

---

---

# A Case Study of De Novo Sequence Analysis of N-Sulfonated Peptides by MALDI TOF/TOF Mass Spectrometry

Bart Samyn, Griet Debysers, Kjell Sergeant, Bart Devreese, and Jozef Van Beeumen

Department of Biochemistry, Physiology, and Microbiology, Laboratory of Protein Biochemistry and Protein Engineering, University of Gent, Gent, Belgium

---

The simplicity and sensitivity of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry have increased its application in recent years. The most common method of "peptide mass fingerprint" analysis often does not provide robust identification. Additional sequence information, obtained by post-source decay or collision induced dissociation, provides additional constraints for database searches. However, de novo sequencing by mass spectrometry is not yet common practice, most likely because of the difficulties associated with the interpretation of high and low energy CID spectra. Success with this type of sequencing requires full sequence coverage and demands better quality spectra than those typically used for data base searching. In this report we show that full-length de novo sequencing is possible using MALDI TOF/TOF analysis. The interpretation of MS/MS data is facilitated by N-terminal sulfonation after protection of lysine side chains (Keough et al., Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 7131–7136). Reliable de novo sequence analysis has been obtained using sub-picamol quantities of peptides and peptide sequences of up to 16 amino acid residues in length have been determined. The simple, predictable fragmentation pattern allows routine de novo interpretation, either manually or using software. Characterization of the complete primary structure of a peptide is often hindered because of differences in fragmentation efficiencies and in specific fragmentation patterns for different peptides. These differences are controlled by various structural parameters including the nature of the residues present. The influence of the presence of internal Pro, acidic and basic residues on the TOF/TOF fragmentation pattern will be discussed, both for underivatized and guanidinated/sulfonated peptides. (J Am Soc Mass Spectrom 2004, 15, 1838–1852) © 2004 American Society for Mass Spectrometry

---

---

Protein identification protocols in proteomics currently rely on the peptide mass fingerprinting (PMF) technique in which a (mostly gel-purified) protein is digested with an endoprotease of known cleavage specificity. The masses of the resulting peptides are measured by mass spectrometry, usually matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), and matched to peptide masses that have been generated theoretically from proteins in databases. Notwithstanding the fact that the PMF approach is useful for identifying proteins in simple mixtures (e.g., 2D-PAGE gel spots), the identification of proteins in more complex mixtures often requires partial peptide sequence data obtained by tandem mass spectrometry (MS/MS) methods. If accurate genome sequence information is available, data-

base search algorithms can be used in conjunction with MS/MS to readily identify proteins. For proteins not contained within sequence databases, it is necessary to determine partial or complete amino acid sequences using either manual or automated de novo peptide sequence analysis methods. This general approach has been applied with both electrospray (ESI) and MALDI ionization.

ESI MS has found use for de novo sequencing under a variety of conditions, which include collision-induced dissociation (CID) in the triple quadrupole mass analyzer [1] and the QqTOF mass analyzer [2], resonance excitation in the quadrupole ion trap mass analyzer [3], and electron capture dissociation in the Fourier transform ion cyclotron resonance mass analyzer [4]. Regardless of the technique employed, interpretation of MS/MS spectra often requires manual interpretation which remains prohibitively challenging because of the variation in favored ion fragmentation sites, the chemical nature of amino acid side chains, and their relative order in the peptide backbone. Because of the complex

---

Published online November 2, 2004

Address reprint requests to Dr. J. Van Beeumen, Department of Biochemistry, Physiology, and Microbiology, Laboratory of Protein Biochemistry and Protein Engineering, University of Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium. E-mail: Jozef.vanbeeumen@UGent.be

nature of peptide fragmentation, a number of sophisticated search algorithms, such as Sequest [5] and Mascot [6], have been developed to facilitate the interpretation of MS/MS data. While these database search algorithms have proved very useful, their performance is directly related to the quality of the product ion spectra. In many cases, the algorithms will yield search scores that are below the accepted thresholds, requiring the data to be manually validated.

MALDI de novo sequencing has been carried out using CID in QqTOF instruments [7, 8] and by resonance excitation of singly charged peptides in a home-built MALDI-quadrupole-ion trap mass spectrometer [9]. In the past, limited sequence information was also obtained through the use of post-source decay (PSD) MALDI-TOF. A drawback of peptide sequencing by the latter technique is the relatively complex and labor-intensive interpretation of fragment ion spectra, mainly due to the many types of fragment ions that arise during fragmentation, i.e., immonium ions, N-terminal fragments ( $a_n^-$ ,  $b_n^-$ ,  $c_n^-$ , and  $d_n^-$ -type ions), C-terminal fragment ions ( $x_n^-$ ,  $y_n^-$  and  $z_n^-$ -type ions), and internal fragment ions resulting from double chain cleavage.

Attempts to simplify and improve the fragmentation pattern, both for PSD and MS/MS approaches, have largely concentrated on the introduction of charged groups at the N- or C-terminus of the peptide (for a review, see [10]). For these purposes chemical derivatizations that add acidic [11, 12], fixed-charge [13], or basic [14] moieties have been used. Alternatively, other derivatives have provided specific, but less complete, sequence information. Gaskell and co-workers converted peptide N-termini to phenylthiocarbonyl (PTC) derivatives [15]. This modification induces selective dissociation of N-terminal peptide bonds in gas phase fragmentation experiments, enabling the facile identification of N-terminal residues. More recently, Beardsley and Reilly [16] illustrated that the use of the N-terminal residue of a peptide as a search constraint reduced the number of candidates extracted from a sequence database and improved database-matching techniques by reducing the occurrence of false-positive assignments. In their study, free amino groups were selectively modified with an amidine label, a form of labeling that has also been employed as a tag in comparative proteomics [17]. CID analysis of singly charged acetamidinated peptides indicated a predominance of  $b\text{-NH}_3$  fragment ions in the fragmentation spectra. While this unique type of fragmentation may facilitate de novo sequencing, the authors noted that not every peptide generated continuous ( $b\text{-NH}_3$ )-ion series [16].

Keough et al. developed a peptide derivatization strategy that facilitates de novo sequencing of singly charged tryptic peptides [11]. Introduction of a strong acid group at the N-terminus of tryptic peptides was shown to facilitate protonation of backbone amide bonds. Protonation destabilizes amide bonds leading to extensive structure-specific fragmentation under post-source decay MALDI and electrospray tandem mass spectrometry conditions [11, 12]. Derivatization of the

N-terminus was accomplished by acylation with 2-sulfobenzoic acid cyclic anhydride or chlorosulfonylacetate chloride [11]. These reagents are commercially available and allow direct sulfonation of subpicomol quantities of tryptic peptides. Later, the method was refined by using a water-soluble, sulfopropionic acid NHS-ester [18], a reagent that recently became available in a commercial kit (Ettan CAF MALDI sequencing kit, Amersham Biosciences, Uppsala, Sweden) [19]. Very recently, a similar derivatization procedure was reported using 4-sulfophenyl isothiocyanate as the derivatization reagent [20, 21].

With the introduction of the 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA) a new MALDI MS/MS configuration became available recently. Dissociation of precursor ions in the TOF/TOF system is based on the combination of two independent mechanisms, metastable decay and collision induced dissociation [22]. Ions formed by MALDI in high vacuum have elevated internal energies and a portion of this "hot" ion population may dissociate on the time-scale of TOF detection by metastable decay. For the recording of MS/MS data, the ionization conditions are similar to those used for post-source decay, i.e., a "hot" matrix and a higher-than-threshold laser power to actively promote metastable decay on short ( $<150 \mu\text{s}$ ) time-scales. The second method of ion activation in the TOF/TOF system is high-energy CID, promoted by laboratory frame collision energies of 1–2 keV and nominally single collision conditions. In a recent application MALDI TOF/TOF was utilized for protein identification via analysis of a tryptic mass map and the authors illustrated that the increased structural information gained from the use of this technique leads to the identification of rare posttranslational modifications on tryptophan [23]. However, high energy CID spectra contain a lot more fragment ions, which are not simple backbone cleavage products and therefore do not provide straightforward sequence information for de novo work. In the present study, we compare the tandem mass spectrometric behavior of guanidinated sulfonated peptides with that of their native analogues using a MALDI TOF/TOF instrument. Fragmentation of both sulfonated and underivatized peptides occurred under metastable decay conditions and a frame collision energy of 1 keV (no gas in the collision cell). Rather than using a number of "test" peptides, we carried out a case study in which a set of peptides, generated by RPLC separation of enzymatically digested proteins, was used. The observed preferential fragmentation of underivatized and derivatized fragments is discussed with respect to several, recently published, studies.

## Experimental

### Materials

Horse heart cytochrome *c* (purity 97%), bovine milk  $\alpha$ -casein (purity 85%), synthetic peptides, and  $\alpha$ -CHCA

were from Sigma (Bornem, Belgium). Endoprotease Lys-C was from Roche (Brussels, Belgium) and trypsin was from (Promega, Leiden, The Netherlands). Dry tetrahydrofuran and 2-sulfolobenzoic acid cyclic anhydride were from Fluka (Bornem, Belgium). HPLC-grade acetonitrile (ACN) was obtained from BioSolve (Valkenswaard, The Netherlands).

### Tryptic Digestions of Proteins

The peptides used in this study were obtained from the proteolytic digestion of two test proteins: horse heart cytochrome *c* and bovine milk  $\alpha$ -casein. Both proteins were incubated with endoprotease Lys-C or trypsin in 10 mM  $\text{NH}_4\text{HCO}_3$ , pH 8, for four hours at 37 °C, at an E/S ratio of 1/40 (wt/wt). After digestion, the sample was acidified (1% TFA/MQ), centrifuged, and separated on an analytical RPLC column (Brownlee, 2.1 × 220 mm, 5  $\mu\text{m}$ , C18) using a SMART system (Pharmacia, Uppsala, Sweden). Individual peptide fractions were collected manually in 500  $\mu\text{l}$  Eppendorf vials and dried in a Speedvac (ThermoSavant, Brussels, Belgium). All fractions were redissolved in 0.1%TFA/MQ-water at an estimated concentration of 25 pmol/ $\mu\text{l}$ . The individual RPLC fractions were analyzed on the 4700 Proteomics Analyzer (Applied Biosystems) in the positive reflectron mode. Only those fractions containing one or two principal peptide components were judged to be pure and used in further experiments (Table 1).

