Journal of proteome research

Technical Note

pubs.acs.org/jpr

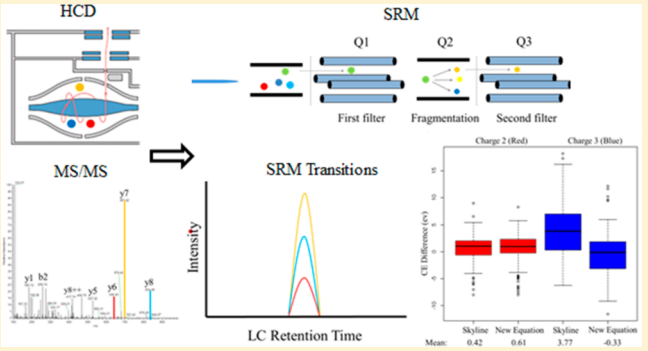
Expediting SRM Assay Development for Large-Scale Targeted Proteomics Experiments

Chaochao Wu,[†] Tujin Shi,[†] Joseph N. Brown,[†] Jintang He,[†] Yuqian Gao,[†] Thomas L. Fillmore,[‡] Anil K. Shukla,[†] Ronald J. Moore,[†] David G. Camp, II,[†] Karin D. Rodland,[†] Wei-Jun Qian,[†] Tao Liu,*[,†] and Richard D. Smith*[,†]

[†]Biological Sciences Division and [‡]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, United States

Supporting Information

ABSTRACT: Because of its high sensitivity and specificity, selected reaction monitoring (SRM)-based targeted proteomics has become increasingly popular for biological and translational applications. Selection of optimal transitions and optimization of collision energy (CE) are important assay development steps for achieving sensitive detection and accurate quantification; however, these steps can be labor-intensive, especially for large-scale applications. Herein, we explored several options for accelerating SRM assay development evaluated in the context of a relatively large set of 215 synthetic peptide targets. We first showed that HCD fragmentation is very similar to that of CID in triple quadrupole (QQQ) instrumentation and that by selection of the top 6 y fragment ions from HCD spectra, >86% of the top transitions optimized from direct infusion with QQQ instrumentation are covered. We also demonstrated that the CE calculated by existing prediction tools was less accurate for 3+ precursors and that a significant increase in intensity for transitions could be obtained using a new CE prediction equation constructed from the present experimental data. Overall, our study illustrated the feasibility of expediting the development of larger numbers of high-sensitivity SRM assays through automation of transition selection and accurate prediction of optimal CE to improve both SRM throughput and measurement quality.



KEYWORDS: *SRM, MRM, HCD, QQQ, transition selection, optimization, CE prediction, targeted quantification*

INTRODUCTION

Targeted measurements, using selected reaction monitoring (SRM; aka multiple reaction monitoring (MRM)) that is typically performed on triple quadruple (QQQ) instruments, have emerged as a powerful, multiplexed quantification technology for biomarker development and systems biology studies.[1−3] By the application of two stages of mass filters, specific to a target peptide and avoiding the constraints of an ion trapping step, SRM using QQQ instruments arguably provides the best overall combination of sensitivity and specificity presently achievable by mass spectrometry for proteomics.[4] Moreover, with prior knowledge of peptide liquid chromatography retention times, highly multiplexed analysis (e.g., hundreds of peptides in a single analysis) can be performed using a scheduled SRM strategy,[5] and this method is now widely used in biomedical and systems biology research, such as biomarker verification in biofluids,[6−8] protein post-translational modifications (PTMs),[9] and studies of signaling pathways.[10]

Despite the increasing popularity, an underappreciated cost of routine targeted proteomics experiments lies in the time and effort necessary to establish a sensitive SRM assay. In general, for SRM assay development the first step is the peptide transition selection, which can be carried out either by extracting information from an existing shotgun proteomics peptide library and data repository[11,12] or actual analysis of synthetic peptides on the QQQ instrument.[13] Most shotgun proteomics peptide libraries and data are typically constituted from ion trap collision induced dissociation (IT-CID) data[14−18] and thus are often used to facilitate SRM method development.[19] However, peptide fragmentation patterns from IT-CID (resonance excitation of a specified m/z window and involving many low-energy collisions) are generally significantly different from the peptide spectra obtained from QQQ (beam-type) CID, which uses ions that have significantly higher initial energies. While the higher energies are gradually dissipated by collisions, further fragmentation contributions are introduced due to additional fragmentation of fragments that are initially formed by the higher-energy collisions.[20] Recently, a

Received: May 20, 2014
Published: August 22, 2014

   dx.doi.org/10.1021/pr500500d | J. Proteome Res. 2014, 13, 4479−4487

fragmentation approach, termed higher-energy collision dissociation (HCD), has been made available with Orbitrap analyzers that provides higher mass resolution and mass accuracy fragmentation data.[21] Given the benefit of high mass accuracy and no low mass cutoff, shotgun proteomics data has been increasingly produced using HCD fragmentation.[22] It has also been pointed out that HCD MS/MS data share a higher degree of similarity with QQQ-based MS/MS data than with IT-CID data,[23] but it also has differences due to details of the collision conditions and different m/z biases of the mass analyzers employed. However, a detailed evaluation comparing the HCD-derived and QQQ-optimized results has not been reported for better understanding the feasibility of using HCD data for accelerating and automating SRM assay development.

In addition to transition selection, it is also important to optimize the collision energy (CE) for each transition to improve SRM sensitivity.[24] CE can be selected using existing empirical formulas (available within the widely used Skyline software[25]) or obtained for individual case using synthetic peptides and experimental optimization.[26] For large-scale accurate quantification of high- and moderate-abundance proteins, peptide transitions are often selected from either existing discovery data or by using CEs predicted by an empirical formula. By contrast, for quantification of extremely low-abundance proteins or protein PTMs in complex biological samples such as human blood plasma/serum,[6] highly sensitive SRM assays are required and typically achieved by optimization using synthetic peptides through direct infusion or LC−SRM analysis using the actual QQQ instrumentation. However, such QQQ-based peptide optimization is low-throughput and labor-intensive, and it greatly increases the cost and time required for the development of large-scale sensitive SRM assays.

Herein, we utilized a relatively large set of 215 synthetic peptides to explore the use of HCD MS/MS data to rapidly establish optimal SRM assays for large-scale SRM quantification. We first manually optimized each of 215 synthetic peptides with QQQ instruments by direct infusion (i.e., the conventional method), including selecting the most intense fragment ions and determining the optimal CE for each transition using QQQ MS/MS measurements. Using the QQQ MS/MS spectra and optimal CEs determined in this manner as the standards, we then systematically compared the HCD and CID MS/MS spectra for each peptide from LC−MS/MS analyses to the direct infusion MS/MS results. On the basis of the MS/MS comparisons and a statistical analysis, we also constructed equations for more accurately predicting transition CEs. Taken together, our results demonstrated that the availability of HCD spectra together with refined CE prediction can expedite the development of optimal SRM assays and facilitate large-scale targeted proteomics applications.

