Published in final edited form as:

Anal Chem. 2011 October 15; 83(20): 7651–7656. doi:10.1021/ac201843e.

Increased Throughput of Proteomics Analysis by Multiplexing High-resolution Tandem Mass Spectra

A. R. Ledvina^{2,§}, M. M. Savitski^{1,#,§}, A. R. Zubarev¹, D. M. Good¹, J. J. Coon², and R. A. Zubarev^{1,3,*}

¹Division of Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17 177 Stockholm, Sweden

²Department of Chemistry and Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, 53706-1322, USA

³Science for Life Laboratory, Stockholm, Sweden

Abstract

High-resolution and accuracy Fourier-transform mass spectrometry (FTMS) is becoming increasingly attractive due to its specificity. However, the speed of tandem FTMS analysis severely limits the competitive advantage of this approach relative to faster low-resolution quadrupole ion trap MS/MS instruments. Here we demonstrate an entirely FTMS-based analysis method with a 2.5–3.0 fold greater throughput than a conventional FT MS/MS approach. The method consists of accumulating together the MS/MS fragments ions from multiple precursors, with subsequent high-resolution analysis of the mixture. Following acquisition, the multiplexed spectrum is deconvoluted into individual MS/MS datasets which are separately submitted for peptide identification to a search engine. The method is tested both *in silico* using a database of MS/MS spectra as well as *in situ* using a modified LTQ-Orbitrap mass spectrometer. The performance of the method in the experiment was consistent with theoretical expectations.

Introduction

Mass spectrometry (MS) based proteomics experiments typically begin with proteolytic digestion of a protein-containing sample. This complex mixture is then chromatographically separated and interrogated by MS and tandem MS (MS/MS). Mass accuracy in both the MS and MS/MS mode is closely tied to the confidence of the resulting peptide identifications.^{1–4}

This relationship implies that the use of high-resolution and accuracy mass analyzers (i.e. FT-ICR and FT Orbitrap) should result in higher confidence peptide identification than lower resolution and accuracy mass analyzers. Another important requirement, however, is data acquisition speed, and here the FT analyzers face a fundamental limitation manifested by the strong link between their resolving power, (*R*), and the length of the acquisition period (*T*):

$$T \approx 2f/R, \quad (1)$$

where *f* [Hz] is the resonance frequency of the ion of interest.⁵ At a resolving power of 100,000 and *f* = 50 kHz, the acquisition period (*T*) is 1 s, or the maximum rate of acquisition

*Corresponding author: Roman.Zubarev@ki.se phone/+46 8 524 87 594.

#Current address: Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany

§These authors' contribution is equal

of one scan/s. Other analyzers, such as quadrupole ion traps (QIT), possess higher acquisition rates. Thus, many researchers operating hybrid QIT-FT MS instruments utilize the FT analyzer only for MS scans. However, the analytical advantages of high-resolution and mass accuracy MS/MS spectra are compelling, especially for reliable peptide identification⁴, unbiased PTM determination⁶ and *de novo* sequencing.⁷

One way to improve the acquisition rate during FTMS analysis of MS/MS spectra is to employ spectral multiplexing. Several methods of multiplexing MS/MS spectra have been described in the literature, with most methods utilizing physical multiplexing, *i.e.* several precursors are fragmented simultaneously. One approach developed by McLafferty and co-workers⁸ aimed to increase signal-to-noise ratio by repeating MS/MS experiments n times on appropriate subsets of n precursor ions with subsequent assignment of product ions to the corresponding precursor ions. Each time an acquisition was made, $n-1$ precursor ions were fragmented simultaneously, and each time a new n -th precursor was left intact. A Hadamard transform (HT) was then used to deconvolute the resulting data and thus determine MS/MS spectra for each of the precursor ions.

Another multiplexed MS/MS approach which aimed at increasing signal-to-noise relied on FT techniques similar to those used in 2D NMR experiments. In this method, precursor ions were excited into their cyclotron orbits by an RF excitation pulse, which was then followed by an RF de-excitation pulse after a specific delay time (td). This de-excitation pulse changed the abundance of specific precursor ions according to the phase difference, which is dependent on the precursor ion m/z ratio. The product ion spectra were then recorded as a function of td , and the FT of this function allowed the product ions to be related to their corresponding precursor ions by their abundance changes. Two variations to this approach have been described, one by Bodenhausen and Gaumann⁹ and another by Marshall and co-workers¹⁰.

