Sandrine Uttenweiler-Joseph[1]*
Gitte Neubauer[1]
Savvas Christoforidis[2]
Marino Zerial[2]
Matthias Wilm[1]

[1]European Molecular Biology Laboratory, Heidelberg, Germany
[2]Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany

Automated *de novo* sequencing of proteins using the differential scanning technique

Despite the progress in genomic DNA sequencing *de novo* sequencing of peptides is still required in a biological research environment since many experiments are done in organisms whose genomes are not sequenced. A way to unambiguously retrieve a peptide sequence from a tandem mass spectrum is to assign the correct ion type to the fragments. Here we describe a method which improves the specificity in y-ion assignment throughout the spectrum. The differential scanning technique requires that the peptides are partially ^{18}O labelled at their *C*-terminus and that two fragment spectra are acquired for each peptide, one selecting the ^{16}O/^{18}O isotopic cluster and a second fragmenting only the ^{18}O labelled ions. When the spectra are acquired with a quadrupole time of flight mass spectrometer y-ions can be very specifically filtered from the spectrum using a computer algorithm. Partial or complete peptide sequences can be assigned automatically simply by finding the most abundant series of fragments spaced by amino acid residue masses. This method was used extensively in a project investigating vesicular transport in bovine brain cells. Human or mouse homologues to the bovine proteins were found in EST databases facilitating rapid cloning of the human homologues.

Keywords: Protein *de novo* sequencing / Mass spectrometry / Isotopic labelling / ^{18}O PRO 0052

1 Introduction

Mass spectrometry is the preferred tool to identify proteins in databases [1–5]. Protein de *novo* sequencing with MS is a more difficult but still required task. Biological experiments are often conducted in organisms which enable the intended research by some specific properties. The genome of many of these model organisms are not sequenced and are not expected to be sequenced in the near future. Reading sequences from fragment spectra of peptides in a reliable way, without confirmation from existing databases, is a difficult task. The overall fragmentation pattern of a peptide depends on its primary structure. Some peptides produce long consistent y- or b-ion series while others fragment only in limited regions of their sequence. This behaviour is not readily recognisable when inspecting a tandem mass spectrum and can lead to false interpretations. Any method developed for *de novo* sequencing should take into account that the sequence of a given peptide may not be completely reflected in its fragment spectrum.

There are different approaches to sequence peptides *de novo* using MS/MS. Some software algorithms try to predict the amino acid sequence of peptides from their native fragment spectrum [6–8]. However, these algorithms tend to require that the entire sequence can be deduced from the fragmentation pattern of the peptide. They use for instance the molecular mass of the peptide and generate a sequence which corresponds to the complete mass. They do not consider sufficiently well that stable regions of the peptide may not have been fragmented so that their sequence information is not contained in the spectrum. Because no additional, chemically encoded information about the ion type is available, it is difficult to recognize whether the fragment spectrum under investigation belongs to a peptide which fragments according to the fragmentation rules used to build the interpretation algorithm, or to a peptide which produced an uncommon fragmentation pattern. Since no additional information about the overall fragmentation behaviour of an unmodified peptide can be obtained from its fragmentation pattern these algorithms cannot differentiate between spectra which they can interpret successfully and those they cannot. This can lead to wrong sequence interpretations.

To change the fragmentation behaviour Pappin [9, 10] proposes to introduce a basic group at the *N*-terminus of all the peptides. B-ions are more favored and the y-ion

Correspondence: Matthias Wilm, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
E-mail: wilm@embl-heidelberg.de
Fax: +49-0-6221-387-306

Abbreviations: nrdb, Nonredundant database; **Q-TOF**, Quadrupole-time of flight; **SNARE**, Soluble *N*-ethyl-maleimide-sensitive factor attachement protein receptor

* Current address: Sanofi-Synthélabo, Labège Innopole, Voie 1, BP 131, 31616 Labege Cedex, France

1615-9853/01/0505–668 $17.50+.50/0

series is longer as for the native peptide. The simultaneous appearance of b- and y-ions helps in the interpretation of the spectrum.

A third approach is to chemically modify the peptides to identify some fragment ion types by the associated mass shifts. This was first achieved by esterification of all acidic groups including the *C*-terminus of a peptide and later by introducing an end-standing ^{18}O isotope upon digestion [11–13]. For the ^{18}O labelling the protein is digested in 50% ^{18}O water so that all cleaved peptides appear as isotopic doublets, 50% with a ^{16}O isotope and 50% with an ^{18}O isotope at their *C*-termini. When the $^{16}O/^{18}O$ isotopic doublet is selected for fragmentation all *C*-terminal fragment ions are represented by the same $^{16}O/^{18}O$ isotopic pattern in contrast to all other fragment ions. This isotopic pattern identifies the y-ions and can be sufficient to call a partial or even a complete peptide sequence from the spectrum [13]. However, it is often not possible to identify all y-ions due to overlapping chemical noise ions in the spectrum. The differential scanning method is an extension to the partial isotopic labelling approach and improves the y-ion recognition. The method has been used on a routine basis in our laboratory to sequence bovine proteins involved in vesicular transport. Mouse or human homologues in EST or protein databases were found with these peptides.

2 Materials and methods

Chemicals were of HPLC grade. ^{18}O water was purchased from Phychem (Düren, Germany) and purified by distillation before use. Trypsin was unmodified bovine trypsin from Boehringer (Mannheim, Germany).

2.1 Protein purification

Glutathione-S-transferase (GST)-Rab5 affinity chromatography was performed as described [14].

2.2 Sample preparation

All proteins were separated on 1-D SDS gels and stained with Coomassie blue or silver [15]. Bands were cut out, reduced, alkylated and in-gel digested using trypsin (15). The digestion buffer contained 33% or 50% ^{18}O water to label the peptides with ^{18}O at the *C*-terminus. Under the conditions chosen no double incorporation of ^{18}O atoms in peptides was observed (< 5%) (13]. Double incorporation of ^{18}O atoms occurs in peptides ending in an arginine if the digestion is performed in 100% ^{18}O water (unpublished results). The degree of incorporation of ^{18}O into peptides as measured from their tandem mass spectra varied between 33% and 50% depending on the amount of ^{18}O water taken during the digest and the purity of the ^{18}O water. Peptides were extracted and dried down. For analysis the peptide mixture was dissolved in 1 μL 80% formic acid and rapidly diluted with 9 μL water. The peptide mixture was desalted on self assembled Poros™ R2 and R3 columns and eluted twice using 0.5 μL 60% methanol, 5% formic acid directly into a nano-electrospray needle [15, 16]. The nano-electrospray needle was mounted on a Q-TOF1 instrument (Micromass, Manchester, UK) and different peptides were investigated in a sequential manner.

2.3 Mass spectrometry

All peptides were analyzed with the Q-TOF1 mass spectrometer equipped with the nano-electrospray ion source (Micromass). To avoid fast evaporation of the sample from the nano-electrospray needle the interface temperature was reduced to 50°C. The collision energy was adjusted individually for each peptide. For every MS/MS analysis the transmission window was adjusted to select the desired isotopic pattern by measuring the transmitted isotopes with the collision energy switched off. Two fragment spectra were acquired *per* peptide, one transmitting the entire $^{16}O/^{18}O$ isotopic pattern into the collision zone and a second with only the ^{18}O isotopes selected. To achieve this the selection for the precursor was shifted to the appropriate higher *m/z* value. Depending on the total mass of the peptide this was either the ion corresponding to the (m+2), the (m+3) or the (m+4) ion with *m* the monoisotopic mass of the peptide. A shift in ion transmission of more than 2 Da was required for peptides larger than 1600 Da to exclude ^{16}O labelled ions which contain ^{13}C or ^{15}N isotopes.

2.4 Data processing

The y-ions scoring filter was programmed using IGOR Pro (Wavemetrics, Lake Oswego, USA). Spectra are exported from MassLynx (Micromass) in ASCII format, processed and reimported. Every peak in the $^{16}O/^{18}O$ spectrum is scored for the likelihood that it is the first isotope of a $^{16}O/^{18}O$ labelled fragment. Two criteria are evaluated: whether the following isotopes correspond to the expected intensities of a $^{16}O/^{18}O$ labelled peptide of the same mass and whether the ^{16}O isotope is suppressed in the second fragment spectrum. After evaluation of every peak the $^{16}O/^{18}O$ spectrum is multiplied with the two scoring functions to generate a y-ion filtered spectrum. The y-ion filtered spectrum is used to read out the amino acid sequence automatically. To display a scored spectrum the dynamic range of the scored spectrum is reduced to about 1000. The displayed intensity I_d is calculated from the scored intensity I_s using the following formula:

$$I_d = (I_s + 1)^{\left(\frac{\log(1000)}{\log\left(\frac{Y_{max}}{Y_{min}}\right)}\right)} - 1 \qquad (1)$$

Y_{max} is the maximal intensity of the scored spectrum, Y_{min} is the minimal intensity which is above the noise level. The y-ion scored spectrum can be reimported into MassLynx.

