Proteomics
www.proteomics-journal.com

Estimating the Reliability of Low-Abundant Signals and Limited Replicate Measurements through MS2 Peak Area in SWATH

Franck Limonier, Sander Willems, Geneviève Waeterloos, Myriam Sneyers, Maarten Dhaenens, and Dieter Deforce*

Sequential windows acquisition of all theoretical fragment ions mass spectrometry (SWATH-MS) provides large-scale protein quantification with high accuracy and selectivity. Nevertheless, reliable quantification of low-abundant signals in complex samples remains challenging, as recently illustrated in a multicenter benchmark study of different label-free software tools. Here, the SWATH Replicates Analysis 2.0 template from Sciex is used to highlight that the relationship between the MS2 peak area and the variability can be described by a function. This functional relationship appears to be largely insensitive to variation in samples or acquisition conditions, suggesting a device-intrinsic property. By using a power regression, it is shown that the MS2 peak area can be used to predict the quantification repeatability without relying on replicate injections, thus contributing to high-throughput confident quantification of low-abundant signals with SWATH-MS.

Sequential windows acquisition of all theoretical fragment ions mass spectrometry (SWATH-MS) aims to provide consistent and accurate protein quantification at large scale by combining data-independent acquisition (DIA) with broad data extraction strategy. A comprehensive tandem spectral record of the injected sample is produced, without requiring any target-specific information prior to analysis. The MS1 mass-over-charge (m/z) range is divided into m/z isolation windows which are consecutively fragmented. A single SWATH-MS injection thus theoretically contains all the spectral information that a sample can deliver at a given limit of detection.[1] Spectral assay libraries generated on a high resolution MS instrument are generally used to perform data extraction by matching SWATH spectral data with annotated spectra from the library.[2] New or extended libraries can be used to re-mine the data, providing wide-scale screening capabilities. Various successful discovery-based applications using SWATH-MS have been described so far, demonstrating SWATH's capability to provide high throughput relative quantification of up to thousands of proteins in complex samples.[3–10] Numerous studies, including a recent inter-laboratory study, report that proteins can be reliably quantified across four orders of dynamic range in complex samples using SWATH-MS on a TripleTOF (TT) 5600 system.[1,9,11,12] Still, reliable quantification of lower abundant signals in complex samples remains challenging.

In common practice, multiple injections are performed so that peptide or protein level coefficients of variation (CV) can be estimated to assess SWATH quantification repeatability. Importantly, these CV are inferred by summing the MS2 ion intensities, the root level of SWATH quantification. Ideally, the repeatability criterion should be assessed and weighed at the MS2 level. Currently, proteins with CV below 20% can actually rely on a majority of ion quantifications with very poor repeatability. Two main phenomena contribute to this inference bias: i) A "weight effect": the elements with the highest peak area skew the error of the combined output because of their larger relative weight on the final summed peak area. ii) The inferred CV from ion to peptide and peptide to protein in part reduces by summing the upwards and downwards deviations in the measurements of ion or peptide areas, thus averaging out and reducing the CV of the higher level. Together, while summing peak areas is indeed the most robust metric available today, it is at least misleading to only report the summed CV.

An intrinsic property of analytical instruments is that their accuracy (i.e., variability in measurement) within a certain signal range can be predefined. In other domains, this is depicted in the instrument specifications, as it is the case for example, a balance. In mass spectrometry however, this property is less well-defined. This was recently illustrated by the multicenter study benchmarking software tools for SWATH label-free proteome quantification from Navarro and colleagues.[13] Briefly, they

F. Limonier, G. Waeterloos, Dr. M. Sneyers
Sciensano
Juliette Wytsman 14, B-1050 Brussels, Belgium
E-mail: franck.limonier@sciensano.be
F. Limonier, S. Willems, Dr. M. Dhaenens, Prof. D. Deforce
Laboratory of Pharmaceutical Biotechnology
Faculty of Pharmaceutical Sciences
Ghent University
Ottergemsesteenweg 460, B-9000 Ghent, Belgium

