Journal of Chromatography B, 877 (2009) 2952–2960

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

# Assessment of sampling strategies for gas chromatography–mass spectrometry (GC–MS) based metabolomics of cyanobacteria

Leonard Krall[1], Jan Huege[1], Gareth Catchpole, Dirk Steinhauser*,[1], Lothar Willmitzer

*Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*

ARTICLE INFO

*Article history:*
Received 6 May 2009
Accepted 5 July 2009
Available online 10 July 2009

*Keywords:*
Metabolite profiling
Cyanobacteria
Metabolite quenching
Fast filtering
GC–MS

ABSTRACT

Metabolomics is the comprehensive analysis of the small molecules that compose an organism's metabolism. The main limiting step in microbial metabolomics is the requirement for fast and efficient separation of microbes from the culture medium under conditions in which metabolism is rapidly halted. In this article we compare three different sampling strategies, quenching, filtering, and centrifugation, for arresting the metabolic activities of two morphologically diverse cyanobacteria, the unicellular Synechocystis sp. PCC 6803 and the filamentous Nostoc sp. PCC 7120 for GC–MS analysis. We demonstrate that each sampling technique produces internally consistent and reproducible data, however, cold methanol–water quenching caused leakage and substantial loss of metabolites from various compound classes, while fast filtering and centrifugation produced quite similar metabolite pool sizes, even for metabolites with predicted high turnover. This indicates that cyanobacterial metabolic pools, as measured by GC–MS, do not show high turnover under standard growing conditions. As well, using stable $^{13}C$ labeling we show the biological origin of some of the consistently observed unknown analytes. With the development of these techniques, we establish the basis for broad scale comparative metabolite profiling of cyanobacteria.



## 1. Introduction

Cyanobacteria are a versatile and ancient group of prokaryotic organisms with huge diversity in terms of morphology and molecular properties, such as the ability to assimilate atmospheric nitrogen [1]. This diversity has made the cyanobacteria ubiquitous, existing anywhere where light and water are present [2]. Cyanobacteria are also a biologically facile model for the photosynthetic cell, especially regarding the organization and regulation of photosynthesis [3,4], as well as possessing some plant signaling properties [5]. Being prokaryotic, and thus lacking compartmentalization, makes them especially suited for metabolite analysis.

In recent years, technology and method developments in metabolite profiling have supported metabolomics, the concept of comprehensive metabolite analysis, to become a cornerstone of functional genomics and systems biology [6,7]. Gas chromatography combined with mass spectrometry (GC–MS) has a proven record as a metabolite profiling method [8–11]. It routinely and robustly provides semi-quantitative information pertaining to in excess of hundreds of different chemical analytes, depending upon the biological material. Today these methods are widely applied to various organisms including microbes [12–14].

Metabolomics from a conceptual standpoint aims at the non-biased analysis of metabolites regardless if their chemical structures are known. Therefore it is imperative that the compounds detected are of proven biological origin. With the sensitivity of GC–MS analysis, numerous contaminant peaks or artifacts can be present [15]. Usually, the biological origin of peaks can be achieved through stable isotope labeling by comparing mass spectra from photoautotrophic cells cultured under atmospheric $^{12}CO_2$ and $^{13}CO_2$ conditions [16,17] or by U-$^{13}$C-glucose labeling [18].

For metabolite profiling it is imperative that a reliable snapshot of the biological system is reflected in the final analysis. To accomplish this, the cellular metabolism must be halted rapidly after removing a sample from the growth environment. Bulk samples, such as plant tissues, can simply be arrested by snapshot freezing in liquid nitrogen. However, to analyze microbes grown in liquid cultures necessitates either centrifugation or an alternative procedure for rapid removal of the microbes from their growth media prior to halting metabolism. It is believed that centrifugation of biologically active samples is undesirable as significant metabolic changes can occur during sample preparation, even at 4 °C, especially regarding sugar phosphates levels, which are considered to be among the most unstable in biological systems [19,20].

* Corresponding author. Tel.: +49 331 5678218; fax: +49 331 5678236.
*E-mail address:* steinhauser@mpimp-golm.mpg.de (D. Steinhauser).
[1] These authors contributed equally to this work.

1570-0232/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2009.07.006

Therefore, a fast and efficient separation of microbes from the culture medium under conditions in which metabolite levels remain stable is required. Traditionally, for unicellular organisms a quenching procedure has been employed [21,22]. This involves the sampling of liquid culture directly into a supercooled (−50 °C) buffer which therefore demands a substantial organic proportion to maintain fluidity. The microbes can then be separated by centrifugation at low temperature (−10 °C to −20 °C). This method, however, is unwieldy and expensive, as well as having the potential to extract metabolites from samples [14,20,23]. Other quenching compositions have recently been used showing a marked improvement in metabolite recovery [24].

