Published in final edited form as:

J Proteome Res. 2009 June ; 8(6): 2733–2739. doi:10.1021/pr801028b.

Expediting the Development of Targeted SRM Assays: Using Data from Shotgun Proteomics to Automate Method Development

Amol Prakash¹, Daniela M. Tomazela², Barbara Frewen², Brendan MacLean², Gennifer Merrihew², Scott Peterman³, and Michael J. MacCoss²

¹Thermo Fisher Scientific, Biomarker Research Initiatives in Mass Spectrometry, Cambridge, MA U.S.A.

²Department of Genome Sciences, University of Washington, Seattle, WA, USA

³Thermo Fisher Scientific, Somerset, NJ, USA

Abstract

Selected reaction monitoring (SRM) is a powerful tandem mass spectrometry method that can be used to monitor target peptides within a complex protein digest. The specificity and sensitivity of the approach, as well as its capability to multiplex the measurement of many analytes in parallel, has made it a technology of particular promise for hypothesis driven proteomics. An underappreciated step in the development of an assay to measure many peptides in parallel is the time and effort necessary to establish a usable assay. Here we report the use of shotgun proteomics data to expedite the selection of SRM transitions for target peptides of interest. The use of tandem mass spectrometry data acquired on an LTQ ion trap mass spectrometer can accurately predict which fragment ions will produce the greatest signal in an SRM assay using a triple quadrupole mass spectrometer. Furthermore, we present a scoring routine that can compare the targeted SRM chromatogram data with an MS/MS spectrum acquired by data-dependent acquisition and stored in a library. This scoring routine is invaluable in determining which signal in the chromatogram from a complex mixture best represents the target peptide. These algorithmic developments have been implemented in a software package that is available from the authors upon request.

Introduction

There has been a recent trend in proteomics towards the development and application of technologies for the targeted analysis of proteins within complex mixtures. Numerous derivations of targeted mass spectrometry using the specific acquisition of tandem mass spectra of peptides predicted *in silico* have been reported¹⁻³ and more recently these methods have been based on the use of selected reaction monitoring (SRM) on triple quadrupole mass spectrometers⁴⁻⁷. The concept of monitoring specific peptides from proteins of interest is well established. These methods have high specificity and sensitivity within a complex mixture and, thus, can be performed in a fraction of the instrument time relative to discovery based methods. Ultimately, targeted studies are intended to complement discovery based analysis and facilitate hypothesis-driven proteomic experiments.

Using the selectivity of multiple stages of mass selection of a tandem mass spectrometer, these targeted SRM assays are the mass spectrometry equivalent of a Western blot¹. This

Address Correspondence to: Michael J. MacCoss, Ph.D., Department of Genome Sciences, University of Washington, Foege Building, Box 355065, Seattle, WA 98195-5065, maccoss@u.washington.edu, Voice: (206) 616-7451, Fax: (206) 685-7301, Scott Peterman, Ph.D., Thermo Fisher Scientific, 265 Davidson Ave., Suite 101, Somerset, NJ 08873, scott.peterman@thermofisher.com, Voice: 732-627-0220.

combination of precursor and product ion m/z is extremely selective and referred to as an SRM transition. An advantage of using targeted mass spectrometry based assay over a traditional Western blot is that it does not rely on the creation of any immunoaffinity reagent. Thus, targeted SRM assays can be multiplexed in quantitative assays.

An underappreciated challenge of performing these targeted protein analyses is the amount of time spent preparing methods to measure specific transitions for specific peptides^{5,7} that are both selective and sensitive. Because significant amounts of time and resources can be spent to produce quantitative internal standards such as synthetic peptides⁸ or recombinant protein sequences^{9,10}, early development steps are often performed without generating these reagents. The goal then is to find peptides that are specific to the target protein and will produce a strong signal in the context of the sample matrix without having a standard of that peptide. This challenge is exacerbated by a lack of informatics tools to aide in the experiment design process. One of the most complicated steps in establishing these methods is determining which product ions provide an intense signal for a peptide of interest. The product ions monitored by SRM have historically been predicted using heuristics⁵ or combined with a refinement step after an initial experiment.

