Protein sequencing by ISD and PSD MALDI-TOF MS

D. Suckau and D.S. Cornett

Bruker Daltonik GmbH, Fahrenheitstrasse 4, D-28359 Bremen, Germany

Peptide and protein sequencing techniques are discussed which are available on a MALDI-TOF mass spectrometer equipped with an ion reflector. They allow the fast assignment of at least partial sequences that are sufficient to identify a protein uniquely by library searches and do not require additional enzymatic or chromatographic steps.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been used for molecular weight determinations of proteins and peptides for almost a decade [1]. More recently, also fragment ions, which are produced after leaving the ion source during the flight in the field-free region were used for peptide sequencing. These ions are called "post-source decay" or PSD ions [2] and can be observed in instruments equipped with an ion reflector. PSD spectra of peptides contain a wealth of different fragment ions: the sequence-specific a, b, y and in some cases also c-type ions [3]. In contrast to classical Edman sequencing, PSD-sequence determinations are possible even from complex mixtures like proteolytic digests if a gated electrostatic precursor ion selector is used.

While the verification of a known structure from sequence-specific ions is fast and straightforward in many cases, data interpretation of unknowns is more tedious and frequently only partial sequences are obtained. Therefore additional sources of information are desired. Contrasting other mass spectrometric sequencing techniques, PSD spectra contain abundant low molecular weight immonium or *i*-type ions $(NH_2 = CHR)^+$. *i*-type ions are not sequence-specific but allow to estimate the presence or absence of certain building blocks in a peptide sequence, because their mass m_i equals the residue mass $m_R - 27$ ($m_i = m_R - m_{CO} + m_H$). The use of internal fragment ions to analyze peptide sequences is discussed.

A recent extension of the MALDI ion source design dramatically increased the ion resolution and the mass accuracy. The technique was dubbed "time lag focusing" (TLF), "delayed extraction" (DE) or "pulsed ion extraction" (PIE) [4]. After desorption/ionisation, the ions are kept in the ion source under a field-free conditions for a short period of time, like 100 – 400 ns, before they are extracted with a high electrical field and accelerated towards the detector. During the brief time between ion desorption and the extraction pulse, peptide ions can undergo prompt fragmentation within the ion source (ISD, "in source decomposition"). All previous reports of this phenomenon utilise linear TOF instruments [5]. Because the fragmentation occurs in the source, product ions experience some of the effects of pulsed ion extraction, but due to the higher laser power required to induce fragmentation it is very difficult to obtain high resolution and mass accuracy for product ions in linear mode. We present here results from ISD measurements made using an instrument with ion reflector, which provides isotopic resolution for product ions up to several thousand Da. A wide range of compounds are found to produce ISD fragment ions up to the molecular weight of BSA (66 kDa). We show that ISD data can be used to provide rapid database identification of the precursor compound without additional chemistry required.

Cystine-cross linked peptides are known to suffer disulfide-bond cleavage under the high laser fluence conditions typical for PSD analyses [6]. Here, the influence of intrachain disulfide bonds on the efficiency to obtain sequence data by PSD and ISD measurements was evaluated.

Methods

The disulfide bonds of oxytocin were reduced by a 10-fold excess of dithioerythritol in 50 mM NH_4HCO_3 at 37 °C for 60 min followed by acidification with TFA. Peptide spectra were obtained in α-cyano-4-hydroxycinnamic acid on a REFLEX II MALDI-TOF mass spectrometer equipped with a gridless pulsed ion extraction source with visualization optics, precursor ion selector, two-stage gridless ion reflector and 1 GHz digitization board (Bruker Daltonik, Germany). 2 – 5 pmol of sample were used for PSD analysis and spectra were obtained at 20 kV accelerating voltage. The reflector voltage was reduced in 25% steps to obtain c.a. 10 segment spectra. The individual segments were calibrated with ACTH (18–39) [7] with a mass error of less than 0.3 Da and pasted under control of the data system. For ISD spectra, 5 – 10 pmol of proteins was prepared in sinapinic acid and the laser fluence adjusted to about 20% above threshold. Spectra were externally calibrated with a peptide mixture and 28.5 and 30 kV were the respective acceleration and reflector voltages.

