
On the Benefits of Acquiring Peptide Fragment Ions at High Measured Mass Accuracy

Alexander Scherl,^a Scott A. Shaffer,^a Gregory K. Taylor,^a
Patricia Hernandez,^b Ron D. Appel,^b Pierre-Alain Binz,^{b,c} and
David R. Goodlett^a

^a Department of Medicinal Chemistry, University of Washington, Seattle, Washington, USA

^b Swiss Institute of Bioinformatics, University of Geneva, Geneva, Switzerland

^c GeneBio SA, Geneva, Switzerland

The advantages and disadvantages of acquiring tandem mass spectra by collision-induced dissociation (CID) of peptides in linear ion trap Fourier-transform hybrid instruments are described. These instruments offer the possibility to transfer fragment ions from the linear ion trap to the FT-based analyzer for analysis with both high resolution and high mass accuracy. In addition, performing CID during the transfer of ions from the linear ion trap (LTQ) to the FT analyzer is also possible in instruments containing an additional collision cell (i.e., the "C-trap" in the LTQ-Orbitrap), resulting in tandem mass spectra over the full m/z range and not limited by the ejection q value of the LTQ. Our results show that these scan modes have lower duty cycles than tandem mass spectra acquired in the LTQ with nominal mass resolution, and typically result in fewer peptide identifications during data-dependent analysis of complex samples. However, the higher measured mass accuracy and resolution provides more specificity and hence provides a lower false positive ratio for the same number of true positives during database search of peptide tandem mass spectra. In addition, the search for modified and unexpected peptides is greatly facilitated with this data acquisition mode. It is therefore concluded that acquisition of tandem mass spectral data with high measured mass accuracy and resolution is a competitive alternative to "classical" data acquisition strategies, especially in situations of complex searches from large databases, searches for modified peptides, or for peptides resulting from unspecific cleavages. (J Am Soc Mass Spectrom 2008, 19, 891–901) © 2008 American Society for Mass Spectrometry

Presently, tandem mass spectrometry combined with liquid chromatography (i.e., LC-MS/MS) is a standard procedure for high throughput proteomic analysis. Modern instruments are capable of identifying hundreds, if not thousands, of peptides per data-dependent LC-MS/MS analysis. Recently, hybrid instruments combining linear ion traps with Fourier-transform (FT) based mass analysis such as FT-ion cyclotron resonance (ICR) [1] or FT-Orbitrap [2] mass analyzers were introduced. These instruments greatly expand opportunities for experimental design because they combine two mass analyzers working in series. Also important to note is that even with external calibration, they deliver mass accuracies in the 2–5 ppm range on a routine basis. Furthermore, even better mass accuracies down to 1–2 ppm have been demonstrated by using the mass difference between adjacent peptide fragment ions in the FT-ICR analyzer [3] as well as injection of a normalized, stable amount of calibrant

into the Orbitrap [4] analyzer. The mass accuracy of these instruments is therefore similar to the prior generations sector instruments used at the beginning of the era of mass spectrometry based peptide sequencing [5], but now at unprecedented sensitivity and data acquisition rates. While the community has embraced these new instrument platforms capable of very high mass accuracy, little attention has been paid to what is actually gained by this additional information, especially when acquired on fragment ions.

Unlike other hybrid mass spectrometers such as Q-TOFs [6] or Q-FTICRs [7], which have a single detector, the LTQ-FT and LTQ-OT instruments allow parallel data acquisition in both mass analyzers by use of dual detectors. This silent revolution in mass analyzers opens up new data acquisition schemes. Typically, the FT-based analyzer performs a survey scan of MS1 ions while the linear ion trap (LTQ) performs MSⁿ scans based upon data-dependent information. This acquisition mode results in the highest available duty cycle (i.e., highest number of acquired CID spectra) and is therefore often recommended for high throughput data-dependent proteomic analysis. Nevertheless, other

Address reprint requests to Dr. D. R. Goodlett, Department of Medicinal Chemistry, University of Washington, Box 357610, Seattle, WA 98195-7610, USA. E-mail: goodlett@u.washington.edu

data acquisition modes are possible with these instruments. For example, CID fragment ions created in the linear ion trap can be transferred to the FT-based analyzer and detected with high measured mass accuracy. Additionally, the LTQ-OT instrument allows dissociating precursor ions during the ion transfer process via a “C-trap” [4]. The C-trap functions as a quadrupole-like ion storage device between the LTQ and the Orbitrap analyzer, and is operated at typical ion trap gas pressure (10^2 Pa) [4, 8]. Following C-trap CID, fragment ions are transferred to the Orbitrap, where they are detected at high measured mass accuracy and resolution. Transferring CID fragment ions from the linear ion trap to the FT-based analyzer was successfully used on hybrid Orbitrap instruments for the top-down identification of whole proteins (using direct infusion), where the high measured mass accuracy and resolution adds more specificity during the database search [9]. In addition, acquisition of tandem mass spectra with high measured mass accuracy and resolution was used at the peptide level in data-dependent acquisition mode to confirm the presence of post-translational modifications [10].

After data acquisition, a high throughput data analysis pipeline, usually consisting of database search followed by a probabilistic assessment of a false-positive ratio is necessary to handle the large and complex datasets generated by these hybrid instruments. Software widely used in these pipelines to match peptide tandem mass spectra to sequence in databases include Sequest [11], Mascot [12], Spectrum Mill, Sonar [13], and Phenix [14]. Interestingly, in spite of the popularity of these recently developed LTQ-hybrid [1, 2] and current Q-TOF hybrid instruments [15, 16], most of the available software packages do not take advantage of the high mass accuracy recorded for fragment ions, and among those that do there is great variation in how it is allowed. For example, Sequest uses a default tolerance of 1 Da for matching fragment ions while Mascot allows the user to restrict the tolerance on fragment ion matching to 0.01 Da and X!tandem allows the user to restrict this tolerance to any user-defined relative value. Alternatively, Phenix uses fragment ion tolerance to optimize the scoring function for a given instrument type, but this parameter is not directly accessible in the submission interface page [14]. The fact that not every database search software uses a user-defined fragment ion mass tolerance is surprising, especially since (1) high mass accuracy of fragment ions has been shown to be of great benefit for peptide sequencing [17], (2) this type data has been available for several decades before development of recent hybrid instruments [5], and (3) the quality of the tandem mass spectrum has been shown to be an important variable during the database search process [18].

