Published in final edited form as:

Anal Chem. 2017 March 07; 89(5): 2956–2963. doi:10.1021/acs.analchem.6b04415.

Extensive Peptide Fractionation and y_1 Ion-based Interference Detection Enable Accurate Quantification by Isobaric Labeling and Mass Spectrometry

Mingming Niu^{†,‡,§}, Ji-Hoon Cho^{†,§}, Kiran Kodali[‡], Vishwajeeth Pagala[‡], Anthony A. High[‡], Hong Wang^{†,‡}, Zhiping Wu[†], Yuxin Li[†], Wenjian Bi[§], Hui Zhang[§], Xusheng Wang[‡], Wei Zou^{||}, and Junmin Peng^{†,‡,*}

[†]Departments of Structural Biology and Developmental Neurobiology, Memphis, TN 38105, USA

[‡]St. Jude Proteomics Facility, Memphis, TN 38105, USA

[§]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

^{||}Heilongjiang University of Chinese Medicine, Harbin, Heilongjiang, 150040, China

[‡]Integrated Biomedical Sciences Program, University of Tennessee Health Science Center, Memphis, TN 38163, USA

Summary

Isobaric labeling quantification by mass spectrometry (MS) has emerged as a powerful technology for multiplexed large-scale protein profiling, but measurement accuracy in complex mixtures is confounded by the interference from co-isolated ions, resulting in ratio compression. Here we report that the ratio compression can be essentially resolved by the combination of pre-MS peptide fractionation, MS2-based interference detection and post-MS computational interference correction. To recapitulate the complexity of biological samples, we pooled tandem mass tag (TMT) labeled *E. coli* peptides at 1 : 3 : 10 ratios, and added in ~20-fold more rat peptides as background, followed by the analysis of two dimensional liquid chromatography (LC)-MS/MS. Systematic investigation show that quantitative interference was impacted by LC fractionation depth, MS isolation window and peptide loading amount. Exhaustive fractionation (320 × 4 h) can nearly eliminate the interference and achieve results comparable to the MS3-based method. Importantly, the interference in MS2 scans can be estimated by the intensity of contaminated y_1 product ions, and we thus developed an algorithm to correct reporter ion ratios of tryptic peptides. Our data indicate that intermediate fractionation (40 × 2 h) and y_1 ion-based correction allow accurate and deep TMT profiling of more than 10,000 proteins, which represents a straightforward and affordable strategy in isobaric labeling proteomics.

Graphical abstract

*Corresponding Author: junmin.peng@stjude.org (J. P.).

#These authors contributed equally.

Author Contributions: The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes: The authors declare no competing financial interest.

Keywords

proteomics; mass spectrometry; LC-MS/MS; isobaric labeling; TMT; iTRAQ; and ratio compression

Introduction

Quantitative proteomics is an essential tool in biomedical research^{1,2} and shows high potential for clinical application³. The integration of liquid chromatography and tandem mass spectrometry (LC-MS/MS) is the mainstream approach for global measurement of proteins and posttranslational modifications. Numerous MS strategies have been developed for large-scale profiling, including label free quantification and stable isotope labeling technologies⁴. More recently, isobaric labeling methods, such as isobaric tags for relative and absolute quantitation (iTRAQ)⁵, tandem mass tags (TMT)⁶ and DiLeu isobaric tags⁷, have gained popularity largely due to multiplexed capacity of processing up to 12 samples⁸. For example, isobaric labeling enables the analysis of hundreds of protein samples, detecting a total of more than 15K proteins (from 12K genes) and 60K phosphosites in mammalian samples⁹⁻¹¹.

Despite the advances of isobaric labeling, the method often suffers from high noise levels due to co-eluted interfering ions, leading to quantitative ratio compression that underestimates the difference, particularly in complex protein samples¹²⁻¹⁶. This drawback is ameliorated by some proposed approaches, which may be classified into three categories: pre-MS fractionation, MS setting modification, and post-MS correction. While pre-MS fractionation (e.g. 2D LC) partially reduced the co-elution of interfering peptides¹⁷, 3D LC of basic pH reversed-phase (RP), strong anion exchange and acidic pH RPLC¹⁸, yielded a more efficient platform for peptide separation. However, the 3D LC platform involves complex settings that are not commonly used in other labs. The co-eluting ions can also be reduced by narrow MS2 isolation window¹⁵, gas-phase purification¹⁹ and complement reporter ion cluster quantification during MS analysis²⁰, usually at the expense of decreased identification of peptides and proteins. Some post-MS corrections have also been reported by subtracting interference to enhance quantitation²¹⁻²³. Moreover, multistage MS3-based technique has been developed to nearly eliminate the ratio compression, but has slightly lower sensitivity to detect weak peptide ions and requires expensive MS instrumentation^{9,15,16,24,25}. Although all of these approaches improve quantitative accuracy, the application is still limited by instrument dependency, time consumption, and computer algorithm availability.

