

# CalibraCurve: A Tool for Calibration of Targeted MS-Based Measurements

Michael Kohl,\* Markus Stepath, Thilo Bracht, Dominik A Megger, Barbara Sitek, Katrin Marcus, and Martin Eisenacher

Targeted proteomics techniques allow accurate quantitative measurements of analytes in complex matrices with dynamic linear ranges that span up to 4–5 orders of magnitude. Hence, targeted methods are promising for the development of robust protein assays in several sensitive areas, for example, in health care. However, exploiting the full method potential requires reliable determination of the dynamic range along with related quantification limits for each analyte. Here, a software named CalibraCurve that enables an automated batch-mode determination of dynamic linear ranges and quantification limits for both targeted proteomics and similar assays is presented. The software uses a variety of measures to assess the accuracy of the calibration, namely precision and trueness. Two different kinds of customizable graphs are created (calibration curves and response factor plots). The accuracy measures and the graphs offer an intuitive, detailed, and reliable opportunity to assess the quality of the model fit. Thus, CalibraCurve is deemed a highly useful and flexible tool to facilitate the development and control of reliable SRM/MRM-MS-based proteomics assays.

It has been demonstrated that reproducible and accurate LC-SRM/MRM-MS measurements can be harmonized across several laboratories.<sup>[3]</sup>

The above-mentioned features are crucial to establish these approaches for the development of robust protein assays in several sensitive areas, for example, in health care.

However, in order to realize the full potential of targeted proteomics methods an impeccable statistical assessment of the conducted data analysis is essential. An important task in this regard is the determination of the dynamic range along with the related quantification limits and the creation of calibration curves for each assay.

Virtually any technique in analytical chemistry that aims at quantification of the amount or the concentration of chemical species is limited to a constrained concentration range.

Usually, calibration curves show three regimes.<sup>[4]</sup> In the first regime (low concentrations), the analyte intensities are highly influenced by noise. In the subsequent middle part of the concentration range, the calibration curve shows a linear trend. This part is often referred to as “response curve.” Response curves are confined by the lower (LLOQ) and the upper limit of quantification (ULOQ). Beyond the ULOQ (saturation regime), the curve shows no longer a linear relationship between increasing concentrations and measured intensities. Reliable inference of concentrations is possible only in the linear part of the calibration curve.

Frequently, only LLOQ is estimated and determination of ULOQ is neglected. However, we observed reduction of product ion intensities at higher analyte concentrations (Figure 1), which consequently impairs estimated quantities: In an IgG subtype quantification study, we spiked in corresponding heavy labeled unique peptides. The native (light) samples were submitted to PRM measurement with a constant total peptide amount and heavy stable labeled peptides were spiked in with increasing concentration per sample. Detailed information regarding the experimental setup is provided in the Supporting Information. While the measured peak areas of the most intense product ions of the native peptide were stable over the first six concentrations of heavy spike-ins (0–53.3 fmol  $\mu\text{L}^{-1}$ ), at higher spike-in concentrations a decrease in respective peak areas of native

## 1. Introduction

In comparison with classical shotgun techniques, targeted proteomics approaches (SRM/MRM and PRM) show higher sensitivity, reliability, and reproducibility. Targeted methods allow accurate quantitative measurements of analytes in complex matrices in broad concentration ranges. SRM/MRM has a dynamic linear range that span up to 4–5 orders of magnitude.<sup>[1,2]</sup>

