Supplementary Information

Delayed fragmentation and optimized isolation width settings improve protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers

Mikhail M. Savitski¹, Gavain Sweetman¹, Manor Askenazi²,³, Jarrod A. Marto²,³, Manja Lang¹, Nico Zinn¹, Marcus Bantscheff¹,*

¹Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany

² Department of Cancer Biology and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115–6084

³ Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

* Corresponding author: marcus.bantscheff@cellzome.com
Supplementary Methods

Figure S-1: Distribution of fold-changes observed in the kinase enriched sample.

Figure S-2: Distribution of peak widths observed in an LC-MS/MS analysis of the kinase enriched sample using a 4 hr gradient.

Table S-1: Peptide and protein identifications achieved for the kinase enriched sample with and without delayed fragmentation (c=0.6).

Figure S-3: Correlation of measured target-to-signal ratios (TSR) with S2I values estimated from MS1 spectra.

Table S-2: Peptide and protein identifications achieved for the kinase enriched sample acquired with different isolation widths on an Orbitrap Velos mass spectrometer using a targeted data acquisition approach.

Figure S-4: Log2 ratios (of TMT 126 to TMT 131) of kinases in the replicate experiments performed A) with isolation width 1 Th and B) with isolation width 2.5 Th on an Orbitrap Velos.

Figure S-5: Log2 ratios (of TMT 126 to TMT 131) of kinases in the replicate experiments performed A) before and B) after optimization of isolation width and chromatography function on an Orbitrap Velos.

Figure S-6: Spike-in experiments of two standard proteins into the kinase sample.

Figure S-7: Scatter plot comparison of log2 protein ratios of spike-in experiments performed with standard and optimized settings.

Table S-3: Peptide and protein identifications achieved for the kinase enriched sample using conditions before (no_chr_f, 2.5 Th IW) and after optimization (c=0.6, 1 Th IW).

References
SUPPLEMENTARY METHODS

Mass spectrometry and protein identification

Experiments were performed on Thermo LTQ Orbitrap XL and Thermo LTQ Orbitrap Velos mass spectrometers (Thermofisher, USA) coupled to split-less Eksigent Nano-LCs (Eksigent, USA). Peptides were separated on a custom made 50 cm x 75uM (ID) reversed phase column (Reprosil, Maisch, Germany) and gradient elution was performed from 2% acetonitrile to 40% acetonitrile in 0.1% formic acid within 260 min using a flow rate of 250 nl/min, at 40 °C. The Orbitrap XL mass spectrometer was operated using XCalibur Developers kit 2.0.7. The Orbitrap Velos mass spectrometer was operated using XCalibur 2.1.0. Intact peptides were detected in the Orbitraps at 30,000 resolution. Internal calibration was performed using the ion signal of (Si(CH3)2O)6H+ at m/z 445.120025 as a lock mass. Automatic gain control was used to prevent over-filling of the ion trap and the minimum MS signal for triggering peptide fragmentation was set to 1000.

Tandem mass spectra were acquired using a CID/HCD dual scan approach described by Koecher et al. CID spectra were used for peptide identification and acquired in the LTQ ion trap using the following parameters: Target value LTQ: 1e4 ions, max LTQ fill time 100 ms, collision energy: 35%. Isolation width was 1 Th, 1.5 Th, 2 Th, or 2.5 Th depending on the experiment. TMT quantification was facilitated by subsequently acquired HCD spectra using the following settings: Target value FT: 5E4 ions for the Orbitrap XL and 2.5E4 for the Orbitrap Velos, collision energy: 70% or 1% as indicated in the results section, maximum FT fill time 350 ms, isolation width was 0.5 Th, 1 Th, 1.5 Th, 2 Th, 2.5 Th for HCD spectra and 2.5 Th or 1 Th for CID spectra, depending on the experiment. Fragment ions were detected in the Orbitrap at a resolution of 7500.
Peptide fragmentation was delayed using the chromatography function in XCalibur. We explored the following settings: correlation coefficient 0.6, 0.8, 1, 1.5 or phase value: 20, 30, 50, 80. Peak width was set to 30 seconds.