In this study, we seek to address the ratio compression issue by extensive high resolution fractionation and a novel y_1 ion-based interference correction method. To mimic real biological samples, we mixed TMT-labeled *E.coli* proteins at known ratios, in the presence of a 20-fold excess amount of background peptides from rat proteins. The mix was analyzed under multiple LC-MS/MS conditions by adjusting key parameters, including fraction number collected in the offline pre-fractionation, MS2 isolation window, peptide loading amount and online RP fractionation depth (gradient length). We also developed a computational method that used the known *E.coli* protein ratios to estimate interference levels from rat proteins. Finally, the interference can be essentially eliminated by pre-MS fractionation, optimization of MS parameters, and post-MS y_1 ion-based correction, leading to a general pipeline for accurate isobaric labeling quantification.

Experimental Section

Preparation of *E. coli* and rat protein samples

Proteins in *E. coli* cells or adult rat brain were extracted as previously described²⁶ in lysis buffer (50 mM HEPES, pH 8.5, 8 M urea, 0.5% sodium deoxycholate), and digested in two steps (1:100 w/w Lys-C, Wako, for 2 h, diluted 4-fold and incubated with 1:50 w/w Trypsin, Promega, overnight at room temperature). Protein concentration was measured by the BCA method (Thermo Scientific) and confirmed by Coomassie-stained short SDS gel²⁷. After digestion, the peptides were desalted, resuspended in HEPES buffer (50 mM, pH 8.5) and labeled by individual 10-plex MT reagents (Thermo Scientific, *E.coli* peptides by 10 channels, rat peptides by 8 channels from 126 to 130N) following manufacturer's instruction^{28,29}. Finally, these peptides were then mixed as specified (Figure 1), desalted again and dried.

Basic pH LC prefractionation

The analysis was performed with a previously reported protocol³⁰ with slight modification. The pooled TMT-labeled sample was solubilized and fractionated on a long reverse phase column (concatenated two Waters 4.6 mm $\times$ 25 cm Xbridge C18 columns, totaling 50 cm, 3.5 μ m beads, Agilent 1270 HPLC, flow rate of $\sim$ 0.4 ml/min). The gradient included 5 min of 95% buffer A (10 mM ammonium, pH 8.0), 215 min of 13%-35% buffer B (10 mM ammonium and 90% acetonitrile, pH 8.0), 90 min of 35-55% buffer B, and 10 min of 55-95% buffer B. A total of 320 fractions (one min each) were collected, dried and stored at -80°C for further analysis.

Acidic pH LC-MS/MS analysis

Dried peptides were dissolved in 0.2% formic acid for LC-MS/MS analysis on an optimized platform^{28,31,32} with modifications. Peptides were separated on a 75 μ m $\times$ $\sim$ 50 cm column (1.9 μ m C18 beads, Dr. Maisch GmbH) and operated at 70°C to reduce back pressure (solvent A: 0.2% formic acid, solvent B: 0.2% formic acid and 70% acetonitrile, 240 min gradient from 12-65% solvent B unless specified). The analysis used an Ultimate 3000 RSLCnano system coupled with an Orbitrap Fusion mass spectrometer (Thermo Scientific). The Orbitrap Fusion acquired data in a data-dependent manner alternating between full scan MS and MS2 scans. The MS spectra (400-1600 m/z) were collected with 60,000 resolution,

AGC of 2×10^5 and 50 ms maximal injection time. Selected ions were sequentially fragmented in a 3 sec cycle by HCD with 38% normalized collision energy, specified isolated windows (0.4 m/z , 0.8-1.6 m/z , 0.3 m/z offset), 60,000 resolution. AGC of 1×10^5 and 150 ms maximal injection time. Dynamic exclusion was set to 20 sec.

For MS3 analysis¹⁶, the precursors for MS2 analysis were isolated with a 1.6 m/z window (0.3 m/z offset). The CID-MS2 spectra were acquired in the ion trap with AGC of 1×10^4 , 50 ms maximal injection time and 35% normalized collision energy. For HCD-MS3 mode of the strongest MS2 ion, the setting was 65% normalized collision energy, 2 m/z isolated windows (0.3 m/z offset), 60,000 resolution. AGC of 1×10^5 and 500 ms maximal injection time. Dynamic exclusion was set to 30 sec.

