Sequence Dependent Fragmentation of Peptides Generated by MALDI Quadrupole Time-of-Flight (MALDI Q-TOF) Mass Spectrometry and its Implications for Protein Identification

Andreas Wattenberg, Andrew J. Organ, and Klaus Schneider
Bioanalytical Sciences, CASS, GlaxoSmithKline, Harlow, Essex, United Kingdom

Richard Tyldesley, Robert Bordoli, and Robert H. Bateman
Micromass UK Ltd., Floats Road, Manchester, United Kingdom

A study has been undertaken to evaluate the usefulness of MALDI Q-TOF data for protein identification. The comparison of MS data of protein digests obtained on a conventional MALDI TOF instrument to the MS data from the MALDI Q-TOF reveal peptide patterns with similar intensity ratios. However, comparison of MS/MS Q-TOF data produced by nanoelectrospray versus MALDI reveals striking differences. Peptide fragment ions obtained from doubly charged precursors produced by nanoelectrospray are mainly y-type ions with some b-ions in the lower mass range. In contrast, peptide fragment ions produced from the singly charged ions originating from the MALDI source are a mixture of y-, b- and a-ions accompanied by ions resulting from neutral loss of ammonia or water. The ratio and intensity of these fragment ions is found to be strongly sequence dependent for MALDI generated ions. The singly charged peptides generated by MALDI show a preferential cleavage of the C-terminal bond of acidic residues aspartic and glutamic acid and the N-terminal bond of proline. This preferential cleavage can be explained by the mobile proton model and is present in peptides that contain both arginine and an acidic amino acid. The MALDI Q-TOF MS/MS data of 24 out of 26 proteolytic peptides produced by trypsin or Asp-N digestions were successfully used for protein identification via database searching, thus indicating the general usefulness of the data for protein identification. De novo sequencing using a mixture of $^{16}O/^{18}O$ water during digestion has been explored and de novo sequences for a number of peptides have been obtained. (J Am Soc Mass Spectrom 2002, 13, 772–783) © 2002 American Society for Mass Spectrometry

The development of protein identification by mass spectrometry has fuelled the rise of proteomics as a new field of research. The high sensitivity of modern mass spectrometers and the high selectivity of protein database searches has enabled the identification of proteins in the sub-picomole level on a larger scale. Two approaches for protein identification have been developed: (i) peptide mass fingerprinting and (ii) fragment ion mass searches. In the first approach, peptide masses of proteolytic protein digests are determined by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS). Measured peptide masses are matched to calculated masses of proteolytic peptides from proteins [1]. This requires only small amounts of protein, provides a relatively fast identification and can easily be automated. In the second approach, fragment ion masses of one or more proteolytic peptides from a protein digest are determined, usually by electrospray ionization tandem mass spectrometry (ESI MS/MS). Database searches can then be performed with different algorithms; either by using partly interpreted data or by submitting the list of fragment ions taken from the MS/MS spectra without any data interpretation [2, 3]. However, both approaches suffer from some limitations. In peptide mass fingerprinting, results obtained are not always unambiguous and the method does not cope well with protein mixtures. On the other hand,

Published online June 11, 2002

Dedicated to the memory of Bob Bordoli

Address reprint requests to: Dr. Klaus Schneider, GlaxoSmithKline, Cold Harbour Road, New Frontiers Science Park (North), Harlow, Essex CM19 5AD, United Kingdom. E-mail: Klaus_Schneider-1@gsk.com

1044-0305/02/$20.00
PII S1044-0305(02)00414-2

Received December 3, 2001
Revised April 11, 2002
Accepted April 11, 2002

fragment ion-based searches are comparably slow, requiring either the powerful, but tedious manual process of nanoelectrospray-MS or the more easily automated, but relatively low throughput method of nanoscale liquid chromatography (nanoLC) ESI MS/MS. The two approaches for protein identification are used either independently, or in a combined fashion exploiting the strengths of both approaches [4]. In the combination of both, the high sensitivity and high mass accuracy obtainable by MALDI-TOF MS is employed for a first high throughput analysis utilizing only a part of the sample. If the protein identification in this first step is not successful, or many major peptide peaks can not be attributed to the protein identified, then the remainder of the sample is subjected to nanoLC ESI MS/MS, which results in high confidence identification even from expressed sequence tag (EST) databases with fragment ion data from just one peptide.

A downside of this two-tiered process lies in the requirement for an initial division of the sample. For this reason, the use of only fragment ion-based searches using nanoLC MS/MS has become increasingly popular. The downside is that it lacks the throughput to cope with rising demands for fast protein identification. Thus, it would be highly desirable to combine data acquisition in MS and MS/MS mode from one sample preparation maintaining the high speed of MALDI-TOF MS.

The recent introduction of a MALDI ion source coupled to a mass spectrometer with a tandem quadrupole-time-of-flight analyzer has provided a means to obtain reliable MS and MS/MS data from the same sample on the same instrument. This has been explored for protein identification by Loboda et al. [5], Shevchenko et al. [6, 7], Krutchinsky et al. [8] and Baldwin et al. [9]. Loboda et al. showed that matrix assisted laser desorption and ionisation quadrupole-time of flight (MALDI Q-TOF) mass spectrometry can be used to successfully identify proteins from in-gel digests using database search engines. Shevchenko et al. demonstrated that MALDI Q-TOF data can be used for BLAST searching in order to identify proteins from organisms for which the genome has not yet been sequenced. Baldwin et al. demonstrated that MALDI QTOF data is suited for protein identification as well as detailed protein structural elucidation including posttranslational modifications such as phosphorylation. Krutchinsky et al. used MALDI and nanoelectrospray ionization in conjunction with a Q-TOF analyzer and demonstrated the complementary use of two different ionization techniques to improve the success rate of protein identification.

Here, we report observations made in the evaluation of the MALDI Q-TOF data in comparison to data from conventional MALDI-TOF and from nanoelectrospray Q-TOF. The purpose of this paper is to investigate the fragmentation behavior of peptides in the new MALDI Q-TOF set-up and to evaluate the usefulness of the MALDI Q-TOF MS and MS/MS data for protein identification from databases as well as for de novo sequencing.

Experimental

MALDI-Q-TOF Set-up

An experimental MALDI source has been fitted to a Q-TOF (Micromass, Manchester, UK) instrument as described in reference [10]. In brief, the standard electrospray source housing was removed and a new housing fitted in its place. This housing allowed a target plate mounted on a probe to be introduced into the housing via a vacuum lock. It also included a window to provide direct line of sight access to the sample plate. A nitrogen ultraviolet (UV) laser (Laser Science, Inc. VSL-337I, 337 nm) was mounted such as to allow reflection of the UV light off an adjustable mirror and through the window onto the target. The mirror was adjusted to optimize the positioning of the laser beam on the surface. A lens was included in the optical path to focus the laser beam on the target plate surface to a spot approximately 300 μm in diameter. The housing included provision for introducing gas via a capillary line and adjustable needle valve. Air was introduced into the ion source to give a pressure of 0.1 mbar. The sample was introduced and the laser was operated at 20 Hz under full power (300 μJ/pulse). Spectra were accumulated until a satisfactory signal/noise ratio had been obtained.

