Statistical Evaluation of Electrospray Tandem Mass Spectra for Optimized Peptide Fragmentation

Jason C. Rogalski, Michael S. Lin, Matthew J. Sniatynski, and Robert J. Taylor
The Biomedical Research Center, University of British Columbia, Vancouver, British Columbia, Canada

Nikolay Youhnovski and Michael Przybylski
Department of Chemistry, University of Konstanz, Konstanz, Germany

Juergen Kast*
Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) has become the method of choice for the analysis of complex peptide mixtures. It combines the separation power of nanoflow LC with highly specific sequence analysis, allowing automated peptide sequencing with high resolution and throughput. For peptide fragmentation, the current experimental setup uses predefined parameters based on the mass-to-charge ratio of the individual precursor. Suitable parameters are typically established by empirical evaluation of fragment spectra of individual peptides used as standards. As a result, nonoptimal fragment spectra are obtained if peptides show fragmentation behavior different from these standards, which often result in the loss of sequence-specific fragment ion information. Here we describe a statistical approach for the systematic evaluation of the quality of individual peptide fragment spectra based on the calculation of their arithmetic mean and standard deviation. The method utilizes the dependence of these parameters on the difference in electric potential across the collision cell to determine the value that results in maximum information content. We show that the method is applicable to fragment spectra generated from a variety of multiply-charged tryptic peptides, over a wide concentration range, and on different types of mass analyzers. We also show how this novel approach can be used to define optimized collision energy settings over a wide mass-to-charge range.   (J Am Soc Mass Spectrom 2005, 16, 505–514) © 2005 American Society for Mass Spectrometry

Matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) based mass spectrometry has developed into an indispensable tool for researchers in molecular biology and biotechnology, and a key technology in proteomics [1]. Two different approaches incorporating mass spectrometry have emerged for the identification of proteins after enzymatic digestion into peptides. Peptide mass fingerprinting determines the masses of all peptides present in the digestion mixture and compares the resulting patterns to those predicted for all entries in protein or genome sequence databases. This approach usually employs MALDI-MS because of its inherent simplicity, speed, and reliability, but is limited to digests of isolated proteins or simple protein mixtures [2].

For more complex digestion mixtures, peptide sequencing using tandem mass spectrometry combined with liquid chromatography (LC-MS/MS) is applied [3]. In this approach, individual peptides are separated, isolated from the mixture, and fragmented using low or high-energy collisions. In order to identify the best-matching sequence, the fragmentation patterns obtained for a specific peptide are compared to those anticipated for each hypothetical peptide predicted from all database entries. Although recently introduced MALDI-MS/MS techniques now also permit selection and sequencing of individual peptides [4, 5], the vast majority of peptide sequencing experiments are still carried out by ESI-MS/MS. Particularly, the generation of intense fragment ions from multiply-charged peptide ions and the possibility of on-line separation by multi-dimensional chromatography make ESI-MS/MS valuable for peptide sequencing [6]. Consequently, several high-throughput techniques such

as isotope-coded affinity tagging (ICAT) [7] and multidimensional protein identification technology (MudPit) [8] rely on this approach to sequence thousands of peptides in one experiment.

The successful identification of proteins by peptide sequencing is dependent on three conditions: (1) the existence, correctness, and uniqueness of the matching database entry, (2) the identification and selection of the precursor peptide ions for fragmentation, and (3) the generation of sufficient tandem mass spectrometry information. Only if all of them are met are proteins identified with high confidence. Consequently, several approaches have been developed to identify proteins that are not present in databases [9–14] and to increase the sensitivity for detecting precursor ions [15–17]; however, the quality of fragment spectra has rarely been addressed. Although methods to exclude fragment spectra of insufficient quality prior to database searching [18] and to evaluate the resulting scores [19–23] have been reported, approaches that allow systematic optimization of acquisition parameters such as the electric potential that induces peptide fragmentation (collision energy) are lacking. Instead, they are usually established empirically by fragmenting standard peptides at various collision energies, selecting the fragmentation spectrum that appears to contain to most sequence information, and transferring the corresponding collision energy and mass-to-charge (*m/z*) ratio of the standard peptide into a reference table. In an LC-MS/MS experiment, instrument software then determines the *m/z* value of an eluting peptide and sets the collision energy for fragmentation according to this table. This can result in nonoptimal fragment spectra if a peptide shows fragmentation behavior different from that of the standard peptides, which would cause loss of sequence-specific fragment ions and compromised sequence information. In more recent approaches, a range of electric potentials is therefore applied by stepping or rolling the collision energy around the predefined value, which increases the likelihood of randomly generating sufficient peptide fragment ions [13, 24]. Due to the lack of statistical data, the actual size of the range is still determined empirically.

Here we describe an alternative approach based on the relationship between the collision energy and the information content of peptide fragment spectra. We make use of the arithmetic mean and the standard deviation as two parameters that can be used to determine the median and spread of data distributions. By systematically assessing the dependence of these parameters on the potential difference between the quadrupole ion guide (Q0) and the collision cell (Q2), we show that they are sufficient to describe the quality of fragmentation spectra. These calculations are not significantly affected by the peptide type, its sequence, its charge state, or the generation of abundant low-mass fragment ions. Furthermore, we show that our method is compatible with a range of peptide concentrations and instrument types. The method readily lends itself to the optimization of collision energy settings over a wide mass-to-charge range.

Experimental

Chemicals

Bradykinin fragment 1-9 (RPPGFSPFR, average MW = 1060.22 Da), Bradykinin fragment 2-9 (PPGFSPFR, average MW = 904.03 Da), bovine serum albumin (BSA), α-casein, fetuin (all bovine), and iodoacetamide were purchased from Sigma-Aldrich (St. Louis, MO), sequencing grade trypsin was obtained from Promega (Madison, WI), and dithiothreitol was from Amersham (Piscataway, NJ). All solvents used were of HPLC grade. For tryptic digests, proteins were denatured at 90 °C for 30 min, reduced with dithiothreitol, alkylated with iodoacetamide, and digested in solution using a enzyme/substrate ratio of 1:50 (wt/wt) overnight as described [25]. Protein digests and individual peptides were dissolved in methanol/water/formic acid (50:45:5) to final concentrations of 2 pmol/μl (digests) and 20 fmol/μl, 200 fmol/μl, and 2 pmol/μl (peptides), prior to mass spectrometric analysis.

