

---

---

# De Novo Sequencing, Peptide Composition Analysis, and Composition-Based Sequencing: A New Strategy Employing Accurate Mass Determination by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

Bernhard Spengler

Institute of Inorganic and Analytical Chemistry, Justus Liebig University Giessen, Giessen, Germany

---

A new strategy is described for the determination of amino acid sequences of unknown peptides. Different from the well-known but often inefficient de novo sequencing approach, the new method is based on a two-step process. In the first step the amino acid composition of an unknown peptide is determined on the basis of accurate mass values of the peptide precursor ion and a small number of accurate fragment ion mass values, and, as in de novo sequencing, without employing protein database information or other pre-information. In the second step the sequence of the found amino acids of the peptide is determined by scoring the agreement between expected and observed fragment ion signals of the permuted sequences. It was found that the new approach is highly efficient if accurate mass values are available and that it easily outstrips common approaches of de novo sequencing being based on lower accuracies and detailed knowledge of fragmentation behavior. Simple permutation and calculation of all possible amino acid sequences, however, is only efficient if the composition is known or if possible compositions are at least reduced to a small list. The latter requires the highest possible instrumental mass accuracy, which is currently provided only by fourier transform ion cyclotron resonance mass spectrometry. The connection between mass accuracy and peptide composition variability is described and an example of peptide compositioning and composition-based sequencing is presented. (J Am Soc Mass Spectrom 2004, 15, 703–714)  
© 2004 American Society for Mass Spectrometry

---

---

Mass spectrometry as a tool for the determination of the amino acid sequence (AAS) of peptides (called peptide sequencing) so far has been employed basically in its form of nominal-mass spectrometry rather than accurate-mass spectrometry [1, 2]. Algorithms for descrambling fragmentation patterns were mainly based on probability functions and combinatorial calculations regarding an assumed fragmentation mechanism, rather than on evaluation of precise and accurate mass values of precursor and fragment ions. Graph representations were used to describe and perform peptide sequencing based on non-accurate mass values [3–5] with reasonable success, but additional neutral losses common in peptide ion fragmentation are hard to include and combinations thereof cannot fully be represented.

When instruments became more precise and more accurate in mass determination, users and evaluation

software tools to some extent started to include the first and second digit of mass values to distinguish different structural possibilities. Taking full advantage of exact mass information in bioanalytical mass spectrometry [6, 7], however, is still in its infant state.

Fourier transform mass spectrometry provides highest mass resolving power, but instruments, so far, were laborious to use and complicated in handling and data evaluation. Only recently, FTMS became available in a form as easy and efficient in operation as for example quadrupole ion trap mass spectrometry. Accurate mass values are now available in routine analytical laboratory work and certainly will dominate data evaluation protocols in the near future.

One of the primary targets of bioanalytical mass spectrometry is peptide sequencing for protein identification, protein characterization or functional proteomics. Nature composes peptides of 20 different amino acids while amino acids are composed of chemical elements C, H, N, O, and S. Only two amino acids, isoleucine (I/Ile) and leucine (L/Leu) have identical elemental compositions while all others can be clearly distinguished by their composition. Although amino

---

Published online March 2, 2004

Address reprint requests to Dr. B. Spengler, Institute of Inorganic and Analytical Chemistry, Justus Liebig University Giessen, Schubertstrasse 60, D-35392 Giessen, Germany. E-mail: Bernhard.spengler@uni-giessen.de

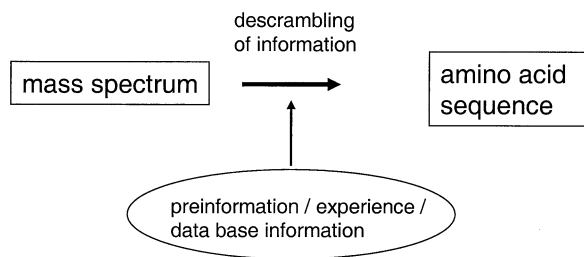


Figure 1. De novo sequencing scheme.

acids (K and Q) and especially amino acid combinations (e.g., AY and PH, or PF and LM) can easily have the same *nominal* mass, the probability of having the same elemental composition (and thus the identical accurate mass) is highly reduced. Taking advantage of the full information contained in accurate mass values is certainly known to be useful [8–16], but has not been fully exploited so far.

The purpose of this paper is to show the power of accurate mass values in sequence determination, and to define the requirements for a new analytical strategy based on accurate-mass spectrometry, resulting in a simplified and highly reliable spectrum interpretation and peptide sequencing independent of protein data base information.

De novo sequencing [17–22] has become a method of prominent importance in the field of functional proteomics. De novo sequencing is necessary because database search procedures [23, 24] often fail if the proteins are modified, unknown, mutated, artificially created via, e.g., combinatorial techniques, are from unknown species or cancerous cells [25]. In this case the amino acid sequence of peptides has to be determined without employing database searching procedures. This procedure, called de novo sequencing, often works reliably and fast, if peptides under investigation are tryptic and thus have basic residues at the C-terminus. They generally fail, however, if the cleavage enzyme is unknown or peptides are not digestion products. Basic residues (as strong proton acceptors) in the middle of a peptide sequence almost regularly lead to highly complicated fragmentation spectra which can be very difficult to interpret unequivocally. Several algorithms for de novo sequencing were reported [17–22], but none of them has been shown to be absolutely reliable and efficient if MS-MS data were obtained from non-ideal fragmentation processes. They use experience based or learned information about fragmentation behavior of peptides. Mass values are used with only limited de-

mand for accuracy, and accurate mass values are typically not used as exclusion or inclusion criteria. Such algorithms are complicated and sometimes slow in processing speed. They are, on the other hand, highly susceptible to failures based on misinterpretation of complicated fragmentation processes. The process of de novo sequencing is summarized in Figure 1.

In this paper a new strategy is reported for non-database assisted peptide sequencing, which is based on the determination of the amino acid composition in a first step (Figure 2). In analogy to the term peptide sequencing this step is called peptide compositioning or peptide composition analysis. It is shown that the interpretation of the fragmentation spectrum is very simple and reliable, if the amino acid composition of the analyte is known and the mass accuracy is high. This second step (based on the first) is called composition-based sequencing (CBS) or de compositione sequencing as an approach different from de novo sequencing. Determination of the composition is possible, if the mass of the peptide is not too high, if the instrumental mass accuracy is sufficient and in some cases only if additional characterizing information is obtained, e.g., by employing hydrogen–deuterium exchange.

## Experimental

### *Nomenclature for Amino Acid Compositions Including Partial Sequence Information*

In order to distinguish amino acid compositions (AAC) from amino acid sequences (AAS), the following nomenclature for AAC descriptors is suggested.

1. AAS of peptides on the basis of single letter codes are usually described by using upper case letters directly following each other from left to right in the sequence order beginning at the N-terminus and ending at the C-terminus (e.g., RLV meaning H-Arginine-Leucine-Valine-OH). To distinguish AAC from AAS it is suggested that compositions instead should be described by using indices for the number of units, by using alphabetical order, and by enclosing permutable (not position-defined) units in parentheses, brackets and braces. The AAC descriptor of the above peptide would therefore be  $H(L_1R_1V_1)OH$  assuming that the peptide has an amine N-terminus and an acid C-terminus.
2. Partial sequence information should be included in the AAC descriptor by using nested and hierarchical

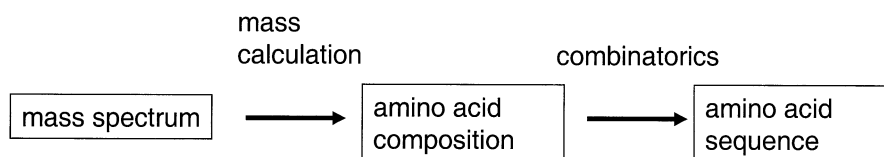
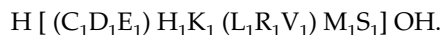


Figure 2. Composition-based sequencing (CBS) scheme.

brackets. A peptide for which it is known that the N-terminus contains the composition ( $L_1R_1V_1$ ) and the C-terminus contains the composition ( $C_1D_1E_1$ ), while the rest of the peptide has the composition ( $H_1K_1M_1S_1$ ) would be described as

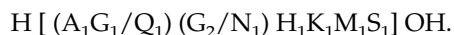


Units in brackets have to be permuted completely for listing all possible amino acid sequences for a given composition. Another peptide for which ( $L_1R_1V_1$ ) and ( $C_1D_1E_1$ ) are not necessarily terminal but possibly internal groups would instead be described by nested and hierarchical parentheses and brackets as



Alphabetical order should be used in this case regarding the first unit of a group. In this case the two groups are undefined in position within the rest of the peptide but do not overlap. So six units in outer brackets are permutable, two groups (each consisting of three units) and 4 amino acids.

