

Top-Down Sequencing of Intact Proteins by MALDI In Source Decay (ISD)

Protein Characterization using the AB SCIEX TOF/TOF™ 5800 System

Dietrich Merkel and Matthias Glückmann
AB SCIEX, Germany

Although MALDI mass spectrometry is a soft ionization technique which enables desorption of high molecular weight biomolecules exceeding several hundred kDa, it became apparent that some analyte fragmentation occurs during the MALDI process as well. The processes of fragmentation initially characterized on MALDI-TOF mass spectrometers were designated as either; (i) in-source decay (ISD)¹, the combination of laser-induced fragmentation and rapidly-occurring metastable decay, or (ii) post-source decay (PSD)², the metastable decay occurring during TOF mass separation. Both the ISD- method and PSD-method were used to obtain primary sequence information on peptides and small proteins. On modern TOF/TOF systems the terminology cannot be simply applied, due to the complex ion optics of the TOF section after the ionization source. Unimolecular or metastable decay may for example be used to describe the PSD-type processes active on TOF/TOF systems.

The direct fragmentation either by unimolecular decay and CID of small proteins on TOF/TOF systems allows direct identification of proteins from mixtures³, while ISD needs to be performed on

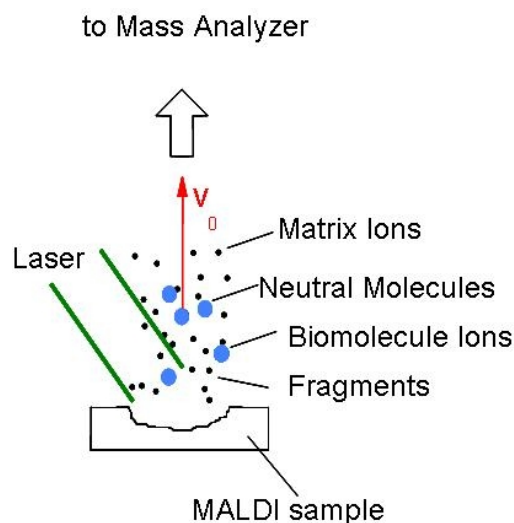


Figure 1. MALDI Process. During laser irradiation, ionization of both matrix and the co-crystallized biomolecules occurs, as well as generation of laser-induced fragments. This fragmentation, together with rapidly-occurring metastable decay, is named ISD fragmentation.



purified proteins, as the low m/z region of protein mixtures is obscured by matrix clusters and peptide fragment ion signals, and the yield of fragmentation in ISD is generally lower than for metastable decay.

The ISD technique can be performed both in single TOF mode, as well as on TOF/TOF systems. Additional MS/MS fragment information obtained on ISD-fragments (Figure 1) allows to further determination of protein sequence information⁴.

Key Features of Top-down Sequencing using AB SCIEX TOF/TOF™ Systems

- Within protein identification workflows, the in-source decay (ISD) fragmentation of proteins provides additional sequence information for protein standards.
- Specific fragments obtained by ISD could be either used to verify the intact N- and C-terminus of the protein or could even be used for identification using the “virtual precursor” approach, where a C-type ion is used as precursor mass for a no-enzyme search (described below).
- High-energy CID and metastable fragmentation of selected precursors of protein mixtures can be used to identify proteins in top-down approach.

Experimental Conditions

Sample Preparation: Proteins were used as commercially supplied without further purification (Sigma-Aldrich). The MALDI matrices used for the ISD experiments were Sinapinic acid (SINA) for Bovine Serum Albumin and 1,5-Diaminonaphtalene (DAN) for horse Myoglobin. Samples were prepared using the dried droplet preparation method by mixing 0.5 μ L of analyte solution with 0.5 μ L of matrix solution [at a concentration of 10 mg/mL in 50% acetonitrile /50% water (v:v)].

