Least-squares fitting mass spectrometry enables accurate and precise highly-multiplexed proteomics at the MS2-level

Anton N. Kozhinov[1], Konstantin O. Nagornov[1], Alex Johnson[2], Michael Stadlmeier[2], Lance Martin[2], Martin Wühr[2], Yury O. Tsybin[1]

[1]Spectroswiss, EPFL Innovation Park, Lausanne, Switzerland; [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ

Introduction

The complement reporter ion approach was developed to improve multiplexed quantitation at the MS/MS level [1]. The currently available TMTpro™ reagents yield complement reporter ion (TMTproC) clusters with 12 quantitation channels. However, unlike the low-mass TMT™ reporters, resolution of the heavy highly-plexed TMTproC reporters requires MS/MS time-domain signals of substantial lengths (e.g. 3 seconds and more), even for the state-of-the-art FTMS instruments. Previously, we presented super-resolution analysis of TMTproC data [2], which was based on the least-squares-fitting (LSF) method for FTMS [3], to address this challenge by reducing the required transient length. Here, we extend the initial LSF implementation for complement reporter ions to enhance the LSF quantitation accuracy in TMTproC workflows with highly-multiplexed channels.

Methods

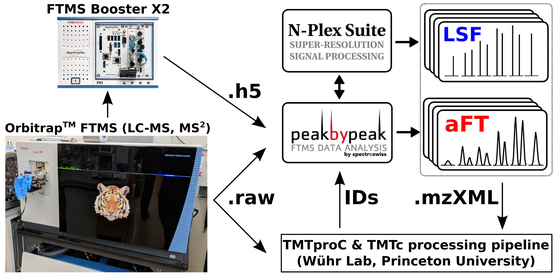

Figure 1. Data acquisition and processing workflow. Cell lysates were prepared as described previously and labeled with TMTpro™ reagents to yield 12 TMTproC channels (Figure 2). LC-MS/MS experiments were conducted on Orbitrap™ Fusion™ Lumos™ FTMS (Thermo Scientific). Time-domain signals were acquired in parallel to RAW files using an external high-performance data acquisition system (FTMS Booster X2, Spectroswiss) [4]. MS/MS time-domain signals were acquired at the resolution presets of 50k, 60k, and 120k, with corresponding time-domain transient lengths of 100 ms, 128 ms, and 256 ms. MS/MS spectra were analyzed with SEQUEST. Calculations of the reference *m/z* values of TMTproC ions, absorption-mode FT (aFT) and LSF processing were performed using Peak-by-Peak (Spectroswiss) running on 8-core desktop computer(s) with 32GB RAM and graphics-card (GPU) data processing capabilities.

LSF for complement reporter ion clusters

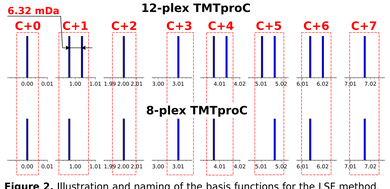

Figure 2. Illustration and naming of the basis functions for the LSF method, corresponding to: (**top**) 12 TMTproC channels with four 6.32 mDa doublets and four singlets, (**bottom**) 8 TMTproC channels with singlets.

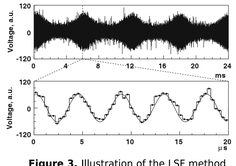

Figure 3. Illustration of the LSF method for FTMS: a transient signal (**top panel**) and a magnified view of the curve fit to the transient (**bottom panel**).

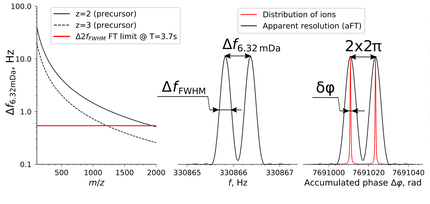

Figure 4. The uncertainty principle of Fourier transform, represented for the spectral components corresponding to 6.32 mDa doublets in complement reporter ion clusters.
Left panel: the frequency difference for a doublet as a function of the doublet's *m/z*. The FT resolution (soft-) limit, T=3.7s, is shown by way of example (it corresponds to the doublet at *m/z*=~1900, 2+ precursor).
Center panel: apparent FT resolution for the doublet in question, at T=3.7 s.
Right panel: comparison of the apparent FT resolution (in the phase scale) and the fundamental limit to ion separation (distribution of ions over their total phase accumulated during ion detection).

