Supporting Information

Improved Precision of iTRAQ and TMT Quantification by an Axial Extraction Field in an Orbitrap HCD Cell

Peter Pichler[1*], Thomas Köcher[2], Johann Holzmann[2], Thomas Möhring[3], Gustav Ammerer[1,4], Karl Mechtler[2,5]

[1]Christian Doppler Laboratory for Proteome Analysis, University of Vienna, Austria
[2]Research Institute of Molecular Pathology, Vienna, Austria
[3]Thermo Fisher Scientific, Bremen, Germany
[4]Max F Perutz Laboratories, University of Vienna, Austria
[5]Institute of Molecular Biotechnology, Vienna, Austria

* corresponding author
Christian Doppler Laboratory for Proteome Analysis
Dr. Bohr-Gasse 3
1030 Vienna, Austria
Email: pichler@imp.univie.ac.at

Table of Contents

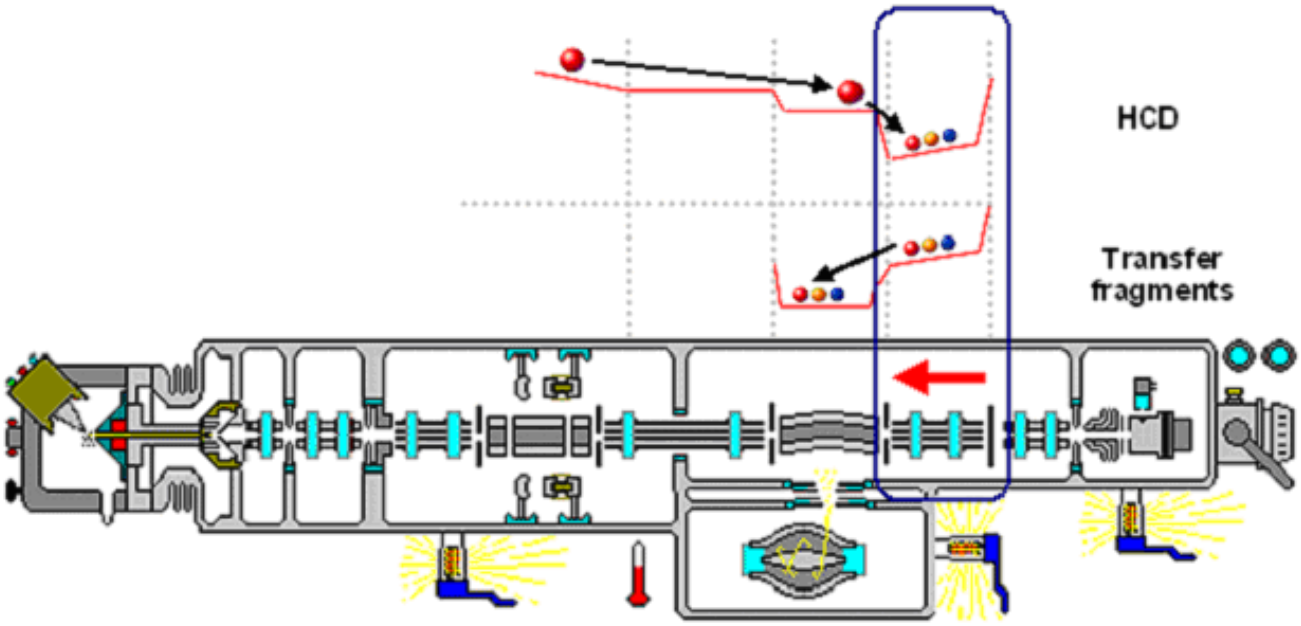

Figure S-1. Schematic diagram of HCD collision cell with axial field.

In LTQ Orbitrap XL ETD a version of the HCD collision cell is mounted where an axial electric field pushes fragment ions into the C-trap for improved ion transmission and read-out. A similar type of HCD collision cell is mounted in LTQ Orbitrap Velos instruments. Red lines indicate potential, red arrow indicates the axial extraction field in the HCD cell.