- Variability in composition should be included in the AAC descriptor by using slashes and for clarity including the respective units in brackets. If in a composition a unit  $N_1$  could as well be  $G_2$  and unit  $Q_1$  could as well be  $G_1A_1$  (due to isoelemental composition) a valid AAC descriptor would be



### *Acquisition of Accurate Mass Values by Mass Spectrometry*

Mass spectra were acquired by electrospray ionization on a new Finnigan LTQ FT (Thermo Electron, Bremen, Germany) hybrid instrument consisting of a linear quadrupolar ion trap and a Fourier transform ion cyclotron resonance mass spectrometer. Mass accuracies are specified by the manufacturer to be better than  $\pm 2$  ppm with external calibration carried out once a week. In practice it was found that mass accuracies are better than  $\pm 1$  ppm if the ion spray is sufficiently stable. A mass resolving power of 400,000 at  $m/z = 400$  u was chosen, which corresponds to a mass resolving power of 160,000 at  $m/z = 1000$  u in FT-ICR mass spectrometry. Mass values were determined by the standard software of the instrument, which is very similar in appearance to the software of the LCQ or LTQ ion trap instruments of the company, and were transferred to the Peptide LEGO software through the Windows XP clipboard.

### *Calculation of Accurate Mass Values of Peptides and Fragment Ions*

All calculations of accurate mass values of peptide ions and peptide fragment ions were performed using the

computer program Peptide LEGO (copyright Bernhard Spengler, 1996–2003) under Windows XP or Linux, which is especially designed for the determination and evaluation of accurate mass values of precursor ions and possible fragment ions, elemental compositions and amino acid compositions. Peptide composition analysis, CBS, and manual de novo sequencing is possible with this program, based on the strategy described in this paper.

## Results and Discussion

### *The Example Peptide: Substance P*

To show the power of accurate mass values and of the described procedure, the peptide Substance P ( $M_{\text{mono}} = 1346.7281$  u,  $[M + H]^+_{\text{mono}} = 1347.7354$  u, 11 amino acids, RPKPQQFFGLM-NH<sub>2</sub>) was chosen. The following procedure was performed semi-automatically using Peptide LEGO. At mass 1347 u the number of possible AAC per peptidogenic elemental composition (PEC) is in the range of 100 and the accuracy that would be required to determine the PEC unequivocally from the precursor mass alone is in the range of 20 ppb, as described in the following.

The peptide Substance P has been the subject of numerous fundamental investigations in the field of peptide fragmentation mechanisms [28–32]. The fragmentation pattern of Substance P is especially complicated because of the appearance of intense and numerous internal fragment ions, because of sequence repetitions and because of multiple possibilities of isobaric sequence doublets. As a result, sequencing of Substance P, if handled as an unknown peptide, usually does not lead to an unequivocal result, either by manual approaches or by automatic algorithms, but to a large number of possible structures. The following demonstration shows, that the determination of the correct amino acid composition (AAC) results in a small number of ambiguities, while the subsequent sequence analysis on the basis of the correct AA composition easily gives the correct sequence result.

The prerequisite for AAC determination on the basis of precursor and product ion mass values is to achieve sufficient reliability and precision (i.e., sufficient accuracy) in the determination of these data. Today, only Fourier transform mass spectrometry is able to provide mass accuracies in the range below 2 ppm using external calibration. Using a combination of a quadrupolar ion trap and a Fourier transform ion cyclotron resonance cell, MS<sup>n</sup> data can be acquired with high operational simplicity using routine ion formation and fragmentation protocols established for ion trap mass spectrometry.

As an example of the usefulness of high mass accuracy in the determination of primary structure, the mass region of smaller fragment ions of Substance P shall be described. It was a matter of discussion earlier [28], what the real nature is of fragment ions that appear at

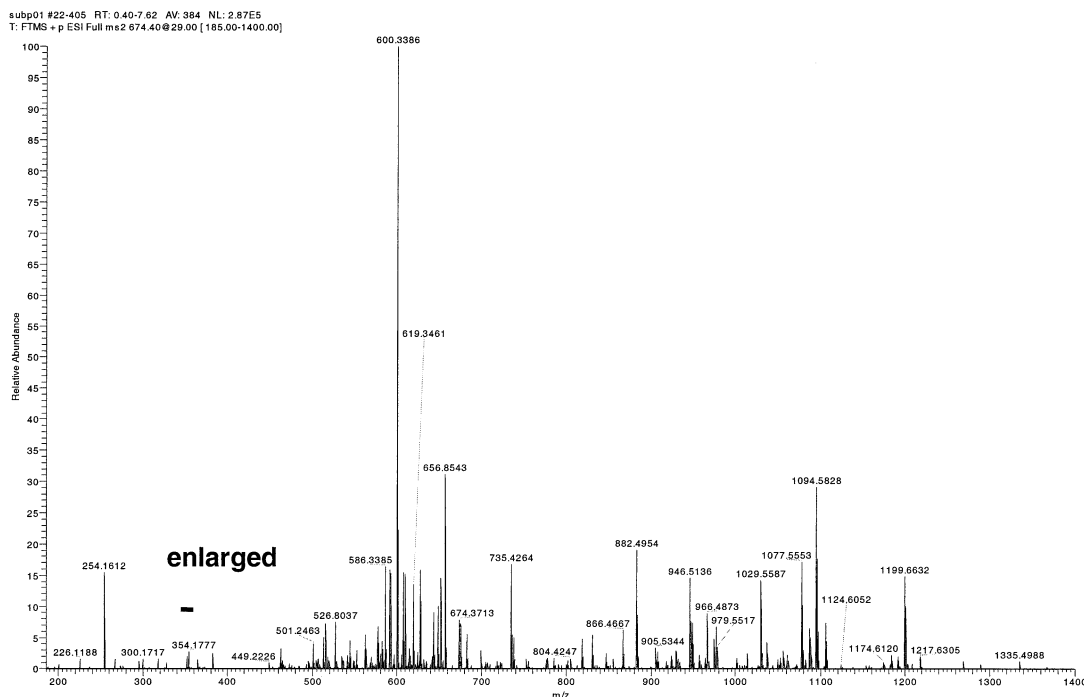


Figure 3. ESI-IT-FTMS spectrum of Substance P (RPKPQQFFGLM-NH<sub>2</sub>,  $M_{\text{mono}} = 1346.7281$  u)

the nominal mass values of calculated fragments  $a_3$ ,  $a_2$ , and  $a_1$ . Employing hydrogen–deuterium exchange of acidic protons, it was found that the number of exchangeable protons of these fragment ions does not match the expected number. It was concluded that the observed ions cannot be  $a_3$ ,  $a_2$ , and  $a_1$  ions, but in fact are either internal ions or some unexpected rearrangement ions, while a-type ions 3 to 1 instead are missing. The lack of formation of fragment ions  $a_3$ ,  $a_2$ , and  $a_1$  was especially surprising, since the complete list of fragment ions  $a_{11}$  down to  $a_4$  were obtained in high intensity [28]. ESI-IT-FTMS of Substance P resulted in the spectrum shown in Figure 3.