Cyanobacterial metabolome studies have mainly concentrated on Synechocystis sp. PCC 6803 [17,25]; very few studies have been undertaken regarding metabolite profiling of other diverse cyanobacterial strains. The morphological diversity of cyanobacteria, ranging from simple unicellular to complex filamentous necessitates optimizations in metabolite analysis strategies, especially in terms of sampling.

Here we compare the effects of three different sampling strategies, classical quenching, fast filtering, and centrifugation, for GC–MS based metabolite profiling of two different cyanobacterial strains, namely the unicellular Synechocystis sp. PCC 6803 (afterwards called Synechocystis) and the filamentous Nostoc sp. PCC 7120 (afterwards called Nostoc).

We show that the metabolite data are internally consistent within each sampling technique, and furthermore discuss the advantages and disadvantages of the techniques for downstream applications. Finally, we show that stable isotope labeling can aide in determining the biological relevance of unknown analytes.

## 2. Material and method

### 2.1. Chemicals

All chemicals were obtained from Sigma–Aldrich (Taufkirchen, Germany) in the highest quality available.

### 2.2. Strains and growth conditions

The unicellular Synechocystis and the heterocyst-forming filamentous Nostoc strain were obtained from the Pasteur Culture Collection (Pasteur Institute, Paris, France). Axenic cultures were grown photoautotrophically in 125 mL Erlenmeyer flasks shaking at 100 rpm with 25 μmol/m$^2$ s continuous soft white illumination in a growth chamber (GFL 3033, Progen Scientific, London) at 30 °C. Three independent cultures were inoculated with 1 mL of a medium-dense pre-culture in 40 mL of BG11 and BG11o medium, respectively [1]. Synechocystis cultures were in the exponential growth phase ($OD_{730}$ ~0.8). Well shaped Nostoc cultures were of medium-density with an average of 4.29 μg/mL chlorophyll A content and were also in the exponential growth phase.

### 2.3. Cell sampling

For centrifugation, 1 mL of cells was centrifuged at RT (Synechocystis) or 4 °C (Nostoc) for 2 min at 14,000 rpm (Eppendorf Microcentrifuge 5417C, Hamburg, Germany). The supernatant was removed and the pellet frozen in liquid nitrogen. Each sample was harvested within about 30 s (w/o centrifugation).

Fast filtering was performed with a vacuum filtration manifold (Pall Corporation, NY, USA). 1 mL cells was filtered using 2.5 cm diameter, 0.22 μm pore size GVWP Durapore filter disks (Millipore Corporation, MA, USA). The filter was then transferred into a 2 mL tube and frozen in liquid nitrogen. This procedure required about 10 s per sample.

Quenching was accomplished as already described [22]. 1 mL of cells was injected into 5 mL of a −50 °C cold quenching solution, consisting of 60% methanol, in a 15 mL tube. The quenched cells were centrifuged at 4600 rpm for 6 min in a −9 °C cold centrifuge (Heraeus Multifuge 3 SR, Thermo Fisher Scientific Inc., Waltham, MA, USA) using tube adapters cooled to −50 °C. The supernatant was removed and the pellet frozen in liquid nitrogen. To determine quenching leakage, 1 mL of the pooled quenching solution (Synechocystis) or individual quenching supernatants (Nostoc) were completely dried in a rotary vacuum and processed for GC–MS measurement.

For each of the above sampling methods and for both organisms under investigation, four individual sampling replicates were taken for analysis from each of the three independent cultures (biological replicates).

### 2.4. Sample preparation

Sample extraction was performed with 600 μL of a cold 10:3:1 (v/v/v) methanol:chloroform:water solution (MCW), supplemented with 0.1 μg/mL of U-$^{13}$C-sorbitol as internal standard, in the presence of glass beads (diameter 0.75–1.0 mm, ROTH, Karlsruhe, Germany) by vortexing for 1 min. Nostoc samples were additionally treated for 5 min in a bath-type sonicator. All samples were then further extracted by strong shaking for >10 min at 4 °C. The extracts were then centrifuged at 4 °C for 2 min at 14,000 rpm (Eppendorf Microcentrifuge 5417R) and 500 μL of supernatant was concentrated to dryness in a rotary vacuum. A mixture of fatty acid methyl esters was used as internal retention index (RI) markers [22]. Sample derivatization was performed as a variation of the two-stage technique used by Roessner et al. [9]. First, 5 μL of 40 mg/mL methoxyamine hydrochloride in pyridine was added and samples shaken for 90 min at 30 °C. The samples were then derivatized by trimethylsilylation of acidic protons by addition of 45 μL MSTFA (N-methyl-N-trimethylsilyltrifluoroacetamide) and incubation for 30 min at 37 °C under shaking.