Empirically we have observed that the product ion spectra on an LTQ or LTQ hybrid mass spectrometer can be a strong predictor of which fragment ions are most intense in a triple quadrupole mass spectrometer. This observation is not entirely surprising since Yates and colleagues demonstrated that tandem mass spectra of peptides acquired on a triple quadrupole mass spectrometer could be identified by searching the spectra against a library of ion trap mass spectra¹¹. Thus, because large amounts of shotgun proteomics data have been acquired using LTQ based mass spectrometers and are organized into libraries or databases¹²⁻¹⁶, we can use these spectra to determine which product ions will provide the most intense signals in the tandem mass spectrum.

Here we describe a method of using MS/MS spectra of peptides from an LTQ linear ion trap mass spectrometer to automatically select which transitions to monitor on a TSQ Quantum triple quadrupole mass spectrometer. We demonstrate that our selection method outperforms transitions chosen by heuristics as measured by total signal intensity and signal-to-noise. We also demonstrate that by comparing the relative signal intensities of the monitored transitions to the intensities in library spectra, we can determine specifically which signal in the chromatogram is derived from the target peptide of interest.

Methods

Materials

C. elegans and bacterial strains were obtained from the *Caenorhabditis* Genetics Center at the University of Minnesota. A yeast 5-protein mixture was purchased from Michrom Bioresources (Auburn, CA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless specified otherwise.

Preparation of a *C. elegans* Protein Digest

C. elegans (N2 strain) were grown on enriched peptone plates seeded with the OP50 strain of *E. coli* at 20°C. A mixed stage culture of worms was washed off plates with M9 buffer (22 mM KH₂PO₄, 22 mM NA₂HPO₄, 85 mM NaCl, 1 mM MgSO₄, all chemicals VWR, West Chester, PA) and sucrose floated to remove bacterial contamination. The worms were lysed in 50 mM ammonium bicarbonate (pH 7.8) using a small probe from Fisher Scientific (Sonic Dismembrator Model 100, Pittsburgh, PA) for 5 cycles of a 30 seconds of continuous pulse followed by 30 seconds of ice incubation. The lysate was then centrifuged at 4000 rpm for 10

minutes at 4°C in an Eppendorf 5417R microcentrifuge (Westbury, NY) to remove cell debris. A second centrifugation at 14000 rpm for 10 minutes at 4°C was performed to separate the soluble lysate from the insoluble lysate. Protein concentration of the soluble lysate was determined using the BCA Protein Assay (Rockford, IL). Protein samples were denatured using 0.1% RapiGest SF (Waters Corporation, Milford, MA) in 50 mM ammonium bicarbonate pH 7.8. The samples were vortexed and boiled at 100°C for 5 minutes to increase denaturation. After samples were cooled, they were reduced with 5 mM DTT and alkylated with 15 mM IAA. Fractions were digested to peptides using trypsin (sequence grade, modified, Promega, Madison, WI) at a substrate to enzyme ratio of 100:1 for 4 hours at 37°C with shaking. The sample was then treated with 100 mM HCl to hydrolyse the RapiGest in the sample. After a final spin, the supernatant was stored at -20°C until ready for analysis by mass spectrometry.

Predicting Peptide SRM Transitions Using Previously Acquired Shotgun Proteomic Data

A library of 67,047 nonredundant MS/MS spectra acquired from a shotgun proteomics analysis of *C. elegans*¹⁷ was used by a software package called Pinpoint to develop methods for targeted SRM analysis on a triple quadrupole mass spectrometer. Similar capabilities have been added to an academic open source software called Skyline, and more information about it can be obtained from our website <http://proteome.gs.washington.edu/software/skyline>. The library spectrum can then be used to identify fragment ion peaks ranked by intensity, and enable the user to define how many product ions ($n = 3, 4, 5$, and 6 in this comparison) are required to provide a specific and selective measurement given the target sample.