EXPERIMENTAL SECTION

Peptide Selection

All peptides were selected using the following criteria[12] (unless indicated otherwise): fully tryptic, less than 25 amino acids in length, no miscleavage (lysine and arginine before proline are acceptable), unique for target proteins, moderate hydrophobicity, and peptides with methionine are not preferred but were retained if there is no better choice. Peptides detected in our own previous shotgun analyses were preferred, followed by those reported in public data repositories such as PeptideAtlas and Global Proteome Machine (GPM). For target proteins without the existing MS/MS data, the surrogate peptides were selected on the basis of *in silico* digestion and then evaluated by using prediction tools such as CONSeQuence software.[27] Details of all of the peptides used in this study and their corresponding proteins are available in Supporting Information Table 1.

Sample Preparation

Crude synthetic peptides labeled with either heavy lysine ($[^{13}C_6,^{15}N_2]$-lysine) or arginine ($[^{13}C_6,^{15}N_4]$-arginine) were purchased from Thermo Scientific (San Jose, CA). In total, 215 crude heavy peptides were used for method development. A small subset of peptides (78) was first used to validate the previously reported conclusion that HCD is more similar to beam-type CID than CID[23] (analyzed by both TSQ (SRM) and Orbitrap (both CID and HCD)); another larger subset of peptides (137) was further analyzed to extensively compare how similar it is between HCD and beam-type CID (analyzed by TSQ (SRM) and Orbitrap (HCD only)), and new CE prediction equations were built based on this combined result from 215 peptides. We then further validated the effectiveness of new CE equations using an independent set of peptides (92) (analyzed by TSQ (SRM)) (see Supporting Information Table 1 for details). For direct infusion analysis, synthetic peptides were analyzed individually, and the nominal concentration of crude peptides was 200 fmol/μL in 0.1% formic acid and 30% acetonitrile. For LC−MS/MS analysis, all crude peptides were diluted and pooled together for one sample with a final nominal concentration of 50 fmol/μL in 0.1% formic acid.

Direct Infusion Experiment

Direct infusion was performed using a 250 μL syringe (Hamilton, Reno, NV) directly connected to a TSQ-Quantum QQQ mass spectrometer (Thermo Scientific, San Jose, CA) with a homemade spray tip[28] (20 μm i.d.). The flow rate of direct infusion was set at 0.7 μL/min. The QQQ was first operated in Q3 MS mode with unit resolution (peak width 0.7) for observation of potential precursors with the following parameters: spray voltage, 2400 V; capillary temperature, 325 °C; capillary offset, 35 V; tube lens offset, 185 V; and collision gas pressure, 1.5 mTorr. For each peptide with different charge states, the precursor with highest intensity was selected for peptide optimization. If multiple precursors with different charge states existed with similar intensities for a given peptide, then all were retained. The QQQ was then operated in MS/MS mode targeting the selected precursors for fragment information with a minimum of 10 MS/MS spectrum acquired using collision gas pressure of 1.5 mTorr. The top 6 to 7 fragment ions from MS/MS spectra were selected for further CE ramping, and the QQQ SRM optimization mode was used to determine the optimal CE values. For each peptide, 4 to 5 transitions were retained as the QQQ-optimized transitions based on their intensities.

LC−MS/MS Analysis Using the LTQ-Orbitrap Velos

The in-house LC system was custom-built using Agilent 1200 nanoflow pumps (Agilent Technologies, Santa Clara, CA), Valco valves (Valco Instruments Co., Houston, TX), and a PAL autosampler (Leap Technologies, Carrboro, NC). Reversed-phase columns were prepared in-house by slurry packing 3 μm Jupiter C18 (Phenomenex, Torrence, CA) into 35 cm × 360 μm o.d. × 75 μm i.d fused silica (Polymicro Technologies Inc., Phoenix, AZ) using a 1 cm sol−gel frit for media retention. Mobile phases consisted of 0.1% formic acid in water (mobile

dx.doi.org/10.1021/pr500500d | J. Proteome Res. 2014, 13, 4479−4487

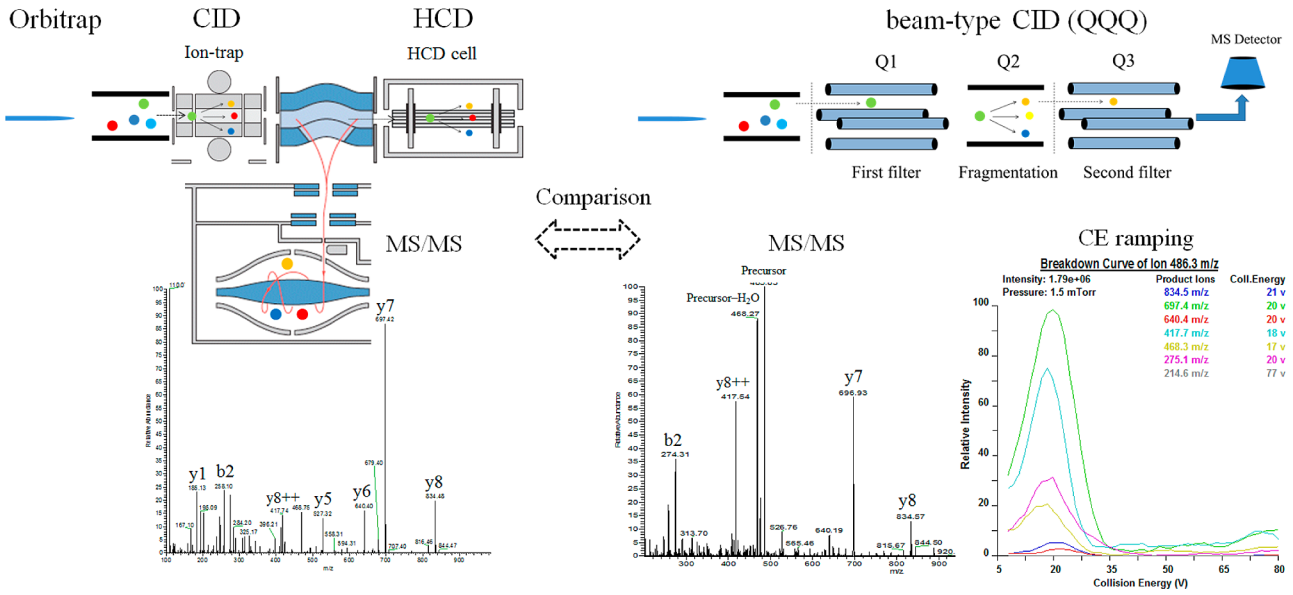


Figure 1. Instrumental arrangements used for HCD/CID and QQQ CID. A systematic comparison was made between Orbitrap HCD/CID and beam-type QQQ CID. (Left) MS/MS spectrum (HHGLLASA**R**, arginine +10) was acquired in HCD/CID mode by LTQ-Orbitrap Velos; (right) MS/MS spectrum (HHGLLASA**R**, arginine +10), as well as optimal CE, was acquired in the QQQ using CID.

phase A) and 0.1% formic acid in acetonitrile (mobile phase B) with a 100 min gradient using the follow profile (minute: %B): 0:0, 2:8, 20:12, 75:35, 97:60, and 100:85. After 100 min, mobile phase B was maintained at 100% for 10 min, and then mobile phase A was set at 100% and maintained for 10 min to recondition the column. The solvent gradient was started at 40 min after the sample was injected, and mass spectra acquisition began 10 min later to account for the column dead volume.

