analytical chemistry

pubs.acs.org/ac Article

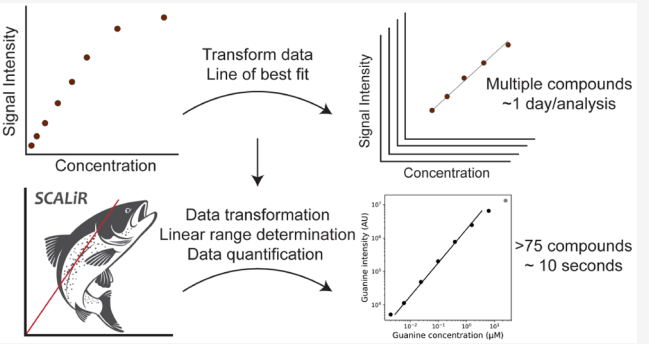
SCALiR: A Web Application for Automating Absolute Quantification of Mass Spectrometry-Based Metabolomics Data

Luis F. Ponce,[‡] Stephanie L. Bishop,[‡] Soren Wacker, Ryan A. Groves, and Ian A. Lewis*

Cite This: https://doi.org/10.1021/acs.analchem.3c04988 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Quantitative liquid chromatography−mass spectrometry (LC-MS)-based metabolomics is becoming an important approach for studying complex biological systems but presents several technical challenges that limit its widespread use. Computing metabolite concentrations using standard curves generated from standard mixtures of known concentrations is a labor-intensive process that is often performed manually. Currently, there are few options for open-source software tools that can automatically calculate metabolite concentrations. Herein, we introduce SCALiR (standard curve application for determining linear ranges), a new web-based software tool specifically built for this task, which allows users to automatically transform LC-MS signals into absolute quantitative data (https://www.lewisresearchgroup.org/software). SCALiR uses an algorithm that automatically finds the equation of the line of best fit for each standard curve and uses this equation to calculate compound concentrations from the LC-MS signal. Using a standard mix containing 77 metabolites, we show a close correlation between the concentrations calculated by SCALiR and the expected concentrations of each compound (R^2 = 0.99 for a $y = x$ curve fitting). Moreover, we demonstrate that SCALiR reproducibly calculates concentrations of midrange standards across ten analytical batches (average coefficient of variation 0.091). SCALiR can be used to calculate metabolite concentrations either using external calibration curves or by using internal standards to correct for matrix effects. This open-source and vendor agnostic software offers users several advantages in that (1) it requires only 10 s of analysis time to compute concentrations of >75 compounds, (2) it facilitates automation of quantitative workflows, and (3) it performs deterministic evaluations of compound quantification limits. SCALiR therefore provides the metabolomics community with a simple and rapid tool that enables rigorous and reproducible quantitative metabolomics studies.



Metabolomics is a mainstream approach for studying complex biological systems, ranging from cancer,[1] infectious diseases,[2] host-microbiome interactions,[3] and microbial engineering.[4,5] A common thread among these diverse disciplines is that they all need to accurately identify and quantify molecules in complex biological mixtures. Although such analyses are becoming more common in many metabolomics facilities, collecting absolute quantitative data for metabolites analyzed via routine liquid chromatography−mass spectrometry (LC-MS) analyses remains challenging.[6−8]

One of the main complications of LC-MS is that the response factor of each instrument varies day-to-day and sample-to-sample, and as a result, standard reference compounds used to calibrate signal intensities need to be acquired frequently to enable robust quantification.[9] The additional logistical complications of acquiring and analyzing data for standard reference materials currently make absolute quantitative metabolomics studies the exception rather than the rule, and the subsequent use of nonquantitative methods limits direct comparisons between data sets obtained in different batches. Conversely, absolute quantification of metabolite signals makes it possible to directly compare data across large-cohort studies and facilitates the analysis of metabolic flux through networks.[6] The conventional practice for acquiring these quantitative data is to prepare matrix-matched mixtures of metabolite standards at a range of physiologically relevant concentrations and use the empirical regression coefficients observed in intensity versus concentration plots of the standards to compute the concentrations of analytes in test mixtures. This external calibration process is generally robust when analytes in the test mixtures are within the linear range of intensity versus concentration responses and when the matrix conditions are well matched between external calibrants and test samples. These calibrations can be further

Received: November 4, 2023
Revised: February 29, 2024
Accepted: April 12, 2024

improved via isotope dilution, wherein stable isotope-labeled standards are included in each test sample to correct for matrix effects.

Although there are many tools and resources available for metabolomics data visualization and interpretation,[10] there is a critical gap in software for performing absolute quantification of LC-MS metabolomics data. Several commercial and open-source tools (e.g., Thermo Fisher's TraceFinder, Skyline for small molecules,[11] targeted metabolome batch quantification (TMBQ),[12] MRMkit,[13] MRMPROBS,[14] SWATHtoMRM,[15] and XCMS-MRM coupled to METLIN-MRM[16]) can determine quantitative metabolomics values from raw instrumental data. While these tools are important resources for metabolomics, many require extensive user input or coding to set up the workflows, internal standard data, or vendor-specific raw data files to complete these analyses. For these reasons, most investigators manually fit standard curves for their analyses. This process is time-consuming and requires a degree of expertise to perform it reproducibly. Moreover, there are no universally accepted parameters for determining the upper and lower limits of quantification (ULOQ and LLOQ, respectively) of a compound and ensuring the linearity of the curve fitting. The LLOQ is often defined by instrumental baseline noise and calculated with a numerical parameter such as 10% or a fixed multiple of the relative standard deviation of a blank matrix.[17−19] The LLOQ can also be calculated with a formula based on the percent deviation from the nominal concentration at the lowest calibrator concentration.[20] Furthermore, the process of determining the ULOQ can be complicated because numerical parameters of the curve fitting, such as homoscedasticity (equal distribution of residuals) and standard deviation of the regression, need to be considered to ensure a valid linear fit across the range of standard concentrations.[21] These challenges make the process of fitting standard curves difficult to automate and encourage investigators to omit metabolite quantification from their studies.

Therefore, there is an urgent need for software that automatically detects the upper and lower limits of quantification and the line of best fit for a series of standards and uses these data to calculate the concentrations of metabolites in a sample. Herein, we introduce SCALiR (standard curve application for determining linear ranges), a new tool designed specifically for this task. This software uses a novel algorithm that leverages a basic logarithmic property and general characteristics of LC-MS data to automatically determine the ULOQ, LLOQ, and line of best fit for external calibration standards. Using these values, SCALiR automatically transforms LC-MS peak data into absolute quantitative data. Unlike some existing tools, SCALiR is vendor- and mass-analyzer agnostic, requires only 10 s to analyze large data sets, and allows users to visualize resulting data directly on the user interface to ensure validity of the standard curve fitting. We show that SCALiR automatically generates data comparable to those obtained via a manual curve fitting by an expert analyst. We then illustrate how the tool can be used to correct batch effects in large cohorts and to facilitate metabolic boundary flux analyses.