Both the HT and 2D approaches to multiplexing on a FTICR require acquisition of n MS/MS spectra to obtain the product ion spectra of n precursor ions. Thus, these methods provide signal-to-noise ratio enhancements but no increase in the analysis speed. A third approach developed by Masselon *et al.*¹¹ allowed multiplexed MS/MS to be performed in a single spectral acquisition. This approach relies on database searching and the high mass accuracy and resolution capabilities of FT-ICR. Fragmentation products of n precursor ions were detected in one mass spectrum. The precursor and fragment m/z values were then compared to hypothetical MS/MS spectra of peptides from a protein database. The high mass accuracy of the FTICR mass spectrometer limited the number of potential candidate peptides. Ultimately, assignment of the product ions to their respective precursor ions was accomplished by matching the peptide mass tag derived from experimental data to candidate peptide sequence. The efficiency of this approach with a mainstream search engine has not been reported, and therefore the efficacy of this technique remains somewhat ambiguous. Our experiments with physical multiplexing of high-resolution mass spectra and matching the resultant spectrum against the database showed a dramatic drop in matching efficiency as the multiplexing factor n grows.

A multiplexing method for performing MS/MS on n peptide ions simultaneously in a quadrupole ion trap mass spectrometer (QIT) was developed by Wilson *et al.*¹² This method takes advantage of the inherent mass bias associated with ion accumulation in the QIT to encode the intensity of precursor ions in a way that allows the corresponding product ions to be identified. The intensity encoding scheme utilizes the Gaussian distributions that characterize the relationship between ion intensities and RF trapping voltages during ion accumulation. This approach uses two arbitrary waveforms, one for isolation and one for dissociation, to gather product ion spectra from n precursor ions with as few as two product

ion spectra. This approach is not suitable for high-resolution mass spectra that are not acquired using an QIT.

Another approach to multiplexing was devised by Waters and termed MS^E.¹³ All ions eluting to the instrument were fragmented, with the resulting fragments grouped together according to shared elution profiles. The utility of this method is limited when there are many co-eluting molecular species, a common occurrence in the analysis of complex protein mixtures, *e.g.* full or partially separated proteome digests.

Finally, direct database submission of multiplexed mass spectra have also been discussed in literature.^{14–15} Such methods usually rely strongly on the mass accuracy of fragment ions and generally suffer from interference between true MS/MS fragments of a given precursor and fragments of other molecules present in the same precursor selection window.

Here we report on a novel multiplexing approach that is free from the above shortcomings. The method utilizes physical multiplexing of n spectra, followed by software-based deconvolution, resulting in n individual, database searchable MS/MS datasets. The deconvolution procedure, the key element of the approach, relies on accurate measurement of fragment ions, and the fact that the charge separation reaction, which produces complementary ionic products, is the most frequent outcome of low-energy collisional dissociation of multiply-protonated tryptic peptides.

We demonstrate, using the SwedCAD database of 15,000 high-resolution collision-dissociation mass spectra of doubly charged precursors¹⁶, that this approach has solid theoretical basis, and that the efficiency of database search using the extracted individual MS/MS spectra exceeds 90%. Furthermore, practical implementation of the method on an LTQ Orbitrap (Thermo Fisher Scientific, Bremen, Germany) improves the number of identified peptides and proteins by more than a factor of two compared to conventional FTMS MS/MS scans.

Experimental

Multiplexing

The method is intended for hybrid FT MS instruments, in which the possibility exists to accumulate fragment ions from several independent MS/MS events in a separate trap, *e.g.*, the C-trap of the LTQ Orbitrap. The analysis cycle (Figure 1) consists of: (1) acquisition of a high-resolution “survey” MS spectrum followed by identification of m/z and z values of n multiply-protonated precursors. Such identification is usually performed based on a short (64 ms or shorter) fraction of the total transient, and therefore subsequent MS/MS events can proceed while the rest of the transient (lasting 0.5–2.0 s) is being acquired. (2) The identified precursors are isolated and fragmented individually in the ion trap, with the product ions of different precursors being mixed together and stored in another ion trap (*i.e.* C-trap). (3) After completing the last MS/MS event, the fragment mix is injected into the FT analyzer and detected with high resolution. Therefore, the shortest cycle duration (disregarding the “overhead” instrumental time between different cycle stages) is equal to the duration of two high-resolution mass spectra, while the acquired information is equivalent to one MS and n MS/MS high-resolution spectra. This should provide an improvement in the acquisition speed by a factor of $k*(n+1)/2$ (where k is the efficiency of deconvolution) compared to a cycle with one survey and n individually acquired high-resolution MS/MS spectra. With $n = 4$ and $k=0.9$ (realistic values, see below), an improvement in speed of $0.9*5/2 = 2.3$ is achieved.