### Guanidination and Sulfonation

An *O*-methylisourea stock solution was prepared by dissolving 0.050 g of the compound in 51  $\mu\text{l}$  of MQ

water. A freshly prepared *O*-methylisourea stock solution was used in every reaction. For the guanidination modification 1.5  $\mu\text{l}$  of this stock solution was mixed with 2  $\mu\text{l}$  of the peptide solution (50 pmol), 3  $\mu\text{l}$  MQ, 5.5  $\mu\text{l}$  7N  $\text{NH}_4\text{OH}$ , and incubated for 10 min at 65 °C. After guanidination, the peptides were desalted on a Prosorb device (PVDF) with two subsequent washes of 100  $\mu\text{l}$  MQ, and extracted with 10  $\mu\text{l}$  50 % ACN/12.5 mM  $\text{NH}_4\text{HCO}_3$ . 2-Sulfolobenzoic acid cyclic anhydride was prepared at a concentration of 2 mg/ml in dry tetrahydrofuran prior to use. A volume of 2  $\mu\text{l}$  of the sulfonation reagent was mixed with 2  $\mu\text{l}$  of the peptide extract, briefly vortexed, and reacted for 5 min at room temperature.

### Matrix-Assisted Laser Desorption/Ionization TOF/TOF Mass Spectrometry

The Applied Biosystems 4700 Proteomics Analyzer with TOF/TOF optics [22] was used in this study for reflectron analysis and MALDI MS/MS applications (Applied Biosystems, Framingham, MA). The mass spectrometer uses a 200-Hz frequency tripled Nd:YAG laser operating at a wavelength of 355 nm. For MS/MS, ions generated by the MALDI process were accelerated at 8 kV through a grid at 6.7 kV into a short, linear, field-free drift region. In this region, the ions pass through a timed-ion-selector device that is able to select one peptide from a mixture of peptides at different *m/z* values for subsequent fragmentation in the collision cell. After a peptide at a given *m/z* was selected by the timed-ion-selector, it passed through a retarding lens where the ions were decelerated and then passed into

**Table 1.** Overview of the peptide sequences and their masses (monoisotopic values) used in this case study

Peptide fragment	Sequence	Theoretical masses (Da)			Experimentally observed masses (Da)		
		Unmod. <sup>b</sup>	Guanid. <sup>c</sup>	Modif. <sup>d</sup>	Unmod. <sup>b</sup>	Guanid. <sup>c</sup>	Modif. <sup>d</sup>
$\alpha$ -S2 cas (A96-K106)	<u>ALNEINQFYQK</u>	1366.695	1408.695	1592.695	1367.643	1409.676	1591.270 <sup>a</sup>
$\alpha$ -S2 cas (N174-K180)	<u>NRLNFK</u>	903.536	945.536	1129.536	904.512	946.518	1130.490
$\alpha$ -S1 cas (Y106-R115)	<u>YLGYLEQLLR</u>	1266.705	1266.705	1450.705	1267.699	1267.639	1449.274 <sup>a</sup>
$\alpha$ -S1 cas (E140-K147)	<u>EGIHAQQK</u>	909.474	951.474	1135.474	910.530	952.538	1136.516
$\alpha$ -S2 cas (T197-K203)	<u>TVYQHQQK</u>	902.468	944.468	1128.48	903.521	945.531	1129.508
$\alpha$ -S2 cas (L168-R175)	<u>LTEEEKNR</u>	1017.516	1059.516	1243.516	1018.579	1060.582	1244.571
$\alpha$ -S1 cas (H23-R37)	<u>HQGLPQEVLNENLLR</u>	1758.945	1758.945	1942.945	1760.010	1759.979	1943.955
$\alpha$ -S2 cas (F189-K196)	<u>FALPQYLK</u>	978.561	1020.561	1204.561	979.640	1021.519	1203.248 <sup>a</sup>
HHC (T40-K53)	<u>TGOAPGFTYTDANK</u>	1469.686	1511.686	1695.686	1470.723	1512.692	1694.529 <sup>a</sup>
Synthetic	<u>ELAQYNVEVHPYTVRK</u>	1945.013	1987.013	2171.013	1945.971	1987.906	2171.638
Synthetic	<u>APWFHHQNGK</u>	1220.591	1262.591	1446.591	1221.361	1263.379	1447.345
Synthetic	<u>QAQVYPNRFPLWK</u>	1645.880	1687.880	1871.880	1646.790	1688.574	1872.549
$\alpha$ -S1 cas (H95-R105)	<u>HIQKEDVPSER</u>	1336.681	1378.681	1562.681	1337.745	1379.783	1563.781
HHC (T28-K39)	<u>TGPNLHGLFGRK</u>	1295.717	1337.717	1521.717	1296.773	1338.715	1522.806

The numbering in parentheses is according to the numbering of the  $\alpha$ -casein ( $\alpha$ -cas) and horse heart cytochrome *c* (HCC) sequences in the latest Swiss-Prot release. Underlined peptide sequences could be deduced from the MS/MS spectra of the modified peptides.

<sup>a</sup>Ions observed in negative mode reflectron analysis;

<sup>b</sup>Mass of the unmodified peptides;

<sup>c</sup>Mass of the guanidinated peptides;

<sup>d</sup>Mass of the guanidinated and sulfonated peptides (modified).

the collision cell, which was operated at 7 kV. The collision energy, defined by the potential difference between the source and the collision cell, was 1 kV. Fragmentation was performed in the simple metastable decomposition mode (no collision gas and with the collision energy set at 1 keV). Under these experimental conditions, the major peaks in MS/MS spectra typically correspond to  $y$ - and  $b$ -series ions. The fragment ions were accelerated in the second source region at 15 kV, passed through a second, field-free, linear drift region, into the reflector, and finally to the detector. The detector amplifies and converts the signal to an electric current, which is observed and manipulated by a PC-based operating system. For high resolution mass analysis the instrument was operated in the reflector mode. After the MALDI process generates the peptide ions, the latter are accelerated at 20 kV through a grid at 14 kV into the first, short, linear, field-free drift region. After this point, the rest of the instrument can be treated as a continuation of this region until the ions enter the reflector and then reach the detector, whereas before, the signal at the detector was amplified and converted to electrical current.

The matrix solution was prepared as a 7 mg/ml  $\alpha$ -cyano-4-hydroxycinnamic acid solution in 50% ACN containing 0.1% TFA. A volume of 1  $\mu$ l of the sulfonated peptide was mixed with 1.5  $\mu$ l matrix solution, vortexed, and 0.5  $\mu$ l of the mixture was spotted on a 192-well stainless steel target plate (500 fmol peptide). The samples were allowed to air-dry at room temperature and were then inserted into the mass spectrometer and subjected to MALDI-MS analysis. Prior to analysis, the mass spectrometer was externally calibrated with a mixture of Angiotensin I, Glu-fibrino-peptide B, ACTH (1-17), and ACTH (18-39). For MS/MS experiments, the instrument was externally calibrated with fragments of Glu-fibrino-peptide. MS and MS/MS data were further processed using DataExplorer 4.0 (Applied Biosystems) or by manual interpretation.

## Results and Discussion

### *TOF/TOF Analysis of Native and Derivatized Peptides*

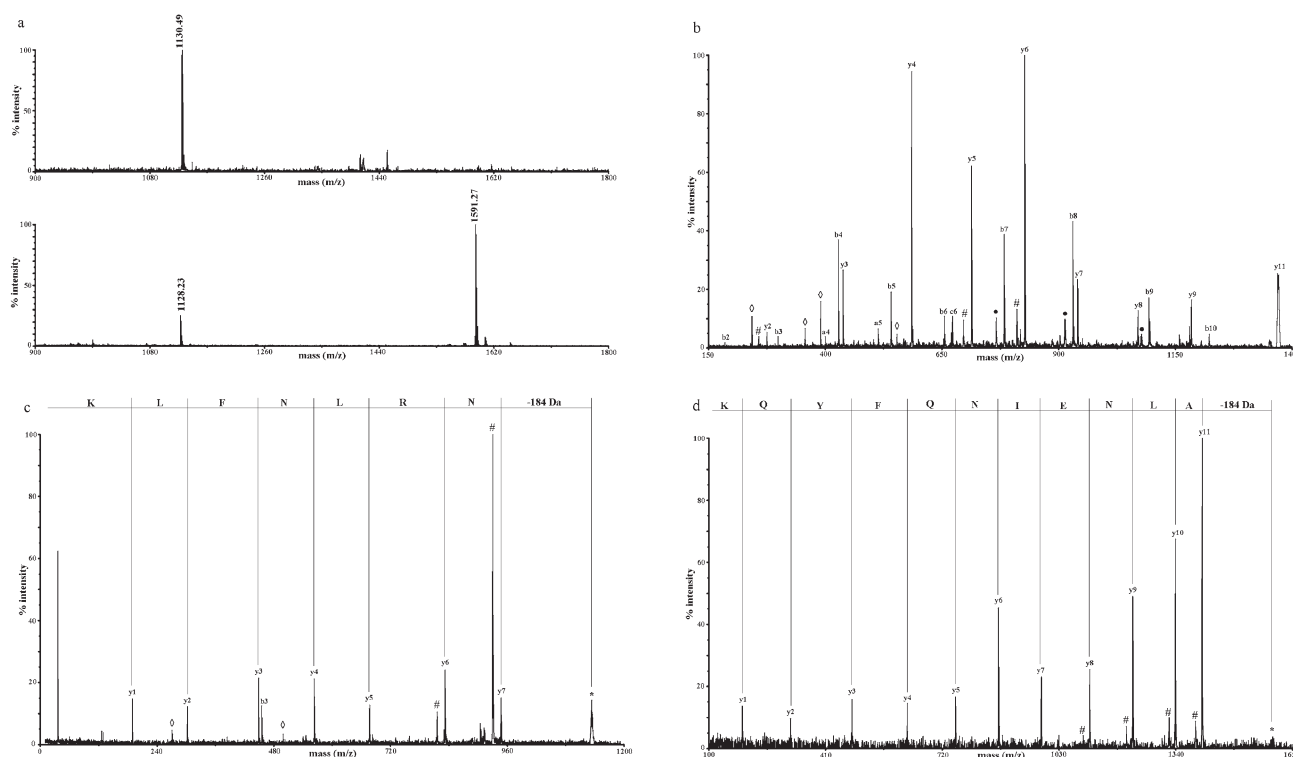
Peptides resulting from proteolytic digests with endoprotease Lys-C or trypsin were separated by HPLC. Individual fractions were analyzed by MALDI-MS (reflectron mode) and judged to be pure if they contained only one or two major fragments. A summary of the peptide fragments used in these experiments is given in Table 1. The numbering of the amino acids of the peptide fragments is according the latest Swiss-Prot release (version 42.10). Derivatization of the individual peptide fragments was performed as described before. After isolation, 50 pmol of RPLC-purified peptide was guanidinated and desalted on a Prosorb-device. After extraction and assuming 100% recovery, 10 pmol of the peptide was sulfonated using 2-sulfo-benzoic acid cyclic anhydride as the derivatization