2.5 Sequence determination

The sequences are determined automatically from the y-ion scored spectrum. In the first step of the interpretation the most abundant ion is detected and considered a y-ion. The second most abundant ion is taken as another y-ion if it can be embedded into one series of ions spaced by amino acid residue masses with the most abundant ion. This process continues until a fragment ion is detected which cannot be aligned with the already selected ions. At this point the selection routine has reached an intensity level where the y-ions in the spectrum are so weak that their positive score could not enhance them enough to get a bigger intensity than other more intense fragment ions which partially fulfil the scoring criteria. The so called "obvious y-ions" are selected. These ions will be part of every sequence proposal generated by the programme at the end. In a second step of the sequence determination algorithm, the remaining mass distances between selected y-ions are filled with sequence proposals by testing all possible amino acid combinations which breech the gap. Extra care is invested to make certain that the gaps are not bigger than nine amino acids in length. The scoring value of every sequence is the sum of all the scored ion intensities. Finally, the 200 best scored sequences are edited to evaluate the positioning of prolines and to test whether two adjacent amino acids should, in fact, be considered to be the amino acid with the combined mass. In a last step the flow of the intensities of the y-ions in the scored spectrum is evaluated. If the intensities suddenly collapse without the presence of a proline or an acidic residue and do not recover immediately, the sequence is marked as uncertain.

3 Results and discussion

3.1 The differential scanning method

De novo peptide sequencing is still required on a daily basis in a biological research environment. Specific biological systems are often studied in dedicated organisms whose genome is not sequenced. The protein sequences from these organisms are sufficiently different from genomically sequenced organisms, so that identification tools fail in many instances [17, 18]. In these cases a *de novo* sequencing approach is required. When the fragmentation behaviour of a peptide is not changed by modifying its primary structure a single tandem mass spectrum of an unlabelled peptide often does not contain sufficient information to deduce its sequence in an unambiguous way. There is no software tool able to recognisze to a satisfactory degree which fragment spectrum can be reliably interpreted for *de novo* sequencing and which cannot. Therefore, we adopted methods that supply the investigator with additional primary structure based information to read out the peptide sequence from its fragment spectrum.

When using partial ^{18}O *C*-terminal labelling y-ion recognition is based on its specific ^{16}O/^{18}O isotopic pattern [13]. However partial ^{18}O labelling has limitations. Particularly in the *m/z* range below the precursor ion the ^{16}O/^{18}O isotopic pattern is often very difficult to recognise. Y-ions can have such a low abundance that peaks may consist of only a few ions. This limits the faithful representation of isotopic distributions so that the ^{16}O and ^{18}O isotopes may be represented unequally. Additionally, this part of the spectrum is often densely populated by fragment and chemical noise ions. They can overlap with y-ions and distort the ^{16}O/^{18}O isotopic pattern or they can mimic their isotopic distribution and the correct y-ion cannot be found (Fig. 1).

In these cases the differential scanning approach can supply additional information which allows unambiguous identification of the y-ions in the spectrum and call additional sequence. The basic principle of the method is that two fragment spectra are acquired from the same peptide, one selecting the entire ^{16}O/^{18}O isotopic envelope and the second selecting only the ^{18}O labelled peptide ions. Y-ions are represented in a different way in the two spectra; in the first as an ^{16}O/^{18}O cluster and in the second only as an ^{18}O ion. A similar method was used by Takao *et al*. [19]. They selected individual isotopes of partially ^{18}O labelled peptides for the generation of fragment spectra using a sector field instrument. Y-ions were identified by their different isotopic representation in both fragment spectra. The advantage of using a Q-TOF instrument and selecting isotopic clusters instead of single isotopes is that the ion transmission is much higher. The higher sensitivity makes the method applicable to standard biological experiments where only low amounts of separated proteins are available (1 pmol or less of protein purified on SDS-PAGE).

The exclusive selection of the ^{18}O isotopes of the peptide can be done by simply shifting the precursor ion selection to a higher *m/z* value (Fig. 2). The interesting observation is that a ^{18}O only fragment spectrum can be acquired without increasing the resolution of the selecting quadrupole.

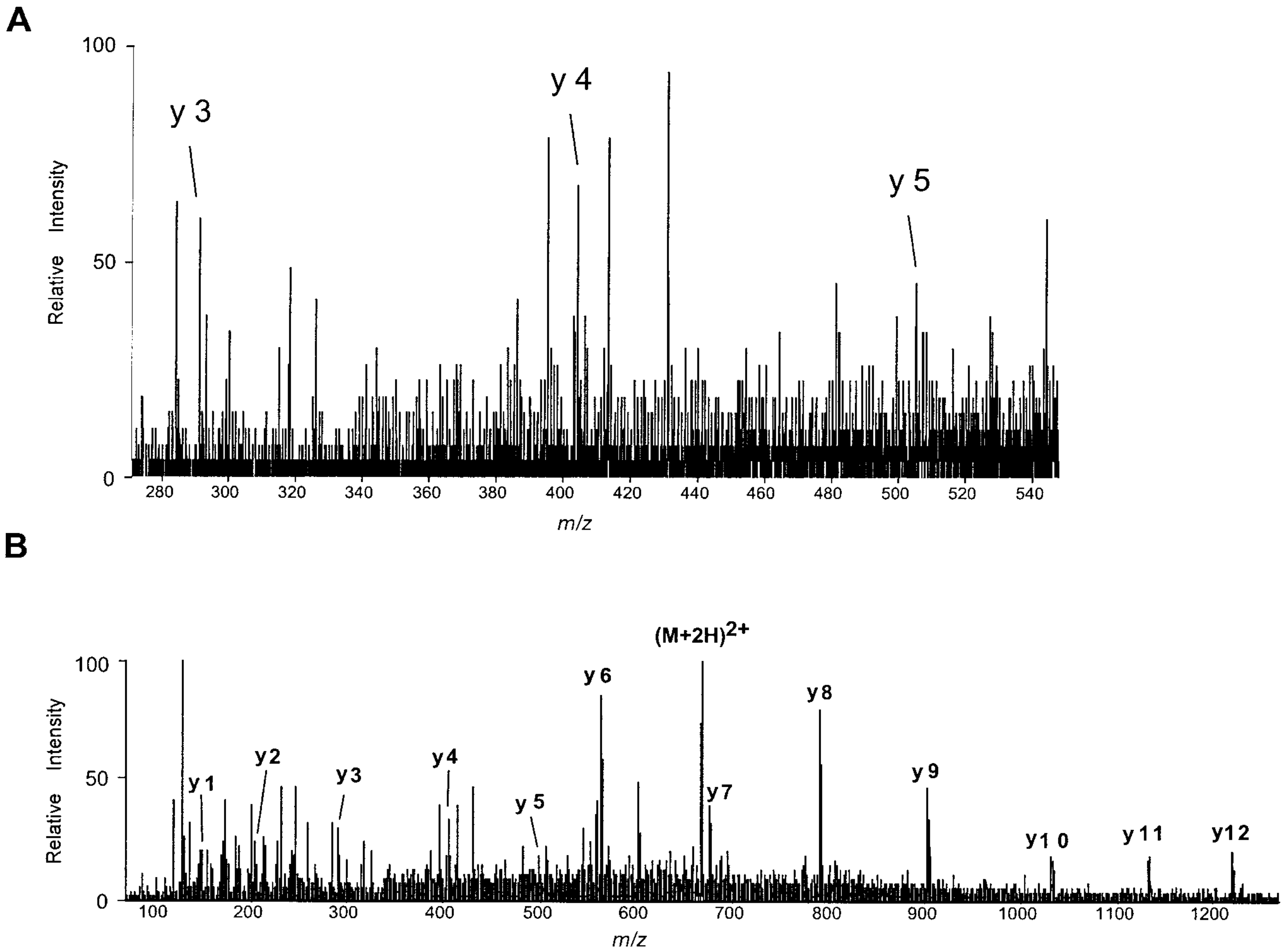

Figure 1. The Q-TOF fragment spectrum of the ^{18}O labelled doubly charged peptide ISTEINLGTLSGK. Panel A shows a part, Panel B the complete spectrum. 35% of the fragmented peptides were labelled with ^{18}O at their *C*-terminus. Even though the y-ions show the correct degree of labelling they cannot be readily identified in the spectrum due to their limited abundance and the large number of other fragment ions in their vicinity.

Although the width of the ion selection window is 3 Thomson the ^{16}O isotopes can be excluded from transmission into the collision cell without reducing the transmission of the adjacent ^{18}O isotopes extensively. In this way a *pseudo* isotope specific investigation can be performed with a reasonable high sensitivity. If the quadrupole were adjusted to exclusively transmit an individual isotope by narrowing the transmission window to 1 Thomson or less, the overall ion transmission would be reduced by at least a factor of 10 and the sensitivity of the analysis would be compromised.

The two fragment spectra, one selecting the entire ^{16}O/^{18}O isotopic cluster and the second selecting only the ^{18}O isotopes, are acquired under identical conditions. Y-ions are identified by comparing the two spectra using the two criteria that they should be represented in the first spectrum by a ^{16}O/^{18}O isotopic cluster and that the ^{16}O isotopes should be relatively suppressed in the second spectrum (Fig. 3). B-ions and internal fragment ions do not contain the *C*-terminus and are therefore represented by their natural isotopic distribution in the first and the second spectrum. The differential scanning method generates an additional criterion for y-ion identification – the different isotopic representation in the second spectrum when compared with the first. This criterion represents the information gained with the acquisition of the second fragment spectrum. It can be decisive to identify y-ions in the low *m/z* region when they coincide in mass with other fragment ions.