The data were compared with fragment ions chosen using a heuristic similar to what has been described previously⁵. First, we chose the number of total ions (x) to be selected using the heuristic. Then we selected any y -ions N-terminal to proline (p) and the remaining ions ($x - p$) were chosen from y -ions starting with $m/z >$ precursor m/z . If the TSQ Quantum m/z limit (1,500 on the Ultra) was exceeded in adding the total number of transition product ions, then we added additional y -ions below the precursor m/z , preferentially adding ions of higher m/z first, until the total number of ions was reached. For the purpose of this study, all targeted peptides were monitored by the automatic selection of ions y_3 through y_{n-1} (where n is the largest y -type ion) for SRM experiments. The comparison of the intensity of the respective selection process was performed post data acquisition.

Microcapillary Liquid Chromatography Selected Reaction Monitoring Mass Spectrometry

For each experiment, 2 μ L (10 μ g) of the *C. elegans* digest was loaded onto an in-house packed capillary column with a $\sim 5 \mu$ m pulled tip, 30 cm length, and 0.075 mm internal diameter using a NanoAcquity UPLC (Waters, Milford). The column was packed with Jupiter Proteo (C12, 4 μ m particle size, 90 Å pore size) using an in-house constructed pressure bomb. A binary solvent system used reverse-phase buffers consisting of an A buffer containing 95% water, 4.9% acetonitrile, and 0.1% formic acid and a B buffer composed of 20% water, 79.9% acetonitrile and 0.1% formic acid. The UPLC was operated at a flow rate of 300 nL/min. Peptides were separated over a gradient from 2% to 35% B over 60 minutes.

Targeted peptide detection and verification experiments were performed with either a TSQ Quantum Ultra or TSQ Quantum Access (Thermo Fisher Scientific, San Jose, CA) equipped with a home built nanospray stage and source operated in positive-mode ESI. A spray voltage of +2400 V was used with a heated ion transfer tube setting of 235 °C for desolvation. SRM experiments were performed with both Q1 and Q3 resolution settings at unit resolution on both mass spectrometers (FWHM = 0.7 Da).

Targeted peptides were selected from fifteen proteins identified previously in *C. elegans* digests analyzed by shotgun proteomics. A total of 18 different injections of the *C. elegans*

digest were performed to monitor 127 peptides containing the y_3 to y_{n-1} transitions for each peptide. Each injection contained ca. a maximum of 60 SRM transitions acquired using a 20 msec dwell time setting. All collision energies were calculated using the formula

$$\text{Normalized Collision Energy} = (\text{parent } m/z) * 0.035 + 3V \quad (\text{Equation 1})$$

and used without further optimization. The q2 collision gas pressure was held constant at 1.5 mtorr of argon.

Library Searching with SRM Data

To compare the targeted SRM data generated from a triple quadrupole mass spectrometer with ion trap spectra of a peptide stored in a spectrum library, we developed a novel spectrum library searching method. This method evaluates whether the signals detected at any one point in time belong to the same molecular species that produced an individual ion trap mass spectrum. The peak area intensities for all SRM transitions that belonged to the same precursor ion m/z and at the same retention time were ranked in decreasing order of intensity. Then, the same product ions from the ion trap library spectrum were also ranked in decreasing order of intensity. Product ions monitored in the targeted SRM experiment that were not present in the library spectra, due to their intrinsic peptide fragmentation mechanism, were assigned a random intensity lower than the minimum intensity for the other measured transitions. Given both these rank order, we used a Costa-Soares rank correlation¹⁸ as a metric of similarity between the SRM data and the LTQ spectrum stored in the library for the target peptide. The Costa-Soares correlation was computed using the equation:

$$\rho = 1 - \frac{6 \sum (p_i - q_i)^2 (n+1 - p_i + n+1 - q_i)}{(n^2 + n)(n^2 - 1)} \quad (\text{Equation 2})$$

where ρ is the correlation coefficient (expected range of values are $1 \geq \rho \geq -1$), p_i and q_i are the ranks of product ion i in the LTQ spectrum and the signal intensities from the SRM data respectively, and n is the total number of product ions under consideration. To provide an estimate of the null distribution, the correlation coefficient was compared with library spectra whose rank intensities were shuffled and then rescored. The null distribution was used to compute a p-value using standard statistical methods.