Peptide sequencing by PSD

PSD spectra were obtained from native oxytocin with intact disulfide bond and after its reductive cleavage. The crosslinked peptide predominantly generated b_6 and y_3 ions and no sequence ions from within the crosslinked part were observed (Fig. 1a). In contrast, extensive *b* and *y* ion series
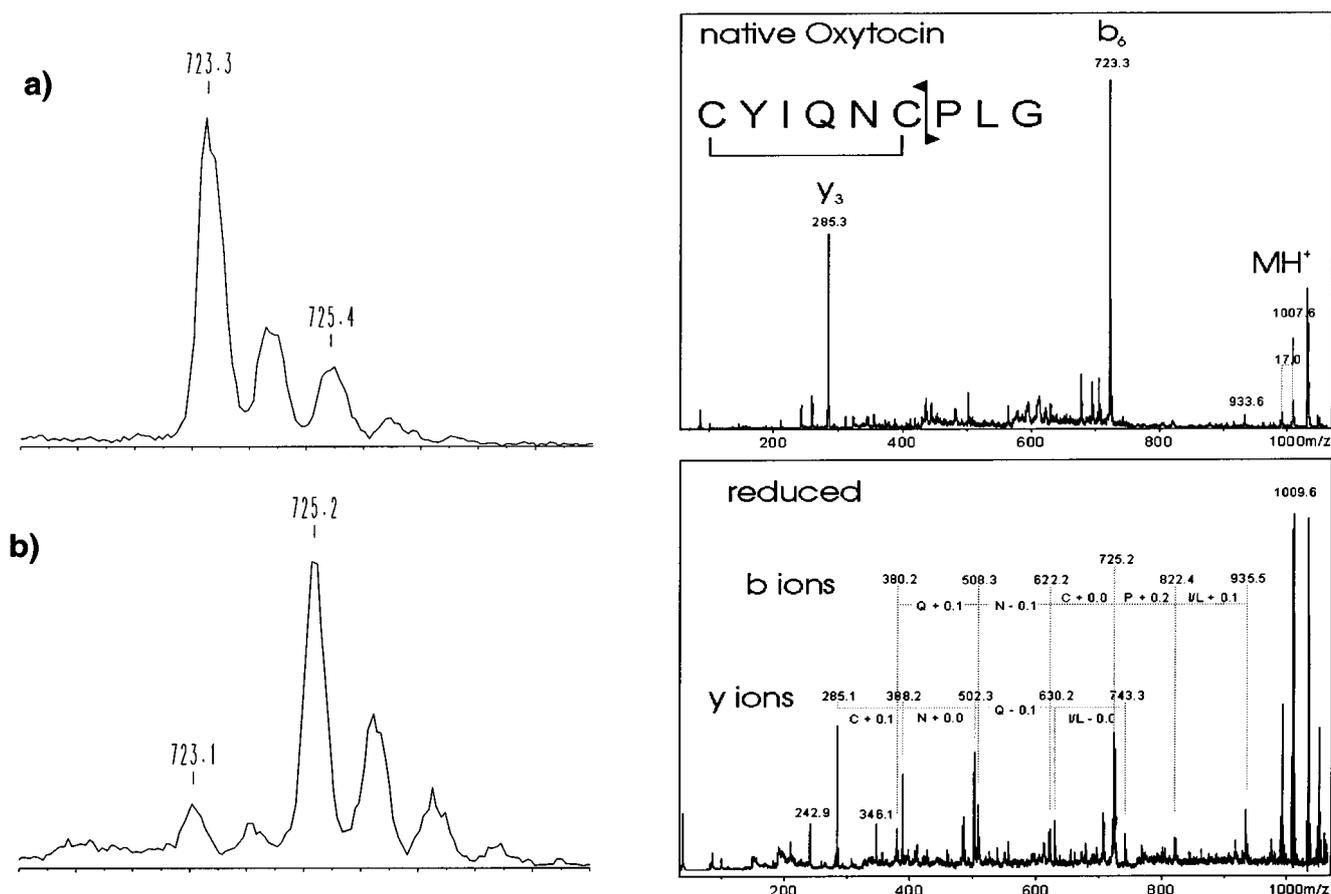
Figure 1. PSD spectra of oxytocin in the native cyteine-oxidized form (a) and after reduction (b).

were observed after reduction (Fig. 1b). The sequence ions in this spectrum leave only 1 C- and 2 N-terminal residues without assignments. However, these residues can be obtained by calculation of the mass differences between b_8 (m/z 935.5) or y_3 (m/z 285.1), respectively, and MH$^+$ (m/z 1009.6) which demonstrates that *de novo* sequencing of peptides by PSD is not always a linear procedure but may require additional analytical steps.