MS/MS analysis of the peptide mixture was based on an inclusion list built by Skyline[25] for each peptide with both charge 2+ and 3+ precursors using an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific) interfaced with a home-built electrospray ionization (ESI) source. The ESI emitter was custom-made by chemically etching 150 μm o.d. × 20 μm i.d. fused silica capillary.[28] The heated capillary temperature and spray voltage were 350 °C and 2.4 kV, respectively. Orbitrap spectra were recorded from 300 to 1800 m/z at a resolution of 100 000 followed by HCD of the 10 most abundant ions in each scan that matched with the target list of ions provided. For HCD settings, isolation width of 2 m/z, normalized collision energy 32%, activation time 0.1 s, and automated gain control (AGC) set at 3×10^4 were used. A dynamic exclusion time of 60 s was used to discriminate against previously analyzed ions.

LC–SRM Analysis Using the QQQ

A mixture of 92 synthetic peptides was separated by reversed-phase LC column (Waters, 10 cm × 75 μm i.d) using a nanoACQUITY UPLC system (Waters, Milford, MA) with the following 50 min gradient (minute:%B): 0:0, 13.5:10,17:15, 38:25, 49:38.5, and 50:95. After 50 min, mobile phase B was maintained at 100% for 10 min, and then mobile phase A was set at 100% and maintained for 10 min to recondition the column. The eluent was analyzed using a TSQ-Vantage (Thermo Scientific) operated in SRM mode with the following parameters: spray voltage, 2400 V; capillary temperature, 325 °C; collision gas pressure, 1.5 mTorr; using tuned S-lens value; Q1 and Q3 unit resolution. The CE ramping method was constructed according to Skyline instruction, while the ramping step was set to 2 V to cover a larger CE range. The method was then imported to the instrument, and the optimal CE was defined as that giving the highest intensity for the individual transition.

Although TSQ-Quantum and TSQ-Vantage instruments were used for direct infusion and LC–SRM analysis, respectively, our previous data have shown that at 1.5 mTorr collision gas pressure (used in this study) the difference of optimal CE values obtained from these two instruments is rather small. To ensure consistency, we recently randomly selected 10 from the list of 215 peptides and reoptimized their CE values on both the TSQ-Quantum and TSQ-Vantage instruments. A comparison of their optimal CE values showed their median difference between these two QQQ instruments was less than 2 V (Supporting Information Table 2).

Data Analysis

For direct infusion analysis, the best transition for each peptide and the optimal CE for each transition were recorded manually, and the peptide sequences, including the optimized transitions, were imported to Skyline. For LC–MS/MS data (HCD or CID), MS/MS spectra of the peptide mixtures were searched against the Human Swiss-Prot database (released on September 18, 2013) using the MS-GF+ algorithm[29] with the following parameters: precursor tolerance, 25 ppm; fixed modifications, cysteine carbamidomethylation (C +57.0215 Da), heavy lysine (K +8.014 Da), heavy arginine (R +10.009 Da); partially tryptic cleavage rule; and decoy search. After database searching, identified peptides were filtered to a false discovery rate (FDR) of <1% at the spectrum level by applying a MSGF spectrum probability filtering criterion ($<1.52 \times 10^{-9}$). A peptide library was then built by importing the HCD results into Skyline in pepXML format, followed by manual inspection in Skyline. The ranking of the fragment ions in HCD MS/MS spectra was then exported and analyzed for correlation to the direct infusion optimized result. In cases where one fragment ion was missing in the HCD spectra, the rank of this product ion was set as 50. For the calculation of Pearson correlations between HCD/CID and QQQ data, we included both y and b fragment ions for the 78 crude peptides, with fragment peaks across 10 MS/MS scans