### 2.5. Data acquisition by GC-EI-TOF-MS

GC–MS analysis was performed on an Agilent 6890 gas chromatograph with deactivated standard split/splitless liners fitted with an Agilent 7683 autosampler injector (Agilent Technologies GmbH, Waldbronn, Germany). A derivatized sample volume of 1 μL was injected in splitless mode at 230 °C injector temperature onto a MDN-35 column (30 m × 0.32 mm I.D.) (Supelco, Bellefonte, PA, USA). The GC was operated at constant flow of 2 mL/min helium. The temperature program started isocratic at 85 °C for 2 min, followed by temperature ramping of 15 °C/min to a final temperature of 360 °C, which was held constant for 2 min. Data acquisition was performed on a Leco Pegasus II TOF mass spectrometer (LECO, St. Joseph, MI, USA) with an acquisition rate of 20 scans/s in a mass range ($m/z$) of 85–500 (Synechocystis)/750 (Nostoc).

### 2.6. Data processing and compound identification

All chromatograms were processed using Leco ChromaTOF software (version 3.25) with two different processing methods using baseline subtraction just above the noise, and automatic deconvolution and peak detection at signal-to-noise (S/N) levels of 10:1. The two methods used differed in smoothing, i.e. (I) auto-smoothing and (II) 5 data point average, and by peak width broadening with width-to-time (w/t) of (I) 4.0/0 s–8.0/1400 s and (II) 3.5/0 s–6.0/1400 s. The peak list was exported as text file containing retention times and raw spectra with absolute intensities.

Linear retention index calculation and peak annotation were performed in the $m/z$ range of 85–500 with a similar algorithm

previously described [15]. The used algorithm is based on a complex spectral evaluation to library entries by combination of (I) retention index deviation, (II) complete mass spectral and pattern matching, as well as (III) peak height and unique mass order (Steinhauser et al., unpublished). Library entries revealing no or non-strict hits regarding the matching criteria were post-evaluated to replace missing quantitative values similarly as described [15]. The consensus library used covered about 3100 evaluated spectra of known chemical compounds (standards) assigned to 886 non-redundant consensus analytes. Moreover, 234 species-related consensus spectra, frequently observed in Synechocystis, were added (Steinhauser et al., unpublished).

### 2.7. $^{13}C$ labeling of cyanobacterial cells

The labeling of cyanobacterial metabolites with $^{13}CO_2$ was performed in a BioBox (GMS GmbH, Berlin, Germany) [16]. 20 mL of medium-dense Synechocystis and Nostoc pre-cultures were washed twice and re-suspended in 20 mL BG11 and BG11o medium, respectively, with and w/o addition of 10 mg/mL $^{13}C$-carbonate. All cultures were grown without shaking under 2500 ppm (2.5%) $^{13}CO_2$. Cells were fast filtered, processed, and analyzed as described in Sections 2.3–2.5. Mass spectral peaks of biological origin were identified manually by searching for mass shifts between $^{12}C$ and $^{13}C$ mass spectra.

### 2.8. Data analyses

The quantification results from the two processing methods were averaged for each strain separately. To avoid bias and reducing missing or ambiguous spectral assignments, arbitrary filter options were applied. First, all analytes found in less than one-third of the entire chromatogram list (w/o leakage samples) were excluded. Secondly, analytes with less than 75% assigned quantification values were removed. The filtered datasets comprises 40 samples × 168 analytes for Synechocystis and 47 samples × 165 analytes for Nostoc (Supplementary data S-1). Both datasets, w/o the internal standard, contain 280 (4.2%) and 439 (5.7%) missing values for Synechocystis and Nostoc, respectively. For all analyses the quantification values were pre-normalized either to both the U-$^{13}C$-sorbitol internal standard and the $OD_{730}$ values for Synechocystis, or to the sucrose peak height for Nostoc. A variance-stabilized post-normalization was then performed [26]. All raw data is provided as an Excel file as Supplementary data S-1.

### 2.9. Statistical analyses and visualization

Statistical analyses were executed in R 2.6.1 [27] according to [28]. The *t*-test was performed two-sided with equal or unequal variance according to Hartley's $F_{max}$-test [29]. Mantel test was performed as Pearson's correlation ($r$) with 9999 bootstrap samples. The coefficient of variation was calculated as the ratio of the standard deviation to the mean and expressed as percentage (relative standard deviation). Principle component analyses (PCA) were performed as probabilistic PCA [30]. Class enrichment analysis was computed with Fisher's exact test. $p$-values ($p$) were adjusted ($p'$) by Bonferroni correction. Multi-dimensional outliers were detected using the robust Mahalanobis distance with the minimum covariance determinant (MCD) estimator [31]. Graphs were created using Sigma Plot 10 (Systat Software Inc., San Jose, CA, USA) or R. Heat maps were post-processed with Adobe Photoshop 7.0 (Adobe Systems Inc., Mountain View, CA, USA). Cluster trees depicted in the heat maps were generated by hierarchical cluster analyses (HCA) with the average linkage cluster algorithm (UPGMA) [32].