Mass Spectrometry: Spectra were either acquired in linear mode to determine the average protein molecular weight or in reflector mode to analyze the ISD fragment ion signals. All spectra were acquired using the AB SCIEX TOF/TOF™ 5800 System or the AB SCIEX 4800 MALDI TOF/TOF™ Analyzer. The low-mass detector cutoff was set to 600 Da. The laser fluence for ISD experiments was increased over the laser threshold for protein detection by about 20%. The obtained reflector spectra were calibrated externally using peptide mixtures and analyzed using MASCOT® Distiller software (Matrixscience, UK) or DataExplorer® Software.

Top-Down Sequencing for Sequence Verification

Top-down sequencing using MALDI in-source decay can be applied to either verify a known sequence, especially an intact N- or C-terminus or it could be used for the identification of the protein itself. The first approach, sequence verification, is an important tool for the quality control of biopharmaceutical proteins. A prerequisite is the knowledge of the exact sequence of the protein. In this case, the known sequence can be pasted into the ion fragmentation calculator in DataExplorer® software. The labeling options within the ion fragmentation calculator need to be adjusted to reflect the specific fragmentation observed within MALDI in-source decay spectra, as mainly c- and y-type ion signals are detected. The option of labeling of internal fragments should not be activated.

Figure 2 shows an in-source decay spectrum of Myoglobin opened in the DataExplorer® software and the adjusted options for the fragment labeling. The software tool calculates all theoretical fragment masses and labels the corresponding peaks automatically within the in-source decay spectrum. As shown in Figure 3, a long series of c-ion signals verify the correct N-terminus of the analyzed protein. Even though c-ion signals from the protein N-terminus are the predominant fragment species, the detection of several y-ions also verify the intact c-terminus of the protein.

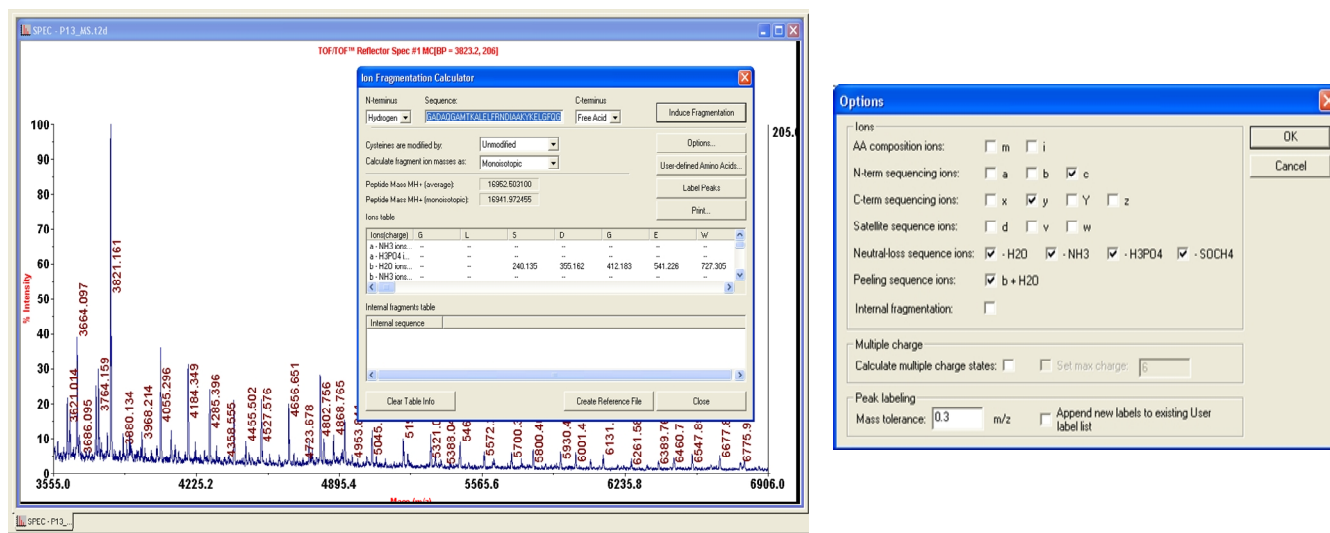


Figure 2: Top-Down Sequencing of Myoglobin for Sequence Confirmation. The in-source decay spectrum was opened in the DataExplorer® software, the Ion Fragmentation Calculator application was used to annotate the sequence specific ions by annotation using theoretically calculated masses of c- and y-type ions from the known sequence.

