



computational proteomics

## Laboratory for Computational Proteomics

[www.FenyoLab.org](http://www.FenyoLab.org)

E-mail: [Info@FenyoLab.org](mailto:Info@FenyoLab.org)

Facebook: [NYUMC Computational Proteomics Laboratory](#)

Twitter: [@CompProteomics](#)



## Detection and correction of interference in SRM analysis<sup>☆</sup>

Y. Bao<sup>a,b</sup>, S. Waldemarson<sup>c</sup>, G. Zhang<sup>c</sup>, A. Wahlander<sup>c</sup>, B. Ueberheide<sup>d</sup>, S. Myung<sup>e</sup>, B. Reed<sup>e</sup>, K. Molloy<sup>e</sup>, J.C. Padovan<sup>e</sup>, J. Eriksson<sup>f</sup>, T.A. Neubert<sup>c,d</sup>, B.T. Chait<sup>e</sup>, D. Fenyö<sup>a,d,\*</sup>

<sup>a</sup> Laboratory of Computational Proteomics, Center for Health Informatics and Bioinformatics, New York University School of Medicine, New York, NY, USA

<sup>b</sup> Stevens Institute of Technology, Hoboken, NJ, USA

<sup>c</sup> Kimmel Center for Biology and Medicine at the Skirball Institute, New York University School of Medicine, New York, NY, USA

<sup>d</sup> Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY, USA

<sup>e</sup> Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA

<sup>f</sup> Department of Chemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden

### ARTICLE INFO

#### Article history:

Available online 23 May 2013

#### Keywords:

Mass spectrometry  
Selected reaction monitoring  
Interference detection  
Interference correction  
Linear range detection

### ABSTRACT

Selected Reaction Monitoring (SRM) is a method of choice for accurate quantitation of low-abundance proteins in complex backgrounds. This strategy is, however, sensitive to interference from other components in the sample that have the same precursor and fragment masses as the monitored transitions. We present here an approach to detect interference by using the expected relative intensity of SRM transitions. We also designed an algorithm to automatically detect the linear range of calibration curves. These approaches were applied to the experimental data of Clinical Proteomic Tumor Analysis Consortium (CPTAC) Verification Work Group Study 7 and show that the corrected measurements provide more accurate quantitation than the uncorrected data.

© 2013 The Authors. Published by Elsevier Inc. All rights reserved.

### 1. Introduction

Selected Reaction Monitoring (SRM) mass spectrometry has increasingly been used to develop assays for precise quantitation of low-abundance proteins in complex biological matrices [1], [2]. The technique was introduced in the late 1970s [3] for analyzing small molecules [4], and its application for protein quantitation has been an active research topic during the past decade. Fig. 1 below shows the schematic work-flow of SRM-MS assays for protein quantitation. Proteins are extracted from biological samples and enzymatically digested. Heavy-isotope versions of the peptides of interest are added to the samples, and the mixture is analyzed by Liquid Chromatography Mass Spectrometry (LC-MS). Most typically a triple quadrupole mass spectrometer [5] is used for SRM assays. In a triple quadrupole mass spectrometer, quadrupole 1 (Q1) and quadrupole 3 (Q3) serve as mass filters for selecting precursor ions and fragment ions with  $m/z$  values that define the transitions that are monitored and quadrupole 2 (Q2) is used as a collision cell for fragmenting the peptides. Because only the selected transitions