Results

Comparison of product ion spectra between an LTQ ion trap and a TSQ Quantum triple quadrupole mass spectrometer

The use of previously acquired discovery based product ion spectra has been used to help to determine which peptides would make good proteotypic peptide candidates^{5;16;19-21}. While these initial studies used these peptide identifications to determine which peptides might be a useful proxy of the protein, the tandem mass spectrum itself was not used in the building of the final targeted SRM method. Our approach is to use not only the previously identified peptide sequence information from libraries but also the MS/MS spectrum for the selection of high intensity transitions. Previous experiments were performed on complex enzymatic digests of *C. elegans* and analyzed on an LTQ mass spectrometer, resulting in the acquisition of >6,000,000 data-dependent MS/MS spectra from ~30 different biochemical fractions. These MS/MS spectra resulted in >67,000 unique peptides mapping to 6,779 independent gene loci¹⁷. Spectra from this shotgun analysis were then stored and filtered in a BiblioSpec library

as described previously^{12;22}. This spectrum library can be downloaded on-line at <http://proteome.gs.washington.edu/software/bibliospec/documentation/libs.html> and the programs Skyline and Pinpoint, used to read these libraries and to perform the analyses described in this work can be obtained by request from the authors.

To illustrate the similarity of the data stored in our libraries and the relative intensities of the product ions on the TSQ Quantum, consider the tryptic peptide ASGAFTGENSVDDQIK (+2). This peptide arises from the digestion of the *S. cerevisiae* protein triose phosphate isomerase and its collision induced dissociation (CID) spectrum is shown in Figure 1. In this figure, the LTQ-CID spectrum of the peptide shows that the most abundant product ions are the y-type ions: y₉, y₁₀, and y₄. The SRM data acquired on the TSQ Quantum for the same peptide is also shown in Figure 1 and its product ions relative abundances are represented by the green stars. Note that the most abundant transitions in the TSQ data correspond to the ions y₉, y₁₀, and y₄. The Costa-Soares rank correlation for the peptide ASGAFTGENSVDDQIK was calculated to be 0.92 with a p-value < 0.001, indicating an excellent match between the library and SRM data. Of primary significance is that the top three product ions in the LTQ spectrum matched the top three product ions from the SRM experiment and the fourth most abundant LTQ product ion was the sixth most abundant in the SRM experiment. If only the top four product ions were monitored in the targeted SRM experiment, a Costa-Soares' rank correlation coefficient of 0.6 would have been computed between the SRM data and the LTQ spectrum.

Comparison of library-selected transitions versus heuristic-based transitions

To determine if MS/MS spectra produced by frequency based activation on an LTQ linear ion trap mass spectrometer could be used to predict transitions on a triple quadrupole mass spectrometer we performed targeted analysis on the y₃ to y_{n-1} transitions for 127 peptides using their doubly charged precursor ions. To determine the effectiveness of library-based selections for finding the most abundant product ions, we processed the SRM data post-acquisition to compare the difference between library- and heuristic-selected peptide transitions. We used heuristics as described in the methods. These heuristics are similar to those described previously⁵. The SRM intensity of these ions were then summed and recorded. For the library-based selection, the three y-type ions that had the most abundant LTQ-CID product ion intensities were chosen and their SRM response summed. The 127 peptides used in this work as well as their heuristic- and library-selected peak areas calculated during the post-acquisition processing can be found in Supplementary Table 1.

To determine the benefit of using library-based selection of SRM transitions, a comparison was made between peak areas using the library- and heuristic-based selection of transitions from the same respective peptides in a *C. elegans* lysate. Because the collected data set included the y₃ to y_{n-1} transitions per peptide, reconstructed chromatograms were produced using a subset of the collected transitions. The summed area for the three most intense library-based SRM transitions was calculated and compared to that using the three heuristic transitions. An example of the improvement was found for the peptide FLAAGLNADNLVATYQK (rpl-5 F54C9.5) (Supplementary Figure 1). In this example, the precursor is the doubly-charged ion at *m/z* 904.9 and, using the heuristic selection criteria proposed by Anderson and Hunter⁵, the product ions selected were y₈ (*m/z* 936.5), y₉ (*m/z* 1051.5) and y₁₀ (*m/z* 1122.6). For the library-based selections, the ions y₄ (*m/z* 539.3), y₅ (*m/z* 610.3) and y₁₁ (*m/z* 1236.6) were chosen and their sum represents a 2.8-fold increase in intensity over what was obtained using transitions estimated using heuristics.