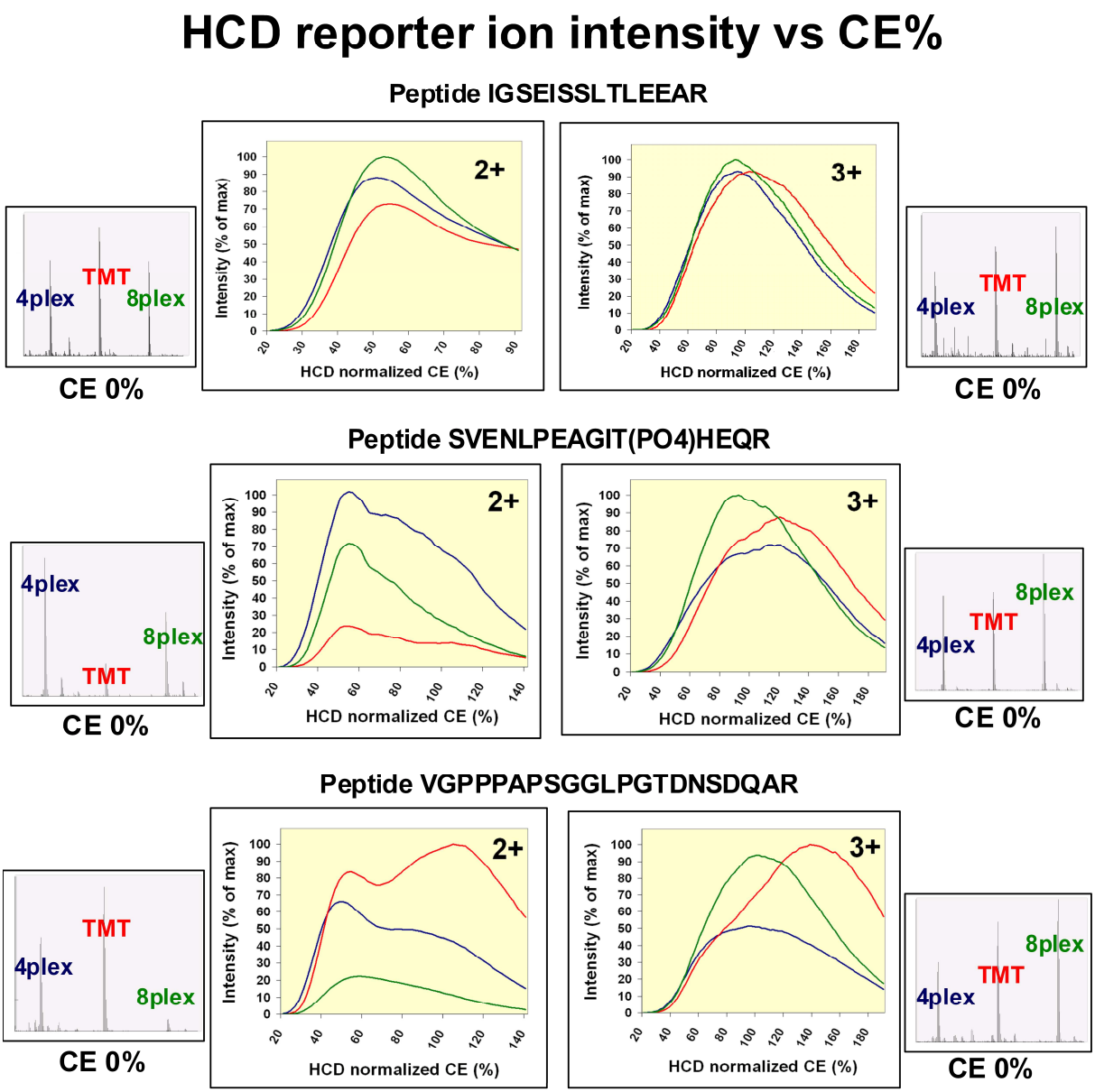

Figure S-2. Reporter ion intensity vs HCD collision energy CE%.