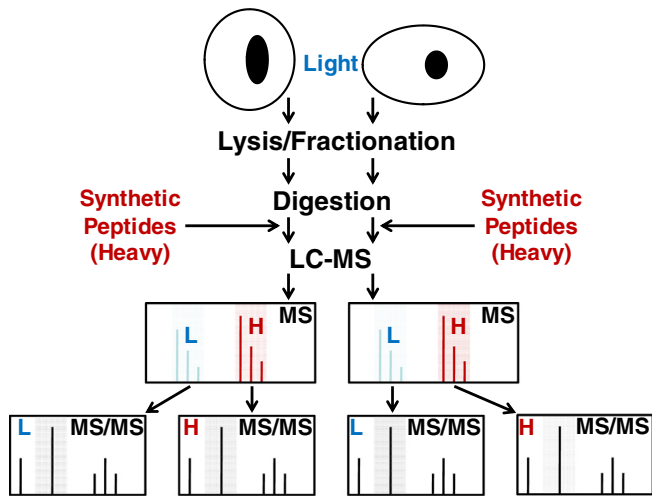
are monitored instead of acquiring the entire tandem mass spectra, the speed of analysis is very fast. During the sample preparation and measurement, there is always variation in the experimental conditions that is difficult to control. Therefore, isotopically labeled reference peptides [6,7] are often used as internal standards in SRM assays in order to reduce the effects of this variation [8]. These isotopically labeled reference peptides are created by replacing a 4–10 atoms of <sup>12</sup>C and <sup>14</sup>N in a selected amino acid (e.g. Arginine, Lysine, Valine, Isoleucine) with <sup>13</sup>C and <sup>15</sup>N, respectively. The isotopically labeled version of the analyte (Fig. 1) behaves identically to the analyte peptides in the different experimental steps except that in the mass spectrometer they are distinguishable based on the masses of the precursor and fragment ions [9–14].

Over the past decade, SRM has been the technology of choice for accurate quantitation of selected peptides in complex biological samples. However, the frequent occurrence of interferences is a significant problem which causes inaccurate quantitation of peptides [15,16]. The key factor contributing to this problem is that other components in a sample may have the same precursor and fragment masses as the monitored transitions. The interference problem only gets worse with some of the newer strategies that attempt to collect fragmentation information on all peptides simultaneously by using wider isolation windows that allow scanning of the mass range of interest in a data independent fashion [17–20]. With the current generation of AB Sciex time-of-flight instruments, the smallest size of isolation window that can be accommodated in these data independent fragmentation strategies

<sup>☆</sup> This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

\* Corresponding author at: Laboratory of Computational Proteomics, Center for Health Informatics and Bioinformatics, New York University School of Medicine, New York, NY, USA.

E-mail address: [David@FenyoLab.org](mailto:David@FenyoLab.org) (D. Fenyö).



**Fig. 1.** Schematic work flow of quantitation by SRM coupled with stable isotope dilution (SID). The heavy labeled peptides are shown in red and the light ones in blue.

is approximately 20–25 Da [18] – in most such cases many peptides will be isolated and fragmented together, and new software is needed for effective analysis of these mixed spectra. The Thermo Scientific Orbitrap-based Q Exactive now allows the use of discontinuous isolation windows (e.g.  $5 \times 4$  Da windows). If these 5 discontinuous windows are selected at random, each window will be grouped with different windows each time, and this makes it possible to deconvolute the mixed spectra [21] more easily. The advantage of these data independent acquisition methods that use wide isolation windows is that they enable data collection without making decisions during data acquisition, thus making the measurements more versatile. However, the price paid for this versatility is an increase in potential interferences.

Currently, manual inspection is typically used to identify the interferences. To address the time-intensive and error-prone nature of manual inspection, we have developed an algorithm to automatically detect and correct the interference using the expected relative intensity of SRM transitions. Previous studies [11,15,22,23] have devised methods to detect interferences in SRM assays. In one of these methods, AuDIT [15], the relative ratio of the analyte and the relative ratio of stable isotope-labeled internal standard are compared to detect interference, and the method requires the use of a stable isotope-labeled internal standard. The novel aspects of the present approach include that it can be applied to experiments where stable isotope standards are not used. Our method detects outliers in the relative intensity of SRM transitions and automatically detect the linear range of the measurements to detect interferences, which can be easily understood and implemented compared with using multiple properties of SRM in the algorithm. The threshold for the outliers were determined by computational simulation.