Deconvolution of the multiplexed MS/MS spectra into n virtual individual MS/MS datasets is performed by software developed in house which creates a text input file for Mascot in the .mgf format from the raw spectral data. First, all the ions in the multiplexed MS/MS spectra are de-charged and de-isotoped². This yields a list of neutral masses and abundances. In cases when the charge state cannot be established because of the lack of isotopic peaks, all possible charge states, from 1 to z (z being the charge state of the precursor) are assigned to the single peak present. After de-charging and de-isotoping, complementary fragment pairs are derived for each precursor molecular mass. To this end, the mass-sorted list of neutral fragments is checked for masses which when summed with a larger mass from the same list yield a value equal to the precursor mass within given mass accuracy, typically ± 20 mDa.

Simulation of multiplexing and deconvolution

Figure 2 shows the workflow for an *in silico* experiment in which 1,000 high-resolution MS/MS spectra of doubly charged tryptic peptides were randomly selected from the online SwedCAD¹² database. These datasets were submitted to Mascot search engine, where they received identifications. The Mowse score (M-score)¹⁷ distribution for these identifications is shown in Figure 3, bottom panel. Totally, 980 peptide identifications were above the threshold for a confident identification. The remaining 20 mass spectra (2%) were not identified mainly because the protein database has changed after the identified spectra were put in the MS/MS database, but also because the original identification was performed using both CAD and ECD data for the same peptide⁴. In parallel, groups of four spectra were mixed together to produce 250 convolved datasets that were supplemented by 250 sets each comprising four precursor masses. These convolved datasets were submitted to Mascot using 1,000 precursor masses (each convolved dataset was submitted four times, each time with a new precursor mass). The distribution of the obtained M-scores is shown in the top panel of Figure 3 (average M-score 17.5). Overall, 551 above-threshold identifications were obtained, or 61% of the original number.

Using the above deconvolution procedure described above, the 250 convolved sets were deconvolved back into 1000 individual datasets which were also submitted to Mascot. The M-score distribution of the deconvolved dataset, shown in the top panel of Figure 2, gave the average M-score 43.7, which is not significantly different from the value 44.5 (the value obtained from the original data and displayed on the bottom panel). In total, 899 peptides were identified after deconvolution, giving the efficiency $k \approx 90\%$. Sequences of only two identified peptides after deconvolution did not coincide with the sequence identified before deconvolution, corresponding to $<0.22\%$ disagreement rate with the original Mascot identifications. This value compared favorably with a typical for many proteomics studies $\approx 1\%$ false discovery rate.

Therefore, utilization of the deconvolution procedure produced superior results as compared to a direct search of convoluted data. Removal of the fragments of the best Mascot-assigned peptide from the 4-convoluted dataset, and subsequent resubmission of the remaining dataset¹⁵ is still inferior to the deconvolution procedure, since the average M-score of the best assignment out of four for each of the 250 convoluted datasets was ≈ 35 , far below the average value obtained with the deconvolution procedure.

Figure 4 shows the comparison of the average Mascot scores in the deconvolution procedure and direct submission of convoluted spectra as a function of the number n of multiplexed spectra. The slight increase in value for the deconvolved spectra, as compared to direct submission at $n=1$, is due to the improved dataset quality due to the removal of many unrelated peaks. Overall, the results of direct submission of the convoluted spectra rapidly degrade with n , while the fall-off for deconvoluted spectra is much less steep.

The role of mass accuracy

Figure 5 shows the true number of the complementary pairs present in all 1000 mass spectra versus the number of deconvolved complementary pairs. At ± 20 mDa mass window, *ca.* 6300 complementary pairs are found, of which *ca.* 4500 are true pairs. Obviously, the fact that 25% of all found complementary pairs were spurious matches did not affect Mascot search too badly, but it follows from the trends in Figure 5 that an improvement in mass accuracy in the MS/MS datasets would further increase the efficiency of the deconvolution method.

Experimental verification

Cell Culture and Protein Harvesting—Wild-type *Saccharomyces cerevisiae* were grown and proteins harvested as previously described.¹⁸ Briefly, yeast was grown in rich media to an OD600 of 0.97, spun down, washed with sterile water, and pelleted *via* centrifugation at 10,000 *g* for 5 min. The cell pellet was added to a volume of lysis buffer containing protease inhibitors. The sample was French-pressed three times and centrifuged for 15 min at 30,000 *g* at 4 °C.

Digestion—To reduce and alkylate cysteine residues, protein was incubated in 2.5 mM DTT for 25 min at 60 °C followed by incubation in 7 mM iodoacetamide in the dark at room temperature for 30 min. Alkylation was capped by incubation in 2.5 mM DTT for 15 min at room temperature. The samples were digested with Trypsin (Promega, Madison, WI), desalted, and the eluent lyophilized following established protocol.