Digestion

The digestion of the proteins were carried out using trypsin (Promega, Madison, WI, USA) and Asp-N (Roche Molecular Biochemicals, Mannheim, Germany) as follows: for the trypsin digestion a solution containing 100mM ammonium bicarbonate buffer (pH 8), 5mM $CaCl_2$ and 1μM trypsin was prepared. 1 nmol of protein was dissolved in 10 μL of this buffer and incubated for 2h at 37 °C. After drying down, the samples were resuspended in 10 μL 0.1% aqueous TFA and cleaned with ZipTip®s (Millipore, Bedford, MA, USA) according to the protocol supplied by the manufacturer. The samples were diluted 1:200 before mixing the peptide solution with equal volumes of saturated DHB matrix solution and spotting 2 μL of the mixture onto the MALDI target plate. For digestion with AspN, the proteins were digested in 50 mM ammonium bicarbonate with an enzyme/substrate ratio of 1:50 at 37 °C for 4 hours. The digest solution were then evaporated to dryness and reconstituted in 0.1% TFA to a concentration of 1 μM.

The digestions for isotopical labeling of peptides were performed as described, except for the use of a 1:1 mixture of $H_2^{16}O$ and $H_2^{18}O$ (Fluka, Chemica, Buchs, Switzerland) as solvent.

Nanoelectrospray MS/MS was obtained on a Q-TOF instrument (Micromass, Manchester, UK) fitted with a nanoelectrospray source. A few microliter of the protein digest in H_2O/methanol containing 5% formic acid at a concentration of 500 nM were loaded in a nanoelectro-

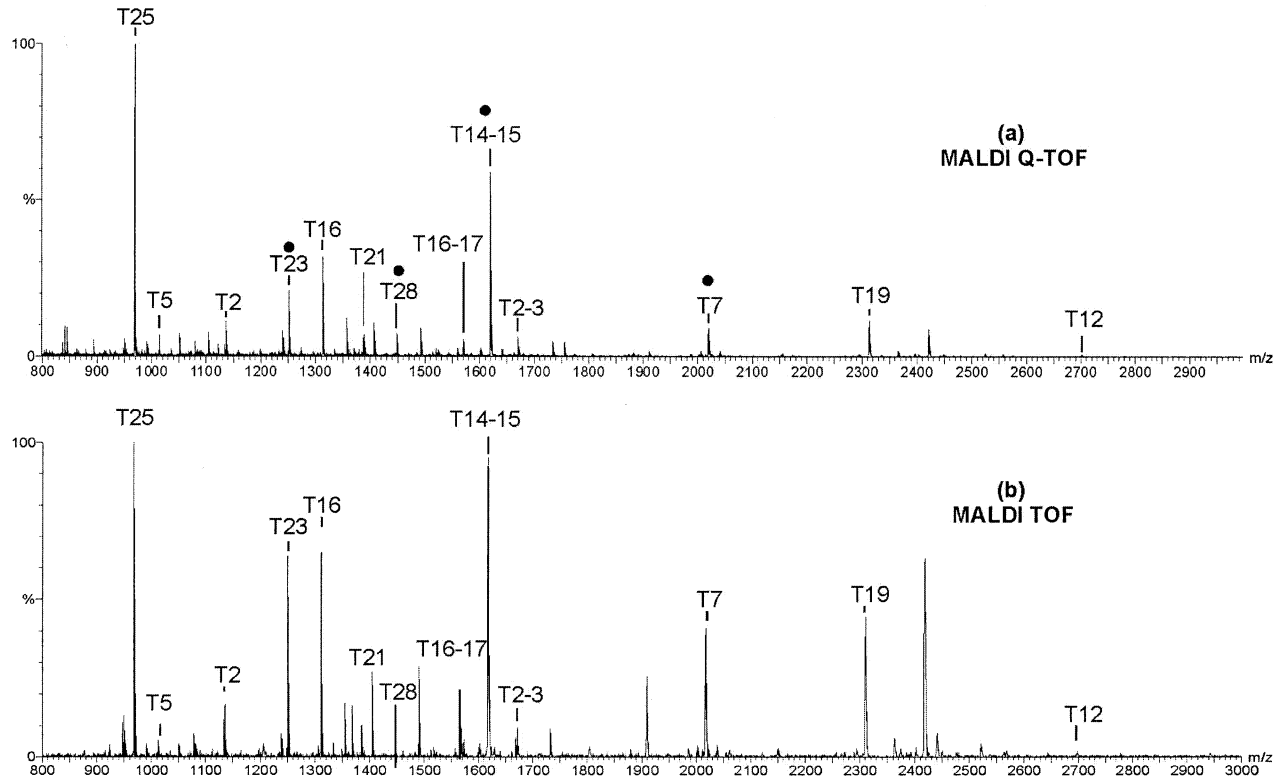

Figure 1. Comparison of MALDI MS spectra acquired using (a) the prototype MALDI Q-TOF and (b) a conventional MALDI TOF mass spectrometer. In both cases 1 pmol of a tryptic digest of alcohol dehydrogenase was spotted onto the target. The peaks arising from the tryptic digest are labeled according to the sequence position in the protein. The MS/MS spectra of the peaks labeled with the large dot symbol are shown in Figure 2.

spray needle (Protana, Odense, Denmark). MS/MS data was acquired over a time of about one to three minutes.

Results

The aim of the work presented in this paper was to compare the data quality obtained using a MALDI Q-TOF to that obtained by the established techniques of MALDI TOF and nanoelectrospray Q-TOF and thus to assess the usefulness of the MALDI Q-TOF for protein identification.

In a first step, the performance of the MALDI Q-TOF in the MS mode was compared to a conventional MALDI TOF. For this, equal amounts of a tryptic digest of alcohol dehydrogenase (ADH) were spotted and analyzed. A comparison of mass spectra acquired with the new MALDI Q-TOF and a conventional MALDI TOF is shown in Figure 1. Both mass analyzers give rise to very similar mass spectra in terms of tryptic peptides observed with a higher signal-to-noise ratio in the MALDI TOF spectrum indicating a higher sensitivity of the conventional MALDI TOF set-up. The extracted peptide masses were searched with the Profound (Proteometrics, New York) search engine against a non-redundant protein database and result in an unambiguous identification of ADH with a sequence coverage of 50% (14 peptides identified). In both cases, the mass error after internal calibration was generally below 20 ppm.