12-plex (Sx4, Dx4) TMTproC, *in silico*

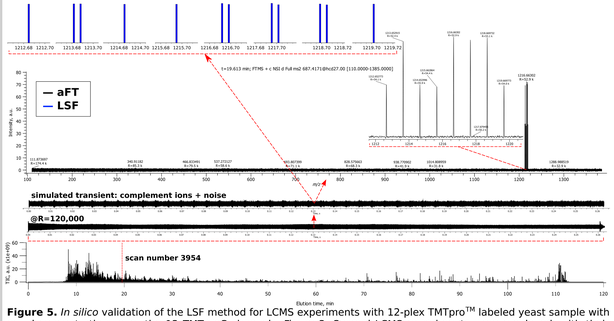

Figure 5. *In silico* validation of the LSF method for LCMS experiments with 12-plex TMTpro™ labeled yeast sample with equal concentrations over the 12 TMTproC channels, Figure 2. Several LCMS experiments were employed, with their MS/MS scans acquired at the resolution setting R=120,000 (256 ms detection period), R=60,000 (128 ms), and R=50,000 (~100 ms). The noise component in each simulated transient is simulated thermal noise (Johnson–Nyquist noise) of the Orbitrap pre-amplifier with a standard deviation according to the thermal noise in experimental transients.

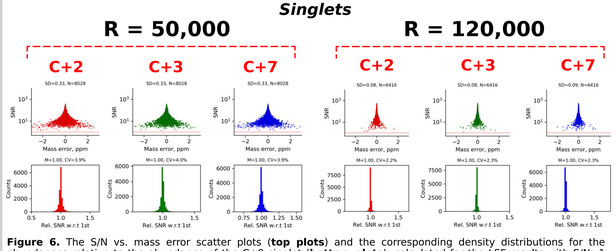

Figure 6. The S/N vs. mass error scatter plots (**top plots**) and the corresponding density distributions for the abundances relative to the abundance of the C+0 singlet (**bottom plots**), calculated for the LSF results with S/N>1 obtained in the LSF analysis of 4 singlets (only) in the simulated TMTproC transients according to the 12-plex TMTproC LCMS experiments with MS/MS scans acquired at R=50,000 (**left panel**) and R=120,000 (**right panel**), Figure 5. The plot titles contain the following values: standard deviation of mass errors (SD), number of identifications (N), mean value of relative abundances (M), coefficient of variation of relative abundances (CV).

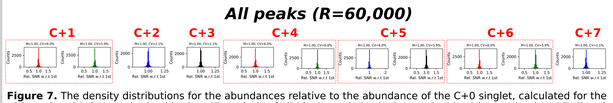

Figure 7. The density distributions for the abundances relative to the abundance of the C+0 singlet, calculated for the LSF results with S/N>1 obtained in the LSF analysis of all 12 channels (4 singlets, 4 doublets) in the simulated TMTproC transients according to the 12-plex TMTproC LCMS experiment with MS/MS scans acquired at R=60,000, Figure 5. The plot titles contain the following values: mean value of relative abundances (M), coefficient of variation of relative abundances (CV).

