produces c- and z+2-ions, which reduces ISD spectra complexity and accelerates data interpretation. DAN also acts as reducing agent that causes the partial reduction of disulfide bridges in proteins (see Figure 4).

## ISD spectra interpretation

The MASCOT TDS Search in BioTools 3.2 Service Release 1 uses a new approach that facilitates the top-down sequencing of proteins that are in a database and are searchable by Mascot. This function is accessed in BioTools through the MSMS Analysis tab of an ISD spectrum in the Top Down Analysis Mode. The TDS Mascot Search (Fig. 5) allows the assisted selection of several ISD fragments as "virtual parent masses" for a database search from the displayed mass range of an ISD spectrum. The search uses a new instrument type "MALDI ISD" that is configured on the MASCOT in-house server. All selected peaks within the zoom region can be used as virtual precursors for a MASCOT TDS MS/MS Ions Search as if they were different peptides in a mixture.

In contrast to Edman Sequencing, MALDI-TDS can be used for analysis of proteins with modified N-termini. However, MASCOT TDS searches are not successful if these modifications are undefined and not accounted for in the definition of the search. To overcome this possible limitation BioTools allows the fast generation of sequence tags that can be used for MASCOT sequence queries in such cases.

A selected tag is automatically transferred into the required sequence string format by Mascot SeqQuery. Use of this feature allowed identification of CA from its IDS spectrum (Fig. 6) and correctly suggested that its N-terminus was acetylated.

If neither of the procedures described above leads to an identification of the protein, the homology search via MS-BLAST (Harvard Medical School) may be used as a further option. It is easily accessible from BioTools and provides a high degree of error tolerance but it is the least straightforward approach and more time-consuming.

## Summary

Using the DAN matrix and the new ultrafleXtreme MALDI-TOF/TOF, an automatic procedure for acquisition of protein top-down sequencing spectra has been established that can be directly used to determine protein N-terminal and (less frequently) C-terminal sequences. Dedicated software tools facilitate the identification and annotation of protein structures elucidated using such spectra for different analytical applications.

- Protein QC: Proposed protein sequences can simply be overlaid with the experimental ISD spectrum to directly confirm the sequence. BioTools software features allow sequences to be screened for truncation variants or terminal modifications [5].
- Protein ID: In the case of unknown proteins, the TDS-Mascot search of a sequence database (triggered through BioTools) provides the ID and annotation of a sequence to the spectrum.

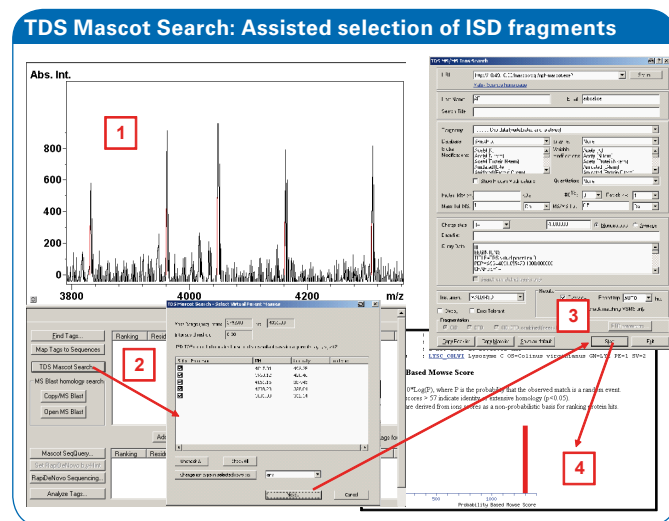


Fig. 5: New interface for Mascot searches using top-down sequencing data. Step 1: zoom in high mass range of an ISD spectrum (peaks should be properly annotated). Step 2: Press the button "TDS MASCOT Search": so-called "virtual precursors" are calculated. Step 3: TDS MS/MS Ions Search is started using "MALDI-MS/MS" as instrument type. Step 4: Protein is identified and sequence is accessible.

## Automatic generation of sequence tags for modified N-termini

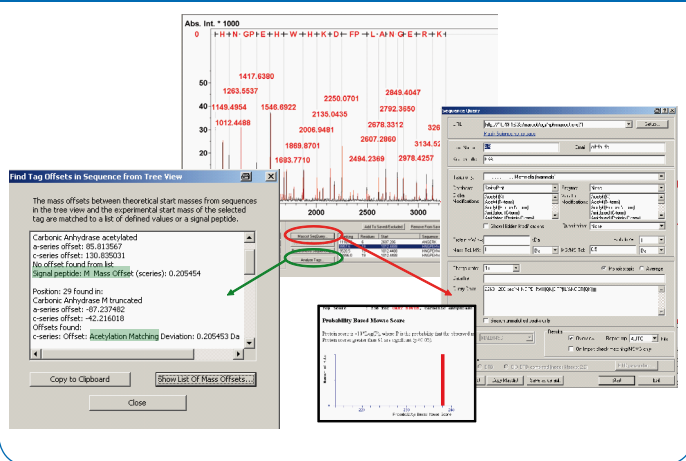


Fig. 6: A sequence tag was automatically generated by the BioTools top-down tool and used for a MASCOT Sequence Query. Carbonic anhydrase was identified and the N-terminal truncation of methionine and N-terminal acetylation were identified.

- Error-tolerant protein ID 1: Sequence tags are automatically generated in BioTools. These are submitted to Mascot sequence queries that provide clear results in those cases where unexpected sequence variations/modifications occur upstream or downstream of the determined sequence tags. The m/z position of the sequence tags in the spectrum guides the further elucidation of unexpected structural features.
- Error-tolerant protein ID 2: Sequence tags generated with BioTools can be directly used for MS-BLAST homology searches. This permits the use of top-down de novo sequencing data for cross-species or protein isoform IDs in databases that contain homologous sequences.

Using the procedures described above enables MALDI-TDS analysis to be applied to a broad range of applications, from extremely fast recombinant protein assignments to more sophisticated de novo sequencing of proteins that are not in any database.

## Conclusion

Thanks to the outstanding performance characteristics of the ultrafleXtreme system and the new DAN matrix, automation of MALDI-TDS makes it an even more powerful tool for N- and C-terminal sequencing.

## Authors

Anja Resemann, Detlev Suckau, Bruker Daltonik, Bremen Germany

## References

- [1] J. Hardouin: Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry, *Mass Spectrometry Reviews* 26, 5 672-682 (2007) doi:10.1002/mas.20142
- [2] Bruker Daltonics Technical Note TN-31: Redefining MALDI-TOF/TOF Performance: ultrafleXtreme.
- [3] K. Demeure et al: Rational Selection of the Optimum MALDI Matrix for Top-Down Proteomics by In-Source Decay, *Analytical Chemistry* 79 (22), 8678-8685 (2007). doi:10.1021/ac070849z
- [4] Bruker Daltonics Application Note MT-96: Calling N- and C-terminal Protein Sequences with High Confidence and Speed: MALDI-TDS applied to the ABRF-ESRG 2009 Research Study.
- [5] Bruker Daltonics Application Note MT-57: Reflector In-Source-Decay (ISD) MALDI-TOF MS: A Powerful Tool for N-terminal Sequence Characterization of Proteins.

Keywords
Edman Sequencing
In-Source Decay
smartbeam II
kHz-Laser
MALDI Perpetual
resolution
mass range
accuracy
robustness
Edmass

Instrumentation & Software
ultrafleXtreme
BioTools

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

[www.bdal.com](http://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