## 2. Materials and methods

### 2.1. Experimental data

The experimental data were taken from CPTAC Verification Work Group Study 7 [16] (<http://cptac.tranche.proteomecommons.org/study7.html>). This data set contains results from a multi-laboratory study designed to assess the performance of SRM assays. Eight laboratories measured 10 peptides in human plasma in the concentration range of 1–500 fmol/μl, and corresponding stable isotope labeled internal standard peptides were

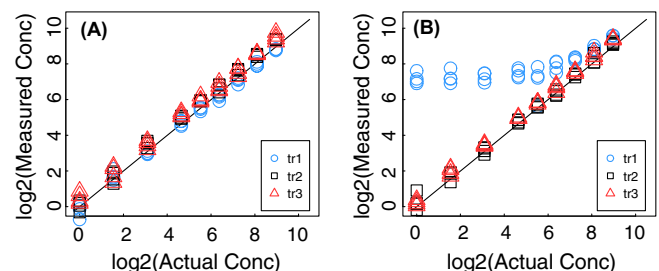
added to each sample. Three SRM transitions for each peptide were monitored. Each analysis was repeated 4 times at each laboratory. Fig. 2 shows examples of experimental data for two peptides: (A) a peptide with no apparent interference (i.e. for all transitions the measured and actual concentrations are close), and (B) a peptide with interference in one of its monitored transitions (i.e. the measured concentrations are higher than the actual concentrations, especially for low concentrations).

### 2.2. Interference detection

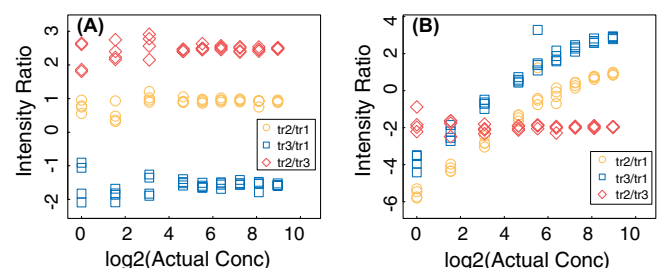
In the absence of interference, the relative intensity of different transitions for a peptide is a property of the peptide sequence and the mass spectrometric method (the energy and method of fragmentation and the analysis time), and it is independent of peptide concentration. Therefore, we use the relative intensities of transition to detect interference. In Fig. 3 the relative intensities are shown for the two peptides in Fig. 2. In the absence of interference, the ratios are constant, with increasing noise at lower amounts. In contrast, when there is interference for one transition, the corresponding ratios deviate from the expected (constant) ratios. Based on these observations, an approach was developed to detect the interference by comparing the ratio of the intensity of pairs of transitions with the expected ratio from which a Z-score for the deviation is calculated:

$$Z_i = \max_{i \neq j} Z_{ji} = \max_{i \neq j} \frac{r_{ij} - I_j/I_i}{\sigma_{ji}} \quad (1)$$

where  $I_j$  is the measured log intensity of transition  $j$ ,  $I_i$  is the measured log intensity of transition  $i$ ,  $r_{ji}$  is the expected transition ratio, which is equal to the median of transition ratios from the measurements of all different concentrations of one peptide and  $\sigma_{ji}$  is the



**Fig. 2.** Experimental data: The measured concentration as a function of the actual concentration for two different peptides with three transitions and 4 technical replicates. The three transitions are shown as blue circles, black squares, and red triangle for transitions tr1, tr2, and tr3, respectively. (A) Example of a peptide with no interference. (B) Example of peptide with interference in transition one.



**Fig. 3.** Relative intensities of transitions: The relative intensity as a function of the actual concentration for two different peptides with three transitions and 4 technical replicates (same peptides as in Fig. 2). The three ratios are shown as yellow circles, blue squares, and red diamonds for transitions tr2/tr1, tr3/tr1, and tr2/tr3, respectively. (A) In the absence of interference, the ratios are constant, with increasing noise at lower concentrations. (B) With interference in one transition, the corresponding ratios deviate from the expected constant ratios.

standard deviation of relative intensities calculated from the four repeated analyses for each concentration of samples of each peptide.  $Z_{ji}$  represents the number of standard deviations that the ratio between the intensities of transitions  $j$  and  $i$  deviate from the expected transition ratio. If interference occurs to transition  $i$ , it will cause an increase in  $I_i$ . Hence, the worse interference is, the larger  $Z_{ji}$  is. The maximum function guarantees that the transition which has the largest interference is detected. A transition is considered to have no interference only when it does not have outliers of intensity ratios, i.e.  $Z_i$  is smaller than a threshold  $Z_{th}$ .