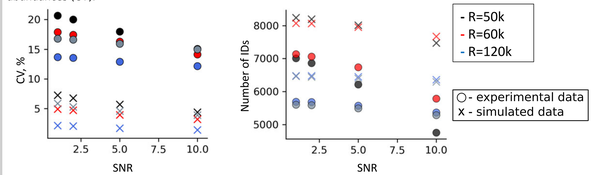

Figure 8. The CV values (**left plot**) and numbers of identifications (**right plot**), plotted as functions of the S/N threshold, calculated for the LSF results obtained in the LSF analysis of all 12 channels in the simulated TMTproC transients according to the 12-plex TMTproC LCMS experiments with MS/MS scans at R=120,000, R=60,000, and R=50,000, Figure 5. Additionally, these plots include the CV and ID results of the LSF analysis of the experimental transients from these LCMS data sets, Figures 9 and 12.

12-plex (Sx4, Dx4) TMTproC

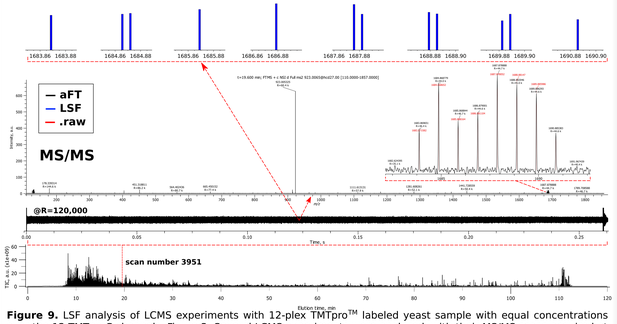

Figure 9. LSF analysis of LCMS experiments with 12-plex TMTpro™ labeled yeast sample with equal concentrations over the 12 TMTproC channels, Figure 2. Several LCMS experiments were analyzed, with their MS/MS scans acquired at the resolution setting R=120,000 (256 ms detection period), R=60,000 (128 ms), and R=50,000 (~100 ms).

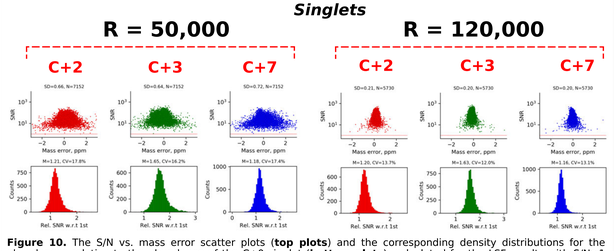

Figure 10. The S/N vs. mass error scatter plots (**top plots**) and the corresponding density distributions for the abundances relative to the abundance of the C+0 singlet (**bottom plots**), calculated for the LSF results with S/N>1 obtained in the LSF analysis of 4 singlets (only) in the MS/MS transients from the 12-plex TMTproC LCMS experiments with MS/MS scans acquired at R=50,000 (**left panel**) and R=120,000 (**right panel**), Figure 9. The plot titles contain the following values: standard deviation of mass errors (SD), number of identifications (N), mean value of relative abundances (M), coefficient of variation of relative abundances (CV).

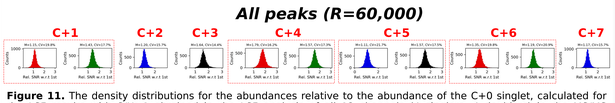

Figure 11. The density distributions for the abundances relative to the abundance of the C+0 singlet, calculated for the LSF results with S/N>1 obtained in the LSF analysis of all 12 channels (4 singlets, 4 doublets) in the MS/MS transients from the 12-plex TMTproC LCMS experiment with MS/MS scans acquired at R=60,000, Figure 9. The plot titles contain the following values: mean value of relative abundances (M), coefficient of variation of relative abundances (CV).

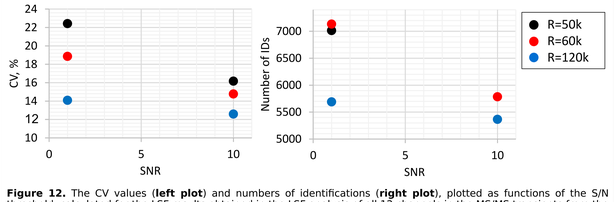

Figure 12. The CV values (**left plot**) and numbers of identifications (**right plot**), plotted as functions of the S/N threshold, calculated for the LSF results obtained in the LSF analysis of all 12 channels in the MS/MS transients from the 12-plex TMTproC LCMS experiments with MS/MS scans at R=120,000, R=60,000, R=50,000, Figure 9.