Here we present results from a comparison of tandem mass spectral data acquired with typical ion-trap mass accuracy and resolution to data acquired with much higher measured mass accuracy and resolution available in FT based mass analyzers. In both cases,

complex biological samples are analyzed by data-dependent LC-MS/MS using Fourier-transform hybrid dual detector instruments. We discuss the effect of the mass accuracy, duty cycle (defined as the number of tandem mass spectra acquired per time unit), and instrument sensitivity on matching peptide sequence to tandem mass spectra. To gauge the sensitivity and specificity of database search results for each data acquisition mode we calculate the false positive ratio estimated using searches of reverse databases [19]. In addition, the results of high-energy C-trap fragmentation in the Orbitrap instrument are presented. Finally, the potential benefit of this type data on searches of modified peptides using high measured mass accuracy on fragment ions is also investigated using the software Popitam [20]. In contrast to most peptide fragment ion fingerprint software, Popitam allows the user to search for an undefined mass shift representing any possible modification, i.e., biologic post-translational modification (PTM) or chemical artifact at any amino acid. In addition, fragment ion tolerance is a variable parameter with this software, allowing us to characterize easily the added value of high measured mass accuracy without need to specify a specific mass shift.

Experimental

Sample Preparation

Soluble protein extracts were prepared from a clinical strain of *Pseudomonas aeruginosa*, isolated from a cystic fibrosis patient [21]. Briefly, the bacteria were grown to 1.2 o.d. in LB broth. Alternatively, *Pseudomonas aeruginosa* were grown on solid media and removed with a cell scraper and washed with ammonium bicarbonate buffer. They were then spun down and washed with ammonium bicarbonate buffer, lysed with a French press, and the insoluble material removed by ultracentrifugation.

Protein concentration was determined using a Coomassie-based protein assay (Pierce/ThermoFisher, San Jose, CA). Proteins were reduced with dithiothreitol (DTE), alkylated with iodoacetamide, and digested with sequencing-grade porcine Trypsin (Promega, Madison, WI). Peptides were then desalted on a Vydac C18 microspin column (The Nest Group, Southborough, MA) according to the manufacturer's instructions.

Mass Spectrometry and HPLC

Unless otherwise specified, all data were acquired in three technical replicates. Peptide digests were analyzed by electrospray ionization in the positive ion mode on two FT-based mass analyzers (ThermoFisher, San Jose, CA): (1) a linear ion trap-Orbitrap, known as the LTQ-Orbitrap [2] and (2) a linear ion trap-FT-ion cyclotron resonance mass spectrometer, LTQ-FT [1]. Both instruments were equipped with a nanospray ion source. The LTQ-Orbitrap was equipped with a nano-flow HPLC system (NanoAcquity; Waters Corp., Mil-

ford, MA) fitted with a home-built helium-degasser. The LTQ-FT was equipped with a nanoflow HPLC system (Paradigm MS4B; Michrom Bioresources, Auburn, CA) coupled to an autosampler (Paradigm Endurance; Michrom). Peptides were trapped on a home-made $100\ \mu\text{m}$ i.d. \times 18 mm long pre-column packed with $200\ \text{\AA}$ ($5\ \mu$) Magic C18 particles (C18AQ; Michrom), and subsequently separated on a home-made gravity-pulled $75\ \mu\text{m}$ i.d. \times 150 mm long analytical column packed with $100\ \text{\AA}$ ($5\ \mu\text{m}$, C18AQ; Michrom) coupled to the mass spectrometer.

For each injection, an estimated amount of $0.5\ \mu\text{g}$ of digested *P. aeruginosa* proteins grown in suspension for FT-Orbitrap analysis or $2.0\ \mu\text{g}$ of digested *P. aeruginosa* grown on solid media for FT-ICR analysis ($5\ \mu\text{L}$ injection volume) was loaded onto the pre-column at $5\ \mu\text{L}/\text{min}$ in water/acetonitrile (95/5) with 0.1% (vol/vol) formic acid. Peptides were eluted using an acetonitrile gradient flowing at $250\ \text{nL}/\text{min}$ using mobile phase consisting of: A, water, 0.1% formic acid; B, acetonitrile, 0.1% formic acid. The gradient program was: 0 min: A (95%), B (5%), 55 min: A (65%), B (35%), 60 min: A (15%), B (85%), 65 min: A (5%), B (95%), 75–90 min: A (95%), B (5%); (stop). The electrospray voltage was applied via a liquid junction using a gold wire inserted into a micro-tee union (Upchurch Scientific, Oak Harbor, WA) located in between the pre-column and analytical column. Ion sources conditions were optimized using the tuning and calibration solution recommended by the instrument provider. For the dilution experiment, the same sample volume was injected. Before the injection, the samples were diluted 10 times and 100 times. The dilution experiment was performed on a new chromatographic column and pre-column, from the least concentrated sample to the most concentrated sample. This experiment was performed in duplicate.

The Orbitrap instrument was used in the following manner: For the MS survey scans, the Orbitrap resolution was set to 60,000 (m/z 400) and ion populations were held at 5×10^5 through the use of automatic gain control (AGC). All MS survey scans were performed from m/z 400 to 2000. For tandem mass spectrometry in the linear ion trap, the ion population was set to 1×10^4 and the precursor isolation width was set to $2\ m/z$. For tandem mass spectrometry with ion detection in the Orbitrap, the ion population was set to 2×10^5 and the precursor isolation width to $4\ m/z$. Maximum ion injection time was set to 800 ms. Collision energy was set to 35% for CID in the LTQ. For C-trap fragmentation, collision energies of 20% to 80% of the maximal available value were used. Injection waveforms were turned on for all ion injections.

The FT-ICR instrument was used with the same parameters with the following exceptions: For MS survey scans, the resolution was set to 100,000 (m/z 400) and the ICR ion populations were held at 1×10^6 through the use of AGC. The ion population for the

acquisition of CID fragment ions in the FT-ICR cell was set to 5×10^5 with resolution set to 25,000.

All data-dependent analyses were acquired using MS survey scans in the Orbitrap and/or ICR cell followed by data-dependent selection of the three most abundant precursors for tandem mass spectrometry. Singly-charged ions were excluded from data-dependent analysis. Data redundancy was minimized by excluding previously selected precursor ions ($-0.1/+1.1\ \text{Da}$) for 60 s following their selection for tandem mass spectrometry. The size of the exclusion list was limited to 140. Data were acquired using Xcalibur, version 2.0 (ThermoFisher).

For direct comparison between CID acquisition in the Orbitrap and in the linear ion trap, the same precursor was acquired in both devices. In a first injection, acquisition was first performed in the Orbitrap, and then in the linear ion trap. In the second injection, the linear ion trap acquisition was performed before the Orbitrap acquisition. The experiment was then repeated (technical duplicate).