A mixture containing a peptide labeled with one "channel" of iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex, respectively, was analyzed by off-line nano-electrospray and simultaneous isolation of the three precursor ions with an appropriately wide precursor selection window while ramping the HCD collision energy. For each peptide, the normalized HCD collision energy vs reporter ion intensity profiles are shown for the 2+ and 3+ charge states. In addition, spectra obtained at a HCD collision energy of 0% CE are shown with an *m/z* window that includes the precursor ions. Data acquired on an LTQ Orbitrap equipped with the original HCD collision cell. Graphs were calculated by averaging ten measurements.

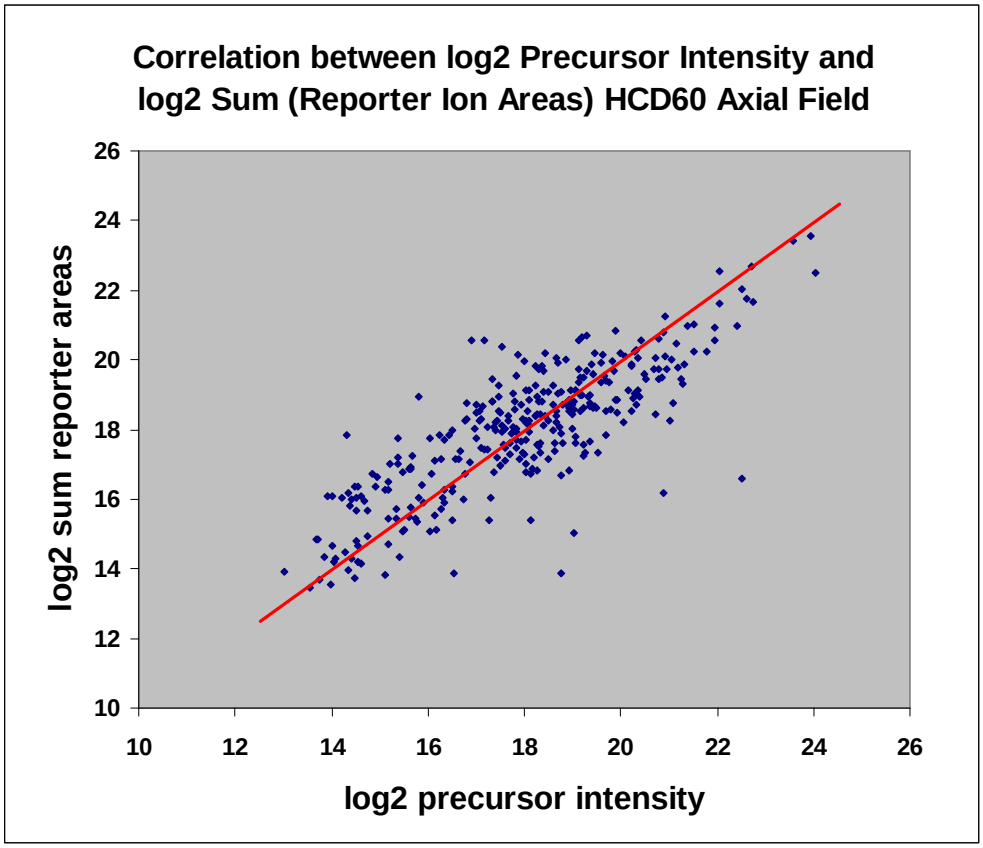

Figure S-3. Correlation between log_2 precursor intensity and log_2 sum (reporter ion areas) at HCD 60 using HCD cell with axial field.

Reporter ion areas from an iTRAQ 4-plex analysis of the protein mixture using a HCD collision energy of 60% were calculated as the sum of all centroid intensities within the integration window of each reporter ion. Precursor ion intensities were corrected for charge state and for the precursor selection window as described in the Additional Experimental Section.