CV results on the peptide level

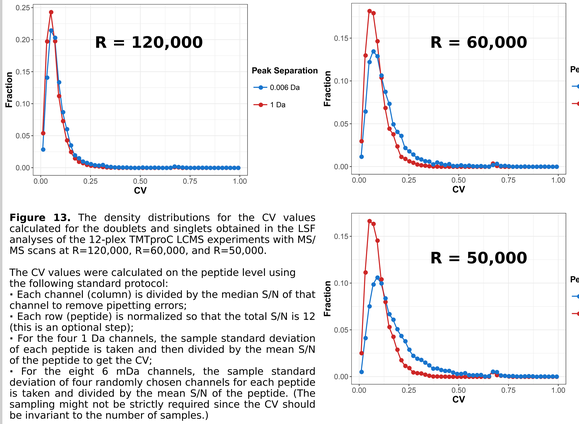

Figure 13. The density distributions for the CV values calculated for the doublets and singlets obtained in the LSF analyses of the 12-plex TMTproC LCMS experiments with MS/MS scans at R=120,000, R=60,000, and R=50,000.

The CV values were calculated on the peptide level using the following standard protocol:
• Each channel (column) is divided by the median S/N of that channel to remove pipetting errors;
• Each row (peptide) is normalized so that the total S/N is 12 (this is an optional step);
• For the four 1 Da channels, the sample standard deviation of each peptide is taken and then divided by the mean S/N of the peptide to get the CV;
• For the eight 6 mDa channels, the sample standard deviation of four randomly chosen channels for each peptide is taken and divided by the mean S/N of the peptide. (The sampling might not be strictly required since the CV should be invariant to the number of samples.)

Conclusions

1. FT requires time-domain signals of substantial lengths (e.g. 3 seconds) to resolve the highly-plexed TMTproC reporters (6 mDa doublets).

2. LSF analysis of time-domain signals allows reducing the required transient times as long as ion interactions within the doublets are sufficiently below the ion coalescence threshold.

3. The LSF analysis of the simulated transients (each generated with the sinusoidal components representing the TMTproC ions and the noise component with a standard deviation according to the thermal noise in experimental transients) measures the contribution of the LSF method's thermal noise-limited performance in achievable CV values.

4. The LSF analysis of the experimental data showed results that are sufficiently far from the thermal noise-limited performance, validating the applicability of the LSF method to the experimental data in this work (i.e. the error distributions due to the thermal noise component are narrow enough relative to the total error distributions obtained for the experimental data).

5. The obtained CV values for the TMTproC ions are a function of the transient time (resolution setting) of the MS/MS scans. For the data acquired at the resolution setting of 120,000 the LSF analysis yields acceptable CVs, for both doublets and singlets in the TMTproC clusters. At this resolution setting, the CV distributions for the doublets practically overlap with those for the singlets, putting forward the LSF approach with R=120k for accurate quantitative analysis in TMTproC workflows with highly-multiplexed channels.

References

1. Johnson A., Stadlmeier M., Wühr M. *J. Proteome Res.* **2021**, 20, 6, 3043.
2. Kozhinov A.N., *et al. The 69th ASMS Conference*, **2021**.
3. Aushev T., Kozhinov A.N., Tsybin Y.O. *JASMS* **2014**, 25, 1263.
4. Nagornov K.O, Zennegg M., Kozhinov A.N., Tsybin Y.O., Bleiner D. *JASMS* **2020**, 31, 2, 257.

Acknowledgements

This work was supported by the European Horizon 2020 research and innovation program under the grant agreement #964553 (ARIADNE).

Poster presented at the 70th ASMS Conference, Minneapolis, MN, USA, June 5–9, 2022.