Mass Spectrometry

For data acquisition, 1–5 μl of the various peptide solutions were transferred into individual nanoelectrospray needles (Protana, Odense, Demark). Nanoelectrospray mass spectometry experiments of individual peptides were performed on a hybrid quadrupole-time of flight instrument (Q-Star Pulsar i, Applied Biosystems, Framingham, MA) supplemented by a commercial nanoelectrospray interface (Protana). In addition, a triple quadrupole instrument (API III+, Sciex, Concord, ON, Canada) equipped with a custom-built nanoelectrospray interface, and a Fourier Transform-Ion Cyclotron Resonance (FT-ICR) instrument (APEX II, Bruker Daltonik, Bremen, Germany) with an actively shielded 7.0 tesla superconducting magnet (Magnex, Oxford, UK) and built-in nanoelectrospray interface (APOLLO, Bruker Daltonik) were used. A second hybrid quadrupole - time of flight instrument (Q-Star Pulsar XL, Applied Biosystems) was employed for the analysis of protein digests and the calibration plot. Tandem mass spectra acquired on quadrupole-time of flight instruments were averaged for at least one minute, whereas at least five scans were averaged on the triple quadrupole instrument, depending on the peptide concentration. FT-ICR tandem mass spectra were obtained by collecting 16 single scans using experimental conditions as described [26]. For the calibration plot, potential differences were stepped by 2 V with an acquisition time of 15 s per increment for each of the 22 precursor ions. Acquisition of all spectra was performed with the respective instrument software.

Data Processing

All tandem mass spectra were obtained in duplicate (calibration plot) or triplicate (all other experiments)

and converted into ASCII format, processed with in-house software, and imported into Excel (Microsoft, Mississauga/ON, Canada) for data display. The in-house software was written in C++ and used the following equations:

$$\text{Arithmetic Mean: } \mu = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i \cdot y_i) \qquad (1)$$

$$\text{Standard Deviation: } \sigma = \left(\frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \mu)^2 \right)^{1/2} \qquad (2)$$

with: x = individual mass-to-charge ratio, y = intensity at x, N = total number of mass-to-charge ratios, to mathematically describe the weighted average (arithmetic mean μ) and the spread (standard deviation σ) of a data distribution.

For determination of fragment ion coverage, spectra were centroided in Analyst QS (Applied Biosystems, Foster City, CA) using 600 ppm merge distance, 99% height, and 10 and 500 ppm mimimum and maximum peak widths, respectively, and converted into ASCII format. Theoretical masses of all b- and y-type as well as immonium and internal fragment ions were compared to each of these spectra and counted if they fell within 0.1 Da mass error. Fragment ion types were assigned according to the nomenclature of Roepstorff and Fohlman [27] modified by Biemann [28].

For noise filtering of FT-ICR data, a discrete wavelet denoising operation [29] was performed using a 10-point Daubechies wavelet and a hard threshold function that zeroed the lowest 99.9% of points in the wavelet transformed data. This removed much of the noise, leaving signal peaks discernibly above the remaining baseline. The remaining baseline was removed by subtracting 1.5 times the average of local maxima in a 10,001-point window around each data point.

Results and Discussion

Fragment Ion Coverage

In peptide sequencing, the most important criterion for successful protein identification is the generation of sequence-specific fragment ions, which directly relates to the score achieved in a database search [30]. The total number of matching fragment ions is dependent on several experimental parameters, including the individual peptide sequence and the collision energy settings. The collision energy, which reflects the kinetic energies of precursor ions in the collision cell, can be determined by multiplying the charge state of the precursor ion with the potential difference between the quadrupole ion guide (Q0) and the collision cell (Q2). We wished to gain insight into how various potential differences affect the total number of matching fragment ions for peptides with similar masses, but different sequences.

For three doubly-charged peptide ions generated from tryptic digests of α-casein and bovine serum

albumin, we determined the number of matching fragment ions for potential differences increased from 15 to 50 V in 5 V increments (Figure 1). The fragment ions were grouped into those produced by fragmentation of a single peptide bond (b- and y-ions, positive scale) or two peptide bonds (immonium and internal fragment ions, negative scale), respectively [31]. Fragment ions generated by additional side chain fragmentation, such as y-NH$_3$, b-NH$_3$, b-H$_2$O, or by both one- and two-step fragmentation pathways, i.e., a-type ions, could not be assigned to either group and were thus not considered. The resulting plots show similar trends with increasing collision energy for all peptides, namely an initial increase, a maximum, and a subsequent decrease in the number of matching fragments. Closer inspection indicates that the potential difference at which the maximum number of matching fragment ions is obtained, varies from peptide to peptide. For α-casein 38-49 (FFVAPFPEVFGK) observed at *m/z* 692.87, a maximum of 19 matching b- and y-ion fragments is reached at 25 and 30 V (positive scale), while the maximum number of 37 matching immonium and internal fragment ions are observed at 40 V (negative scale). Similar results are found for BSA 569-580 (TVMENFVAFVDK) detected at *m/z* 700.35 with a maximum of 19 matching b- and y-type ions at 25 V and 29 matching immonium and internal fragment ions at 40 V. By contrast, for BSA 89-100 (SLHTLFGDELCK) identified at *m/z* 710.35, a maximum of 19 matching b-and y-type and 34 matching immonium and internal fragment ions are observed at 40 and 50 V, respectively. Fragmenting this peptide at the potential differences that generated the maximum
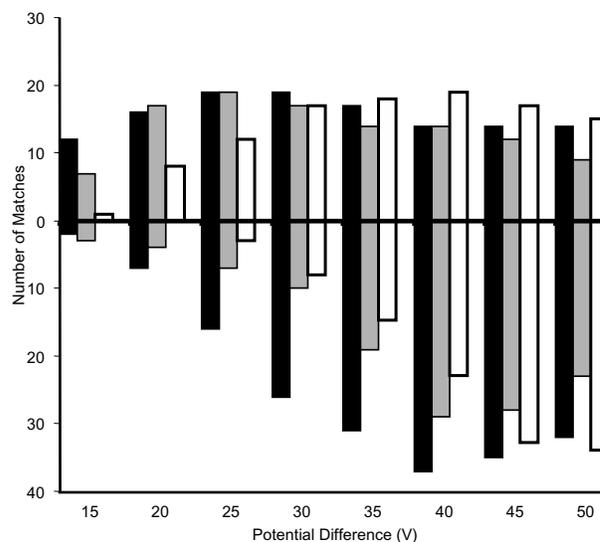

Figure 1. Dependence of the number of matching fragment ions on the potential difference used for peptide fragmentation for the doubly-charged tryptic peptides α-casein 38-49 at *m/z* 692.87, (black), BSA 569-580 at *m/z* 700.35 (grey), and BSA 89-100 at *m/z* 710.35 (white) on a hybrid quadrupole-time of flight instrument. The numbers of matching b- and y-ions are depicted on the positive scale, matching immonium and internal fragment ions on the negative scale.

number of matching fragment ions for the aforementioned two peptides, i.e., 25 and 40 V, results in 7 fewer b- and y-type and 11 fewer immonium and internal fragment ions. Consequently, applying conditions that are based on the fragmentation of peptides with similar *m/z* ratio corresponds to a significant loss of peptide-specific fragment ions. Moreover, for all three peptides, breaking two peptide bonds requires 10–15 V higher potential differences than breaking single peptide bonds. Although this is easily explained by the increase in intrinsic energy required for this process, the practical consequences for peptide sequencing are still noteworthy.