3.2 The y-ion filtering algorithm

The high resolution of fragment spectra acquired with a Q-TOF mass spectrometer allows the method to be exploited to its full depth. However, the richness of

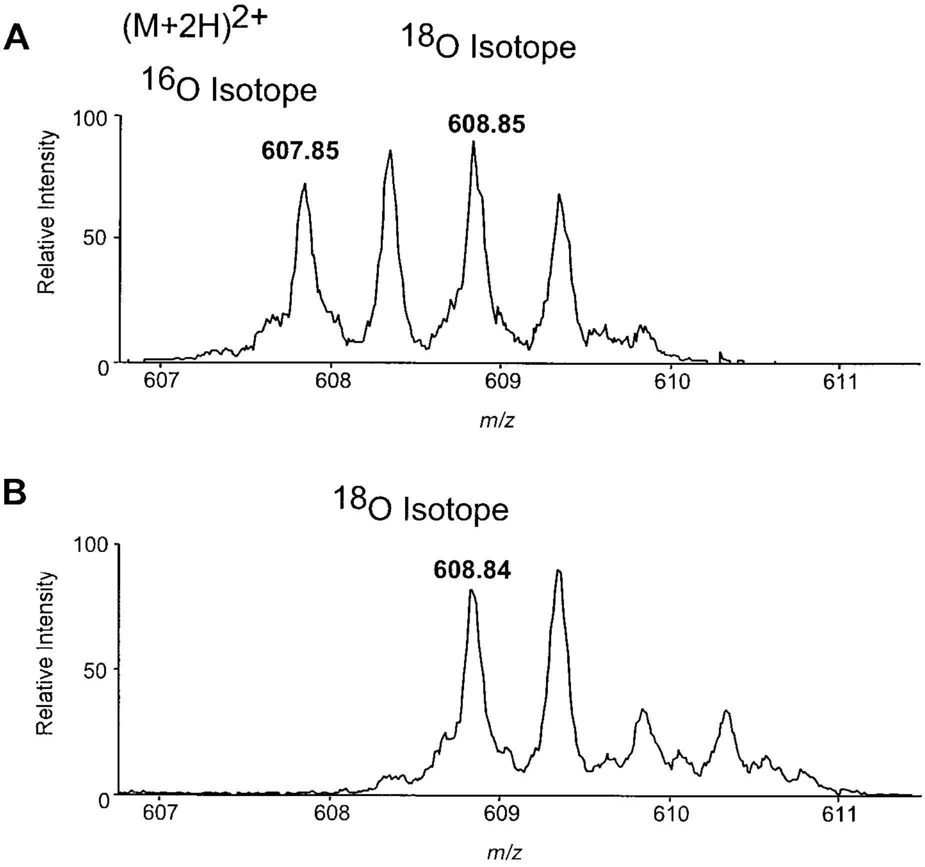

Figure 2. The ion transmission into the collision cell of a doubly charged ^{18}O labelled peptide before fragmentation. Panel A shows that when the precursor ion selection is set to the first isotope of a $^{16}O/^{18}O$ labelled peptide the entire isotopic cluster is transmitted into the collision cell if the quadrupole is adjusted to an appropriate low resolution. When the precursor ion selection is moved to the *m/z* value of the ^{18}O isotope, the ^{16}O isotope can be excluded from the fragmentation experiment without affecting the transmission of the ^{18}O isotope to a major degree. Since the resolution of the quadrupole was not increased for the second MS/MS experiment and the width of the window of transmission did not decrease, more chemical noise ions reach the collision zone (here between *m/z* 609.5 and 611).

the spectrum can make the interpretation very time consuming. Therefore, an algorithm has been developed to score each peak in the $^{16}O/^{18}O$ fragment spectrum for its likelihood to be a *C*-terminal fragment ion. The scoring function consists of two parts. In the first step every peak in the $^{16}O/^{18}O$ fragment spectrum is evaluated whether it is the first isotope of a $^{16}O/^{18}O$ labelled fragment. The intensities of the following peaks are compared with the expected isotope abundance of an ^{18}O labelled peptide. To calculate the expected isotope distribution the elemental composition of the fragment and the effective degree of ^{18}O labelling must be known. The elemental composition is approximated by the elemental composition of a peptide having the same mass with an average amino acid composition determined from the statistical abundance of amino acids of all proteins in the nonredundant database (NRDB). The actual degree of ^{18}O labelling is measured on an obvious y-ion in the fragment spectrum itself. In a second step the relative suppression of the ^{16}O isotope in the second spectrum in comparison with the first is determined. The two scoring functions are multiplied with the $^{16}O/^{18}O$ spectrum to produce a spectrum which is filtered for y-ions. The dynamic range of the scored spectra is reduced to approximately 1000 to display them on the computer screen (see Section 2.4).

Before evaluating the spectrum the charge state of every ion is determined. It is derived from the best correspondence between the intensities of *m/z* values of putative isotopes with the expected isotopic abundance for the charges 1 to 4. The expected isotopic distribution is calculated from the average amino acid abundance determined from the NRDB (EBI, Cambridge, UK, http://www.ebi.ac.uk).

The algorithm to measure the actual degree of ^{18}O labelling consists of three parts, detection of a series of putative ^{18}O labelled fragments, determination of their degree of ^{18}O labelling and scoring of the measured values. Only the upper 40% of the spectrum's *m/z* scale is considered to find ^{18}O labelled fragments. This restriction reduces the probability of overlap with other fragment ions. To detect ^{18}O labelled fragments an autocorrelated spectrum is calculated. The intensity of every *m/z* value is multiplied with the intensities of (*m/z* + 1), (*m/z* + 2) and (*m/z* + 3). This enhances considerably the first isotope of an ^{18}O labelled fragment. The peak with the maximum intensity in every 70 Da wide window is considered a putative ^{18}O labelled fragment. Its degree of ^{18}O labelling is calculated from its isotopic distribution in comparison with the expected isotopic distribution of a peptide of the same

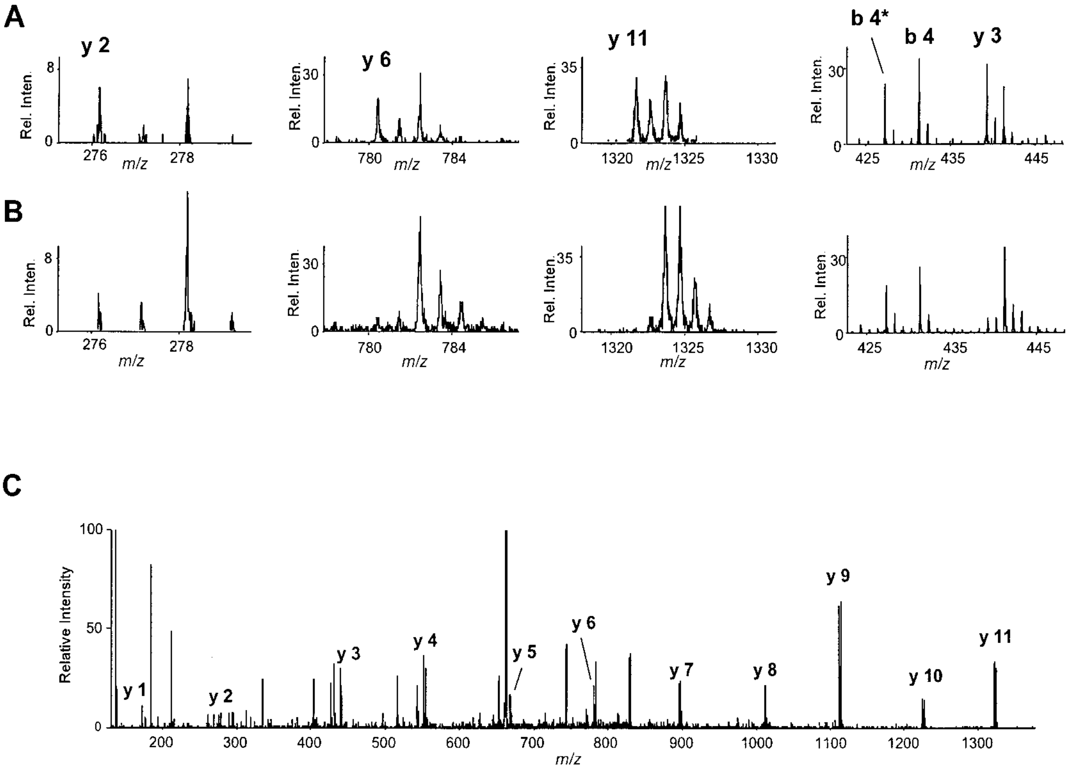

Figure 3. MS/MS sequencing of the peptide LGYPITDDLDIYTR using the differential scanning method. 37% of the peptide is labelled on its *C*-terminus with an ^{18}O isotope. Panel C shows the entire fragment spectrum. Panel A displays selected views from the fragment spectrum of the ^{16}O/^{18}O isotopic cluster. Panel B shows the corresponding parts from the ^{18}O fragment spectrum. y-ions can consistently be identified by their isotopic pattern in Panel A and the relative suppression of the ^{16}O isotope in Panel B. B-ions or internal b-ions (b 4*) do not change their relative isotopic abundance in the second spectrum.

mass with an average amino acid composition. To exclude all obviously unlabelled fragment ions from further consideration only peaks with a labelling efficiency between 12.5% and 75% are taken into account. They are scored using the intensity of the first isotope in theoriginal spectrum and in the autocorrelated spectrum. The score is a linear function of the intensity. The effective degree of labelling is the scored average of the determined degree of labelling of the different peaks.