## 3. Results

Our initial metabolite profiling attempts of Synechocystis and Nostoc suggested a similar degree of leakage by quenching (Fig. 1, Supplementary data S-1). As this phenomenon of metabolite leakage has been previously observed for *Escherichia coli* as well as numerous other microorganisms [14,33] we therefore decided to compare different sampling methods that allow a fast and easy separation of cyanobacteria from the growth medium. To assess variation, three independent biological cultures were used and four replicate samples were harvested from each culture by centrifugation (C), quenching (Q), and fast filtering (F). All subsequent analyses were performed identically to reduce further variations.

### 3.1. $^{13}C$ labeling validates the biological origin of measured analytes

To extend our knowledge of cyanobacteria beyond known analytes, we analyzed Synechocystis and Nostoc samples with $^{13}CO_2$ incorporated into metabolism. A typical GC–MS chromatogram of unperturbed Synechocystis or Nostoc samples contains about 400–600 deconvoluted spectra. As described [15] some of these peaks may result from potential contaminants during sample handling, chemical leeching, deconvolution errors due to

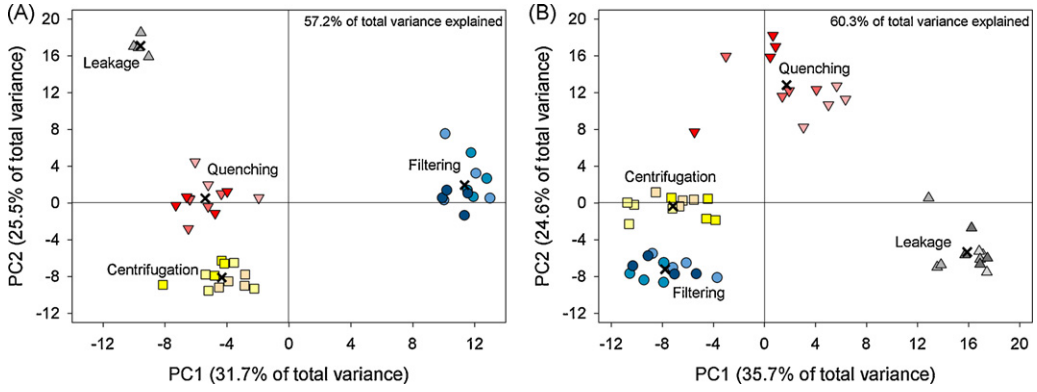


**Fig. 1.** Principal component analyses of the (A) Synechocystis and the (B) Nostoc dataset show a high cohesion within and good separation between the different sampling groups for principle components 1 and 2 based on the scores. The sampling groups are color coded: blue = filtering, yellow = centrifugation, red = quenching, and leakage = grey, with color shades according the independent biological replicates. The group centre is marked as a black cross for each sampling technique. The Synechocystis dataset used for PCA generation was pre-normalized to both the internal standard and the $OD_{730}$ values. The Nostoc dataset was pre-normalized to the sucrose peak. Both datasets were variance-stabilized post-normalized, and data were expressed as log2 transformed ratios to the dataset specific average from the filtering samples for each analyte. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

chromatogram complexity and peak proximity, column bleeding, or derivitization artifacts.

Usually, observed spectra are matched against libraries of curated compounds, such as the Golm Metabolome Database [34,35]. However, such an approach is restricted to known metabolites. A reliable criterion for assigning a compound to be of biological origin is thus a distinct shift between the $^{12}C$ and $^{13}C$ mass spectra observed at the same RI. Isotopic shifts of molecular fragments containing C atoms within the mass spectrum is thus enough to demonstrate the *in vivo* synthesis and the biological origin of an analyte. Manual annotation of chromatograms from both strains revealed isotopic mass shifts for 56 out of 80 analytes with known chemical structure and for 35 out of 109 unknown analytes examined (Supplementary data S-1).

### 3.2. Alternative pre-normalization strategies support similar downstream findings

In GC–MS metabolite profiling experiments quantification values are usually expressed relative to both the internal standard and the amount of sample used. As the fresh weight is difficult to measure for microbes growing in liquid cultures, we used the optical density ($OD_{730}$) for Synechocystis. Due to the small amount of sample that is sufficient for GC–MS measurements, using the dry weight is not an option. As well, for filamentous and some unicellular strains, the optical density cannot be used as reliable approximation due to colonial or aggregate growth. To overcome this we pre-normalized the Nostoc dataset exclusively to the observed sucrose peak height as stable internal pool size for this particular growth condition.

To test if both pre-normalization strategies support similar findings we additionally pre-normalized the Synechocystis data to the sucrose peak height. Comparison of both Synechocystis matrices by Mantel test on the Euclidean distances among the samples and the detected analytes revealed a highly significant correlation (both $r < 0.997$, $p < 1e\text{-}04$). Therefore, both pre-normalization strategies support similar findings and thus, similar conclusions can be drawn. Consequently, for all further analyses on the Synechocystis dataset we refer to the quantification matrix pre-normalized to the internal standard and the $OD_{730}$. Results for the alternative strategy on the Synechocystis dataset can be found as supplementary data (Fig. S-1).