LC-MS/MS Analysis—The LTQ Orbitrap firmware was modified to allow implementation of the *n*-multiplexing experiment. The scan cycle began with a high-resolution ($R=100,000$) MS¹ followed by *n* MS/MS events in the LTQ on the most abundant precursor ions with $z \geq 2$. Precursors were subject to dynamic exclusion for 45 s, with an exclusion window of -0.5 Th to $+1.5$ Th. Precursors with unassigned charge states or charge states of one were also excluded. Following each MS/MS event, the product ion population was sent to the C-trap for storage. After all *n* fragment populations were accumulated in the C-trap, the summed ion population was injected into the Orbitrap analyzer for detection ($R=60,000$). The firmware recorded accurate monoisotopic m/z values of all *n* precursors in the scan header of each multiplexed MS/MS event.

As a control, an LC-MS/MS experiment was performed on the same sample using a standard data-dependent FTMS method. Briefly, following a high-resolution survey scan, the *n* most intense precursors were interrogated. Precursors were subjected to the same charge state and dynamic exclusion settings as the earlier experiment. Each MS/MS event consisted of precursor activation followed by m/z analysis in the Orbitrap.

Spectral File Generation and Database Searching—All .raw files resulting from MS/MS analysis were transformed to .mgf files using in-house software as described above. Mascot Generic Format files corresponding to control runs were created using a multiplexing *n* of 1, whereas those created from experimental analyses used *n* corresponding to the level of multiplexing (3 or 4). The m/z tolerance for fragment ions was set to 0.01 Da and no intensity cut-off was implemented. Database search was performed against a concatenated forward and reversed *Saccharomyces cerevisiae* (*Saccharomyces* Genome Database; version modified Feb. 3, 2011; concatenated version generated in-house) using MASCOT version 2.3 (Matrix Science, London, UK). The search was set for tryptic peptides, using a peptide tolerance of 5ppm with an MS/MS tolerance of 0.005 Da. A fixed modification of carbamidomethylation and a variable modification of Methionine oxidation were used, with a maximum of 2 missed cleavages.

In a series of separate experiments (data not shown), the efficiency of multiplexing, defined as the percent retention of ions from the first activation retained in the c-trap through the n th activation, was found to have a dependence upon both the number of multiplexed events (n) and the number of ions activated in each MS/MS event (automatic gain control, AGC). To investigate these variables, a series of experiments comparing multiplexed and the control analyses for different AGC target values and number of multiplexed MS/MS spectra was performed. The results are summarized in Table 1. For each comparison, multiplexing yielded 2–3 times more identified peptides (with M-score above the threshold value of 20) and proteins compared to the control experiment. At the same time, the size of the .mgf file (Mascot search file) was reduced by 67–75%, and the search time was reduced by *ca.* 25%. These results confirm the theoretical expectations of the efficiency of deconvolution procedure and provides basis for implementation of the deconvolution procedure in routine proteomics applications.

Conclusions

Here we demonstrated that an entirely FTMS-based method of MS/MS spectra multiplexing provides a 2.5–3.0 fold greater overall throughput than a typical FTMS method, with a negligible loss of information. The method is tested both *in silico* as well as experimentally, and it showed the expected performance. We hope that this convolution-deconvolution scheme will be widely used in practice for increasing the throughput of proteomics experiments.

Acknowledgments

The authors are grateful to Alexander Makarov and Stevan Horning of Thermo Fisher Scientific for fruitful discussions. This work was supported by the grants to RZ (VR 621-2008-10; EU project PredictAD) and JC (RO1 GM080148). DMG is grateful for support from a Wenner-Gren post-doctoral fellowship.