For the evaluation of the ^{16}O/^{18}O isotopic pattern the relative deviation Δ of the measured intensities of the isotopes of the peptide from the expected intensity of a ^{18}O labelled peptide with an average amino acid composition is calculated. Below a mass value of 300 only one additional isotope is considered (m+2); between 300 and 1000, three (m+1, m+2, m+3); and above 1000, four isotopes (m+1, m+2, m+3, m+4) are taken into account. The isotope intensities are normalized by dividing them through their average intensity. The deviation of the measured to the expected isotopic distribution is the square root of the sum of the squares of their differences. The scoring value p_Δ of the deviation is calculated using the following formula:

$$p_\Delta = S_\Delta \left(e^{\left(-\frac{1}{W_\Delta}\Delta^2 \right)} + 0.001 \right) \qquad (2)$$

The parameters S_Δ (strength) and W_Δ (width) determine which weight the evaluation of the isotopic distribution should contribute to the overall evaluation (S_Δ) and how fast the scoring factor p_Δ drops to 1/1000 of its maximal value when the isotopic distribution deviates from the expected isotopic distribution (W_Δ). For the evaluation of the spectra S_Δ was set to 10 and W_Δ to 0.05.

The suppression of the first isotopes in the second spectrum is evaluated by calculating the difference of the relative height of the first isotope (m) for masses below 900 Da, the first and the second for masses between 900 Da and 1800 Da (m, m+1) and the first three isotopes for masses above 1800 Da (m, m+1, m+2). The suppression factor is the product of the differences in intensity

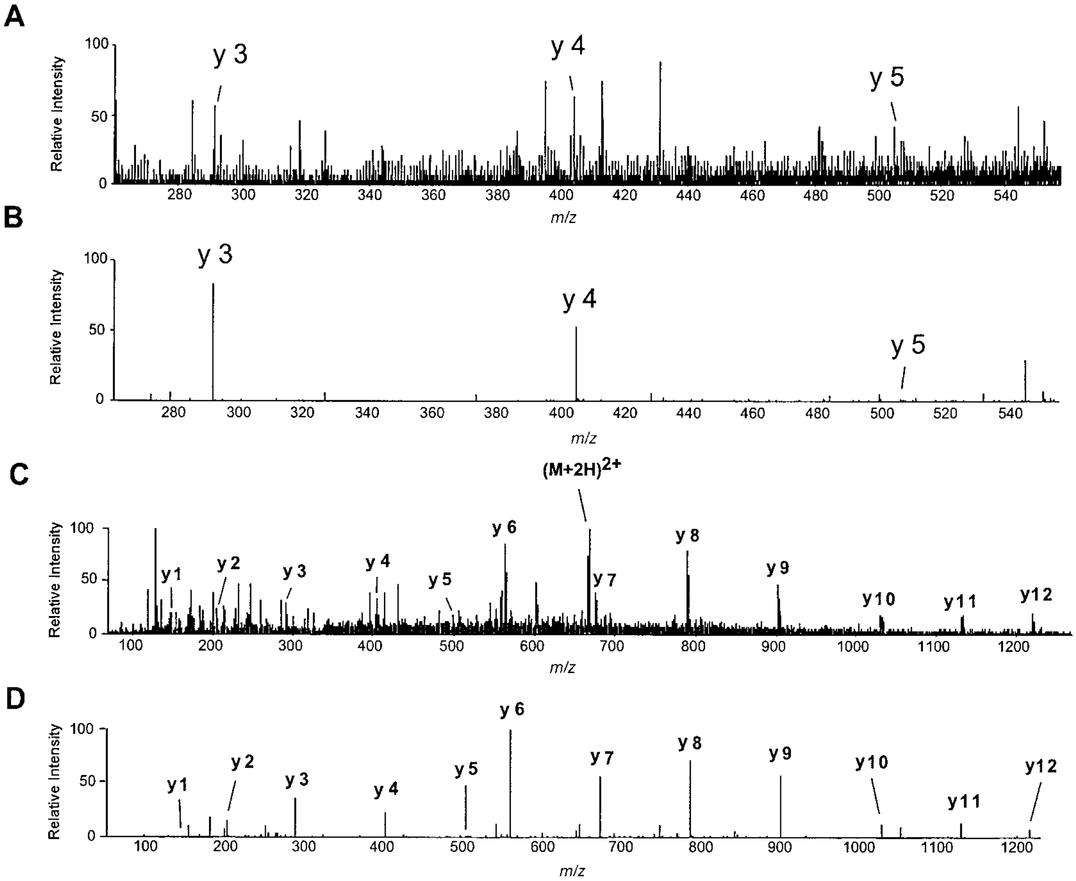

Figure 4. The fragment spectrum of the peptide ISTEINLGTLSGK investigated with the differential scanning technique. Panel C shows the $^{16}O/^{18}O$ fragment spectrum before scoring for y-ions and Panel D afterwards. Panels A and B show subsets comparable to those of Fig. 1. The dynamic range of the scored spectra had been reduced. The sequence could be automatically read by aligning the most abundant peaks in a series spaced by amino acid masses (Table 1, peptide 5). The comparison between Panel A and B demonstrates the specificity with which y-ions can be identified using the differential scanning approach.

of the individual isotopes. The differences in intensity are taken to the power of exponents, such that the suppression of the first isotope has a higher significance than the suppression of the later isotopes and that the sum of the exponents is always 1. In the case of a negative relative suppression of one of the considered isotopes the suppression factor σ is set to 0. The scoring factor p_σ of the relative isotope suppression is calculated with the following formula:

$$p_\sigma = S_\sigma\left(1 - e^{\left(-\frac{1}{W_\sigma}\sigma^2\right)}\right) \tag{3}$$

The parameters S_σ (strength) and W_σ (width) have the same significance as S_Δ and W_Δ. For the evaluation of the spectra S_σ was set to 5 and W_σ to 50 000. The y-ion scored spectrum is calculated by multiplying the original spectrum with the two scoring values p_σ and p_Δ.

3.3 The automatic sequence determination

A representative number of different peptides which have been investigated with this method and interpreted automatically are listed in Table 1. The table shows the three best proposals generated for every peptide. The algorithm does not yet consider b- or a-ions. Therefore, it clearly shows the merits and limitations of the differential scanning technique in y-ion identification and sequence prediction.

The scoring functions have a major influence on the scored fragment ion "intensities" and the resulting spectrum cannot be considered to reflect the chemical stability of individual bonds. Fig. 4 shows the fragment spectrum of the peptide ISTEINLGTLSGK. The y-ion series is complete but could not be identified to its full extent in the $^{16}O/^{18}O$ fragment spectrum (see Fig. 1). The scoring algorithm filters for y-ions so efficiently that the amino

Table 1.

No.	Charge state	Sequence	Predicted sequence	Comment
1	2	FAENAYFIK	**FAENAYFLK**	
2	2	SNTFVAELK	**SNTFVAELK**	
3	2	LFGVTTLDIIR	**LFGVTTLDLLR**	
4	2	IGSDAYNQGLSER	**LGSDAYNQGLSER**	
5	2	ISTEINLGTLSGK	**LSTELNLGTLSGK** **LSTELNL**SA**LSGK** **LS**VM**LNLGTLSGK**	
6	2	GIPADKISAR	**GLPADQLSAR** **GLPADQL**TG**R**	
7	2	TNSTFNQVVLK	**TNSTFNQVVLK** GTG**STFNQVVLK** **TNSTFNGAVVLK**	* No fragmentation between TN
8	2	SDVLFNFNK	**SDVLFNFNK** DS**VLFNFNK**	* No fragmentation between SD
9	2	DGSVVVLGYTDR	**DGSVVVLGYTDR** GSD**VVVLGYTDR** STA**VVVLGYTDR**	* No fragmentation between DGS
10	2	TAVVVGTVTDDVR	**TAVVVGTVTDDVR** **ATVVVGTVTDDVR** GD**VVVGTVTDDVR**	* No fragmentation between TA
11	2	ILTFDQLAELESPK	**LLTFDQLALESPK** **TLLFDQLALESPK** LTL**FDQLALESPK**	
12	2	LGYPITDDLDIYTR	**LGYPLTDDLDLYTR** YGL**PLTDDLDLYTR** AFD**PLTDDLDLYTR**	* No fragmentation between LGY
13	2	HTAEFAAHLVK	QNVP**FAAHLVK** QNPV**FAAHLVK** QPRG**FAAHLVK**	* b-ion dominated spectrum
14	3	FGQGEAAPVVAPA-PAPAPEVQTK	SW**APAPAPAPEVQTK** WS**APAPAPAPEVQTK** QS**APAPAPAPEVQTK**	* No y-ions beyond y 13 (APAP)
15	2	AALIDCLAPDR	**AALLDCLA**LV**R** **AALLDCLAPDR** **AALLDCLA**VL**R**	
16	2	IPYVGLTGNYR	**LPYVGLT**NG**YR** **LPYVGLTGNYR** **LPYVGLT**YGN**R**	
17	3	LLPHIPADQFPAQA-LACELYK	MA**FPAQALA**MTG**LYK** MA**FPAQALA**MSA**LYK** MA**FPAQALA**MG**TLYK**	* No fragmentation between CE * No y-ions beyond y 13 (FPA)
18	2	AVEIGSFLLGRDPK	**AVELGSFLL**AGSL**PK** VA**ELGSFLL**AGSL**PK** GL**ELGSFLL**AGSL**PK**	* No fragmentation between GRD and AV
19	3	LFVRPFPLDVQESEL-NEIFGPFGPFK	GWFMLDW**SEL**QD**LFGPFG**DE**K** QGFMLDW**SEL**QD**LFGPFG**DE**K** WGFMNDW**SEL**QD**LFGPFGPFK**	* Very little fragmentation between PF (y 2) * y 10 overlaps with a multiply charged fragment (NE) * No y-ions beyond y 15 (ESE)