Protein/peptide identification and quantification

The analysis was performed by our recently developed JUMP engine to improve sensitivity and specificity, which combines the advantages of pattern matching with *de novo* sequencing during database search^{33,34}. The JUMP hybrid algorithm was used to process numerous published large datasets^{32,34-36}. RAW files were converted to mzXML format and MS2 spectra were searched against rat and *E.coli* target-decoy Uniprot databases to estimate false discovery rate (FDR)^{37,38}. Search parameters included precursor and product ion mass tolerance (6 ppm), fully tryptic restriction, two maximal missed cleavages, static TMT modification (+229.162932 Da on N-termini and Lys residues), dynamic Met oxidation (+15.99492 Da), and three maximal dynamic modification sites. Only *a*, *b*, and *y* ions were considered during the search. Peptide-spectrum matches (PSMs) were filtered by 7 minimal peptide length, mass accuracy (~ 2.5 ppm) and matching scores to achieve 1% protein FDR. For each accepted PSM, the peaks of TMT reporter ions were extracted for quantification (Supporting Information, Figure S1-S2).

Quantitative data analysis and post-MS computational correction approach

To evaluate the levels of interference, TMT reporter ion intensities of each PSM were converted into relative intensities. For rat peptides that were equally mixed, the relative intensities were calculated by dividing individual reporter ion intensity by the mean intensity of eight reporters (126-130N). For *E. coli* peptides of three groups with known ratio, (126, 128C) : (127N, 128N, 129N, 130N) : (127C, 129C) to be 1 : 3 : 10 (Figure 1), the relative intensities were converted by dividing each channel intensity by the mean intensity of 126 and 128C. Then the relative intensity of each group was averaged in two steps:

For each PSM,

$$m_{1,i} = \frac{m_{126,i} + m_{128C,i}}{2}, m_{2,i} = \frac{m_{127N,i} + m_{128N,i} + m_{129N,i} + m_{130N,i}}{4} \text{ and } m_{3,i} = \frac{m_{127C,i} + m_{129C,i}}{2},$$

where $m_{\text{reporter},i}$ and $m_{g,i}$ represent relative intensities of a reporter channel and the *g*-th group, respectively, for the *i*-th PSM.

Then for each group,

$\overline{m}_g = \sum_{i=1}^N m_{g,i}$, where $\overline{m}_g$ is the mean by averaging all PSM relative intensities; and N is the total number of PSMs.

Finally, the group mean was used to calculate interference level by comparison to the expected ratio of the three groups ($r_1 = 1$, $r_2 = 3$ and $r_3 = 10$, Supporting Information).

We also developed a post-MS computational approach to correct interference based on y_1 ion in MS2 scans. As K-TMT- and R-C-terminal tryptic peptides generate different y_1 ions (376.27574 Da and 175.11895 Da, respectively). If only one y_1 ion is detected and consistent with the identified peptide, the MS2 is termed a clean scan. If both y_1 ions are detected, the MS2 is termed a noisy scan. Assuming that the y_1 ion intensity is proportional to the reporter ion intensity, we computed their linear relationship from the clean scans, and then used the contaminated y_1 ion intensity in the noisy scans to derive the interference level (Supporting Information, Figure S3-S7).

Results and Discussion

Generation of a cross-species peptide mix to mimic complex biological samples

As ratio compression in isobaric labeling is largely influenced by sample complexity, we attempted to replicate the complexity of real biological samples by mixing cross-species peptides from *E. coli* and rat brains in a 10-plex TMT experiment (Figure 1). The *E. coli* peptides were labeled by 10 TMT reagents with known ratios (1 : 3 : 10 : 3 : 1 : 3 : 10 : 3 : 1 : 10), while rat peptides were labeled in 8 channels and mixed equally as background, leaving 2 channels without interference.

In the vast majority of comparative proteomic studies, proteins of high abundance are usually expressed from house-keeping genes and show no significant alteration under experimental conditions, whereas proteins with differing expression patterns are likely to play regulatory roles and exist at low abundance. To simulate this scenario, we markedly increased the levels of background rat peptides, approximately 20-fold more than the targeted *E. coli* peptides, although in many of previous reports^{24,39}, the background and targeted peptides were pooled at comparable amounts.

This cross-species peptide mix was used to systematically dissect the effect on ratio compression in three major steps, including pre-MS fractionation, MS settings and post-MS correction. The pre-MS fractionation was carried out by the combination of basic pH RPLC and acidic pH RPLC. During post-MS analysis, peptides shared between *E. coli* and rat were removed, and only species-specific peptides were considered (Figure 1).