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/pmic.201800186

DOI: 10.1002/pmic.201800186

measured two hybrid proteome samples, A and B, consisting of tryptic digests of human, yeast, and *Escherichia coli* (HYE) proteins mixed in defined proportions to yield expected peptide and protein ratios of 1:1 (A/B) for human, 2:1 for yeast, and 1:4 for *E. coli* proteins ("HYE124"). These samples were measured on two TT systems, respectively 5600+ and 6600, with either 32 fixed or 64 variable windows. A second sample with higher ratio differences (1:1 for human, 10:1 for yeast, and 1:10 for *E. coli*, "HYE110") was also measured in four acquisition modes, exclusively on TT 6600: with 32 or 64, fixed or variable, windows. They point out that in contrast to the SWATH-MS Gold Standard data set, the use of hybrid proteome samples provides several thousand proteins present at defined relative ratios, enabling in-depth statistical evaluation of quantification across a dynamic range several orders of magnitude. In this high-complexity SWATH benchmarking samples, peptide and protein ratios within the lowest intensity tertile displayed the highest variance and differed most from the expected values in all five software methods tested. Moreover, results improved with increasing peptide signals and selection of the most intense peptides generally resulted in lower variance and better quantification accuracy at protein level.[13]

"Intensity score" is incorporated in the mProphet data processing algorithm (now pyProphet) that is used by all these software tools for separating targets from decoys, that is, "identification" by peptide-centric scoring.[14,15] Its weight in the confidence measure is calculated by semi-supervised learning. Quantification of peptides on the other hand is always done by summing the intensities of the transitions, wherein their weight is given by their peak areas. For MS1 label-free quantification of proteins, it was recently shown that modeling at the peptide level outperforms classical summarization-based approaches.[16]

We studied the relationship between intensity and variance in the benchmark dataset with the SWATH Replicates Analysis 2.0 Excel template (Sciex, 2016) to plot average peak area against corresponding CV for up to forty thousand extracted MS2 ions drawn uniformly across the data. We imported the Peakview SWATH 2.0 results from this study (samples A of each sample set—ProteomeXchange identifier PXD002952) into the Replicates Analysis tool. **Figure 1** shows how the relationship between variability (as estimated through the CV), and the root measurement signal (here the MS2 peak area) can be visualized by using the template. Interestingly, the functional dependence between the 80th percentile CV and the log-transformed peak area can be modeled by a power regression ($y = ax^b$) ($R^2 > 0.9$, see Figures S1 and S2, Supporting Information).

In order to explore the applicability of this relationship, we first defined the power regression of our instrument, a TT5600, using in-house SWATH data generated from a commercial human protein extract (Promega, Madison, USA). Next, we applied it to estimate the reliability of quantification of undesired residual—that is, low concentration plasma proteins in fractionated immunoglobulin (Ig) samples. Ig are the leading product of the current plasma fractionation market, being used in immune replacement therapies but also for immunomodulation in various diseases (e.g., Idiopathic Thrombocytopenic Purpura, Guillain-Barré, or Kawasaki syndromes).[17,18] Accurate profiling of residual plasma proteins from plasma pool source material is more thoroughly investigated since co-purified coagulation factor XI caused a serious thromboembolic events outbreak in 2010–2011.[19–21] (For in-house SWATH data generation details, refer to Note S1, Supporting Information). As hypothesized, equivalent relationships between CV and MS2 peak area were observed in both sample types (Figure S2, Supporting Information).

Importantly, the overall power regression is greatly conserved (correlation coefficient [r] > 0.99) across all tested samples (HYE124 and HYE110 from Navarro or in-house human protein extract and Ig samples), conditions (sample load, LC gradient length, number of windows, MS2 acquisition mode, and spectral library), and devices (TT 5600, 5600+, or 6600) (**Figure 2**). This supports the hypothesis that the functional relation between repeatability and signal intensity is an intrinsic property related to the instrument, such as detector capabilities. This in turn implies that power can be borrowed from this function. Indeed, the depicted functional relationship integrates around 40K data points,