Database search programs aim to correlate all experimentally observed fragment ions with those that are generated in silico for each sequence based on predefined fragment ion types [30], therefore the ideal potential difference for each peptide should coincide with the highest total number of matching fragment ions. We found the number of matching immonium and internal fragment ions to be significantly higher than the b- and y-type fragment ions in all cases, therefore the sum of both is biased towards fragment ions that contain less sequence information. On the other hand, considering only b- and y-type fragment ions would discard valuable information that can be used to validate the peptide sequence. Our data indicate that maximum sequence confidence is achieved at potential differences slightly higher than indicated by the b- and y-type fragment ion maximum, because a sharp increase in the number of matching immonium and internal fragment ions compensates the slight decrease in the number of matching b- and y-type fragment ions. In our study, using three peptides with almost identical molecular weight and charge state, we found that the potential difference required to achieve optimal fragmentation can differ by more than 10 V. We concluded that the mass-to-charge ratio alone is not a good indicator to determine the optimal potential difference, and taking into account the individual fragment spectra might be more suited for this purpose.

Evaluation of Individual Fragment Ion Distributions

In order to investigate what the characteristics of the individual spectra are and how they change with increasing potential difference, a series of tandem mass spectra using a model peptide mimicking the sequence of a tryptic peptide (Bradykinin fragment 2-9, sequence PPGFSPFR) were generated. Its doubly-charged molecular ion at *m/z* 452.74 was selected and fragmented on a hybrid quadrupole-time of flight instrument altering the potential difference from 10 to 50 V in 5 V increments.

At 20 V potential difference (Figure 2a), mostly the precursor ion at *m/z* 452.74 is visible. Under these conditions, the initial threshold of fragmentation is
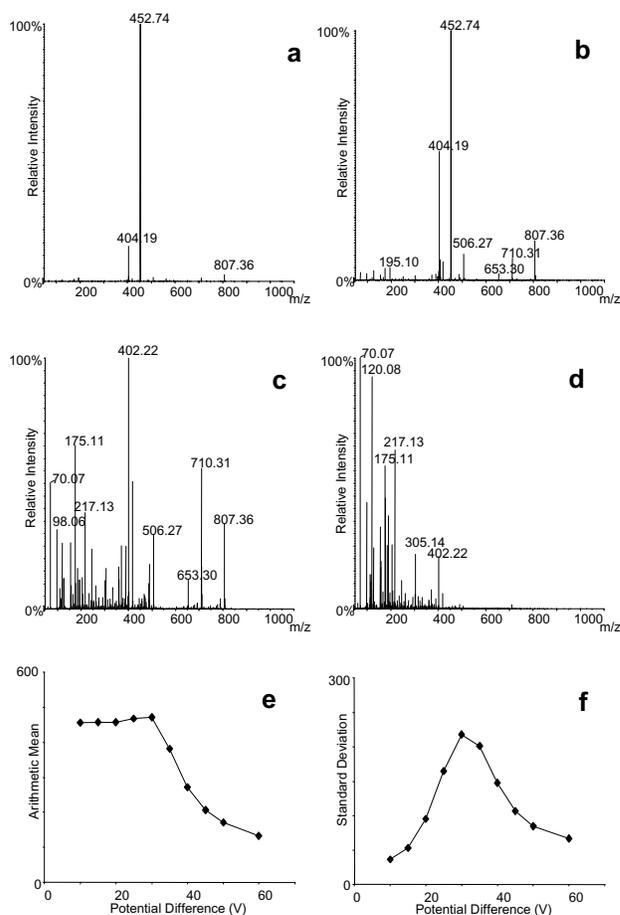

Figure 2. Representative fragment spectra of doubly-charged Bradykinin 2-9 (*m/z* 452.74) obtained on a hybrid quadrupole-time of flight instrument at (**a**) 20 V, (**b**) 25 V, (**c**) 35 V, and (**d**) 45 V potential difference. Fragment spectra were acquired in triplicate, increasing the potential difference in 5 V increments, arithmetic mean (**e**) and standard deviation (**f**) were calculated according to eqs 1 and 2 and plotted against the potential difference.

barely overcome, leaving the precursor ion largely intact while breaking only the most labile bonds. Due to the nature of the tryptic peptide, the basic carboxy-terminal amino acid residue will retain one proton creating a fragment ion of the y-series while additional protons are mobile and distributed over the entire peptide [32]. A neutral amino-terminal fragment will only be created if the bond breakage occurs close to the amino terminus of the peptide, otherwise two singly-charged fragments with complementary masses will be obtained. A doubly-charged fragment ion at *m/z* 404.19 (y7^{++}) with significant intensity is due to the loss of the amino-terminal proline residue, which proves the proline-proline bond to be the most labile [31]. A number of singly-charged fragment ions of the y series are present in the high-mass part of the spectrum but show very low intensity. Raising the potential difference to 25 V (Figure 2b) significantly increases the intensities of the previously observed fragment ions at *m/z* 506.27 (y4), 653.30 (y5), 710.31 (y6), and 807.36 (y7),

and yields additional ions in the low-mass end of the spectrum. These fragment ions are generated with similar probabilities and have similar intensities because the fragmenting bonds have comparable stabilities. An intense signal is still observed for the nonfragmented precursor ion, indicating that the kinetic energy, which will follow a Maxwell-Boltzmann distribution, is insufficient to induce fragmentation for a significant portion of the precursor ion.