The fragment ions at low m/z values can provide such additional information. The low m/z part of the PSD spectrum of a pentapeptide containing pyridoethylated cystein (peC), all amino acid residues give rise to the corresponding *i*-type ions (Fig. 2). The peC signal at m/z 181 is accompanied by a strong vinylpyridinium ion at m/z 106. This feature allows the screening for cystein containing peptides in a protein digest by the mere PSD analysis of the mass segment containing the relevant ions. However, some fragment ions can be misassigned due to ambiguities. E.g., m/z 70 can be a prolin *i*-type ion or an internal fragment derived from the Arg *i*-type ion. In this spectrum, the wealth of Arg-derived fragments (60, 70, 112, 140 and 157; see Tab. Ia) rather safely allows to assign R as building block, but in fact, it does not allow to exclude Pro (m/z 70) or Ser (m/z 60) from the sequence. Based on the PSD spectra of 30 peptides, in table Ib the approximate probabilities that a certain *i*-type ion is found in a PSD spectrum if the respective amino acid residue is present in the peptide is compiled. While Gly is
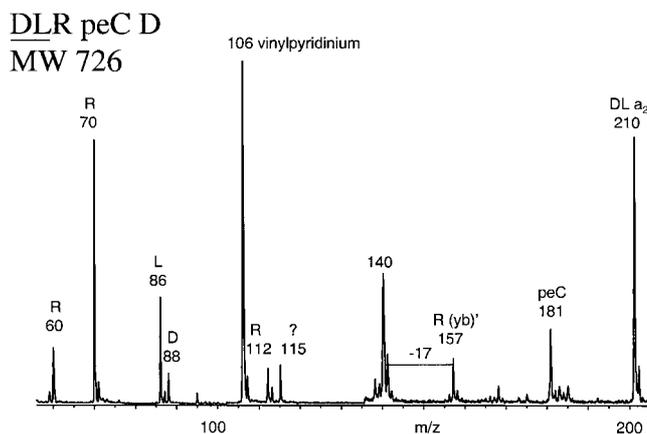

Figure 2. *i*-type ions and derived fragments in the PSD spectrum of a pentapeptide containing pyridoethylated cystein. *i*-type ions are labeled according to the single letter code of amino acids.

never detected, the residues in the >80% row of table I serve very well as an indication of the presence or absence of a certain amino acid residue in a peptide sequence. The *i*-type ions of cysteine/cystine could be a good indication of the reductive state, however they are usually very weak. Moreover, cysteic acid residues in oxidized insulin B chain

did not form *i*-ions but induced a strong fragmentation at the respective C-terminal peptide bonds (data not shown).

The mass range between c.a. *m/z* 200-400 contains the internal (*by*)´ ions, which provide additional information to verify partial sequences. These ions may comprise of di-, tri- or tetrameric partial peptide sequences which are helpful for *de novo* peptide sequencing, since they help to confirm unsafely assigned sequences (see Fig. 3). Their N- and C-termini resemble those of *y* and *b* ions, respectively.
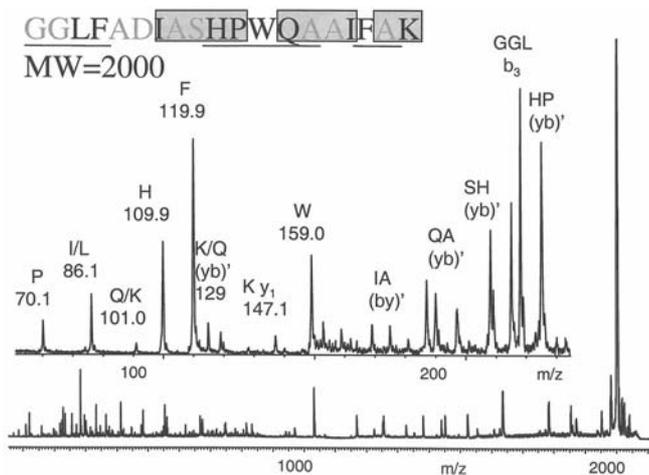

Figure 3. *i*-type and other internal fragment ions (*by*)´ from a PSD spectrum of a 19 mer peptide. Residues are labelled according to ion type of identification: *i*-type ions (black), (*by*)´ ions (boxed) and sequence-specific ions (underlined, data not shown). *I/L* and *K/Q* were not distinguished.