Database Search

Raw data from mass spectrometric acquisitions were converted to .dta files format using Bioworks 3.0 software (Thermo Fisher). In the situation where tandem mass spectra were acquired in both devices during the same analysis (i.e., linear ion trap and Orbitrap), the .dta files were separated according to the analyzer with an in-house written perl script. Database search was performed with Phenix (GeneBio SA, Geneva, Switzerland) against the *Pseudomonas aeruginosa* PA01 database (<http://www.Pseudomonas.com>), originally annotated in September 2000. The database contains 5570 protein entries. All false positive ratios were evaluated using the corresponding reverse protein database. Precursor ion tolerance was set to 2.1 Da. As specified in the text, either trypsin specificity with one missed cleavage or half-tryptic specificity (one specific cleavage on either C-terminal or N-terminal, the other nonspecific) with up to two missed cleavage sites were specified. Methionine residues were considered as being present in reduced and oxidized form. Cysteine residues were considered alkylated with iodoacetamide. ROC-like curves were constructed by plotting sensitivity as a function of $1 - \text{specificity}$. Sensitivity is defined as the number of true positives divided by the number of true positives plus false negatives. Specificity is defined as the number of true negatives divided by the number of true negatives plus false positives. The true positives are approximated as the number of identifications in the forward database above the considered z-score. True negatives are approximated as the number of identifications in the reverse database below the z-score. False positives are approximated as the number of identifications in the reverse database above the z-score. False negatives are approximated as the number of identifications in the forward database below the z-score.

Open modification search (open gap mode) was performed using Popitam software [20], accessible through the ExPASy server (<http://expasy.org/tools/popitam/>). Briefly, any mass modification (gap) between -100 and $+500$ Da was searched, using trypsin specificity and one missed cleavage. The smallest possible tolerance (0.01 Da) was used for fragment ion matching. The potentially modified peptides were searched in a database containing all previously identified proteins. Rather than a simple reversed sequence database search as a control, a database was used containing the double number of proteins, randomly retrieved from the Swiss-Prot database. Indeed, since Popitam is searching for a sequence tag in the spectrum, searching in a reverse database would result with an inversion of C-terminal and N-terminal ions.

Results

LTQ-FT Acquisitions

In a first set of experiments, data-dependent tandem MS acquisition was performed in both the linear ion trap and the ICR cell of an LTQ-FT instrument. To reach the desired ion population in the ICR cell, the maximum ion injection time was set to 800 ms and the precursor ion isolation window was opened to 4 m/z . The ion population was set to 500,000 to reach sufficient signal to noise for accurate fragment ion detection. Typical tandem mass spectra, acquired in a data-dependent manner in both the FT-ICR and linear ion trap are shown in Figure 1. The spectra display similar quality in terms of ion abundance and ion series. However, some low and high m/z fragment ions are of less abundance and/or are missing in the FT-ICR spectra. The ion injection times as well as the number of acquired tandem mass spectra are shown in Table 1.

A review of the spectra reveals that only a few good quality spectra (in terms of number and abundance of ions) were obtained from the FT-ICR acquisitions of fragment ions. Indeed, the maximum ion injection time was reached for a large proportion of the tandem mass spectra, resulting in noisy spectra containing only very abundant ions. Often, only the harmonic noise peaks (due for example to turbo pumps and local radio transmitter frequencies) are visible in those spectra. In contrast, the corresponding spectra acquired in the linear ion trap are often of high quality in terms of ion abundance and number of ions interpretable as peptide fragment ions.

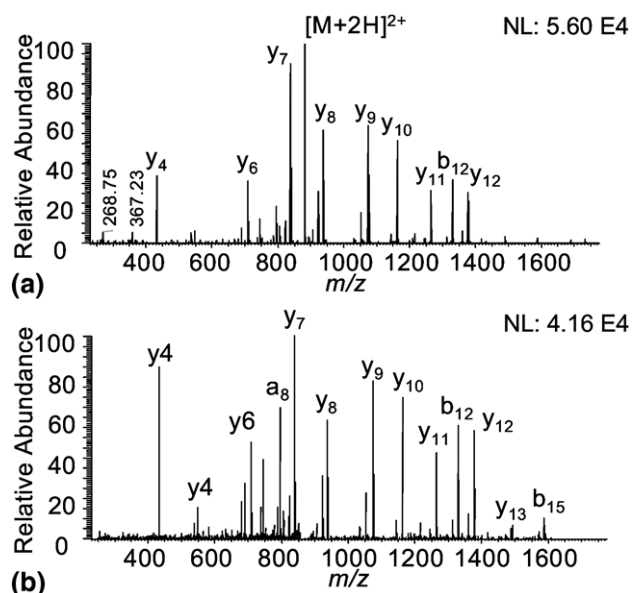


Figure 1. CID spectra of $[M + 2H]^{2+} = 632.88$ Da, resulting with the identification of peptide GITINTSHVEYDSAVR from *Pseudomonas aeruginosa* protein PA4265, coded by the gene *tufA*. (a) FT-ICR acquisition. The ion injection time was 319 ms, the total elapsed scan time was 750 ms at resolution 25,000. (b) Linear ion trap acquisition. The ion injection time was 5 ms, the total elapsed scan time was 230 ms.

After the database search, the number of identified peptides was about three times higher in the linear ion trap acquisitions than in the FT-ICR acquisitions (data not shown).

LTQ-Orbitrap Acquisitions

In a second set of experiments, data-dependent LC-MS/MS acquisitions were performed in both the linear ion trap and the Orbitrap analyzer of an LTQ-Orbitrap instrument. The same precursor ion isolation window and same maximum ion injection time as used previously were used for the FT-Orbitrap acquisition (4 m/z , 800 ms maximum injection time). A typical tandem mass spectrum where fragment ions are acquired in the Orbitrap versus the linear ion trap is shown in Figure 2. Here, both tandem mass spectra show similar relative ion abundance and richness of spectra between the different fragment ions over the entire m/z range. The only major difference, aside from mass accuracy and

Table 1. Peptide amount loaded on column, number of acquired tandem mass spectra and ion injection time for different scanning methods

Analyzer	Activation method	Peptide amount loaded (mg)	Mean number of tandem mass spectra	SD	Mean ion injection time	SD
FT	CID	2.0	1623	64	321	58
OT	CID	0.5	2579	87	378	58
OT	HCD 40%*	0.5	2062	104	530	58
LIT	CID	0.5	4202	12	41	2

* indicates % of maximal available collision energy.

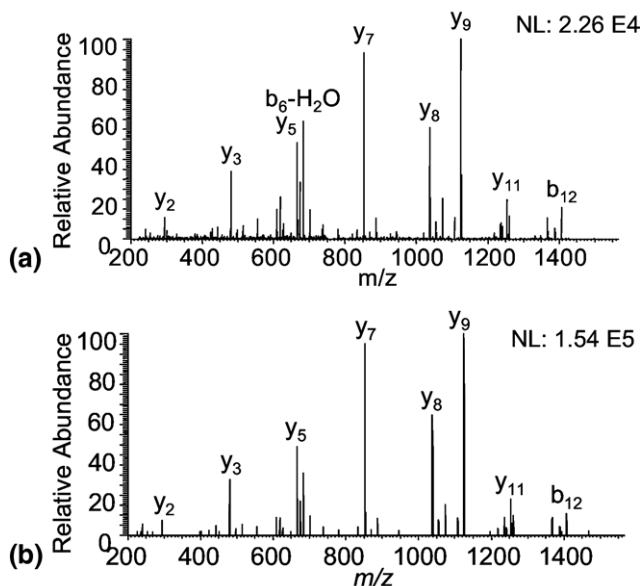


Figure 2. CID spectra of $[M + 2H]^{2+} = 776.87$ Da, resulting with the identification of peptide SVLQSWADGEWFK from *Pseudomonas aeruginosa* protein PA1787, coded by the gene *acnB*. (a) Orbitrap acquisition. The ion injection time was 244 ms, the total elapsed scan time was 490 ms for a resolution of 7500. (b) Linear ion trap acquisition. The ion injection time was 14 ms, the total elapsed scan time was 240 ms.

resolution, is the ion injection time and the total scan time (Table 1).