The raw data was processed with in-house developed software written in Python that utilizes some core elements from the open source Multiplierz Python-based environment. Mgf files were created and submitted to the Mascot search engine using 7.5 ppm precursor mass tolerance (monoisotopic mass) and 0.8 Th fragment ion mass.

The following modifications were selected as variable modifications: oxidation (methionine), TMT (N-term), Acetyl (Protein N-term); and as fixed modifications: TMT (Lysine), Carbamidomethyl (Cysteine). The instrument type was set to ‘ESI-TRAP’, the enzyme specificity to Trypsin/P allowing up to 3 missed cleavages. The data was searched against an in-house curated version of the human International Protein Index database combined with a decoy version thereof. Our database contains a total of 163,476 protein sequences (50% forward, 50% reverse) and represents a non-redundant composite of International Protein Index versions 1.0–3.54. Protein identification acceptance criteria for kinase enriched samples were based on spectrum to peptide sequence assignments that had a 10x higher Mascot probability than the second best match. For protein identifications with only one single peptide meeting these criteria, we required the Mascot score to be at least 33. Experiments with a decoy database indicate a <1% false discovery rate for protein identification. For identifications where multiple peptides met these criteria, the decoy search results indicate <0.1% false discovery rate on protein level.

**Chromatography function settings, phase and correlation**

To determine the peak apex with the correlation function the instrument software creates an XIC for each m/z over a certain threshold in the MS1. The XIC is correlated to the first half of a Gaussian function which is defined by peak width at FWHM. Then a weighting function based on the XIC correlation value is applied in order to obtain a current-weighted intensity which emphasizes mass spectral peaks that occur
at apexes of chromatographic peaks. Based on the current-weighted intensity for each \textit{m/z} a weighted mass spectrum is constructed. The final real-time decision for the data dependent acquisition is based on the area ratio defined as XIC intensity of the previous scan divided by the XIC intensity of the current scan. Thus a value of 1 should trigger the MS\textsuperscript{2} at the peak apex while smaller values shift the data acquisition prior to the apex and higher after the apex.

The phase apex detection creates an XIC of each \textit{m/z} just like the correlation function but the XIC intensity during the set trigger window which is equal to a percentage of the expected peak width is Fourier transformed into the frequency domain. When the frequency domain data matches a phase criterion indicating a peak apex an MS\textsuperscript{2} is triggered \textsuperscript{8}. Since the phase is essentially independent of abundance, this method should work equally well for large and small chromatographic peaks.

\textit{XIC peak extraction and Full Width Half Maximum, FWHM calculation}

The chromatographic peak was determined by first generating an extracted ion chromatograms (XIC) corresponding to the precursor \textit{m/z} (as determined by the Xcalibur acquisition software, called monoisotopic MZ by Xcalibur). The XIC was generated by extracting all ions within a 10 ppm window from the precursor \textit{m/z}. Peaks were found by smoothing the intensity data using a 3 point mean smooth and the local maxima were detected, the half maximum intensity points were then determined to provide a measure of the peak width (full width, half maximum, FWHM). Additional XICs were created corresponding to isotope cluster ions. Peaks were detected in these XICs and matched to peaks from the precursor \textit{m/z} XIC. Isotope peaks were considered matching if the portion of the peak above half maximum intensity overlapped with the same region of the precursor peak. Any peak from the precursor XIC without isotopic data was discarded as were peaks where the intensity pattern of the isotopes didn’t match to the theoretical intensity data calculated using poly-averagine. The MS\textsuperscript{2} event was then assigned to the most intense of the remaining peaks whose apex retention time was within 30 s of the time of the MS/MS event.
**Determination of signal-to-interference (S2I) and precursor abundance-to-threshold (P2T) values**

