Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry

Brendan MacLean[1†], Daniela M. Tomazela[1†], Susan E. Abbatiello[2], Steven A. Carr[2], and Michael J. MacCoss[1]

1. Department of Genome Sciences, University of Washington, Seattle, WA, USA.
2. Broad Institute of MIT and Harvard, Cambridge, MA, USA.
† These authors contributed equally to this work.

Outline:

Proteomics experiments based on Selected Reaction Monitoring (SRM, also referred to as Multiple Reaction Monitoring or MRM) are being used to target large numbers of protein candidates in complex mixtures. At present, instrument parameters are often optimized for each peptide, a time and resource intensive process. Large SRM experiments are greatly facilitated by having the ability to predict MS instrument parameters that work well with the broad diversity of peptides they target. For this reason, we :

✓ investigated the impact of using simple linear equations to predict the collision energy (CE) on peptide signal intensity and compared it with the empirical optimization of the CE for each peptide and transition individually;

✓ report a fully automated workflow for optimized method development, compatible with MS platforms from Agilent (pending a software patch), Applied Biosystems, Waters and Thermo Fisher Scientific. This workflow has been implemented in version 0.6 of the open source Skyline software (http://proteome.gs.washington.edu/software/skyline).

Methods:

The data were acquired at two sites (University of Washington and The Broad Institute).

Sample: Tryptic digest of six bovine proteins purchased from Michrom (PN: PTD/00006/63)

Collision Energy optimization replicates: the sample was reconstituted and diluted to 50 fmol/µl (or 100 fmol/µl for the 4000 Q Trap) using 97 % Water, 3 % Acetonitrile, and 0.1 % Formic Acid. 2 µl of the digest was injected into the LC system (100 fmol on column) for data acquisition.

Dilution analysis: the sample was initially reconstituted to 200 fmols/µl using the same buffer and diluted to produce the following concentrations: 6.25 fmols/µl, 12.5 fmols/µl, 25 fmols/µl, 50 fmols/µl, 75 fmols/µl and 100 fmols/µl. A 2 µl aliquot of each sample was injected on column for data acquisition.

Skyline method set up for collision energy optimization

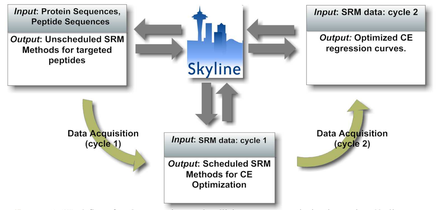


Figure 1: Workflow for the experimental collision energy optimization using Skyline.

Data Acquisition Cycle 1:
✓20 doubly charged and 10 triply charged peptides using 4 transitions per precursor ion.

Data Acquisition Cycle 2:
✓Scheduling time windows of 4 or 5 min and maximum number of 132 or 110 allowable concurrent measurements.
✓4 or 5 scheduled methods, depending on LC gradient conditions and system capabilities.
✓Collision energy optimization parameters were set to use 5 steps on either side of the value predicted by the default equation (Table 1), with the step size set to 1 V. In total, 11 collision energy voltage values were considered for each fragment ion, yielding 1320 transitions per replicate.
✓Consecutive mass variations of one hundredth of a mass unit were used for each fragment ion as a vendor-neutral method of allowing software tools like Skyline to specify and recognize variation in secondary parameters like CE. The true product m/z was assigned to the CE for the default equation, with no other product m/z value varying more than 5 hundredths from that.

Differential Experiment:
Using a single set of measurements for each transition and CE values (4000 Q-trap) from step 2, two unscheduled methods were generated: i) Values predicted by linear equations calculated from all precursors; ii) Values that produced the maximum area for each transition. (Methods run in randomized order over 8 replicates).

Results:

Data Analysis:

Raw data from cycle 2 was imported into Skyline for peak area integration and for the determination of the updated CE equations. See Figure 2 and 3 bellow.