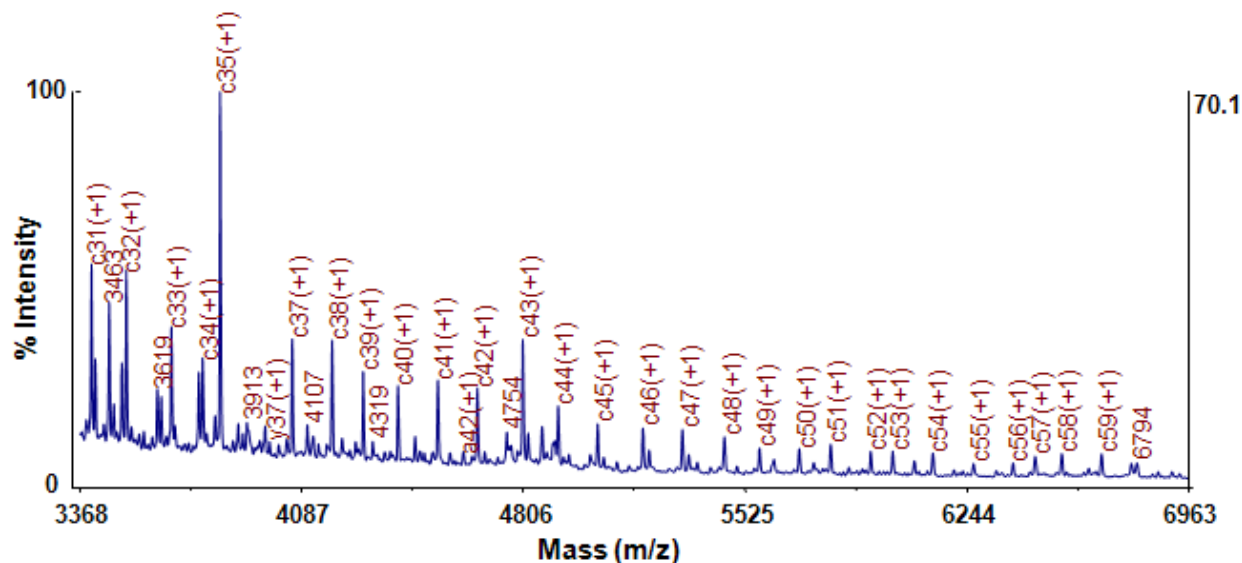


Figure 3: Detected C-ions from In-Source Decay of Myoglobin. As the theoretical fragment ion m/z values are matching the ion signals observed, this approach can be used to verify the correct N-terminus of the protein. Shown are the first 60 residues from the N-terminus of horse Myoglobin, corresponding to the sequence GLSDGEWQQVLNVWGKVEADIAGHGQEVLRIRLFTGHPETLEKFDKFKHLKTEAEMKASED--.

Top-Down Sequencing for Protein ID using MS BLAST

In order to identify a protein using ISD data, there are two possible ways of data evaluation available. The first possible approach is using sequence tags, which are blasted against any protein database to identify the protein. In the example we show here, an in-source decay spectrum of BSA, which was analyzed with the MASCOT Distiller software. This software generates a

de novo sequencing tag based on the in-source fragments (Figure 4). The obtained sequence was submitted to an integrated MS BLAST search, which provided a clear identification of BSA (Figure 5). The de novo sequence tag could also be generated by any other available software tools, like the De Novo Tool (integrated into the acquisition software of the TOF/TOF™ systems) or an available DataExplorer® software macro⁵.