Confirmation of ratio compression and interference computation by known *E. coli* peptide ratios

We initially analyzed the pooled peptide sample by one dimensional (1D) LC-MS/MS (Figure 2). As expected, rat peptide intensities in the 8 channels were almost equal, suggesting that these rat peptide measurements were not significantly affected by the *E. coli* peptides of low abundance (Figure 2A). In contrast, a clear ratio compression was observed

for the *E. coli* peptides in the 8 channels in the presence of ~20-fold background rat peptides, but not in the 2 channels without rat peptides (i.e. 130C and 131, Figure 2B). Averaged ratios of 1 : 3 : 10 *E. coli* channels were found to be 1 : 1.6 : 4.1 (Figure 2C), consistently with previously reported ratio compression effect¹⁴⁻¹⁶.

We then developed a method to calculate the level of interference based on known peptide ratios. Assuming the level of interference was stable among channels, defined as the percentage of maximum reporter ion intensity (e.g. 10, Figure 2D), if there was no interference, the relative intensities were 1, 3 and 10; if the interference was 20% ($10 \times 20\% = 2$), the relative intensities increased to 3, 5 and 12, resulting in compressed ratios (1 : 1.7 : 4.0); and so on. Thus, standard curves were generated to describe the relationship between experimental ratios and interference, with respect to different theoretical ratios (Figure 2E). With this standard curve, experimental and theoretical ratios, the interference could be calculated. For example, when experimental 1 : 4 ratio was detected for theoretical 1 : 10, we concluded that the interference was about 20% of maximum reporter ion intensity. If there were multiple experimental ratios (e.g. 1 : 1.7 : 4) for theoretical ratios (e.g. 1 : 3 : 10), the overall level of interference could be derived by summarizing multiple comparisons (Supporting Information).

Interference level affected by core LC/LC-MS/MS parameters

With the interference computation method, we examined the effects of a number of core parameters in LC/LC-MS/MS, including the resolution of the offline high pH RPLC, MS2 isolation window, the loading amount on the online acidic pH RPLC, and the online LC resolution. To change the resolution during the offline LC, we separated the complex TMT labeled peptide mixture into 320 fractions, and then combined some of the fractions together to adjust the separation power. For instance, combination of every 4 adjacent fractions would yield 80 fractions, and so on. Thus, the offline LC resolution was reflected by different number of collected fractions (1, 5, 10, 20, 40, 80, and 320), under which the interference levels decreased gradually from 16.4% to 2.8%, suggesting that extensive fraction in the offline LC alleviated co-eluted peptides but could not remove the interference completely (Figure 3A).

As to the online RPLC-MS/MS analysis, we examined the effect of MS2 isolation window and found that the interferences were almost proportional to the size of the isolation window (Figure 3B), in agreement with previous studies^{15,23}. For example, 4-fold difference of window size (1.6 to 0.4 Da) resulted in ~4-fold difference of inference level (14.4% to 3.7%). We also observed a visible impact of peptide loading amount on the interference. When the loading level decreased from 900 ng to 100 ng, the interference was reduced from 9.4% to 3.3%, implying that high loading might lead to peak broadening²⁷ and therefore raised the interference level (Figure 3C). Finally, we adjusted the online RPLC resolution by gradient length (1, 2, and 4 h) on a long column. The 4 h gradient nearly eliminated the interference (down to 0.4%) and gave the best result (averaged ratios of 1 : 2.8 : 9.9, Figure 3D). This result was comparable with that of the accurate MS3 strategy¹⁵ (the interference of 0.6%, averaged ratios of 1 : 2.8 : 9.7, Figure 3E), although the MS3 analysis may have slightly lower sensitivity, requires more MS acquisition time and uses low-resolution MS2

for protein identification². In general, this MS3 analysis (without the function of synchronous precursor selection³⁹) identified ~20% less proteins than the MS2 method, similar to reported estimation¹⁶. Taken together, our data demonstrate that the combination of extreme fractionation ($320 \times 4 \text{ h} = 1,280 \text{ h}$, 53 days) and narrow isolation window is able to solve the issue of ratio compression.