Underestimation of fold-changes in iTRAQ/TMT experiments depends on the purity of the isolated and subsequently fragmented precursor ion. To estimate the purity of isolated peptide ions directly from MS\(^1\) spectra we have recently introduced the S2I measure\(^9,10\). Briefly, for each tandem mass spectrum, in-house software was used to identify the precursor ion and the respective m/z range within the isolation width in corresponding MS\(^1\) scans and to then determine precursor abundance by integrating signal intensities of the precursor ion and its isotopes within this m/z range. S2I was then calculated by dividing precursor abundance by the sum of all ion signals observed within the isolation window. Consequently, values close to one indicate little and values close to zero a high degree of interference caused by co-eluting components. In the previously described procedure we calculated S2I values only from the preceding MS\(^1\) scan, S2I\(_{E}\). Here, we have elaborated the calculation by also determining the S2I value from the MS\(^1\) scan following the MS\(^2\) (late scan, termed S2I\(_{L}\)), and calculating the extrapolated, S2I value as a time weighted linear combination of the two:

\[
S2I = \left( RT_M - RT_E \right) \left( \frac{S2I_L - S2I_E}{RT_L - RT_E} \right) + S2I_E
\]

where RT\(_M\) denotes the retention time of the MS\(^2\) event, and RT\(_E\) and RT\(_L\) are the retention times of the early and late survey scans, respectively.

The XCalibur software automatically applies a noise filter in order to remove all (electronic and chemical noise) ions below an intensity cutoff of 2.4 standard deviations of all detected signals within several segments across the spectrum (personal communication with Thermofisher Scientific). These noise threshold values are stored in the raw file and are accessible to software extraction (e.g. via the freely supplied XDK). For precursor ions with abundances very close to this noise threshold S2I necessarily becomes very inaccurate. Precursor abundance-to-threshold ratios (P2T) were determined by dividing
precursor ion abundance for each acquired tandem mass spectrum by the noise threshold. Extrapolated P2T values were calculated in the same way as described for S2I:

\[
P2T = (RT_M - RT_E) \left( \frac{P2T_E - P2T_L}{RT_L - RT_E} \right) + P2T_E
\]

Where \( RT_M \) is the retention time of the MS\(^2\) spectrum, \( RT_E \) is the retention time of the preceding MS\(^1\) scan. \( P2T_E \) is the precursor abundance-to-threshold ratio determined from the preceding MS\(^1\) scan and \( P2T_L \) is the corresponding value calculated from the next MS\(^1\) scan.
Figure S-1: Distribution of fold-changes observed with the kinase enriched sample using an isolation width of 2.5 Th without delayed fragmentation using the chromatography function. Binned ratios of TMT126-130 over TMT131 are displayed on a log2 scale. Out of the 246 proteins quantified in the sample only 27 (11%) display a larger than two-fold change for TMT126 (red line), whereas the vast majority of proteins do not change significantly. Hence, co-fragmentation will in almost all cases lead to fold-changes closer to one.
Figure S-2: Distribution of peak widths observed in an LC-MS/MS analysis of the kinase enriched sample using a 4 hr gradient.

Average peak width: 29 s
Table S-1: Peptide and protein identifications achieved for the kinase enriched sample with and without delayed fragmentation (c=0.6).

<table>
<thead>
<tr>
<th></th>
<th>no_chr_f</th>
<th></th>
<th>c=0.6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>replicate 1</td>
<td>replication 2</td>
<td>replicate 1</td>
<td>replicate 2</td>
</tr>
<tr>
<td>CID spectra</td>
<td>19502</td>
<td>19386</td>
<td>17827</td>
<td>17707</td>
</tr>
<tr>
<td>PSM</td>
<td>4942</td>
<td>4950</td>
<td>5017</td>
<td>4945</td>
</tr>
<tr>
<td>unique peptides</td>
<td>3332</td>
<td>3317</td>
<td>3377</td>
<td>3337</td>
</tr>
<tr>
<td>proteins</td>
<td>361</td>
<td>335</td>
<td>363</td>
<td>354</td>
</tr>
<tr>
<td>kinases</td>
<td>114</td>
<td>114</td>
<td>118</td>
<td>117</td>
</tr>
</tbody>
</table>
Figure S-3: Correlation of measured target-to-signal ratios (TSR) with S2I values estimated from MS1 spectra. A: S2I values were estimated solely from the MS1 scan directly preceding fragmentation ($R^2 = 0.48$), B: Refined determination of S2i values using a linear interpolation of MS1 scans before and after the MS2 scan yields near unity slope with experimentally determined TSR ($R^2 = 0.62$).
Table S-2: Peptide and protein identifications achieved for the kinase enriched sample acquired with different isolation widths on an Orbitrap Velos mass spectrometer using a targeted data acquisition approach.