### 3.3. Filtering, quenching, and centrifugation are distinct

To get an overview regarding the similarity of the sampling techniques used, we first performed PCA analysis on datasets with the log 2-transformed ratios expressed to the group mean of the respective fast filtering samples. Visualization of the principle components (PC) 1 and 2, which represent 57.2% (Synechocystis) and 60.3% (Nostoc) of the total variance, demonstrate a clear separation of the sampling techniques and leakage group from each other (Fig. 1A and B). This confirms that the individual sampling techniques produce distinct results, irrespective of the cyanobacterial strain. The PCA as well as HCA graphs (not shown), revealed a relatively tight grouping of the sampling and independent biological replicates for both strains (Fig. 1A and B).

To confirm the cohesion within and the separation among the techniques we calculated the silhouette coefficients ($s$) on the Euclidean distances based on the scores of PCs 1–3, which explain 68.7% and 73.3% of the total variance for the Synechocystis and Nostoc dataset, respectively. For both, relatively high silhouette coefficients were obtained with an overall average of $s > 0.75$ and an average within cluster distance ($d_E$) of about 0.14 (Table S-1). With exception for the quenching samples, all samples revealed a relatively high cohesion within their designated group demonstrating the grouping validity. Group assignments were further tested by nonparametric analysis of variance (ANOVA) performed as Mantel test between the Euclidean distances among the samples and their assigned groups. These comparisons showed high and significant ($p < 1e\text{-}04$) Pearson matrix correlations of $r = 0.843$ and $r = 0.884$ for the Synechocystis and Nostoc dataset, respectively. In contrast, a separation of the sampling and independent biological replicates within each sampling group could not be confirmed by any of the abovementioned statistical analyses.

### 3.4. Filtering, quenching, and centrifugation yield highly reproducible data

Reproducibility is one of the key factors for any analytical method. To address this in more detail we performed linear regression analyses for all pairwise sample comparisons. For this, log 2-transformed normalized datasets were used and peak intensities from each sample were linearly fitted to those of each other sample. The resulting adjusted pairwise coefficients of determination ($R^2$) were clustered and visualized (Fig. 2A and B). Independent of the strain, the graphs demonstrate that samples of each sampling technique are in good approximation to each other, resulting in large $R^2$ values with an overall mean of $R^2 = 0.876 \pm 0.046$ with $m = 0.95 \pm 0.043$ for Synechocystis, and $R^2 = 0.856 \pm 0.073$ with $m = 0.856 \pm 0.073$ (both as mean ± SD) for Nostoc. In contrast, the observed $R^2$ values are much lower between the individual sampling techniques (Table S-2) resulting in distinct hierarchical clusters (Fig. 2A and B), which is in agreement with the PCA results (Fig. 1A and B). In further agreement, the sampling and independent biological replicates of each group are not clearly distinguishable from each other. However, the observed $R^2$ values among the sampling replicates are on average slightly higher as compared to the average between the biological replicates (Table S-2), independent of the strain and technique considered. The observed large $R^2$ values, which explain on average more than 85% of the observed variance, and the slope near 1 validates the normalization and demonstrates the high reproducibility of the analytical procedure independent of the sampling technique.

To further investigate the precision and reproducibility of the sampling techniques and metabolite profile analyses in relation to the individual analytes, we calculated the relative standard deviation (RSD) between the (I) $3 \times 4$ sampling and (II) the averaged sampling replicates from the three independent biological replicate sets (Table S-3). Between the sampling techniques and strains investigated, an overall median RSD of about 26% for the sampling and 22% for the independent biological replicates was determined. The individual results are slightly strain dependent. For Synechocystis no significant differences of the median RSD could be observed among the independent biological replicates, but centrifugation performed significantly better regarding sampling variation (Table S-3). Fast filtering performed significantly better for Nostoc in comparison to both quenching and centrifugation, between both the sampling and independent biological replicates (Table S-3).