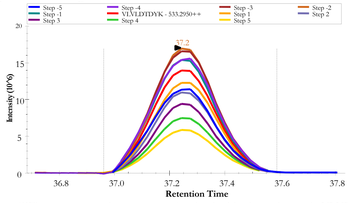


Figure 2: Collision energy (CE) optimization chromatogram for the doubly charged peptide VLVLDTDYK at m/z 533.3. The colored lines correspond to the Total Ion Current (TIC) for the sum of four transitions (533.3>853.4, 754.4, 641.3 and 526.2) at eleven different voltage steps. The red trace corresponds to the CE value calculated using the instrument default equation (TSQ Ultra, Thermo Fisher Scientific, See Table 1); the positive-value steps represent 1V increments above the red line CE value and the negative-value steps represent 1V decrements.

Reproducibility:

Figure 3 shows the reproducibility of the relative peak areas, calculated by integrating the TIC chromatograms from ten replicate analyses, for the doubly charged peptide VLVLDTDYK.

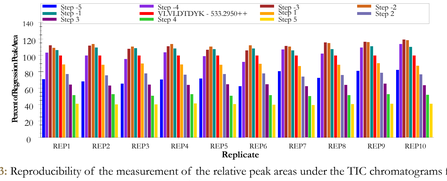


Figure 3: Reproducibility of the measurement of the relative peak areas under the TIC chromatograms for ten replicate analyses for the doubly charged peptide VLVLDTDYK. The areas obtained at each voltage step were normalized within each replicate relative to the peak area values obtained at the reference voltage setting (red bar, 100%) calculated using the default CE equation (TSQ Ultra, Thermo Fisher Scientific).

Although we collected our data from 10 technical mass spectrometry replicates, most laboratories derive "optimal" CE values from a single measurement. In these cases, peptides are injected once, and intensities are measured over a range of CE values. The CE yielding the maximum intensity, by peak area, is chosen for all future experiments. For this approach to work, the effect of CE on peak area must be reproducible enough that a single measurement produces an accurate representation of the optimal CE value. See Figure 3 bellow.

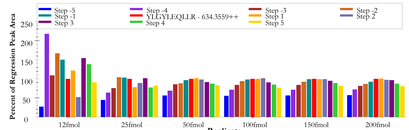


Figure 3: Concentration dependence of the peak area measurement reproducibility for peptide YLGYLEQLLR under the same conditions described in Figure 2. Varying on-column amounts of peptide YLGYLEQLLR were injected (12 - 200 fmol). Reproducible area measurements were observed at 50 fmol on-column or above.

Comparing Regression Coefficients

A comparison of the default and optimized linear equations for our selected peptides, acquired on a TSQ Ultra, is shown in the **Figure 4**. These best fit linear regressions were computed separately for the doubly and triply charged peptide precursors. The solid lines represent simple linear regressions of the experimental values obtained during the CE optimization study. The dashed lines represent the plots for their respective default linear equations used for data acquisition (Table 1). The same plots were generated for each of the instrument platforms used in our analysis.

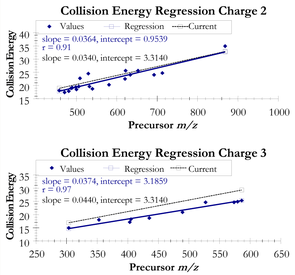


Figure 4: Comparison of the default and optimized linear equations for our selected peptides, acquired on a TSQ-Ultra.

Table 1: Default and updated CE equations with 95% confidence intervals for doubly and triply charged peptides for different MS platforms. Displayed in parenthesis after each equation is the number of precursor measurements used to calculate the regression.