This same comparison between transitions selected using heuristics and those selected using the respective LTQ spectrum was performed for a larger set of targeted peptides from *C. elegans*. A criterion for the selection of the 127 peptides was that the peptide must be present in our LTQ spectrum library and must be readily detectable in an unfractionated protein digest.

Because we monitored all y-ion product ions that ranged from y_3 to y_{n-1} , the covarying transitions were manually evaluated to assess if the peptide was detected. Figure 2A shows the ratio of the library transitions area/heuristic transitions area calculated for the selected peptides using three transitions. The mean ratio is 1.90 and the median is 1.20. Thus, most peptides show an increase in signal intensity when transitions were chosen using the LTQ spectrum library. In fact, the library-based SRM transition selection showed a greater area calculation for 75.6% of the peptides and no change for 8.7%. Only 15.7% of the peptides had greater signal intensity using the transitions chosen based on the heuristics as opposed to the LTQ spectrum library. Of these 20 peptides that had greater intensity with the heuristic, 15 of them were within 20% of the signal intensity obtained from selecting ions by the spectrum library. Thus for these 127 peptides, 84.3% of the time it was beneficial or equivalent to use the spectrum library for choosing the SRM transitions.

It is important to point out that the heuristic itself does not do that poorly. Of the 381 transitions considered, 203 (53.3%) are shared between the heuristic and the library. Likewise, the mode in Figure 3A is 1.1 – indicating that the transitions chosen using the heuristic method are often equivalent to the transitions chosen using the spectral library information.

Another metric for reporting the use of the LTQ library for the selection of transitions on the TSQ is shown in Figure 2B, where we compare the fraction of summed signal from the n most intense transitions that are accounted for by the respective method of selecting the product ions. A score of 100% means that the selected transitions were the n most intense transitions monitored. In this comparison we compared three different heuristics. Heuristic 1 is implemented exactly as described in the methods and begins choosing y-ions at $m/z >$ the precursor m/z ; whereas, Heuristic 2 and Heuristic 3 began at the first and second y-ion below the precursor m/z , respectively. By using the LTQ spectrum library, >90% of the maximum signal intensity can be obtained regardless of the number of transitions selected. Interestingly, in the absence of an LTQ spectrum library, choosing y-ions using Heuristic 1 represents an acceptable alternative.

Costa-Soares coefficients and the measurement of p-values

We used the Costa-Soares rank correlation as a metric of similarity between the SRM peak intensities and the LTQ ion trap mass spectra. The effect of the number of SRM transitions used in computing the similarity metric was investigated. This comparison was done by choosing the most intense SRM peaks matched with their corresponding LTQ peaks. The distribution of the Costa Soares score for 10 peptides (each having at least 11 amino acids) as we vary the number of SRM transitions from 4 to 8 can be found in Supplementary Figure 2. By increasing the number of transitions we are able to match the ranks between LTQ spectra and SRM transitions more strongly, and this provides more confidence to the identification.

To determine the likelihood that a given Costa Soares rank correlation could have been derived by chance, we calculated p-values for each peptide. Assuming a null hypothesis of uniform distribution of signal intensities, the p-values were calculated exhaustively, i.e., by taking every possible rank permutation and computing the corresponding Costa-Soares score. This null hypothesis is independent of the sample complexity, and just assumes a uniform distribution of signal intensities (and thereby, ranks) Then we calculated the number of randomly permuted rank permutations that score higher than experimentally measured Costa-Soares score to give a p-value measure. To evaluate the effect of the Costa-Soares score on the p-value these two scores were plotted relative to one another with varying number of measured SRM transitions in Figure 3. As expected, the p-value is improved (i.e. gets smaller) at the same Costa-Soares score when more transitions are used to evaluate the rank correlation. Thus, as expected the likelihood that a specific rank order of transitions could have occurred by chance decreases with increasing numbers of transitions.