By contrast, at 35 V (Figure 2c), the precursor ion is completely absent and the singly-charged fragment ion y3 with an additional loss of ammonia (y3-NH$_3$) at m/z 402.22 becomes the base peak of the spectrum. In addition to the fragment ions observed at 25 V, there is a large number of low-mass ions. These are immonium ions such as those of proline (m/z 70.07) and phenylalanine (m/z 120.08), terminal fragment ions of the a-, b-, and y-series such as those at m/z 175.11 (y1) and 98.06 (b1), and internal fragment ions such as those of PG (m/z 155.08) and PF with an additional loss of carbon monoxide CO (m/z 217.13). These ions are formed by secondary fragmentation of the large peptide fragments. Therefore, a concurrent decrease of large fragment ions can be noticed for the increase in intensity of the low-mass fragment ions. At 45 V (Figure 2d), the low-mass fragment ions dominate the spectrum while high-mass fragment ions are virtually absent, indicating secondary fragmentation is commonplace.

In order to determine whether the varying distribution of fragment ions coincides with a change in the total number of sequence-specific fragment ions, we created a peak list for each of these fragment spectra and compared them to the list of all possible b-, y-, internal, and immonium ions we predicted for Bradykinin fragment 2-9, using the previously described criteria. At 20 V, only one out of a maximum of 35 possible fragment ions is found, whereas 11 possible fragment ions are observed at 25 V. The number of matching fragment ions reaches a maximum of 26 at 35 V potential difference before declining to 17 at 45 V. From these experiments, we concluded that conditions that result in the spectra shown in Figure 2a , b, and d likely do not generate enough sequence-specific information for unambiguous identification. By contrast, conditions that lead to the spectrum in Figure 2c produced the highest number of sequence-specific peptide fragments and therefore correspond to the highest likelihood of unambiguous identification.

Arithmetic Mean and Standard Deviation

We found a wide distribution of fragment ions to be a key feature of high-scoring fragment spectra. Therefore, we tested whether the arithmetic mean μ, a mathematical variable that describes the weighted average of a data distribution (eq 1), and the standard deviation σ, which is used to quantify the spread of such data distributions (eq 2), could be applied for the evaluation of fragment spectra. For the series of spectra obtained in the previously described experiment, the arithmetic mean and standard deviation were calculated and plotted against the individual potential difference settings. The arithmetic mean (Figure 2e) remains almost constant for potential differences from 15 to 30 V with a value similar to the mass-to-charge ratio of the precursor ion, followed by a sharp drop from 35 to 50 V. This indicates that the centre of the fragment ion distribution becomes progressively lower than the precursor m/z due to the increasing presence of low-mass fragments.

By contrast, the standard deviation (Figure 2f) is low at 10 V due to only a few fragment ions being present. It continuously increases from 15 to 30 V where it reaches a maximum, indicative of the widest spread of information in the spectrum, then drops off at 35 to 50 V. From these results, we concluded that the fragment spectrum possessing the most even distribution of fragment ions, depicted in Figure 2c, will show an arithmetic mean similar to the precursor m/z, and a standard deviation that is maximal. Therefore, calculating arithmetic mean and standard deviation is sufficient to distinguish potential high- and low-scoring fragment spectra.

Evaluation of Standard Deviation Maxima

The previous data indicated that the optimal potential difference is between 25 and 35 V. Therefore, we acquired a second set of fragment spectra using the same conditions but altering the potential difference from 25 to 35 V in 1 V increments. By comparing the spectra at 25 V (Figure 2b) and 35 V (Figure 2c), it is noticeable that the major difference between these spectra is in the relative intensity of the various fragment ions rather than their mere presence. This trend is confirmed in representative fragment spectra shown for increasing potential difference.

At 27 V (Figure 3a), the base peak of the spectrum is the doubly-charged fragment ion at m/z 404.19 (y7^{++}) with the precursor ion at m/z 452.74 having comparable intensity. While high-mass fragment ions at m/z 506.27 (y4), 653.30 (y5), 710.31 (y6), and 807.36 (y7) show significant intensities, low-mass fragment ions at m/z 70.07 (immonium ion, proline), m/z 155.08 (internal fragment PG), and m/z 175.11 (y1) show lower intensities. When the potential difference is raised to 29 V (Figure 3b), the intensity of the doubly-charged fragment ion at m/z 404.19 (y7^{++}) decreases while high- and low-mass fragment ions gain in intensity, with the one at 710.31 (y6) becoming the base peak. At 31 V (Figure 3c), the low-mass fragment ion at m/z 70.07 (immonium ion of proline) becomes the base peak, with intense low- and high-mass fragment ions and the almost complete disappearance of the precursor ion. Finally, when the potential difference is increased to 33 V (Figure 3d), the very low-mass fragment ions gain in intensity while all other fragment ions decrease. The arithmetic mean and the standard deviation were again calculated and plot-
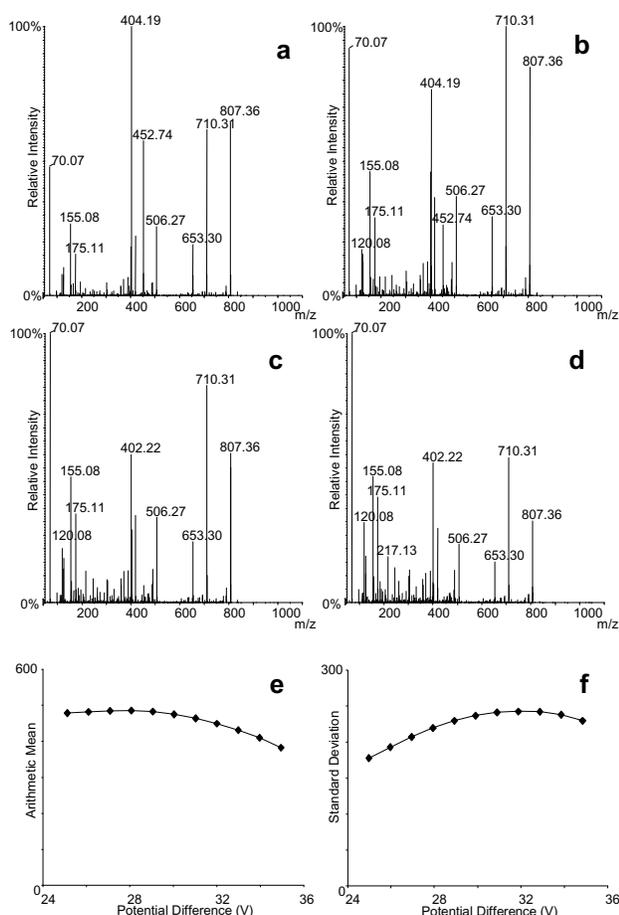
investigated whether the standard deviation approach can also be applied to evaluate fragment spectra of more complex precursor ions that regularly occur in tryptic digests. Specifically, we used another model peptide mimicking the sequence of a tryptic peptide containing a second basic residue due to a missed cleavage site (Bradykinin fragment 1-9, sequence RPPGFSPFR). For this peptide both the doubly and triply charged precursor ions were observed at m/z 530.79 and 354.19, respectively, as commonly found in tryptic digests. Each of them was fragmented individually on a hybrid quadrupole-time of flight instrument using a potential difference range from 10 to 50 V in 5 V increments. After calculating arithmetic mean and standard deviation for each of these fragmentation spectra, the resulting curves were plotted against those obtained previously for Bradykinin fragment 2-9.