reagent rather than chlorosulfonyl acetyl chloride, which is known to hydrolyze quickly with water. In all experiments, 500 fmol of the derivatized peptide was applied on the MALDI stainless steel probe. The minimum amount of protein required to obtain good quality MS/MS spectra varies from 10 pmol to 200 fmol, as recently reported by Lin et al. [24]. Fragmentation spectra were obtained from approximately 5000 laser shots of 500 fmol of the peptide.

MALDI reflectron analysis of a fraction from the endoprotease Lys-C digest of  $\alpha$ -casein indicated the presence of two peptides: ALNEINQFYQK (Ala96-Lys106,  $\alpha$ -S2 casein) and NRLNFLK (Asn174-Lys180) at 1367.643 and 904.512 Da, respectively. Upon guanidination (+42 Da) and sulfonation (+184 Da) the smallest fragment was clearly observed at 1130.490 Da but the modified, larger, fragment was not detected in positive ion mode analysis. Analysis in negative mode clearly showed the two fragments at 1128.228 and 1591.270 Da (Figure 1a). TOF/TOF analysis, in the positive mode, on the smaller precursor with  $m/z$  1130.49, yielded a complete  $y$ -ion series (Figure 1c). The precursor ion first loses the sulfonation label ( $-184$  Da) followed by formation of a complete series of  $y$ -ions. By simple manual calculation of the differences between the adjacent  $y$ -ion fragments, or by using suitable software, the complete sequence of the peptide could easily be determined. Fragment ions  $y_6$  and  $y_7$  are accompanied by their ( $y_i - 17$ ) satellite ions, most likely formed by neutral loss of an ammonia molecule from the internal arginine as discussed below. TOF/TOF analysis of the unmodified precursor from the larger fragment ( $m/z$  1367.643) yielded a complex fragmentation pattern including incomplete series of  $b$ -,  $a$ -, and  $y$ -ions and several internal fragment and immonium ions (Figure 1b). Even though it is possible to derive sequence information from such MS/MS spectra in some cases, the complexity of the fragment spectra often prevents an unambiguous sequence determination. After derivatization, the largest peptide was not observed in positive mode analysis. However, setting the first TOF analyzer to select the theoretical 1593.27 precursor ( $1591.27 + 2 H^+$ ), followed by fragmentation, also yielded the complete  $y$ -ion series (Figure 1d). The same phenomenon was observed during analysis of the tryptic peptide fragment YLGYLEQLLR (Tyr106-Arg115) from  $\alpha$ -S1 casein (P02662) (theoretical mass 1266.705 Da). As this peptide contains no Lys its mass remained unchanged after guanidination (experimental mass 1267.64 Da, Table 1). Upon sulfonation, the expected mass (1451.70 Da) was not observed in positive mode reflectron analysis but was clearly detected as its deprotonated form in negative mode analysis (Table 1). However, selection of the 1451.27 ion as precursor in positive mode fragmentation analysis yielded a spectrum containing exclusively  $y$ -ions (results not shown).

Figure 2a shows the analysis of a second tryptic fraction wherein three components were observed: two major ones, EGIHAQQK (Glu140-Lys147,  $\alpha$ -S1 casein) and TVYQHQQK (Thr197-Lys203,  $\alpha$ -S2 casein), and a minor one, LTEEEKNR (Leu168-Arg175,  $\alpha$ -S2 casein) at 910.530, 903.521 and 1018.579 Da, respectively (Table 1). After





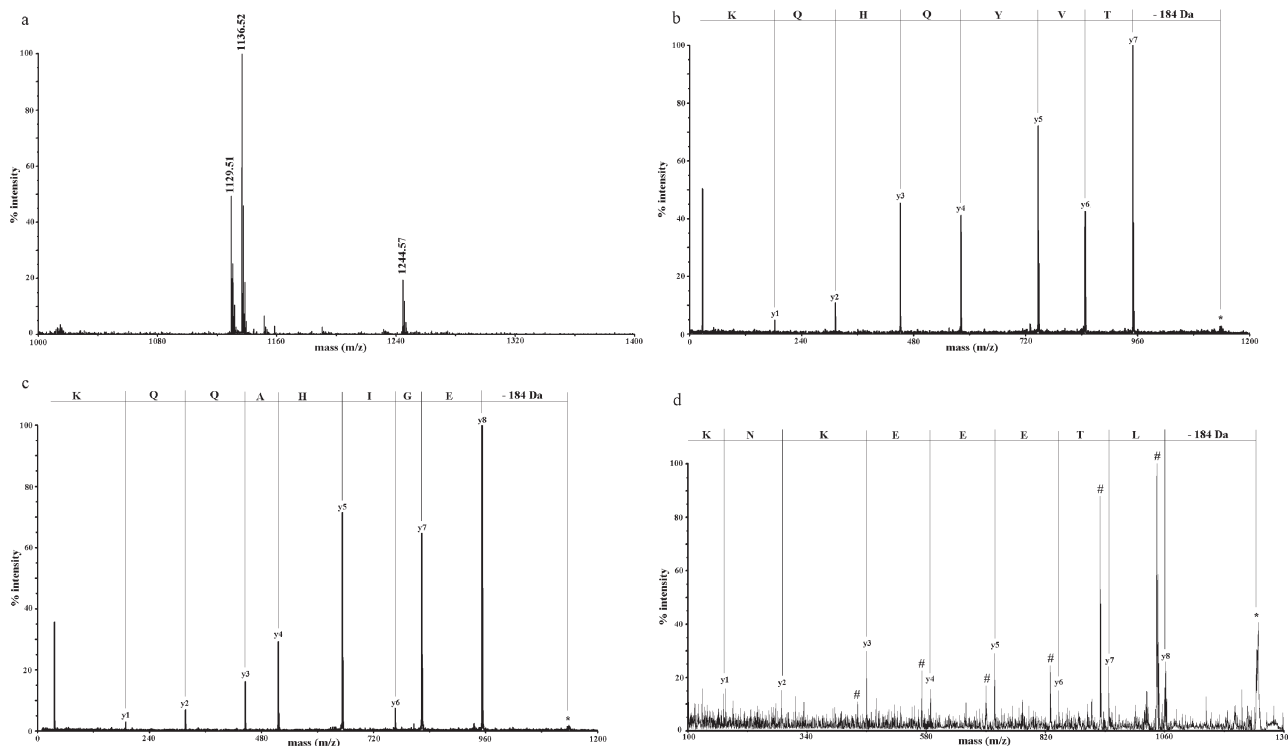
**Figure 1.** (a) The positive (upper) and negative (lower) MALDI reflectron MS spectrum of a RPLC fraction from  $\alpha$ -casein cleaved with endoprotease Lys-C. (b) MALDI MS/MS spectrum of the  $m/z$  1367.64 ion of the unmodified peptide. (c) MALDI MS/MS spectrum of the  $m/z$  1130.49 ion of the derivatized peptide. (d) MALDI MS/MS spectrum of the  $m/z$  1593.27 ion of the derivatized peptide (positive mode). In each spectrum only a-, b-, and y-ions are labeled. The derivatized precursors are labeled with an asterisk. Where appropriate, (y-17)-ions (number symbol), (b-17)-ions (filled circle), and internal fragment ions (open diamond) are shown as indicated.

guanidination and N-terminal sulfonation the masses of all fragments increased 42 and 184 Da, respectively, (Figure 2a). Fragmentation of the underivatized peptides yielded complex fragmentation patterns consisting of N-terminal a- and b-ions, C-terminal y-ions and internal fragment ions. Again, these spectra would be difficult to interpret de novo. Fragmentation of the completely derivatized major components resulted in two complete y-ions series (Figure 2b and c) whereas fragmentation of the minor fragment precursor ( $m/z$  1244.57), being approximately 5-fold lower in intensity (Figure 2a and d), resulted in a more complicated fragment spectrum. The y-ion series was evident in the spectrum, but all y-ions, except for  $y_1$  and  $y_2$ , were accompanied by (y - 17)-ions approximately twice as large as their corresponding  $y_i$  ions. The formation of this second series is due to the presence of an internal homoarginine which is known to lose ammonia (-17 Da). The loss of neutral molecules such as ammonia has also been observed during MALDI-analysis of peptides containing internal Arg residues.