The MALDI Q-TOF allows acquiring MS/MS data on the same sample preparation that was used for MALDI peptide mass fingerprinting. MALDI MS/MS spectra of some singly charged ions of tryptic peptides identified in Figure 1a (selected parent ions are labeled with black circles) are shown in Figures 2a–d (upper spectra). For these spectra the collision energy was between 60–115 eV (see Table 1). The MS/MS spectra of the multiply charged ions of the same peptides analyzed by nanoelectrospray Q-TOF are also shown (lower spectra). For these spectra the collision voltage was set to 20–32 volts, and the collision energy was between 40–72 eV (see Table 1). The spectra are ordered according to the molecular weight of the parent ions. Peptide fragment ions obtained from doubly-charged precursors produced by nanoelectrospray are mainly y-type ions with some b-ions in the lower mass range. In contrast, peptide fragment ions produced from the singly charged ions originating from the MALDI source are a mixture of y-, b- and a-ions accompanied by ions resulting from neutral loss of ammonia or water. In direct comparison there are some other differences noticeable, especially the intensity of the fragment ions formed is very different between the two desorption techniques. In Figure 2a, in the case of nanoelectros-

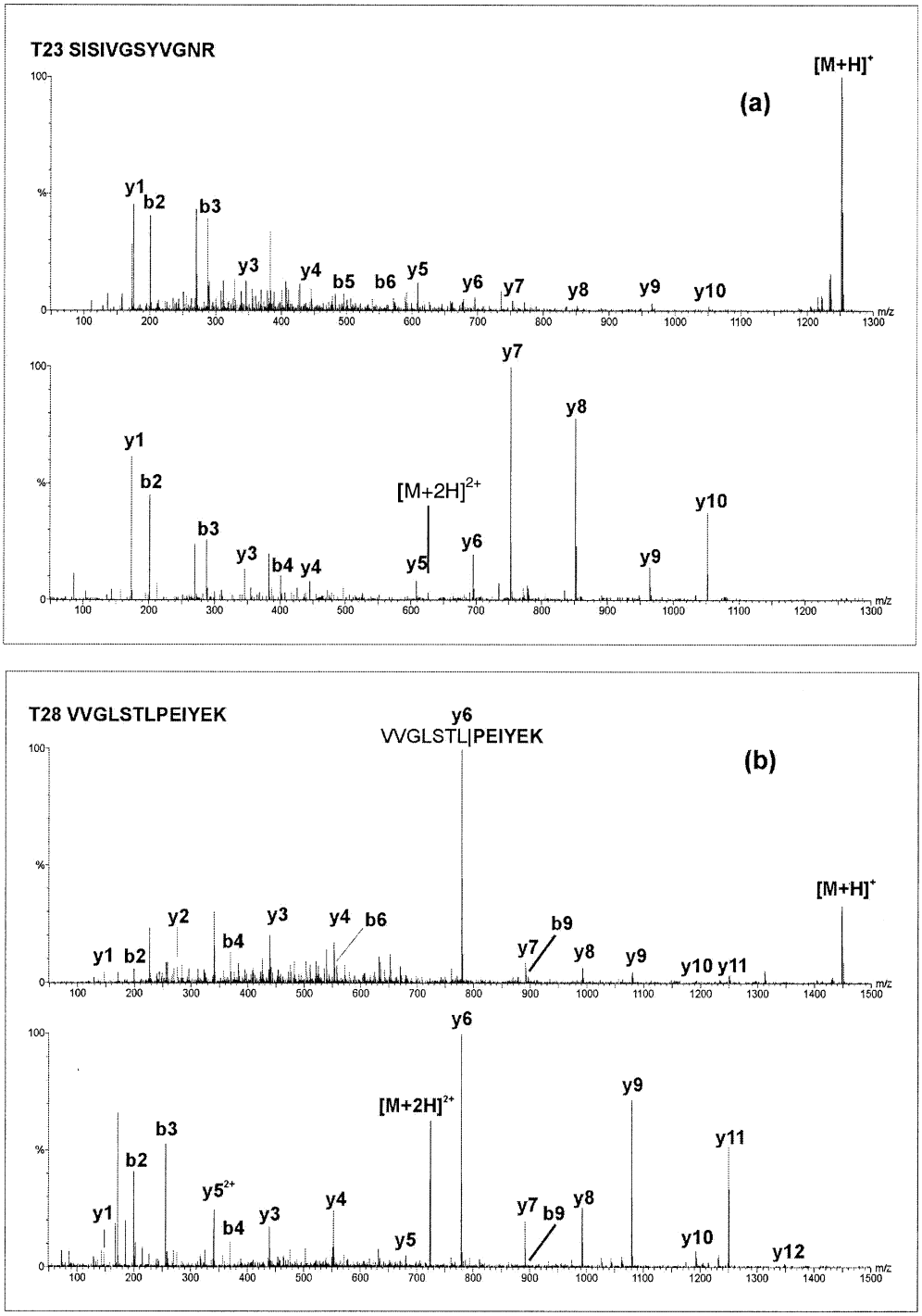

Figure 2. Comparison of the MS/MS spectra of four peptides of a tryptic digest of alcohol dehydrogenase. In each case, the upper spectrum shows the data acquired from the singly charged ion of a peptide using a MALDI Q-TOF, the lower spectrum from the doubly charged ion from of an tryptic digest of alcohol dehydrogenase analyzed using a nanoelectrospray Q-TOF. The fragment ion peaks are labeled with their ion series label. Peaks labeled with an inverted filled triangle arise from neutral losses of ammonia or water. Peaks arising from preferential cleavage are annotated with the peptide sequence. A vertical line indicates the cleavage site, the bold font the peptide observed.

pray, there are very intense y-ions above the parent ion mass-to-charge ratio, but in contrast there are only very small signals for the corresponding y_7-y_{10} ions in the MALDI spectrum. In Figure 2b, the y_6 ion is the most intense fragment peak in both cases. This fragment arises from a cleavage of the N-terminal proline (P) bond. Again, the high mass y-ions (y_7-y_{11}) are more intense in the nanoelectrospray spectrum. Figure 2c