### 3.5. Quantitative analyte differences between the three techniques

To address relative changes between the sampling techniques, we performed two sample $t$-tests for each reported analyte by comparing the relative quantification values of filtering to both quenching and centrifugation. In addition to the $p$-values for statistical significance estimation, we determined the ratios per analyte to the respective fast filtering average. A log 2-fold change of 1 was assumed to be of potential relevance. Visualization of the p-values versus the log 2 ratios ($R$) revealed, depending on the significance

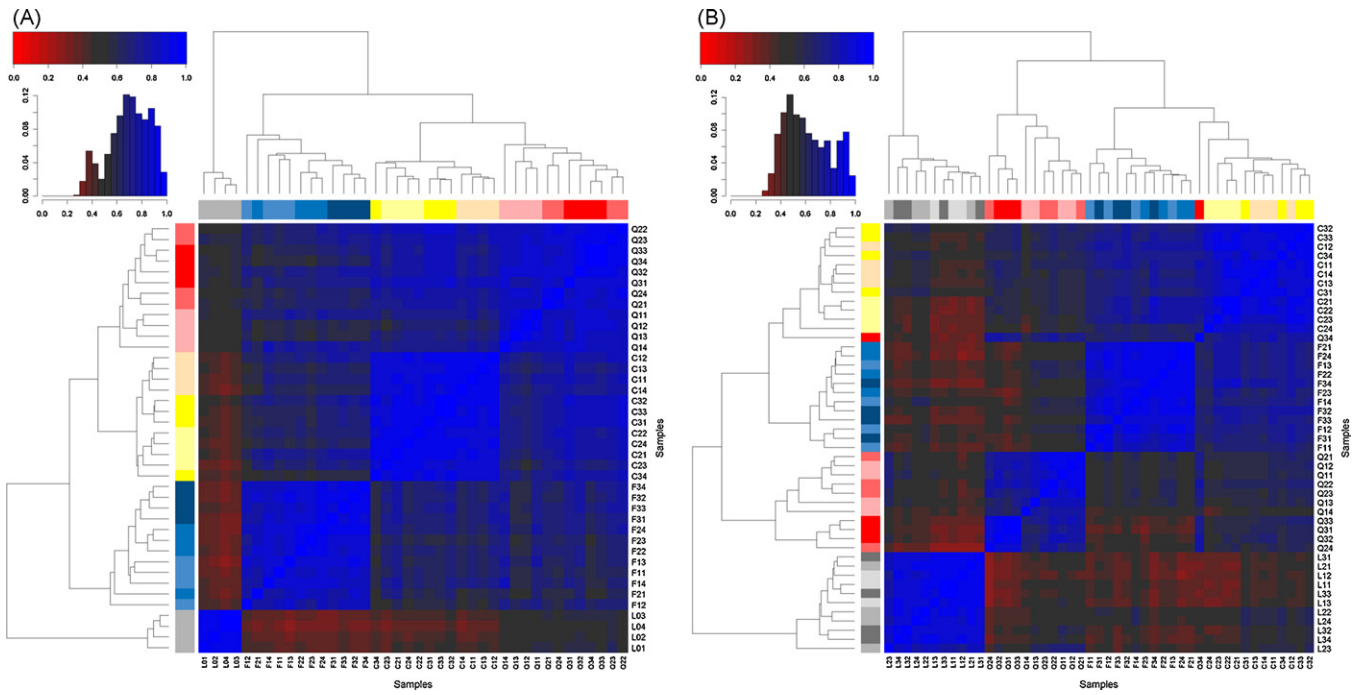


**Fig. 2.** Heat map visualization and cluster tree representation of the coefficients of determination ($R^2$) calculated by linear regression among the sampling and independent biological replicates and each sampling group for the (A) Synechocystis and (B) Nostoc dataset. The left top-side color bar represents the color code ranging from 0 to 1 for the obtained $R^2$ values drawn in each heat map cell. The distribution of the observed $R^2$ values is depicted as a histogram. The sampling and replicate group assignments are drawn as top and left sided color coded columns: blue = filtering, yellow = centrifugation, red = quenching, and leakage = grey. The independent biological replicates of each sampling group are represented by different shades of the respective sampling group color. Basically, sampling replicates per sample group reveal a slightly higher $R^2$ (darker blue colored) compared to the independent biological replicates. The within group $R^2$ are quite specific whereas a less strong agreement can be found between the sampling groups, especially in the quenching samples from Nostoc. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

threshold chosen, significant changes between filtering and both centrifugation and quenching (Fig. 3A and B).

For the Synechocystis dataset, 70 (41.7%) and 41 (24.4%) analytes were significantly changed with $p < 0.05$ between filtering and both quenching and centrifugation, respectively (Fig. 3A). However, if the applied $p$-value is adjusted for multiple comparisons ($p_{0.05}' < 2.98e\text{-}04$), only two significant and relevant (abs($R$) > 1) changes for each of both comparisons can be detected; the amino acids alanine ($R = -2.37$) as well as beta-alanine ($R = -1.42$) for the quenching comparison, and two unknown analytes (DSLK-GCTOF1-MCWP1-522572-7306ka38_1-SYN and DSLK-GCTOF1-MCWP1-319549-7306ka38_1-SYN with $R = -2.41$ and 3.55, respectively) for the comparison to centrifugation. As no $^{13}C$ labeled spectra could be observed for both unknowns their biological origin is questionable (Supplementary data S-1).