#### *Peptides Having Internal Amino Acids Known to Induce Specific Fragmentation*

Two tryptic peptide fragments, HQGLPQEVLENLLR (His23-Arg37,  $\alpha$ -S1 casein) and FALPQYLK

(Phe189-Lys196,  $\alpha$ -S2 casein), were subjected to modification and subsequent MALDI TOF/TOF analysis. Upon guanidination and sulfonation of the second peptide fragment, FALPQYLK, again the peptide was not observed during MALDI reflectron analysis in the positive mode but only as its deprotonated ion in the negative mode (1203.248 Da, Table 1). Both the unmodified ( $m/z$  979.640) and the completely modified peptide ( $m/z$  1205.25, not observed in positive mode analysis) were fragmented. As described above, fragmentation of the native peptide yielded a mixture of incomplete b- and y-ion series (wherein the  $y_5$  ion is the most abundant) and a number of internal fragment ions, whereas fragmentation of the modified peptide ( $m/z$  1205.25) yielded the complete y-ion series (Figure 3a). As for the underivatized peptide, a preferential cleavage at the N-terminal side of Pro was observed, resulting in the presence of a dominant  $y_5$ -ion and a very weak, although detectable,  $y_4$ -ion. Since fragment His23-Arg37 contains no lysine, its mass remained unchanged upon guanidination (1759.979 Da). In the MS/MS fragment spectrum of the underivatized peptide, three major y-ions were observed,  $y_{11}$ ,  $y_{8'}$  and  $y_{4'}$ , indicating preferential fragmentation, respectively N-terminal of Pro5 and C-terminal of Glu7 and Glu11. Fragmentation of the



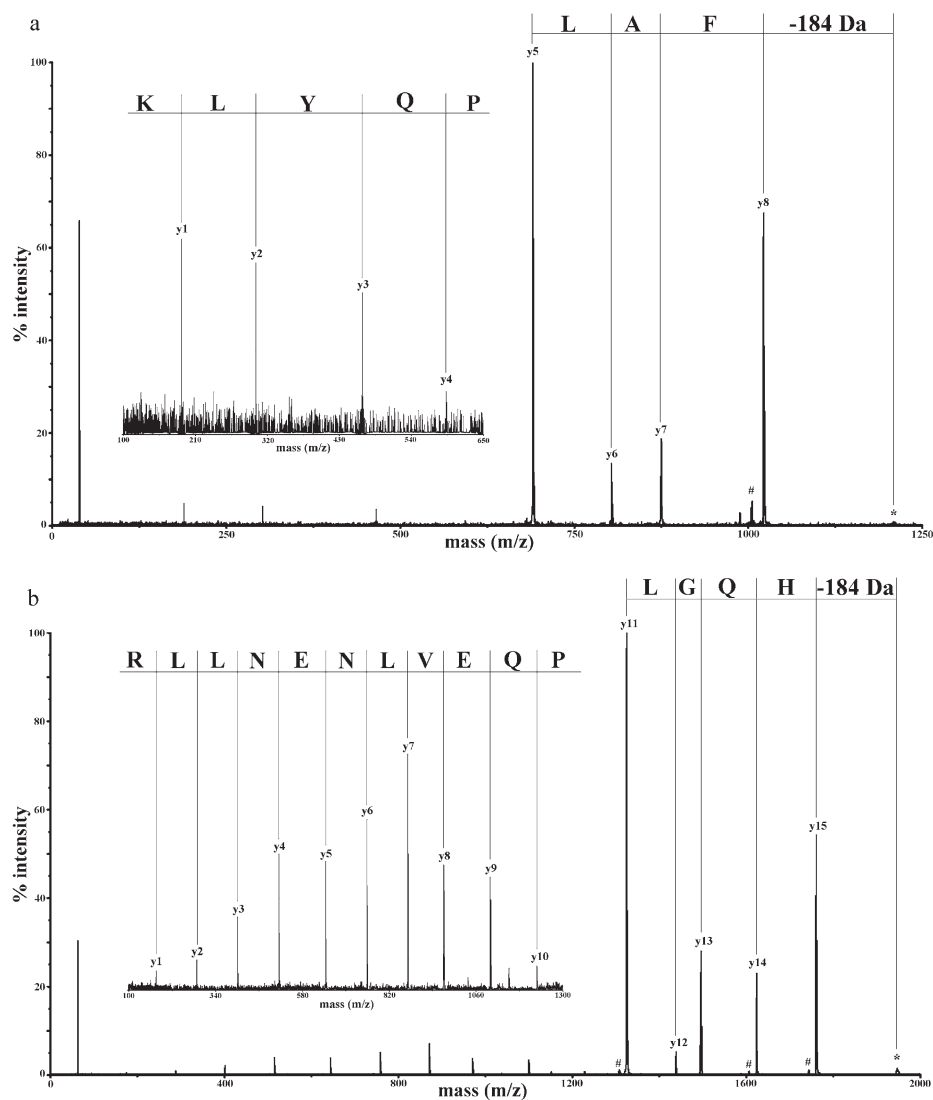
**Figure 2.** (a) MALDI reflectron MS spectrum (positive mode) of a tryptic  $\alpha$ -casein fraction after derivatization. (b), (c), and (d) show, respectively, the MALDI MS/MS spectra of the modified peptides at  $m/z$  1129.51,  $m/z$  1136.52, and  $m/z$  1244.57. The loss of the sulfonation label ( $-184$  Da) is indicated and ( $y$ -17)-ions are labeled with the number symbol.

sulfonated precursor ion (1943.955 Da) yielded the complete  $y$ -ion series ( $y_1$ - $y_{15}$ ) wherein the  $y_{11}$ -ion was the most dominant fragment-ion (Figure 3b). Although the intensity of the  $y_{10}$ -ion was low the ion could be observed in the fragmentation spectrum. Preferential cleavage C-terminal of the Glu residues was not longer observed (Figure 3b). Upon complete derivatization, peptide TGQAPGFYTDANK (Thr40-Lys53, horse heart cytochrome *c*) was only observed in negative mode analysis (1694.53 Da). Again, fragmentation of the hypothetical protonated precursor (1696.53 Da) yielded the complete  $y$ -ion series and indicated a preferential cleavage at the Ala-Pro bond. No preferential cleavage C-terminal of the Asp residue was observed, neither in the fragment spectra of the underivatized fragment nor in that of the sulfonated peptide.

From recent studies it is known that peptides containing internal basic residues (Arg, homo-Arg, or His) do not fragment as readily as typical tryptic fragments [25]. MALDI TOF/TOF analysis of the underivatized synthetic peptide ELAQYNVEVH-PYTVRK yielded an incomplete  $y$ -ion series, in which the  $y_6$ -,  $y_7$ -,  $y_8$ -, and  $y_{15}$ -ion were the most abundant fragment ions (Figure 4b). The predominance of  $y_8$  and  $y_{15}$  indicates a preferential cleavage of the Glu-Xxx bonds. The  $y_6$ -ion is attributable to preferential cleavage of the His-Pro bond whereas the  $y_7$  results from a preferential cleavage of the Val-His bond.

Fragmentation of the derivatized peptide (precursor  $m/z$  2171.64) yielded the complete  $y$ -ion series accompanied by a more pronounced ( $y - 17$ )-ion series due to the presence of Arg at the penultimate position. Some of these ( $y - 17$ )-ions were also observed in the fragmentation spectrum of the underivatized peptide (Figure 4b). The  $y_7$ -,  $y_8$ -, and  $y_{15}$ -ions are no longer more pronounced in the spectrum, in contrast to the  $y_6$ -ion, indicating that the preferential His-Pro bond cleavage still occurs (Figure 4d). Although weaker, the  $y_1$ - to  $y_5$ -ion series could still be observed. The influence of the presence of an internal His on the fragmentation was also observed during analysis of the synthetic peptide APWFHHQNGK. MALDI TOF/TOF analysis of the underivatized fragment ( $m/z$  1221.36) showed a mixture of  $b$ -,  $a$ - and  $y$ -ion series. The most dominant ions are the  $y_5$ - and  $y_6$ -ions, indicating preferential cleavage at the Xxx-His bonds (Figure 4a). In contrast to previous observations, the  $y_9$ -ion, indicating a preferential Ala-Pro bond cleavage, was not dominantly present in this fragmentation spectrum. Fragmentation of the derivatized peptide (precursor 1447.34 Da) yielded a completely different fragmentation spectrum with the  $y_9$ -ion being the most dominant fragment ion in the  $y$ -ion series (Figure 4c). The  $y_5$ - and  $y_6$ -ions were only slightly more pronounced in this  $y$ -ion series.

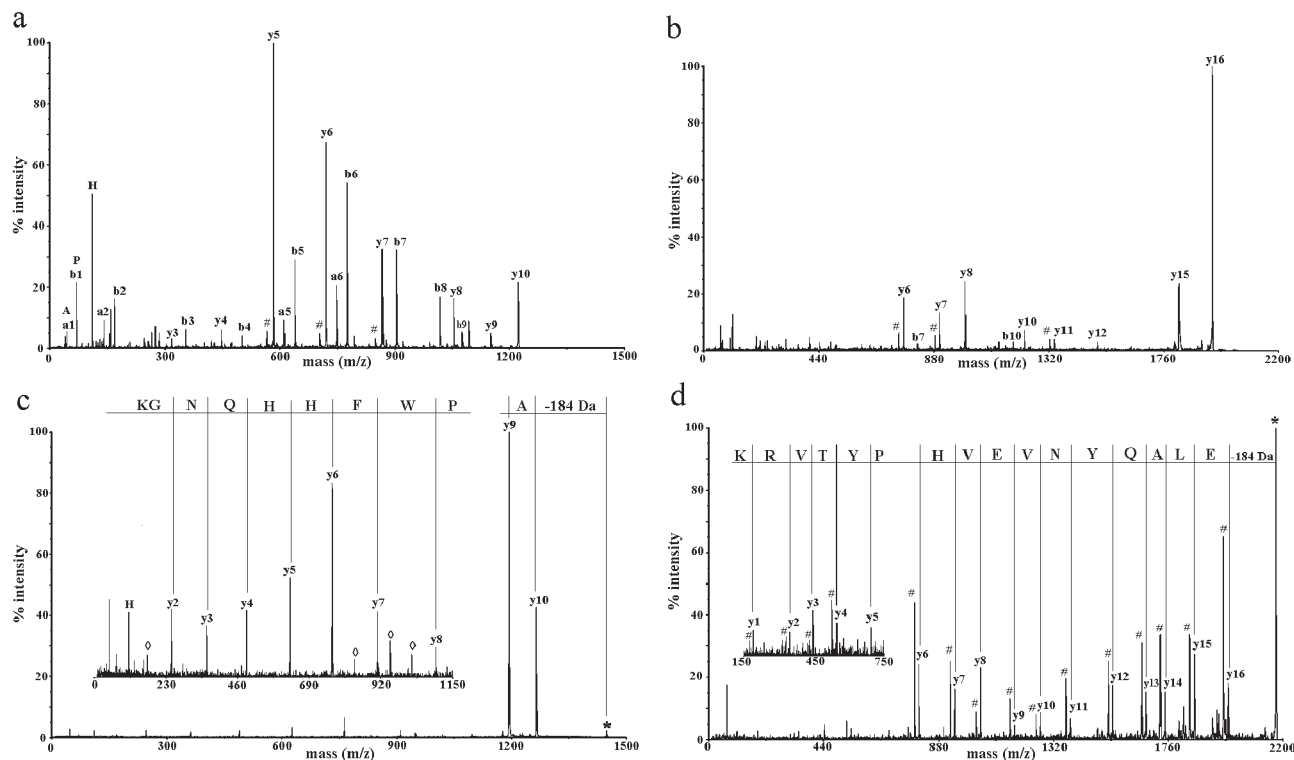
TOF/TOF analysis of the underivatized synthetic peptide QAQVYPNRFPLWK yielded a rather complex