CID Versus C-Trap Fragmentation

The Orbitrap instrument allows CID during ion transfer from the linear ion trap to the Orbitrap via a device known as the C-trap. Fragmentation in the C-trap is referred to as “C-trap fragmentation” or “higher collision decomposition” (HCD). This C-trap CID mode with fragment ions acquired in the Orbitrap was therefore compared to CID in the linear ion trap with Orbitrap acquisition of fragment ions. There are two major differences between CID in the C-trap versus the linear ion trap: (1) collision energy is higher and (2) there is no loss of low m/z ions as a function of the ejection q -value that limits the lower m/z range detected from an ion trap scan [22]. However, in contrast to the results described by Olsen et al. [4], precursor ion activation is performed during ion transfer from the linear ion trap to the C-trap, and not in a supplemental hexapole located at the back-end of the C-trap. Typical spectra are shown in Figure 3. The ion injection times and the total scan times are similar between both data-acquisition modes (Table 1).

The collision energy used for C-trap fragmentation is a critical parameter that affects tandem mass spectral quality. For example, precursor ions are generally base peaks at 20% to 30% normalized collision energy. Precursor ions are still abundant at 40–50% collision energy, but typical peptide fragment ions (such as b - and

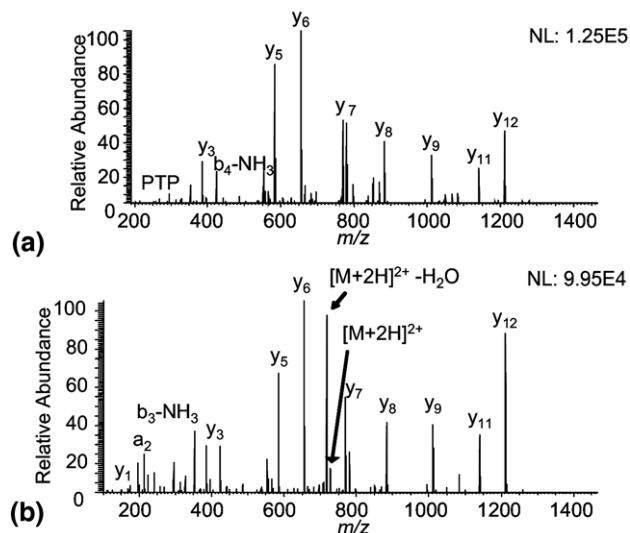


Figure 3. CID spectra of $[M + 2H]^{2+} = 726.87$ Da, resulting with the identification of peptide QIAGAEIDAPTNR from the *Pseudomonas aeruginosa* protein PA2009, coded by the gene *hmgA*. (a) CID spectrum with Orbitrap acquisition. The ion injection time was 527 ms, the total elapsed scan time was 770 ms at a resolution of 7500. (b) C-trap fragmentation spectrum at 50% of the normalized collision energy. The ion injection time was 649 ms, the total elapsed scan time was 890 ms for a resolution of 7500.

y -ions) are base peaks. If collision energy is further increased (e.g., 60% to 80% of the maximum available value), then previously undetected low mass ions (such as immonium and the a_2/b_2 ion pair) become dominant (e.g., the base peak shift towards lower m/z). To compare results, database searching was performed for all data acquired at three different collision energy ranges. A maximum number of unique peptide identifications were observed at values between 40% to 50% normalized C-trap collision energy with minimal numbers of immonium ions present (Table 2). At higher collision energy, the proportion of spectra with detected immonium ions is higher, but the total number of spectra leading to a peptide identification decreases (Table 2).

Influence of High Measured Mass Accuracy

Qualitatively, data acquisition in the Orbitrap results in relatively high quality spectra. However, the ion injection times and therefore the duty cycle of the instrument is longer in this acquisition mode. Therefore, the performance of this data acquisition with regard to peptide/protein identification after database search was investigated and compared to linear ion trap acquisitions. In particular, the added value of the higher resolution and measured mass accuracy was investigated in the context of the loss in sensitivity and duty cycle. For this purpose, tandem mass spectra of the same precursor ions were simultaneously acquired in the Orbitrap and in the linear ion trap.

The database search was performed on both datasets using a forward and a reverse database, and corre-

Table 2. Number of identified unique peptides and total number of detected immonium ion signals as function of collision energy with C-trap fragmentation

Collision energy (% of max)	# Identified unique peptides	SD	Mean # immonium ions					
			K immonium ion	SD	F immonium ion	SD	R immonium ion	SD
20	177	235	28	38	25	35	0	0
30	603	26	82	47	42	30	0	0
40	1028	39	273	15	248	25	0	0
50	1024	18	445	18	303	4	1	1
60	893	28	370	1	210	4	1	0
70	523	39	267	7	130	13	5	1
80	48	5	254	28	133	35	7	2

sponding ROC-like curves derived. Since the spectra are qualitatively similar (besides the higher mass accuracy and resolution), only the m/z tolerance window for fragment ion matching was modified for Orbitrap acquisition data. This parameter was varied between 2 and 30 ppm. The ROC-like curves showed that error values between 8 and 12 ppm display the lowest false-positive/true-positive ratio (Figure 4a). A value of

12 ppm was further used for all other database searches with Orbitrap acquisitions. Then, the Orbitrap acquisitions were directly compared with the linear ion trap acquisitions. Full tryptic specificity (two tryptic termini), half-tryptic specificity (only one tryptic terminus), and half-tryptic specificity with the addition of possible methyl-ester modifications on amino acid D, E, S, and T were used. The corresponding ROC-like curves