Similarly, 71 (43.3%, F vs. Q) and 37 (22.6%, F vs. C) analytes were significantly changed with $p < 0.05$ within the Nostoc dataset (Fig. 3B). After correction for multiple comparisons ($p_{0.05}' < 3.05e\text{-}04$) five significant changes compared to quenching can be observed; one unknown analyte (DSLK-GCTOF1-MCWP1-601583-

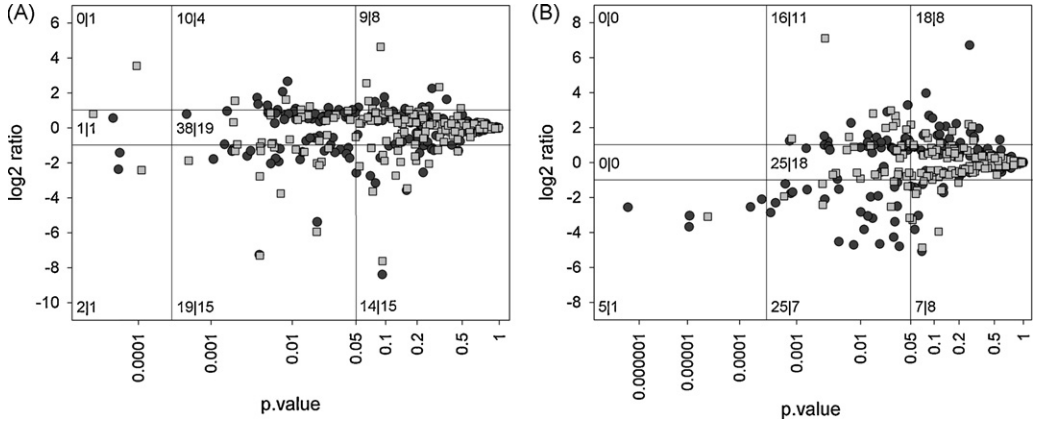


**Fig. 3.** Probability log 2 ratio plot demonstrating the relation of significance to fold change of metabolites by quenching and centrifugation compared to filtering for the (A) Synechocystis and (B) Nostoc dataset. Each graph was subdivided into nine grids. The $p$-value, calculated by t-test on the independent biological replicates, was binned accordingly: not significant, significant with $p < 0.05$ and significant with adjusted $p < 0.05$ ($p_{0.05}'$). The log 2 ratios were binned according to: above 1, between 1 and −1, and below 1. Absolute log 2-fold changes above 1 (>2-fold increase or decrease) were treated as relevant changes. The number of counts per grid is given for the comparison of filtering to both quenching (Q, black circles), and centrifugation (C, grey squares) as Q|C counts. For potentially significant changes ($p < 0.05$, not adjusted) a larger number of decreased relative peak heights in the quenched samples, compared to filtering, can be found.

7306ka38_1-SYN, $R = -3.05$), urea ($R = -3.68$), the amino acids lysine ($R = -2.57$), serine ($R = -2.54$), and the $C_4H_4O_4$ carboxylic acid maleate and/or fumarate ($R = -2.10$; both undistinguishable by RI and spectral identity). Only one, the same unknown analyte as for the quenching comparison ($R = -3.09$) is significantly changed compared to centrifugation. Again, no $^{13}C$ mass shift could be identified for this analyte.

Independent of the strain used, only few significant (adjusted $p$) analyte changes can be detected between filtering and centrifugation, again demonstrating that both techniques yielded similar results. For Synechocystis, however, a significantly ($p < 0.01$, Fisher-exact test) larger portion of analytes had significantly reduced relative peak heights, probably due to metabolic activities during centrifugation at RT (Fig. 3A). In contrast, the quenching technique clearly showed a significantly larger proportion (Fisher-exact test $p \leq 0.002$) of analytes with reduced peak height in both strains.

### 3.6. Primarily amino and organic acids drive the sampling group separation and are depleted by quenching

As only few changes were detectable by $t$-tests with $p$-value correction the question arose as to which analytes were responsible for the sampling group separation. To determine this, we re-performed the PCA without the leakage samples. As shown in Fig. 1A and B, all samples of one sampling technique grouped together and separately from the other techniques used (Fig. S2). To detect analytes responsible for the separation we performed a multi-dimensional outlier test, pairwise or overall, on the first three PCs, which covered 65.9% and 68.5% of the total variance, for the Synechocystis and Nostoc dataset, respectively (Fig. 4A and B).

A total of 23 known and 15 unknown analytes were identified in the Synechocystis dataset as being responsible for the separation (Table S-4A). Out of the 23 known analytes, 16 out of a total of 21 amino acids and 3 out of a total of 12 carboxylic acids were found to drive the separation (Table S-4A). To test for enrichment we performed Fisher's exact test, which revealed a highly significant enrichment of $p < 2.3e\text{-}08$ for the amino acid compound class. This enrichment is still significant ($p < 8.0e\text{-}07$) if we perform the test on the fused compound classes of carboxylic and amino acids. Comparing filtering versus quenching, 16 of the total 23 known analytes showed a greater than 2-fold reduction ($p < 4.5e\text{-}07$) (Table S-4A, see Fig. 5A). Only two known analytes, namely ketoglutarate (major) and spermidine showed a 2-fold or greater increase.