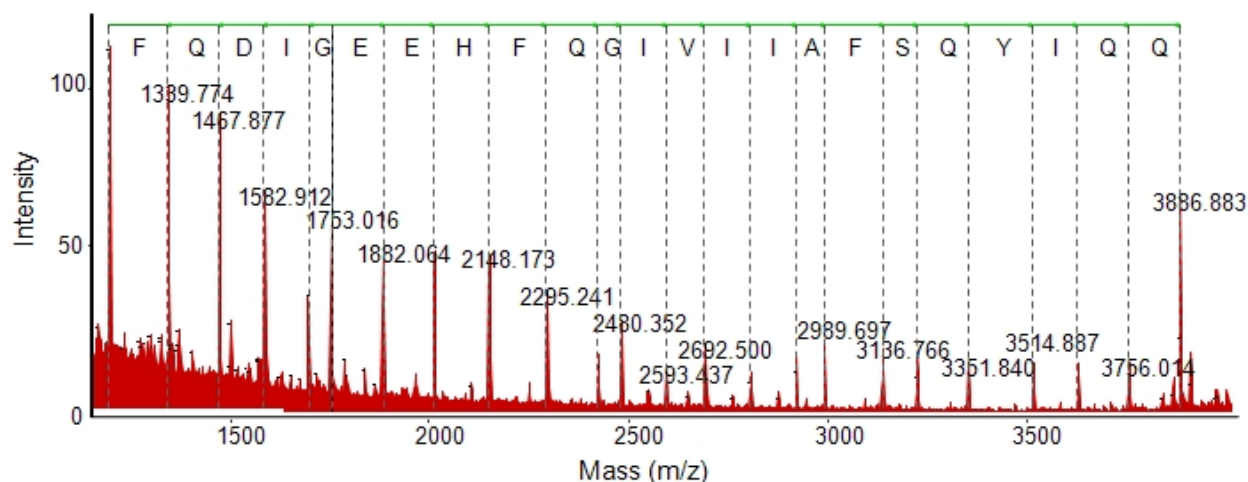


Figure 4. Generation of a De Novo Sequence Tag with the MASCOT Distiller Software obtained for BSA. The highlighted sequence can be used for a subsequent BLAST search.

Top-Down Sequencing for Protein ID using MS/MS Searching

De novo sequencing and BLAST searching is one way to use in-source decay data for protein identification. But it is also possible to do a MASCOT software MS/MS ion search, as it would be performed with any tandem MS/MS data. In a first step, it is necessary to introduce a new MS instrument within the MASCOT configuration of a local MASCOT software server installation. The instrument could be named MALDI ISD, and generates predominantly singly charged c-, y- and a- fragment ion signals as well as doubly charged z-ions. Figure 6 shows the instrument configuration page of MASCOT software version 2.2 with a new configured MALDI ISD instrument and the corresponding search settings for an MSMS ion search for in-source decay data.

Color Key: red = positive hit; green = borderline hit; black = negative result

Summary:

Sequences producing High-scoring Segment Pairs:

	High Score	Total Score
/:sptrembl Q757G0 Q757G0 AER053Cp.//:sptrembl Q757G0 Q757...	51	138
/:swiss P02769 ALBU_BOVIN Serum albumin precursor (Allerg...	137	137
/:swiss Q28522 ALBU_MACMU Serum albumin precursor (Fragme...	132	132
/:swiss P02770 ALBU_RAT Serum albumin precursor [Contains...	130	130
/:swiss P07724 ALBU_MOUSE Serum albumin precursor.//:sptr...	126	126

^ = swiss|P02769|ALBU_BOVIN Serum albumin precursor (Allergen Bos d 6) (BSA) //:trembl|M73993|BTALBUMIN_1 gene: "ALB"; product: "albumin"; Bos taurus albumin (ALB) mRNA, complete cds. //:gp|M73993|162648 albumin [Bos taurus] Length = 607

Total Score: 137

Score = 137 (74.0 bits)
Identities = 16/22 (72%), Positives = 22/22 (100%)

Query: 2 ZDLGEEHFZGLVLLAFSZVLZ2 23
+DLGEEHF+GLVL+AFS+YL++
Sbjct: 36 KDLGEEHFZGLVLLAFSQQYLQ 57

Figure 5. MS BLAST Results. After determination of the sequence tags using the Mascot Distiller, the sequence is submitted for MS BLAST searching. A confident hit for Albumin is obtained.