**Figure 3.** MALDI MS/MS spectrum of the modified tryptic peptides FALPQYLK ( $m/z$  1205.25) (a) and HQGLPQEVLNENLLR ( $m/z$  1943.96) (b). The derivatized precursors are labeled with an asterisk and the loss of the sulfonation label ( $-184$  Da) is indicated. The insets show an expanded view of the y-ion series obtained after preferential cleavage at the Xxx-Pro bond.

fragmentation spectrum wherein no dominant fragment ions were observed. The TOF/TOF fragment spectra obtained from the sulfonated homoarginine-terminated peptide is shown in Figure 5a. The y-ion fragment series  $y_6$ - $y_{13}$  is accompanied by the more intense ( $y_1 - 17$ )-ion series due to the loss of ammonia from the internal Arg. Due to preferential fragmentation at the Xxx-Pro amide bonds the spectrum contains predominantly the  $y_4$ - and the ( $y_8 - 17$ )-ions, whereas the  $y_3$ - and the  $y_7$ -ions are not or hardly observed. Due to the presence of a consecutive Pro-Asn sequence in the peptide care must be taken to interpretate this spectrum for assigning the correct sequence (Asn, 114 Da = [Pro, 97 Da + ammonia, 17 Da]). The same problem was encountered during TOF/TOF analysis of peptide TGPNLHGLFGRK (Thr28-Lys39, HHC). Due to the presence of Arg as the penultimate amino acid, the

fragmentation spectrum of the derivatized peptide ( $m/z$  1522.81) was dominated by the ( $y_{17}$ )-ion series ( $y_2$ - $y_{12}$ ). Together with the presence of a consecutive Pro-Asn ( $\Delta m = 17$  Da) in the sequence this might lead to an incorrect sequence assignment (Figure 5b). Fragmentation of the underivatized tryptic fragment HIQKEDVPSEK (His95-Arg105,  $\alpha$ -S1 casein) yielded a fragmentation spectrum with a b- and y-ion series in which the  $y_4$  and  $y_5$  were the most intense fragment ions (Figure 6a), indicating a preferential cleavage of the Val-Pro and the Asp-Val peptide bond, respectively. In the fragmentation spectrum of the derivatized peptide ( $m/z$  1563.78) the same ions were observed as the most intense fragment-ions (Figure 6b). Apart from the complete y-ion series, a more pronounced ( $y_i - 17$ )-ion series was observed due to the presence of an internal homo-Arg in the peptide.



**Figure 4.** MALDI MS/MS spectra of the underivatized (a) and modified (c) peptide APWFH-HQNGK. MALDI MS/MS spectra of the underivatized (b) and modified (d) peptide ELAQYNVEVH-PYTVRK. The insets show an expanded view of the y-ion series obtained after preferential cleavage at the Xxx-Pro bond. In each spectrum only a-, b-, and y-ions are labeled. The derivatized precursors are labeled with an asterisk. The loss of the sulfonation label (–184 Da) is indicated and, where appropriate, (y-17)-ions (number symbol) and internal fragment ions (open diamond) are shown as indicated.

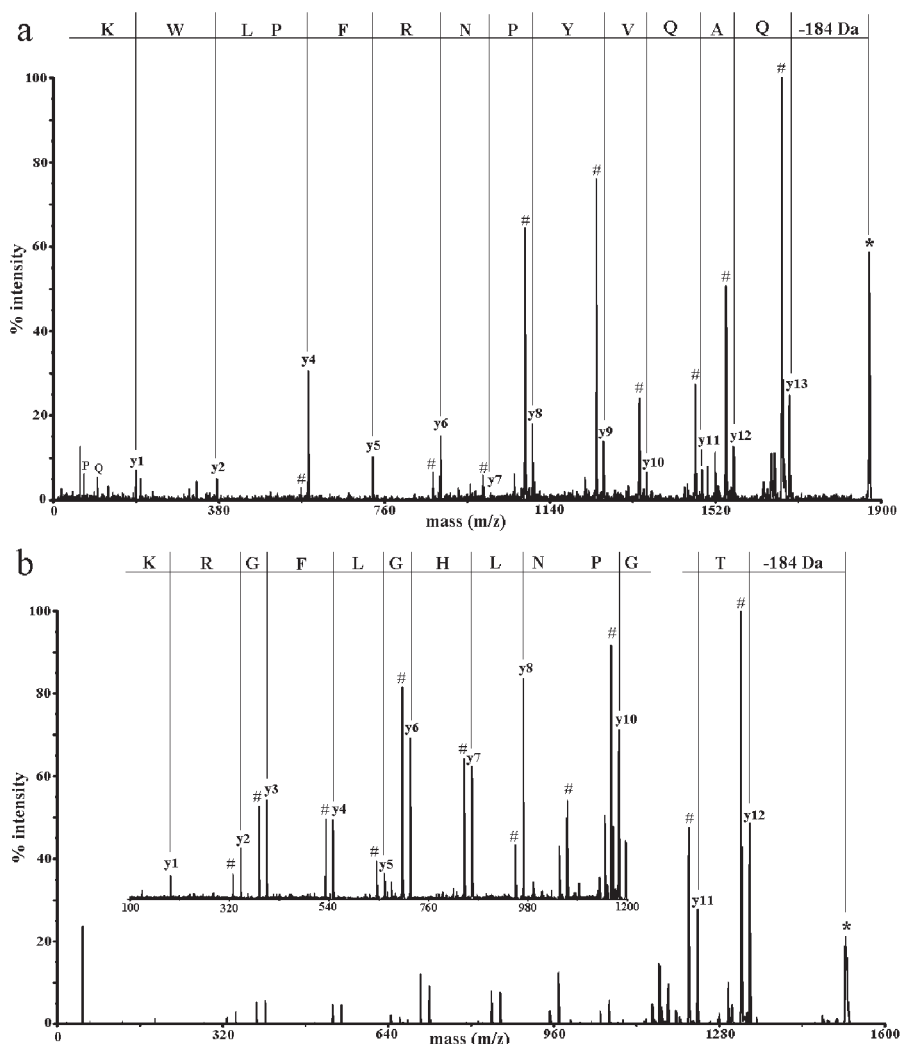
### De Novo Sequence Analysis

It is known that the introduction of a strong acidic group at the N-terminus of tryptic peptides facilitates protonation of backbone amide bonds, leading to extensive structure-specific fragmentation under PSD MALDI and electrospray MS/MS conditions [11, 12, 26]. In the present study, the effect of the N-terminal sulfonation on the CID TOF/TOF fragmentation pattern was investigated. The sulfonation reaction results in the modification of both the N-termini and the  $\epsilon$ -amino groups of lysine-containing peptides. For unprotected Lys-terminated peptides this will result in the formation of disulfonate derivatives. The formation of such derivatives was previously shown to be undesirable because they exhibit poor sensitivity in the positive-ion mode and relatively poor fragmentation under negative-ion analysis conditions. Negative-ion PSD spectra of several Lys-containing disulfonate derivatives showed both low-product ion yields and complex fragmentation patterns containing b- and y-type ions linked to the sulfonate group [26]. Therefore, this approach requires a preliminary modification of the  $\epsilon$ -amino group of lysine residues. It was demonstrated that, following guanidination of lysine  $\epsilon$ -amines, introduction of sulfonic acid groups to tryptic peptides is possible solely at the N-terminus. The guanidination, leading to C-terminal ly-

sines being converted into homoarginines (+42 Da), can be selectively and quantitatively performed with *O*-methylisourea at high pH and does not affect the peptide amino terminus or other side groups. The resulting homoarginine has an even higher pKa than arginine, and it has previously been shown that a rise of the basicity of lysines increases their relative abundances in MALDI mass spectra by enabling enhanced charge retention [27, 28]. Furthermore, it allows unambiguous differentiation of Lys (128.18 Da) and Gln (128.13 Da) by mass spectrometry.