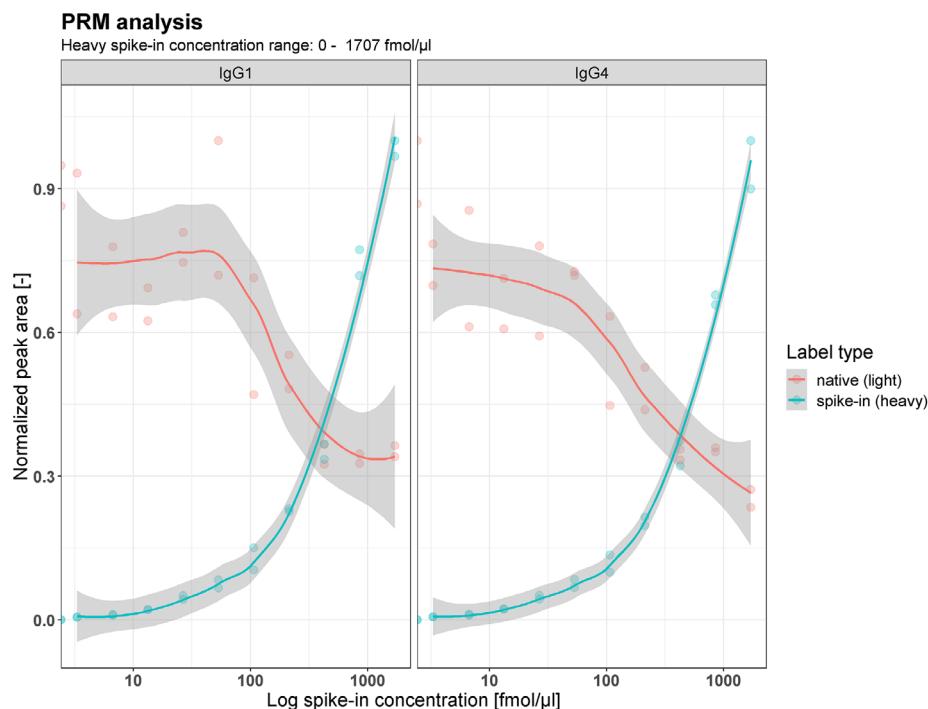
Dr. M. Kohl, Dr. M. Stepath, Dr. T. Bracht, Dr. D. A Megger, Prof. B. Sitek, Prof. K. Marcus, Dr. M. Eisenacher  
Medizinisches Proteom-Center, Medical Faculty  
Ruhr University Bochum  
Gesundheitscampus 4, Bochum D-44 801, Germany  
E-mail: michael.kohl@rub.de

Dr. D. A Megger  
Institute for Virology, University Hospital Essen  
University of Duisburg-Essen  
Essen D-45 147 Germany

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/pmic.201900143>

© 2020 The Authors. *Proteomics* published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/pmic.201900143



**Figure 1.** PRM analysis on a Q Exactive instrument (Thermo Scientific, Bremen, Germany). A medium complex mixture of tryptic digested IgG proteins was measured with constant peptide amount (140 ng). Heavy labeled peptides (IgG1: GPSVFPLAPSSR, IgG4: GPSVFPLASSSR) were spiked in ( $n = 2$ ) with concentrations that double with each step of the calibration curve from 50 to 1707 fmol  $\mu\text{L}^{-1}$ . Raw data was inspected with the Skyline software, following normalization and curve fitting in R. Colored circles (red: light peptide, cyan: heavy peptide) represent normalized peak areas of the most intense product ions ( $y7$ ). Areas were normalized to the maximum values with respect to the peptide and label type. Colored solid lines illustrate regression curves (loess). 0.95 confidence intervals are shown as grey ribbon.

product ions was observed, indicating the need of ULOQ determination.

## 2. Software Development

A software named CalibraCurve is presented. The software is written in R<sup>[5]</sup> and additionally provided as KNIME<sup>[6]</sup> workflow(s). In comparison with utilization as an R script, the CalibraCurve KNIME workflow implementation facilitates usage of the tool for users who are not accustomed to command line applications. CalibraCurve calculates calibration curves and both LLOQ and ULOQ. The software computes several measures that can be used to assess the accuracy of measurements.

We start with some definitions of central terms that are used in the manuscript: According to the definition of the International Organization for Standardization,<sup>[7]</sup> accuracy is an umbrella term for the overall measurement uncertainty and comprises both precision and trueness.

Precision assesses measurement reproducibility, that is, the variance of measurement results performed repeatedly with unaltered conditions.

The measure of trueness is frequently given in terms of bias, which is defined as the difference between a measurement result and the true value.<sup>[7]</sup> Here, the measurement result is the calculated concentration and the true value is the known concentration

of an analyte (established, e.g., by spiking in a known peptide amount).