EXPERIMENTAL SECTION

Preparation of Mixed Metabolite Standards. All standard stock solutions were prepared from compounds ordered from Sigma-Aldrich (Oakville, ON, Canada), VWR (Edmonton, AB, Canada), or Acros Organics (now Thermo Scientific Chemicals, Waltham, MA, USA), see Table S1 for compound CAS numbers. To evaluate the linear dynamic range of each compound manually, we developed a mixed standard containing 77 compounds with variable starting concentrations and ran a 10-point dilution series with three technical replicates, where each sequential standard was diluted 4-fold (1:4) with 50/50 (v/v) methanol (Fisher Optima LC/MS grade; Toronto, ON, Canada) and water (Fisher Optima LC/MS grade). Stock standard concentrations for each compound are found in Table S1.

Liquid Chromatography−Mass Spectrometry. All liquid chromatography−mass spectrometry (LC-MS) metabolomics data were acquired at the Calgary Metabolomics Research Facility (CMRF), according to the methods described in detail in refs[2,9]. Metabolite samples were resolved via a Thermo Scientific Vanquish UHPLC (Thermo Fisher Scientific) platform using hydrophilic interaction liquid chromatography (HILIC) with a 15 min gradient. Chromatographic separation was attained using a binary solvent mixture of 20 mM ammonium formate at pH 3.0 in LC-MS grade water (Solvent A) and 0.1% formic acid (% v/v) in LC-MS grade acetonitrile (Solvent B) in conjunction with a 100 mm (length) × 2.1 mm (diameter) Syncronis HILIC column (Thermo Fisher Scientific) with a 1.7 μm particle size. Data were acquired on a Thermo Scientific Q Exactive HF (Thermo Fisher Scientific) mass spectrometer in negative ionization mode.

Manual Data Analysis. All LC-MS raw data files were converted to mzXML format via MSConvert GUI software.[22] LC-MS analysis was conducted in El-Maven, v.0.12.0[23] with manual peak selection using default parameters (Peak Quantitation Type: Area top; Mass Cutoff unit: 10.0 ppm). All data were then collated in Microsoft Excel, where the linear range for each compound was calculated according to visual inspection of the linearity and goodness of fit. Select performance characteristics according to the U.S. FDA guidelines for bioanalytical method validation[20] for measured compounds can be found in Table S2.

Bacterial Sample Preparation and Analysis. *Staphylococcus aureus* isolates were collected by the Alberta Precision Laboratories (Calgary, AB, Canada) and were prepared as described in ref[9]. Briefly, microbes were inoculated in Mueller-Hinton broth in 96-well plates and cultured aerobically overnight at 37 °C in a humidified incubator with a 5% CO_2 and 21% O_2 atmosphere. After reaching an optical density between 0.35 and 0.4, culture supernatants were fixed in methanol (1:1 v/v) and centrifuged for 5 min at 4,000 × *g*. The supernatants were then diluted 1:10 with 50% methanol before LC-MS analysis.

The concentrations of metabolites in these samples were quantified using the mixed metabolite standards (see above), with each point prepared as an 8-point dilution series. The highest concentration standard was prepared as a 2-fold (1:2) dilution in 50/50 (v/v) methanol (Fisher Optima LC/MS grade) and water (Fisher Optima LC/MS grade) from the stock standard concentrations described in Table S1. The next seven dilutions were prepared as 4-fold (1:4) dilutions in the same 50% methanol solvent. LC-MS data were acquired as described above. Two 100 × 2.1 mm Syncronis HILIC columns (Thermo Fisher Scientific) were used to analyze multiple batches. All data were quantified manually, as described above, and compared to automated data quantification via SCALiR.

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

Implementation, Statistical Analysis, and Visualization. The SCALiR backend was implemented in Python (Version 3) using the standard libraries pandas, numpy, and matplotlib. The user interface was created on Streamlit, and the source code is available in GitHub (https://github.com/LewisResearchGroup/ms-conc). All statistical analysis of data was performed in Python (Version 3) or GraphPad Prism (Version 9.5.1). Figures were created using Python, GraphPad Prism, Adobe Illustrator (Version 27.2), and Inkscape (Version 1.2) and by directly downloading images and screenshots from the SCALiR app.

Safety Considerations. Reagents used in this investigation do not pose any significant safety risks outside of those experienced regularly when working with moderately strong acids and solvents.

RESULTS AND DISCUSSION

SCALiR Algorithm. One of the principal challenges in quantitative metabolomics is the lack of a standardized or deterministic method for computing the LLOQ and ULOQ for a series of standards. The algorithm incorporated into SCALiR takes advantage of a basic logarithmic property and an intrinsic characteristic of LC-MS signals that allows the curve fitting process to be generalized. Specifically, in linear data sets where the slope is much greater than the intercept, the log−log transformation of the linear series will have a slope approaching one. This transformation is detailed in eq 1.

Robust calibration reference standards follow a relationship wherein signal intensities (y) and metabolite concentrations (x) are described by a linear model:

$$y = m \times x + b \tag{1}$$

where m denotes the slope and b denotes the y-intercept. This relationship can be transformed as follows:

$$\ln(y) = \ln(m \times x + b) \tag{2}$$

$$\ln(y) = \ln\left(m\left(x + \frac{b}{m}\right)\right) \tag{3}$$

$$\ln(y) = \ln(\hat{x}) + \ln(m) \tag{4}$$

where

$$\hat{x} = x + \frac{b}{m} = \left(1 + \frac{b}{(m \times x)}\right)$$

When $b \ll (m \times x)$, in eq 4, $\hat{x} \to x$ and becomes

$$\ln(y) = \ln(x) + \ln(m) \tag{5}$$

which can be rewritten as

$$y^* = 1 \times x + b^* \tag{6}$$

where y^* is the LC-MS signal in the *ln* scale, the slope is equal to 1, and b^* is ln (m).

To illustrate this point, we calculated the intercept fraction value, $b/(m \times \bar{x})$, for ten metabolites across ten batches using the average concentration of the five lowest concentration standards of an 8-point standard curve to ensure points fell below the ULOQ (excluding points with no signal). As shown in Figure 1a, these values fall within the range (−0.1, 0.1), and this distribution range holds for 80% of the metabolites we measured (Figure 1b). When $b \ll (m \times x)$, the curve in the linear scale (eq 1) is transformed into a linear curve in the *ln* scale with a slope equal to 1 and the intercept equal to the ln of the slope in the linear scale (Figure 1c). Compounds that exhibit deviations from this linear relationship, which result in a less adequate fit after the ln transformation with a slope far from 1 and an intercept fraction value ≫0 (examples shown in Figure S1), are automatically flagged as a "failed fit" by SCALiR.