The change in arithmetic mean with increasing potential difference is remarkably similar for all three precursor ions (Figure 4a). All of them show almost constant values until they reach a threshold followed by a sharp decrease. The onset of the decrease is noticeable at 35 V for doubly-charged Bradykinin fragment 2-9 and at 30 V for doubly-charged Bradykinin fragment 1-9, while that for triply-charged Bradykinin fragment 1-9 is at 25 V. Again, a second set of experiments was performed to explore these trends in more detail (Figure 4b). For doubly-charged Bradykinin fragment 2-9, a maximum of the arithmetic mean is observed at 28 V. This does not occur in the other two plots. Instead, a plateau followed by a decrease is observed for both
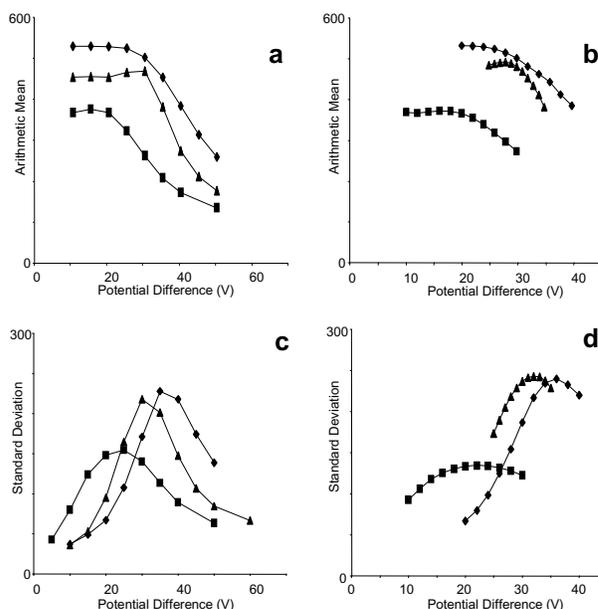
Figure 3. Representative fragment spectra of doubly-charged Bradykinin 2-9 (m/z 452.74) obtained on a hybrid quadrupole-time of flight instrument at (**a**) 27 V, (**b**) 29 V, (**c**) 31 V, and (**d**) 33 V potential difference. Fragment spectra were acquired in triplicate, increasing the potential difference in 1 V increments, arithmetic mean (**e**) and standard deviation (**f**) were calculated according to eqs 1 and 2 and plotted against the potential difference.

ted for this set of fragmentation spectra. The arithmetic mean (Figure 3e) remains almost constant between 25 and 28 V, but decreases at higher potential differences with an increasing slope. By contrast, the standard deviation (Figure 3f) increases and reaches a maximum at 33 V before it starts to decrease.

Although the changes in the 25 to 35 V range seem rather subtle, comparison of the individual spectra indicates that they differ significantly in the intensities of the individual fragment ions. It is noteworthy that the shift in intensities from high-mass to low-mass fragment ions is more obvious by the drop in the arithmetic mean than by the minimal changes in the standard deviation. However, the maximum of the standard deviation at 33 V does coincide with the spectrum that is the most diverse in fragment ions.

Charge State and Peptide Sequence

Since it is well known that different types of precursor ions generate very different fragmentation patterns, we

Figure 4. Dependence of (**a**, **b**) arithmetic mean and (**c**, **d**) standard deviation on the potential difference varied in (**a**, **c**) 5 V increments and (**b**, **d**) 2 V or 3 V increments for: doubly-charged Bradykinin 2-9 (triangles), doubly-charged Bradykinin 1-9 (diamonds), and triply-charged Bradykinin 1-9 (squares) using a hybrid quadrupole-time of flight instrument.

starting at 24 V for doubly-charged Bradykinin fragment 1-9 and 18 V for triply-charged Bradykinin fragment 1-9.

By contrast, all three peptides reveal similar trends in the standard deviation plots, including a maximum value for a specific potential difference setting (Figure 4c). This maximum is observed at 30 V for doubly-charged Bradykinin fragment 2-9, at 35 V for doubly-charged Bradykinin fragment 1-9, and at 25 V for triply-charged Bradykinin fragment 1-9 when stepping from 10 to 50 V in 5 V increments. Closer inspection of the individual ranges using smaller increments confirms these findings (Figure 4d). For doubly-charged Bradykinin fragment 2-9, a potential difference of 32 V results in maximal standard deviation, whereas that of doubly- and triply-charged Bradykinin fragment 1-9 is determined at 36 and 22 V, respectively. The similarity of these plots demonstrates that this method is also applicable for higher charge states and tryptic peptides containing missed cleavage sites.

Low-Mass Region

Fragment ions of the a-, b-, and y-ion series as well as internal fragment ions are distributed over the entire mass range, whereas immonium ions representing single amino acid residues can only be found in the low mass region. Due to this disproportionate representation in the spectrum, immonium ions could shift or alter the values of arithmetic mean and standard deviation. Therefore, we eliminated increasing portions of the low mass range in the fragment spectra of doubly-charged Bradykinin fragment 2-9 for the various potential differences. The recalculated arithmetic mean and standard deviation were compared to those obtained for the unchanged spectra with a low-mass cut-off at m/z 50.

The arithmetic mean of the original spectrum shows constant values until 30 V followed by a sharp drop at 35 V and higher (Figure 5a). Cutting the low-mass region up to m/z 100 results in a slight increase of the arithmetic mean to a maximum at 30 V, while the drop at 35 V and higher is unchanged. Similar results are obtained for cut-offs at m/z 150 and 200, respectively, although the changes are more pronounced. Performing the same processing on the narrow potential difference range data confirms the appearance of a maximum at 30 V instead of an immediate decrease at higher mass cut-offs (Figure 5b). The low-mass cut-off has an even more subtle effect on the standard deviation at various potential differences. For the data obtained in 5 V increments, no difference in the position of the maximum at 30 V and the trend of the plots is visible (Figure 5c). The only noticeable difference is the decrease in the value of the standard deviation with increasing cut-off masses. Closer inspection using the more narrow increment data confirms a gradual flattening of the curves with increasing cut-off (Figure 5d).