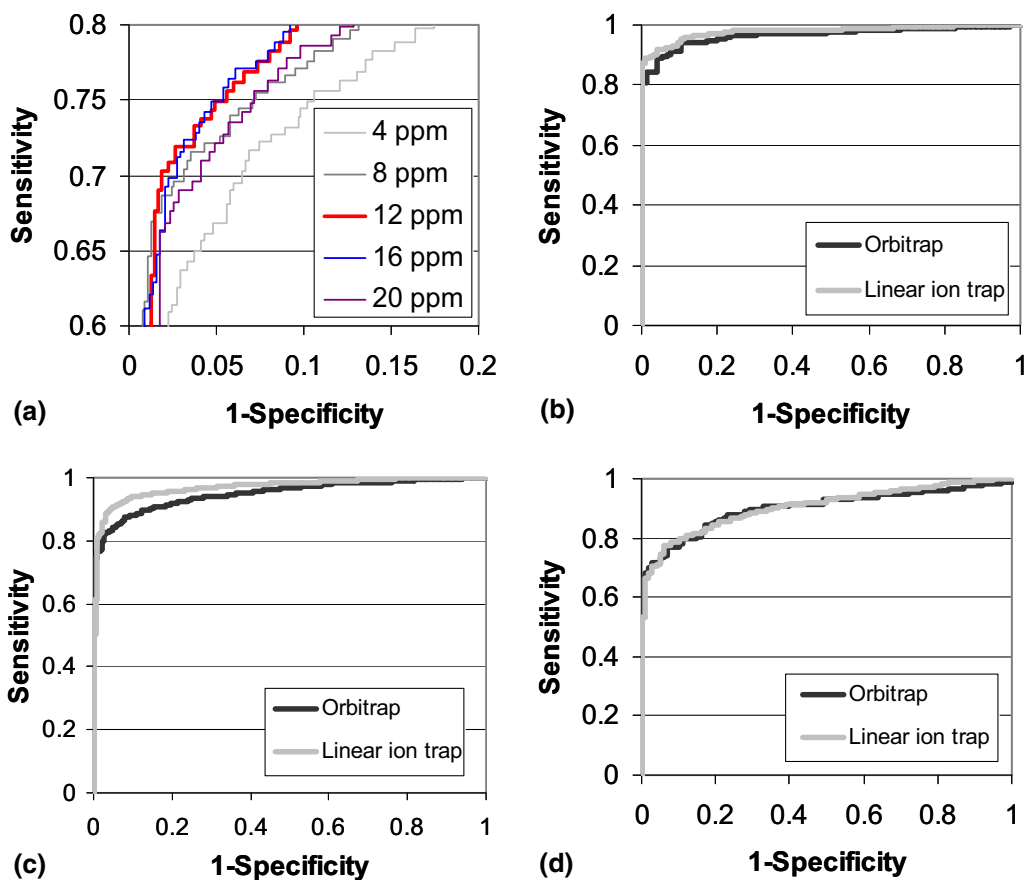


Figure 4. ROC-like curves showing sensitivity as a function of 1-specificity of database searches in different conditions and with different CID spectrum acquisition modes. (a) Orbitrap acquisitions using different fragment ion m/z tolerance. (b) Orbitrap acquisitions versus linear ion trap acquisitions with tryptic specificity and maximum one missed cleavage. (c) Orbitrap acquisitions versus linear ion trap acquisitions with half-tryptic specificity and maximum two missed cleavages. (d) Orbitrap acquisitions versus linear ion trap acquisitions with half-tryptic specificity, maximum two missed cleavages, and potential modification (methyl-ester) on the amino acids D, E, S, and T.

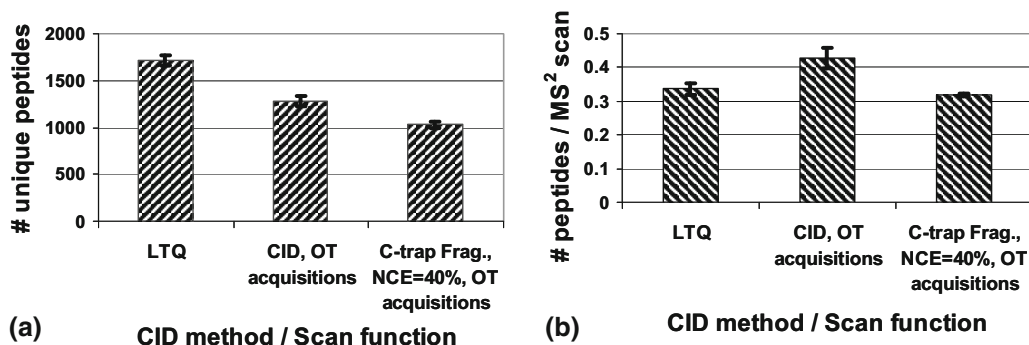


Figure 5. (a) Average \pm standard deviation of unique peptide identifications as function of the CID method and data acquisition mode. (b) Average \pm standard deviation of identification efficiency (number of identified unique peptides per acquired tandem mass spectrum) as function of the CID method and data acquisition mode.

are shown in Figure 4b, c, and d. With full tryptic specificity and at the same false positive rate, the linear ion trap acquisition identified more peptides than the Orbitrap acquisitions (Figure 4b). When the search is repeated with reduced enzyme specificity (i.e., the number of possible candidate peptides increase), the Orbitrap performs better at a very low false-positive rate (Figure 4c). Finally, with less enzyme specificity and potential modifications, both analyzers performed in a similar manner at a higher false-positive rate, but the Orbitrap performs better at a very low false-positive rate (Figure 4d).

Performance of Data Acquisition Modes

After direct comparison between the scan modes, data-dependent analysis was performed using all three acquisition modes. The results were compared as a function of the number of peptides matched to tandem mass spectra for each acquisition mode. As shown in Table 2, tandem mass spectral data acquired in the linear ion trap were compared to CID acquired in the Orbitrap as well as C-trap fragmentation at 40% collision energy with Orbitrap acquisition. The average number of unique identified peptides was, respectively, 1710 for the linear ion trap acquisition, 1283 for CID with Orbitrap acquisition, and 1028 for the C-trap fragmentation (Figure 5a). To determine the identification efficiency, these numbers were divided by the total number of tandem mass spectra acquired by the instrument during each entire data-dependent analysis. Expressed in these terms, CID carried out in the linear ion trap with Orbitrap data acquisition performs best, with an average of 0.43 unique peptides identified per acquired tandem mass spectrum. Values of 0.34 and 0.32 unique peptides per acquired tandem mass spectrum in the linear ion trap and C-trap fragmentation with Orbitrap acquisition were similar, but lower than the Orbitrap value (Figure 5b).

Sensitivity and Detection Limit of Orbitrap Acquisition

As mentioned above, the required ion population and therefore the ion injection times for Orbitrap acquisitions are higher compared with linear ion trap acquisitions (Table 1). To assess how this difference translates into differences of detection limits, a dilution experiment was performed. For this, the total sample amount injected into the instrument was diluted 10 \times (corresponding to 0.05 μ g of peptides) and 100 \times (corresponding to 0.005 μ g of injected peptides). The average number of peptides identified by CID with Orbitrap acquisition decreased by 53% after a 10-fold dilution (average of 547 identified unique peptides), and by 96% after a 100-fold dilution (average of 54 identified peptides). In contrast, the identifications decreased respectively by 46% (average of 922 identified unique peptides) and 91% (average of 156 identified unique peptides) with linear ion trap acquisitions.