CID experiments on several underivatized peptides with a prototype MALDI TOF/TOF mass spectrometer indicated that the nature of the MALDI-CID spectra is quite dependent upon the amino acid composition of the peptide, the peptide size, the matrix, and the collision gas used [22]. Different combinations of MALDI matrix and collision gas determine the amount of internal energy deposited by the MALDI and the CID processes. In our study, all experiments were performed using  $\alpha$ -cyano-4-hydroxycinnamic acid ( $\alpha$ -CHCA), the most commonly used matrix for peptide analysis.  $\alpha$ -CHCA is a relative hot matrix and is known to produce unimolecular decomposition (also called PSD) which is known to vary with the amino acid

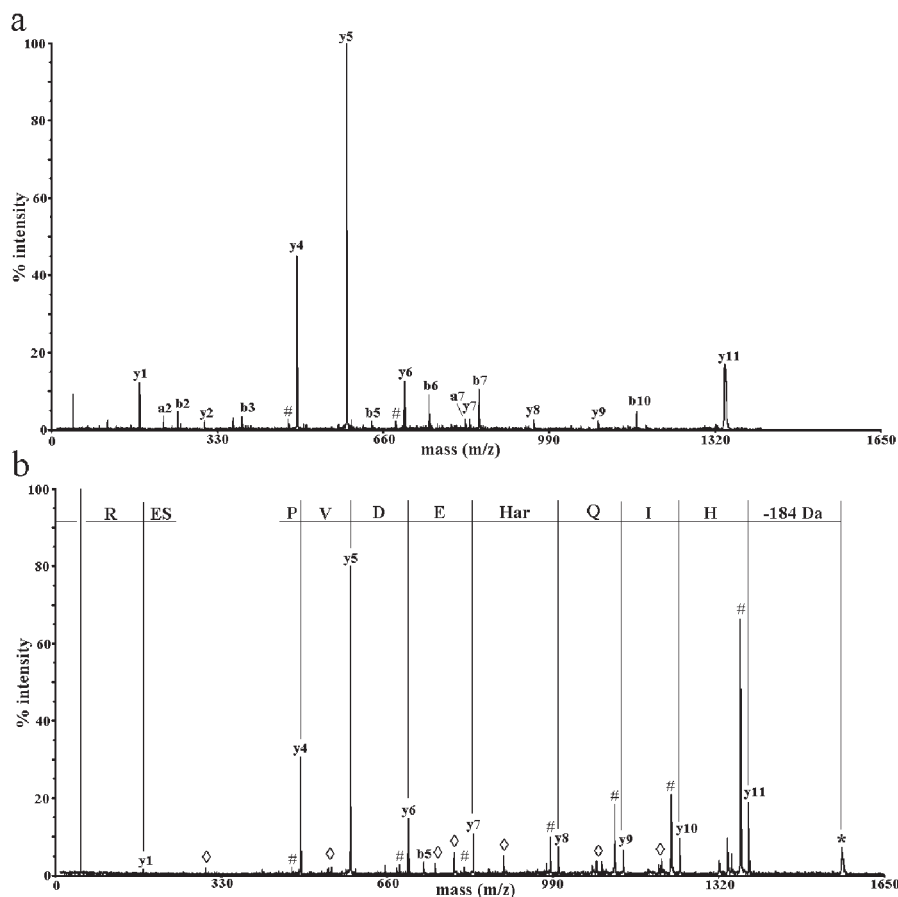




**Figure 5.** MALDI MS/MS spectrum of the modified peptides QAQVYPNRFPLWK ( $m/z$  1872.55) (a) and TGPLNLHGLFGRK ( $m/z$  1522.81) (b). The derivatized precursors are labeled with an asterisk and the loss of the sulfonation label ( $-184$  Da) is indicated. The inset shows an expanded view of the y-ion series obtained after preferential cleavage at the Xxx-Pro bond. (y-17)-ions (number symbol) are shown as indicated.

sequence. In a preliminary study, fragmentation spectra resulting from high- and low-energy CID were compared [29]. The authors concluded that the difference in fragmentation and the effect on database search results was surprisingly small. The major difference observed is the presence of high-energy fragment ions (w-ions) in the high-energy CID spectra of some peptides. When the collision induced dissociation mode (gas on, collision energy 0.5 to  $>3.5$  keV) is utilized, a larger number of low molecular weight fragments (immonium ions, internal fragments) have been observed [30]. However, it has also been shown that the use of high energy CID results in a loss of sequence information as the y-ion abundance decreases at both higher gas pressure and higher collision energy [31]. Therefore, we performed all fragmentation experiments with the collision energy set at 1 keV and no gas in the collision chamber (low-energy CID).

In the present case study, we used a set of peptides purified from proteolytically cleaved proteins with known sequences, rather than using test peptides. We found this set to be more representative for “real life” samples as some of the peptides had internal basic residues resulting from missed cleavages, some RPLC fractions contained mixtures (non-stoichiometric amounts), and most peptides had one or more amino acids in their sequence known to induce specific fragmentation. TOF/TOF analysis of underivatized peptides typically results in complex fragment spectra containing incomplete a-, b-, and y-ion series (Figure 4a and b, Figure 6a). The complexity of the spectra is often increased by the presence of internal fragment ions (Figure 1b). Similar complex fragmentation spectra have also been observed in fragmentation studies of singly charged precursor ions using a MALDI quadrupole TOF instrument [8]. The guanidinated lysine-terminated peptides showed better positive ion MALDI



**Figure 6.** MALDI MS/MS spectra of the underivatized (a) and modified (b) peptide HIQKEDVPSER. In each spectrum only a-, b-, and y-ions are labeled. The derivatized precursor is labeled with an asterisk. Where appropriate, (y-17)-ions (number symbol) and internal fragment ions (open diamond) are shown as indicated.

response than native lysine-terminated peptides. From MS/MS experiments of homoarginine-terminated peptides it was apparent that guanidination does not enhance peptide fragmentation efficiency (results not shown). This is consistent with earlier findings that increasing the basicity of the peptide increases the internal energy required for fragmentation [32].

Cleavage of amide bonds under low-energy collision activation conditions is generally thought to be initiated by migration of the charge from the initial site of protonation (e.g., the N-terminal amino group or the side chains of basic amino acids such as arginine, lysine, and histidine) to an amide carbonyl oxygen along the peptide backbone. This “mobile proton” termed hypothesis is one of the central tenets in peptide fragmentation mechanisms [33]. Most ESI MS/MS sequencing experiments generate doubly charged protonated peptides in which the first proton is sequestered at the most basic site of the peptide. This leads to a more labile bonding of the second proton that is more mobile to induce dissociation of other bonds, leading to more complete fragment ions series. Fragmentation of the peptide amide bond occurs by neighboring group attack from an adjacent nucleophilic amide carbonyl

moiety to yield complementary b- and/or y-ions (a charge-directed process). Unfortunately, doubly protonated molecules are not formed in high yield under MALDI conditions and, therefore, these labile ions are not available for sequencing studies using this technique. The addition of a sulfonic group to the N-terminus allows to create a peptide with an extra proton, but not a double charge. This strategy concedes that the basic C-terminal residue will be protonated under MALDI conditions. The strong acid is chosen such that it would remain deprotonated under MALDI conditions, counter-balancing the C-terminal positive charge. The additional proton, required to ionize the peptide for MS analysis, is more or less free to randomly protonate amide bonds, as the most basic site in the molecule is already occupied [34].

After guanidination and sulfonation, a contiguous series of y-ions in all of the fragmentation spectra was observed. The y-ion series could easily be interpreted (manually or by using an algorithm) facilitating de novo sequencing. Although there is a dramatic difference in speed, cost, and sensitivity, the ease by which de novo sequencing can be performed by MALDI TOF/TOF analysis of derivatized peptides can be compared

with classic Edman degradation analysis. In all fragment spectra an initial loss of the sulfonic acid derivative is observed ( $\Delta m = 184$  Da). A significant loss of the derivative was also observed by others using vacuum MALDI MS when the analyses were conducted using  $\alpha$ -CHCA as matrix [11]. The results indicated that some of the sulfonic acid-derivatized peptides had poorer positive-ion sensitivity than the corresponding native peptides and, after derivatization, four of the peptides were no longer observed in positive mode reflectron analysis. These fragments could be detected as their deprotonated ions when the analysis was performed in the negative mode (Table 1). However, selection of the corresponding protonated precursor ion for TOF/TOF analysis (positive mode) results in the formation of complete series of  $y$  fragment-ions (Figures 1d and 3a). Most likely, the protonated precursor is not detected due to metastable decomposition. Keough et al. [11] already noticed a decreased intensity of sulfonated peptides in the positive ion mode, compared to the negative ion mode and, apparently, some peptides show no signal above the noise level. However, the metastable decomposition yields excellent MS/MS spectra in the positive ion mode, as illustrated in this work.

#### *Amino Acid Dependent Specific Fragmentation*

For many years, low-energy CID has been the activation method of choice in the attempt to identify proteins by means of gas phase fragmentation of one or more of their peptides [35]. However, many factors can influence peptide dissociation in the gas phase, including the nature of the residues present, the charge state of the precursor ion, and the size and conformation of the peptide. In some instances, a number of these factors can lead to enhanced or specific cleavage at certain peptide bonds, limiting the information obtained from the MS/MS spectrum. It has long been recognized that mass spectrometric peptide sequencing techniques would be improved if the relative fragmentation efficiencies of the cleavage sites in a peptide could be predicted. Such studies included the identification of residues that enhance specific ion fragmentation pathways. Most of these studies have focused on fragment spectra resulting from MS/MS of doubly protonated tryptic peptides. In this context, a highly efficient fragmentation of backbone amide bonds on the N-terminal side of proline residues has been described by several authors [36, 37]. Other investigators have elucidated dissociation pathways that are promoted by acidic residues [38]. Preferential cleavages at Pro and His [39] and at the acidic residues Asp and Glu [40] have been indicated. The selective cleavage of protonated peptides, derivatized with a fixed-charge, containing internal Asp or His residues has been reported [13, 41]. The influence of the basic residue content on the fragmentation behavior has also been studied [25, 42–44].