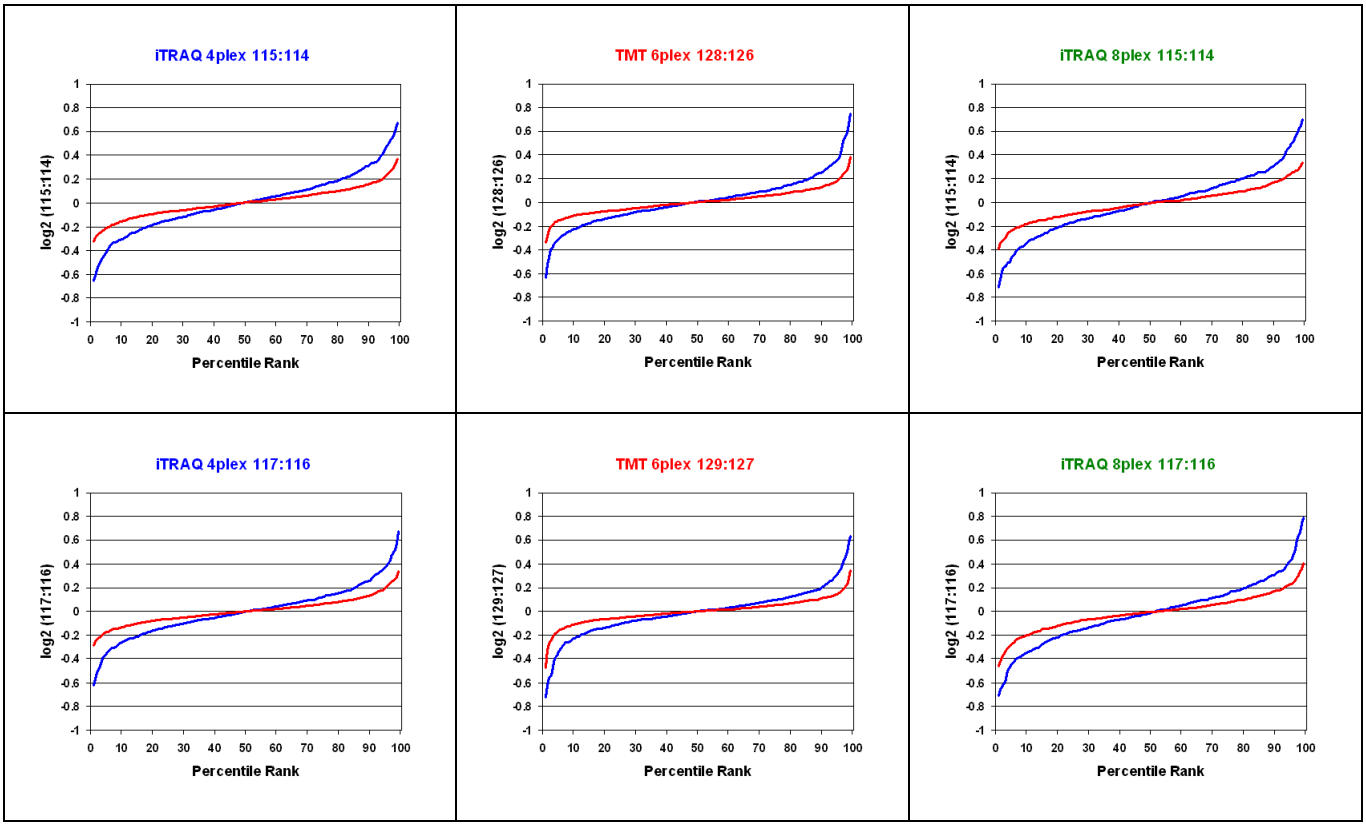

Figure S-4. An axial field improves analytical precision for all three types of tested isobaric labeling reagents.