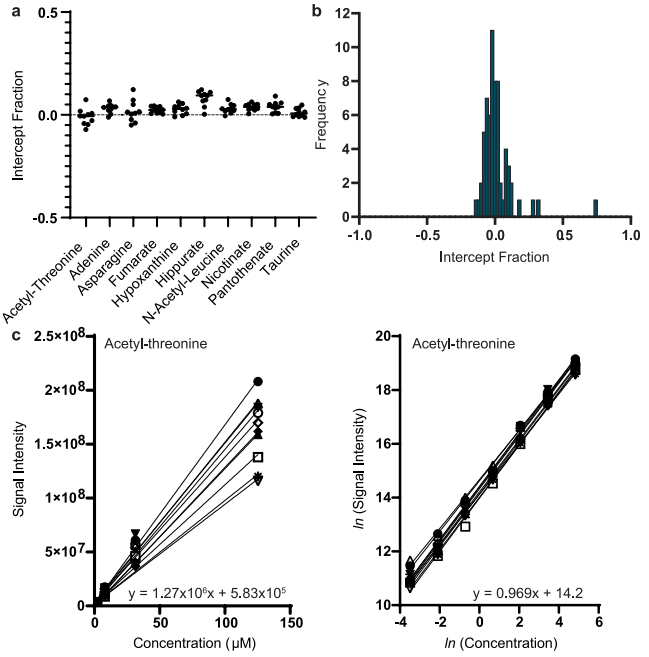


Figure 1. (a) Examples of intercept fraction values ($b/(m \times \bar{x})$) for the standard curves of ten metabolites run over ten batches. (b) The distribution of intercept fraction values for standard curves of all metabolites measured. (c) Examples of a set of ten standard curves for acetyl-threonine in the linear and *ln* scale. Each curve corresponds to one batch, and the line of best fit equation shows an example of the slope and y-intercept from one standard curve.

The algorithm used by SCALiR incorporates these properties to determine the ULOQ and LLOQ for the standard curve as well as an iterative fitting process to find the line of best fit. The goodness of fit for the standard curve is controlled by the residual function (r) between the expected (e) and observed (o) values defined as

$$r = \frac{1}{N}\sum_{i=1}^{N}(e - o)^2 \tag{7}$$

where N is the number of points in the data set.

When r is calculated to be above the empirically determined threshold value of 0.01 (Figure S2), the point with the largest distance between the predicted and the experimental value is removed from the data set (Figure 2). This step is repeated until the value of r falls below the set threshold, indicating acceptable goodness of fit or until a minimum of three points remain on the curve. SCALiR reports the number of points used in the fitting in the standard curve parameters download file (N_points column).

SCALiR App Design. The web-based application SCALiR (https://www.lewisresearchgroup.org/software) is designed to work with peak data files generated from open-source software El-Maven[23] or MINT (https://mint.resistancedb.org/) without further modifications. Peak data from any software can be used if the peak data file follows the formatting in the sample

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

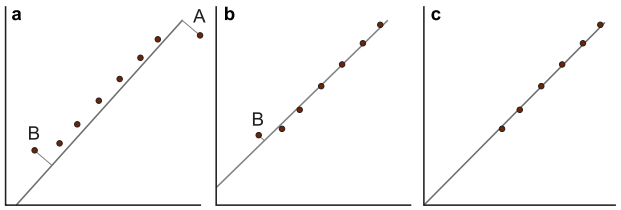
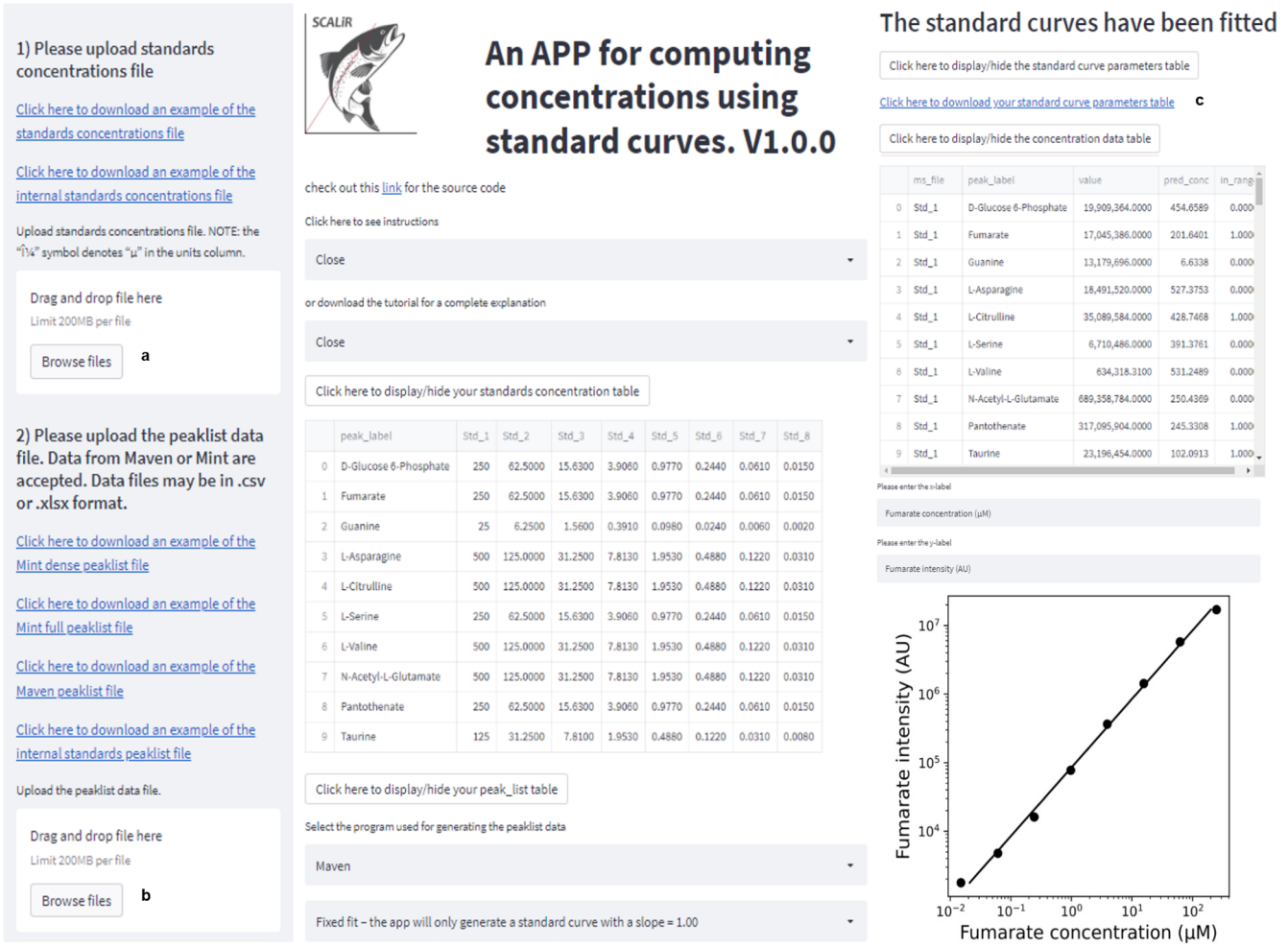