Today, using high-accuracy mass spectrometry, the open question raised by MALDI-PSD-TOF mass spectrometry in year 1993 is easily solved. Looking at the section of the product ion spectrum containing the region of the expected  $a_3$  ion, as an example, (Figure 4) mass determination unequivocally showed that the observed fragment ion is an internal ion PQQ =  $(B_6Y_8)_3 + 1$ , not an internal ion KPQ =  $(B_5Y_9)_3 + 1$  or any other possible internal ion or rearrangement product, and that the expected n-terminal ion  $a_3$  is indeed completely missing.

#### The Relation Between Peptide Ion Masses and Composition

Peptide mass values are not dispersed homogeneously across the complete mass range but are located in clusters [26, 27] around a mean value of  $n * 1.000458$  with  $n$  being a whole positive number. For the cluster of

all possible peptide masses around mass 1005.46 u, as an example, the distribution is shown in Figure 5. Determination of the amino acid composition from the experimentally determined accurate mass is much simpler if the target mass is located in the boundaries of the peptide mass distribution with only a low number of possible amino acid compositions, compared to the case that the target mass is located in the peak of the mass distribution function. For all calculations in the following, example mass values were therefore chosen from the center of the distribution function in order to describe the worst case with respect to variability of compositions.

The variability of peptide compositions was determined as a function of the given accuracy of mass determination. In the mass range typical for tryptic peptides or, e.g., for MHC-restricted peptides [25] it was found that mass accuracies of 0.1 to 1 ppm are necessary to achieve the highest possible certainty in determination of the amino acid composition.

Figure 6 as an example shows the number of possible amino acid compositions of a peptide whose protonated singly charged parent ion has a monomolecular mass of 1005.4433 u, if the ion mass was determined with different accuracies. When determined with 0.1 ppm accuracy the peptide ion mass points to a set of seven possible amino acid compositions (AAC). This number cannot be further reduced since all of the seven AAC have the same elemental composition and thus the same accurate mass. At 1 ppm accuracy five different peptidogenic elemental compositions (PEC) can be found within the boundaries of mass determination

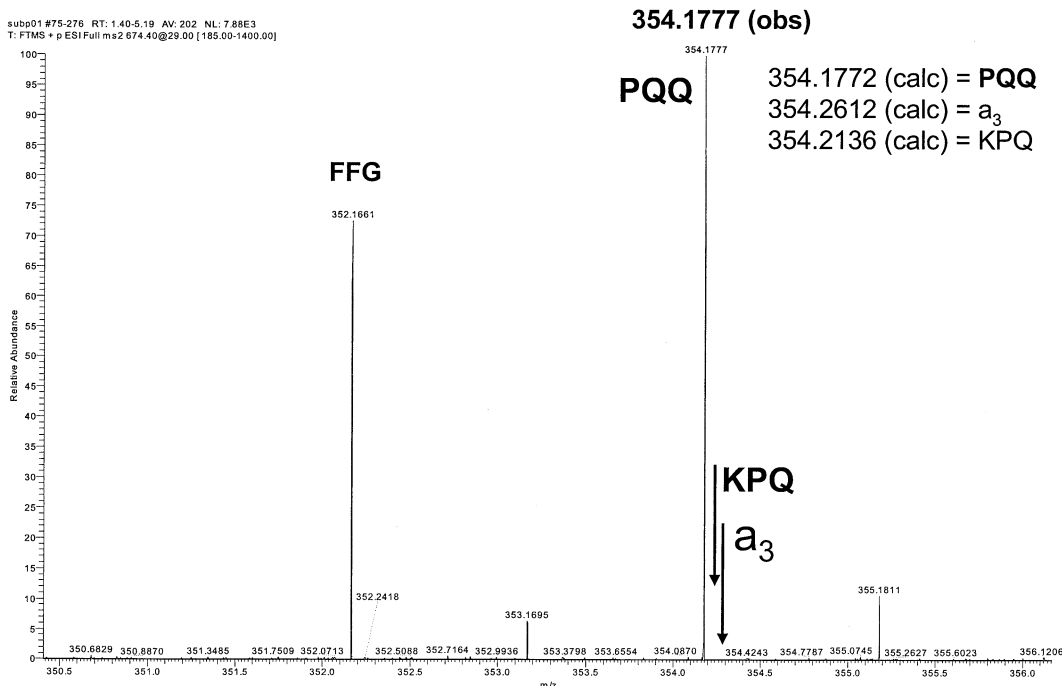


Figure 4. Section of ESI-IT-FTMS product ion spectrum of Substance P.

accuracy. Table 1 lists the results of calculation for this example. This example shows that the mass accuracies available with state-of-the-art mass spectrometers lie exactly within the areas of common analytical requirements and that there is still some room and need for instrumental improvements.

The isobaric nature of different amino acid compositions is certainly a serious limitation of mass spectrometrical information obtainable even by an ideal instrument. It is, however, demonstrated in this paper, that using combinatorial logic, this limitation can typically be overcome to a large extent.

Amino acid variability per elemental composition is shown in Figure 7 as a function of peptide mass. For a

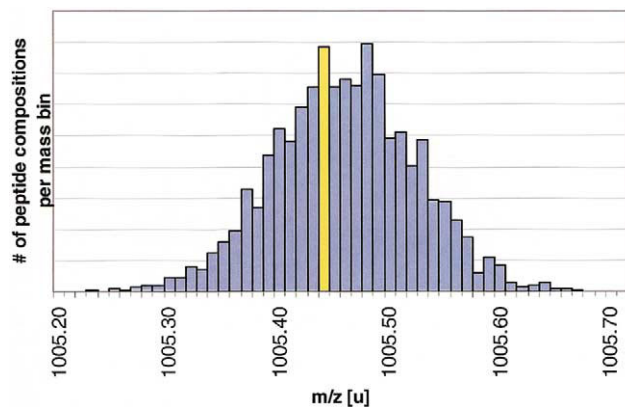


Figure 5. Number of possible peptide amino acid compositions in arbitrary units for the cluster around  $m/z = 1005.46$  u. A mass value from the segment displayed highlighted was taken for the following calculation.