CalibraCurve adopts this concept and calculates metrics for both trueness and precision. As a measure for precision, coefficient of variation (CV) values ( $\text{Meas}_{\text{CV}}$ ) are calculated for each concentration level from the intensities of the replicates.

Bias is given as difference between the true and the calculated value expressed as percentage. In the following, this value is referred to as “percent bias” (PB) and calculated by

$$\text{PB} = \frac{100 * (|x_e - x_c|)}{x_e} \quad (1)$$

where  $x_e$  is the expected concentration and  $x_c$  is the calculated concentration, which is derived from the coefficients of a fitted linear model. PB is computed for each measurement. Additionally, the PB values are used to compute  $\text{PB}_{\text{av}}$  (i.e., either the mean or the median value [user setting]) and the corresponding CV value ( $\text{PB}_{\text{CV}}$ ) for each concentration level.

Besides calibration curves, CalibraCurve calculates so-called response factor plots. Response factor plots are graphs of response factors versus the concentration.

Similar to the method described by ref. [8], response factors (RF) are calculated by

$$\text{RF} = \frac{\text{SR} - y_{\text{intersect}}}{C_A} \quad (2)$$

where SR is signal response (i.e., for example the peak area or the peak area ratio [i.e., heavy area/light area]),  $y_{\text{intersect}}$  is the intersect of the regression line, and  $C_A$  is the concentration of the corresponding analyte. RF plots are efficient and intuitive means to determine the dynamic range from available calibration data.<sup>[8,9]</sup> Ideally, within the linear range the mean RF values calculated for each concentration level  $i$  ( $RF_{\text{mean}_i}$ ) should have a near zero-slope (i.e., a horizontal line). Individual RFs should be located within a narrow distance from  $RF_{\text{mean\_total}}$  (i.e., the mean value of all RFs included in the final linear range that is calculated by CalibraCurve).

Development of CalibraCurve is oriented toward the following objectives.

- Implementation of basic quality checks to ensure reliability of the calculations.
- Usability: CalibraCurve is currently implemented as a command line tool and as KNIME workflow. This enables adaption of the software features for experienced R users and via the KNIME workflow utilization of the software for users that do not want to use the R script directly.
- Adaptability: The software is highly customizable. A large set of optional parameters that govern both program execution and the output (including visualization) can be adapted if required. However, the adaption of mandatory parameters is reduced to the absolute minimum in order to facilitate the basic application of the tool. The parameter settings of CalibraCurve are extensively discussed in the Supporting Information, which also includes the CalibraCurve 2.0 R script (CalibraCurve\_v2.0.R) and two KNIME workflows that deal with different input formats.
- Automation: CalibraCurve supports batch-wise processing of an arbitrary number of input files.
- Generation of result files with the most important metrics and visualizations enabling evaluation of the calibration quality by the user.

### 3. Software Architecture and Workflow

In the following, the working steps of the CalibraCurve algorithm (Figure 2) are detailed. Valid input data are plain text/csv files and xls/xlsx files (supported by one of the CalibraCurve KNIME workflows). The following working steps were performed for each of the read data files. CalibraCurve starts with several preprocessing procedures. Rows that contain either zero or missing values are identified and removed. Afterward, CalibraCurve checks whether a user-defined number of replicates is available for each concentration level. Levels with insufficient replicate numbers are removed from the dataset.

The now validated dataset is used to assess accuracy for each concentration level in order to estimate the dynamic linear range. To this end, the algorithm consecutively evaluates the precision and the percentage bias of the data. First, CalibraCurve calculates  $Meas_{CV}$  values for each level in order to compute the preliminary linear range (PLR). PLR is defined as a range of consecutive levels with  $Meas_{CV}$  that pass a user-defined threshold ( $TMeas_{CV}$ ). If several PLRs exist, which are separated by concentration levels with  $Meas_{CV} > TMeas_{CV}$ , CalibraCurve selects the PLR with the

highest number of consecutive levels by default. Alternatively, users can also choose that levels with  $Meas_{CV} > TMeas_{CV}$  are allowed within the PLR. In this case, the PLR extends from the lowest to the highest concentration level that pass the  $TMeas_{CV}$  criterion. For more details of this part of the algorithm, please refer to the CalibraCurve manual.