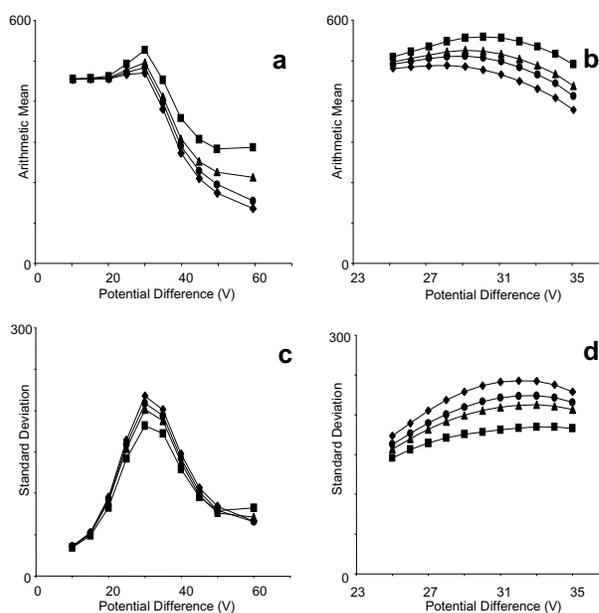Similar results were obtained for triply-charged Bradykinin fragment 1-9 (data not shown). The only difference was a shift in the maximum of the standard deviation for the m/z 200 cut-off compared to the others. This can be explained by the presence of additional doubly-charged fragment ions in this region. Our data therefore shows that the low-mass region has only minor influence on the arithmetic mean and standard deviation of both doubly- and triply-charged precursor ions, demonstrating the general applicability of this method.

Figure 5. Dependence of (**a**, **b**) arithmetic mean and (**c**, **d**) standard deviation on the potential difference varied in (**a**, **c**) 5 V increments and (**b**, **d**) 1 V increments for doubly-charged Bradykinin 2-9 peptide using a low-mass cut-off of m/z 50 (diamonds), m/z 100 (circles), m/z 150 (triangles), and m/z 200 (squares) using a hybrid quadrupole-time of flight instrument.

Peptide Concentration

All data used for the various calculations described above were obtained using peptide concentrations of 2 μM. In order to determine whether this approach is also applicable at lower peptide concentrations, we acquired fragment spectra of doubly-charged Bradykinin fragment 2-9 at 200 and 20 nM concentrations at various potential differences. Previous work had shown that the switching criteria are a limiting factor for the sensitivity of LC-MS/MS experiments, giving nanoelectrospray experiments at least comparable sensitivity and making them a valid substitute for sensitivity studies [15–17]. For each spectrum, we calculated arithmetic mean and standard deviation, and compared the resulting plots to those obtained for the same peptide at 2 μM concentration.

Using 5 V increments (Figure 6a), the arithmetic mean remains nearly constant at low, but shows a drop at higher potential differences. For 2 μM, this drop starts at 30 V, whereas at 200 nM it shifts to 25 V. At
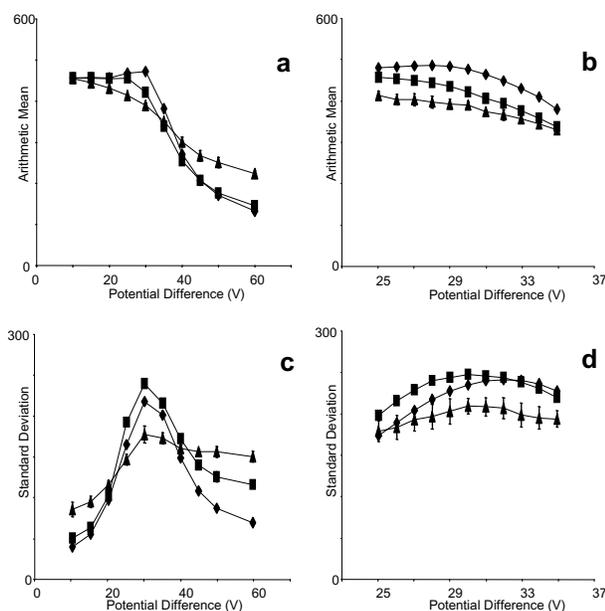
Figure 6. Dependence of (**a**, **b**) arithmetic mean and (**c**, **d**) standard deviation on the potential difference varied in (**a**, **c**) 5 V increments and (**b**, **d**) 1 V increments for doubly-charged Bradykinin 2-9 peptide using concentrations of 2 μM (diamonds), 200 nM (squares), and 20 nM (triangles) using a hybrid quadrupole-time of flight instrument.

20 nM, a minor drop is already visible at 10 V, which becomes more noticeable at 25 V. Closer inspection of the region form 25 to 35 V confirms these findings (Figure 6b). At 2μM concentration, a maximum is reached at 28 V before the arithmetic mean starts to decrease, whereas at 200 nM there is not discernable maximum, but a decrease is obtained beginning at 27V. By contrast, at 20 nM a decrease is observed over the entire range, which becomes more pronounced between 25 and 30 V. Since the values of the arithmetic mean differ by only 15% in this region, it becomes difficult to determine a trend. Similar results are obtained for the plots of the standard deviation at varying potential differences. The clear maximum observed at 30 V using peptide concentrations of 2 μM is also observed at 200 and 20 nM, although the trend is less noticeable due to increased flattening of the curves (Figure 6c). Closer inspection of the 30 V region shows that the maximum is shifting towards lower potential differences with decreasing concentrations. At 2 μM, the maximum is observed at 32 V, whereas at 200 and 20 nM it shifts to 30 V. Again, it is difficult to locate the maximum in the plot of 20 nM because the individual values change by only 15%.

By comparing the individual spectra, we concluded that the flattening of the curves is due to a drastic change of the signal-to-noise ratio caused by the decrease in peptide signal and the concurrent increase in chemical noise. In fact, this is a well-known effect in nanoelectrospray and LC-MS experiments due to the generation of large numbers of droplets that do not contain any peptide molecules, resulting in increased

intensities of charged solvent clusters [33]. Application of established methods to increase signal-to-noise ratios may be used to further extend the sensitivity of this approach [16].