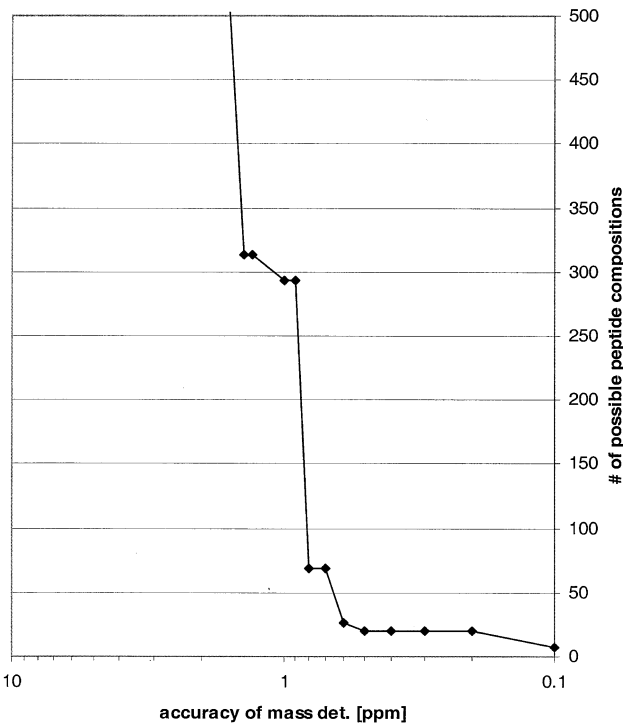


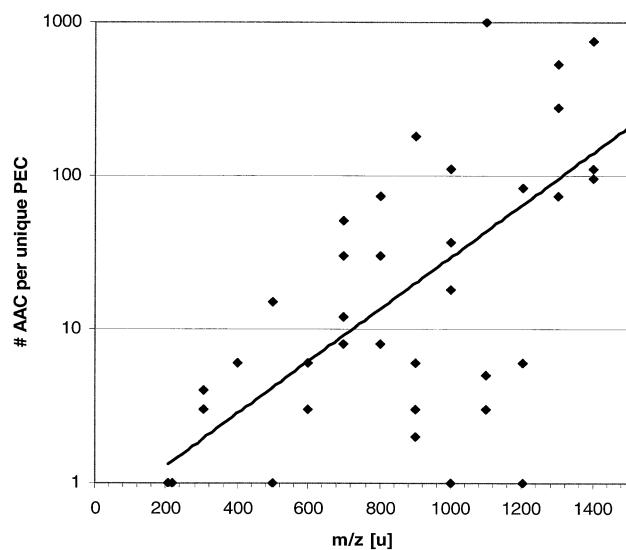
Figure 6. Number of possible amino acid compositions of peptides for a given example mass of 1005.4433 u as depending on the accuracy of mass determination. Calculations were performed including all of the 19 amino acids A C D E F G H (I = L) K M N P Q R S T V W and Y. Mass calculations correspond to the monomolecular isotopomers of the singly positively charged protonated peptides (C terminus -OH, N terminus -H).

**Table 1.** Number of possible amino acid compositions (AAC) and amino acid based (peptidogenic) elemental compositions (PEC) of peptides for a given example mass of 1005.4433 u as depending on the accuracy of mass determination

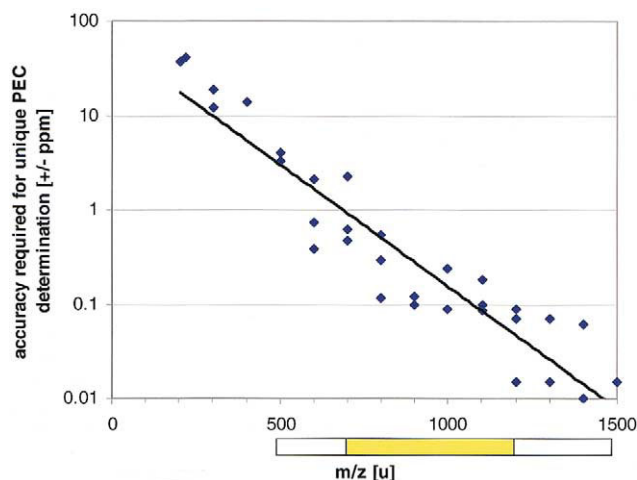
Accuracy [ $\pm$ ppm]	# of AAC	# of PEC
9	3927	not determined
8	3577	not determined
7	3480	not determined
6	3022	not determined
5	1962	19
4	1582	16
3	1449	9
2	868	7
1.4	314	6
1.3	314	6
1	293	5
0.9	293	5
0.8	69	4
0.7	69	4
0.6	27	3
0.5	20	2
0.4	20	2
0.3	20	2
0.2	20	2
0.1	7	1

Calculations were performed for the 19 amino acids A C D E F G H (I = L) K M N P Q R S T V W and Y. Mass calculations correspond to the monomolecular isotopomers of the singly positively charged protonated peptides (C terminus –OH, N terminus –H).

set of arbitrarily chosen peptide mass values, whose accurate masses fall into the maxima of the mass distribution functions (see Figure 5), the numbers of possible amino acid compositions per unique peptidogenic elemental composition were calculated. The results show that for peptides composed of 8 to 10 amino acids (masses around 1000 u) the numbers of possible



**Figure 7.** Number of possible amino acid combinations for a unique peptidogenic elemental composition. Plotted values are those for selected accurate masses chosen from the maxima of the mass distribution function.



**Figure 8.** Calculated accuracies of mass determination required to achieve unique peptidogenic elemental compositions (PEC), as a function of peptide masses. Plotted values are those for selected accurate masses chosen at the maxima of the mass distribution function. The bar below the mass axis indicates the (complete and preferred) mass range of peptides composed of eight amino acids.

AAC per unique PEC are between 10 and 100. A method for determination of the amino acid composition of such peptides should ideally be able to differentiate between iso-PEC compositions.

The mass spectrometric goal (and limit) is to determine a unique PEC from the obtained experimental mass values. Figure 8 shows the calculated dependence of the accuracies of mass determination necessary for achieving unique PEC, as a function of the peptide mass. For nominal-mass intervals of 100 u, three to four accurate mass values were arbitrarily chosen from the maxima of the mass distribution functions (see Figure 5), and distances to nearest neighbors were calculated in ppm units. The calculation indicates that determination of the PEC of a peptide from the mass of the precursor ion alone is possible up to a mass of ca. 300 u if the instrumental accuracy is  $\pm 10$  ppm (routinely obtainable with e.g., a quadrupole time-of-flight instrument) and up to a mass of ca. 700 if the instrumental accuracy is  $\pm 1$  ppm (routinely obtainable with an FT-ICR instrument). Only with a (currently unavailable) instrumental mass accuracy of  $\pm 0.1$  ppm, the PEC of a peptide can be determined directly from the precursor ion mass up to a mass of ca. 1100 u. Even if the described calculation does not take into account each possible neighborhood of accurate masses it still shows the mean values of accuracy needed for PEC determination as a function of mass.

It has to be noted here that all calculations were performed for singly charged peptide ions. For multiply charged ions the mass limits for determination of unique PEC have to be divided by the charge of the ion, so that for example the mass limit for an instrumental accuracy of  $\pm 1$  ppm is ca.  $m/z = 700$  u for a singly charged ion while it is ca.  $m/z = 350$  u for a doubly charged ion.

**Table 2.** Observed, calculated and corrected mass values used for AAC determination as obtained on the basis of instrumental external calibration, computation and peptide-internal fine calibration respectively

$F^+_{\text{obs}}$ [u]	$F^+_{\text{calc}}$ [u]	Error <sub>obs/calc</sub> [ppm]	$F^+_{\text{corrected}}$ [u]	Error <sub>corr/calc</sub> [ppm]
352.16610	352.165568	+1.51	352.16552	-0.14
382.25662	382.256115	+1.32	382.25599	-0.33
465.25049	465.249632	+1.84	465.24972	+0.19
501.24631	501.245609	+1.40	501.24548	-0.25
449.22259	449.221703	+1.97	449.22185	+0.32

From these results it is clear that determination of the accurate precursor ion mass alone is sufficient for the unequivocal determination of the PEC or AAC of peptides up to mass 700 u, with currently available instrumental accuracies. For peptides above mass 700 u it is described in the following how accurate fragment ion mass values can additionally be employed to still identify the PEC.

### Combination with MS-MS Data: Peptide Compositioning

In this section it is shown that, in addition to the accurate precursor ion mass, only a few accurate masses of MS/MS fragment ions are necessary to obtain a unique PEC and a small number of candidate AACs for an unknown peptide.

The general strategy for this is to use an iterative process consisting of basically three iteration tasks:

#### Task 1.

Determine the amino acid composition of a fragment ion in the mass range of 300 to 400 u. In this mass range the PEC can usually be determined unequivocally from the accurate mass if the instrumental accuracy is better than  $\pm 1$  ppm (see Figure 8). Besides the PEC, the amino acid composition often can be determined unequivocally as well under these conditions. If there is more than one candidate AAC for a determined PEC, the list of these candidate AACs can be evaluated in terms of ranges of appearance for each amino acid. These ranges can be used as a criterion for the following task.

#### Task 2.

Use the resulting AAC restriction (number or range of numbers, respectively, of amino acids within the candidate AACs) determined for the fragment ion to restrict the possible numbers of amino acids of the precursor peptide ion. For each type of amino acids its number in the fragment (or its minimum number within the list of candidate AACs of the fragment, resp.) is taken as a minimum condition for the precursor peptide. Using the restricted range of AACs the list of possible AACs of the peptide can be computed from its accurate mass value. This typically leads to an already limited number of possibilities.