Table 1. Continued

No.	Charge state	Sequence	Predicted sequence	Comment
20	2	RAQSVVDYLISK	HLPTH**LSK** HLPHT**LSK** HLGYD**LSK**	* b-ion dominated spectrum * No fragmentation ofthe *N*-terminal amino acids (RAQS)
21	3	LEESLSIISEKVPFNDTK	**LEESLSLLSEQV**GLGM**DTK** **LEESLSLL**TD**QV**GLGM**DTK** **LEESLSLLSEQV**LNM**DTK**	* Very little fragmentation between PFN
22	3	ALLNFHTYTEQR	**ALLNFHTYTEQR** **ALLNFHTYTE**AG**R** **ALLN**HFTTY**EQR**	

The table lists a series of peptides which were analyzed with the differential scanning technique and interpreted automatically. If several sequence proposals were made the three top scoring ones are displayed. Underlined sequences are correct, bold sequences had been marked as relatively certain since the amino acids are flanked by highly scored fragment ions. The method filters y-ions to a sufficient degree that the automatic interpretation routine can pick them up with an acceptable degree of reliability. Note that the amino acids which had been marked as certain are, with very few exceptions, correct. At this stage of the development, the interpretation algorithm looks exclusively for y-ions. If no y-ions are generated the sequence cannot be called. Integration of b-ion based interpretation can increase the length of retrieved sequences in individual cases (peptide 13, 14, 19, 20).

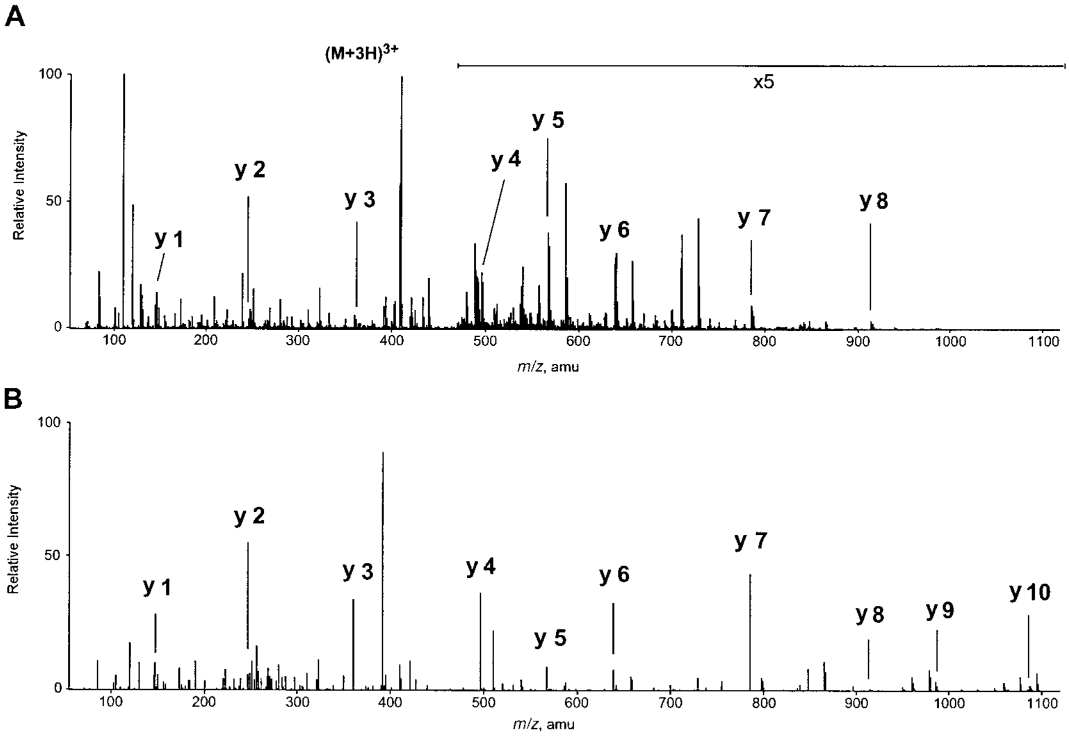

Figure 5. MS/MS investigation of the peptide HTAEFAAHLVK using the differential scanning approach. The triply charged precursor was fragmented. Even though most of the y-ions are not the most abundant ions after scoring y1 to y7 can be identified automatically by aligning the y-ions into a series of amino acid mass spaced ions (Table 1, peptide 13).

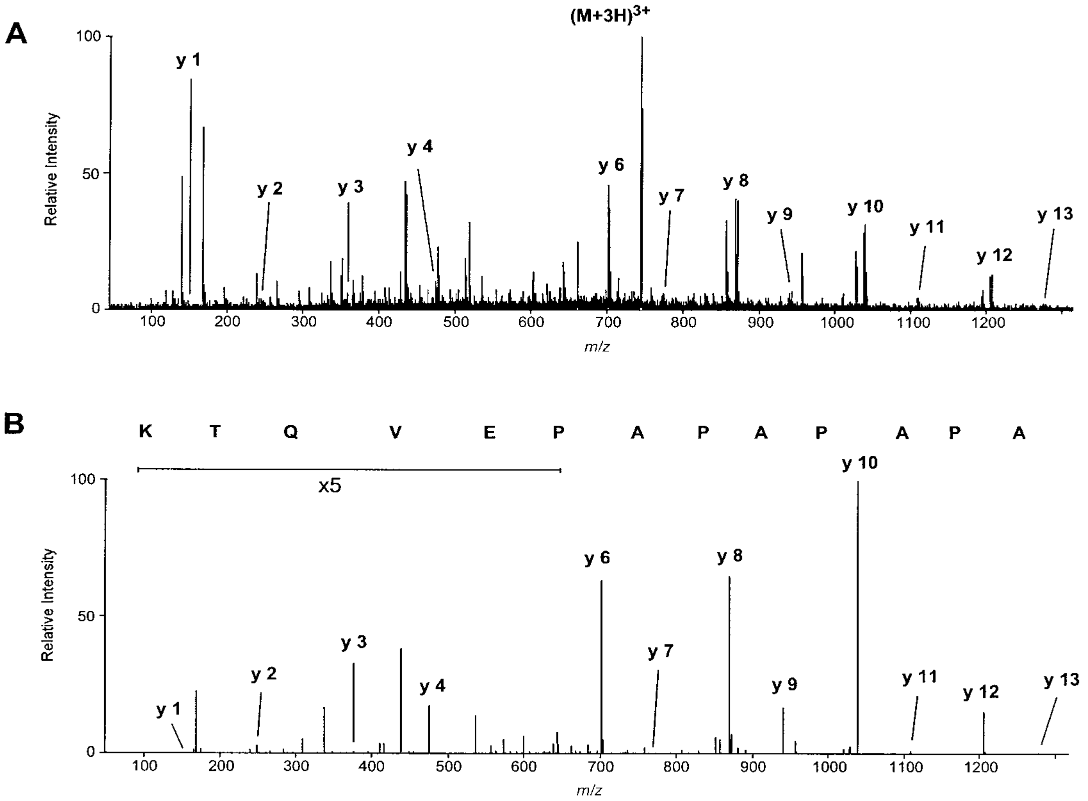

Figure 6. MS/MS investigation of the peptide FGQGEAAPVVAPAPAPAPEVQTK with the differential scanning technique. The triply charged precursor was fragmented. Panel A shows the original and Panel B the scored fragment spectrum of the $^{16}O/^{18}O$ isotopic cluster. Because of the fragile repetitive AP region the fragmentation pattern is incomplete and difficult to interpret. The original spectrum cannot be interpreted correctly. The series of y-ions between y1 and y13 are identified automatically ((1205.6)PAPAPAPEVQTK) (Table 1, peptide 14). The algorithm selects the y-ions based on their intensity and their amino acid spacings. The complete sequence cannot be deduced from the spectrum since the *N*-terminal part of the peptide is not fragmenting.

acid sequence can be read automatically from the scored spectrum (Table 1, peptide 5).

The merits of this technique are tested when used on peptides that show a more complex fragmentation behaviour. Fig. 5 shows the fragment spectrum of the triply charged precursor of the peptide HTAEFAAHLVK. Even though the y-on series is not dominating the low *m/z* region the y1–y7 ions are correctly identified with the highest score when aligning amino acid spaced peaks (Table 1, peptide 13). The advantage of an automatic sequence read out from a spectrum is that many spectra can be processed. The disadvantage is that an algorithm is much less flexible than an experienced individual to recognize a particular situation in the spectrum. In Fig. 5 the ion at 391 Da is scored as a y-ion. Closer inspection of the original spectrum shows that the putative ^{16}O and ^{18}O isotopes have a different primary structure (390.970 Da and 393.115 Da) in contrast to the y3 ion (359.159 Da and 361.165 Da). This difference was not recognized by the automatic routine. The mass tolerance set for the general evaluation is just a little big larger than required in this case. This was necessary to accommodate correct assignment of multiple charged fragments after projection to their singly charged *m/z* value. Theoretically it is possible to keep the charge state of every peak in memory and adjust the mass tolerance accordingly. However, the situation described in Fig. 5 is a rare event and even in this case did not lead to a false interpretation. It is obvious that with instruments having a higher resolution than the Q-Tof1 like those that are made available today will improve the filtering algorithm and with it the interpretation routine.