### 2.3. Threshold estimation

To maximize the accuracy of quantitation, a threshold,  $Z_{th}$ , needs to be selected which balances the need for removal of interference with the need for retaining data to obtain sufficient statistics. Computer simulations were used to investigate the effect on the quantitation of the noise and the interference detection threshold. It was assumed that the data set had no interference but did have noise. The expected relative intensity of the transitions for the sample was set to be 1:1:1 and the relative noise was assumed to be 0.2 and normally distributed. 10,000 simulations were executed by adding noise to transitions randomly based on the noise rate. Interference detection was performed on the simulated data (even though there was no interference but only noise). Fig. 4 below shows the results of computer simulations using the approach from Section 2.2 above that detects interference by measuring the deviation of the intensity ratios of transitions from the expected ratios.  $Z_{th}$  equal to infinity corresponds to no interference detection (black curve with centroid equal to zero since there is only noise and no interference). If a low threshold (e.g.  $Z_{th} = 0$ ) is applied, the distribution will be distorted since high intensity values are selectively removed. As the value of  $Z_{th}$  is increased from 0 to larger numbers, the distortion of the distribution decreases (Fig. 4A) and the centroid (Fig. 4B) moves closer to zero. In Supplementary Fig. 1.1 ([http://fenyolab.org/data/12\\_bao\\_methods/12\\_bao\\_methods\\_suppl\\_fig1.pdf](http://fenyolab.org/data/12_bao_methods/12_bao_methods_suppl_fig1.pdf)), more noise ratios and relative intensities are simulated. The three noise ratios: 0.2, 0.4 and 0.6 are colored and shaped differently. The simulation shows the centroid and width of corrected relative error distributions as different thresholds are selected under the assumption that there is no interference. The result is consistent with Fig. 4. Supplementary Figs. 1.2–1.7 show the effect of selecting different interference removal thresholds when there is both noise and interference. Based on these simulations, an interference detection threshold corresponding to two standard deviations of the noise was selected as a reasonable trade-off between interference detection sensitivity and specificity.

### 2.4. Linear range detection

In cases when a calibration curve has been obtained for an SRM assay it is useful to automatically find the linear range of the assay.

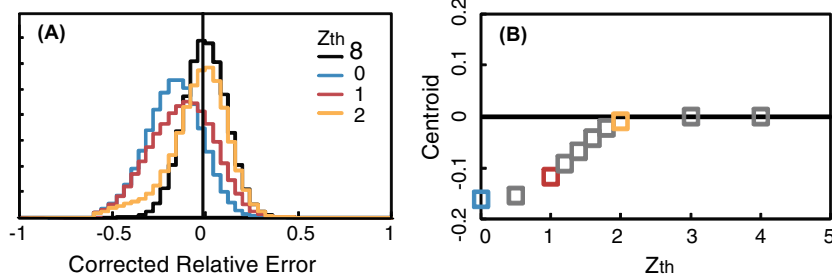


Fig. 4. Simulation for interference detection threshold. (A) Relative error distributions for threshold  $Z_{th} = \infty$ ,  $Z_{th} = 0$ ,  $Z_{th} = 1$  and  $Z_{th} = 2$ , respectively. (B) The distribution centroid as a function of threshold.

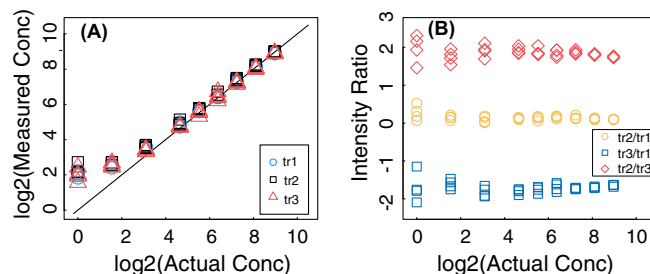


Fig. 5. Example of peptide where the concentration dependence deviates from linear for low concentrations. (A) Intensities of transitions. (B) Relative intensities of transitions.