#### Task 3.

From the list of candidate AACs of the peptide determine the highest numbers of amino acids of each type for the precursor and use them as maximum restrictions for the next iteration. The same fragment ion of task 1 can be evaluated again using this new maximum condition, often leading to a stronger restriction of candidate AACs within a unique PEC of the fragment ion. If the repeated evaluation of the fragment ion does not lead to any modification of the restrictions, go to another fragment ion and repeat tasks 1 to 3 until no further change in the list of possible AAC of the precursor peptide is obtained.

It was found that using this three-step iteration procedure the PEC of peptides can often be determined unequivocally and the list of possible AAC can be reduced to a small number, even if (for the given mass range and instrumental accuracy) the AAC determined from the accurate mass of the precursor ion alone is rather ambiguous (see Figures 7 and 8). Additionally, partial sequence information can be extracted from the above procedure, on the basis of determination of fragment type and composition of the evaluated fragments. The probability of obtaining a unique PEC depends on the mass of the peptide and the instrumental mass accuracy.

*Determination of the AAC of Substance P.* On the basis of accurate MS-MS data, a list of product ion masses was obtained. Isotopomeric patterns were evaluated to determine the charge state of the observed fragment ions.

For the following determination of AACs, the peptide was treated as a complete unknown. It was undefined as well, whether the unknown peptide under investigation was a C-terminal amide or acid. The observed precursor ion  $m/z$  value was 674.37143 u for a doubly charged ion  $[M + 2H]^{2+}$ . The corresponding observed  $m/z$  value for a singly charged ion  $[M + H]^+$  would thus be 1347.73558 u.

Fragment ion signals were mass determined by using the once-a-week external calibration of the instrument. The partial list of observed mass values is shown in Table 2. Additional internal fine calibration is described further down.

The following steps were necessary to achieve the final AAC information:

First iteration:

Iteration task 1.

Fragment ion 352.16610 u (external calibration; final internally calibrated mass value 352.16552 u) was unequivocally identified by Peptide LEGO as an internal b-type ion of composition ( $F_2G_1$ ). No other compositions are possible within  $\pm 2$  ppm accuracy of any fragment ion type y, a, or b. With mass accuracies of less than  $\pm 9$  ppm the identification would have been ambiguous, since at mass 352.1689 u two a-type ion compositions exist, ( $F_1M_1T_1$ ) and ( $C_1M_1Y_1$ ), which are 8.9 ppm off the observed mass. A mass resolving power of 113,600 is necessary to resolve the possible neighboring mass signals. It is obvious that only FTMS is able to provide a reliable result in this case.

Iteration task 2.

Using the minimum condition that the unknown peptide contains at least one G and two F, the number of possible AAC for mass 1347.73558 within  $\pm 2$  ppm accuracy reduced to **97** assuming an amide peptide and another **74** assuming an acid peptide. This corresponds to a reduction of possibilities by two orders of magnitude compared with the unconstrained calculation: The number of all possible AAC for mass 1347.73558 within 2 ppm accuracy is 8648 assuming an amide peptide and 7535 assuming an acid peptide.

Iteration task 3.

Since there is only one candidate AAC for fragment ion mass 352.16610 u, iteration task 3 did not lead to a further reduction of possibilities within this iteration. The maximum condition was used for the next iteration.

Second iteration:

Iteration task 1.

Fragment ion 382.25597 u was identified as a b-type ion of composition ( $K_1P_1R_1$ ). Other possibilities within  $\pm 2$  ppm accuracy would be a-type ion composition ( $G_1P_1R_1V_1$ ) or y-type ion composition ( $H_1K_1V_1$ ) (amide at C-terminus). No y-type ion composition was found assuming a C-terminal acid instead of an amide. The a-type ion was rejected since its appearance is quite unlikely without appearance of a corresponding b-type ion ( $m/z = 410$  u). The possible y-type ion was rejected since an amide-type y-ion should be accompanied by appearance of a y-NH<sub>3</sub> ion ( $m/z = 365$  u) which was not observed. Therefore ( $K_1P_1R_1$ ) was included in the minimum condition of AAC of the unknown peptide.

Iteration task 2.

Using these constraints,  $m/z$  value 674.37143 u (doubly charged precursor ion) or 1347.73558 u (recalculated singly charged precursor ion) resulted in **19** possible AAC assuming an amide peptide and another **two** possible AAC assuming an acid peptide.

Iteration task 3.

Since there is only one candidate AAC for fragment ion mass 382.25597 u, iteration task 3 did not lead to a further reduction of possibilities within this iteration. The maximum condition was used for the next iteration.

Third iteration:

Iteration task 1.

Evaluation of observed fragment ion 465.24973 u resulted in the b-type ion composition possibilities ( $A_1F_2V_1$ ) and ( $F_2G_1L_1$ ), the y-type ion composition possibility ( $K_1M_1S_1T_1$ ) and the a-type ion possibility ( $F_2P_1T_1$ ).

Iteration task 2.

Using ( $K_1M_1S_1T_1$ ) or ( $F_2P_1T_1$ ) as an additional minimum constraint in the determination of AAC of the unknown peptide resulted in zero possibilities, thus the fragment ion 465.24973 u cannot be a y- or a-type ion but has to be a b-type ion (internal or n-terminal). Using ( $A_1F_2V_1$ ) or ( $F_2G_1L_1$ ) as additional constraints in the determination of AAC of the unknown peptide reduced the list of possible compositions from **19** to **14** (amide peptides) and from **two** to **one** (acid peptide). The list of remaining possibilities of AAC of the unknown peptide is:

H ( $A_3F_2G_2K_1M_1P_2R_1V_1$ ) NH<sub>2</sub>  
 H ( $A_2F_2G_3K_1L_1M_1P_2R_1$ ) NH<sub>2</sub>  
 H ( $A_2F_2G_1K_1L_1M_1N_1P_2R_1$ ) NH<sub>2</sub>  
 H ( $A_2F_2G_1K_1M_1P_2Q_1R_1V_1$ ) NH<sub>2</sub>  
 H ( $A_1F_2G_2K_1L_1M_1P_2Q_1R_1$ ) NH<sub>2</sub>  
 H ( $A_4C_1F_2G_1K_1L_1P_2R_1$ ) NH<sub>2</sub>  
 H ( $A_2C_1F_2G_2K_1P_2R_1V_2$ ) NH<sub>2</sub>  
 H ( $A_1C_1F_2G_3K_1L_1P_2R_1V_1$ ) NH<sub>2</sub>  
 H ( $C_1F_2G_4K_1L_2P_2R_1$ ) NH<sub>2</sub>  
 H ( $F_2G_1K_1L_1M_1P_2Q_2R_1$ ) NH<sub>2</sub>  
 H ( $A_1C_1F_2G_1K_1L_1N_1P_2R_1V_1$ ) NH<sub>21</sub>  
 H ( $A_1C_1F_2G_1K_1P_2Q_1R_1V_2$ ) NH<sub>2</sub>  
 H ( $C_1F_2G_2K_1L_2N_1P_2R_1$ ) NH<sub>2</sub>  
 H ( $C_1F_2G_2K_1L_1P_2Q_1R_1V_1$ ) NH<sub>2</sub>  
 H ( $F_2G_1H_1K_2L_1M_1P_1R_1S_1$ ) OH

Iteration task 3.

Iteration task 3 did not lead to rejection of possible AACs ( $A_1F_2V_1$ ) or ( $F_2G_1L_1$ ) of fragment ion 465.24973 u within this iteration. The maximum condition was used for the next iteration.



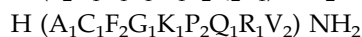
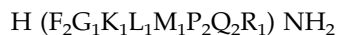
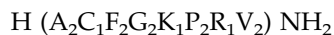
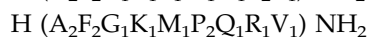
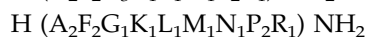
Fourth iteration:

Iteration task 1.