The same is true for judging peak overlaps. Care was invested to recognize multiple charged fragments correctly. But there are situations when the ^{18}O isotope of a singly charged fragment overlaps with the first isotope of a multiple charged fragment. The ^{18}O isotope will disappear from the spectrum since it is projected to a higher *m/z* value as part of a multiple charged fragment. If the spectrum would perfectly well represent the relative abundance of ions this could be recognised as such by the computer programme. However, this is not the case, particularly for very small peaks which consist only of 10 ions or less. For the overall result it is important to allow the algorithm to interpret very small peaks. For this high sensitivity the requirement of a very precise quantitative representation of every peak was sacrificed.

Fig. 6 gives an even more complex example. The triply charged precursor of the peptide FGQGEAAPWAPAPA-PAPEVQTK was fragmented. Since the *N*-terminal bond of a proline is labile the peptide has a fragile region in the repetitive AP sequence. Because the *C*-terminal bond of a proline is very stable the fragment spectrum is dominated by abundant fragments followed by very small ions. This renders the interpretation of the original data very difficult and error prone. The automatic interpretation routine determines the *C*-terminal 13 amino acids correctly, they are marked as being relatively certain and it is realized that there are no more sufficiently good fragmentation data available to support a much longer sequence proposal (Table 1, peptide 14). There are several doubly charged y-ions which are recognized as such and the y13 ion is nearly exclusively based on a doubly charged fragment ion. It should be noted that the sequence of the *C*-terminal 12 amino acids can only be partially confirmed by b-ions. It is not possible to determine the entire sequence of the peptide from the spectrum because the *N*-terminal region does not fragment to a sufficient degree.

Figs. 5 and 6 demonstrate that the scoring mechanism does not generate spectra which consist only of y-ions. The existing y-ions may not represent the peaks with the highest intensity in every region of the scored spectrum. The mathematical reduction of the dynamic range of the spectra exaggerates the impression that there are many other ions of similar intensity still in the scored spectrum since the relative differences between the peaks are reduced. But it is the case that existing y-ions are not always receiving the highest scores. This is due to the high sensitivity in picking up y-ions which is required to optimize the automatic sequence read out. Fig. 7 shows extracts from the original fragment spectra and the scored spectrum of the peptide FGQGEAAPVVAPAPAPAPEVQTK. Each isotope of the y2 ion which was correctly assigned consists of less than 10 ions. The scoring functions allow enough tolerance so that these peaks still receive a positive score. But the same tolerance renders the spectrum more noisy. In contrast to y2 the isotopes of y12 contain 20 isotopes or more *per* peak. This ion is recognized as an “obvious y-ion” by the interpretation routine and is therefore part of every possible sequence proposal.

It may be mentioned that from the two criteria to identify y-ions using the differential scanning method, the $^{16}O/^{18}O$ isotopic labelling in the first spectrum and the relative suppression of the ^{16}O isotope in the second, none of the two is a sufficient criterion by itself to pick up y-ions unambiguously. The $^{16}O/^{18}O$ isotopic distribution can be mimicked or distorted by other fragment ions particularly in the low *m/z* region. On the other hand, the suppression of the ^{16}O isotope is observed for all fragment ions in the high *m/z* region when a peptide with a mass above 1800 Da is analyzed. Fig. 8 shows the contribution of ions labelled with ^{16}O and ^{18}O to the overall isotopic distribution of a large peptide. To exclude sufficiently well the ^{16}O labelled ions from the second experiment the (m+2) ion must be completely excluded from the MS/MS investigation. Ions with the mass of (m+2) are ^{18}O labelled ions but as well ^{16}O labelled ions which contain somewhere in the peptide two heavier isotopes like ^{13}C or ^{15}N. If the (m+2) ions are excluded from the second experiment some ^{18}O labelled peptide ions are excluded as well. These are precisely those which contain beside the ^{18}O at their *C*-terminus no other heavier isotope (so called ^{12}C-only ions). A specific subgroup of all ^{18}O labelled peptide ions are excluded from the second experiment. This will be visible when this specific subgroup contributes a large proportion to a particular isotope. This is the case for the first isotope of any large fragment. For small fragments the first isotope is coming from two groups of ions, the ^{12}C-only ions and other ions of the peptide whose ^{13}C or ^{15}N isotopes are located in the other part of the peptide. This second group for generating the first isotope of a fragment is getting rapidly smaller when the fragment size increases. This is so because the size of the other part of the peptide where the ^{13}C and ^{15}N isotopes are to be located decreases. Therefore, the ^{12}C-only ions of the peptide contribute considerably to the first isotope of any large fragment. If these ions are excluded from the second experiment the first isotope of all larger fragments will be suppressed (Fig. 9). The criterion that the first isotope for y-ions is suppressed is not sufficient to filter them out unambiguously. In this region only the combination of the two criteria, ^{18}O labelling and suppression, is specific for y-ions. Luckily, this region of the spectrum is generally not populated with too many fragment ions so that the true isotopic ratios of the peaks are often conserved.

3.4 Characterization of proteins involved in vesicular targeting and fusion

Endocytosis, the incorporation of a small vesicle formed from the outer cellular membrane, is the main mechanism by which cells transport material from the outside of the cell to their interior. The opposite process, the melting of an internally generated vesicle with the outer cell

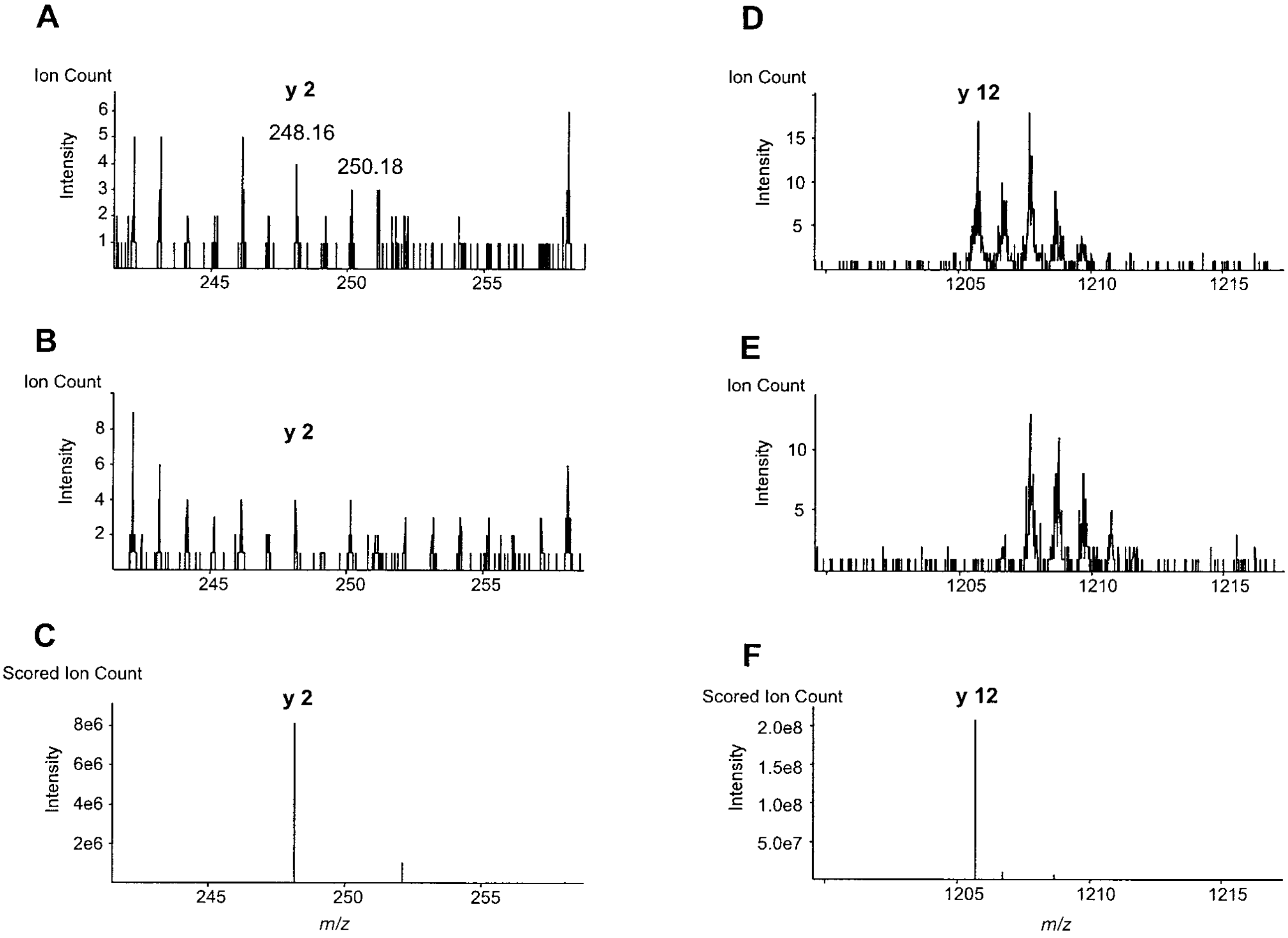

Figure 7. Details from the fragment spectra of the peptide FGQGEAAPWAPAPAPAPEVQTK (see Fig. 6). Panels A and D show details from the fragment spectra after selection of the entire $^{16}O/^{18}O$ isotopic cluster, Panels B and E details from the ^{18}O isotope fragmentation experiment and Panel C and F show the scored spectra. y2 and y12 were correctly assigned by the sequence determination programme (Table 1, peptide 14), y2 because it is part of the best scoring sequence proposal and y12 because it is recognized as an "obvious y-ion" and therefore part of every proposed sequence. The example demonstrates the level of sensitivity to which the scoring algorithm is adjusted to optimise the quality of the sequence proposals.