Sophisticated algorithms (e.g., SEQUEST, Mascot) have been developed for identifying proteins from peptide MS/MS data. Peptides are hereby identified by correlating the uninterpreted MS/MS spectra with simulated product ion spectra using a relative simple set of previously defined parameters regarding the expected fragmentation behavior of protonated peptide ions. However, the current understanding of the fragmentation mechanisms is still insufficient to ensure a high correlation between theoretically predicted MS/MS spectra and experimental results. Furthermore, the fact that selective/enhanced cleavage does not always occur might have discouraged investigators from including them in the existing computer-based interpretation of peptide MS/MS spectra. Therefore, any improved understanding of the enhancement or absence of certain fragment ions in MS/MS experiments provides additional and better predictive rules for the interpretation of peptide MS/MS spectra. Recently, several groups have systematically examined databases of tryptic MS/MS spectra to evaluate the effects of specific residues or charge states on the peptide fragmentation and fragment ion intensity [38, 39, 45–47].

In a preliminary study, in which MALDI TOF/TOF was used to produce MS/MS spectra of singly charged underivatized peptides, dominant product ions resulting from Xxx-Pro fragmentation were observed [30]. These authors studied the effects of the MALDI matrix and the CID gas on the fragmentation efficiency of larger peptides and small proteins. They observed that fragmentation becomes more sequence specific, with a cleavage selectively at the N-terminus of Pro as well as at the C-terminus of the acidic residues, as the molecular weight of the peptide is increased ( $> 4000$  Da). MS/MS spectra generated by MALDI quadrupole TOF mass spectrometry of singly charged peptides also showed preferential fragmentation at the N-terminal bond of Pro and at the C-terminal bond of the acidic residues Asp and Glu [8]. The same observation was made during CID analysis of singly charged peptide ions using an in-house assembled MALDI-quadrupole ion trap [9]. Several of the peptides used in our study contained Pro and/or acidic residues and some had, due to missed cleavages, internal Lys or Arg (Table 1). TOF/TOF analysis of the underivatized peptides was performed to explore any known or novel preferential fragmentation patterns. The influence of the sulfonation reaction on the observed preferential fragmentation, and hence the possibility to obtain complete  $y$ -ion series, was determined by comparing the TOF/TOF spectra before and after derivatization.

#### *Fragmentation of Peptides Containing Proline Residues*

When the ionizing proton is freely “mobile”, enhanced cleavage is commonly observed at Xxx-Pro bonds, presumably due to higher local proton affinity of the

proline imide bond compared to a conventional amide bond. During PSD experiments with derivatized peptides, Keough et al. observed enhanced fragmentation at Pro residues and a reduced abundance in fragmentation at the C-terminal side of Pro [26]. Therefore, a Pro residue near the N-terminus of a peptide can limit the amount of sequence information that can be derived from the peptide. In our study, TOF/TOF analysis of peptides containing an internal Pro indicated that  $y$ -ions resulting from cleavage on the N-terminal side of Pro are enhanced while  $y$ -ions resulting from cleavage on the C-terminal side of Pro are less abundant or almost completely depleted (Table 1). MS/MS analysis of Pro-containing underivatized peptides indicated a preferential cleavage at the Xxx-Pro bond except for peptides APWFHHQNGK and TGPNLHGLFGRK. MS/MS analysis of the underivatized fragment QAQVYPNRFPLWK, having two internal Pro residues, indicated a cleavage only at the second Pro ( $y_4$ ) (Figure 5a). In contrast, upon complete derivatization, all peptides showed preferential cleavage at all Xxx-Pro bonds, with the exception of the peptide TGPNLHGLFGRK from HHC (Figure 5b). In a recent study, the unusual fragmentation effect in tandem mass analysis of doubly charged Pro-containing peptides was systematically analyzed [37]. From this study it became apparent that ions formed at positions C-terminal to Pro, where Xxx is any amino acid, produce low Pro-Xxx cleavage. In contrast, preferential cleavage N-terminal to Pro occurred with strong ions formed at Xxx-Pro when Xxx was His, Asp, Ile, Leu, and weak or no ions formed when Xxx was Gly or Pro as is the case for the peptide TGPNLHGLFGRK.

#### *Fragmentation of Peptides with Internal Basic Residues (Arg/homo-Arg/His)*

Peptides resulting from incomplete protein digests show differences in their fragmentation behavior due to the presence of internal basic residues. This type of residues can affect fragmentation in a variety of ways. Their presence in a peptide raises the minimum energy requirement for fragmentation [32]. Their side chains may be the sites in a peptide for which protons show the greatest affinity, in some cases resulting in proton sequestration and enabling charge-remote fragmentation mechanisms [13, 42]. The protonated side chains of His and Arg may also cleave adjacent amide bonds to form b-ions of non-oxazolone structure [33]. Tabb et al. recently described a statistical analysis of the influence of basic residue position and identity on fragment ion intensities in tandem mass spectra of non-tryptic peptides [25].

After derivatization with the commercially available CAF-reagent, incomplete  $y$ -ion series were observed during PSD experiments of peptides containing internal Arg [19]. In contrast to these results, we observed that TOF/TOF analysis of derivatized peptides with an

internal Arg or homo-Arg results in the formation of a complete  $y$ -ion series. However, in all these spectra the  $y$ -ion series were accompanied by a more abundant ( $y_i - 17$ )-ion series due to the neutral loss of  $\text{NH}_3$ . Although this ( $y_i - 17$ ) series shows up in the fragment spectra until the basic residue is cleaved off (Figures 1c, 2d, 4d, 5a and b, 6b), its presence does not interfere with the sequence interpretation, so that the complete sequence of all the peptides could be determined (Table 1). However, as illustrated in Figure 5, the neutral loss ( $\Delta m = 17\text{Da}$ ) might prevent a correct sequence interpretation for peptides having a Asn-Pro bond in their sequence ( $\Delta m \text{ Asn/Pro} = 17 \text{ Da}$ ). The presence of a ( $y_i - 17$ )-ion series, twice as large as the corresponding  $y_i$ -ions, was also observed by Keough et al. during PSD analysis of derivatized peptides containing two arginine residues [11].

Recently, it has also been noted that product ion spectra may be influenced by the presence of histidine in a peptide [25, 42]. Data mining of a set of peptide spectra showed that histidine is the amino acid residue most likely to give preferential cleavage at its C-terminal side in doubly protonated tryptic peptides [39]. In contrast, TOF/TOF analysis of several underivatized peptides containing internal His residues shows a preferential cleavage at Xxx-His (Figure 4a and b). More abundant  $y$ -ions were observed in the MS/MS spectra of EGIHAQQK ( $y_5$ ), TVYQHQQ ( $y_3$ ), ELAQYNVEVHPYTVRK ( $y_7$ ), and APWFHHQNGK ( $y_5$  and  $y_6$ ). This preferential cleavage was not observed, or less abundantly, in the fragment spectra of the derivatized peptides (Figure 2b and c, Figure 4c and d).

#### *Fragmentation of Peptides Containing Acidic Residues*

Some MS/MS spectra of native peptides containing Asp or Glu residues are dominated by cleavage on the C-terminal side of the acidic residues. The fragment spectra of the underivatized peptides YLGYLEQLLR, ELAQYNVEVHPYTVRK, and HQGLPQEVLENLLR were dominated by the  $y_{4-7}$ ,  $y_8/y_{15}$ -ions (Figure 4b) and the  $y_4/y_8$ -ions, respectively, whereas the fragment spectrum of the native peptide HIQKEDVPSEK showed a dominant  $y_5$ -ion (Figure 6a). In contrast, some of the peptides resulting from endoprotease Lys-C cleavage, or containing an internal Lys (ALNEINQFYQK, LTEEEKNR and TGQAPGFTYTDANK), did not show preferential cleavage at the acidic residues. Preferential fragmentation at Asp-Xxx bonds of underivatized peptides on a TOF/TOF instrument has also been observed by Lin et al. [24]. Tandem mass spectra of the derivatized peptides containing Asp or Glu show a more uniform fragmentation along the peptide backbone. Only one peptide, HIQKEDVPSEK, showed a slightly more abundant cleavage C-terminal to Asp ( $y_5$ ) after derivatization (Figure 6b). A more uniform  $y$ -ion series in the MS/MS of sulfonic acid derivatized tryptic



peptides spectra using atmospheric pressure MALDI ion trap MS was also observed by others [48].

C-terminal cleavage at aspartic acid residues, a low-energy fragmentation pathway, was first observed in peptides containing Asp-Pro and Asp-Xxx bonds [38]. Selective gas phase cleavage at the peptide bond C-terminal of an Asp residue in protonated peptides has been investigated by several research groups. Recent investigations show that when the ionizing protons are not mobile, i.e., when they are sequestered by Arg, cleavage catalyzed by the acidic hydrogen of the Asp side chain becomes pronounced [33, 44]. This is supported by data for related peptides that have been derivatized to add a fixed charge to them instead of a proton; even those peptides that contain no added proton fragmented selectively at the aspartic acid [13]. Inspection of the singly charged spectra of 10 tryptic peptides containing Arg plus Asp and/or Glu showed that all of them fragment selectively at the acidic residues [33]. However, peptides containing lysine did not show the same trend, thus spectra of singly charged peptides that contain Lys and/or Asp/Glu showed no enhancement of cleavage at acidic residues [43]. When protons in excess of the number of Arg are present, enhanced/selective cleavages of Asp-Xxx bonds are not expected [13]. This is the case for the N-sulfonated peptides, as the introduction of a negative charge is counterbalanced by a second mobile proton. This proton is more or less free to randomly ionize the peptide backbone amide groups independently of the presence of acidic residues.