Using this list of possible AAC of the unknown peptide, the observed fragment ion at 501.24547 u was evaluated. No possible  $\gamma$ -type ions (amide or acid type) and no possible  $\alpha$ -type ions were found within  $\pm 2$  ppm deviation. The four possible  $\beta$ -type ions (internal or n-terminal) are  $(A_2F_1G_2P_1)$ ,  $(A_2F_1N_1P_1)$ ,  $(A_1F_1G_1P_1Q_1)$ , and  $(F_1P_1Q_2)$ .

Iteration task 2.

The list of possible AAC of amide peptides thus reduces to the following 8 AACs:



all of them having the same elemental composition of  $C_{63}H_{98}N_{17}O_{14}S_1$ .

No possible AACs for fragment ion 501.24547 u were found when assuming that the unknown peptide is an acid peptide. Therefore at this stage it is definite, that the peptide is an amide peptide (neglecting other possible modifications).

The resulting list of 8 possible AAC of the unknown peptide can be combined in terms of common components to the AAC descriptor:  $H\{K_1R_1F_2P_2[(A_1V_1M_1)/(G_1L_1M_1)/(V_2C_1)][(G_2A_2)/(N_1A_2)/(G_1A_1Q_1)/(Q_2)]\}NH_2$ , with units separated by slashes being interchangeable.

Iteration task 3.

Iteration task 3 did not lead to rejection of possible AACs  $(A_2F_1G_2P_1)$ ,  $(A_2F_1N_1P_1)$ ,  $(A_1F_1G_1P_1Q_1)$ , or  $(F_1P_1Q_2)$  of fragment ion 501.24547 u within this iteration.

This final list cannot be further reduced without employing additional methods, due to the isobaric nature of the pairs (Q/GA), (N/GG), (VC/AM), and (GL/AV). Still this short list of eight possible AACs can easily be used to determine the most probable amino acid sequence, as shown in the section on composition-based sequencing.

*Employing H/D exchange and exchangeable hydrogen analysis.* If it is necessary to achieve a unique AAC, an extremely useful method for obtaining further information to reduce the list of possibilities of AAC, hydrogen–deuterium exchange can be employed [28] as a very simple, quick, reversible, and sensitive method of sample derivatization. Determination of the number of exchangeable hydrogens using this

method in the example of the above peptide leads to a reduction of possible AAC to only the two matching compositions



and



having 23 exchangeable hydrogens, as measured. These two possibilities can easily be distinguished by evaluating fragment ion 449.22186 u, leading to possible AACs of  $\beta$ -type  $(F_1G_1L_1M_1)$  or  $(C_1F_1V_2)$  as a subset of the two possible peptide compositions. While  $(F_1G_1L_1M_1)$  contains five exchangeable hydrogens,  $(C_1F_1V_2)$  contains six exchangeable hydrogens. With the additional information obtained by hydrogen–deuterium exchange, therefore, the amino acid composition of the unknown peptide can unequivocally be determined to



which is indeed the composition of Substance P.

*Internal fine calibration procedure.* The mass accuracy of the Finnigan LTQ-FT instrument is specified as being better than  $\pm 2$  ppm. As can be seen from Table 2 all of the values used for the AAC determination procedure lie well within these limits. It is, however, obvious that external calibration at the time of measurement was already shifted systematically, since all deviations have positive values above +1 ppm. It is, therefore, reasonable to employ an additional fine tuning procedure to shift values back to an averaged situation. Since mass and frequency are linearly correlated in FT-ICR mass spectrometry, a linear shift of mass values should already lead to a considerable improvement of mass accuracy. This shift can be controlled peptide-internally, by using the exact calculated mass values of unequivocally defined PEC of the peptide fragments. In the first step of determining the AAC of the peptide under investigation, evaluating mass value 352.1661 u, the instrumental external calibration was used. The determined PEC and AAC resulted in a calculated mass value of 352.165568 u. The shift between observed and calculated mass value was used to define a first correction factor of 0.999998489. Evaluating the second fragment ion mass value in the following step was performed using the corrected mass values, and the determined PEC mass value was used together with the calculated mass of the first fragment to form a second order, averaged correction factor. After evaluating all of the listed fragment ions, a final averaged correction factor of 0.999998350 resulted in errors of the corrected versus the calculated fragment ion masses well below  $\pm 0.4$  ppm.

**Table 3.** Weighting factors used to score sequences

Fragment ion type a, b, y	1.0
Fragment ion type z, b + H <sub>2</sub> O	0.5
Fragment ion type c, d, d', x	0.1
Fragment ion type v, w, w'	0.05
Internal fragment ion	0.3
Neutral losses (NH <sub>3</sub> , H <sub>2</sub> O, ...)	0.5 .. 0.05
Ratio of numbers of explained observed ions and theoretical number of calculated ions	$N_{\text{explained-observed}}/N_{\text{theoretical}}$
Ratio of numbers of explained observed ions and observed ions	$N_{\text{explained-observed}}/N_{\text{observed}}$

The score factor for each possible sequence was calculated by averaging the applicable weighting factors for each explained observed fragment ion, by multiplying the result by 100 and by the two number ratios.

### Composition-Based Sequencing (CBS)

As soon as the amino acid composition of the target peptide is determined or the list of possible compositions is reduced to a low number, possible amino acid sequences (AAS) can be assessed by appropriate algorithms.

The number of possible AAS (i.e., number of possible peptides) is  $v = n^k$  where  $v$  = the number of variations (number of possible sequences),  $k$  = peptide length (number of amino acid positions) and  $n$  = number of different amino acids. For a peptide of unknown amino acid composition and a length of 8 amino acids, the number of possible sequences is thus  $v = 20^8 = 2.56 \times 10^{10}$ .

This is a number of list components which can typically not be computed and assessed in reasonable time by standard computers. If, on the other hand, the amino acid composition is known, the number of possible sequences is only  $P = k!$  where  $P$  = the number of permutations (number of possible sequences for a given AAC) and  $k$  = the peptide length (number of amino acid positions). For the same peptide length of eight amino acids, the number of possible sequences then is only  $P = 8! = 40320$ .

It is obvious that this number of possibilities can easily be computed in short time by regular computers. Therefore in this example, accurate mass determination results in a reduction of computing time by a factor of 634,920 when compared with permutation based de novo sequencing. This might be quite astonishing and certainly has not been realized to its full extent so far.

Having the amino acid composition of the target peptide in hand, the number of sequences to be tested is 40320 for a 8 AA peptide and 39,916,800 for a 11 AA peptide like Substance P. This number can be further reduced if (as in the case of Substance P) several amino acids are present in multiple copies. Including MS-MS preinformation obtained for certain fragments as described above, can be used to group several amino acids within the permutation procedure, which reduces the number of permutations considerably. Assuming, for example, that the fragment identified as composition (K<sub>1</sub>P<sub>1</sub>R<sub>1</sub>) is an N-terminal ion (rather than an internal ion) reduces the number of permutations (sequences to be assessed) from  $11! = 39,916,800$  to  $P = 3! \times 8! =$

241,920 for the AAC descriptor H-(K<sub>1</sub>P<sub>1</sub>R<sub>1</sub>)(X<sub>8</sub>)-NH<sub>2</sub> with X denoting various amino acids.

Assuming that the same fragment is an internal ion rather than an N-terminal ion results in

$P = 3! \times 9! = 2,177,280$  permutations for the AAC descriptor H-[(K<sub>1</sub>P<sub>1</sub>R<sub>1</sub>)X<sub>8</sub>]-NH<sub>2</sub>.

Another (non-overlapping) group like (F<sub>2</sub>G<sub>1</sub>), assumed to be an internal ion, results in  $P = (3!/2!) \times 3! \times 7! = 90,720$  permutations for the corresponding AAC descriptor H-[(K<sub>1</sub>P<sub>1</sub>R<sub>1</sub>)(F<sub>2</sub>G<sub>1</sub>)X<sub>5</sub>]-NH<sub>2</sub>.