	Starting Linear Equations for Predicting Peptide CE (Default equations)	
TSQ Ultra	CE=0.034*m/z + 3.314	CE=0.044*m/z + 3.314
TSQ Vantage		
4000 Q Trap	CE=0.043*m/z + 4.756	CE=0.043*m/z + 4.756

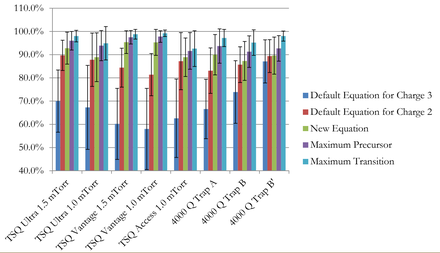
	Calculated Linear Equations	
TSQ Ultra	CE=(0.055±0.02*m/z +	CE=(0.027±0.022*m/z +
1.0 mTorr	-8.01±11.5 (n = 18)	4.492±10.379 (n = 9)
TSQ Ultra	CE=(0.056±0.008*m/z +	CE=(0.037±0.007*m/z +
1.5 mTorr	0.954±4.818 (n = 18)	3.525±3.384 (n = 9)
TSQ Vantage	CE=(0.041±0.01*m/z +	CE=(0.04±0.005*m/z +
1.0 mTorr	-3.442±5.765 (n = 18)	0.773±2.214 (n = 9)
TSQ Vantage	CE=(0.03±0.005*m/z +	CE=(0.038±0.004*m/z +
1.5 mTorr	2.905±3.151 (n = 18)	2.281±1.808 (n = 9)
TSQ Access	CE=(0.049±0.009*m/z +	CE=(0.039±0.012*m/z +
1.0 mTorr	-5.75±5.428 (n = 16)	3.314±5.835 (n = 7)
4000 Q Trap	CE=(0.052±0.008*m/z +	CE=(0.036±0.008*m/z +
(Instrument A)	-2.919±5.514 (n = 13)	4.106±4.444 (n = 9)
4000 Q Trap	CE=(0.057±0.01*m/z +	CE=(0.035±0.015*m/z +
(Instrument B)	-4.815±6.384 (n = 14)	6.494±8.615 (n = 9)
4000 Q Trap	CE=(0.057±0.009*m/z +	CE=(0.031±0.018*m/z +
(Instrument B')	-4.256±5.752 (n = 18)	7.082±10.3 (n = 9)

Comparing Approaches

To compare the relative effect of different approaches for choosing CE values on the sensitivity of measurement, we simulated these approaches using our data. Because CE values are commonly chosen from a single set of measurements, we first chose one of our 10 replicates, designating it the training set. The training set was used to calculate CE values for the remaining 9 replicates, designated as test sets. The four approaches for calculating the CE we tested were:

✓the CE value predicted by the default linear equation;
✓the CE value predicted by a new linear equation calculated from the training set;
✓the CE value that produced the maximum total peak area for the current precursor in the training set;
✓the CE value that produced the maximum peak area for the current transition in the training set.

In each of the 9 test sets, the peak areas corresponding to the calculated CE values were selected, separately for each approach, summed by precursor and normalized by the sum of the maximum measured areas for the precursor transitions in the test set. This process was repeated, using each of the 10 replicates as the training set, to produce a total of 90 trials for each peptide. The mean and standard deviation of the normalized area percentage values, for each approach, on each platform, are illustrated in following Figure.



Conclusions:

✓ In these experiments we have used a fully automated workflow, implemented in version 0.6 of the open source software Skyline, to evaluate the effect of optimizing CE values across multiple instrument platforms. By comparing peak areas at different CE values for different peptide sequences and precursor charge-states we have quantified the improvement in sensitivity obtained by optimizing each transition for each peptide individually compared to using a simple linear equation to predict CE based on the peptide precursor mass.

✓ We have derived optimized linear equations for predicting CE values that are different from the default equations recommended by the manufacturer or reported in the literature. These new equations for predicting CE are available in the latest release of Skyline.

✓ With well optimized linear equations, we have shown losses of 8.4% of total peak area on average compared with fully optimizing each transition.