In the next workflow step, a weighted and an unweighted linear model are fitted to the PLR dataset. CalibraCurve provides two weighting methods ( $1/x$  and  $1/x^2$ ), that are often used in LC-MS/MS assays.<sup>[10]</sup> Then, the software calculates PB values for each measurement as well as  $PB_{av}$  and  $PB_{cv}$  for each level of the PLR dataset. The algorithm checks if both the  $PB_{av}$  value of the highest ( $C_h$ ) and the lowest concentration level ( $C_l$ ) is below a user-defined threshold ( $TPB_{av}$ ). This is considered as valid condition for the final linear range.<sup>[10]</sup>

If either  $C_l$  or  $C_h$  do not pass the  $TPB_{av}$  criteria, the algorithm selects this level for removal. If both  $C_l$  and  $C_h$  show  $PB_{av}$  values above  $TPB_{av}$ , CalibraCurve provides several opportunities for level removal. Per default, the level with the higher  $PB_{av}$  is selected for removal. Alternatively, the user may also choose a method that penalizes high variance of PB within a level (measured by  $PB_{cv}$ ). If the difference between  $PB_{av}$  values of  $C_l$  and  $C_h$  is sufficiently small, the algorithm removes the level with the higher  $PB_{cv}$ . Users can define a threshold for the  $PB_{av}$  difference between  $C_l$  and  $C_h$ . Calculated  $PB_{av}$  difference values below the threshold give rise to removal of the level with higher  $PB_{cv}$ .

Please note that CalibraCurve only checks whether  $C_h$  or  $C_l$  shows  $PB_{av}$  values that meet the threshold criteria. The rationale for this approach is that  $PB_{av}$  is supposed to drop only below LLOQ and above ULOQ. However, individual PB values for each measurement and both  $PB_{av}$  and  $PB_{cv}$  values for each level are written to the result file (along with warnings if  $PB_{av}$  values indicate poor trueness). It is recommended to review this data in order to check the consistency of the input data.

After level removal, the hereby-reduced dataset is again used to fit new linear models and to recalculate PBs,  $PB_{av}$ , and  $PB_{cv}$ . The algorithm carries out these procedures iteratively until the remaining dataset is "final," that is,  $PB_{av}$  values of both  $C_l$  and  $C_h < TPB_{av}$ .

Finally, CalibraCurve writes a result file that comprises  $Meas_{CV}$  (for the preliminary final range) and PB,  $PB_{av}$ , and  $PB_{cv}$  values (calculated from the data of the final linear range). Another result file stores the summary information for both the weighted and the unweighted linear model (calculated from the final dataset). For each sample, calibration graphs (Figure 3a) and response factor plots (Figure 3b,c) are computed.

### 4. Interpretation of CalibraCurve Results

Interpretation of calibration curves is exemplified using data obtained from MRM validation experiments carried out as part of a proteomics discovery study aiming at biomarker identification for hepatic fibrosis.<sup>[11]</sup>

The example dataset analyzed here was measured using an Agilent 6490 triple quadrupole mass spectrometer (detailed description of the experimental setup is given in ref. [11] and data of this study is available from PASSEL via the identifier PASS00653). Measurements are used to determine the dynamic range of