Deviation from linear dependence at low amounts is usually caused by increase in the uncertainty of the measurements when the limit of quantitation is approached, or that the peptide studied is already present in the matrix at low amounts. At high amounts, saturation can cause deviation from linear behavior. Fig. 5A shows an example where the measurements at low amounts deviate from the actual concentrations, probably because of the presence of the peptide in the matrix at low amounts. In this case the measurements of all three transitions shift from the diagonal, and Fig. 5B shows that the relative intensities of transitions are almost constant on the low concentrations side. Hence, it is a different type of interference that cannot be detected using the method described in Section 2.2. To address this problem, we developed an algorithm to automatically detect the linear range of a calibration curve.

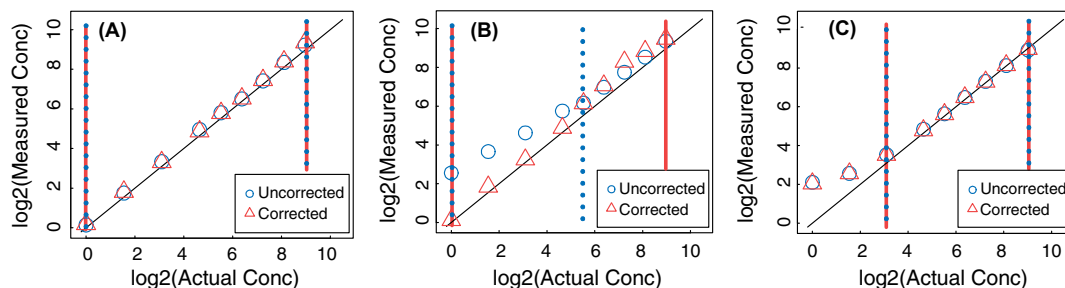
It is assumed that a data set  $\Theta = \{(x, y)\} | x \in \text{actual\_conc}, y \in \text{measured\_conc}\}$  where *actual\_conc* represents the set of log base 2 of actual concentrations, and *measured\_conc* represents the set of log base 2 of measured concentrations. First, three elements are selected from  $\Theta$  and fitted to a linear function  $f(x) = kx + b$  to determine the slope,  $k$ . Second, the errors for all the three which represents the deviation between the slope of linear range and the expected line  $x = y$  are calculated using formula (2) below.

$$\text{Error}(x) = k - 1 \tag{2}$$

If the error of the three elements are all smaller than a set threshold (here selected to be 30%), an attempt is made to elongate the linear range by moving one element forward in the set  $\Theta$  and calculate the linear regression for the next three elements. The linear range will be extended until no more elements can be added and all the data in  $\Theta$  have been tested. Finally, the algorithm returns the longest linear range.

### 2.5. Implementation

Data preprocessing, interference detection, and linear range detection were programmed in Perl ([perldoc.perl.org](http://perldoc.perl.org)) on a Windows platform. Statements of R [24] (a statistical programming



**Fig. 6.** Correction of interference and linear range detection. Uncorrected (blue circles) and corrected (red triangles) measurements, calculated by averaging over the transitions and the technical replicates for the three example peptides from Fig. 2A and B and Fig 5A. In the corrected data, the transitions with interference were removed and the peptide quantity was calculated using only the transitions without interference. The blue dotted line and the red solid line show the linear range of uncorrected and corrected measurements, respectively.

language) were called from Perl in order to plot the data. After obtaining the necessary data and the required data format from the preprocessing step, the transition ratios for the measurements of each peptide from each laboratory were calculated. Based on the algorithm of Section 2.2, outliers of the transitions ratios were found. The measurements were corrected by removing the transitions with interference and the peptide quantity was calculated using only the transitions without interference, and subsequently the linear range detection algorithm was applied to the data.

### 3. Results and discussion

#### 3.1. Results of the interference detection and correction approach applied to experimental data

The method was applied to the CPTAC study Verification Work Group Study 7 data to detect and correct interference (Fig. 6 and Supplementary Fig. 2 [http://fenyolab.org/data/12\\_bao\\_methods/12\\_bao\\_methods\\_suppl\\_fig2.pdf](http://fenyolab.org/data/12_bao_methods/12_bao_methods_suppl_fig2.pdf)). Fig. 6 shows three examples of the performance of our interference detection and correction method with regard to three aspects: measurements without interference are not distorted and are left uncorrected (Fig. 6A); interference can be detected and corrected (Fig. 6B); and the linear range can be detected (Fig. 6C).