Plots of the distribution of duplicate channel $\log_2$-ratios from HeLa samples labeled with iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex respectively. Duplicate channel ratios derived from split samples should be as close as possible to the x-axis which reflects the theoretical ratio of 1:1 on a logarithmic scale. The observed reduction of the deviation of duplicate channel ratios from $\log_2$ (1:1) = 0 reflects improved precision with the HCD cell with an axial field (red) as compared to the original HCD cell (blue). The effect was evident for all three types of isobaric labeling reagents.

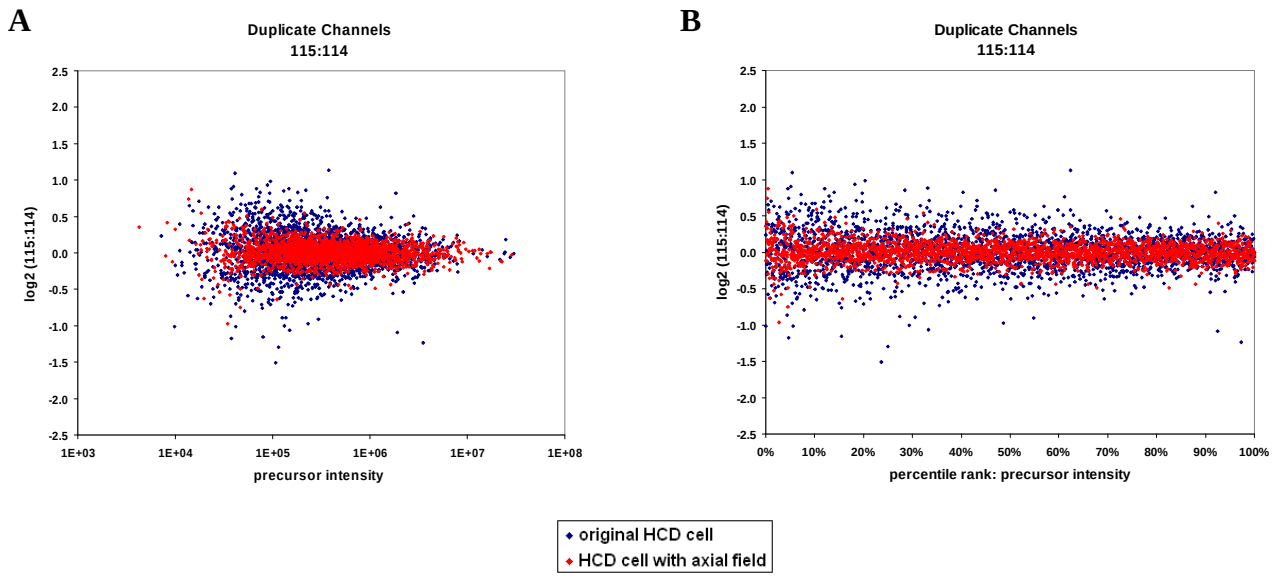

Figure S-5. Duplicate channel variation versus precursor intensity.

Scatter plots illustrating the variation of duplicate channel ratios 115:114 in relationship to the precursor ion intensity. Each data point represents a peptide-spectrum match. The $\log_2$ of duplicate channel ratios is plotted on the y-axis. In panel A the x-axis is proportional to $\log_{10}$ of the precursor intensity, whereas in panel B it is proportional to the percentile rank of the precursor intensity. Data are from the analyses of the iTRAQ 4-plex labeled HeLa sample (nocodazole sample split before labeling).

Expected number of outliers per 1000 PSMs					
cutoff	1.5	317	122	26	0
	2.0	87	8	0	0
		1.5	1.3	1.2	1.1
		Geometric Standard Deviation			

Table S-1. Expected number of outliers in relation to the geometric standard deviation.
Assuming a log-normal distribution of measurement error, the number of false positives with regard to regulation that arise as a consequence of measurement variance alone can be estimated. The expected number of outliers per 1000 unregulated (1:1) peptide-spectrum matches is given for two cutoffs namely 1.5-fold and 2-fold regulation. The number of outliers is highly elastic to ostensibly small changes in the geometric standard deviation.