Figure 2. Iterative regression algorithm for finding the linear range of a series of standards. (a) The best fit curve is generated from the data set. (b) When the residual (r) of the curve is >0.01, the farthest point from the curve ("A") is removed and this step is repeated for the next furthest point from the curve ("B"). This process is carried out until (c) the data allow a linear regression fitting that achieves the quality threshold for r. Standard curves containing <5 points should be considered semiquantitative.

files accessed directly from the app. For quantification via spiking of an internal standard in both the standard mix and samples, another peak data file format enables users to upload response ratio values calculated from the ratio of the peak area of the analyte to the peak area of the internal standard. Sample data upload files for standard concentrations (or the ratio of the analyte concentration to the internal standard concentration) can also be accessed directly from the app. Instructions on how to use the app are included on the web application interface, and a tutorial with demo data files for upload can be accessed via the web interface and are included in the Supporting Information. Users can also document issues or suggestions for the app at https://github.com/LewisResearchGroup/ms-conc/issues.

Figure 3 shows key steps and features of the app. Users upload a comma-separated values (CSV) or Excel file containing concentrations of each standard (Figure 3a) as well as a CSV or Excel file with peak list information for standards and samples (Figure 3b). The user then selects which program was used to generate the peak list information, chooses the fixed slope, wide slope, or interval slope option, and can adjust the permissible values of the slope for the interval slope option. Once the program is run, the user can download the standard curve parameters, including slope, intercept, and linear range minimum and maximum for each compound, as well as the calculated concentrations for standards and samples (Figure 3c). Additionally, the user can view log−log plots showing the standard curve for each



	peak_label	Std_1	Std_2	Std_3	Std_4	Std_5	Std_6	Std_7	Std_8
0	D-Glucose 6-Phosphate	250	62.5000	15.6300	3.9060	0.9770	0.2440	0.0610	0.0150
1	Fumarate	250	62.5000	15.6300	3.9060	0.9770	0.2440	0.0610	0.0150
2	Guanine	25	6.2500	1.5600	0.3910	0.0980	0.0240	0.0060	0.0020
3	L-Asparagine	500	125.0000	31.2500	7.8130	1.9530	0.4880	0.1220	0.0310
4	L-Citrulline	500	125.0000	31.2500	7.8130	1.9530	0.4880	0.1220	0.0310
5	L-Serine	250	62.5000	15.6300	3.9060	0.9770	0.2440	0.0610	0.0150
6	L-Valine	500	125.0000	31.2500	7.8130	1.9530	0.4880	0.1220	0.0310
7	N-Acetyl-L-Glutamate	500	125.0000	31.2500	7.8130	1.9530	0.4880	0.1220	0.0310
8	Pantothenate	250	62.5000	15.6300	3.9060	0.9770	0.2440	0.0610	0.0150
9	Taurine	125	31.2500	7.8100	1.9530	0.4880	0.1220	0.0310	0.0080

	ms_file	peak_label	value	pred_conc	in_rang
0	Std_1	D-Glucose 6-Phosphate	19,909,364.0000	454.6589	0.000
1	Std_1	Fumarate	17,045,386.0000	201.6401	1.000
2	Std_1	Guanine	13,179,696.0000	6.6338	0.000
3	Std_1	L-Asparagine	18,491,520.0000	527.3753	0.000
4	Std_1	L-Citrulline	35,089,584.0000	428.7468	1.000
5	Std_1	L-Serine	6,710,486.0000	391.3761	0.000
6	Std_1	L-Valine	634,318.3100	531.2489	0.000
7	Std_1	N-Acetyl-L-Glutamate	689,358,784.0000	250.4369	0.000
8	Std_1	Pantothenate	317,095,904.0000	245.3308	1.000
9	Std_1	Taurine	23,196,454.0000	102.0913	1.000

Figure 3. Features of the web-based application SCALiR. Users upload standard concentrations (a) and peak list data files (b) and select settings for generating the standard curves. (c) SCALiR performs its iterative fitting algorithm, stopping when the fitting reaches the residual threshold ($r \leq 0.01$). The remaining maximum and minimum concentration values are reported as the ranges for the linear behavior. Users can download the results for standard curve parameters and concentration data and visualize individual standard curves as log−log plots. All calculations were performed with the fixed slope feature.

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

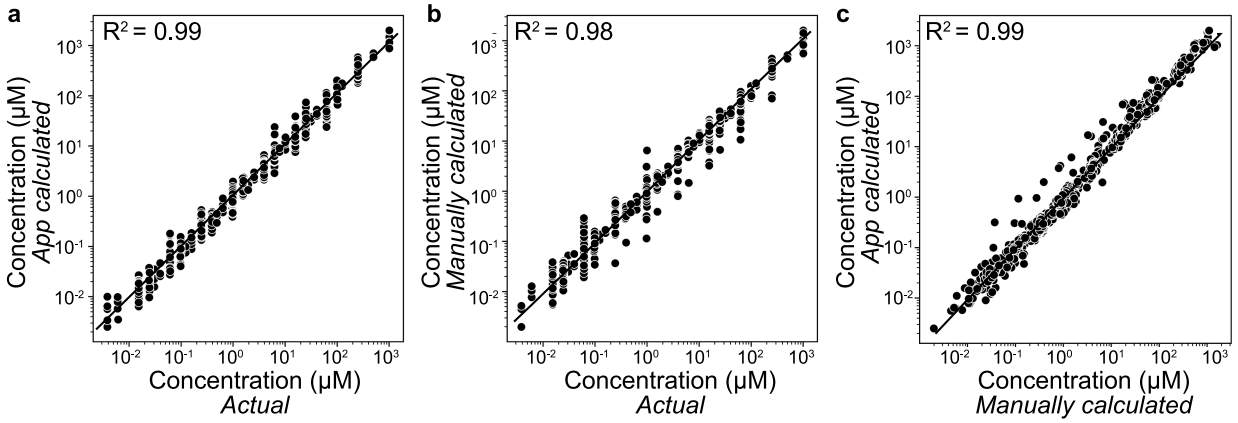


Figure 4. Validation of the algorithm used in SCALiR for calculating concentrations of compounds in a standard mix. (a) Comparison between the app-calculated concentrations in the standards and the actual concentrations for each compound. (b) Comparison between the manually calculated concentrations and actual concentrations for each compound. (c) Comparison between the app-calculated concentrations in the standard samples and manually calculated concentrations.