We also analyzed the effects of these core parameters on peptide/protein identification in the analysis (Table 1). As anticipated, pre-MS fractionation reduced sample complexity, leading to low peptide/protein identification per fraction. However, if all fractions were analyzed, the combined number of identifications would be improved^{18,32}. Interestingly, the reduction of MS2 isolation window had minor effect on peptide/protein identification, which may be due to high isolation efficiency of quadrupole mass filters that were installed in newly developed MS instruments (e.g. Q Exactive HF)⁴⁰. The effect of sample loading was also relatively small, as even 100 ng of fractionated peptides (one of 80 fractions, Table 1, Figure 3C) contained concentrated peptides from the initial 8,000 ng total peptides ($80 \times 100 \text{ ng}$). When some of the 320 fractions were analyzed by different gradient time, the number of peptides detected were similar from 1 h to 2 h, but decreased in 4 h, likely because the long gradient resulted in peak broadening and decreased the sensitivity of identifying weak peptides²⁷.

In summary, we could eliminate ratio compression by utilizing extreme separation power (LC/LC fractionation and narrow isolation window), but this strategy required >50 days instrument time. To compromise the effects of these core parameters on the interference and protein identification, we recommend the final setting of narrow isolation window (0.4 Da), medium fractionation ($\sim 40 \times 2 \text{ h}$, 3.3 days) and $\sim 100 \text{ ng}$ of fraction sample loading, which resulted in a low level of interference (3.4%, Figure 3A).

Interference correction by y_1 ion-based post-MS method

We devised a post-MS computational approach to calculate and remove the interference based on the information in MS2 scans, in contrast to MS1-based correction methods^{21,22}. In the MS1-based strategies, the intensity of co-eluted peptides is estimated from the precursor isolation window in MS1 scans, but these MS1 survey scans are acquired at different time points from the MS2 scans during real time online LC-MS/MS analysis. Therefore, the exact co-eluted peptides in MS2 are not directly measured. In contrast, our method directly analyzes the level of interference in the same MS2 scan for reporter ion quantification. When examining TMT MS2 scans of tryptic peptides (Figure 4A), we found that some MS2 scans (clean scans) displayed only one y_1 ion (K-TMT or R residue) consistent with the matched peptide sequences. The other MS2 scans (noisy scans) had the two y_1 ions, indicative of contaminated peptides. The contaminated y_1 ion could be generated from numerous co-eluting peptides of low abundance in the MS2 isolation window, even if these weak peptides may not be detected in corresponding MS1 scans, suggesting that this y_1 ion-based method is a direct, sensitive approach to monitor co-eluting TMT-labeled ions.

In the clean scans, the y_1 ion intensity tended to be proportional to the measured TMT reporter intensities because they originated from the same precursor ions. Therefore, the relationship between intensities of either K-TMT- or R- y_1 ion and TMT reporter ions was

modeled as a linear form (Supporting Information Figure S3). During the analysis, we found that this linear relationship was dependent on peptide charge state and the number of TMT labels (e.g. equivalent to the number of amine groups) on the peptides. As exemplified in one dataset, although the TMT reporter ion intensities were globally correlated with y_1 ion intensities, the correlation was poor ($R^2 = 0.38$, Figure S4A). By contrast, when dividing the peptide ions into multiple groups based on charge state and TMT labels, the correction was significantly improved ($R^2 = 0.7-0.8$, Figure S4B). Thus we derived the y_1 – TMT reporter relationship in these different groups (Figure S5). In noisy cans, this y_1 – TMT reporter relationship enable the calculation of the interference level based on the contaminated y_1 ion, which is subtracted for correcting the compressed ratios (Figure 4A, Figure S3).

The performance of this post-MS correction approach was evaluated in the datasets of varying LC resolution (e.g. different fractions in the offline LC), and showed unanimous improvement in quantitative precision (Figure 4B). The percentage of clean scans was affected by the separation power (e.g. low in 1-fraction and high in 320-fraction). In the case of 40-fraction, 55% of accepted PSMs are clean scans. As expected, the clean scans showed highly accurate quantification with a very low interference level, centered around 9 for expected ratio of 10 (Figure S7). The noisy scans displayed ratio compression, centered around 7 for expected ratio of 10. After post-MS correction, the noisy scan results appeared to be highly similar the clean scan data (Figure S7), suggesting that this method is effective. The global interference level decreased from 3.4% to 1.3%, with corrected ratio of 1 : 2.7 : 9.1 (Figure 4B), nearly reaching the expected ratio of 1 : 3 : 10). It should be noted that the post-MS correction alone is not sufficient to release ratio compression, particularly for the condition of highly limited separation (e.g. 1-fraction, Figure 4B). It is important to combine the extensive fractionation and post-MS correction to achieve the desired outcome.