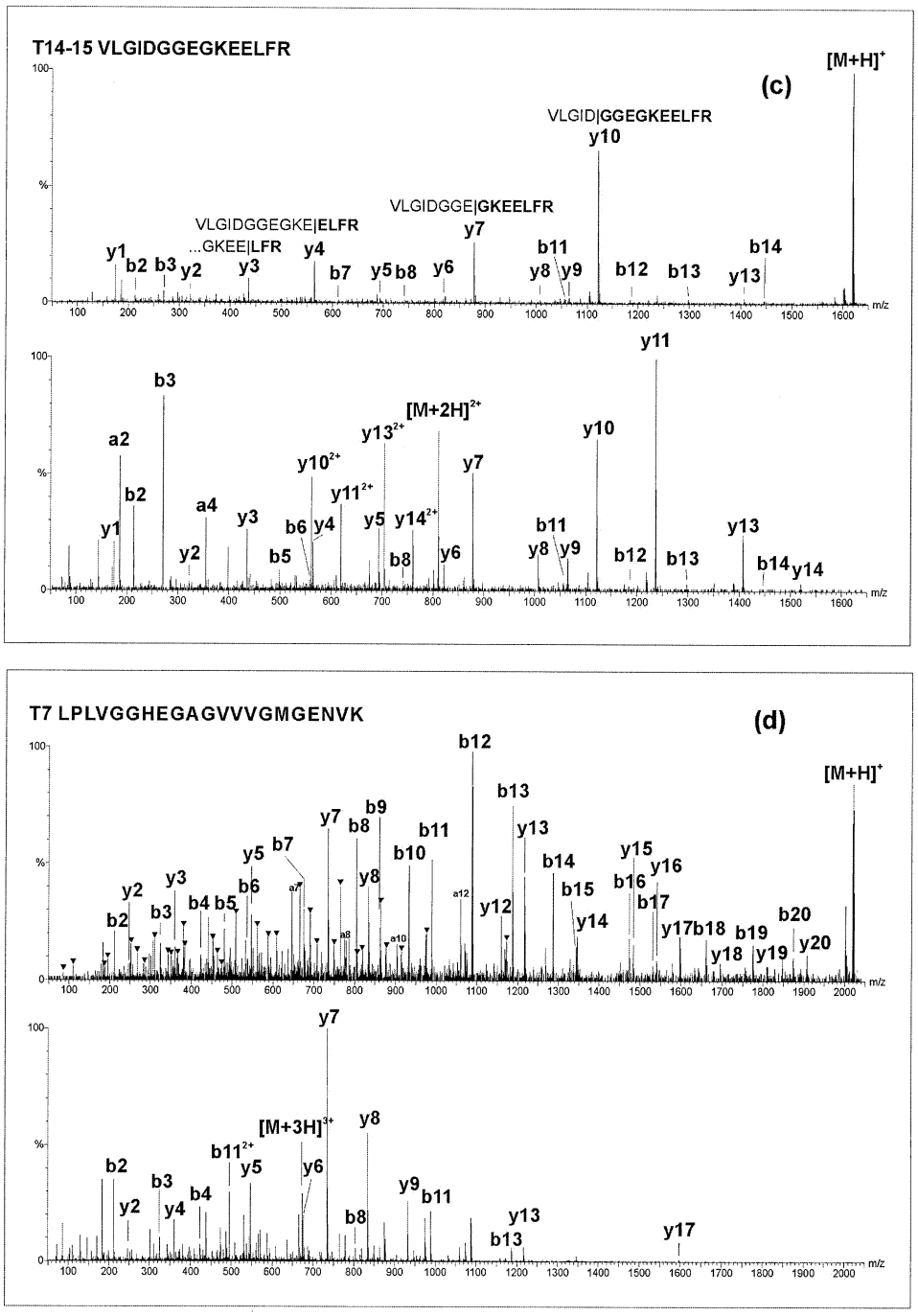

Figure 2. Continued

shows that the fragment ion intensities vary significantly between the two desorption methods. Whereas nanoelectrospray gives rise to fragment ions of overall similar intensity, the MALDI spectrum clearly shows that certain fragments are very dominant (y_3, y_4, y_7 and y_{10}). These fragment ions arise from cleavages of the C-terminal bond of acidic amino acids glutamic acid (D) and aspartic acid (E). The fragmentation of the largest tryptic peptide in Figure 2d reveals a very complex MALDI MS/MS spectrum. y- and b-ions are very intense throughout the mass range and peaks arising from neutral loss of ammonia (−17) and water (−18) show up. Also, a series of a-ion peaks can be detected. The nanoelectrospray spectrum shows virtually no fragment ions above *m/z* 1300, whereas the lower molecular weight region shows b- and y-ions.

Tryptic peptides from another protein were also studied. In Figure 3a, the MALDI Q-TOF spectrum of a

Table 1. Results of database searches using the MS/MS spectra of the peptides resulting from digestion by trypsin or Asp-N proteases

Protein	Digest	Mass	Protein-Lynx	MASCOT	Collision voltage
ADH	Asp-N	835.4	identified	identified	37
ADH	Asp-N	1087.5		identified[1]	49
ADH	Asp-N	1623.05	no id	no id	85
ADH	Asp-N	2495.17	identified[2]	identified[2]	115
ADH	Trypsin	968.5	identified	identified	45
ADH	Trypsin	1251.7	identified	identified	60
ADH	Trypsin	1312.7	identified	identified	60
ADH	Trypsin	1447.8	identified	identified	62
ADH	Trypsin	1618.9	identified	identified	77
ADH	Trypsin	1754.9	no id	no id	87
ADH	Trypsin	2020.3	identified	identified	93
ADH	Trypsin	2313.5	identified	identified	115
Myoglobin	Asp-N	1423.8	identified	identified	67
Myoglobin	Asp-N	1440	identified	identified	67
Myoglobin	Asp-N	1625		identified[1]	82
Myoglobin	Asp-N	1857.9	identified	identified	85
Myoglobin	Asp-N	1895.56	identified	identified	85
Myoglobin	Trypsin	1272.2	identified	identified	60
Myoglobin	Trypsin	1378.7	identified	identified	67
Myoglobin	Trypsin	1502	identified	identified	72
Myoglobin	Trypsin	1507		identified	72
Myoglobin	Trypsin	1607	identified	identified	75
Myoglobin	Trypsin	1815.9	identified	identified	82
Myoglobin	Trypsin	1854	identified	identified	82
Myoglobin	Trypsin	1982	identified	identified	90
Myoglobin	Trypsin	2110.2	identified[3]	identified[3]	105

For the database search the entire mass spectra were submitted. Proteins digested are alcohol dehydrogenase (ADH), and myoglobin. The entry "identified" means that the database search returned ADH1_YEAST or MYG_HORSE, respectively.
[1]Identified by MASCOT search with enzyme set to none
[2]Identified with N-terminal acetylation and H → Y mutation
[3]Identified with missed cleavages set to 2

digest peptide from horse heart myoglobin is shown. The most obvious feature in this spectrum is the pronounced loss of water from the parent and the b_8 and b_3 ions. A preferential fragmentation leading to the y_5 and y_7 ions can be observed. Still, these ions are not as dominant as, for example, in Figure 1c, as some b-ions are as intense as these y-ions indicating a much lesser degree of preference. In Figure 3b, the preferential formation of the y_4 and y_{11} ions is much more obvious, as these ions are the most intense fragment ions in the spectrum.

In order to further study the fragmentation behavior of peptides in the MALDI Q-TOF, the same proteins as above were digested using a different enzyme, in this case Asp-N, and analyzed by both MALDI Q-TOF and nanoelectrospray Q-TOF. Whereas peptides generated by trypsin have basic groups at their C-terminus, Asp-N cleaves N-terminally to aspartic acid, leaving a negatively charged side-chain at the N-terminus. The MALDI MS/MS spectra of peptides of the Asp-N digest of myoglobin are shown in Figure 4a and b, again ordered in ascending parent ion mass. In Figure 4a, the MALDI Q-TOF spectrum is dominated by b-ions, whereas the y-ion series generally is of lower intensity. In Figure 4b, the largest peptide fragmented in this study shows a much more complex fragmentation behavior with peaks arising from neutral loss in the lower mass region.

When applying the MALDI Q-TOF technology in proteomics research, one of the crucial points is whether the data allows for an unambiguous protein identification through database searching. To test the usefulness of MALDI MS/MS data for protein identification purposes, two proteins (alcohol dehydrogenase and myoglobin) were digested with trypsin and Asp-N and MS/MS spectra were recorded from the resulting peptides with the MALDI Q-TOF MS/MS. This resulted in 26 MS/MS spectra, including those shown in Figure 2, 3 and 4, which were submitted to the ProteinLynx program (Micromass, Manchester, UK) and searched against a SWISSPROT database. In parallel, the spectra were searched against a non-redundant protein database using the MASCOT search engine (Matrix Science, London, UK). These two search engines were chosen because they accept uninterpreted mass spectra, minimizing the user interference and thus providing unbiased results. Through this procedure, 21 peptides were identified (Table 1) in a first pass. One more peptide was identified after taking into account an N-terminal acetylation and a mutated amino acid. Two other peptides were products of unspecific cleavages and were identified by searching the database using the