<table>
<thead>
<tr>
<th></th>
<th>Isolation width 1 Th</th>
<th></th>
<th>Isolation width 2.5 Th</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Replicate 1</td>
<td>Replicate 2</td>
<td>Replicate 1</td>
<td>Replicate 2</td>
</tr>
<tr>
<td>acquired spectra</td>
<td>8418</td>
<td>8488</td>
<td>8631</td>
<td>8512</td>
</tr>
<tr>
<td>kin spectra</td>
<td>3484</td>
<td>3435</td>
<td>3541</td>
<td>3515</td>
</tr>
<tr>
<td>unique kinase</td>
<td>997</td>
<td>969</td>
<td>987</td>
<td>981</td>
</tr>
<tr>
<td>kinases</td>
<td>129</td>
<td>127</td>
<td>129</td>
<td>132</td>
</tr>
<tr>
<td>kinases &gt; 3</td>
<td>125</td>
<td>125</td>
<td>126</td>
<td>125</td>
</tr>
<tr>
<td>quantifiable spectra</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure S-4

Figure S-4: Log2 ratios (of TMT 126 to TMT 131) of kinases in the replicate experiments performed A) with isolation width 1 Th and B) with isolation width 2.5 Th on an Orbitrap Velos.

Figure S-5

Figure S-5: Log2 ratios (of TMT 126 to TMT 131) of kinases in the replicate experiments performed A) before and B) optimization of isolation width and chromatography function on an Orbitrap Velos.
Figure S-6:

A) The measured ratios of TMT 126-130 relative to TMT 131 for the YOL086C protein. Green circles represent measurements from the “pure”, not spiked in sample (YOL086C and HBA1) run with 1 Th isolation window and chromatography function set to c=0.6, error bars represent 95% bootstrap confidence intervals (22 peptide sequence matches). Blue circles represent protein ratios obtained from “spike in” sample (kinase enriched sample as background) run with 1 Th isolation window and chromatography function set to c=0.6 (9 peptide sequence matches). Red circles represent protein ratios obtained from the “spike in” sample run with 2.5 Th isolation window and no chromatography function (10 peptide sequence matches). Dashed lines correspond to the expected TMT ratios. The six TMT colors samples were mixed to achieve the following proportions: 126:0, 127:1, 128:2, 129:4, 130:8, 131:1 B) same as A on the log2 scale. C) same as A for the HBA1 protein, 13 peptide sequence matches for pure sample, 9 peptide sequence matches for “spike in” sample run with 1 Th isolation width and c=0.6, and 8 peptide sequence matches for “spike in” sample run with 2.5 Th isolation width and no chromatography function. D) same as C on the log2 scale.
Figure S-7

Figure S-7: Scatter plot comparison of the log2 ratios for TMT 126-130 vs TMT 131 of the two proteins, YOL086C and HBA1, where the x-axis values correspond to measurements using the 1 Th isolation window and chromatography function setting c=0.6, and the y-axis values correspond to measurements with 2.5 Th isolation window and no chromatography function.
Table S-3: Peptide and protein identifications achieved for the kinase enriched sample using conditions before (no_chr_f, 2.5 Th IW) and after optimization (c=0.6, 1 Th IW).

<table>
<thead>
<tr>
<th></th>
<th>Before optimization</th>
<th></th>
<th>After optimization</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>replicate 1</td>
<td>replicate 2</td>
<td>replicate 1</td>
<td>replicate 2</td>
</tr>
<tr>
<td>CID spectra</td>
<td>17182</td>
<td>19083</td>
<td>13698</td>
<td>14958</td>
</tr>
<tr>
<td>PSM</td>
<td>3973</td>
<td>4297</td>
<td>3686</td>
<td>3996</td>
</tr>
<tr>
<td>unique peptides</td>
<td>2542</td>
<td>2791</td>
<td>2775</td>
<td>2993</td>
</tr>
<tr>
<td>proteins</td>
<td>247</td>
<td>293</td>
<td>322</td>
<td>320</td>
</tr>
<tr>
<td>kinases</td>
<td>101</td>
<td>109</td>
<td>115</td>
<td>109</td>
</tr>
</tbody>
</table>
References