## Conclusions

Although de novo sequencing of underivatized peptides using MALDI TOF/TOF has recently been demonstrated by Yergey et al. [49], the interpretation of fragment spectra from peptides originating from unknown proteins strongly depends on the use of automated search routines or on manual interpretation. TOF/TOF fragmentation analysis of underivatized peptides yields multiple, incomplete fragment ion series, which are often difficult to interpret. There is no guarantee that the information in such MS/MS spectra is sufficient for peptide identification. As proposed by Keough et al., the addition of a sulfonic acid group to the N-terminus of a peptide promotes efficient charge-site-initiated fragmentation of backbone amide bonds [11]. Upon guanidination and sulfonation, all peptides resulting from a proteolytic cleavage with trypsin or endoprotease Lys-C show complete y-ion series in their fragment spectra. By simple manual calculation of the differences between the adjacent y-ion fragments, or by using suitable software, the amino acid sequence can readily be interpreted. In the preceding guanidination reaction, Lys is converted to homo-Arg and therefore can easily be differentiated from Gln; unfortunately, Leu and Ile cannot be distinguished. As the mass of the parent peptide is known from the initial reflectron

analysis, the final assignment can be verified. Furthermore, the use of a TOF/TOF instrument provides improved ion mass measurement accuracy (results not shown). After derivatization, some of the peptides were solely observed in negative ion mode analysis (Table 1). However, selection of the corresponding precursor in positive ion mode analysis allowed to obtain the full y-ion series. The difference in fragmentation behavior between sulfonated and underivatized peptides has important consequences for protein identification via database searches. Search algorithms, such as FASTS and FASTF, which use multiple peptide sequences to identify homologous sequences in protein or DNA databases have recently been described [50]. Given the size and growth of the current databases, many proteins more than likely already have homologues in a database [7]. Database searching with MS-derived de novo peptide sequences allow proteomic identification of proteins from organisms whose genomes have not been sequenced.

MS/MS analysis of several underivatized peptides indicated a preferential cleavage at Xxx-Pro bonds and C-terminal of the acidic residues. These results are supported by previously described fragmentation mechanisms, supporting the mobile proton mode. In contrast to previously published results [41], a preferential fragmentation at several Xxx-His bonds was observed in the fragment spectra of the underivatized peptides. Upon guanidination and sulfonation, preferential fragmentation at His and the acidic residues was no longer observed in the MS/MS spectra. In contrast, preferential cleavage at Xxx-Pro was still observed and sometimes even more pronounced in the MS/MS spectra of the derivatized peptides. Although the y-ions resulting from the Pro-Xxx were sometimes less intense, the complete sequence could be determined de novo for most peptides. The 14 peptides constitute 150 amino acids, of which 142 could be determined (Table 1). The fragment spectra of derivatized peptides containing an internal Arg or homo-Arg all showed y-ion series accompanied by a more dominant ( $y_i - 17$ )-ion series. As proteomic strategies are becoming increasingly reliant on the use of automated database search algorithms, incorporation of "fragmentation rules", such as the observed preferential cleavage at Xxx-Pro peptide bonds, into the database search algorithms will aid in the development of more effective tools for high throughput protein identification. Furthermore, the occurrence of "non-sequence" specific ion fragments, such as the neutral loss of ammonia from peptides with internal Arg or homo-Arg, can be used to improve predictive models of peptide fragmentation for de novo sequence analysis.

## Acknowledgments

J.V.B. is indebted for research grant 3G031202. B.S. is a postdoctoral fellow of the Fund for Scientific Research-Flanders (F.W.O.-



Vlaanderen, Belgium). K.S. is funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (I.W.T.-Vlaanderen).

## References

- Wilm, M.; Shevchenko, A.; Houthaave, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379*, 466–469.
- Shevchenko, A.; Loboda, A.; Ens, W.; Standing, K.G. MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal. Chem.* **2000**, *72*, 2132–2141.
- Arnott, D.; Henzel, W.J.; Stultz, J.T. Rapid identification of comigrating gel-isolated proteins by ion trap-mass spectrometry. *Electrophoresis* **1998**, *19*, 968–980.
- Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10313–10317.
- Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.
- Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevchenko, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal. Chem.* **2001**, *72*, 1917–1923.
- Wattenberg, A.; Organ, A. J.; Schneider, K.; Tyldesley, R.; Bordoli, R.; Bateman, R. H. Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 772–783.
- Zhang, W.; Krutchinsky, A. N.; Chait, B. T. “De novo” peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: A preliminary study. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1012–1021.
- Roth, K. D. W.; Huang, Z. H.; Sadagopan, N.; Watson, J. T. Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom. Rev.* **1998**, *17*, 255–274.
- Keough, T.; Youngquist, R. S.; Lacey, M. P. A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7131–7136.
- Bauer, M. D.; Sun, Y.; Keough, T.; Lacey, M. P. Sequencing of sulfonic acid derivatized peptides by electrospray mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 924–929.
- Gu, C.; Tsapraillis, G.; Brecci, L.; Wysocki, V. H. Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal. Chem.* **2000**, *72*, 5804–5813.
- Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* **2000**, *72*, 4047–4057.
- Summerfield, S. G.; Steen, H.; O'Malley, M.; Gaskell, S. J. Phenyl thiocarbonyl and related derivatives of peptides: Edman chemistry in the gas phase. *Int. J. Mass Spectrom.* **1999**, *188*, 95–103.
- Beardsley, R. L.; Reilly, J. P. Fragmentation of amidated peptide ions. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 158–167.
- Beardsley, R. L.; Reilly, J. P. Quantitation using enhanced signal tags: A technique for comparative proteomics. *J. Proteome Res.* **2003**, *2*, 15–21.
- Keough, T.; Lacey, M. P.; Youngquist, R. S. Solid-phase derivatization of tryptic peptides for rapid protein identification by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1003–1015.
- Hellman, U.; Bhikhabhai. Easy amino acid sequencing of sulfonated peptides using post-source decay on a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer equipped with a variable voltage reflector. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1851–1859.
- Marekov, L. N.; Steinert, P. M. Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* **2003**, *38*, 373–377.
- Wang, D.; Kalb, S. R.; Cotter, R. J. Improved procedures for N-terminal sulfonation of peptides for matrix-assisted laser desorption/ionization post-source decay peptide sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 96–102.
- Medzihradsky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **2000**, *72*, 552–558.
- Bienvenut, W. V.; Deon, C.; Pasquarello, C.; Campbell, J. M.; Sanchez, J.-C.; Vestal, M. L.; Hochstrasser, D. F. Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* **2002**, *2*, 868–876.
- Lin, M.; Campbell, J. M.; Mueller, D. R.; Wirth, U. Intact protein analysis by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1809–1814.
- Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. *Anal. Chem.* **2004**, *76*, 1243–1248.
- Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization procedures to facilitate de novo sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348–2356.
- Beardsley, R. L.; Karty, J. A.; Reilly, J. P. Enhancing the intensities of lysine-terminated tryptic peptide ions in matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147–2153.
- Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. Increased sensitivity of tryptic peptide detection by MALDI-TOF mass spectrometry is achieved by conversion of lysine to homoarginine. *Anal. Biochem.* **2000**, *287*, 110–117.
- Sumpton, D. P.; Thomas, J. R.; Thomas-Oates, J. *Abstracts of the 16th International Mass Spectrometry Conference*; Edinburgh, September, 2003.
- Walker, A. K.; Andrews, P. C. *Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics*; Montreal, June, 2003.
- Campbell, J. M. *Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics*; Montreal, June, 2003.
- Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374.

33. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Brechi, L. A. Mobile and localized protons: A framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.
34. Keough, T.; Youngquist, R. S.; Lacey, M. P. Sulfonic acid derivatives for peptide sequencing by MALDI MS. *Anal. Chem.* **2003**, *156A*–165A.
35. Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.
36. Loo, J. A.; Edmonds, C. G.; Smith, R. D. Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Anal. Chem.* **1993**, *65*, 425–438.
37. Brechi, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-terminal to proline: Analysis of a database of peptide tandem mass spectra. *Anal. Chem.* **2003**, *75*, 1963–1971.
38. Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* **1993**, *65*, 3015–3023.
39. Tabb, D. L.; Smith, L. L.; Brechi, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. *Anal. Chem.* **2003**, *75*, 1155–1163.
40. Bailey, T. H.; Laskin, J.; Futrell, J. H. Energetics of selective cleavage at acidic residues studied by time- and energy-resolved surface-induced dissociation in FT-ICR MS. *Int. J. Mass Spectrom.* **2003**, *222*, 313–327.
41. Tsaprailis, G.; Nair, H.; Zhong, W.; Kuppannan, K.; Futrell, J. H.; Wysocki, V. H. A mechanistic investigation of the enhanced cleavage at histidine in the gas-phase dissociation of protonated peptides. *Anal. Chem.* **2004**, *76*, 2083–2094.
42. Huang, Y.; Wysocki, V. H.; Tabb, D. L.; Yates, J. R. The influence of histidine on cleavage C-terminal to acidic residues in doubly protonated tryptic peptides. *Int. J. Mass Spectrom.* **2002**, *219*, 233–244.
43. Gu, C. G.; Somogyi, A.; Wysocki, V. H.; Medzihradsky, K. F. Fragmentation of protonated oligopeptides XLDVLQ (X = L, H, K, or R) by surface induced dissociation: Additional evidence for the mobile proton model. *Anal. Chim. Acta* **1999**, *397*, 247–256.
44. Tsaprailis, G.; Somogyi, A.; Nikolaev, E. N.; Wysocki, V. H. Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides. *Int. J. Mass Spectrom.* **2000**, *195/196*, 467–479.
45. Shütz, F.; Kapp, E. A.; Simpson, R. J.; Speed, T. P. Deriving statistical models for predicting peptide tandem MS product ion intensities. *Biochem. Soc. Transact.* **2003**, *31*, 1479–1483.
46. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658.
47. Kapp, E. A.; Shütz, F.; Reid, G. E.; Eddes, J. S.; Moritz, R. L.; O'Hair, R. A. J.; Speed, T. P.; Simpson, R. J. Mining a tandem mass spectrometry database to determine the trends and global factors influencing peptide fragmentation. *Anal. Chem.* **2003**, *75*, 6251–6264.
48. Keough, T.; Lacey, M. P.; Strife, R. J. Atmospheric pressure matrix-assisted laser desorption/ionization ion trap mass spectrometry of sulfonic acid derivatized tryptic peptides. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2227–2239.
49. Yergey, A. L.; Coorssen, J. R.; Backlund, P. S.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. De novo sequencing of peptides using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 784–791.
50. Mackey, A. J.; Haystead, T. A. J.; Pearson, W. R. Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences. *Mol. Cell. Proteomics* **2002**, *1*, 139–147.