Instrument Type

Next, we wanted to determine whether our method is platform-independent and compatible with different types of mass analyzers. To this end, we acquired fragment spectra of doubly-charged Bradykinin fragment 2-9 and doubly-charged Bradykinin fragment 1-9, respectively, at 2 μM concentration and varying potential differences on a triple quadrupole and a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. On the triple quadrupole instrument, the arithmetic mean for doubly-charged Bradykinin fragment 2-9 slightly increased with increasing potential difference until a maximum was reached at 20 V, followed by a sharp decrease at higher values, whereas for doubly-charged Bradykinin fragment 1-9 the same trend with a maximum at 17 V occurred (Figure 7a). The standard deviation for doubly-charged Bradykinin fragment 2-9 initially increased with increasing potential differences, reached a maximum at 20V, then decreased again, which was also observed for doubly-charged Bradykinin fragment 1-9 with the maximum shifted to 25 V (Figure 7b).

By contrast, for the unprocessed fragment spectra of doubly-charged Bradykinin fragment 2-9 that were ac-
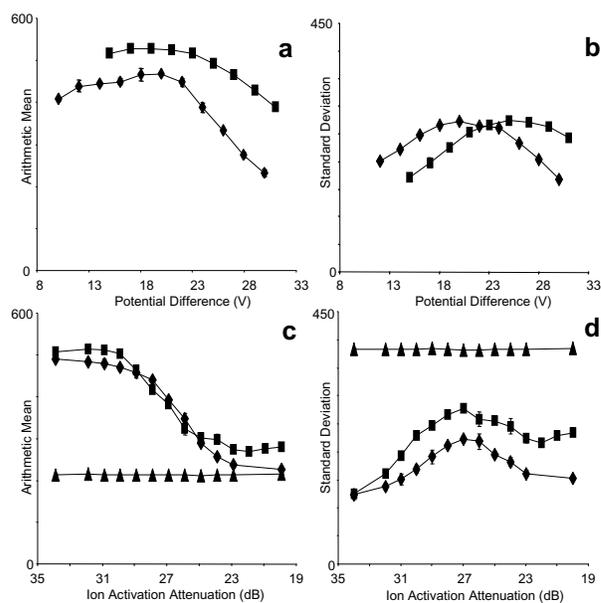

Figure 7. Dependence of (**a**, **c**) arithmetic mean and (**b**, **d**) standard deviation on (**a**, **b**) the potential difference of a triple quadrupole instrument for doubly-charged Bradykinin 2-9 peptide (diamonds) and doubly-charged Bradykinin 1-9 peptide (squares), and on (**c**, **d**) the ion activation attenuation of an FT-ICR instrument for doubly-charged Bradykinin 2-9 peptide when processing raw data (triangles) and data after noise filtering (diamonds) as well as for doubly-charged Bradykinin 1-9 peptide after noise filtering (squares).

quired on the FT-ICR instrument, both arithmetic mean (Figure 7c) and standard deviation (Figure 7d) remained virtually unchanged over the entire range. Closer inspection of the spectra showed that they consist largely of random noise as background to which a few fragment ion signals are superimposed. Consequently, we processed each spectrum to remove the noise using a wavelet denoising function [29] followed by a baseline correction. We then calculated arithmetic mean (Figure 7c) and standard deviation (Figure 7d) for the filtered fragment spectra and plotted them against decreasing ion activation attenuation on the x-axis, which corresponds to increasing kinetic energy of the precursor ions. The arithmetic mean showed a slight decrease with decreasing attenuation, which became much steeper at a value of 28 dB before flattening. By contrast, the standard deviation showed a maximum at a value of 27 dB while it was significantly lower at higher and lower attenuation. For doubly-charged Bradykinin fragment 1-9, which was chosen as a second example, similar trends and maxima with only minor changes in shape were observed, which may be indicative of the different experimental setup when performing precursor isolation, collision-induced fragmentation, and fragment spectra acquisition in an ICR cell. In general, the resulting plots for arithmetic mean and standard deviation obtained for both instruments were indeed comparable to those obtained on a hybrid quadrupole-time of flight instrument (Figures 2 , 3), proving that the method was not limited to a specific type of instrument.

Application for Calibration Plots

Finally, we decided to demonstrate the applicability of this method for the existing "rolling collision energy" approach [13, 24]. In this setup, a specific potential difference is used for each mass-to-charge ratio assuming a linear dependence. Slope and intercept of this equation are predefined and have to be determined using standard peptides. Using tryptic in-solution digests of BSA, α-casein, and fetuin at 2 μM concentration each, we selected 22 different doubly-charged precursor peptide ions for fragmentation and varied the potential difference in increments of 2 V. For each peptide, arithmetic mean and standard deviation were calculated and plotted against the potential difference. The maximal standard deviations were then plotted against the mass-to-charge ratios of the individual peptide precursor ions (Figure 8). The graph showed that each peptide had a unique potential difference that resulted in a maximal standard deviation, which could be approximated by a best-fit function using linear regression. Closer inspection revealed that most values did not fall on the line but differed by as much as 5 to 10 V from their neighbors. Additional statistical analysis revealed that 15 out of 22 peptides had maximum values within one standard deviation of the best-fit line, whereas an additional six peptides were located between one and two standard deviations. Similar results were obtained for
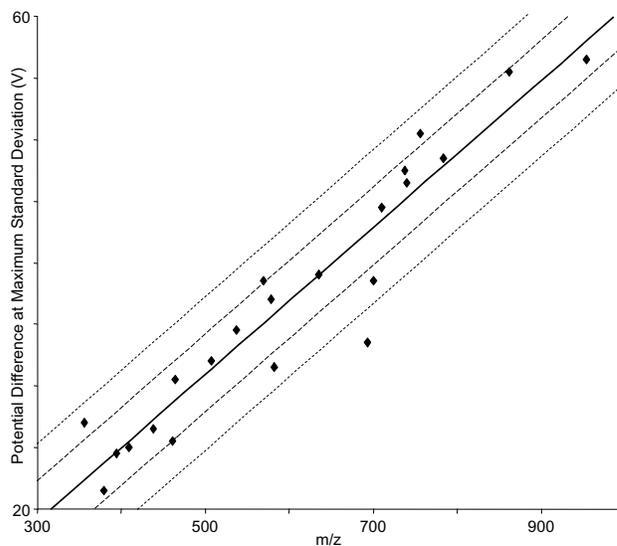

Figure 8. Dependence of the optimal potential difference determined by the standard deviation approach on the mass-to-charge ratio of 22 doubly-charged peptides present in tryptic digests of BSA, α-casein, and fetuin on a hybrid quadrupole-time of flight instrument. Linear regression resulted in a best-fit function (solid line) as well as areas defined by one (dashed line) and two (dotted line) standard deviation limits.

triply-charged peptide ions (data not shown), indicating that the method is also applicable to unknown peptides provided the charge state can be identified correctly. We concluded that the slope and intercept of the best-fit line determined by the standard deviation method defines the center in the rolling collision energy approach, whereas the maximum distance of a value from this line could be used to determine the spread of the collision energy. Using a distance of two standard deviations of the best-fit line, which corresponds to 6.2 V and contains 21 out of 22 peptides in this experiment, will guarantee maximum fragment ion coverage in at least one of the settings that are applied for the vast majority of peptides. Consequently, the statistical method presented here can be used immediately to optimize individual settings for maximal fragment ion coverage on a variety of instruments.