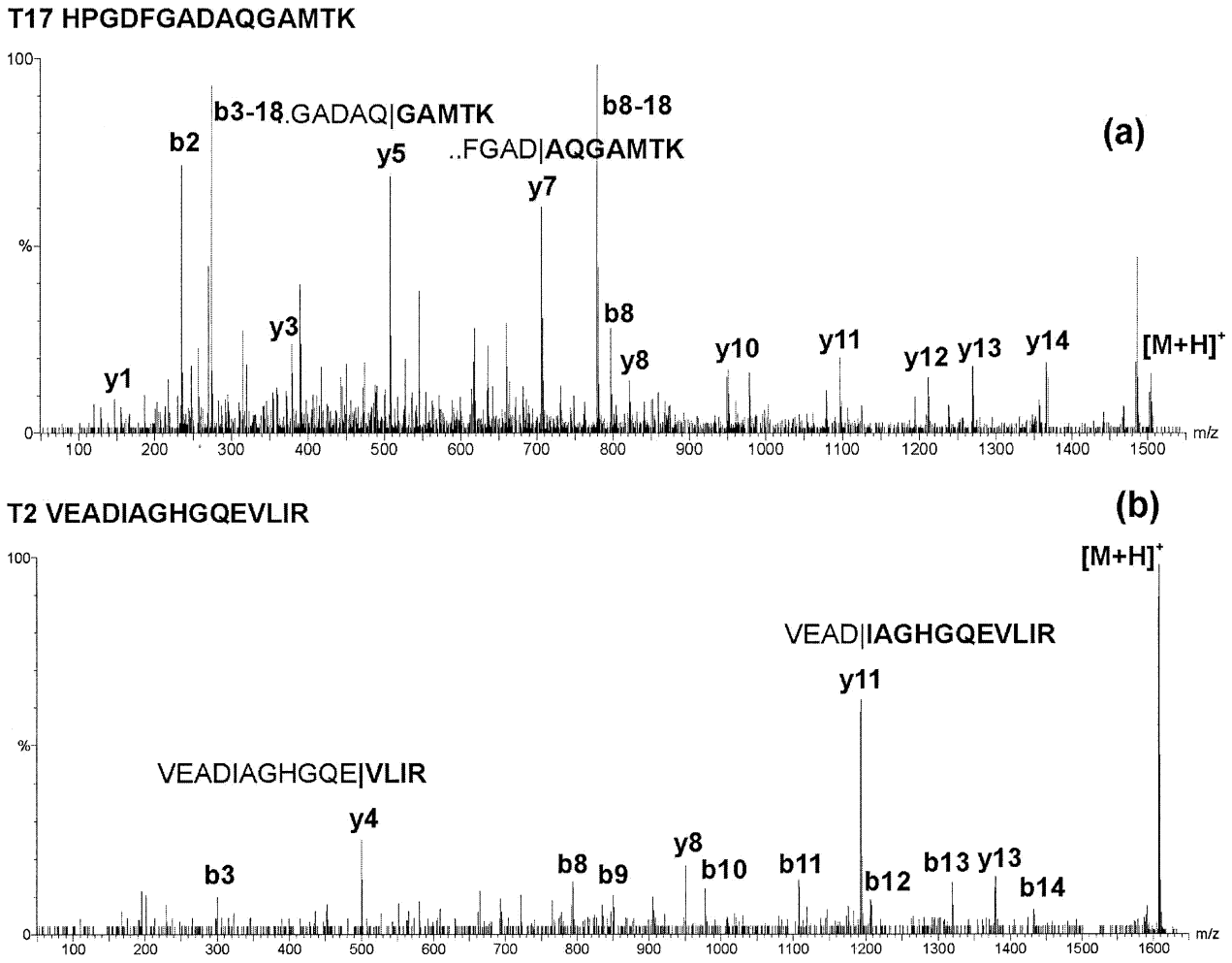

Figure 3. MALDI Q-TOF MS/MS spectra of tryptic digest peptides from horse heart myoglobin.

MASCOT search engine with the enzyme specificity set to "none." One other peptide at *m/z* 2110.2 was identified after allowing for more than one missed cleavage site. Two peptides that gave rise to poor fragment spectra remained unidentified.

Even though deriving sequence information directly from the MS/MS spectrum (de novo sequencing) is possible in some cases, the complex fragmentation often prevents an unambiguous sequence determination. An example of this is the spectrum in Figure 2d that shows y-, b- and a-ion series. A possible problem for de novo sequencing here is that the ions b_8 to b_{12} give rise to very intense peaks, that can easily be mistaken as a part of a y-ion series. A possible readout consists of the ions that are annotated as y_{12}, b_{12}, b_{10} and b_8, which would give rise to the sequence tag ERK. As shown in Figure 2d, that actual sequence is not part of the peptide fragmented nor of the alcohol dehydrogenase protein under study.

In order to enhance the feasibility of de novo sequencing from MALDI Q-TOF data, the enzymatic digest was also conducted in an aqueous solution that contains a 1:1 mixture of ^{16}O and ^{18}O water. The MassLynx software used for the data acquisition is capable of interpreting these isotopically labeled spectra automatically. An example for de novo sequencing is shown in Figure 5, showing the MS/MS spectrum of the same peptide as in Figure 2d. Even though the spectrum is highly complex with many intense b-ions, y-ions and a-ions present, with isotope labeling seven amino acids could be assigned automatically. Also, the isotope labeling prevented the misinterpretation of this spectrum as outlined above. This information was then used to locate a tryptic peptide of alcohol dehydrogenase with the sequence.

All in all five peptides originating from isotopically labeled digestion were studied (Table 2). To four of these five peptides, a partial sequence of 5–10 amino acids length could be assigned without user intervention. Manually, these sequences could easily be elongated as shown in Table 2. A seven amino acid sequence could be assigned manually to the fifth spectrum using the $^{16}O/^{18}O$ isotope information.

Discussion

The MALDI Q-TOF is a relatively new type of mass spectrometer that combines a MALDI ion source with a