Figure 4 plots the distribution of the p-values for 10 peptides as we vary the number of SRM transitions from 4 to 8 (same peptides shown in Supplementary Figure 2). As expected, higher numbers of transitions provide more confidence in the identification, as the p-value goes down. Interestingly, to obtain a significant match between the targeted SRM data and the LTQ library, Figure 4 indicates that we need to monitor >5 transitions per peptide. However, we would like to emphasize that this p-value is conservative because our p-value assumes that there is signal in every transition at a given point in time – a very rare event when 5 or more transitions are monitored. Instead of computing rank correlations using only transitions that had a detectable chromatographic peak, we assign a random intensity less than the smallest detected peak to the transitions without a peak. By forcing a random low intensity value into transitions without peaks, we penalize retention times where there are few covarying peaks.

Because SRM signals at multiple different retention times are compared to the LTQ spectrum, we also needed to perform a multiple hypothesis correction. This multiple hypothesis correction also corrects the p-value for the sample complexity, as there is a higher likelihood of observing measurable signals in multiple transitions at multiple retention times in a complex mixture. A standard Bonferroni correction was used to adjust the p-values based on the number of instances in the LC time gradient that were compared with target LTQ spectrum.

Confirming the presence of peptide in a targeted triple quadrupole measurement assay using ion trap data

A library-based entry can also be used to help verify the retention time of a targeted peptide. The biological matrix provides challenges to identifying the correct retention time if one or more of the SRM transitions contain interference. The more complex the background matrix, the greater the probability that transitions will contain signal from random incorrect analyte present in the sample. An example of this is presented in Supplementary Figure 3 for the peptide QEYDESGPSIVHR (+2) (act-4) where the chromatogram for the summed response shows >11 different peaks spread over 60 minutes. Figure 5 shows the summed (top chromatogram in blue) and individual chromatograms for the y_3 through y_7 transitions. There are only four of the 11 peaks that contain measurable signal in each of the five transitions as indicated by the differently red rectangles, increasing the difficulty in determining the correct targeted peptide. In this case, the benefit of incorporating multiple SRM transitions per peptide to take advantage of the common retention time for all transitions would not be enough for the identification of the correct peptide. Despite measurable signal in multiple transitions, the relative abundance of the intensities is very different for the precursor > product ion pairs.

The product ion spectrum from the LTQ was used to verify the correct retention time of the target peptide. The profile of the product ion intensities from the CID spectrum was compared to each of the four targeted retention times for visual inspection and analysis using a Costa-Soares correlation. Figure 6 shows the four results. Clearly the match for the relative intensities at 21.2 minutes (red bars) is the closest visually and has a Costa-Soares rank coefficient (CS) of 0.8 compared to 0.03, 0.21, and 0.21 for the peaks at retention times 30.7, 36, and 41.2 min respectively. These Costa-Soares rank coefficients result in p-values of 0.01, 0.47, 0.30, and 0.30 for the signals at retention times 21.2, 30.2, 36, and 41.2 min respectively. Thus, the rank order of the peaks at retention time 21.2 min is the only peak that is statistically significant. Furthermore, even after performing multiple hypothesis testing correction the p-value ≤ 0.05 .

Discussion

The present work took advantage of an LTQ-based mass spectrometry peptide spectrum library to support building SRM methods using a TSQ Quantum mass spectrometer. The discovery data identified by shotgun proteomics, contains which peptides have been identified previously, the number of times they had been detected, the charge state they have been identified at, and

the relative abundance of the product ions in the MS/MS spectrum. These data can be used to expedite the generation of rational SRM methods and, to also analyze the data generated. The relative product ion abundances for the y-ion series of tryptic peptides are very comparable between the LTQ and TSQ platforms. Compared to the heuristic based approach for selecting product ions, SRM methods built choosing the most intense y-ions found in the LTQ spectrum of tryptic peptide had a higher overall performance and can be part of automated method development workflows.