As this systematic optimization process was highly time consuming, we only ran representative fractions under each condition to expedite the process. Finally, we used the recommended condition to analyze all 40 offline fractions (Figure 5A), resulting in the acceptance of 363,672 PSMs, 86,878 peptides and 10,379 proteins (~1% protein false discovery rate). The dataset contained 11,181 PSMs, 6,284 peptides and 1,506 proteins from *E.coli* with measured ratio of 1 : 2.8 : 8.7 and interference level of 1.5% (Figure 5B), consistent with our previous analysis. The histograms of detected *E.coli* peptides show distributions that are centered close to the expected ratios of 3 and 10, respectively (Figure 5C). The analysis indicate that the use of our optimized LC-MS parameters and post-MS correction provide deep proteomic analysis and almost eliminate ratio compression commonly observed in the TMT-based quantification.

Conclusion

Multiplex isobaric labeling provides an efficient mass spectrometry technology for quantitative proteomics, but a common limitation of ratio compression leads to quantitative inaccuracy and often constrains its application. We demonstrate that the optimization of LC/LC-MS/MS settings, in combination of y_1 ion-based post-MS correction, is capable of virtually removing the effect of interference and substantially enhancing the precision of measurements. The extensive LC/LC fraction also allows deep profiling of proteome and

protein modifications. Although we only analyzed TMT-labeled samples in this study, the principle of y_1 ion-based correction is anticipated to be applicable to all other isobaric labeling approaches for analyzing trypsinized proteome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all other lab and facility members for helpful discussion. This work was partially supported by National Institutes of Health grants R01GM114260, R01AG047928, R01AG053987, and ALSAC (American Lebanese Syrian Associated Charities). The MS analysis was performed in the St. Jude Children's Research Hospital Proteomics Facility, partially supported by NIH Cancer Center Support Grant (P30CA021765).

References

1. Aebersold R, Mann M. *Nature*. 2016; 537:347–355. [PubMed: 27629641]
2. Yates JR, Ruse CI, Nakorchevsky A. *Annu Rev Biomed Eng*. 2009; 11:49–79. [PubMed: 19400705]
3. Rifai N, Gillette MA, Carr SA. *Nat Biotechnol*. 2006; 24:971–983. [PubMed: 16900146]
4. Altelaar AF, Munoz J, Heck AJ. *Nat Rev Genet*. 2013; 14:35–48. [PubMed: 23207911]
5. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. *Mol Cell Proteomics*. 2004; 3:1154–1169. [PubMed: 15385600]
6. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C. *Anal Chem*. 2003; 75:1895–1904. [PubMed: 12713048]
7. Frost DC, Greer T, Li L. *Anal Chem*. 2015; 87:1646–1654. [PubMed: 25405479]
8. Rauniyar N, Yates JR 3rd. *J Proteome Res*. 2014; 13:5293–5309. [PubMed: 25337643]
9. Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D, Sun S, Yang F, Chen L, Wang J, Shah P, Cha SW, Aiyetan P, Woo S, Tian Y, Gritsenko MA, Clauss TR, Choi C, Monroe ME, Thomas S, Nie S, Wu C, Moore RJ, Yu KH, Tabb DL, Fenyo D, Bafna V, Wang Y, Rodriguez H, Boja ES, Hiltke T, Rivers RC, Sokoll L, Zhu H, Shih Ie M, Cope L, Pandey A, Zhang B, Snyder MP, Levine DA, Smith RD, Chan DW, Rodland KD, Investigators C. *Cell*. 2016; 166:755–765. [PubMed: 27372738]
10. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatza ML, Wilkerson M, Perou CM, Yellapantula V, Huang KL, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, Ding L, Paulovich AG, Fenyo D, Ellis MJ, Carr SA, Nci C. *Nature*. 2016; 534:55–62. [PubMed: 27251275]
11. Chick JM, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP. *Nature*. 2016; 534:500–505. [PubMed: 27309819]
12. Bantscheff M, Boesche M, Eberhard D, Matthieson T, Sweetman G, Kuster B. *Mol Cell Proteomics*. 2008; 7:1702–1713. [PubMed: 18511480]
13. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. *J Proteome Res*. 2009; 8:5347–5355. [PubMed: 19754192]
14. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS. *Mol Cell Proteomics*. 2010; 9:1885–1897. [PubMed: 20382981]
15. Savitski MM, Sweetman G, Askenazi M, Marto JA, Lang M, Zinn N, Bantscheff M. *Anal Chem*. 2011; 83:8959–8967. [PubMed: 22017476]
16. Ting L, Rad R, Gygi SP, Haas W. *Nat Methods*. 2011; 8:937–940. [PubMed: 21963607]
17. Ow SY, Salim M, Noirel J, Evans C, Wright PC. *Proteomics*. 2011; 11:2341–2346. [PubMed: 21548092]