Additional Experimental Section

For off-line nano-electrospray experiments, synthetic peptides were dissolved in 0.5 M TEAB at a concentration of 3 nmol/µl, split into equal parts, and labeled with one channel of iTRAQ 4-plex, TMT 6-plex and iTRAQ 8-plex according to manufacturer's protocol. Labeled peptides were mixed, diluted to a concentration of 0.75 pmol/µl in 50% ACN, 2% FA and analyzed immediately as follows: A Proxeon NanoES spray capillary containing the sample was mounted in lieu of the nano-electrospray ion source with spray voltage 0.8 kV and capillary temperature 200 °C. HCD spectra were acquired at a resolution of 7500 with target value 3E5, maximum inject time 400 ms, 3 microscans, activation time kept constant at 30 ms while ramping the HCD collision energy as indicated. Each measurement was repeated 10 times and the graphs in Figure S-2 illustrate averages of 10 such measurements extracted from raw files using Qual Browser 2.0.7 SP1 (Thermo Scientific).

To compare the efficiency of the generation of reporter ions by fragmentation of the precursor ion and any potential loss of reporter ions before reaching the Orbitrap mass analyzer, we estimated which fraction of the precursor ion current could be detected as reporter ion currents. In LTQ Orbitrap instruments, signals calculated from image current are stored in raw files after correction by dividing through the inject time (ion collection time). This corrects the values so that intensities in raw files reflect ion currents. Signals at or below the noise level are filtered from Orbitrap spectra so that this part of the ion current remains unknown. To calculate the sum of all reporter ion areas for a spectrum, reporter areas were determined as the sum of all centroid intensities within a +/- 5 mmu window around the theoretical *m/z* of the respective reporter ion. Precursor intensity was divided by charge state because image current signal strength in the Orbitrap analyzer is proportional to charge state.[1] In addition the monoisotopic C12 intensity was multiplied by a correction factor calculated individually for each precursor ion to take into account the selection of additional C13 isotope peaks within the precursor isolation window. For all HCD spectra where the corresponding CID scan led to a peptide identification, the sum of the reporter ion areas was plotted against the precursor ion intensity derived from the respective parent MS^1 spectrum after correction in the above described manner. In addition the ratio between the two quantities was calculated for each HCD spectrum, and the median of all such ratios was determined.

Additional Results and Discussion

Our analysis focused on the precision of HCD-based quantification because we consider low measurement variability one of the most critical factors for quantitative proteomics applications such as biomarker research or the search for regulated proteins involved in disease models. Quantifying hundreds or thousands of proteins means testing a high number of hypotheses, which requires appropriate statistical correction algorithms[2-5] to render the data useful for further studies. However, even excellent correction algorithms cannot fully compensate for deficiencies during data acquisition. Outliers arising because of the stochastic nature of the measurement process might be incorrectly classified as regulated targets. Improved analytical precision can help minimize such “false positives” (type I errors). This would reduce the false discovery rate (FDR) with regard to regulation or alternatively this would permit a lower cutoff for classifying peptides and proteins as regulated at an acceptable FDR. It is evident that the false discovery rate of a particular study can have an enormous impact on the feasibility of biological follow-up experiments or on the validity of diagnostic propositions. In addition, inadequate analytical precision would lead to type II errors i.e. to a decrease in statistical power, which means that a study has less chance to discover truly regulated proteins. One remedy in such a situation is to provide technical replicates, however in typical shotgun proteomics studies, these can be time-consuming, inexpedient or simply costly to obtain, given the low overlap of peptides and proteins detected in repeated analyses.[6, 7] For these reasons alone, high overall analytical precision is of paramount importance in quantitative proteomics.