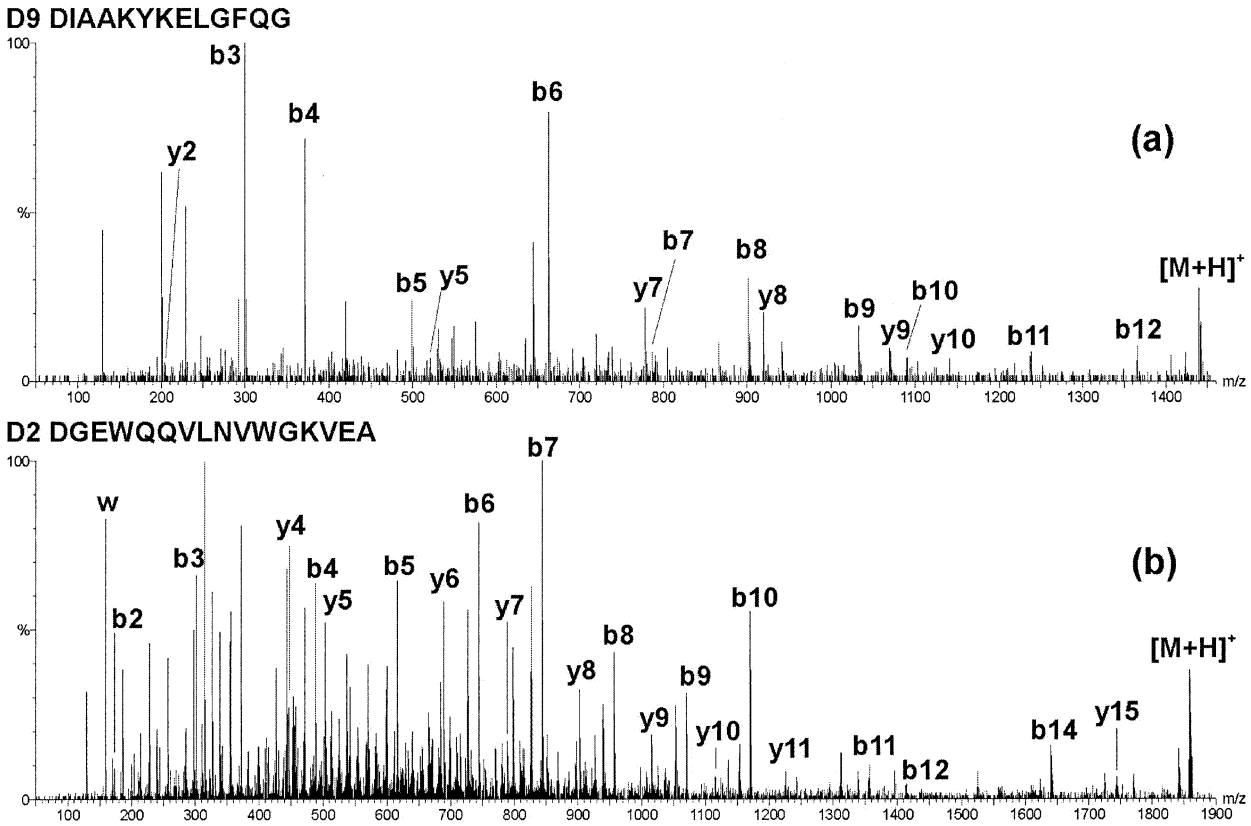

Figure 4. MALDI Q-TOF MS/MS spectra of peptides arising from an AspN digest of horse heart myoglobin.

tandem quadrupole-orthogonal acceleration time-of-flight mass analyzer (Q-TOF), that so far has been exclusively used with atmospheric pressure ionization sources. This new combination of desorption method and mass analyzer opens up new and exciting opportunities for proteomics research, one of the most important being the possibility to acquire MS and MS/MS spectra of ions produced by the MALDI from the same

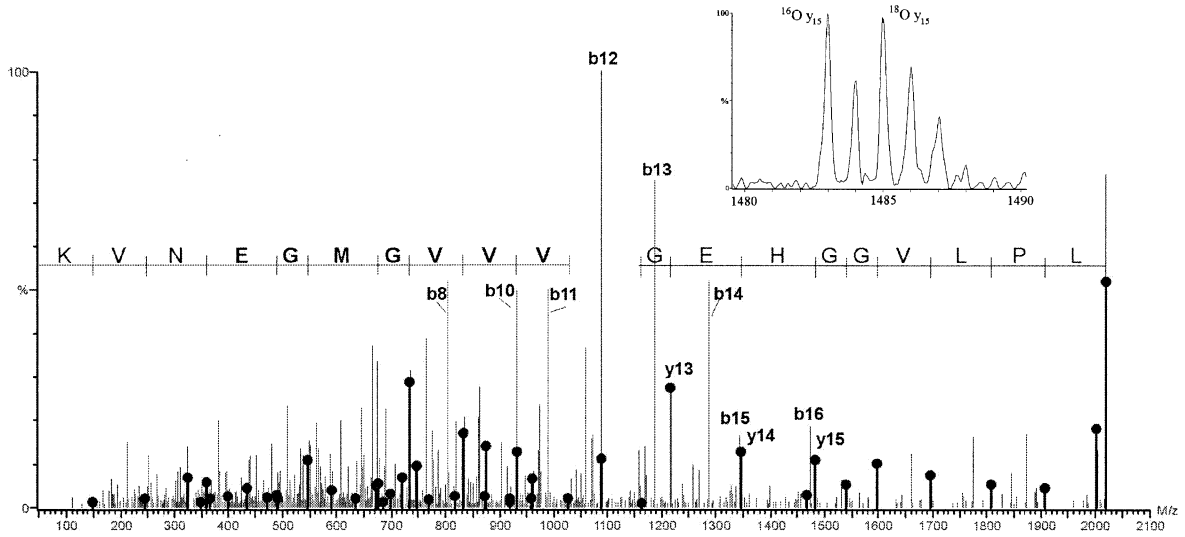

Figure 5. Example of a MALDI Q-TOF MS/MS spectrum of a tryptic peptide of alcohol dehydrogenase digested in a 50:50 mixture of ^{16}O and ^{18}O water. The sequence automatically annotated is shown in the figure and in Table 2. The sequence stretch printed bold was assigned automatically. The ions identified as potential y-ions due their 1:1 isotope ratio are marked with a filled circle. The inset shows an expanded view of the y_{15} ion region.

Table 2. Sequences derived from the ^{18}O labeled digestion of ADH

Enzyme	Mass	Digest peak	Sequencing mode	Sequence
Trypsin	1136.7	x	automatic	vgl**FYESH**GK
Trypsin	1338.8		automatic	ccsD**VFNQVV**K
Trypsin	2019.3	x	automatic	LPLVGGHEGag**VVVGMGE**NVK
Trypsin	2312.4	x	automatic	atdggaHGVIN**VSVSEAAIEA**STR
Asp-N	1679.9		manual	..cELGNESNc..

The bold sequence was assigned automatically by the MassLynx software. The sequence in capital letters was added manually, whereas the amino acids in lower case letters could not be read from the mass spectra.

sample preparation. One of the main questions now is, how the quality of the data—in terms of mass accuracy and fragmentation characteristics—affects the protein identification from protein and/or DNA databases.

From a MALDI MS perspective, the Q-TOF analyzer does not change the overall appearance of the MS spectra from those acquired using more traditional axial time-of-flight mass analyzers. This comparison is shown in Figure 1. The mass accuracy and the peaks identified are comparable.

On the other hand, changing the ion source of the Q-TOF from ESI to MALDI ionization reveals major differences in the MS/MS mode (Figure 2). The most obvious difference between these desorption methods is the charge state of the parent ions. Whereas the MALDI ion source typically produces singly charged ions, the nanoelectrospray source generates multiply charged parent ions. The consequences of the different parent ion charge states for fragmentation is shown in Figure 2. Here, a set of tryptic peptides were ionized both by MALDI or ESI and fragmented in the Q-TOF instrument.

Not surprisingly, MALDI and nanoelectrospray MS/MS experiments give rise to very different fragment ion spectra. Most notably, singly charged peptides are prone to fragmentation at specific peptide bonds as opposed to a more non-selective cleavage of multiply charged ions. When fragmenting a parent ion, the daughter ions produced will always vary in intensity according to the ease with which the peptide bonds are broken and the location of the ionizing charge. In some cases, however, ions are formed with a greater preference thus clearly dominating the fragment ion spectra. These ions are considered to be formed due to preferential cleavage.

Using this criterium, it is apparent that the preferable cleavage of singly charged peptides arise from the cleavage of peptide bonds N-terminal to proline (Figure 2b: y_6)and C-terminal to glutamic acid or aspartic acid (Figure 2c: y_3, y_4, y_7, y_{10} and Figure 3b: y_4, y_{11}).