dx.doi.org/10.1021/pr500500d | J. Proteome Res. 2014, 13, 4479–4487

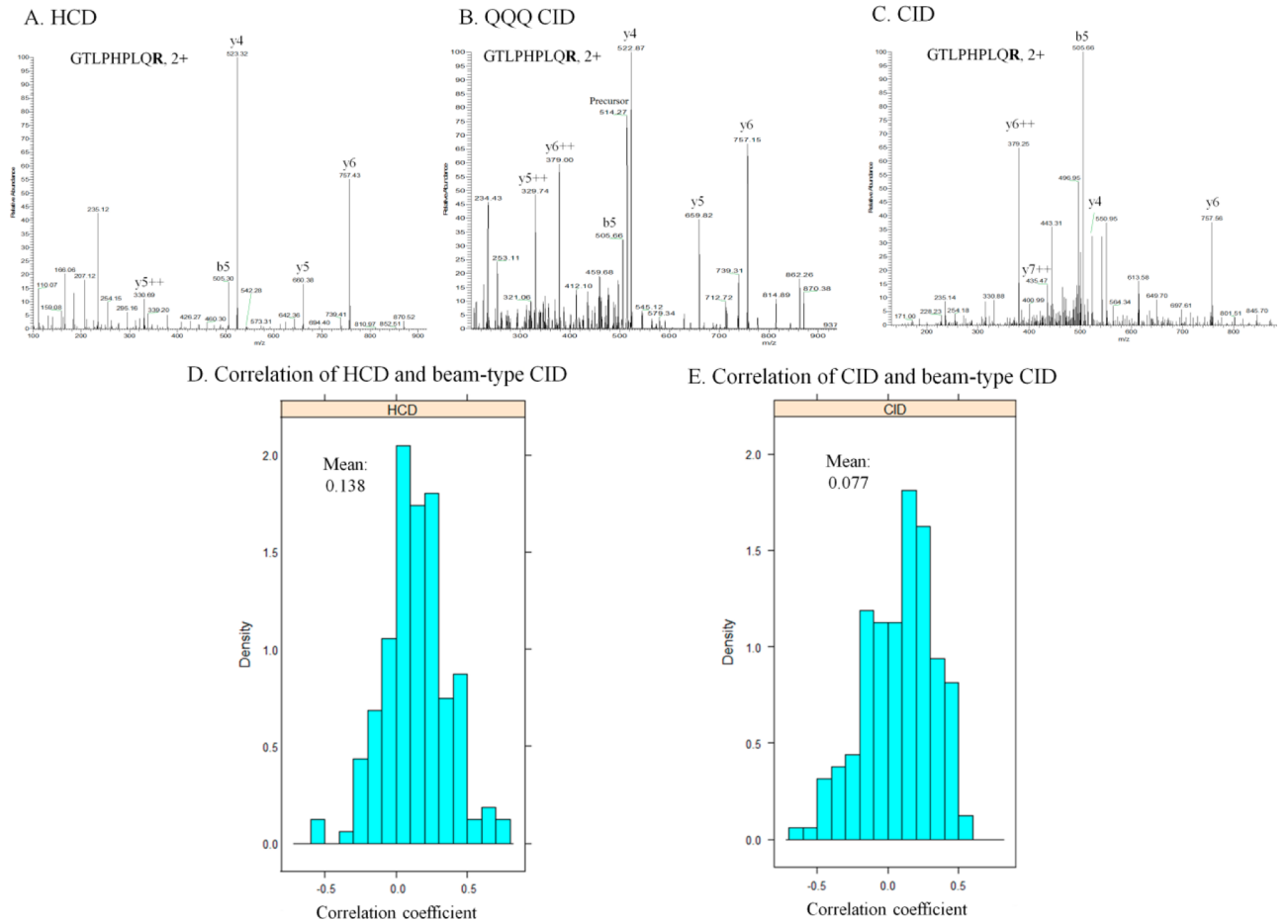


Figure 2. Comparison of spectrum and Pearson correlation between LTQ-Orbitrap Velos CID/QQQ CID and HCD/QQQ CID. (A–C) Comparison of MS/MS spectrum for doubly charged peptide GTLPHPLQR (arginine +10), either HCD (A), QQQ CID (B), or CID (C). (D, E) Pearson product-moment correlation coefficients for HCD/CID versus QQQ CID for 78 crude heavy peptides. Distribution of Pearson correlation displayed in histogram for both HCD/QQQ CID (lower left) and CID/QQQ CID (lower right), which showed a significantly higher correlation in HCD compared to that in CID with a *p*-value of 0.01972 (Welch's Two Sample *t*-test). The *x* axis is the correlation coefficient, and the *y* axis is the data density.

of QQQ direct infusion data being averaged. For Orbitrap HCD and CID data, the MS/MS scan with the best spectrum probability score was retained for each peptide/charge state. For all of the commonly detected product ions between HCD/CID and QQQ MS/MS spectra, the Pearson correlation was calculated using product ion intensities with in-house software. For calculating the intensity ratio of the new prediction formula to the Skyline prediction, the area of each transition with different CE was exported from Skyline and analyzed.

RESULTS

Selecting the best transitions and optimizing their CEs are crucial steps for the development of optimally sensitive SRM assays. To explore the feasibility of accelerating this process of SRM assay development using pre-existing shotgun proteomics data, we compared the MS/MS Orbitrap HCD or CID fragmentation data with the optimized QQQ data (Figure 1). The peptide mixture was analyzed by LC–MS/MS on an LTQ-Orbitrap Velos instrument to generate HCD/CID data. SRM parameters for the 215 peptides were also optimized individually through direct infusion with the QQQ instrument. For Orbitrap data, two synthetic peptide mixtures were analyzed separately: a peptide mixture with 78 peptides was analyzed in both CID and HCD fragmentation mode for comparison of similarity between HCD/CID and QQQ CID fragmentation, and another peptide mixture with 137 peptides was analyzed in HCD mode only for a more detailed comparison between HCD and QQQ CID. Peptides identified on Orbitrap were first filtered to achieve <1% FDR at the spectrum level with a median spectrum probability score of 3.60×10^{-14}, and only spectra with the best spectrum probability values were retained for comparison. For the QQQ data, 215 peptides were optimized individually using direct infusion, including transition selection with highest response as well as optimal CE for each transition (Figure 1, right). These manually optimized parameters (i.e., best transition and optimal CE) were considered as the gold standard for SRM assay development and used for comparison with the Orbitrap HCD/CID results. For each peptide, 4 to 5 product ions were selected on the basis of their intensities, most of which were y ions (90.4%) with some b ions (9.6%) and 2+ product ions (11.4%) (b1 and y1 ions were excluded). At least one product ion with m/z higher than precursor was included for each peptide.

Comparison of Peptide Fragmentation Patterns between Orbitrap HCD/CID and QQQ CID

In general, the QQQ CID had greater similarity with Orbitrap HCD than with the CID, as expected and as illustrated in