References

1. Zubarev RA, Mann M. On the proper use of mass accuracy in proteomics. *Mol Cell Proteomics*. 2007; 6:377–381. [PubMed: 17164402]
2. Bruce J, Anderson GA, Wen J, Harkewicz R, Smith RD. High Mass Measurement Accuracy for 100% Sequence Coverage of Enzymatically Digested Bovine Serum Albumin From an ESI-FTICR Mass Spectrum. *Anal Chem*. 1999; 71:2595–2599. [PubMed: 10424157]
3. Palmblad M, Wetterhall M, Markides K, Håkansson P, Bergquist J. Analysis of enzymatically digested proteins and protein mixtures using a 9.4 Tesla Fourier transform mass spectrometer. *Rapid Commun Mass Spectrom*. 2000; 12:1029–1034. [PubMed: 10861983]
4. Nielsen ML, Savitski MM, Zubarev RA. Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry. *Mol Cell Proteomics*. 2005; 6:835–845. [PubMed: 15772112]
5. Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom Rev*. 1998; 17:1–35. [PubMed: 9768511]
6. Savitski MM, Nielsen ML, Zubarev RA. ModifiComb: A Database-independent Proteomics Tool for Detecting Known Post-translational Modifications, Finding Novel Types of Modifications and Fingerprinting Complex Protein Mixtures. *Mol Cell Proteomics*. 2006; 5:935–948. [PubMed: 16439352]
7. Savitski MM, Nielsen ML, Kjeldsen F, Zubarev RA. Proteomics-grade *de novo* sequencing approach. *J Proteome Res*. 2005; 4:2348–2354. [PubMed: 16335984]
8. Williams ER, Loh SY, McLafferty FW, Cody RB. *Anal Chem*. 1990; 62:698–703. [PubMed: 2327586]
9. Pfandler P, Bodenhausen G, Rapin J, Walser ME, Gaumann T. J 2D FT/ICR/MS. *Am Chem Soc*. 1988; 110:5625–5628.

10. Ross CW III, Guan S, Grosshans PB, Ricca TL, Marshall AG. J Am Chem Soc. 1993; 115:7854–7861.
11. Masselon C, Anderson GA, Harkewicz R, Bruce JE, Pasa-Tolic L, Smith RD. Accurate mass multiplexed tandem mass spectrometry for high-throughput polypeptide identification from mixtures. Anal Chem. 2000; 72:1918–1924. [PubMed: 10784162]
12. Wilson J, Vachet RW. Multiplexed MS/MS in a Quadrupole Ion Trap Mass Spectrometer. Anal Chem. 2004; 76:7346–7353. [PubMed: 15595878]
13. Geromanos SJ. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics. 2009; 9:1683–1695. [PubMed: 19294628]
14. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat Biotechn. 2001; 19:952–957.
15. Venable JD, Xu T, Cociorva D, Yates JR III. Cross-Correlation Algorithm for Calculation of Peptide Molecular Weight from Tandem Mass Spectra. Anal Chem. 2006; 78:1921–1929. [PubMed: 16536429]
16. Fälvh M, Savitski MM, Nielsen ML, Kjeldsen F, Andren PE, Zubarev RA. SwedCADa Database of Annotated High-Mass Accuracy MS/MS Spectra of Tryptic Peptides. J Prot Res. 2007; 6:4063–4067.
17. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–3567. [PubMed: 10612281]
18. Swaney DL, McAlister GC, Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. Nat Methods. 2008; 5(11):959–964. [PubMed: 18931669]

1. Survey MS

2. n MS/MS

3. Multiplexed MS

Figure 1.

Workflow of the proposed multiplexing experiment: 1) high-resolution survey spectrum is taken and n multiply-charged precursors are identified by their m/z and $z > 1$ values; 2) n CAD MS/MS experiments are performed in the LTQ with the fragment ions stored in the C-trap; 3) a single high-resolution MS detection is performed of the multiplexed MS/MS experiments. This experiment is to be followed by deconvolution of n individual MS/MS datasets.

Figure 2.

Workflow for an *in silico* 4-plexing experiment in which 1000 high-resolution MS/MS spectra of doubly charged tryptic peptides were randomly selected from the SwedCAD database.

Figure 3.

The Mowse score (M-score) distributions for the identifications obtained in *in silico* 4-plexing experiment outlined in Figure 2. Black columns: upon deconvolution. Red columns: top panel - without deconvolution; bottom panel – the original MS/MS datasets.

Figure 4. Comparison of the average M-scores obtained after complementary-pair deconvolution (circles) and direct Mascot search without deconvolution (squares) for *in silico* n -plexing experiments with different n .

Figure 5. Number of complementary mass pairs extracted by deconvolution procedure in *in silico* 4-plexing experiment as a function of the mass window for fragment masses.

Table 1

Comparison of the result of multiplexing experiment with n merged MS/MS events, and control experiments with $n = 1$.

n	AGC Target Value	Analysis	Total IDs	IDs; 1% FDR	Exp./Control (Total ID's)	Exp./Control (1% FDR)
1	1×10^4	Control	562	536	1	1
3	1×10^4	Experiment	1404	1057	2.50	1.97
1	5×10^4	Control	723	563	1	1
3	5×10^4	Experiment	1572	1166	2.17	2.07
1	1×10^4	Control	612	487	1	1
4	1×10^4	Experiment	1730	1127	2.83	2.31