membrane is used to secrete material (exocytosis) or to expose molecules at the outside of the cellular membrane. Endocytosis, exocytosis and the transport between organelles of the biosynthetic and endocytic pathways are highly regulated in order to maintain the integrity of the organelles in the cell. The generation of transport vesicles, their targeting and fusion with the appropriate acceptor membrane are controlled by specific molecules. Particularly, two classes of proteins play an essential role in many vesicle transport processes. The first class is represented by the SNAREs (soluble NSF-attachment protein receptors) [20] which are necessary for docking and fusion, and the second by the small GTPases of the Rab family. Rab proteins have first been implicated in vesicle docking as upstream modulators of the SNARE proteins but their functions appear to be more diversified among members of the Rab family [21]. For example, Rab5 functions in the early endocytic pathway where it regulates transport from the plasma membrane to the early endosome as well as homotypic endosome fusion [22]. Rab5 is also involved in the regulation of endosome interactions with the microtubule network of the cytoskeleton [23].

To unravel the molecular mechanisms of the different functions of Rab5 and the identity of Rab5 effectors, an affinity chromatography approach based on recombinant Rab5:GTP immobilized on beads was developed to identify Rab5 effectors (for further experimental details, see reference [14]). This approach led to the purification of over 20 Rab5:GTP interacting proteins, including previously characterized as well as novel molecules.

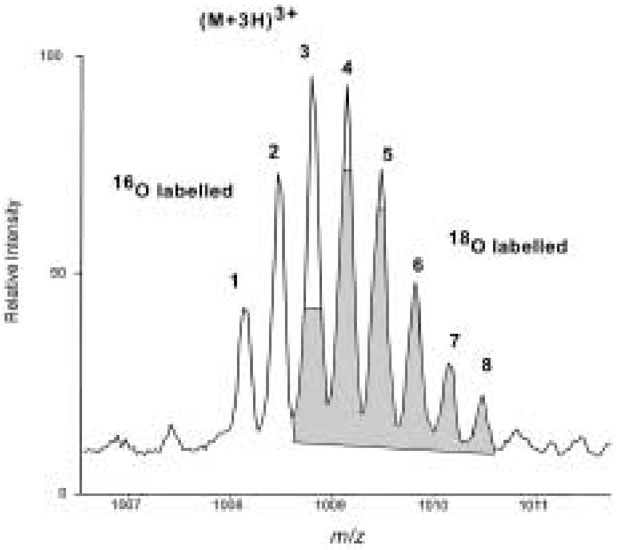

Figure 8. The precursor ion of the triply charged peptide LFVRPFPLDVQESELNEIFGPFGPFK. The ^{18}O labelled ions are shown in a darker color. To exclude efficiently the ^{16}O labelled peptide ions even the (*m*+2) ion needs to be excluded from the fragmentation experiment.

Here we report on the sequencing of a 95 kDa bovine protein from the Rab5:GTP affinity column eluate using the differential scanning technique. After tryptic digestion the peptides were sequenced on a quadrupole time of flight instrument (Q-TOF, Micromass) using the differential scanning technique. Fig. 10 shows the fragment spectrum of one of the ^{18}O labelled peptides. Its triply charged precursor was fragmented. The filtering of the spectrum for y-ions implies the determination of the charge state of every fragment. This allows scoring and deconvolution of the spectrum simultaneously. The deconvolution simplifies the interpretation of the spectrum. Two peptides of the protein were completely sequenced. The automatically generated sequences were controlled manually (Table 1, peptides 21, 22) and resulted in ((L/I)EES(L/I)S (L/I)(UI)**S**EKVPFNDTK and A(L/I)(L/I)NFHTYTEQR). These peptides were used to find similar proteins in searches against the dbEST databank (http://www.ncbi.nlm.nih.gov/BLAST/). With the two peptides, only a mouse EST (gb|AA003226|AA003226) was found by homology. The sequence of two mouse peptides were nearly identical to the sequenced bovine peptides and they were found in

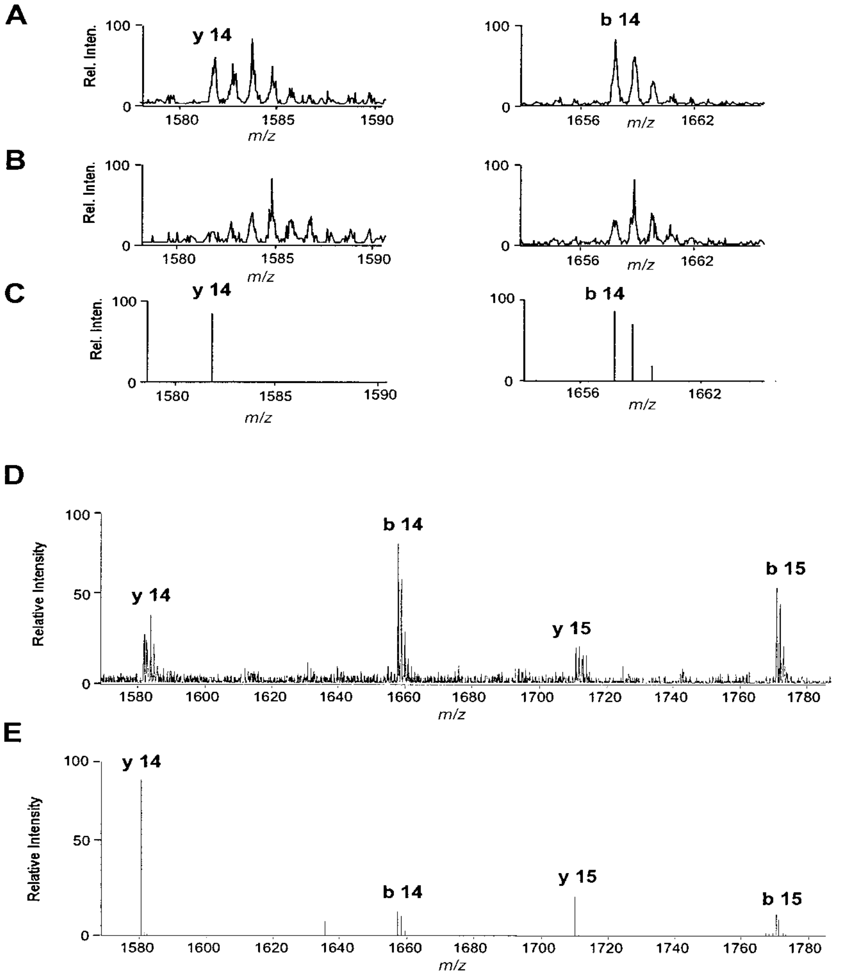

Figure 9. MS/NS investigation of the peptide LFVRPFPLDVQESELNEIF-GPFGPFK analyzed with the differential scanning technique. Panels A and D show the original $^{16}O/^{18}O$, Panel B the original ^{18}O fragment spectrum and panels C and E the scored $^{16}O/^{18}O$ spectra. Since the (*m*+3) Da ion was omitted from the second MS/MS investigation less ^{12}C only ions of the ^{18}O labelled peptide were fragmented. This is visible when the relative isotopic abundance of large b ions are compared (Panels A and B). When scoring the spectra for y-ions one of the two criteria, the relative suppression of the first isotopes is fulfilled for the b-ions as well. This is the reason why these ions are not completely suppressed in this region (see Panels C and E). The differences between the b- and y-ions in the scored spectrum are larger than it seems since the dynamic range of the scored spectrum had been reduced for display. y14 but not y15 is still part of the automatically generated sequence (Table 1, peptide 19). y15 is not part of the automatic sequence proposal because larger b-ions guide the sequence pattern to a different path through the scored spectrum.