dx.doi.org/10.1021/pr500500d | *J. Proteome Res.* 2014, 13, 4479–4487

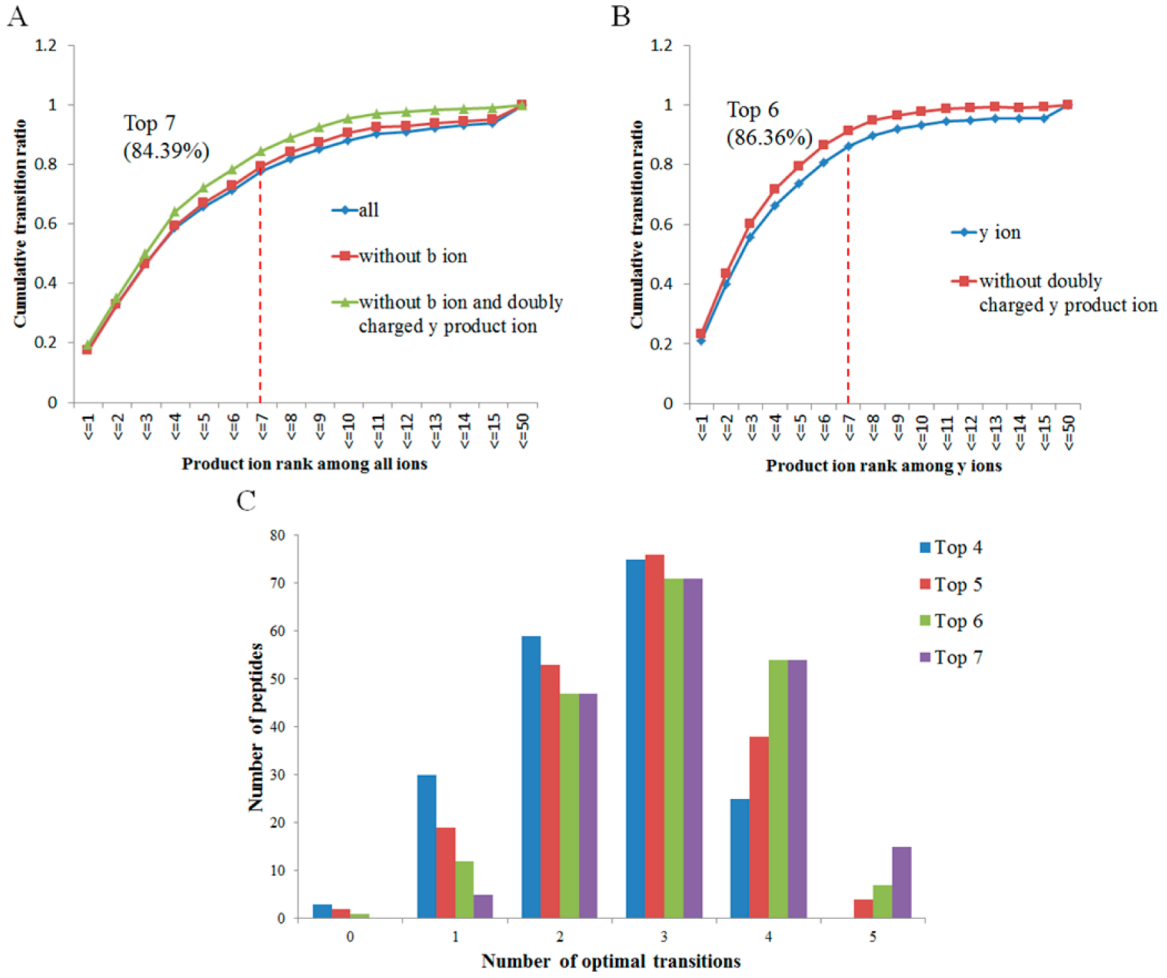


Figure 3. Mapping of top HCD fragment ions to QQQ-optimized transitions. Different types of fragment ions from HCD spectrum were further mapped to QQQ-optimized transitions, either for all fragment ions (A) or y fragment ions only (B), and the distribution of peptides containing specific number of optimal transitions after selection of top y fragment ions is displayed (C). (A, B) The ranks of fragment ions in HCD spectrum were summarized from Skyline and are labeled on the *x* axis; the rank of product ion is set as 50 if it was missing in the HCD spectrum or below noise level. The *y* axis is the cumulative proportion of matched QQQ-optimized transitions among all QQQ-optimized transitions. (C) The *x* axis is the number of optimal transitions for one peptide; the *y* axis is the number of peptides containing specific number of optimal transitions.

Figure 2A–C. When comparing the CID spectrum to the HCD and QQQ CID spectrum, there was a dominant b5 fragment ion present in CID spectrum, whose intensity was significantly lower in either the HCD or QQQ CID spectra. The less intense b fragment ions were most likely due to multiple stages of fragmentation that were inherent with the higher-energy injection of ions into the collision gas to produce the QQQ CID spectrum. More comparison of peptide spectra from Orbitrap HCD/CID versus QQQ CID can be found in Supporting Information Figure 1.

We further calculated the Pearson correlation between the measurements using the 78 synthetic peptide mixtures: Orbitrap CID vs QQQ CID and HCD vs QQQ CID (see Experimental Section). The distributions of Pearson correlation coefficients for CID vs QQQ CID (lower right) and HCD vs QQQ CID (lower left) are shown in Figure 2D and 2E, respectively. Overall, HCD showed a statistically significant higher correlation with QQQ CID spectra compared to CID data (Welch's Two Sample *t* test *p*-value = 0.01972), which was consistent with a previous study.[23] Therefore, choosing Orbitrap HCD data over CID data is more effective for selecting peptide transitions in SRM assay optimization.[23]

Comparison between HCD Data and QQQ Optimization Data

For a more detailed comparison between HCD data and QQQ-optimized results, we included another 137 synthetic peptides for the analysis. Altogether, 182 out of 215 peptides were identified after database searching and filtering from the HCD data (including 11 peptides identified with both 2+ and 3+ precursors). The median peptide length for those identified peptides is 12 amino acid residues. In order to facilitate the comparison between HCD data and QQQ-optimized results, we assembled a peptide library for those crude heavy peptides in Skyline using HCD data, which facilitated selecting product ions and exporting the rank information on each product ion. Then, we mapped the rank of the product ions to the QQQ-optimized result (Figure 3A and Supporting Information Table 3).

As shown in Supporting Information Table 3, 149 out of 193 (77%) top-ranked HCD product ions were among the QQQ-optimized 4 to 5 best transitions, compared to a decreased 65% of the second-ranked HCD product ions that were matched. By including higher-ranked product ions, the cumulative proportion (i.e., the sum of proportions for selected product ion

dx.doi.org/10.1021/pr500500d | *J. Proteome Res.* 2014, 13, 4479–4487

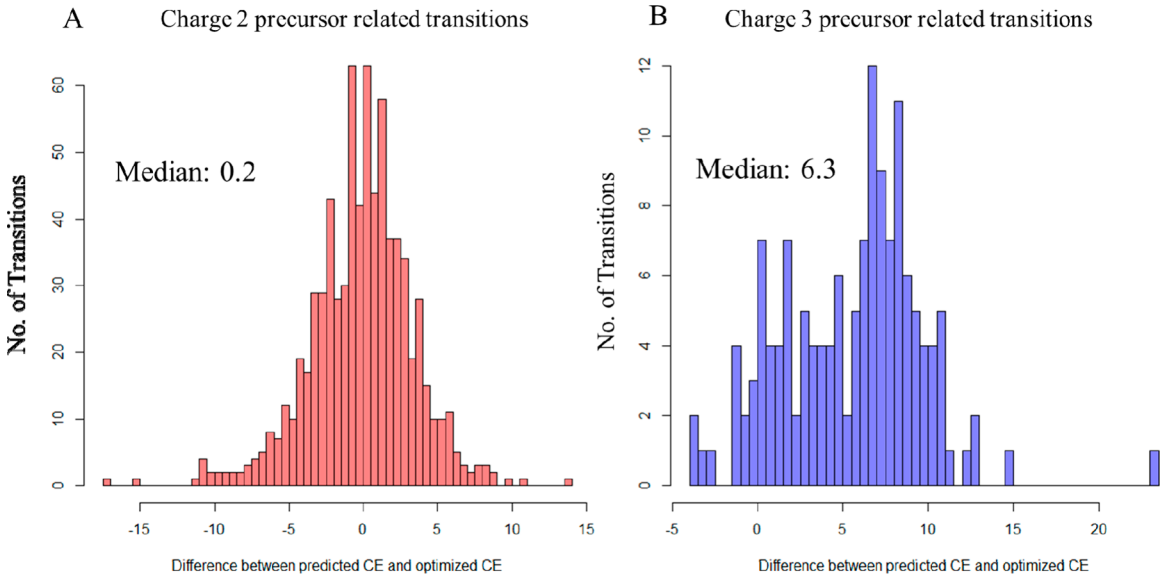


Figure 4. Distribution of the difference between the predicted and optimized CE. Distribution of the difference between the Skyline-predicted CE and optimal CE obtained from a QQQ instrument was demonstrated for all transitions either from charge 2+ precursors (A) or charge 3+ precursors (B). The *x* axis is the difference between predicted CE and optimal CE, and the *y* axis is the number of transitions.

ranked either equal or smaller) of mapped optimal transitions increased. As shown in Figure 3A and Supporting Information Table 3, the matched number as well as a cumulative proportion of optimized results continued to increase as the rank of the product ion increased. However, typical SRM experiments need to limit the number of transitions for each peptide to avoid longer scan times or reduced signal intensities and degraded quantification. When the top 7 HCD product ions were selected, a cumulative 80% of QQQ-optimized transitions were covered.