Mascot Configuration: Instruments

Ion series	Default	ESI QUAD TOF	MALDI TOF PSD	ESI TRAP	ESI QUAD	ESI FTICR	MALDI TOF	ESI 4SECTOR	FTMS ECD	ETD TRAP	MALDI TOF	MALDI TOF	MALDI TOF	Maldi 15D
1+	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2+	X	X		X	X	X	X	X	X	X	X	X	X	X
2+ (precursors>3+)														
immonium			X				X	X				X	X	
a	X	X					X	X				X	X	X
a*	X	X					X	X				X	X	
a0		X	X				X	X				X	X	
b	X	X	X	X	X	X	X	X				X	X	X
b*	X	X	X	X	X	X	X	X				X	X	X
b0		X	X	X	X	X	X	X				X	X	
c									X	X				X
w														
y	X	X	X	X	X	X	X	X	X	X	X	X	X	X
y*	X	X		X	X	X	X	X				X	X	
y0		X		X	X	X	X	X				X	X	
z								X						
zb							X	X				X	X	
ze							X	X				X	X	
y must be significant														
y must be highest score														
2+1									X	X				
d							X							
v							X							
w							X							
2+2									X	X				X
Minimum mass	700	700	700	700	700	700	700	700	700	700	700	700	700	
Max mass														
	Edit	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete	Delete
	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit	Edit

MASCOT Sequence Query

Your name: Email:

Search title:

Database:

Taxonomy:

Enzyme: Allow up to: missed cleavages

Fixed modifications: Variable modifications:

Quantitation:

Peptide tol. ±: ppm MS/MS tol. ±: Da

Peptide charge: Monoisotopic Average

Query:

Instrument:

Decoy: Report top: hits

Figure 6. MASCOT Software Instrument Configuration and MS/MS Ion Search Settings for ISD data. On the left side, new introduced MALDI ISD instrument within MASCOT configuration of instruments is shown. On the right side, the specific search settings for ISD fragment searches are shown.

Top-down sequencing spectra of intact proteins are measured in MS reflector mode with increased laser fluency to generate in-source decay fragments. The use of such MS reflector data for a MASCOT software MS/MS ion search requires some processing as well a change in the input *.txt file for MASCOT software. It is recommended to truncate and smooth the MS reflector in-source decay spectrum to get a clear series of fragment ion signals in average resolution, which simplifies the correct detection of high mass fragments. The easiest way to build the MASCOT software input txt-file for the MS/MS ion search starting from an MS reflector data is the MASCOT software toolbox macro within the DataExplorer® software (Figure 7). After manual selection of MS/MS ion search as well as manual input of the precursor mass the corresponding MS/MS peak list will be directly send to the MASCOT software server.

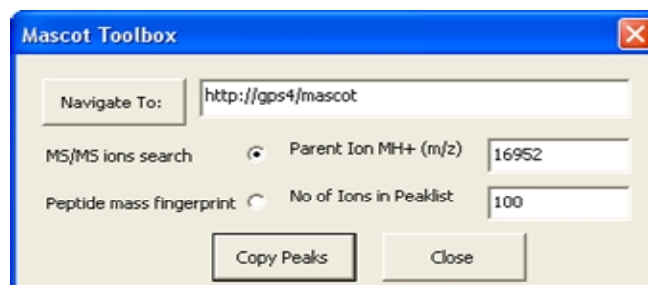


Figure 7. MASCOT Software Toolbox Window within the DataExplorer® Software. MS/MS ion search as well as parent ion MH+ mass has to be filled in to generate a MS/MS ion peak list. If the “big mascot” server is installed the intact protein mass can be used as parent ion. For standard MASCOT® software servers a “virtual precursor” mass can be used as parent ion, which is calculated by a selected any high-mass c-ion and add 1 Da.

A standard MASCOT software server has a maximum precursor mass of 16000 Da. There are two possible ways to overcome this limitation. For MASCOT software version 2.2 or higher, a so called “MASCOT Top-Down” server is available and doesn't have a precursor mass limit⁶. As an alternative you can use the “virtual precursor” approach, which uses a high mass C-ion as a virtual precursor for the MS/MS ion search. Due to the chemical structure of peptide c-ions it is necessary to add 1 Da to the mass of the c-ion to make it searchable as MH+ ion of the virtual precursor.