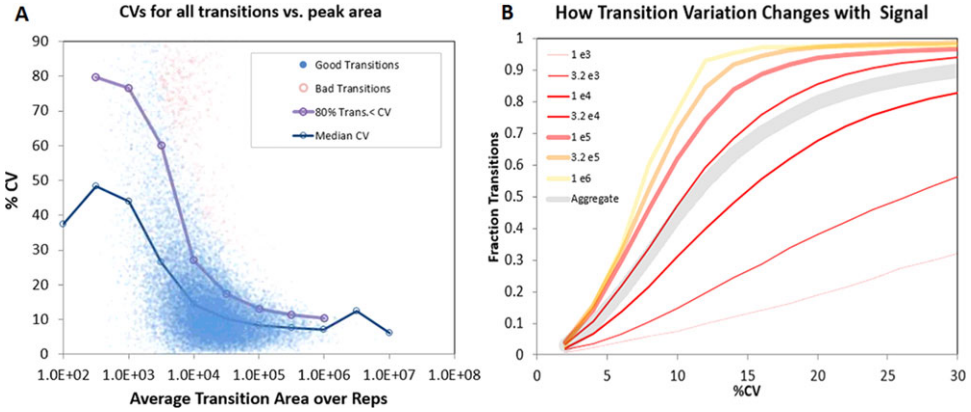

Figure 1. Transition peak area to CV relationship generated with the SWATH Replicate Analysis 2.0 Excel sheet (HYE124 sample A on TT6600, 64 variable windows). A) Dotplot representing the %CV measurements over triplicate injections. Transitions are defined by the software as "good" or "bad" based on whether the %CV is more or less than three times the median CV for area. Nevertheless, all the ions are included for the calculation of the median and 80th percentile lines. The percentile value is user-defined. B) How transition variation changes with signal (transitions are clustered into signal intensity groups).