We noted that while the elimination of b ions and doubly charged y product ions resulted in a slight decrease in the matched number of optimized results (Supporting Information Figure 2A), the mapping of the remaining results (rank unchanged) still covered a larger proportion of the optimized results (Figure 3A), which suggested a preference of y ions over b ions and doubly charged y product ions. As shown in Figure 3A for mapping the top 7 ranked HCD product ions, the cumulative proportion was increased from 77 to 79% by eliminating b ions and was further increased to 84% by the elimination of doubly charged y product ions. Because b ions were less favored due to the potentially less accurate quantitative information from the heavy-isotope labeled synthetic peptides, we examined the scenario where only y ions were considered, which shifted the ranking of product ions as well. As shown in Figure 3B and Supporting Information Table 4, 82% of top-ranked y ions matched to the optimized product ions; this proportion decreased to 74% for the second-ranked y ions. Similarly, only 21% of optimized product ions were covered if only top-ranked y fragment ions were retained. In comparison, 81% of all optimized product ions were covered by the selection of the top 6 y ions. The proportion of mapped optimized product ions was further increased from 81 to 86% (Figure 1B and Supporting Information Table 4) when doubly charged y ions were not considered, although the absolute number decreased (see Supporting Information Figure 2B).

In summary, by selection of the top 6 y product ions in the HCD data, up to 86% of all QQQ-optimized results (y fragment ion) could be retained, thus allowing for rapid, automated selection of optimal transitions for most SRM assays. To reduce the chance of having interference in the transition list, especially in complex sample matrices such as blood plasma/serum, we made sure that there is at least one transition with m/z higher than that of the precursor for each peptide. However, detailed tests are warranted for a given sample matrix in order to completely rule out potential interferences.

Distribution of Peptides with Certain Transitions

After showing that the selection of the top 6 y product ions from HCD data coincided with 86% of QQQ-optimized results, we next examined the distribution of peptides having specific numbers of optimized transitions (it is generally desired for SRM assays that each peptide have at least three transitions). Figure 3C shows the distribution of peptides versus the number of optimal transitions after the selection of a different number of top-ranked HCD product ions (y ion only) (Figure 3C). The distribution of peptides for all product ions mapping is illustrated in Supporting Information Figure 3.

As shown in Figure 3C, when the top 4 product ions were selected, 100 out of 193 peptides (52%) contained equal or greater than three optimal transitions, whereas 34 peptides (18%) had equal or less than one transition; in contrast, when the top 7 product ions were selected, more than 140 peptides (72%) could have at least three transitions, whereas only 6 peptides (3%) had equal or less than one transition. We also noticed that the result was quite comparable for top 6 and top 7 transition selection from y product ions, with a very similar distribution. In contrast, by selection of the top 7 ions from all product ions (Supporting Information Figure 3), about 122 (63%) peptides could have at least three transitions, which was much lower compared to that for selection of the top 6 y ions. Therefore, combining the information from the transition coverage and peptide distribution, we concluded that selecting the top 6 HCD y product ions was sufficient for automated SRM assay development.