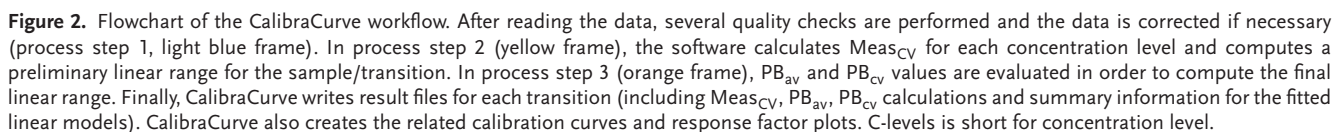
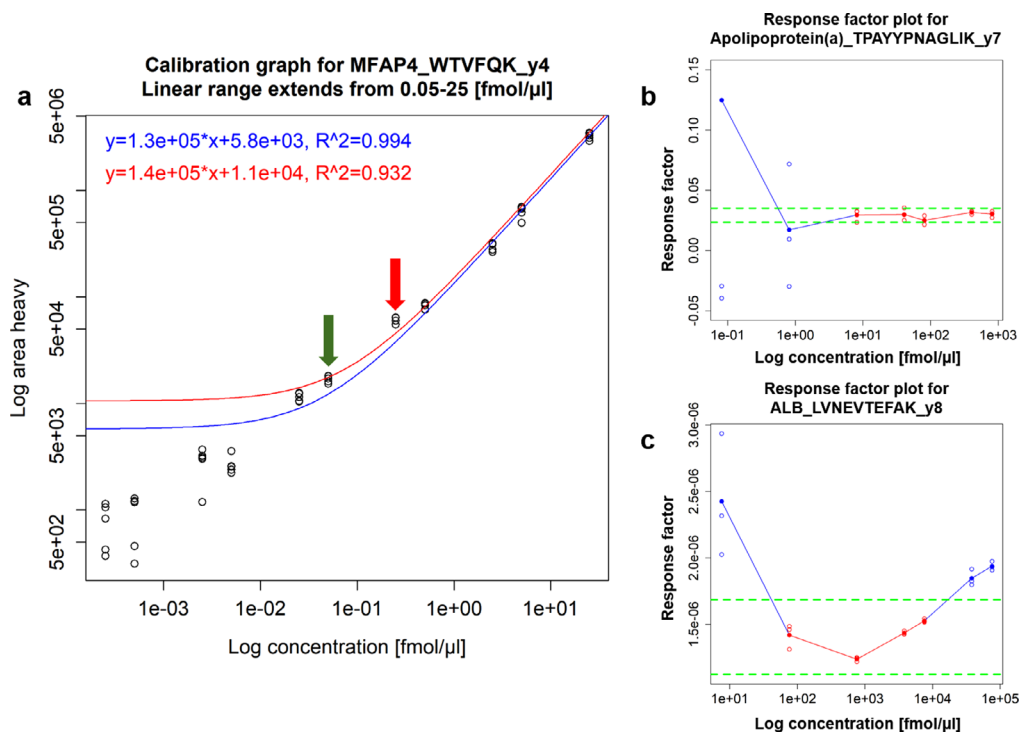


Figure 3a shows the calibration graph for the y4 product ion. The linear range extends from 0.05 to 25 fmol  $\mu\text{L}^{-1}$  (maximum spike-in concentration). Naturally, coefficient of determination ( $R^2$ ) for unweighted linear regression is always higher than for the weighted linear regression. In the considered case, the unweighted linear model (blue regression line) yields  $R^2 = 0.994$  and the weighted linear model ( $1/x^2$  weighting, red regression line) shows an  $R^2$  of 0.932. However, despite the higher  $R^2$  value, some of the  $\text{PB}_{\text{av}}$  values calculated for the unweighted model are very high: For the lowest level of the final linear range (0.05 fmol  $\mu\text{L}^{-1}$ ), the unweighted linear model yields  $\text{PB}_{\text{av}} = 73.75\%$ . In contrast, for the weighted model,  $\text{PB}_{\text{av}}$  is 14.65%. This confirms the statement of Green<sup>[8]</sup> that considering  $R^2$  alone is misleading for

Two datasets obtained from ref. [13] are used to illustrate application of response factor plots (Figure 3b,c). In Figure 3b



**Figure 3.** a) Calibration curve for a transition of the WTVFQK peptide (protein MFAP4). The red regression line and red equation are calculated from the weighted linear model, the blue regression line and the blue equation are obtained from the unweighted linear model. The green arrow indicates LLOQ, the red arrow a concentration level with high  $PB_{av}$ . b,c) Response factor plots for apolipoprotein A (transition TPAYPNAGLIK\_y7) and for serum albumin (ALB, transition LVNEVTEFAK\_y8). Blue points (RFs) and lines ( $RF_{mean,i}$  values) are calculated from the validated dataset, red points and lines from the final linear range dataset. Green dashed lines indicate thresholds for distances from the  $RF_{mean,total}$ .