Database Search for Modified Peptides

In a final experiment, the potential benefit of fragment ions acquired at high mass accuracy used to identify peptides with expected and unexpected covalent chemical modifications was investigated. To investigate this, the unmatched tandem spectra acquired in the Orbitrap and in the linear ion trap were searched using the software Popitam [20]. This software is designed to search for any mass modifications (i.e., whether from biologic post-translational events or chemical artifactual ones) within a user-specified mass window, and within all possible tryptic peptides from a user-specified protein database. A typical workflow consists of creating a database of all proteins initially identified during a first database search of unmodified sequences. In other words, the search space for the modified peptides consists of all tryptic peptides from all previously identified proteins without modifications. The remaining unmatched spectra are then searched for possible

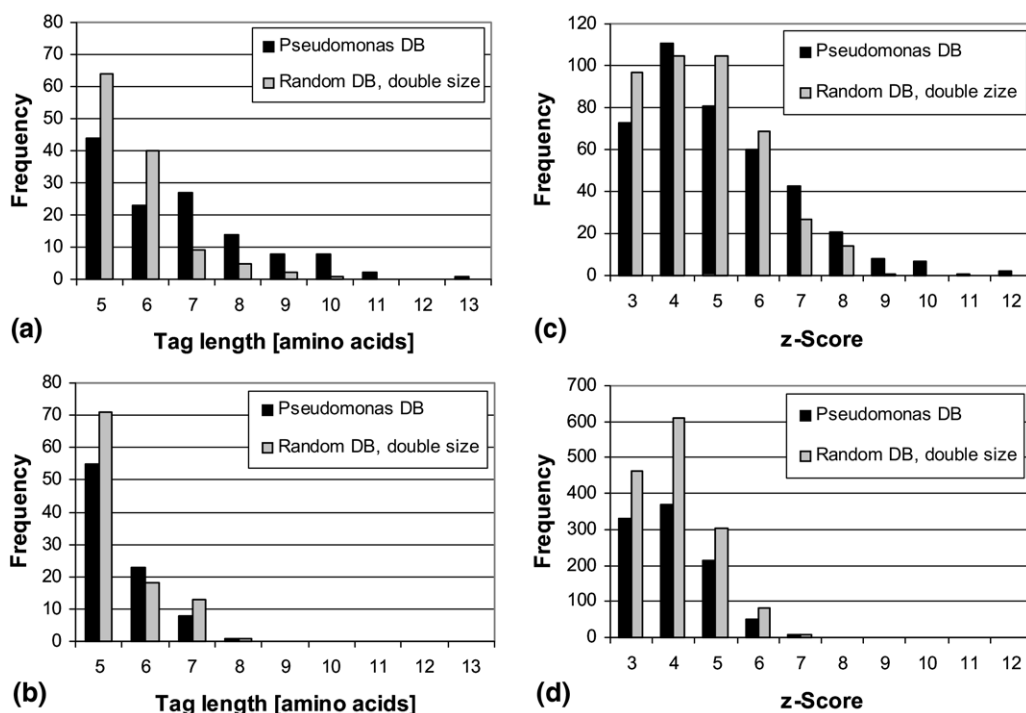


Figure 6. Distribution of identification quality after database search in open-gap mode (search for unexpected modifications). The search is performed in a database containing all identified *Pseudomonas aeruginosa* proteins after database search for unmodified peptides, as well as in a random database (containing the double number of randomly retrieved proteins). (a) Identified non-sequential sequence tag length distribution with CID and Orbitrap acquisition, searched with 0.01 Da tolerance on fragment ions. (b) Identified non-sequential sequence tag length distribution with linear ion trap acquisition, searched with 0.5 Da tolerance on CID fragments. (c) Recalculated Phenyx z-score distribution with CID and Orbitrap acquisition, searched with 0.01 Da tolerance on fragment ions. (d) Recalculated Phenyx z-score distribution with linear ion trap acquisition, searched with 0.5 Da tolerance on CID fragments.

modification resulting in a mass shift of -100 to $+500$ Da. Data from the Orbitrap acquisitions were searched with a 0.01 Da tolerance on fragment ions, whereas the linear ion trap data were searched within a 0.5 Da tolerance. To differentiate between true and false positives, the search was performed in the created first-pass database containing the previously identified proteins and a “random database,” containing twice as many proteins, randomly retrieved from the SwissProt database. Two metrics were used to differentiate peptides identified in the real versus random database: identified nonsequential sequence tags and the recalculated Phenyx z-score. The first one, nonsequential sequence tags, corresponds to the number of amino acids in the peptide sequence identified by adjacent *b*- and/or *y*-ions observed in the spectrum. The second one, recalculated Phenyx z-score, corresponds to the z-score the database software would have given in the event that the observed modification would have been considered during the search. Searching the linear ion trap data did not show any discrimination between the number of hits in the real and random database. However, a high discrimination was observed for sequence tags of six amino acids and more with the Orbitrap acquisitions. The distribution of peptide sequence tag length in both

real and random databases for both data acquisition modes is shown in Figure 6a and b. Most peptides identified from Orbitrap acquisitions with sequence tags of six or more amino acids revealed common biological modifications and artifacts due to sample preparation and handling. Indeed, mass shifts corresponding to oxidation, adduction of sodium, methylation, and dethiomethylation (loss of 48 Da) were commonly observed. In contrast, the corresponding linear ion trap acquired peptides have equally distributed modifications over the entire specified mass range (data not shown).

Discussion

LTQ CID with FT-ICR Acquisitions

The potential benefits and drawbacks of data-dependent CID acquisition with high measured mass accuracy were evaluated. All data were acquired on an LC-MS time scale. Since the number of ions transmitted to the FT-ICR analyzer is required to be high compared with the linear ion trap acquisitions, the ion injection times are typically in the half-second time-scale or

higher. This required the target ion population to be 50 times higher for FT-ICR acquisitions than for the linear ion trap and for ion injection times to be high with longer spectral acquisition times as well. These requirements resulted in a duty cycle > 1 s and a concomitant decrease in the volume of data acquired. Furthermore, the spectral quality in terms of relative ion abundance over the entire m/z range varies compared with linear ion trap acquired data. In fact, spectra acquired in the FT-ICR show low and high m/z regions as under-represented, due to the so-called “time-of-flight effect” during the transfer of ions from the ion trap to the FT-ICR cell. It was previously suggested that this could be corrected by separate data acquisitions, in a “stitched” fashion, optimized either for low or high m/z regions and then combined post-acquisition [23]. However, this would make the duty cycle even lower, and is therefore not compatible on an LC timescale for data-dependent acquisitions intended to profile proteomes efficiently. On the other hand, sufficient sensitivity can be obtained with a large enough precursor ion isolation window and a limited ion population in the ICR cell. Data acquisition time can also be reduced by lowering the spectral resolution. We used $R = 25,000$ in this study, but this value could be further reduced without sacrificing measured mass accuracy. In any case, data-dependent LC-MS/MS acquisition is possible in the FT-ICR cell, although the overall performance in terms of number of identifications is limited, when compared to linear ion trap acquisition. Therefore, such a data acquisition strategy might be suited for particular samples and purposes like de novo sequencing, identification or validation of a particular post-translational modification, or any other assay where high measured mass accuracy or resolution might be necessary.