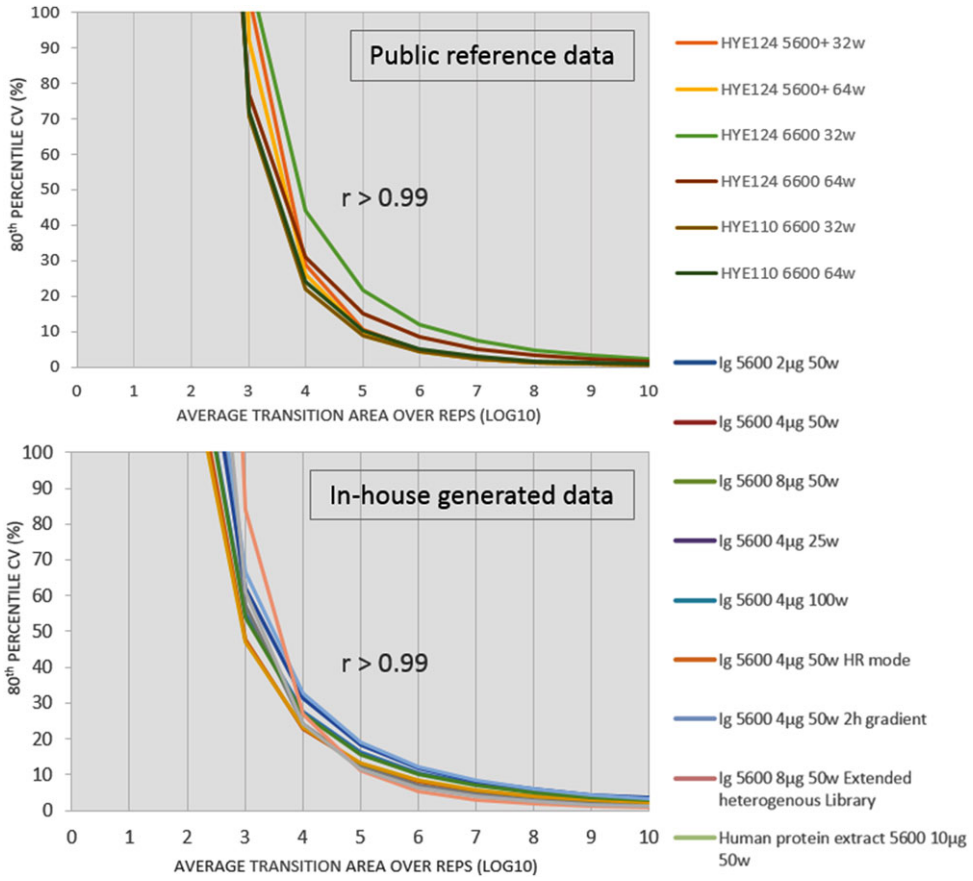

Figure 2. MS2 peak area (log10) to 80th percentile CV (%) functional relationship modeling with power regression. Correlation of the power regression across the different conditions tested: different samples (public reference data–HYE124 and HYE110 or in-house generated data–human protein extract and Ig) sample load (in μg), LC gradient length, MS2 acquisition mode (high sensitivity by default or high resolution, HR), number of windows (w), spectral library and devices (TripleTOF 5600, 5600+, or 6600). Correlation coefficient (r) is above 0.99.

whereas usual CV approaches only consider the number of replicates for each point individually. Note that we were not able to model the instrument performance in the lowest peak area range ($\log_{10}$ peak area < 3), and we neglected the sometimes observed re-decrease in variability as well (orange dots in Figure S1, Supporting Information), because there is a lack of data points in this region (as seen in Figure 1A). On the other extremity of the peak area scale, the sometimes observed re-increase in variability might be linked to detector saturation, but there are again too few data points in the region to confirm this hypothesis (Figures 1A and S2, Supporting Information).

Once the user has outlined the functional dependence of the employed device in his/her own lab, he/she can now estimate the probability that a given measurement will meet given repeatability requirements, for example, 80% chance that the data has a CV below 20%. This can be illustrated by a concrete example. In the data of Navarro and colleagues, the estimate function between the 80th percentile CV (y) and MS2 peak area (x in $\log_{10}$ scale) for the TT 6600 is $y = 2530.2 \times x^{-3.18}$ using the triplicate injection of the sample HYE124. This means that an ion with a peak area of 1000 has 80% chance that the CV is below $y = 2530.2 \times 3^{-3.18} = 76.90\%$. On the other hand, an ion with a peak area of 100 000 has 80% chance that the CV is below $y = 2530.2 \times 5^{-3.18} = 15.15\%$. In other words, one would consider an ion with a peak area of 100 000 as more reliable for quantification than an ion with a peak area of 1000. Obviously, the function also enables to estimate the peak area value from which a certain level of repeatability is reached. Here, we can estimate that the MS2 ions with a peak area equal to or greater than $x = \sqrt[b]{y/a} = \sqrt[-3.18]{20/2530.2} = 4.58$, that is, around 38 000, have at least 80% chance to have a CV $< 20\%$. This gives the user a better idea on how low-abundant signals can be quantified without compromising the repeatability and quantification confidence.

To verify these predictions, we applied the power regression function to the other sample batch, HYE110, that was independently acquired on the same device in triplicate. For this, we picked the transitions with an area between 950 and 1050 as well as the transitions with a peak area between 95 000 and 105 000 and compared their CV on triplicate injections (Table 1, Supporting Information). 79% of the first group had a CV below 77%, while 95% of the second group had a CV below 15%. Then, we selected all the transitions with a peak area above 38 000. The increased reliability of this subset of transitions was clear-cut: 96% of these transitions had a CV $< 20\%$, while only 55% of all the

extracted transitions passed the same CV threshold (Table 2, Supporting Information).