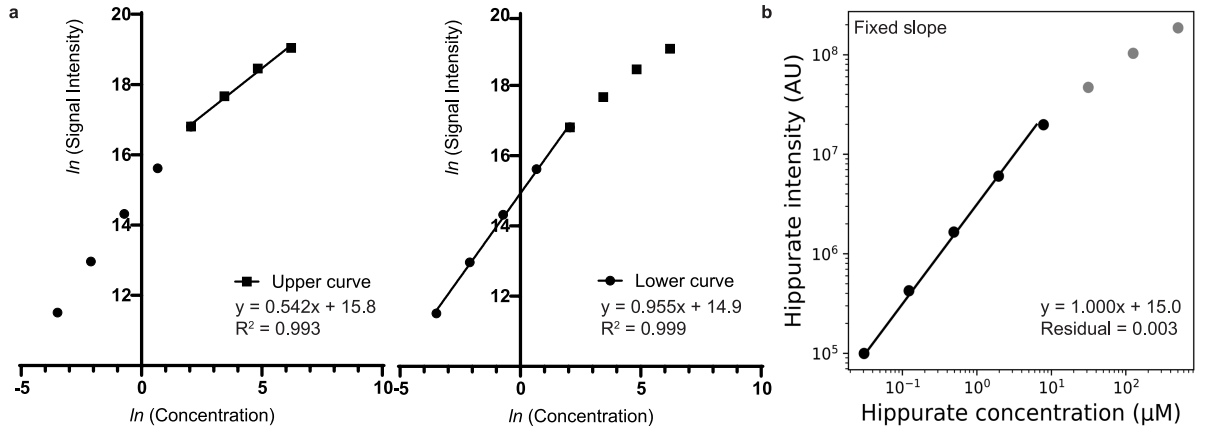


Figure 5. SCALiR simplifies the process for determining the linear range of a standard series in which there are two or more distinct linear regimes. The 8-point standard curve for hippurate in the natural logarithmic scale is shown as an example in (a), which can result in two possible linear regions with a high R^2 value (>0.99). (b) Log−log plot downloaded from SCALiR, which uses a deterministic fitting algorithm that sets the slope of the regression curve to 1.000 in the *ln* scale with the fixed slope option and stops the fitting when the residual of the curve <0.01.

compound and download individual images of each plot with the option to either include axis labels or leave them blank before downloading the images.

Performance Evaluation. To assess the performance of SCALiR, we compared the concentrations calculated by the app to the actual concentration values of each metabolite in an 8-point standard dilution series (Figure 4a; supplementary data file). The data shown here are from the mixed standard containing 77 metabolites, and all validation data are shown with SCALiR's fixed slope feature. Similar results were obtained when we used the wide slope option. The concentration values calculated by the app correlated closely with the actual concentration values of the standards (R^2 = 0.99 in the *ln* scale; Figure 4a), demonstrating that SCALiR's algorithm accurately determines the linear range and concentrations of metabolites. This value was comparable to the correlation between calculated and actual concentrations obtained by an expert analyst that manually fitted standard curves (R^2 = 0.98 in the *ln* scale; Figure 4b). Third, we compared SCALiR against concentration values generated by an expert analyst manually (Figure 4c). Again, we found an excellent correlation between the concentration values calculated by SCALiR and those calculated by manual inspection (R^2 = 0.99 in the *ln* scale; Figure 4c), indicating that SCALiR's algorithm performs very similarly to a trained analyst.

SCALiR also demonstrated reproducible results across ten separate analytical batches (Table S3). We calculated that the average coefficient of variation (CV) for the middle four standards in the 8-point standard curves (Standards 3−6) of ten representative compounds was between 0.054 and 0.152, with an average CV of 0.091 for all ten compounds. Additionally, 85% of the SCALiR-calculated concentrations of these standards met the US Food and Drug Administration (FDA) guidelines stating that nonzero calibrators should be within ±15% of the nominal (expected) concentration[20] (Table S3).

Analytical Challenges Solved by SCALiR. One of the major advantages of using SCALiR is its deterministic process for evaluating the linear range of a standard series, which is based on empirically determined parameters. Manual evaluation of quantification limits not only is time-consuming but also can lead to variation in data analysis results between analysts and over separate batches. This can result in

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

differences in quantification limits and can affect the accuracy of the calculated sample concentrations. Figure 5 shows an example of challenges that can occur when a standard series displays multiple linear regimes or regions of the curve that display a linear fit with a coefficient of determination (R^2) > 0.99. Using hippurate as an example, we show that an 8-point standard series transformed to the *ln* scale can display two or more regions that would have a valid linear fit based on the coefficient of determination parameter (Figure 5a). However, only two of these regions satisfy the linear relationship between signal intensity and concentration required for the valid *ln* transformation showing a slope of the regression curve equal to approximately 1 (lower curve in Figure 5a).

SCALiR automatically determines the region of the curve that both satisfies the linear relationship requirement (slope of the *ln* transformed curve = 1) and provides a high-quality fit (residual of the curve <0.01) (Figure 5b). An example of a log−log plot downloaded from the app shows the standard curve of an individual metabolite, hippurate. In summary, SCALiR's deterministic algorithm standardizes the process of determining a linear range for a standard series when multiple valid linear fittings are possible and thus can lead to more reproducible quantitative results in a fraction of the analysis time of traditional methods.

Applications. SCALiR can alleviate common quantitative LC-MS challenges including batch effects and has clear utility for a variety of quantitative applications, including analyzing cellular boundary flux,[24−26] an approach which is gaining traction as a foundation for clinical diagnostics,[2,27] analyzing metabolic interactions within microbial communities,[28,29] and assessing the microbial production of biofuels.[30,31] In this section, we show two examples of applications of SCALiR in common analytical and biological problems. Batch effects are mainly caused by instrumental drift in signal over time and are exacerbated by varied analytical conditions including different chromatography columns or solvent batches.[9] In this first example, we show how SCALiR can minimize a column-induced data offset arising from a multiplexed strategy that uses two separate chromatography columns to enable high-throughput analysis of over 3,000 injections of the same sample of *S. aureus* growth medium[9] (Figures 6 and S3). Clear column-specific effects were apparent when we compared only the signal intensity values for arginine in a data set (Figure 6a). Using SCALiR to calculate concentrations of arginine in each batch, we generated a stable result across repeated injections of the sample irrespective of the column used (Figure 6b).