Table I. Identification of amino acids by low mass PSD fragments. a) fragment ion masses indicative for selected amino acid residues which are difficult to distinguish; bold: often strong signal, underlined: unique to this amino acid residue. b) probability to detect the presence of amino acid residues in a peptide by the corresponding *i*-type ions.

a)

R	129 **$\underline{112}$** 100 87 **70 60** 43 /$\underline{175}$ y_1 158, $\underline{157}$ 140 b_1/yb
K	129 **$\underline{101}$** $\underline{84}$ / **147** y_1
Q	129 **$\underline{101}$** $\underline{84}$
S	60 (weak and rare)
P	**70** 98

b)

Detection Probability	Amino acid residues
>80%	peC, F, H, I/L, K/Q, R, T, V, W, Y
c.a. 50%	E, M, N, P
c.a. 20%	A, D, S
0%	G

Protein sequencing by ISD

The ISD spectra of proteins contain primarily fragment ions of the *c*-type in accordance with previous reports [5]. The linear ISD spectrum of cytochrome *c* (Fig. 4) is essentially identical to the published cytochrome spectrum where >90% of the sequence can be identified by "reading" the *c*-ion ladder. The reflector mode ISD spectrum of cytochrome (data not shown) yields similar information. The major difference being that the reflector provides isotopic resolution for fragment ions up to several thousand kDa. The importance of this is illustrated in the ISD spectrum of myoglobin (Fig. 5). In this example other ion series are clearly apparent at much lower intensity relative to the *c*-ion series. The expanded inset shows the c_{31} ion clearly resolved from a neighboring peak with it's isotope distribution. This illustrates the increased specificity of ion assignments and therefore the higher quality of sequence ladders from reflector ISD measurements. In our hands, cytochrome *c* appears unique among all compounds evaluated in that nearly the full sequence is represented by the *c*-ion series. We usually obtain a *c*-ion series representing only c.a. 10% of the sequence.

A special case is RNAse. The reflector ISD spectrum of bovine RNAse (Fig. 6) shows a clear *c*-type ion ladder from Arg^{10} to Tyr^{25}. The lack of sequence ladder information beyond Tyr^{25} is due to Cys^{26} which is crosslinked to Cys^{84} in the native molecule. This result is similar to the PSD spectrum of oxytocin and suggests that sequence information by these MS/MS techniques is not directly available from crosslinked sequence stretches of peptides and proteins.

BSA (66 kDa) is the largest compound that was examined. ISD fragments have been observed as high as 7 – 8 kDa representing roughly 10% of the sequence. Figure 7 shows an expansion of the high resolution region in the ISD spectrum. As with RNAse A, a sequence library search (BLASTP) for the sequence motif in figure 7 allowed the identification of bovine BSA [8]. ISD is therefore well suited for partial (internal) sequencing and possibly useful in proteome research for the identification of proteins in 2 D gels or the identification of point mutations like in aberrant haemoglobins.
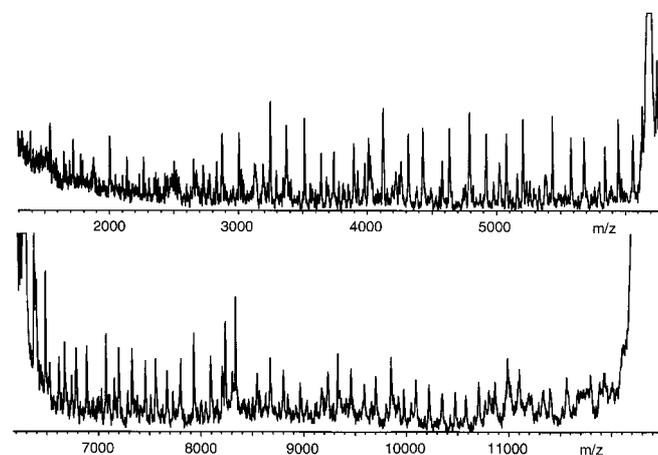

Figure 4. Linear ISD spectrum of equine cytochrome *c*. Sum of 75 shots.

Figure 5. Reflector ISD spectrum of equine myoglobin. Sum of 75 shots.

Figure 6. Reflector ISD spectrum of bovine pancreatic RNAse. Sum of 75 shots. Sequence ladder is assigned from Arg^{10} to Tyr^{25}; average residue mass error = 0.04 Da.

Figure 7. Reflector ISD spectrum of BSA. Sum of 75 shots. c ion ladder is assigned; average residue mass error = 0.01 Da.

Conclusions

In addition to the common MS/MS sequence-specific a, b and y ions, PSD spectra contain a wealth of internal fragment ions like i-type ions and $(by)'$ ions which help to analyze the sometimes complex data. i-type ions allow to estimate the presence or absence of certain residues in a peptide and $(by)'$ ions allow to verify suggested sequence patterns by the masses of dipeptide fragments. Although these ions complicate PSD spectra, they are valuable features of these spectra in *de novo* peptide sequencing.

All proteins investigated by reflector ISD up to the size of BSA (66 kDa) were found to undergo fragmentation in a time period of less than 200 ns. The additional energy focusing provided in reflector ISD measurements resulted in mass resolutions in excess of 10 000 for fragment ions as large as m/z 5000 and a residue mass accuracy of better than 0.05 Da, which increased the specificity of peaks as compared to linear ISD measurements [9]. Significant amounts of sequence information are routinely obtained from the c ion sequence ladders which can be used in library searches like BLASTP providing a fast method of protein identification.