Finally, to illustrate the applicability, we built a Python tool that enables to automatically flag transitions falling in a user-defined peak area range corresponding to a desired repeatability, as calculated through the power regression (script available in Supporting Information). This way, peptides and proteins that integrate at least a certain number of such transitions can be selected. This script was used to estimate the reliability of quantification of low concentration residual plasma proteins in the Ig samples, where the initial concentration dynamic range reaches at least six orders of magnitude.[22,23] Across all the different Ig samples tested, 82–100% of the proteins having at least two peptides with at least two ions in the desired area range (2.0E+04–1.0E+06, where 80% of data has a CV $<$ 20%) demonstrated a CV below 20% over the replicates. At the same time, only 32–49% of the other proteins passed the same CV threshold. In total, 94% of all flagged residual plasma proteins quantifications had a CV below 20% (see Table S3, Supporting Information). The script also enables to select the acquisition parameters that maximize the number of MS2 ions in the adequate area range without compromising the total number of ions extracted (e.g., increasing the sample load, see Figure S3, Supporting Information). While such distinction of different subgroups should never be seen as a hard cut-off criterion, the highlighted relationship thus appears as an actionable point for increased confidence in low concentration proteins quantification with SWATH.

In conclusion, we here refer to a well-known concept in analytical chemistry and bring it to a new level of applicability. More specifically, it is well established that the repeatability is linked to the measurement intensity. It was evidenced by Navarro and colleagues that this equally holds for SWATH-MS. Transition intensity by means of the MS2 peak area is already included in data processing pipelines by selecting the top *n* (usually top 3–6) most abundant transitions. As observed in the Navarro manuscript however, this is not sufficient to guarantee a reliable quantification of low-abundant signals. To our knowledge, we are the first to describe the functional relationship between CV and MS2 peak area (by using the relatively under-exploited Replicates Analysis 2.0 template from Sciex). With only a simple power regression to model the relation, we demonstrate that the MS2 peak area can confidently contribute to predict the reliability of MS2 quantification. Importantly, this metric does not rely on the number of replicates and can even be considered for single injections, contributing to increased high-throughput. The robustness of the observed relationship across all the investigated conditions suggests that it is intrinsic to the instrument. Assuming that its stability would be regularly verified, this functional relationship could thus be incorporated into the scoring algorithm of SWATH-MS and other DIA software tools.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

M.D. and D.D. contributed equally to this work. The authors would like to thank Jörg Dojahn from Sciex for his help with Peakview SWATH 2.0. This study was supported by FWO Grant G013916N (SW) and FWO Mandate 12E9716N (MD).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

low-abundant signals, MS2 peak area, repeatability, sequential windows acquisition of all theoretical fragment ions mass spectrometry

Received: May 2, 2018
Revised: September 25, 2018
Published online:

[1] L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.
[2] O. T. Schubert, L. C. Gillet, B. C. Collins, P. Navarro, G. Rosenberger, W. E. Wolski, H. Lam, D. Amodei, P. Mallick, B. MacLean, R. Aebersold, *Nat. Protocols* **2015**, *10*, 426.
[3] B. C. Collins, L. C. Gillet, G. Rosenberger, H. L. Röst, A. Vichalkovski, M. Gstaiger, R. Aebersold, *Nat. Methods* **2013**, *10*, 1246.
[4] J.-P. Lambert, G. Ivosev, A. L. Couzens, B. Larsen, M. Taipale, Z.-Y. Lin, Q. Zhong, S. Lindquist, M. Vidal, R. Aebersold, T. Pawson, R. Bonner, S. Tate, A.-C. Gingras, *Nat. Methods* **2013**, *10*, 1239.
[5] Y. Liu, A. Buil, B. C. Collins, L. C. J. Gillet, L. C. Blum, L.-Y. Cheng, O. Vitek, J. Mouritsen, G. Lachance, T. D. Spector, E. T. Dermitzakis, R. Aebersold, *Mol. Syst. Biol.* **2015**, *11*, 786.
[6] T. Guo, P. Kouvonen, C. C. Koh, L. C. Gillet, W. E. Wolski, H. L. Röst, G. Rosenberger, B. C. Collins, L. C. Blum, S. Gillessen, M. Joerger, W. Jochum, R. Aebersold, *Nat. Med.* **2015**, *21*, 407.
[7] R. Bruderer, O. M. Bernhardt, T. Gandhi, S. M. Miladinović, L.-Y. Cheng, S. Messner, T. Ehrenberger, V. Zanotelli, Y. Butscheid, C. Escher, O. Vitek, O. Rinner, L. Reiter, *Mol. Cell. Proteomics* **2015**, *14*, 1400.
[8] O. T. Schubert, C. Ludwig, M. Kogadeeva, M. Zimmermann, G. Rosenberger, M. Gengenbacher, L. C. Gillet, B. C. Collins, H. L. Rost, S. H. E. Kaufmann, U. Sauer, R. Aebersold, *Cell Host Microbe* **2015**, *18*, 96.
[9] N. Selevsek, C.-Y. Chang, L. C. Gillet, P. Navarro, O. M. Bernhardt, L. Reiter, L.-Y. Cheng, O. Vitek, R. Aebersold, *Mol. Cell. Proteomics* **2015**, *14*, 739.
[10] E. G. Williams, Y. Wu, P. Jha, S. Dubuis, P. Blattmann, C. A. Argmann, S. M. Houten, T. Amariuta, W. Wolski, N. Zamboni, R. Aebersold, J. Auwerx, *Science* **2016**, *352*, aad0189.
[11] J. M. B. Simbürger, K. Dettmer, P. J. Oefner, J. Reinders, *J. Proteomics* **2016**, *145*, 137.
[12] B. C. Collins, C. L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, S. L. Bader, D. W. Chan, B. W. Gibson, A.-C. Gingras, J. M. Held, M. Hirayama-Kurogi, G. Hou, C. Krisp, B. Larsen, L. Lin, S. Liu, M. P. Molloy, R. L. Moritz, S. Ohtsuki, et al., *Nat. Commun.* **2017**, *8*, 291.
[13] P. Navarro, J. Kuharev, L. C. Gillet, O. M. Bernhardt, B. MacLean, H. L. Röst, S. A. Tate, C.-C. Tsou, L. Reiter, U. Distler, G. Rosenberger, Y.

Perez-Riverol, A. I. Nesvizhskii, R. Aebersold, S. Tenzer, *Nat. Biotechnol.* **2016**, *34*, 1130.
[14] L. Reiter, O. Rinner, P. Picotti, R. Hüttenhain, M. Beck, M.-Y. Brusniak, M. O. Hengartner, R. Aebersold, *Nat. Methods* **2011**, *8*, 430.
[15] G. Rosenberger, I. Bludau, U. Schmitt, M. Heusel, C. L. Hunter, Y. Liu, M. J. MacCoss, B. X. MacLean, A. I. Nesvizhskii, P. G. A. Pedrioli, L. Reiter, H. L. Röst, S. Tate, Y. S. Ting, B. C. Collins, R. Aebersold, *Nat. Methods* **2017**, *14*, 921.
[16] L. J. E. Goeminne, K. Gevaert, L. Clement, *Mol. Cell. Proteomics* **2016**, *15*, 657.
[17] Creativ-Ceutical/EU Commission, **2015**. https://ec.europa.eu/health//sites/health/files/blood_tissues_organs/docs/20150408_cc_report_en.pdf (Accessed on 2018-01-24)
[18] WHO, vol. 19, World Health Organization, **2015**. http://www.who.int/medicines/publications/essentialmedicines/en/ (Accessed on 2018-01-24)
[19] A. S. Wolberg, R. H. Kon, D. M. Monroe, M. Hoffman, *Am. J. Hematol.* **2000**, *65*, 30.
[20] M. Etscheid, S. Breitner-Ruddock, S. Gross, A. Hunfeld, R. Seitz, J. Dodt, *Vox Sang.* **2012**, *102*, 40.
[21] M. B. Funk, N. Gross, S. Gross, A. Hunfeld, A. Lohmann, S. Guenay, K. M. Hanschmann, B. Keller-Stanislawski, *Vox Sang.* **2013**, *105*, 54.
[22] European Pharmacopoeia Commission: Council of Europe (European Department for the Quality of Medicines), *Eur. Pharmacopoeia Online* **2012**, *9.2*, 0918.
[23] W. A. Germishuizen, D. C. Gyure, D. Stubbings, T. Burnouf, *Biologicals* **2014**, *42*, 260.