#### 3.2. Overcorrection when the estimated standard deviation is low

Because the standard deviation,  $\sigma$ , in Eq. (1) is estimated using technical replicates, this estimation can sometimes be too low, and lead to overcorrection and rejection of good measurements. Fig. 7A shows an example in which the measurements at the highest concentrations were all flagged as incorrect by our interference detection method, because the estimation of the standard deviation of the measurements for the highest concentration was too low; hence, the Z-score was bigger than the threshold, and all transi-

tions were rejected. To overcome this problem of underestimating the standard deviation, we introduce a minimum allowable value for the standard deviation, which is a property of the instrumentation used and the experimental design. To find an appropriate minimum value for these experiments, the standard deviations of the four technical replicates were studied. Fig. 7B shows the frequency distribution for the standard deviation, and the minimum was selected to be 0.02 so that 95% of the standard deviation measurements were higher. This choice of minimum removed the false corrections due to underestimation of the standard deviation for the whole data set, but it did not affect the capability of the algorithm to detect interference.

### 4. Conclusions

In conclusion, we presented a straight-forward method for detection and correction of interference that uses the expected relative intensity of SRM transitions, and an estimation of the noise from measurements of technical replicates. Computer simulations were used to select the optimal interference detection threshold. In addition, an algorithm to automatically detect the linear range of a calibration curve was developed and combined with the method for interference detection. Information about the interferences can be used either to correct or to reject measurements, yielding a straightforward strategy that improves SRM quantitation. The tool for detecting and correcting interference in SRM analysis is freely available <http://fenyolab.org/tools/srm-interference/>.

### Acknowledgements

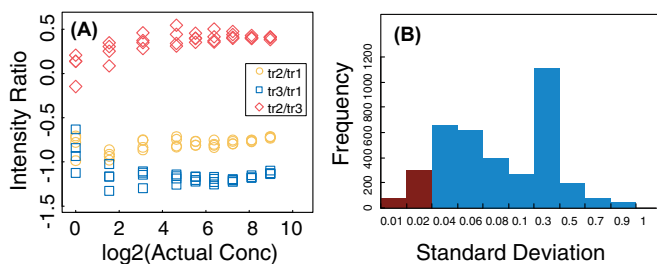
This work was supported by funding provided by the National Institutes of Health Grants CA126485, CA160035, GM103314, GM098256 and GM103362.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ymeth.2013.05.008>.

### References

- [1] S. Gallien, E. Duriez, B. Doman, J. Mass Spectrom. 46 (3) (Mar. 2011) 298–312.
- [2] V. Marx, Nat. Methods 10 (1) (2013) 19–22.
- [3] J.D. Baty, P.R. Robinson, Biomed. Mass Spectrom. 4 (1) (1977) 36–41.
- [4] S. Hoke, K. Morand, K. Greis, T. Baker, K. Harbol, R. Dobson, Int. J. Mass Spectrom. 212 (1) (2001) 135–196.
- [5] R.A. Yost, C.G. Enke, Anal. Chem. 51 (12) (1979) 1251–1264.
- [6] A. Van Langenhove, C.E. Costello, J.E. Biller, K. Biemann, T.R. Browne, Int. J. Clin. Chem. 115 (3) (Sep. 1981) 263–275.



**Fig. 7.** (A) Example of a peptide where the variation of relative transitions is very small between the technical replicates at high concentrations. (B) The frequency of standard deviation values for an example of peptide.