18. Zhou F, Lu Y, Ficarro SB, Adelmant G, Jiang W, Luckey CJ, Marto JA. *Nat Commun.* 2013; 4:2171. [PubMed: 23863870]
19. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ. *Nat Methods.* 2011; 8:933–935. [PubMed: 21963608]
20. Wuhr M, Haas W, McAlister GC, Peshkin L, Rad R, Kirschner MW, Gygi SP. *Anal Chem.* 2012; 84:9214–9221. [PubMed: 23098179]
21. Savitski MM, Mathieson T, Zinn N, Sweetman G, Doce C, Becher I, Pachi F, Kuster B, Bantscheff M. *J Proteome Res.* 2013; 12:3586–3598. [PubMed: 23768245]
22. Sandberg A, Branca RM, Lehtio J, Forshed J. *J Proteomics.* 2014; 96:133–144. [PubMed: 24211767]
23. Ahrne E, Glatter T, Vigano C, Schubert C, Nigg EA, Schmidt A. *J Proteome Res.* 2016; 15:2537–2547. [PubMed: 27345528]
24. Erickson BK, Jedrychowski MP, McAlister GC, Everley RA, Kunz R, Gygi SP. *Anal Chem.* 2015; 87:1241–1249. [PubMed: 25521595]
25. Liu JM, Sweredoski MJ, Hess S. *Anal Chem.* 2016; 88:7471–7475. [PubMed: 27377715]
26. Na CH, Jones DR, Yang Y, Wang X, Xu Y, Peng J. *J Proteome Res.* 2012; 11:4722–4732. [PubMed: 22871113]
27. Xu P, Duong DM, Peng JM. *J Proteome Res.* 2009; 8:3944–3950. [PubMed: 19566079]
28. Bai B, Tan H, Pagala VR, High AA, Ichhaporis VP, Hendershot L, Peng J. *Methods Enzymol.* 2017; 585:377–395. [PubMed: 28109439]
29. Pagala VR, High AA, Wang X, Tan H, Kodali K, Mishra A, Kavdia K, Xu Y, Wu Z, Peng J. *Methods in molecular biology.* 2015; 1278:281–305. [PubMed: 25859956]
30. Wang H, Yang Y, Li Y, Bai B, Wang X, Tan H, Liu T, Beach TG, Peng J, Wu Z. *J Proteome Res.* 2015; 14:829–838. [PubMed: 25455107]
31. Tan H, Wu Z, Wang H, Bai B, Li Y, Wang X, Zhai B, Beach TG, Peng J. *Proteomics.* 2015; 15:500–507. [PubMed: 25307156]
32. Mertz J, Tan HY, Pagala V, Bai B, Chen PC, Li YX, Cho JH, Shaw T, Wang XS, Peng JM. *Mol Cell Proteomics.* 2015; 14:1898–1910. [PubMed: 25931508]
33. Wang X, Li Y, Wu Z, Wang H, Tan H, Peng J. *Mol Cell Proteomics.* 2014; 13:3663–3673. [PubMed: 25202125]
34. Li Y, Wang X, Cho JH, Shaw TI, Wu Z, Bai B, Wang H, Zhou S, Beach TG, Wu G, Zhang J, Peng J. *J Proteome Res.* 2016; 15:2309–2320. [PubMed: 27225868]
35. Hanna JA, Garcia MR, Go JC, Finkelstein D, Kodali K, Pagala V, Wang X, Peng J, Hatley ME. *Cell Death Dis.* 2016; 7:e2256. [PubMed: 27277678]
36. Churchman ML, Low J, Qu C, Paietta EM, Kasper LH, Chang Y, Payne-Turner D, Althoff MJ, Song G, Chen SC, Ma J, Rusch M, McGoldrick D, Edmonson M, Gupta P, Wang YD, Caulfield W, Freeman B, Li L, Panetta JC, Baker S, Yang YL, Roberts KG, McCastlain K, Iacobucci I, Peters JL, Centonze VE, Notta F, Dobson SM, Zandi S, Dick JE, Janke L, Peng J, Kodali K, Pagala V, Min J, Mayasundari A, Williams RT, Willman CL, Rowe J, Luger S, Dickins RA, Guy RK, Chen T, Mullighan CG. *Cancer Cell.* 2015; 28:343–356. [PubMed: 26321221]
37. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. *J Proteome Res.* 2003; 2:43–50. [PubMed: 12643542]
38. Elias JE, Gygi SP. *Nat Methods.* 2007; 4:207–214. [PubMed: 17327847]
39. McAlister GC, Nusinow DP, Jedrychowski MP, Wuhr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. *Anal Chem.* 2014; 86:7150–7158. [PubMed: 24927332]
40. Scheltema RA, Hauschild JP, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M. *Mol Cell Proteomics.* 2014; 13:3698–3708. [PubMed: 25360005]