dx.doi.org/10.1021/pr500500d | J. Proteome Res. 2014, 13, 4479−4487

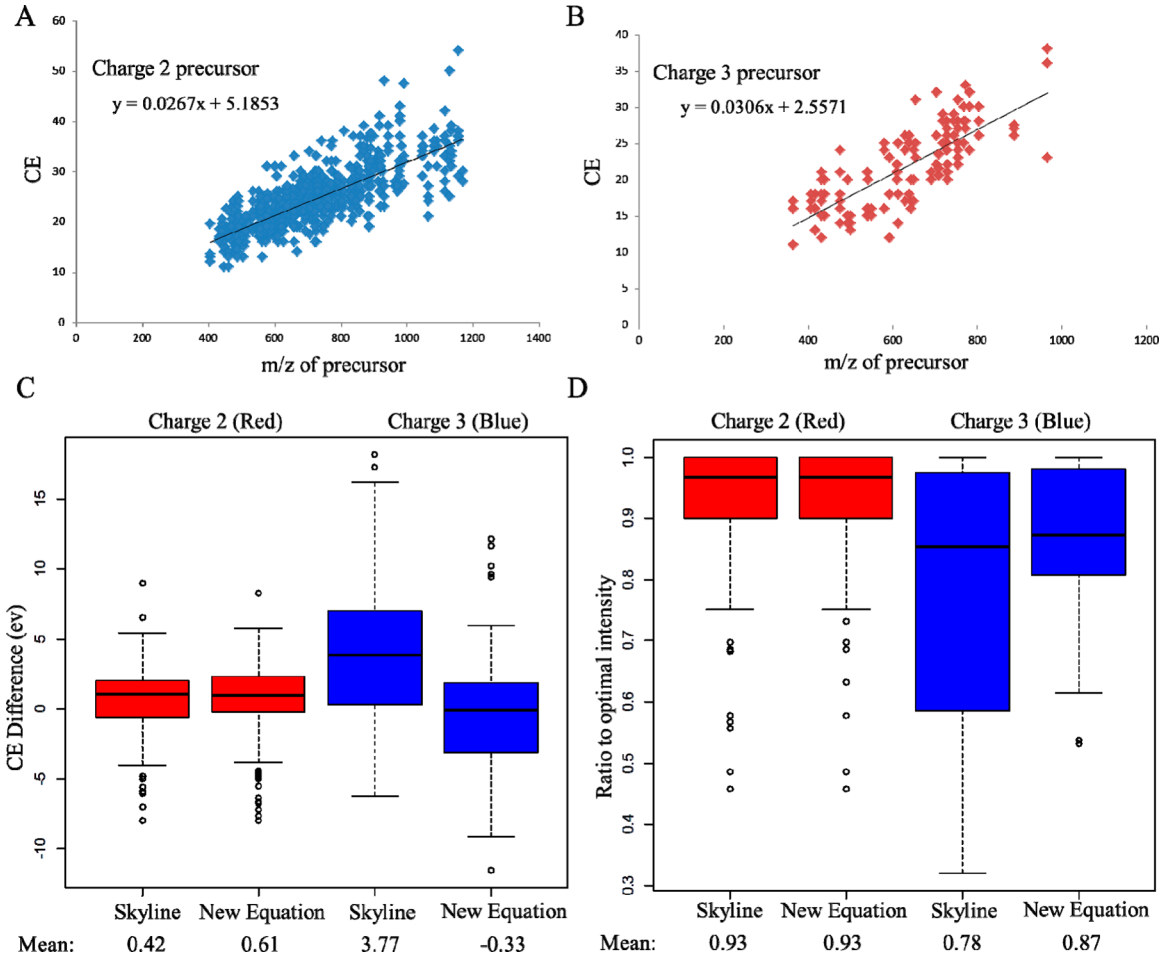


Figure 5. Construction of new CE prediction equations and the comparison to Skyline prediction in an independent data set. A new CE prediction equation was constructed for both charge 2+ and 3+ precursors and further compared to Skyline predicted results in an independent data set consisting of 92 synthetic peptides. (A, B) Construction of new CE prediction equation for charge 2+ (A) and charge 3+ (B) precursor related transitions. The *x* axis shows the *m*/*z* of the precursor, and the *y* axis represents the value of CE. (C) Comparison of the difference between Skyline prediction/new equation prediction and optimal CE for either charge 2+ (red) and charge 3+ (blue) precursor related transitions. The mean of the CE difference is labeled below the *x* axis. The *x* axis plots either Skyline or the new equation, and the *y* axis is the CE difference (in volts) between the predicted CE and the optimal CE. (D) The distribution of intensity ratio for new equation prediction/Skyline prediction to optimal CE, either for charge 2+ precursor (red) or charge 3+ precursor (blue). The mean of ratio to optimal intensity is labeled below the *x* axis. The *x* axis plots either Skyline or new equation, and the *y* axis is the ratio to optimal intensity.

Differences between Skyline-Predicted and QQQ-Optimized CE

Besides the transition selection, CE optimization is also an important step for developing an optimal SRM assay. There are a number of approaches available for optimizing the CE, such as LC−MS-based CE ramping optimization, which enables automated and high-throughput CE optimization using Skyline, as well as theoretical prediction, which is an important complement to manual optimization.[30] However, to obtain a gold standard set of CE values for the comparison, our optimal CE values for peptides were obtained by direct infusion of peptides into QQQ, which was considered to be the most accurate irrespective of the LC elution profile.

We calculated the difference between optimal CE and CE predicted using Skyline for each transition (Figure 4). For charge 2+ precursors (Figure 4A), the difference between predicted CE and optimal CE was quite small: most of the transitions (87%) fell within a ±5 eV window, with a median of 0.2 eV, which was consistent with a previous study;[30] however, for charge 3+ precursors (Figure 4B), the difference between predicted CE and empirical CE was much larger compared to that for the charge 2+ precursors, with a wide and flat distribution and a median of 6.3 eV. This indicated that the CE prediction formula for charge 3+ precursors needed further refinement in order to establish an optimal SRM assay.

Construction and Validation of New CE Prediction Equations

We next constructed two new CE prediction equations based on QQQ-optimized results for charge 2+ and 3+ precursors, respectively (Figure 5A,B), and the improvement was further validated in an independent LC ramping-based CE optimization experiment (using 92 synthetic peptides; see Experimental Section). As shown in Figure 5C, the differences of CE between the predicted values (both Skyline prediction and the new equation) and the LC ramping optimization results are depicted. For charge 2+ precursors (red), the difference between the two prediction approaches and the gold standard was negligible, with a mean of 0.42 for Skyline and 0.61 for new equation prediction. However, for charge 3+ precursors (blue), the mean of the difference between Skyline and the optimized

dx.doi.org/10.1021/pr500500d | J. Proteome Res. 2014, 13, 4479−4487

result was 3.7, demonstrating that the predicted CE was too high for charge 3+ precursors; in contrast, the mean of the difference between the new equation and the optimized result was only −0.33. These results demonstrated that significantly better CE prediction can be achieved from the new equation than when applying Skyline.

We further calculated the gain in signal intensity for the new equation when compared to the Skyline prediction for charge 2+ and charge 3+ precursors using the CE ramping result (92 synthetic peptides). As depicted in Figure 5D, for charge 2+ precursor related transitions (red), there was almost no difference between Skyline and the new equation predictions: both had a mean of ratio 0.93. For charge 3+ precursor (blue), the mean of the new equation was significantly higher than that for Skyline predictions (0.87 vs 0.78; *t* test, *p*-value 0.001724), which confirmed the usefulness of the new equation for charge 3+ precursor related transitions. We also compared intensities from Skyline to the new equation predictions directly, which (Supporting Information Figure 4) had similar results, showing almost no difference for charge 2+ precursor related transitions, whereas for charge 3+ precursor related transitions, the maximum increase in intensity from the new equation was 2.7-fold greater than with Skyline, with an overall mean ratio of 1.2.

DISCUSSION

SRM has been widely used for accurate protein quantification due to its high sensitivity and specificity. The development of optimal SRM assays is of critical importance for sensitive detection and accurate quantification of target proteins in complex biological samples. In this work, we demonstrated the utility of readily obtainable HCD data for the selection of the best transition for a given peptide and the determination of optimal CE, both of which are important steps in streamlining the assay optimization process. We have systematically compared the MS/MS data resulting from shotgun proteomics (CID and HCD) with QQQ manually optimized data through analysis of a relatively large set of 215 synthetic peptides. For transition selection, it was determined that selection of the top 6 y product ions from HCD spectra could cover as much as 86% of the best transitions. We further showed that the selection of the top 6 HCD y product ions was sufficient to cover the optimal transitions for most peptides, with at least 3 of the QQQ optimal transitions for each peptide. For CE optimization, we found that the Skyline-predicted CE was accurate for most charge 2+ precursor transitions; however, it was too high for most charge 3+ precursor transitions. By using the newly constructed prediction equation established from our QQQ experimental data, we obtained as much as a 2.7-fold increase (1.2-fold average) in intensity for charge 3+ precursor transitions, which was higher than previous reports.[23,30] Finally, we validated the usefulness and accuracy of the new, refined formula using a completely independent set of synthetic peptides. We recognize that the CE optimization results may not be applicable to all QQQ instruments from different vendors; however, our work demonstrated the proof of principle for generating more accurate CE prediction equations that require large sets of peptides, and similar experiments can be pursued by applying the same principles for different types of QQQ instruments.

In summary, our results demonstrated the feasibility of automating the selection of the best transitions from the existing HCD results (i.e., discovery results) without using synthetic peptides and obtaining optimal CE conditions for the most responsive transitions, which would accelerate and improve large-scale targeted proteomic experiments with sensitive measurements of hundreds of target proteins in complex biological matrices.

ASSOCIATED CONTENT

Supporting Information

Figure 1: Comparison of MS/MS spectra among HCD, QQQ CID, and CID. Figure 2: Mapping of top HCD fragment ions to TSQ-optimized transitions. Figure 3: Distribution of 193 peptides containing a different number of optimal transitions for all product ions. Figure 4: Distribution of intensity ratio between new equation prediction and Skyline prediction for charge 2 or charge 3 precursor related transitions. Table 1: Detailed list of selected peptides and corresponding protein candidates. Table 2: CE comparison of peptides on TSQ Quantum and TSQ Vantage. Table 3: Mapping of HCD results to the TSQ optimized transitions for all product ions. Table 4: Mapping of HCD results to TSQ optimized transitions for *y* fragment ions only. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*(T.L.) Phone: 509-371-6346; Fax: 509-371-6564; E-mail: tao.liu@pnnl.gov.
*(R.D.S.) Phone: 509-371-6576; Fax: 509-371-6564; E-mail: dick.smith@pnnl.gov.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Portions of this work were supported by NIH grant U24-CA-160019 from the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), NIGMS Biomedical Technology Research Resource P41GM103493, and DP2OD006668. The experimental work described herein was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by United States Department of Energy (DOE) Office of Science Biological and Environmental Research and located at Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the DOE under contract DE-AC05-76RL0 1830.