This behavior is not unique to the MALDI Q-TOF used in this study. In previous reports, groups have used MALDI Q-TOF fragmentation data for protein identification [5–9]. Even though the authors do not explicitly mention the sequence dependent fragmentation, the spectra published in these papers show the same preferential cleavage of the X-P and D/E-X bond. Furthermore, the same fragmentation pattern has been observed using a MALDI ion trap mass spectrometer [11]. Also, MALDI PSD data shows preferential cleavage of the D-X bond [12]. This fragmentation has been shown to be inhibited by esterification of the acidic side chains in the peptide. This strongly indicates that the acidic function of the side chain plays a crucial role in the bond cleavage.

The preferential fragmentation behavior of the peptide bonds adjacent to D/E can be rationalized by using the "mobile" proton model for peptide fragmentation [13, 14]. In this model, the proton responsible for fragmentation is able to move along the peptide backbone and protonate any of the backbone amide bonds. These protonated bonds are then specially susceptible to cleavage. If the proton can move freely along the backbone, all peptide bonds will be cleaved with the same probability. On the other hand, protons can be "trapped" by basic amino acids (especially arginines) and therefore may not be available for fragmentation. This for example can be seen in the ESI MS/MS fragmentation of tryptic peptides, in which doubly charged peptide ions fragment much more easily than singly charged ions. This can be rationalized by a trapping of the first proton by the basic amino acid at the C-terminus. This increases the required energy to fragment a singly charged ion. In doubly charged ions, the second proton is free to move across the peptide backbone and can induce fragmentation easily.

In a study analyzing model peptides containing both acidic amino acids and the basic amino acid arginine, the preferential cleavage of peptide bonds C-terminal of aspartic and glutamic acid could be shown with a variety of fragmentation methods [15]. In the model put forward in that paper, the protonated arginine in the peptides forms a acid-base interaction with the acidic side chains of D or E, solvating and effectively trapping the ionizing proton. If no further proton is available that can act as the mobile proton, a second mechanism involving the acidic proton of the aspartic or glutamic acid cleaves the adjacent bond. This model explains why in our study, preferential cleavage was observed only for some tryptic peptides. As ions generated by MALDI carry just one ionizing proton, this generally will be trapped if the peptide contains a C-terminal arginine and one or more acidic amino acids. As no mobile proton is no longer available, this leads to preferential cleavage adjacent to the acidic amino acids.

The peptide in Figure 2a contains arginine but no

aspartic or glutamic acid. Therefore, no sequence-dependent fragmentation is observed. The peptide in Figure 2b contains no arginine and shows predominantly proline cleavage (y_6) and no cleavage adjacent to aspartic or glutamic acid (y_1, y_4). In Figure 2c, the effect of the mobile proton can be seen for peptides containing a combination of acidic amino acids and terminal arginine. The singly protonated MALDI ion leads to preferential cleavage of the backbone (y_3, y_4, y_7, y_{10}) as the proton is trapped. The doubly charged ion (generated by nanoelectrospray) fragments without any preferences as the second proton can freely move along the peptide backbone and cleave it evenly. As the peptide in Figure 2d contains no arginine in the primary structure, no preferential cleavage is observed, even though acidic amino acids are present. In accordance with the model, another tryptic peptide containing acidic amino acids but no arginine also shows no preferential cleavage (Figure 3a). In the MALDI MS/MS spectrum, the cleavage between glutamine and glycine (y_5) is as pronounced as the cleavage between aspartic acid and alanine (y_{10}). Interestingly, the loss of water from the b_3 and b_8 ions is very pronounced in this example. Figure 3b shows preferential cleavage again as the fragment spectrum is dominated by the y_4 and y_{11} ions. This spectrum arises from a peptide that contains a C-terminal arginine and aspartic as well as glutamic acid.

The fragmentation model also predicts that peptides lacking basic amino acid residues should not show such a preferential cleavage. The endoprotease Asp-N was used to digest the proteins alcohol dehydrogenase and myoglobin. Peptides generated by Asp-N cleavage carry an acidic amino acid on the N-terminus thus changing the chemistry of the digest peptides quite dramatically. Some MALDI Q-TOF spectra of this new set of peptides are shown in Figures 4a, b. As predicted by the model of the "mobile" proton, no preferential cleavage could be observed in this set of peptides. Whereas in the case of the tryptic peptides certain cleavage products were almost completely suppressed (y_5, y_6, y_8, y_9, y_{11}-y_{15} in Figure 2c), the sequence ions in Figure 3 are all identifiable. Especially in Figure 4a, a preferential cleavage of the b_8 ion (produced by cleavage to the E-L bond) is not observed.

Another change that could be observed in this limited data set was a higher tendency of the Asp-N peptides to form b-ions. As the C-terminus of the peptides no longer carries a positively charged amino acid, the formation of y-ions is no longer favored. Therefore, the relative probability of y- and b-ion formation is about the same. This leaves both b- and y-ions present in the MS/MS spectra without one series dominating.

The drawback of the preferential cleavage is that it decreases the information content of the MALDI MS/MS spectra as the complete y-ion series may no longer be detectable. Therefore, some sequence information is no longer available. For database searching, this may not be a problem as this fragmentation behavior can be accounted for in the matching algorithms. It can even be beneficial when it can be used to validate the database search result. For de novo sequencing on the other hand, missing fragment ions can lead to problems in unambiguously identifying the peptide sequence.

Another interesting tendency that can be seen in this MALDI Q-TOF data is that larger parent ions tend to give rise to very complex fragment ion spectra due to abundant neutral loss of water and ammonia. This is especially obvious in Figures 2d and 4b. In both cases, large (>1800 Da) peptides that did not contain any preferential cleavage sites were studied.

So the important question for proteomics is whether the quality of the MS/MS data acquired with a MALDI Q-TOF is such that it allows an unambiguous identification of the protein it was derived from. Even though it has been shown that MALDI Q-TOF data is of sufficient quality to allow protein identification using database search engines [5–7], a detailed assessment of how this different fragmentation behavior affects the success rate of this approach has not been undertaken. As discussed above, the sequence dependent cleavage of tryptic peptides potentially decreases the information content of the MALDI MS/MS spectra. The MALDI MS/MS spectra of tryptic and Asp-N digests of different proteins were submitted to a database search. Twenty two out of 26 peptides (85%) could be identified using standard parameters for the search as outlined in the experimental section. A protein was considered to be identified if the correct database entry was the top hit in the database search. Two of the missed peptides were not identified immediately, as the protease Asp-N gave rise to some unspecific cleavage. These peptides were identified using the MASCOT search engine without specifying a digestion enzyme. In the future the database searches could be made even more effective taking into account the different fragmentation behavior of the MALDI Q-TOF, such as the preferred cleavage sites X-P and E/D-X in arginine containing peptides. Still, two peptides could not be directly identified by this approach. The main reason for this was insufficient fragmentation of the parent ion even though the collision energy was raised well above the expected value.