The strong correlation of the fragmentation spectrum on an LTQ ion trap with a TSQ Quantum triple quadrupole might be counterintuitive. The LTQ uses frequency based activation where once a peptide bond is cleaved the resulting fragments are off-resonance from the excitation frequency and should not experience further activation. In contrast, the triple quadrupole uses traditional beam type low energy collision induced dissociation where multiple collisions are frequent and cleavage of multiple bonds are not unexpected. Furthermore, the LTQ uses helium as the collision gas and the TSQ uses argon. Nevertheless, the similarity of the product ion rank intensities in the y-ion series is significant between the two instruments for a majority of peptides.

This similarity does not mean that there are not differences. First, the presence of b-ions is underrepresented in the TSQ Quantum data from tryptic peptides. The reduced presence of b-ions is probably a result of multiple collisions and long q2 path length in the TSQ Quantum. Thus, the b-ion series continues to fragment to smaller product ions until it stops at the b2 ion – frequently one of the largest but least selective product ions in the spectrum. Another difference between the LTQ and TSQ spectra is the low m/z region. The LTQ, like all ion traps that use frequency based activation and mass selective instability to acquire the product ion spectra, has a low m/z cut-off. In contrast, the TSQ does not have the same low m/z cutoff in the product ion spectrum. For these reasons, all of the data presented here are using the y-ion series from the y_3 to the y_{n-1} .

One of the biggest benefits of a targeted assay on a triple quadrupole mass spectrometer is high throughput. Targeted proteomics has the promise of analyzing samples for multiple targets (proteins/peptides) in a short duration of time. Unfortunately, in the current scenario, building such a method for this workflow requires significant time and effort. Starting from a target list of protein candidates, multiple iterations and optimizations are required to select the best set of peptides and transitions, at times even synthesizing multiple candidate peptides in the process. By no means is this process automated and high throughput.

While the strategy that we suggest will work for most proteins, there will be cases that will be missed, e.g., when the targeted peptides are non-tryptic or have a very uncommon amino acid composition. In most cases the use of ion trap product ion spectra for peptides acquired in large unbiased discovery proteomics experiments can help automate the method development on a triple quadrupole – simplifying the process of generating a sensitive and selective assay. Even in the case when the protein is novel, and not seen in any discovery experiment before, it will be beneficial to analyze the proteins or synthesized peptides on the ion trap to guide the SRM method development.

In this paper, we present a novel methodology and software that can automate this process. The key contribution lies in using the ion trap spectral libraries that labs have been collected by shotgun proteomics to select the product ions to monitor in a SRM experiment on a triple quadrupole mass spectrometer. These libraries can guide the selection of peptides and transitions, and thus automate this process for proteins detected previously. This observation further shows the usefulness of high quality annotated data produced by every lab and re-

emphasized the benefits of maintaining such data sets in a structured way that can be shared between research laboratories in the proteomics community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support for this work was provided in part by National Institutes of Health grants P41 RR011823, R01 DK069386, and U24 CA126479.