Abbreviations

SILAC

stable isotope labeling with amino acids in cell culture

iTRAQ

isobaric tags for relative and absolute quantification

TMT

tandem mass tags

DiLeu isobaric tags, N

N-dimethyl leucine isobaric tags

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Experimental design and procedures for evaluating TMT analysis

Digested rat and *E. coli* peptides were TMT-labeled and mixed at known ratios, fractionated by basic pH RPLC, and analyzed by acidic pH RPLC-MS/MS. The interference of the TMT analysis was assessed by computational approaches, including a novel y_1 ion-based correction method.

Figure 2. Interference analysis of TMT data based on known peptide ratios

(A) The averaged relative intensities of rat peptides in each TMT channel. (B) The averaged relative intensities of *E. coli* peptides in all TMT channels. (C) The summed relative intensities of *E. coli* peptides in three groups: lowest level (126 and 128C), medium level (127N, 128N, 129N and 130N) and highest level (127C and 129C). (D) Schematic diagram showing the interference definition and its effect on ratio compression. The interference is defined as the proportion of maximal reporter intensity. If the interference is equal to 5% of maximal intensity ($5\% \times 10 = 0.5$) in each TMT channel, the intensities are elevated, resulting in ratio compression. The cases of interference of 20% and 50% are also shown. (E) Given theoretical ratios between reporters, the interference can be inferred by experimental ratios. For example, when theoretical and measured ratios are 1 : 10 and 1 : 4 respectively, the interference should be ~20% of maximal reporter intensity (Supporting Information).

Figure 3. Effects of LC-MS parameters on interference in the TMT analysis

The summed relative intensities of *E. coli* peptides in three groups under different conditions of (A) the offline dimensional LC for pre-fractionation, (B) precursor isolation window, (C) peptide loading amount, (D) the online RPLC gradient, and (E) MS3 method with ~500 ng loading and 2 h gradient. The interference level under each condition was computed from the known 1 : 3 : 10 ratios.

Figure 4. Post-MS computational approach for interference removal

(A) Schematic diagram. MS2 scans of tryptic peptides can be divided into clean and noisy scans based on the detected y_1 ions. The relationship between y_1 ions and TMT reporter ion intensities is modeled by the clean scans. Then the relationship is used to calculate the interference in the noisy scans (Supporting Information). (B) The summed relative intensities of *E. coli* peptides in the three groups, before and after the post-MS computational correction.

Figure 5. Quantification of more than 10,000 proteins in one deep proteomics analysis
(A) The procedure of whole proteome analysis using the optimized settings (40 offline fractions, ~100 ng of fraction sample loading for 2 h gradient LC-MS/MS, and 0.4 *m/z* isolation window for MS/MS). (B) Comparison of expected and measured intensities of identified *E. coli* peptides, before and after post-MS correction. (C) Histogram of quantified *E. coli* peptides. Dashed lines indicate expected relative intensities on a log2 scale.

Table 1
Identification summary of different MS analyses¹

MS parameters	Estimated interference level	#PSMs	#Peptides	#Unique Proteins
Offline LC resolution (number of fractions collected, 100 ng, 2 h run, 0.4 m/z isolation)				
1	16.4%	28142±710	15590±84	2389±10
5	9.7%	9274±149	6789±60	2112±5
10	7.2%	6699±874	4523±508	1859±109
20	4.7%	4759±381	2354±204	1437±136
40	3.4%	3830±151	1824±105	1270±98
80	3.3%	3423±110	1524±59	1022±168
320	2.8%	2449±9	1006±13	853±13
Isolation window (<i>m/z</i> , 100 ng, 1 h run, 320 fractions)				
1.6	14.4%	1865±94	844±37	569±972
1.0	9.7%	1806±248	842±87	566±47
0.4	3.7%	1765±53	856±1	571±7
Amount of loading peptides (ng, 2 h run, 0.4 m/z isolation, 80 fractions)				
900	9.4%	4815±59	1671±3	1018±5
300	5.8%	4492±69	1665±26	998±12
100	3.3%	3951±13	1401±9	1022±168
Gradient length (hours, 100 ng, 0.4 m/z isolation, 320 fractions)				
1	3.7%	2432±15	1319±14	571±7
2	2.8%	2671±444	1106±157	853±13
4	0.4%	1630±240	563±66	495±59

¹The analyses were usually based on duplicates of one representative fraction.