Additionally, we used SCALiR to perform a metabolic boundary flux analysis of the Mueller Hinton growth medium components consumed by *S. aureus* in the large-cohort study described previously (Figures 7 and S4). Using our quantitative approach, we revealed different metabolic phenotypes that were not discernible using LC-MS signal data alone. For example, by employing quantitative methods we found that *S. aureus* cells mostly consume serine, arginine, and trehalose as opposed to using other carbon and nitrogen sources identified from nonquantitative methods calculated from the normalized peak area (Figure 7a,b). Using the quantitative approach, we also observed that some metabolites, such as glutamate and aspartate, despite their high availability, are not preferentially consumed by *S. aureus* (Figure 7c), suggesting that they could be removed from growth media without perturbing cell growth. Using SCALiR, we were therefore able to quickly infer the main metabolic routes used

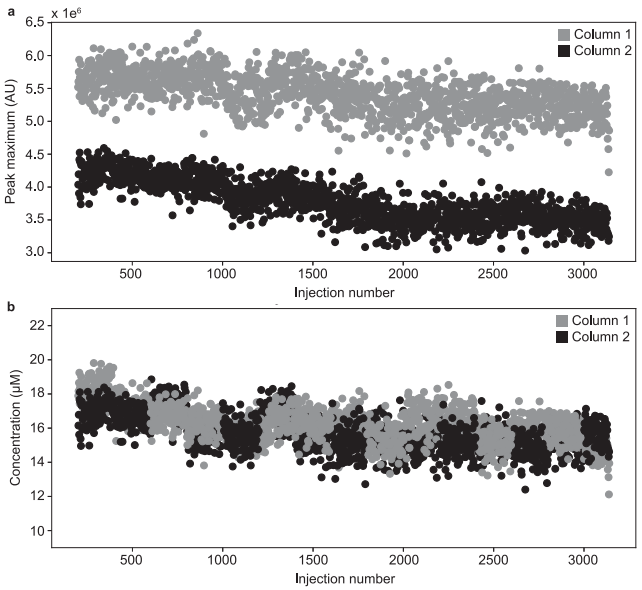


Figure 6. SCALiR reduces column-induced effects across 3,000 injections of the same sample in a *S. aureus* data set with a multiplexed chromatography strategy. The dot plot shows a comparison of arginine (a) peak maximum values (arbitrary units) and (b) concentration values (μM) in the sample injected across two columns.

by the cells for optimal growth (Figure 7d). By facilitating quantitative metabolomics studies, SCALiR represents an important new tool that may soon serve as a new standard for a wide range of metabolomics applications, including metabolic network modeling, which facilitates biomarker discovery for a variety of conditions such as Alzheimer's disease[32] and cancer,[33] as well as studies of plant secondary metabolism.[34]

Limitations. One limitation of SCALiR is that it only computes the line of best fit for linear data using the equation $y = m \times x^1 + b$ and uses a logarithmic transformation which may not be appropriate for all data sets.[21] However, users can also select the wide slope option to include the most points possible in the linear range or the interval defined option, where the user sets a permissible range of slopes to fit the standard curve. We recommend using the parameter with the slope equal to 1 or a narrowly defined interval (0.85−1.15).

As with any method used to calculate concentrations from a standard curve, SCALiR can provide accurate values only for concentrations in the respective linear range for a given compound. The use of external standard curves to calculate metabolite concentrations produces accurate concentrations when the analyte is measured within this linear range and matrix effects including ion suppression are minimal.[6] Therefore, users should carefully consider sample preparation and chromatography methods to limit concentration biases resulting from sample matrix effects; other methods, including standard addition or internal standards, may be more suitable in the case of large matrix-induced biases. If users wish to use an internal standard method, SCALiR provides the option to upload response ratio and concentration ratio data to calculate the concentrations of the analytes. We recommend that users evaluate results carefully if they choose to calculate concentrations of metabolites using the standard curve parameters from a different but structurally similar metabolite, as this can lead to greater than 100% error in calculated