(apolipoprotein), most individual RFs and all  $RF_{mean,i}$  values of the final linear range (red line and points) are within reasonable distances from  $RF_{mean,total}$  (indicated by the green threshold lines). The red line shows a near zero slope indicating the extent of the linear range.

In Figure 3c (albumin), only RFs of four levels are within the thresholds. The lowest and the two highest concentration levels show RFs that are outside of the thresholds. These results are in accordance with LLOQ and ULOQ calculated by CalibraCurve and with the  $PB_{av}$ ,  $Meas_{CV}$  data of the result file. This example shows the importance of ULOQ consideration.

The discussed input data and the CalibraCurve results/plots are included in the Supporting Information.

## 5. Comparison of CalibraCurve with Similar Software

In this section, the features of CalibraCurve are compared with similar, freely available tools.

The first discussed software, QuaSAR, is a tool suite aiming at quality control, analysis, and data visualization of SRM/MRM experiments. The software is available as so-called “Skyline External Tool,” that is, QuaSAR is directly integrated into Skyline<sup>[14]</sup> after installation. This feature is a great benefit for users that already use Skyline. Among other things, QuaSAR<sup>[15]</sup> provides batch-wise creation of calibration curves and calculation of related limits of detection (LODs) and LLOQs. The software also supports

visualization and stores calibration plots within a PDF file. However, QuaSAR does not support detection of ULOQ. LLOQ is simply calculated using the formula  $LLOQ = 3 \times LOD$ .

MSstats<sup>[16]</sup> is another “Skyline External Tool.” MSstats aims at relative quantification of proteins and peptides. The tool provides workflows for detection of limit of blank (LOB) and LOD, but does not intend to detect quantification limits or to create and evaluate calibration curves.

MRMplus<sup>[17]</sup> is a standalone Java application. The software calculates calibration curves and computes LOD, LLOQ, and ULOQ. However, adjustment options are rather limited (e.g., adjustment of thresholds or selection of weighting options for the curve fitting are lacking). It seems that no opportunity exists to export the calibration graphs.

Qualis-SIS<sup>[13]</sup> is a web-based client-server application. The software utilizes calibration strategies that are similar to techniques used by CalibraCurve: Qualis-SIS computes LLOQ, ULOQ and uses both precision and trueness to evaluate the quality of the linear model fit. Besides a simple method that always removes  $C_1$ , Qualis-SIS provides a more complex approach that removes either  $C_1$  or  $C_h$ . To this end, the software uses  $R^2$  values to evaluate the goodness of fit and to select the level for removal. However, as stated before,  $R^2$  values alone are insufficient to evaluate linearity. In contrast to Qualis-SIS, CalibraCurve applies the user-defined  $Meas_{CV}$  and  $PB_{av}$  thresholds for both the assessment of the model fit and for the detection of concentration levels that are outside of the linear range. We consider this approach as a more consistent scheme for computation of the



linear range. The most important advantages of CalibraCurve regard improved visualization features: Neither Qualis-SIS nor any other software discussed provides the possibility to generate response factor plots. Compared with Qualis-SIS, our software provides calibration and response factor graphs, which are ready for publication and customizable in many details (e.g., regarding figure captions, legends, and resolution).

## 6. Outlook and Conclusion

Future tasks of development include implementation of a web-based CalibraCurve version. Furthermore, provision of CalibraCurve as a Skyline External Tool is also intended. Skyline integration is beneficial for users who want to extent their existing Skyline workflows with capabilities provided by CalibraCurve.

To the best of our knowledge, there is no software with a similar feature combination. Thus, we deem CalibraCurve a highly useful tool to facilitate a batch-mode creation of calibration curves, response factor plots, and to determine both dynamic ranges and quantification limits of targeted proteomics assays. Hence, CalibraCurve gives valuable assistance in the development and control of reliable SRM/MS-based proteomics assays. The software is freely available under the three clause BSD license and included in the Supporting Information. Current and future CalibraCurve versions are also available from our website (<https://www.ruhr-uni-bochum.de/mpc/software/CalibraCurve/index.html>, Accessed March 2, 2020) and from GitHub (<https://github.com/mpc-bioinformatics/CalibraCurve>, Accessed March 2, 2020).