LTQ CID with Orbitrap Acquisitions

Transferring fragment ions produced in the linear ion trap into the Orbitrap was shown to be more efficient than transferring ions into the ICR cell. Indeed, mass spectral quality in terms of relative ion abundance over the entire m/z range was similar to acquisition in the linear ion trap, whereas only limited m/z range transmission was observed from the linear ion trap to ICR cell. Also, the required ion population in the Orbitrap analyzer is lower compared with the ICR cell, leading to shorter ion injection times. This makes mass spectral acquisition time at lower resolution (typically 7500) extremely fast (about 110 ms), allowing for a high instrument duty cycle, comparable to linear ion trap acquisitions. It should be noted that acquiring data at a spectral resolution of 7500 is enough to provide high measured mass accuracy as well as monoisotopic resolution for typical peptide fragment ion charge states (e.g., $z = 1, 2, \text{ or } 3$).

C-Trap Fragmentation with Orbitrap Acquisitions

C-trap CID with Orbitrap fragment ion acquisition gave similar results to linear trap CID with Orbitrap acquisition in terms of ion injection times and total scan times. In addition, C-trap fragmentation gave access to the full mass range, whereas the low-mass ions were previously excluded when tandem mass spectra were produced in the linear ion trap. Importantly, to be effective for a given task the collision energy for C-trap fragmentation is a critical experimental parameter that must be optimized. Here we note two basic regimes as useful for peptide analysis: (1) at 60% of the maximum available collision energy where immonium type ions are preferentially produced relative to higher m/z ions, and (2) at 40% of the maximum available collision energy where peptide fragment ions dominate the tandem mass spectra and immonium ions are missing. Indeed, the optimal window between too abundant a precursor ion and too many low mass ions is narrow. In addition, the higher collision energies result in a dramatic decrease in the number of “protein” identifications after database search (Table 2). Exactly why this decrease occurs is difficult to gauge because almost all sequence-associated fragment ions are still present in these tandem mass spectra. It may be caused by ion scattering and poor ion recovery during C-trap fragmentation, which is corroborated by the fact that immonium ions appear only at higher collision energies that are not optimal for matching peptide tandem mass spectra to peptide sequence in a database.

Influence of Mass Accuracy

The Orbitrap analyzer delivers m/z measurement accuracy in the 2 to 4 ppm range. This performance could potentially be improved by the use of the so-called “lock-mass” function [4], although this introduces a supplemental cost in duty cycle, which is not well suited on a LC-MS/MS timescale and was therefore not used in present study. For this study, the optimal fragment ion tolerance for database searching was determined to be in the range 8 to 12 ppm. As expected, most of the ions with highest relative abundance display higher mass accuracy than those of low relative abundance, but inclusion of a large number of these low abundance ions with lower mass accuracy can have a positive outcome on the database search. In fact, their inclusion renders the search result more specific at a given sensitivity (e.g., more true positives at the same false-positive ratio) because the higher number of accounted ions increases the z -score for a given tandem mass spectrum, providing better differentiation between true and false positive matches than without these low abundant ions.

Direct comparison of fragment ion acquisitions in the linear ion trap and in the Orbitrap demonstrates how powerful the latter mode can be. Indeed, the ROC-like curves presented in Figure 4b, c, and d show clearly that

similar performance can be obtained when the data is used for complex database searches. In this situation, the duty cycle is the same, since the same precursor ions are always acquired in both analyzers. The lower sensitivity of the Orbitrap compared with linear ion trap acquisition (due to the necessary higher number of ions in the analyzer) results in poorer performing ROC-like curve for a constrained search (Figure 4b), but this effect is compensated by the higher specificity due to the higher measured mass accuracy. These two parameters balance each other, resulting in similar number of identifications at same false positive ratio even for more complex searches (e.g., less enzyme specificity, possible modifications, etc.).

Instrument Duty Cycle and CID Efficiency

Data-dependent analysis performed with the different data acquisition modes affects the duty cycle of the instrument. Therefore, the total number of precursors potentially selected for data-dependent CID varies as a function of data-acquisition mode. Indeed, fragment ion acquisitions performed in the linear ion trap resulted in the highest number of identifications whereas data-dependent acquisition with C-trap fragmentation provided fewer peptide identifications, but full m/z coverage. In the case where sufficiently high C-trap collision energy is used, full m/z coverage provides access to immonium ion information. The latter point can be important for de novo sequencing or for the correct assignment of modified peptides (for example those containing post-translational modifications). The tandem MS efficiency, in terms of identified unique spectra per acquired tandem mass spectra, was highest in the case of CID followed by Orbitrap acquisition.

Sensitivity and Detection Limits of Orbitrap Acquisitions

Due to the requirement of more ions for Orbitrap acquisitions compared with linear ion trap acquisitions, the sensitivity and detection limit of Orbitrap acquisitions was expected to be lower. Indeed, the dilution experiment showed that the number of identified unique peptides decreases slightly more rapidly in the Orbitrap compared with linear ion trap acquisitions. However, it should be noted that our initial comparisons between tandem mass spectral acquisition in the linear ion trap versus Orbitrap acquisitions (before the dilution experiment) are based on chromatographic injections of 0.5 μg of digested protein mixture. This amount is typical for shotgun proteomic type data-dependent acquisitions with or without relative quantification and provides the highest chromatographic resolution. Moreover, CID with Orbitrap acquisition of tandem mass spectra was shown to be a sensitive technique if the maximum ion injection times were long enough.

Unexpected and Modified Peptides

Finally, the search for modified and unexpected peptides after acquisitions using the linear ion trap and the Orbitrap demonstrated another advantage of measuring fragment ions at high mass accuracy. Indeed, only data acquired with high measured mass accuracy allowed discrimination of matches between the forward database of identified unmodified proteins and a database containing randomly selected sequences. The nature of the identified modifications confirmed this finding. Relatively few but commonly observed modifications were identified from the data acquired at high measured mass accuracy. In contrast, linear ion trap data resulted in many possible mass shifts, which corresponded neither to commonly observed modifications nor artifacts. This fact confirms that there is no discrimination between true and false positive identification of modified peptides when using tandem mass spectra acquired at lower measured mass accuracy. This was observed using both considered "quality metrics," e.g., identified sequence tag length and recalculated Phenix z-score. Obviously, these tandem mass spectra acquired at lower mass accuracy are a disadvantage for discovery of post-translationally modified peptides, which are of great biological importance. In addition, the large part of other modifications such as chemical artifacts produced during sample preparation, atypical cleavages, and amino acid substitution are also more difficult to confidently identify from these low mass accuracy tandem mass spectra. Thus, a software analysis pipeline that utilizes the information encoded in high measured mass accuracy provides more comprehensive data analysis than data acquired at low measured mass accuracy. Here, it should be noted that all tandem mass spectra were acquired in a data-dependent manner. After a first subtraction of spectra matching to expected modifications, all remaining spectra were searched for unexpected modifications. The purpose of this analysis was not to compare acquisitions of the same precursor ions, containing the same modification, in the linear ion trap or in the Orbitrap, but rather to determine if a certain data-dependent acquisition strategy was compatible with identification workflows. In this context, we have shown that high measured mass accuracy greatly improves the extraction of useful information from tandem mass spectral data without significant sacrifice in proteome coverage, provided that the amount of sample analyzed remains above a given threshold.