Assuming that measurement error is log-normal distributed (which is often a good approximation and a viable assumption for relative quantitation in general and for iTRAQ data in particular),[8] one can predict the number of outliers for completely unregulated (1:1) peptides as a consequence of measurement variance alone: Supporting Information, Table S-1 displays the expected number of outliers per 1000 peptide-spectrum matches (PSMs) for two arbitrarily chosen cutoff values 1.5-fold and 2-fold, as detection of regulation at such levels is often considered desirable from a biological point of view. Notably, the expected number of outliers is highly elastic to ostensibly small changes in the geometric standard deviation: For a geometric standard deviation of 1.5, around 317 out of 1000 PSMs would be expected to be classified “pseudo-regulated" more than 1.5-fold and 87 out of 1000 PSMs would appear pseudo-regulated beyond 2-fold. As shot-gun proteomics experiments typically involve the detection and quantification of several thousands of peptides, it is evident that with such a high

variance, even requiring two data points (two PSMs) is unlikely to eliminate outliers at these cutoff levels. The numbers of outliers beyond the 2-fold cutoff drop to 8 out of 1000 PSMs for a geometric standard deviation of 1.3, and to 0.14 out of 1000 PSMs for a geometric standard deviation of 1.2. Finally the expected numbers of outliers are 0.02 and 0.0 out of 1000 PSMs for the cutoffs 1.5- and 2.0-fold respectively with a geometric standard deviation of 1.1. This highlights why we consider high analytical precision such a critical factor in quantitative proteomics.

A second important issue in quantification is accuracy. While precision concerns repeatability i.e. the closeness of repeated measurements to one other, accuracy reflects agreement of measured values with the true value. In case of iTRAQ experiments, accuracy has been reported to be impaired by co-eluting peptides within the precursor isolation window.[9, 10] As the majority of proteins in a differential proteomics experiment are usually unregulated, this background will in most situations lead to less extreme regulatory ratios and therefore to a compression of the observed dynamic range. In addition, it was shown that fragment ions with *m/z* values similar to those of the reporter ions can interfere with iTRAQ quantification.[10] Whereas the first problem could be reduced somewhat by chromatographic measures leading to better separation and by a reduction of the width of the precursor isolation window, the second problem can be minimized by extracting reporter ion ratios from high resolution mass spectra,[2, 10] an advantage that can be realized by the hybrid CID-HCD method due to the high resolution of the Orbitrap mass analyzer particularly in the low *m/z* region of the reporter ions.

Additional References

(1) Makarov, A. *Anal Chem* **2000**, *72*, 1156-1162.
(2) Zhang, Y.; Askenazi, M.; Jiang, J.; Luckey, C. J.; Griffin, J. D.; Marto, J. A. *Mol Cell Proteomics* **2010**, *9*, 780-790.
(3) Karp, N. A.; Huber, W.; Sadowski, P. G.; Charles, P. D.; Hester, S. V.; Lilley, K. S. *Mol Cell Proteomics* **2010**, *9*, 1885-1897.
(4) Hochberg, Y.; Benjamini, Y. *Stat Med* **1990**, *9*, 811-818.
(5) Storey, J. D.; Tibshirani, R. *Proc Natl Acad Sci U S A* **2003**, *100*, 9440-9445.
(6) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P.; Variyath, A.; Ham, A. J.; Bunk, D. M.; Kilpatrick, L.; Billheimer, D.; Blackman, R.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K.; Neubert, T. A.; Regnier, F.; Schilling, B.; Tegeler, T.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A.; Liebler, D. C.; Spiegelman, C. *J Proteome Res* **2009**, *9*, 761-776.
(7) Delmotte, N.; Lasaosa, M.; Tholey, A.; Heinzle, E.; van Dorsselaer, A.; Huber, C. G. *Journal of Separation Science* **2009**, *32*, 1156-1164.
(8) Hill, E. G.; Schwacke, J. H.; Comte-Walters, S.; Slate, E. H.; Oberg, A. L.; Eckel-Passow, J. E.; Therneau, T. M.; Schey, K. L. *J Proteome Res* **2008**, *7*, 3091-3101.
(9) Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. *Mol Cell Proteomics* **2008**, *7*, 1702-1713.
(10) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. *J Proteome Res* **2009**, *8*, 5347-5355.