References

1. Arnott D, Kishiyama A, Luis EA, Ludlum SG, Marsters JC Jr, Stults JT. *Mol Cell Proteomics* 2002;1:148–56. [PubMed: 12096133]
2. Chang EJ, Archambault V, McLachlin DT, Krutchinsky AN, Chait BT. *Anal Chem* 2004;76:4472–83. [PubMed: 15283590]
3. Jin M, Bateup H, Padovan JC, Greengard P, Nairn AC, Chait BT. *Anal Chem* 2005;77:7845–51. [PubMed: 16351129]
4. Barnidge DR, Goodmanson MK, Klee GG, Muddiman DC. *J Proteome Res* 2004;3:644–52. [PubMed: 15253448]
5. Anderson L, Hunter CL. *Mol Cell Proteomics* 2006;5:573–88. [PubMed: 16332733]
6. Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. *Mol Cell Proteomics* 2007;6:2212–29. [PubMed: 17939991]
7. Anderson L. *J Physiol* 2005;563:23–60. [PubMed: 15611012]
8. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. *Proc Natl Acad Sci USA* 2003;100:6940–45. [PubMed: 12771378]
9. Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. *Nat Protoc* 2006;1:1029–43. [PubMed: 17406340]
10. Rivers J, Simpson DM, Robertson DH, Gaskell SJ, Beynon RJ. *Mol Cell Proteomics* 2007;6:1416–27. [PubMed: 17510050]
11. Yates JR III, Morgan SF, Gatlin CL, Griffin PR, Eng JK. *Anal Chem* 1998;70:3557–67. [PubMed: 9737207]
12. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ. *Anal Chem* 2006;78:5678–84. [PubMed: 16906711]
13. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Edes J, Loevenich SN, Aebersold R. *Nucleic Acids Res* 2006;34:D655–D658. [PubMed: 16381952]
14. Prince JT, Carlson MW, Wang R, Lu P, Marcotte EM. *Nat Biotechnol* 2004;22:471–72. [PubMed: 15085804]
15. Craig R, Cortens JP, Beavis RC. *J Proteome Res* 2004;3:1234–42. [PubMed: 15595733]
16. Craig R, Cortens JP, Beavis RC. *Rapid Commun Mass Spectrom* 2005;19:1844–50. [PubMed: 15945033]
17. Merrihew GE, Davis C, Ewing B, Williams G, Kall L, Frewen BE, Noble WS, Green P, Thomas JH, MacCoss MJ. *Genome Res*. 2008
18. Pinto da Costa JF, Soares C. *Australian and New Zealand Journal of Statistics* 2005;47:515–29.
19. Kuster B, Schirle M, Mallick P, Aebersold R. *Nat Rev Mol Cell Biol* 2005;6:577–83. [PubMed: 15957003]
20. King NL, Deutsch EW, Ranish JA, Nesvizhskii AI, Edes JS, Mallick P, Eng J, Desiere F, Flory M, Martin DB, Kim B, Lee H, Raught B, Aebersold R. *Genome Biol* 2006;7:R106. [PubMed: 17101051]
21. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R. *Nat Biotechnol* 2007;25:125–31. [PubMed: 17195840]
22. Frewen B, MacCoss MJ. *Curr Protoc Bioinformatics* 2007;Chapter 13(Unit)

Figure 1.

LTQ-CID spectrum of the doubly charge peptide at m/z 762 from the peptide ASGAFTGENSVDQIK from the *S. cerevisiae* protein triose phosphate isomerase. The green stars represent the TSQ-CID product ion relative abundance for the same peptide measured by selected reaction monitoring.

Figure 2.

Effect of using an LTQ spectrum library in directing the selection of product ions. **A.** Plot of the ratio of the library transitions area/heuristic transitions area calculated for 129 *C. elegans* peptides. The mean ratio is 1.90 and the median is 1.20. Thus, most peptides show an increase in signal intensity when transitions were chosen using the LTQ spectrum library. The library-based SRM transition selection showed a greater area calculation for 75.6% of the peptides and no change for 8.7%. **B.** The fraction of summed signal from the *n* most intense transitions that are accounted for by the respective method of selecting the product ions. A score of 100% means that the selected transitions were the *n* most intense transitions monitored. Heuristic 1 is implemented exactly as described in the methods and begins choosing y-ions at $m/z >$ the precursor m/z ; whereas, Heuristic 2 and Heuristic 3 began at the first and second y-ion below the precursor m/z respectively.

Figure 3.
Plot of the distribution of the Costa-Soares score and the p-value for varying number of transitions.

Figure 4.

Effect of the number of SRM transitions on the match between the SRM data and the LTQ library spectrum. As the number of transitions increases, the likelihood that the Costa-Soares score could have been derived by random chance decreases.

Figure 5. TSQ-SRM chromatogram for the individual transitions 758>411, 524, 611, 708 and 765, corresponding to the formation of the y-ions y_3 , y_4 , y_5 , y_6 and y_7 , respectively. There are four different retention times where there are multiple precursor > product ion transition pairs at a single retention time, complicating the unambiguous determination of which retention time is representative of the target peptide.

Figure 6.

Comparison of the product ion intensities from the LTQ-CID spectrum (black bars) compared to each of the four targeted retention times: 21.2 (red), 30.7, 36 and 41.2. The Costa-Soares score (CS), p-value (PV), and Bonferroni corrected p-value (BPV) are listed to provide a measure of similarity between the SRM transition intensities and the LTQ library spectrum.