Conclusions

The results reported here clearly show that calculation of arithmetic mean and standard deviation lends itself to the evaluation of peptide fragment spectra for different types of peptides, instruments, and conditions. Therefore, this statistical method can be used off-line to optimize individual settings for maximal fragment ion coverage. On-line applications to optimize the collision energy for each peptide on an LC-MS/MS timescale would require significant changes in instrument control software and have not been pursued, but might well be feasible.

Acknowledgments

References

1. Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422,* 198–207.
2. Pappin, D. J. Peptide mass fingerprinting using MALDI-TOF mass spectrometry. *Methods Mol. Biol.* **2003**, *211,* 211–219.
3. McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R., III. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **1997,** *69,* 767–776.
4. Shevchenko, A.; Loboda, A.; Ens, W.; Standing, K. G. MALDI quadrupole time-of-flight mass spectrometry: A powerful tool for proteomic research. *Anal. Chem.* **2000,** *72,* 2132–2141.
5. Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **2000,** *72,* 552–558.
6. Cramer, R.; Corless, S. The nature of collision-induced dissociation processes of doubly protonated peptides: Comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics. *Rapid Commun. Mass Spectrom.* **2001,** *15,* 2058–2066.
7. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999,** *17,* 994–999.
8. Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001,** *19,* 242–247.
9. Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97,* 10313–10317.
10. Shevchenko, A.; Sunyaev, S.; Loboda, A.; Bork, P.; Ens, W.; Standing, K. G. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal. Chem.* **2001,** *73,* 1917–1926.
11. Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, S.; Zerial, M.; Wilm, M. Automated de novo sequencing of proteins using the differential scanning technique. *Proteomics* **2001,** *1,* 668–682.
12. Cagney, G.; Emili, A. De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.* **2002,** *20,* 163–170.
13. Guo, X.; Fell, L. M.; Bloomfield, N.; Lock, C. Optimization of peptide fragmentation patterns for de novo sequencing and protein identification by a rolling stepped collision energy. *Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics*; Orlando, FL, June, 2002.
14. Yergey, A. L.; Coorssen, J. R.; Backlund, P. S., Jr.; Blank, P. S.; Humphrey, G. A.; Zimmerberg, J.; Campbell, J. M.; Vestal, M. L. De novo sequencing of peptides using MALDI/TOF-TOF. *J. Am. Soc. Mass Spectrom.* **2002,** *13,* 784–791.
15. Wilm, M.; Neubauer, G.; Mann, M. Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **1996,** *68,* 527–533.
16. Kast, J.; Gentzel, M.; Wilm, M.; Richardson, K. Noise filtering techniques for electrospray quadrupole time of flight mass spectra. *J. Am. Soc. Mass Spectrom.* **2003,** *14,* 766–776.
17. Kast, J.; Parker, C. E.; van der Drift, K.; Dial, J. M.; Milgram, S. L.; Wilm, M.; Howell, M.; Borchers, C. H. Matrix-assisted laser desorption/ionization directed nano-electrospray ionization tandem mass spectrometric analysis for protein identification. *Rapid Commun. Mass Spectrom.* **2003,** *17,* 1825–1834.
18. Moore, R. E.; Young, M. K.; Lee, T. D. Method for screening peptide fragment ion mass spectra prior to database searching. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 422–426.
19. Bafna, V.; Edwards, N. SCOPE: A probabilistic model for scoring tandem mass spectra against a peptide database. *Bioinformatics* **2001,** *17 Suppl. 1,* S13–S21.
20. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002,** *74,* 5383–5392.
21. MacCoss, M. J.; Wu, C. C.; Yates, J. R., III. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **2002,** *74,* 5593–5599.
22. Moore, R. E.; Young, M. K.; Lee, T. D. Qscore: An algorithm for evaluating SEQUEST database search results. *J. Am. Soc. Mass Spectrom.* **2002,** *13,* 378–386.
23. Zhang, N.; Aebersold, R.; Schwikowski, B. ProbID: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2002,** *2,* 1406–1412.
24. Pace, N.; Gamble, T.; LeBlanc, J. C. Y.; Bloomfield, N. Implementation and benefits of "universal collision energy (CE)" on an hybrid quadrupole linear ion trap. *Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics*; Montreal, Quebec, Canada, June, 2003.
25. Park, Z. Y.; Russell, D. H. Thermal denaturation: A useful technique in peptide mass mapping. *Anal. Chem.* **2000,** *72,* 2667–2670.
26. Schmidt, P.; Youhnovski, N.; Daiber, A.; Balan, A.; Arsic, M.; Bachschmid, M.; Przybylski, M.; Ullrich, V. Specific nitration at tyrosine 430 revealed by high resolution mass spectrometry as basis for redox regulation of bovine prostacyclin synthase. *J. Biol. Chem.* **2003,** *278,* 12813–12819.
27. Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **1984,** *11,* 601.
28. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **1988,** *16,* 99–111.
29. Barclay, V. J.; Bonner, R. F.; Hamilton, I. P. Application of wavelet transforms to experimental spectra: Smoothing, denoising, and data set compression. *Anal. Chem.* **1997,** *69,* 78–90.
30. Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999,** *20,* 3551–3567.
31. Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1986,** *83,* 6233–6237.
32. Tang, X. J.; Thibault, P.; Boyd, R. K. Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal. Chem.* **1993,** *65,* 2824–2834.
33. Juraschek, R.; Dulcks, T.; Karas, M. Nanoelectrospray—more than just a minimized-flow electrospray ionization source. *J. Am. Soc. Mass Spectrom.* **1999,** *10,* 300–308.