Conclusions

Modern linear ion trap-Fourier-transformation mass spectrometers are increasingly used for proteomics studies. These instruments offer the possibility to acquire data, including tandem mass spectra, with high measured mass accuracy and resolution. Acquiring tandem mass spectra with this high measured mass accuracy (1–10 ppm) is an interesting alternative to

classical data acquisition schemes where fragment ions are typically acquired at much lower mass accuracy (100–500 ppm) and resolution in the linear ion trap. In our studies the only obvious negative impact of acquiring fragment ions at high mass accuracy was a slightly decreased number of identifications compared to acquisition in the faster scanning ion trap and a higher detection limit. This effect was offset by the additional specificity in identifications (i.e., more identifications at same false positive ratio) obtained from tandem mass spectra recorded at high mass accuracy. As a bonus, the additional information encoded in the high mass accuracy allowed the data to be used to confidently identify peptides with unexpected modifications not possible from the lower mass accuracy ion trap data acquisition schemes.

Acknowledgments

AS acknowledges support by the Swiss National Science Foundation (grant PBGEA-111259). DRG thanks the NIH for support from the following sources: R33CA099139-01, 5P30ES007033-10, and 1U54 AI57141-01.

References

- Syka, J. E.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-Translational Modifications. *J. Proteome Res.* **2004**, *3*, 621–626.
- Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M. Graham Cooks, R. The Orbitrap: A New Mass Spectrometer. *J. Mass Spectrom.* **2005**, *40*, 430–443.
- Wu, S.; Kaiser, N. K.; Meng, D.; Anderson, G. A.; Zhang, K. Bruce, J. E. Increased Protein Identification Capabilities Through Novel Tandem MS Calibration Strategies. *J. Proteome Res.* **2005**, *4*, 1434–1441.
- Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S. Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat. Methods* **2007**, *4*, 709–712.
- Papayannopoulos, I. A. The Interpretation of Collision-Induced Dissociation Tandem Mass Spectra of Peptides. *Mass Spectrom. Rev.* **1995**, *14*, 49–73.
- Morris, H. R.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J. Bateman, R. H. High Sensitivity Collisionally-Activated Decomposition Tandem Mass Spectrometry on a Novel Quadrupole/Orthogonal-acceleration Time-of-Flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889–896.
- Belov, M. E.; Gorshkov, M. V.; Udseth, H. R. Smith, R. D. Controlled Ion Fragmentation in a 2-D Quadrupole Ion Trap for External Ion Accumulation in ESI FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1312–1319.
- Makarov, A.; Denisov, E.; Lange, O. Horning, S. Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977–982.
- Macek, B.; Waanders, L. F.; Olsen, J. V.; Mann, M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol. Cell Proteomics* **2006**, *5*, 949–958.
- Wisniewski, J. R.; Zougman, A.; Kruger, S. Mann, M. Mass Spectrometric Mapping of Linker Histone H1 Variants Reveals Multiple Acetylations, Methylations, and Phosphorylation as Well as Differences Between Cell Culture and Tissue. *Mol. Cell. Proteom.* **2007**, *6*, 72–87.
- Eng, J. K.; McCormack, A. L. Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
- Perkins, D. N.; Pappin, D. J.; Creasy, D. M. Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20*, 3551–3567.
- Field, H. I.; Fenyo, D. Beavis, R. C. RADARS, a Bioinformatics Solution that Automates Proteome Mass Spectral Analysis, Optimizes Protein Identification, and Archives Data in a Relational Database. *Proteomics* **2002**, *2*, 36–47.
- Colinge, J.; Masselot, A.; Giron, M.; Dessingy, T. Magnin, J. OLAV: Towards High-Throughput Tandem Mass Spectrometry Data Identification. *Proteomics* **2003**, *3*, 1454–1463.
- Wolff, J. C.; Eckers, C.; Sage, A. B.; Giles, K. Bateman, R. Accurate Mass Liquid Chromatography/Mass Spectrometry on Quadrupole Orthogonal Acceleration Time-of-Flight Mass Analyzers Using Switching between Separate Sample and Reference Sprays. II. Applications Using the Dual-Electrospray Ion Source. *Anal. Chem.* **2001**, *73*, 2605–2612.
- Eckers, C.; Wolff, J. C.; Haskins, N. J.; Sage, A. B.; Giles, K. Bateman, R. Accurate Mass Liquid Chromatography/Mass Spectrometry on Orthogonal Acceleration Time-of-Flight Mass Analyzers Using Switching between Separate Sample and Reference Sprays. I. Proof of Concept. *Anal. Chem.* **2000**, *72*, 3683–3688.
- Spengler, B. 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. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 703–714.
- Venable, J. D. Yates, J. R. III. Impact of Ion Trap Tandem Mass Spectra Variability on the Identification of Peptides. *Anal. Chem.* **2004**, *76*, 2928–2937.
- Elias, J. E. Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat. Methods* **2007**, *4*, 207–214.
- Hernandez, P.; Gras, R.; Frey, J. Appel, R. D. Popitam: Towards New Heuristic Strategies to Improve Protein Identification from Tandem Mass Spectrometry data. *Proteomics* **2003**, *3*, 870–878.
- Hixson, K. K.; Adkins, J. N.; Baker, S. E.; Moore, R. J.; Chromy, B. A.; Smith, R. D.; McCutchen-Maloney, S. L. Lipton, M. S. Biomarker Candidate Identification in *Yersinia pestis* Using Organism-Wide Semi-quantitative Proteomics. *J. Proteome Res.* **2006**, *5*, 3008–3017.
- Jonscher, K. R. Yates, J. R. III. The Quadrupole Ion Trap Mass Spectrometer—a Small Solution to a Big Challenge. *Anal. Biochem.* **1997**, *244*, 1–15.
- Macek, B.; Olsen, J. V.; Zhang, Y. Mann, M. Assessment of High Versus Low Mass Accuracy MS/MS for Complex Mixture Analysis in Proteomics. Poster Presentation. *Proceedings of the American Society of Mass Spectrometry*; San Antonio, TX, June 2005.