Spectra of reduced and cystine containing peptides indicate that intrachain disulfide bonds prevent the formation of sequence-specific fragment ions under PSD conditions. Even under ISD conditions, which tend to increase the fragmentation efficiency, no intrachain disulfide bond fission could be found [9], which would allow sequence determinations from the disulfide-crosslinked core region of proteins.

References

1. Andersen, J.S.; Svensson, B.; Roepstorff, P. *Nature Biotech.* **1996**, *14*, 449-457.
2. Spengler, B.; Kirsch, D.; Kaufmann, R. *J. Phys. Chem.* **1992**, *96*, 9678.
3. Papayannopoulos, I.A. *Mass Spectrom. Rev.* **1995**, *14*, 49-73.
4. Brown, R.S.; Lennon, J.J. *Anal. Chem.* **1995**, *67*, 1998-2003.
5. Brown, R.S.; Lennon, J.J. *Anal. Chem.* **1995**, *67*, 3990-3999.
6. Jones, M.D.; Patterson, S.D.; Lu, H.S. *Anal. Chem. 70*, 136-143.
7. Rose, J.C.; Wu, W.; Martin, S.A. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 822-835.
8. Cornett, D.S. Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics, Palm Springs, CA, June 1–5, 1997; p 316.
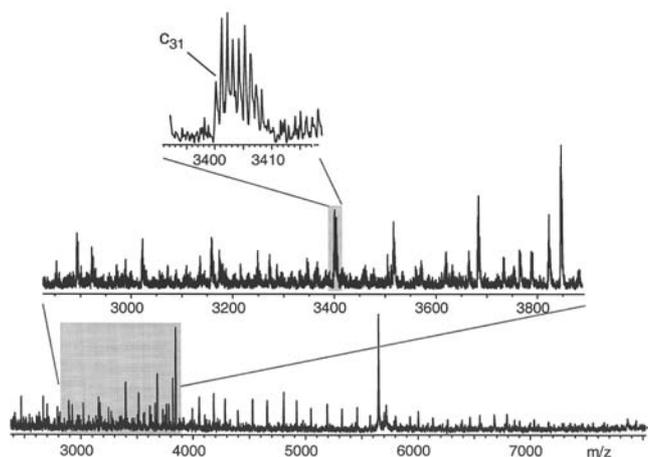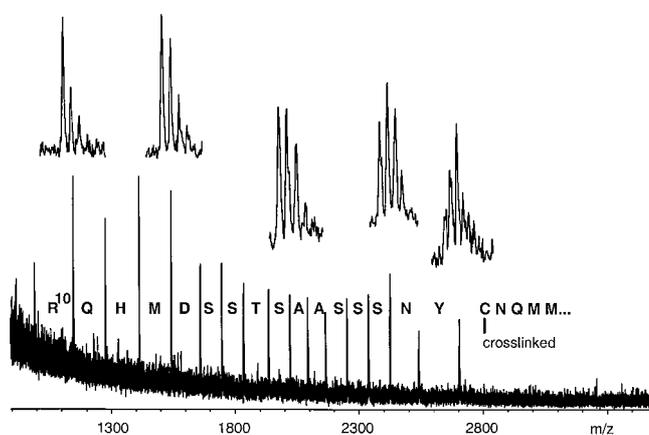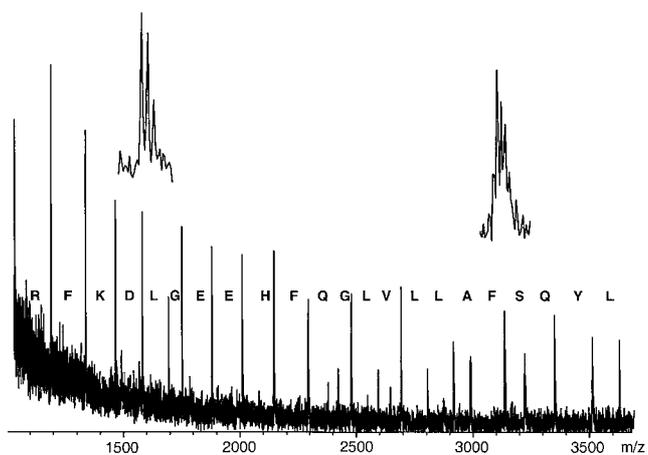9. Lennon, J.J.; Walsh, K.A. *Protein Sci.* **1997**, *6*, 2446-2453.