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

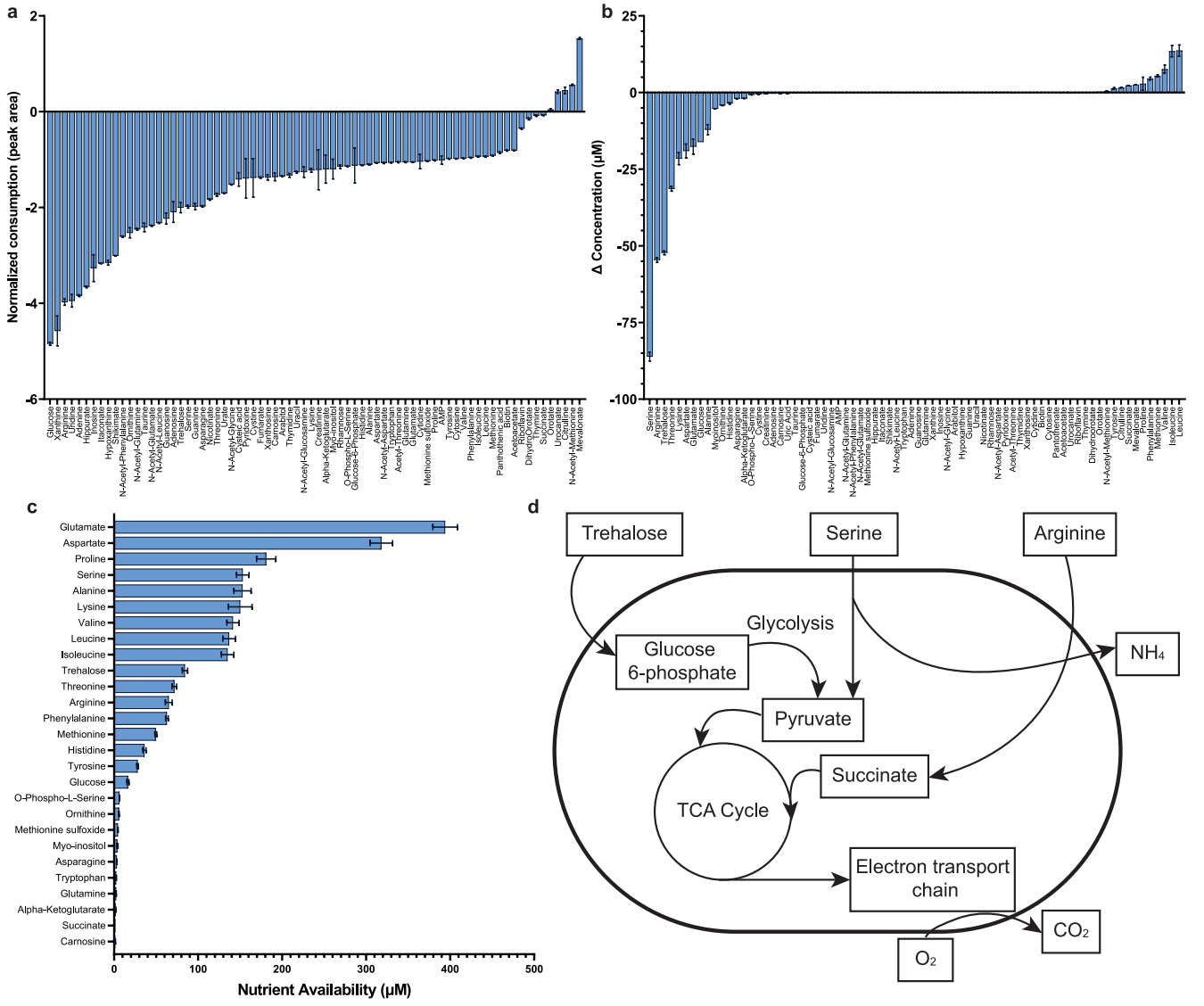


Figure 7. Boundary flux analysis of *S. aureus* clinical isolates grown in Mueller Hinton (MH) growth medium. (a) Metabolite consumption/secretion assessment by the ratio of peak intensities in the microbial supernatant and MH medium. (b) Metabolite consumption/secretion analysis using concentration differences. (c) Nutrient availability in MH medium. (d) Proposed metabolic pathways for optimal *S. aureus* growth according to nutrient consumption and secretion data.

concentration values (see Table S4 for examples). With these considerations in mind, SCALiR provides users with a new tool to generate quantitative metabolomics data over a wide range of distinct compounds and concentration ranges.

CONCLUSION

Quantitative metabolomics is rapidly emerging as an important component of many metabolomics research programs; however, the complexity and time required to quantify metabolites have limited its widespread use. SCALiR provides users with three main advantages compared to the traditional method of manually fitting standard curves for quantitative metabolomics: (1) reduced analysis time, requiring only 10 s to compute concentrations of >75 compounds; (2) automation of quantitative workflows with minimal training or computational expertise; and (3) deterministic evaluation of compound quantification limits and line of best fit, reducing analytical subjectivity. Additionally, SCALiR is open-source and vendor agnostic. SCALiR fits an important need in the metabolomics community and opens the door to the routine and widespread use of quantitative metabolomics workflows.

ASSOCIATED CONTENT

Data Availability Statement

All data are available in the Supporting Information of this manuscript or at https://pubs.acs.org/doi/10.1021/acs.analchem.2c00078?ref=PDF. SCALiR source code is available at https://www.lewisresearchgroup.org/software.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c04988.

Table S1. Compound information and highest standard concentration for mixed metabolite standard dilution series, Table S2. Performance characteristics for visual inspection of standard curves, according to U.S. FDA Bioanalytical Method Validation Guidelines, Figure S1.

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX

Examples of standard curves of compounds that do not exhibit a linear relationship between signal and concentration, Figure S2. Effect of using different threshold residual values (T) to stop the algorithm's fitting process for hypoxanthine, using a 10-point standard curve with 3 technical replicates, Table S3. SCALiR performance evaluation results for 10 representative compounds, showing the concentrations calculated by the SCALiR app compared to the expected concentration in an 8-point standard curve, Figure S3. Dot plots showing comparison of arginine (a) peak maximum values (arbitrary units) and (b) concentration values (μM) in samples across two columns, Figure S4. Heatmap of bacterial metabolomics data set showing normalized metabolite concentrations in one batch of samples, and Table S4. Accuracy errors resulting from using the line of best from the standard curve of a structurally similar compound to calculate concentrations (PDF)

Bacterial metabolomics concentrations comparison manual vs SCALiR data file (XLSX)

Bacterial metabolomics samples and standards SCALiR concentrations data file (XLSX)

Raw signal intensity data manual standard curves Maven format peak list negative ionization mode (XLSX)

SCALiR tutorial (PDF)

SCALiR standards concentrations file (XLSX)

SCALiR internal standards concentrations file (XLSX)

SCALiR MINT peaklist dense peak max (XLSX)

SCALiR MINT peaklist full results (XLSX)

SCALiR Maven peaklist (XLSX)

SCALiR internal standards peaklist (XLSX)

AUTHOR INFORMATION

Corresponding Author

Ian A. Lewis − *Alberta Centre for Advanced Diagnostics, Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada;* orcid.org/0000-0002-5753-499X; Email: ian.lewis2@ucalgary.ca

Authors

Luis F. Ponce − *Alberta Centre for Advanced Diagnostics, Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada*

Stephanie L. Bishop − *Alberta Centre for Advanced Diagnostics, Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada;* orcid.org/0000-0002-8134-7297

Soren Wacker − *Alberta Centre for Advanced Diagnostics, Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada;* orcid.org/0000-0003-3201-1233

Ryan A. Groves − *Alberta Centre for Advanced Diagnostics, Department of Biological Sciences, University of Calgary, Calgary, AB T2N 1N4, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c04988

Author Contributions

‡L.F.P. and S.L.B. contributed equally to this manuscript. I.A.L. and L.F.P. conceived of the project. L.F.P. and S.W. developed the algorithm and web application. S.L.B. and R.A.G. performed all experimental analyses. L.F.P. and S.L.B. performed all data analysis and wrote the primary draft of the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

I.A.L. is supported by the Alberta Centre for Advanced Diagnostics, an Alberta Innovates—Health Solutions (AIHS) Translational Health Chair, a 2016 Genome Canada Genomics Applied Partnership Program (GAPP), and a 2017 Genome Canada Large Scale Applied Research Project award (both administered by Genome Alberta), the Natural Sciences and Engineering Research Council (NSERC) [DG 04547]. S.L.B. is supported by a National Institutes of Health (NIH) grant [Project Number 1R01AI153521-01] and a Canadian Institutes of Health Research (CIHR) Postdoctoral Fellowship [Project Number 202110HIV-477488-87373]. S.W. is directly supported via a Genome Canada 2017 Large Scale Applied Research Project award administered through Genome Alberta. Data were acquired at the Calgary Metabolomics Research Facility (CMRF) at the Alberta Centre for Advanced Diagnostics, which is supported by PrairiesCan (000022734), the International Microbiome Centre and IMPACTT Microbiome Research Core (CIHR IMC-161484), and the Canada Foundation for Innovation (CFI-JELF 34986). This work was made possible in part by a research collaboration agreement with Thermo Fisher Scientific. We thank K. A. Pittman and N. van Bavel for their critical review and editorial assistance in preparing this manuscript.