## Supporting Information

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

## Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) grants “de.NBI” (grant number FKZ 031 A 534A) and “i:DSem-Verbundprojekt: Electronic Patient Path (EPP)” (grant number FKZ 031 L 0025A).

## Conflict of Interest

The authors declare no conflict of interest.

## Keywords

calibration, limit of quantification, linear range, multiple-reaction monitoring, targeted proteomics

Received: June 11, 2019  
Revised: February 7, 2020  
Published online: March 6, 2020

- [1] C. Ludwig, L. Gillet, G. Rosenberger, S. Amon, B. C. Collins, R. Aebersold, *Mol. Syst. Biol.* **2018**, *14*, e8126.
- [2] C. M. Colangelo, L. Chung, C. Bruce, K. H. Cheung, *Methods* **2013**, *61*, 287.
- [3] S. E. Abbatiello, B. Schilling, D. R. Mani, L. J. Zimmerman, S. C. Hall, B. MacLean, M. Albertolle, S. Allen, M. Burgess, M. P. Cusack, M. Gosh, V. Hedrick, J. M. Held, H. D. Inerowicz, A. Jackson, H. Keshishian, C. R. Kinsinger, J. Lyssand, L. Makowski, M. Mesri, H. Rodriguez, P. Rudnick, P. Sadowski, N. Sedransk, K. Shaddox, S. J. Skates, E. Kuhn, D. Smith, J. R. Whiteaker, C. Whitwell, et al., *Mol. Cell. Proteomics* **2015**, *14*, 2357.
- [4] C. Galitzine, J. D. Egerton, S. Abbatiello, C. M. Henderson, L. K. Pino, M. MacCoss, A. N. Hoofnagle, O. Vitek, *Mol. Cell. Proteomics* **2018**, *17*, 913.
- [5] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria **2018**. <https://www.R-project.org/>
- [6] M. R. Berthold, N. Cebon, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Berlin, **2007**.
- [7] ISO 5725-1, **1994**, <https://www.iso.org/standard/11833.html>. Accessed February 27, 2020.
- [8] J. M. Green, *Anal. Chem.* **1996**, *68*, 305A.
- [9] A. G. González, M. Á. Herrador, *TrAC, Trends Anal. Chem.* **2007**, *26*, 227.
- [10] H. Gu, G. Liu, J. Wang, A. F. Aubry, M. E. Arnold, *Anal. Chem.* **2014**, *86*, 8959.
- [11] T. Bracht, V. Schweinsberg, M. Trippler, M. Kohl, M. Ahrens, J. Padden, W. Naboulsi, K. Barkovits, D. A. Megger, M. Eisenacher, C. H. Borchers, J. F. Schlaak, A. C. Hoffmann, F. Weber, H. A. Baba, H. E. Meyer, B. Sitek, *J. Proteome Res.* **2015**, *14*, 2278.
- [12] U.S. Food and Drug Administration, Guidance Document: Bioanalytical Method Validation Guidance for Industry, Center for Drug Evaluation and Research, **2018**. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry>. Accessed February 27, 2020.
- [13] Y. Mohammed, A. J. Percy, A. G. Chambers, C. H. Borchers, *J. Proteome Res.* **2015**, *14*, 1137.
- [14] B. MacLean, D. M. Tomazela, N. Shulman, M. Chambers, G. L. Finney, B. Frewen, R. Kern, D. L. Tabb, D. C. Liebler, M. J. MacCoss, *Bioinformatics* **2010**, *26*, 966.
- [15] QuaSAR, <https://skyline.ms/skyts/home/software/Skyline/tools/details.view?name=QuaSAR>. Accessed February 27, 2020.
- [16] M. Choi, C. Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean, O. Vitek, *Bioinformatics* **2014**, *30*, 2524.
- [17] P. Aiyetan, S. N. Thomas, Z. Zhang, H. Zhang, *BMC Bioinformatics* **2015**, *16*, 411.