The remaining units contain a Q doublet, which reduces the number of permutations to  $P' = P/2 = 45,360$  for the corresponding AAC descriptor H-[(K<sub>1</sub>P<sub>1</sub>R<sub>1</sub>)(F<sub>2</sub>G<sub>1</sub>)L<sub>1</sub>M<sub>1</sub>P<sub>1</sub>Q<sub>2</sub>]-NH<sub>2</sub>.

Without performing further groupings, this number of possibilities can already be computed easily in reasonable time. Using accurate masses within an uncertainty limit of  $\pm 2$  ppm, it was found that finding the correct sequence out of all possible sequences is a rather simple goal. A very simple algorithm is currently used in Peptide LEGO ignoring any experience-based or rule-based validation steps and any intensity correlations. Peptide LEGO permutes possible sequences on the basis of the determined AAC or the list of possible AAC. For each sequence tested, Peptide LEGO simply counts the number of hits when comparing theoretical and observed fragment ion mass signals within the chosen uncertainty limit. Each hit is weighed with a fragment-type specific factor, so that for example internal ions are taken less important than N- or C-terminal ions. Weighting factors used are listed in Table 3. With this simple algorithm, theoretical fragment ion masses of all 45,360 possible sequences for the determined AAC were calculated and were compared to the list of observed fragment ion masses. When starting with the correct AAC of Substance P, Peptide LEGO found the correct sequence RPKPQQFFGLM-NH<sub>2</sub> of the peptide as the first hit within seconds. Thus in combination with hydrogen–deuterium exchange, the sequence of Substance P, when treated as an unknown peptide, can be found easily with high confidence.

Without additional information about the number of exchangeable hydrogens, eight possible AAC remained. Testing all of the eight compositions led to the correct sequence of the correct composition of Substance P, and

**Table 4.** List of top scored sequences for each of the final eight possible AACs of the example peptide, summarized as AAC H {K<sub>1</sub>R<sub>1</sub>F<sub>2</sub>P<sub>2</sub> [(A<sub>1</sub>V<sub>1</sub>M<sub>1</sub>)/(G<sub>1</sub>L<sub>1</sub>M<sub>1</sub>)/(V<sub>2</sub>C<sub>1</sub>)] [(G<sub>2</sub>A<sub>2</sub>)/(N<sub>1</sub>A<sub>2</sub>)/(G<sub>1</sub>A<sub>1</sub>Q<sub>1</sub>)/(Q<sub>2</sub>)] NH<sub>2</sub>

Score [arbitrary units]	Number of equally scored sequences	Sequence (example sequence in italics, resp.)	Number of observed ions matched	AAC
0.1820	1	<b>RPKPQQFFGLM-NH<sub>2</sub></b>	30	H (F <sub>2</sub> G <sub>1</sub> K <sub>1</sub> L <sub>1</sub> M <sub>1</sub> P <sub>2</sub> Q <sub>2</sub> R <sub>1</sub> ) NH <sub>2</sub>
0.1607	4	<i>RPKPQGAFGLM-NH<sub>2</sub></i>	30	H (A <sub>1</sub> F <sub>2</sub> G <sub>2</sub> K <sub>1</sub> L <sub>1</sub> M <sub>1</sub> P <sub>2</sub> Q <sub>1</sub> R <sub>1</sub> ) NH <sub>2</sub>
0.1473	1	RPKNAAPFFGLM-NH <sub>2</sub>	28	H (A <sub>2</sub> F <sub>2</sub> G <sub>1</sub> K <sub>1</sub> L <sub>1</sub> M <sub>1</sub> N <sub>1</sub> P <sub>2</sub> R <sub>1</sub> ) NH <sub>2</sub>
0.1425	9	<i>RPKAPGGAFGLM-NH<sub>2</sub></i>	30	H (A <sub>2</sub> F <sub>2</sub> G <sub>3</sub> K <sub>1</sub> L <sub>1</sub> M <sub>1</sub> P <sub>2</sub> R <sub>1</sub> ) NH <sub>2</sub>
0.1396	1	RPKPQAGFFVAM-NH <sub>2</sub>	28	H (A <sub>2</sub> F <sub>2</sub> G <sub>1</sub> K <sub>1</sub> M <sub>1</sub> P <sub>2</sub> Q <sub>1</sub> R <sub>1</sub> V <sub>1</sub> ) NH <sub>2</sub>
0.1295	1	RPKAPGAGFFVAM-NH <sub>2</sub>	29	H (A <sub>3</sub> F <sub>2</sub> G <sub>2</sub> K <sub>1</sub> M <sub>1</sub> P <sub>2</sub> R <sub>1</sub> V <sub>1</sub> ) NH <sub>2</sub>
0.0760	11	<i>RPKGPAAAGFFVVCV-NH<sub>2</sub></i>	22	H (A <sub>2</sub> C <sub>1</sub> F <sub>2</sub> G <sub>2</sub> K <sub>1</sub> P <sub>2</sub> R <sub>1</sub> V <sub>2</sub> ) NH <sub>2</sub>
0.0810	2	<i>RPKGPQAFFVVCV-NH<sub>2</sub></i>	21	H (A <sub>1</sub> C <sub>1</sub> F <sub>2</sub> G <sub>1</sub> K <sub>1</sub> P <sub>2</sub> Q <sub>1</sub> R <sub>1</sub> V <sub>2</sub> ) NH <sub>2</sub>

"Score" is a relative value resulting from the scoring process as described in the text. "Number of equally scored sequences" describes, how many sequences having the same AAC were scored with the same scoring value, thus cannot be differentiated. Calculations were performed assuming a mass accuracy of  $\pm 2$  ppm.

results were clearly unambiguous using the simplistic approach of Peptide LEGO. The scoring results for the top scored AACs are shown in Table 4. Including intensity information and intensity assumptions, or including charge location expectations, however, would easily improve the confidence of the results further, but appears to be unnecessary at least for the given example.

The procedure described in this paper was used to test other peptides of lengths between eight and ten amino acids. In all cases unequivocal results were obtained for the list of possible AAC and for the sequence. There is no doubt that future sequencing algorithms based on peptide amino acid composition analysis (CBS) will result in considerably faster determination of the peptide sequence, compared with de novo sequencing. To develop such algorithms is a rather straight forward task for the near future. The important result of this paper is that the determination of the sequence is simple and reliable if the composition is already known.

Finally it has to be noted that peptide compositioning can be useful for sequence determination of much larger peptides as well. Even if it is not possible to determine a unique AAC of a large peptide, future sequencing programs could still make use of range information determined from longer lists of candidate AACs and use these restrictions for improved scoring in de novo sequencing.

## Conclusions

High mass accuracy as available from state-of-the-art mass spectrometers such as linear trap-Fourier transform ion cyclotron resonance hybrid mass spectrometers has been shown in this paper to provide a new quality of structural information in mass spectrometry, when employed to determine the amino acid sequence of completely unknown peptides. Determination of the amino acid composition in the first step was found to be the major task and is possible for medium size peptides when including high accuracy MS-MS data. Subsequent

peptide sequencing is a rather simple task which can easily be performed by error-proof permutation based scoring algorithms.

The present study on the other hand shows that high mass accuracy and high mass resolution is of great importance for sequencing of unknown peptides or for structure analysis of other biopolymers. The complexity of operation of existing FT-ICR instruments until recently had resulted in mixed opinions about the necessity and appropriateness of accurate mass determination with this technique. Despite the fact that the utility and value of accurate mass determination has been recognized for some time, there was limited yearning to make use of it. With the advent of a new generation in technology, for example with the combination of FT-ICR mass spectrometry with quadrupolar ion trap technology controlled by user friendly software, this has changed considerably. Accurate mass values and highest mass resolution are now available in combination with the most simple instrument handling and high sample throughput. The ready availability of highly accurate mass values now appears to provoke the development of intelligent software that makes use of this new dimension of structural information.