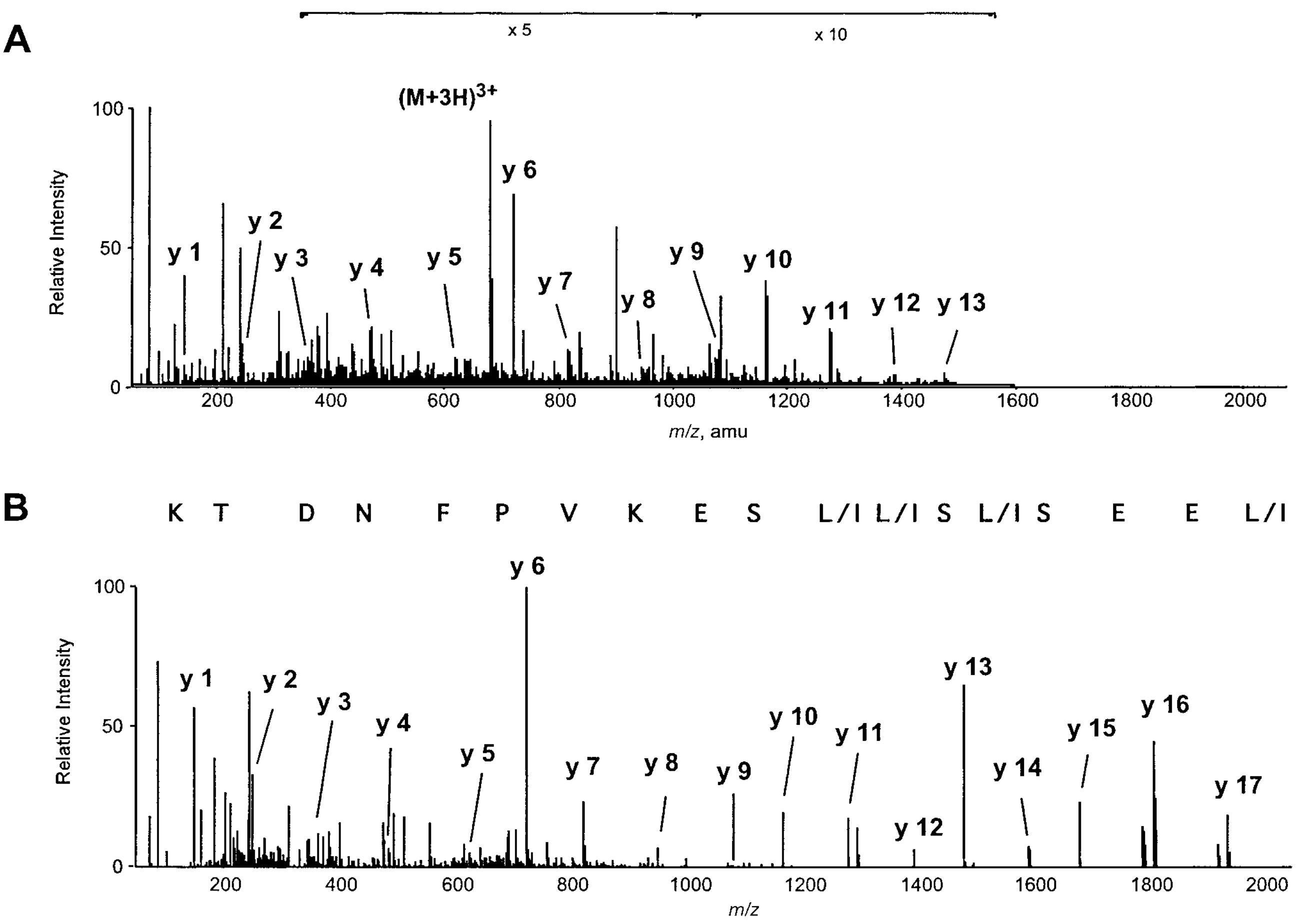

Figure 10. *De novo* sequencing of a peptide from a putative Rab5 effector protein using the differential scanning method. The peptide was labelled with ^{18}O upon digestion. The triply charged ion with an *m/z* of 683.7 Thomson was fragmented. The original fragment spectrum is shown in Panel A. After scoring and charge state deconvolution the spectrum is converted into a linear, easy to interpret spectrum (Panel B). The sequence of the peptide (L/I)EES(L/I)S(L/I)(L/I)SEKVPFNDTK was determined by manual correction of the automatically determined peptide sequence **LEESLSLLSEQV**GLG**MDTK** (Table 1, peptide 21). The automatically generated result indicated that the GLG sequence was not read with a high confidence level. Lysine and glutamine are not differentiated automatically because their mass difference is smaller than the tolerance in mass used by the programme.

the same EST. This indicates that the EST represents a part of a mouse protein which is a very close homologue to the purified bovine protein (EST peptide sequences: LEESLSII**N**EKVPFNDTK and the nontryptic peptide ... **DLVP**ALLNFHTYTEQR). The 110 amino acid long sequence of the EST does not show any clear homology to any known protein. Its cloning is in progress. The recombinant protein will be used for further functional studies of vesicular fusion in *in vivo* and *in vitro* systems.

4 Concluding remarks

The differential scanning technique is a method to sequence isotopically labelled peptides *de novo* in a very reliable manner on Q-TOF instruments. Two tandem mass spectra are acquired of partially ^{18}O labelled peptides, one selecting the entire ^{16}O/^{18}O isotopic envelope and a second selecting only the ^{18}O labelled peptide ions for fragmentation. This can be done with a high sensitivity because the quadrupoles do not need to be adjusted to monoisotopic resolution. Two criteria can be used to identify *C*-terminal fragment ions in the acquired spectra, their ^{16}O/^{18}O isotopic pattern in the first spectrum and the difference in their isotopic representation between the first and the second spectrum. Both criteria together are so specific that an automatic evaluation of the fragment spectrum filters y-ions effectively from the spectrum. This simplifies the interpretation of tandem mass spectra of peptides considerably and can be used to speed up *de novo* sequencing. Even though manual interpretation is still superior, the automatic read out of sequences generates acceptable results. The peptide sequences are mostly used to find highly homologue proteins in databases. A homologue protein from a closely related organism can then be cloned and used in the biological system under investigation for further functional studies. This helps avoiding cloning proteins by degenerate oligonucleotide primers and increases the efficiency to experimentally validate new protein functions. The high quality of the filtered y-ion spectra can be of decisive help when proteins are identified *via* homology in EST databases. A reliable automatic interpretation of the spectra in combination with the identification of homologue proteins based on individual peptides is relevant when protein mixtures are analyzed with on-line HPLC MS/MS experiments [3, 24].

If the y-ion intensity is in the order of 20 or more *per* isotope and there are no peak overlaps which distort the isotopic ratios to more than 50% the identification of y-ions is very clear. If the mass spacing between adjacent fragments does not correspond to amino acid residue masses a structural deviation from the common linear amino acid chain can be assumed. This is an indication for a secondary modification or a structural rearrangements within the peptide. In this way the differential scanning method can be very useful to find unexpected secondary modifications. Since Q-TOF instruments allow the acquisition of fragment spectra from several precursors simultaneously the analysis can be automated using on-line HPLC MS/MS instrumentation with automatic precursor selection.

Financial support by a grant from the Bundesministerium für Bildung und Forschung (BMBF), BE031/11862 is gratefully acknowledged. SU was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG).

Received October 3, 2000

5 References

[1] Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V., Neubauer, G., Shevchenko, A., Mortensen, P., Mann, M., *Biochem. Soc. Trans.* 1996, *24,* 893–896.

[2] Patterson, S. D., Aebersold, R., *Electrophoresis* 1995, *16,* 1791–1814.

[3] McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., Yates, I. R., *Anal. Chem.* 1997, *69,* 767–776.

[4] Jensen, O. N., Mortensen, P., Vorm, O., Mann, M., *Anal. Chem.* 1997, *69,* 1706–1714.

[5] Haynes, P. A., Gygi, S. P., Figeys, D., Aebersold, R., *Electrophoresis* 1998, *19,* 1862–1871.

[6] Skilling, J., Cottrell, J., Green, B., Hoyes, J., Kapp, E., Landgridge, J., Bordoli, R., *Proc.* 47th ASMS Conf. *Mass Spectrom.* Allied Topics, Dallas, Texas 1999.

[7] Taylor, J. A., Johnson, R. S., *Rapid Commun. Mass Spectrom.* 1997, *11,* 1067–1075.

[8] Dancik, V., Addona, T., Clauser, K., Vath, J., Pevzner, P., *J. Comput. Biol.* 1999, *6,* 327–342.

[9] Kondo, H., Rabouille, C., Newman, R., Levine, T. P., Pappin, D., Freemont, P., Warren, G., *Nature* 1997, *338,* 75–78.

[10] Bartlet-Jones, M., Canas, B., Jeffery, W. A., Rahman, R., Hansen, H. F., Pappin, D. J. C., Proc. 46th ASMS Conf. Mass Spectrom. Allied Topics, Orlando, Florida 1998.

[11] Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., Hauer, C. R., *Proc. Natl. Acad. Sci. USA* 1986, *83,* 6233–6237.

[12] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M., *Nature* 1996, *379,* 466–469.

[13] Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomson, B., Wilm, M., Mann, M., *Rapid Commun. Mass Spectrom.* 1997, *11,* 1015–1024.

[14] Christoforidis, S., McBride, H. M., Burgoyne, R. D., Zerial, M., *Nature* 1999, *397,* 621–652.

[15] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68,* 850–858.

[16] Wilm, M., Mann, M., *Anal. Chem.* 1996, *68,* 1–8.

[17] Eng, J. K., McCormack, A. L., Yates III, J. R., *J. Am. Soc. Mass Spectrom.* 1994, *5,* 976–989.

[18] Mann, M., *Trends Biochem. Sci.* 1996, *21,* 494–495.

[19] Takao, T., Gonzalez, J., Yoshidome, K., Sato, K., Asada, T., Kammei, Y., Shimonishi, Y., *Anal. Chem.* 1993, *65,* 2394–2399.

[20] Rothman, J. E., *Nature* 1994, *372,* 55–63.

[21] Novick, P., Zerial, M., *Curr. Opin. Cell Biol.* 1997, *9,* 496–504.

[22] McBride, H. M., Rybin, V., Murphy, C., Giner, A., Teasdale, R., Zerial, M., *Cell* 1999, *98,* 377–386.

[23] Nielsen, E., Severin, F., Backer, J. M., Hyman, A. A., Zerial, M., *Nat. Cell Biol.* 1999, *1,* 376–382.

[24] Blackstock, W. P., Weir, M. P., *Trends Biotechnol.* 1999, *17,* 121–127.