Another approach to protein identification is automatic de novo sequencing directly from the MS/MS spectrum and identification of the protein by homology, e.g., through a BLAST search [7]. Even though this approach has been largely successful, an important problem is that the quality of the BLAST search depends on the quality of the search sequence submitted. Indeed, the authors of Ref. [7] state that ". . .the limited accuracy of automated interpretation of MS/MS spectra, as many sequence proposals as possible need to be included in a query" pointing at this problem. This is due to the high complexity and preferential fragmentation of MALDI generated- ions, which does not make it generally possible to derive sequence information directly from the MS-MS spectra (de novo sequencing).

Especially the complex spectra of high mass precursor ions as shown in Figure 2d are difficult to interpret. A possible way around this problem is the digestion of the peptides in isotopically-enriched water containing 50:50 ^{16}O:^{18}O [16]. As the two isotopes are statistically incorporated into the C-terminal acid group, all fragment ions deriving from the C-terminus will have a distinct 1:1 isotope distribution, allowing for an easy identification of y-ions. The MassLynx software has a built-in automatic sequencing algorithm for such isotope labeled peptides. The benefit of this additional information is shown in Figure 5. The ions that were identified as y-ions by the software are marked with circles in Figure 5. The complexity of the spectrum has been reduced greatly by isotope labeling. Even though some ions were falsely labeled as potential y-ions, the additional isotope information makes it much easier to interpret this complex spectrum.

The MS/MS spectrum of the same peptide as shown in Figure 5 is also shown in Figure 2d, where it was digested under normal conditions. In Figure 2d the correct protein was identified after conventional tryptic digestion by database searching. A manual sequence read-out, however, is not straightforward due to the multitude of possible fragment peaks. Now, the peptide was generated by a tryptic digest in isotope labeled water and identified by de novo sequencing. Automatically, the program can extract the sequence tag VVVGMGE using the isotope information. More importantly, during manual sequencing the isotope pattern can be used to narrow down the possible amino acid assignments. In the mass region above 1200Da, the y-ion series can be identified clearly, providing the sequence tag LPLVGGHE ending on the ion labeled y_{13} (*m/z* 1216.72) in Figure 5. Without the isotope information, it is possible to extend this sequence tag via the ions labeled b_{12} (1087.67), b_{10} (*m/z* 931.57) and b_8 (*m/z* 803.48). This would give rise to a sequence ending in ERK instead of the correct EGAG. The isotope distribution of these prominent peaks shows clearly that these are not y-ions ruling out the wrong sequence interpretation.

The isotope information allowed the automatic (partial) sequence assignment in four out of five peptides (Table 2). A longer stretch sequence could easily be assigned manually taking into account the isotope information. The fifth peptide sequence was assigned manually in the same way. Two of the peptides sequenced were not identified as digest peaks in the MS/MS database search. After sequencing these peptides with the MALDI Q-TOF, they could be identified as cystein-bridged digest products with the sequences CCSDVFNQVVK and CELGNESNC. As the disulphide bonds of the proteins digested for this work were not reduced and alkylated, these peaks contain cross-linked peptides and therefore were not identified on basis of their masses alone. As an example, the peptide of the mass 1338.8 was not identified as a tryptic peptide through its mass alone. The measured mass was 97 Da higher than the theoretical mass for the peptide with unmodified cysteins making the identification via database search impossible. The nature of this mass difference is not easily explained as it does not fit with any of the possible fragment structures.

This shows that using the isotope labeling technique, de novo sequencing with MALDI Q-TOF is possible and can be used to identify those peaks in a mass spectrum that are not automatically related to a database search hit.

Conclusions

MALDI Q-TOF is a relatively new technique that can obtain MALDI MS and MS/MS data from the same sample preparation. The scope of this paper was to evaluate this new instrument with regards to the type of fragmentation observed and its implications for the use in proteomics.

The comparison of MS data of protein digests obtained on a conventional MALDI TOF instrument to the MS data from the MALDI QTOF reveal very similar peptide patterns with very similar intensity ratios. However, comparison of MS/MS QTOF data produced by nanoelectrospray versus MALDI show striking differences. Peptide fragment ions obtained from doubly charged precursors produced by nanoelectrospray are mainly y-type ions with some b-ions in the lower mass range, whereas peptide fragment ions produced from the singly charged ions originating from the MALDI source are a mixture of y-, b-, a-ions and ions thereof after neutral loss of ammonia or water. The ratio and intensity of these different fragment ions is found to be more sequence dependent for MALDI generated ions. The singly charged peptides generated by MALDI show a preferred cleavage C-terminal of acidic residues aspartic and glutamic acid as well as N-terminal of proline. This preferential cleavage can be explained by the mobile proton model and is present only in peptides that contains both arginine and an acidic amino acid. In agreement with the proposed model, digest peptides with an acidic group on the N-terminus generated by an Asp-N digest do not exhibit a sequence dependent fragmentation.

The MALDI Q-TOF MS/MS data of 24 out of 26 proteolytic peptides produced by either trypsin or Asp-N digestions were successfully used for protein identification via database searching, thus indicating the general usefulness of MALDI Q-TOF MS/MS data for protein identification.

Due to the complex nature of MALDI MS/MS spectra de novo sequencing can be a difficult task. Specially abundant b- and a-ions can hinder an unambiguous sequence determination. The use of a mixture of ^{16}O/^{18}O water during digestion proved to be beneficial to this end. Using this approach, four out of five peptides studied in this paper could be assigned a correct sequence tag automatically.

Acknowledgments

Funding of Andreas Wattenberg under the European Union Marie-Curie industrial fellowship program is gratefully acknowledged. Andy West and Neil Jones are gratefully acknowledged for their valuable discussion and support throughout this study.

References

1. Pandey, A.; Mann, M. *Nature* **2000,** *405,* 837–846.
2. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999,** *20,* 3551–3567.
3. Eng, J. K.; McCormack, A. L.; Yates, J. R., III. *J. Am. Soc. Mass Spectrom.* **1994,** *5,* 976–989.
4. Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *73,* 14440–14445.
5. Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000,** *14,* 1047–1057.
6. Shevchenko, A.; Loboda, A.; Shevchenko, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2000,** *72,* 2132–2141.
7. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevchenko, A.; Bork, P.; Ens, W.; Standing, K. G. *Anal. Chem.* **2001,** *72,* 1917–1923.
8. Krutchinsky, A. N.; Zhang, W.; Chait, B. T. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 493–504.
9. Baldwin, M. A.; Medzihradszky, K. F.; Lock, C. M.; Fisher, B.; Settineri, T. A.; Burlingame, A. L. *Anal. Chem.* **2001,** *73,* 1707–1720.
10. Harvey, D. J.; Bateman, R. H.; Bordoli, R. S.; Tyldesley, R. *Rapid Comm. Mass Spectrom.* **2000,** *14,* 2135–2142.
11. Qin, J.; Chait, B. T. *J. Am. Chem. Soc.* **1995,** *117,* 5411–5412.
12. Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. *Anal. Chem.* **1993,** *65,* 3015–3023.
13. McCormack, A. L.; Somogyi, Á.; Dongré, A. R.; Wysocki, V. H. *Anal. Chem.* **1993,** *65,* 2859–2872.
14. Dongré, A. R.; Somogyi, Á.; Wysocki, V. H. *J. Mass Spectrom.* **1996,** *31,* 339–350.
15. Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Chem. Soc.* **1999,** *121,* 5142–5154.
16. Schnoelzer, M.; Jedrzejeski, P.; Lehmann, W. D. *Electrophoresis* **1996,** *17,* 945–953.