✓ Understanding the variance in the transition peak area to CE relationship is critical to understanding how any peak area optimization using CE will perform. We found that low concentrations for certain peptides yielded measurements that were too noisy to support CE optimization. A similar effect occurs when using an optimal CE value determined with a standard at high abundance to measure a low abundance species in a real sample. Even if the measurement is made with the optimal CE, the improvement in sensitivity over a less optimal CE will be indistinguishable at low intensity because the variance from the measurement shot noise will overwhelm the small improvement in peak area.

✓ By simulating a differential proteomics experiment, where the only difference between groups of technical replicates was the method of choosing the CE values, we showed that the loss in peak signal intensity from using an optimized linear equation as opposed to optimizing each peptide was rarely detectable as significant in this type of experiment.

✓ Our data indicate that all linear equations should be derived empirically for each respective charge-state, rather than using linear equations derived from doubly charged precursors across charge-states.

✓ To achieve performance in the optimal range, we have shown linear equations should be reassessed on new instruments. We also found no evidence of significant variation in optimal CE between instruments of the same type, or even varying gas pressure on a single instrument.

A) charge 2 versus charge 3:

	Access 1.0	Ultra 1.0	Ultra 1.5	Vantage 1.0	Vantage 1.5	Q Trap A	Q Trap B	Q Trap B'
Access 1.0 mTorr	0.05							
Ultra 1.0 mTorr		0.21						
Ultra 1.5 mTorr			0.01					
Vantage 1.0 mTorr				0.004				
Vantage 1.5 mTorr					0.001			
Q Trap A						0.09		
Q Trap B							0.14	
Q Trap B'								0.01

B) charge 2 data set comparisons:

	Access 1.0	Ultra 1.0	Ultra 1.5	Vantage 1.0	Vantage 1.5	Q Trap A	Q Trap B	Q Trap B'
Access 1.0 mTorr	1							
Ultra 1.0 mTorr	0.73	1						
Ultra 1.5 mTorr	0.07	0.11	1					
Vantage 1.0 mTorr		0.02	0.02	1				
Vantage 1.5 mTorr			0.006	0.15	1			
Q Trap A				0.00007	0.00003	1		
Q Trap B					0.00002	0.12	1	
Q Trap B'						0.12	0.98	1

Significance tests of difference in linear regression coefficients between pairs of data sets, presenting p-values for the null hypothesis that two sample sets represent the same linear regression between precursor m/z and CE:

C) charge 3 data set comparisons.

	Access 1.0	Ultra 1.0	Ultra 1.5	Vantage 1.0	Vantage 1.5	Q Trap A	Q Trap B	Q Trap B'
Access 1.0 mTorr	1							
Ultra 1.0 mTorr	0.11	1						
Ultra 1.5 mTorr	0.42	0.03	1					
Vantage 1.0 mTorr		0.07	0.05	1				
Vantage 1.5 mTorr			0.17	0.21	1			
Q Trap A				0.03	0.07	1		
Q Trap B					0.04	0.17	1	
Q Trap B'						0.79	0.29	1

Selected References:

Picotti, P.; Bodenmiller, B.; Mueller, L. N.; Domon, B.; Aebersold, R. *Cell* **2009,** *138*, 795-806.

Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. *Nat.Methods* **2010,** *7*, 43-46.

Sherwood, C. A.; Eastham, A.; Lee, L. W.; Risler, J.; Mirzaei, H.; Falkner, J. A.; Martin, D. B. *J.Proteome.Res.* **2009,** *8*, 3746-51.

Acknowledgements:

Funding: This work was supported by a subcontract from Vanderbilt University under NIH/NCI grant number U24CA126479 and by grants to S.A. Carr under NIH/NCI grant number U24CA126476 as part of the National Cancer Institute's Clinical Proteomic Technologies Assessment in Cancer Program. Additional support was provided by NIH grant R01 HL082747, R01 DK069386, and by the University of Washington's Proteomics Resource (UWPR95794).