REFERENCES

(1) Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **2009**, *138*, 795−806.
(2) Keshishian, H.; Addona, T.; Burgess, M.; Kuhn, E.; Carr, S. A. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2007**, *6*, 2212−29.
(3) Qian, W. J.; Liu, T.; Petyuk, V. A.; Gritsenko, M. A.; Petritis, B. O.; Polpitiya, A. D.; Kaushal, A.; Xiao, W.; Finnerty, C. C.; Jeschke, M. G.; Jaitly, N.; Monroe, M. E.; Moore, R. J.; Moldawer, L. L.; Davis, R. W.; Tompkins, R. G.; Herndon, D. N.; Camp, D. G.; Smith, R. D. Large-scale multiplexed quantitative discovery proteomics enabled by the use of an ^{18}O-labeled "universal" reference sample. *J. Proteome Res.* **2009**, *8*, 290−9.

(4) Shi, T.; Su, D.; Liu, T.; Tang, K.; Camp, D. G., II; Qian, W. J.; Smith, R. D. Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* **2012**, *12*, 1074−92.
(5) Huttenhain, R.; Soste, M.; Selevsek, N.; Rost, H.; Sethi, A.; Carapito, C.; Farrah, T.; Deutsch, E. W.; Kusebauch, U.; Moritz, R. L.; Nimeus-Malmstrom, E.; Rinner, O.; Aebersold, R. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* **2012**, *4*, 142ra94.
(6) Shi, T.; Fillmore, T. L.; Sun, X.; Zhao, R.; Schepmoes, A. A.; Hossain, M.; Xie, F.; Wu, S.; Kim, J. S.; Jones, N.; Moore, R. J.; Pasa-Tolic, L.; Kagan, J.; Rodland, K. D.; Liu, T.; Tang, K.; Camp, D. G., II; Smith, R. D.; Qian, W. J. Antibody-free, targeted mass-spectrometric approach for quantification of proteins at low picogram per milliliter levels in human plasma/serum. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15395−400.
(7) Whiteaker, J. R.; Lin, C.; Kennedy, J.; Hou, L.; Trute, M.; Sokal, I.; Yan, P.; Schoenherr, R. M.; Zhao, L.; Voytovich, U. J.; Kelly-Spratt, K. S.; Krasnoselsky, A.; Gafken, P. R.; Hogan, J. M.; Jones, L. A.; Wang, P.; Amon, L.; Chodosh, L. A.; Nelson, P. S.; McIntosh, M. W.; Kemp, C. J.; Paulovich, A. G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **2011**, *29*, 625−34.
(8) Liu, T.; Hossain, M.; Schepmoes, A. A.; Fillmore, T. L.; Sokoll, L. J.; Kronewitter, S. R.; Izmirlian, G.; Shi, T.; Qian, W. J.; Leach, R. J.; Thompson, I. M.; Chan, D. W.; Smith, R. D.; Kagan, J.; Srivastava, S.; Rodland, K. D.; Camp, D. G., II Analysis of serum total and free PSA using immunoaffinity depletion coupled to SRM: correlation with clinical immunoassay tests. *J. Proteomics* **2012**, *75*, 4747−57.
(9) Wu, C.; Wei, W.; Li, C.; Li, Q.; Sheng, Q.; Zeng, R. Delicate analysis of post-translational modifications on Dishevelled 3. *J. Proteome Res.* **2012**, *11*, 3829−37.
(10) Nesvizhskii, A. I.; Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* **2005**, *4*, 1419−40.
(11) Deutsch, E. W. The PeptideAtlas Project. *Methods Mol. Biol.* **2010**, *604*, 285−96.
(12) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222.
(13) Hilpert, K.; Winkler, D. F.; Hancock, R. E. Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat. Protoc.* **2007**, *2*, 1333−49.
(14) Frewen, B. E.; Merrihew, G. E.; Wu, C. C.; Noble, W. S.; MacCoss, M. J. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal. Chem.* **2006**, *78*, 5678−84.
(15) Prince, J. T.; Carlson, M. W.; Wang, R.; Lu, P.; Marcotte, E. M. The need for a public proteomics repository. *Nat. Biotechnol.* **2004**, *22*, 471−2.
(16) Craig, R.; Cortens, J. P.; Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **2004**, *3*, 1234−42.
(17) Craig, R.; Cortens, J. P.; Beavis, R. C. The use of proteotypic peptide libraries for protein identification. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1844−50.
(18) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655−8.
(19) Prakash, A.; Tomazela, D. M.; Frewen, B.; Maclean, B.; Merrihew, G.; Peterman, S.; Maccoss, M. J. Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J. Proteome Res.* **2009**, *8*, 2733−9.
(20) Xia, Y.; Liang, X.; McLuckey, S. A. Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal. Chem.* **2006**, *78*, 1218−27.
(21) 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*, 709−12.
(22) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18132−8.
(23) de Graaf, E. L.; Altelaar, A. F.; van Breukelen, B.; Mohammed, S.; Heck, A. J. Improving SRM assay development: a global comparison between triple quadrupole, ion trap, and higher energy CID peptide fragmentation spectra. *J. Proteome Res.* **2011**, *10*, 4334−41.
(24) Holstein-Sherwood, C. A.; Gafken, P. R.; Martin, D. B. Collision energy optimization of *b*- and *y*-ions for multiple reaction monitoring mass spectrometry. *J. Proteome Res.* **2011**, *10*, 231−40.
(25) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966−8.
(26) Percy, A. J.; Chambers, A. G.; Yang, J.; Hardie, D. B.; Borchers, C. H. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. *Biochim. Biophys. Acta* **2014**, *1844*, 917−26.
(27) Eyers, C. E.; Lawless, C.; Wedge, D. C.; Lau, K. W.; Gaskell, S. J.; Hubbard, S. J. CONSeQuence: prediction of reference peptides for absolute quantitative proteomics using consensus machine learning approaches. *Mol. Cell. Proteomics* **2011**, *10*, M110.003384.
(28) Kelly, R. T.; Page, J. S.; Luo, Q.; Moore, R. J.; Orton, D. J.; Tang, K.; Smith, R. D. Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem.* **2006**, *78*, 7796−801.
(29) Kim, S.; Gupta, N.; Pevzner, P. A. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **2008**, *7*, 3354−63.
(30) Maclean, B.; Tomazela, D. M.; Abbatiello, S. E.; Zhang, S.; Whiteaker, J. R.; Paulovich, A. G.; Carr, S. A.; Maccoss, M. J. Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal. Chem.* **2010**, *82*, 10116−24.