## Acknowledgments

The authors gratefully acknowledge financial support by the Deutsche Forschungsgemeinschaft (grant number Sp 314-3/1-2).

## References

- Spengler, B. Postsource Decay Analysis in Matrix-Assisted Laser Desorption Ionization Mass Spectrometry of Biomolecules. *J. Mass Spectrom.* **1997**, *32*, 1019–1036.
- Dancik, V.; Addona, T.A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De Novo Peptide Sequencing via Tandem Mass Spectrometry. *J. Comp. Biol.* **1999**, *6*, 327–342.
- Bartels, C. Fast Algorithm for Peptide Sequencing by Mass Spectrometry. *Biomed. Environ. Mass Spectrom.* **1990**, *19*, 363–368.
- Hines, C. W.; Falick, A. M.; Burlingame, A. L.; Gibson, B. W. Pattern-Based Algorithm for Peptide Sequencing from Tan-

- dem High Energy Collision-Induced Dissociation Mass Spectra. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 326–336.
- Taylor, J. A.; Johnson, R. S. Sequence Database Searches via de Novo Peptide Sequencing by Tandem Mass Spectrometry. *Rapid. Commun. Mass Spectrom.* **1997**, *11*, 1067–1075.
  - Lehmann, W. D.; Bohne, A.; von der Lieth, C. W. The Information Encrypted in Accurate Peptide Masses—Improved Protein Identification and Assistance in Glycopeptide Identification and Characterization. *J. Mass Spectrom.* **2000**, *35*, 1335–1341.
  - Schlosser, A.; Lehmann, W. D. Patchwork Peptide Sequencing: Extraction of Sequence Information from Accurate Mass Data of Peptide Tandem Mass Spectra Recorded at High Resolution. *Proteomics* **2002**, *2*, 524–533.
  - Ledford, E. B., Jr.; Ghaderi, S.; White, R. L.; Spencer, R. B.; Kulkarni, P. S.; Wilkins, C. L.; Gross, M. L. Exact Mass Measurement by Fourier Transform Mass Spectrometry. *Anal. Chem.* **1980**, *52*, 463–468.
  - White, R. L.; Onyiriuka, E. C.; Wilkins, C. L. Exact Mass Measurement in the Absence of Calibrant by Fourier Transform Mass Spectrometry. *Anal. Chem.* **1983**, *55*, 339–343.
  - Huang, N.; Siegel, M. M.; Kruppa, G. H.; Laukien, F. H. Automation of a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for Acquisition, Analysis, and E-Mailing of High-Resolution Exact-Mass Electrospray Ionization Mass Spectral Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1166–1173.
  - Burton, R.; Matuszak, K. P.; Watson, C. H.; Eyler, J. R. Exact Mass Measurements Using a 7 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer in a Good Laboratory Practices-Regulated Environment. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1291–1297.
  - Ramjit, H. G.; Kruppa, G. H.; Spier, J. P.; Ross, C. W. III; Garsky, V. M. The Significance of Monoisotopic and Carbon-13 Isobars for the Identification of a 19-Component Dodecapeptide Library by Positive Ion Electrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1368–1376.
  - Conrads, T. P.; Anderson, G. A.; Veenstra, T. D.; Pasa-Tolic, L.; Smith, R. D. Utility of Accurate Mass Tags for Proteome-Wide Protein Identification. *Anal. Chem.* **2000**, *72*, 3349–3354.
  - Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements. *Proteomics* **2002**, *2*, 513–523.
  - Strittmatter, E. F.; Ferguson, P. L.; Tang, K.; Smith, R. D. Proteome Analyses Using Accurate Mass and Elution Time Peptide Tags with Capillary LC Time-of-Flight Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 980–991.
  - Nepomuceno, A. I.; Muddiman, D. C.; Bergen, H. R.; Craighead, J. R.; Burke, M. J.; Caskey, P. E.; Allan, J. A. Dual Electrospray Ionization Source for Confident Generation of Accurate Mass Tags Using Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 3411–3418.
  - xSakurai, T.; Matsuo, T.; Matsuda, T.; Katakuse, I. Paas 3: A Computer Program to Determine Probable Sequence of Peptides from Mass Spectrometric Data. *Biomed. Mass Spectrom.* **1984**, *11*, 396–399.
  - Ishikawa, K.; Niva, Y. Computer-Aided Peptide Sequencing by Fast Atom Bombardment Mass Spectrometry. *Biomed. Environ. Mass Spectrom.* **1986**, *13*, 373–380.
  - Siegel, M. M.; Bauman, N. An Efficient Algorithm for Sequencing Peptides Using Fast Atom Bombardment Mass Spectral Data. *Biomed. Environ. Mass Spectrom.* **1988**, *15*, 333–343.
  - Johnson, R. J.; Biemann, K. Computer Program (seqpep) to Aid in the Interpretation of High-Energy Collision Tandem Mass Spectra of Peptides. *Biomed. Environ. Mass Spectrom.* **1989**, *18*, 945–957.
  - Zidarov, D.; Thibault, P.; Evans, M. J.; Bertrand, M. J. Determination of the Primary Structure of Peptides Using Fast Atom Bombardment Mass Spectrometry. *Biomed. Environ. Mass Spectrom.* **1990**, *19*, 13–16.
  - Yates, J. R.; Griffin, P. R.; Hood, L. E.; Zhou, J. X. Computer Aided Interpretation of Low Energy MS/MS Mass Spectra of Peptides. In *Techniques in Protein Chemistry II*; Villafranca, J. J., Ed.; Academic Press: San Diego, CA, 1991; pp 477–485.
  - Mann, M.; Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. *Anal. Chem.* **1994**, *66*, 4390–4399.
  - Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
  - Flad, T.; Spengler, B.; Kalbacher, H.; Brossart, P.; Baier, D.; Kaufmann, R.; Bold, P.; Metzger, S.; Blüggel, M.; Meyer, H. E.; Kurz, B.; Müller, C. A. Direct Identification of Major Histocompatibility Complex Class I-bound Tumor-Associated Peptide Antigens of a Renal Carcinoma Cell Line by a Novel Mass Spectrometric Method. *Cancer Res.* **1998**, *58*, 5803–5811.
  - Mann M. Useful Tables of Possible and Probable Peptide Masses. Proceedings of the 43rd ASMS Conference on Mass Spectrometry and Allied Topics, Atlanta, GA, May, 1995, p 639.
  - Wehofsky, M.; Hoffmann, R.; Hubert, M.; Spengler, B. Isotopic Deconvolution of Matrix-Assisted Laser Desorption/Ionization Mass Spectra for Substance-Class Specific Analysis of Complex Samples. *Eur. J. Mass Spectrom.* **2001**, *7*(39), 46.
  - Spengler, B.; Lutzenkirchen, F.; Kaufmann, R. On-Target Deuteration for Peptide Sequencing by Laser Mass Spectrometry. *Org. Mass Spectrom.* **1993**, *28*, 1482–1490.
  - Perera, I. K.; Candy, J. M.; Hakansson, P.; Oakley, A. E.; Brinkmalm, G.; Sundqvist, B. U. R. UV-Laser Induced Desorption Mass Spectrometry of Insulin, Substance P, and A4 Amyloid Protein Fragments from Synthetic Fibrillary Aggregates. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 527–532.
  - Jai-nhuknan, J.; Cassidy, C. J. Anion and Cation Post-Source Decay Time-of-Flight Mass Spectrometry of Small Peptides: Substance P, Angiotensin II, and Renin Substrate. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1678–1682.
  - Barbacci, D. C.; Russell, D. H. Sequence and Side-Chain Specific Photofragment (193 nm) Ions from Protonated Substance P by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1038–1040.
  - Anderson, W. R.; Frick, W.; Daves, G. D.; Barofsky, D. F.; Yamaguchi, I.; Chang, D.; Folkers, K.; Rosell, S. Mass Spectra of Underivatized Peptide Amides Related to Substance P. *Biochem. Biophys. Res. Commun.* **1977**, *78*, 372–376.