REFERENCES

(1) Schmidt, D. R.; Patel, R.; Kirsch, D. G.; Lewis, C. A.; Locasale, J. W. *Ca Cancer J. Clin.* **2021**, *71* (4), 333−358.

(2) Rydzak, T.; Groves, R. A.; Zhang, R.; Aburashed, R.; Pushpker, R.; Mapar, M.; Lewis, I. A. *Nat. Commun.* **2022**, *13* (1), 2332.

(3) Bishop, S. L.; Drikic, M.; Wacker, S.; Chen, Y. Y.; Kozyrskyj, A. L.; Lewis, I. A. *Mucosal Immunol.* **2022**, *15* (6), 1071−1084.

(4) Jang, Y. S.; Woo, H. M.; Im, J. A.; Kim, I. H.; Lee, S. Y. *Appl. Microbiol. Biotechnol.* **2013**, *97* (21), 9355−9363.

(5) Martien, J. I.; Amador-Noguez, D. *Curr. Opin. Biotechnol.* **2017**, *43*, 118−126.

(6) Alseekh, S.; Aharoni, A.; Brotman, Y.; Contrepois, K.; D'Auria, J.; Ewald, J. C.; Ewald, J.; Fraser, P. D.; Giavalisco, P.; Hall, R. D.; et al. *Nat. Methods* **2021**, *18* (7), 747−756.

(7) Lu, W.; Su, X.; Klein, M. S.; Lewis, I. A.; Fiehn, O.; Rabinowitz, J. D. *Annu. Rev. Biochem.* **2017**, *86*, 277−304.

(8) Kapoore, R. V.; Vaidyanathan, S. *Phil. Trans. R. Soc. A.* **2016**, *374* (2079), 20150363.

(9) Groves, R. A.; Mapar, M.; Aburashed, R.; Ponce, L. F.; Bishop, S. L.; Rydzak, T.; Drikic, M.; Bihan, D. G.; Benediktsson, H.; Clement, F. *Anal. Chem.* **2022**, *24*, 8874−8882.

(10) Misra, B. B. *Metabolomics* **2021**, *17* (5), 49.

(11) Adams, K. J.; Pratt, B.; Bose, N.; Dubois, L. G.; St. John-Williams, L.; Perrott, K. M.; Ky, K.; Kapahi, P.; Sharma, V.; Maccoss, M. J. *J. Proteome Res.* **2020**, *19* (4), 1447−1458.

(12) Xie, G.; Wang, L.; Chen, T.; Zhou, K.; Zhang, Z.; Li, J.; Sun, B.; Guo, Y.; Wang, X.; Wang, Y.; et al. *Anal. Chem.* **2021**, *93* (14), 5709−5717.

(13) Teo, G.; Chew, W. S.; Burla, B. J.; Herr, D.; Tai, E. S.; Wenk, M. R.; Torta, F.; Choi, H. *Anal. Chem.* **2020**, *92* (20), 13677−13682.

(14) Tsugawa, H.; Arita, M.; Kanazawa, M.; Ogiwara, A.; Bamba, T.; Fukusaki, E. *Anal. Chem.* **2013**, *85* (10), 5191−5199.

(15) Zha, H.; Cai, Y.; Yin, Y.; Wang, Z.; Li, K.; Zhu, Z. J. *Anal. Chem.* **2018**, *90* (6), 4062−4070.
(16) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Ivanisevic, J.; Thomas, A.; Sidibé, J.; Teav, T.; Guijas, C.; Aisporna, A. E.; Rinehart, D.; Hoang, L.; et al. *Nat. Methods* **2018**, *15* (9), 681−684.
(17) Bachmanov, A. A.; Reed, D. R.; Li, X.; Beauchamp, G. K. *Pure Appl. Chem.* **2002**, *74* (7), 1135−1140.
(18) EPA *Definition and Procedure for the Determination of the Method Detection Limit, Revision 2*; EPA, 2016.
(19) Betz, J. M.; Brown, P. N.; Roman, M. C. *Fitoterapia* **2011**, *82* (1), 44−52.
(20) U.S. FDA *Bioanalytical Method Validation Guidance For Industry*; U.S. FDA, 2018.
(21) Jurado, J. M.; Alcázar, A.; Muñiz-Valencia, R.; Ceballos-Magaña, S. G.; Raposo, F. *Talanta* **2017**, *172* (May), 221−229.
(22) Chambers, M. C.; MacLean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. *Nat. Biotechnol.* **2012**, *30* (10), 918−920.
(23) Melamud, E.; Vastag, L.; Rabinowitz, J. D. *Anal. Chem.* **2010**, *82* (23), 9818−9826.
(24) Hui, S.; Cowan, A. J.; Zeng, X.; Yang, L.; TeSlaa, T.; Li, X.; Bartman, C.; Zhang, Z.; Jang, C.; Wang, L.; Lu, W. *Cell Metab.* **2020**, *32* (4), 676−688.e4.
(25) Orth, J. D.; Thiele, I.; Palsson, B. O. *Nat. Biotechnol.* **2010**, *28* (3), 245−248.
(26) Pinu, F. R.; Villas-Boas, S. G.; Aggio, R. *Metabolites* **2017**, *7* (4), 53.
(27) Lewis, I. A. *Curr. Opin. Biotechnol.* **2024**, *85*, 103027.
(28) Khandelwal, R. A.; Olivier, B. G.; Röling, W. F. M.; Teusink, B.; Bruggeman, F. J. *PloS One* **2013**, *8* (5), No. e64567.
(29) Perez-Garcia, O.; Lear, G.; Singhal, N. *Front. Microbiol.* **2016**, *7*, 186528.
(30) Hollinshead, W.; He, L.; Tang, Y. J. *Front. Microbiol.* **2014**, *5*, 344.
(31) Carnicer, M.; Vieira, G.; Brautaset, T.; Portais, J. C.; Heux, S. *Microb. Cell Fact.* **2016**, *15* (1), 92.
(32) Varma, V. R.; Oommen, A. M.; Varma, S.; Casanova, R.; An, Y.; Andrews, R. M.; O'Brien, R.; Pletnikova, O.; Troncoso, J. C.; Toledo, J. *PLoS Med.* **2018**, *15*, No. e1002482.
(33) Sullivan, M. R.; Danai, L. V.; Lewis, C. A.; Chan, S. H.; Gui, D. Y.; Kunchok, T.; Dennstedt, E. A.; Vander Heiden, M. G.; Muir, A. *Elife* **2019**, *8*, No. e44235.
(34) Jorge, T. F.; Mata, A. T.; António, C. *Phil. Trans. R. Soc. A.* **2016**, *374* (2079), 20150370.

https://doi.org/10.1021/acs.analchem.3c04988
Anal. Chem. XXXX, XXX, XXX−XXX