- [7] S.J. Gaskell, K. Rollins, R.W. Smith, C.E. Parker, *Biomed. & Environ. Mass Spectrom.* 14 (12) (1987) 717–722.
- [8] L. Anderson, C.L. Hunter, *Mol. Cell. Proteomics MCP* 5 (4) (2006) 573–588.
- [9] J.R. Barr, V.L. Maggio, D.G. Patterson Jr, G.R. Cooper, L.O. Henderson, W.E. Turner, S.J. Smith, W.H. Hannon, L.L. Needham, E.J. Sampson, *Clin. Chem.* 42 (10) (1996) 1676–1682.
- [10] S.A. Gerber, J. Rush, O. Stemman, M.W. Kirschner, S.P. Gygi, *Proc. Natl. Acad. Sci. USA* 100 (12) (2003) 6940–6945.
- [11] H. Keshishian, T. Addona, M. Burgess, E. Kuhn, S.A. Carr, *Mol. Cell. Proteomics* 6 (12) (2007) 2212–2229.
- [12] M. Kalkum, G.J. Lyon, B.T. Chait, *Proc. Natl. Acad. Sci. USA* 100 (5) (2003) 2795–2800.
- [13] O.P. Bondar, D.R. Barnidge, E.W. Klee, B.J. Davis, G.G. Klee, *Clin. Chem.* 53 (4) (2007) 673–678.
- [14] D.M. Desiderio, M. Kai, *Biomed. Mass Spectrom.* 10 (8) (1983) 471–479.
- [15] S.E. Abbatiello, D.R. Mani, H. Keshishian, S.A. Carr, *Clin. Chem.* 56 (2) (2010) 291–305.
- [16] T.A. Addona, S.E. Abbatiello, B. Schilling, S.J. Skates, D.R. Mani, D.M. Bunk, C.H. Spiegelman, L.J. Zimmerman, A.-J.L. Ham, H. Keshishian, S.C. Hall, S. Allen, R.K. Blackman, C.H. Borchers, C. Buck, H.L. Cardasis, M.P. Cusack, N.G. Dodder, B.W. Gibson, J.M. Held, T. Hiltke, A. Jackson, E.B. Johansen, C.R. Kinsinger, J. Li, M. Mesri, T.A. Neubert, R.K. Niles, T.C. Pulsipher, D. Ransohoff, H. Rodriguez, P.A. Rudnick, D. Smith, D.L. Tabb, T.J. Tegeler, A.M. Variyath, L.J. Vega-Montoto, A. Wahlander, S. Waldemarson, M. Wang, J.R. Whiteaker, L. Zhao, N.L. Anderson, S.J. Fisher, D.C. Liebler, A.G. Paulovich, F.E. Regnier, P. Tempst, S.A. Carr, *Nat. Biotechnol.* 27 (7) (2009) 633–641.
- [17] M. Bern, G. Finney, M.R. Hoopmann, G. Merrihew, M.J. Toth, M.J. MacCoss, *Anal. chem.* 82 (3) (2010) 833–841.
- [18] R.H. Bateman, R. Carruthers, J.B. Hoyes, C. Jones, J.I. Langridge, A. Millar, J.P.C. Vissers, *J. Am. Soc. Mass Spectrom.* 13 (7) (2002) 792–803.
- [19] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, *Mol. Cell. Proteomics MCP* 11 (6) (2012). p. O111.016717.
- [20] J.C. Silva, R. Denny, C.A. Dorschel, M. Gorenstein, I.J. Kass, G.-Z. Li, T. McKenna, M.J. Nold, K. Richardson, P. Young, S. Geromanos, *Anal. chem.* 77 (7) (2005) 2187–2200.
- [21] J. Egertson, A. Kuehn, G. Merrihew, N. Bateman, B. Maclean, J.D. Canterbury, M. Kellmann, V. Zabrouskov, Christine Wu, M.J. MacCoss, *Proc. Am. Soc. Mass Spectrom.* 60 (2012).
- [22] H. Röst, L. Malmström, R. Aebersold, *Mol. Cell. Proteomics* 11 (8) (2012) 540–549.
- [23] D. Domanski, A.J. Percy, J. Yang, A.G. Chambers, J.S. Hill, G.V.C. Freue, C.H. Borchers, *Proteomics* 12 (8) (2012) 1222–1243.
- [24] R Development Core Team, R: A Language and Environment for Statistical Computing, vol. 1 (10) (2008) 2673.