Following these processing steps, an identification of the protein based on the top-down sequencing data is possible. Figure 8 shows the result of MASCOT software MS/MS ion search using the in-source decay data of Myoglobin shown before in Figure 3. A clear identification was possible with a MASCOT software MS/MS ion score of 150.

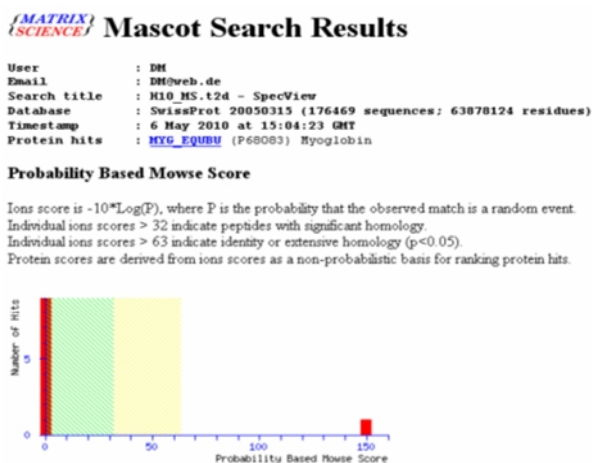


Figure 8: Results of the MASCOT Software MS/MS Ion Search for the ISD spectrum of Myoglobin. To obtain this result, the intact mass of the protein was determined from MS mode analysis, and then this measured mass was used in the MS/MS ion type search.

Conclusions

The ISD spectra of the proteins were acquired by performing MALDI TOF mass spectrometry on a dried droplet preparation with SINA or DAN matrix at elevated laser fluence. The laser fluence used depended on the spatial concentration distribution of the analyte and on the crystal shape, but was set about 25% higher than the threshold for ion formation in linear mode. The combined yield of fragmentation products in this experiment was in the range of 10%, normalized on the peak intensity of the protonated molecular species. As previously observed by Zubarev et al.⁷ a relatively low efficiency of ISD was generally observed in all of our experiments, thus limiting the applicability of ISD to sample quantities in the low picomole range.

In the case that in-source decay spectra of intact proteins should be used for the identification of the protein, it must be underlined that this will only be successful with pure proteins, due to the lack of a MS based precursor selection, which is used in classical tandem MS/MS.

MALDI top down sequencing is an important way to characterize an intact protein without the need of a time consuming enzymatic digest. The only limitation for this approach is the need of a clean sample of sufficient amount. MALDI in-source decay spectra can then be used to verify the correct sequence of a known intact protein, but it also allows the identification of the protein, either by de novo sequencing and blasting or by a MASCOT software MS/MS ion search.

Acknowledgements

We want to thank Juraj Lenco, Inst. Mol. Pathology, Faculty of Health Sciences, UD, Hradec Kralove.

References

1. Brown, Lennon, Anal. Chem. 1995, 67, 3990-3999. Brown, Feng, Reiber, Int. J. Mass Spectrom. Ion Processes 1997, 169/170, 1-18.
2. Spengler, Kirsch, Kaufmann, Jaeger, Rapid Commun. Mass Spectrom. 1992, 6, 105-108.
3. Lin, Campbell, Mueller, Wirth, Rapid Commun. Mass Spectrom. 2003, 17, 1809-1814.
4. Suckau, Reseman, Anal. Chem. 2003, 75, 5817-5824.
5. Please contact support@absciex to obtain the PSD macro.
6. "MASCOT Top Down" server - requires a new license, available on request from Matrixscience (London, UK).
7. Köcher, Engström, Zubarev, Anal. Chem. 2005, 77, 172-177.

For Research Use Only. Not for use in diagnostic procedures.

© 2010 AB SCIEX. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Publication number: 0680310-01

Headquarters

353 Hatch Drive | Foster City CA 94404 USA
 Phone 650-638-5800
www.absciex.com

International Locations

For our office locations please call the division headquarters or refer to our website at www.absciex.com/offices