For the comparison filtering versus centrifugation, however, most of the observed log 2 ratios are similar; only two known analytes showed greater than a 2-fold decrease, malate and a probable breakdown fragment of estrone-3-sulfate, and four analytes showed a 2-fold or greater increase, asparagine, glutamate, ketoglutarate (major) and spermidine (Table S-4A, see Fig. 5A).

Similar results were found for the Nostoc dataset with a total of 33 known and 15 unknown analytes driving the separation (Table S-4B). Significant enrichments were found for the amino acid ($p < 2.6e\text{-}05$, 14 of 19), carboxylic acid ($p < 6.3e\text{-}05$, 18 of 29), and saccharide class ($p < 0.009$, 6 of 8). A total of 23 out of 33 known analytes showed reduced levels greater than 2-fold in quenched compared to filtered samples, which represent a statistically significant enrichment with $p < 4.4e\text{-}09$ (Table S-4B, see Fig. 5B). Such enrichment could not be identified for the comparison of the filtered and centrifuged samples and is consistent with the findings of the Synechocystis dataset. For both comparisons four known fatty acid analytes were identified with greater than a 2-fold increase compared to the filtering. None, however, could be identified as $^{13}C$ labeled spectra (Table S-4B).

Mainly analytes of the amino and carboxylic acid classes, and additionally trisaccharides in Nostoc, are responsible for the separation of the sampling techniques (Fig. 4A and B; Tables S-4A and S-4B). The reduction is in accordance with the enrichment of these analytes in the quenching loss, suggesting a significant leakage by the quenching procedure in both cyanobacterial strains (cf. Fig. 5A and B). Moreover, the general overall decrease in relative analyte peak heights implies a general leaking of the quenching procedure beyond particular compound classes (Fig. 5A and B). Interestingly, with only one exception in the Nostoc dataset, the phosphorylated analytes were not responsible for the separation. These compounds seem to remain stable over a few minutes, as no significant changes can be observed between the sampling techniques (Fig. 5A and B). However for the Nostoc dataset the phosporylated analyte glucose 6-P (and/or galactose 6-P, both undistinguishable by RI and spectral identity) is strongly decreased ($R = -4.8$) between the filtered and quenched samples (Fig. 5B). Since this analyte is enriched in the leakage samples (cf. Fig. 5B), it demonstrates that phosphorylated analytes are also lost through quenching.

## 4. Discussion and conclusion

### 4.1. All sampling techniques show a high replicate consistency

Despite the fact that a clear sample separation can be observed (Fig. 1A and B), the metabolite profiles of cyanobacterial cells grown under standard conditions and isolated from the exponential growth phase are highly comparable across different experiments and replicates irrespective of the sampling technique used (Fig. 2A

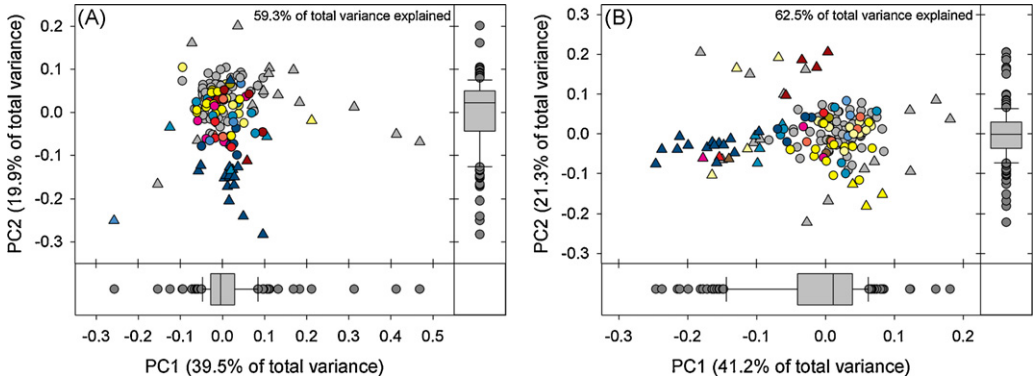


**Fig. 4.** Scatter and boxplot graphs of the loadings computed by PCA analyses of the (A) Synechocystis and (B) Nostoc dataset without the respective leakage samples. The graphs show the loadings for the principle components 1 and 2. Detected outliers, analytes that are mainly driving the sampling group separation are represented by triangles. The assigned analytical class for each analyte is color encoded (cf. Fig. 5A and B). Class color assignments for analytes driving the separation are: unknown compound (grey), amino acid (dark blue), carboxylic acid (blue), mono- (dark red) and oligosaccharides (brown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

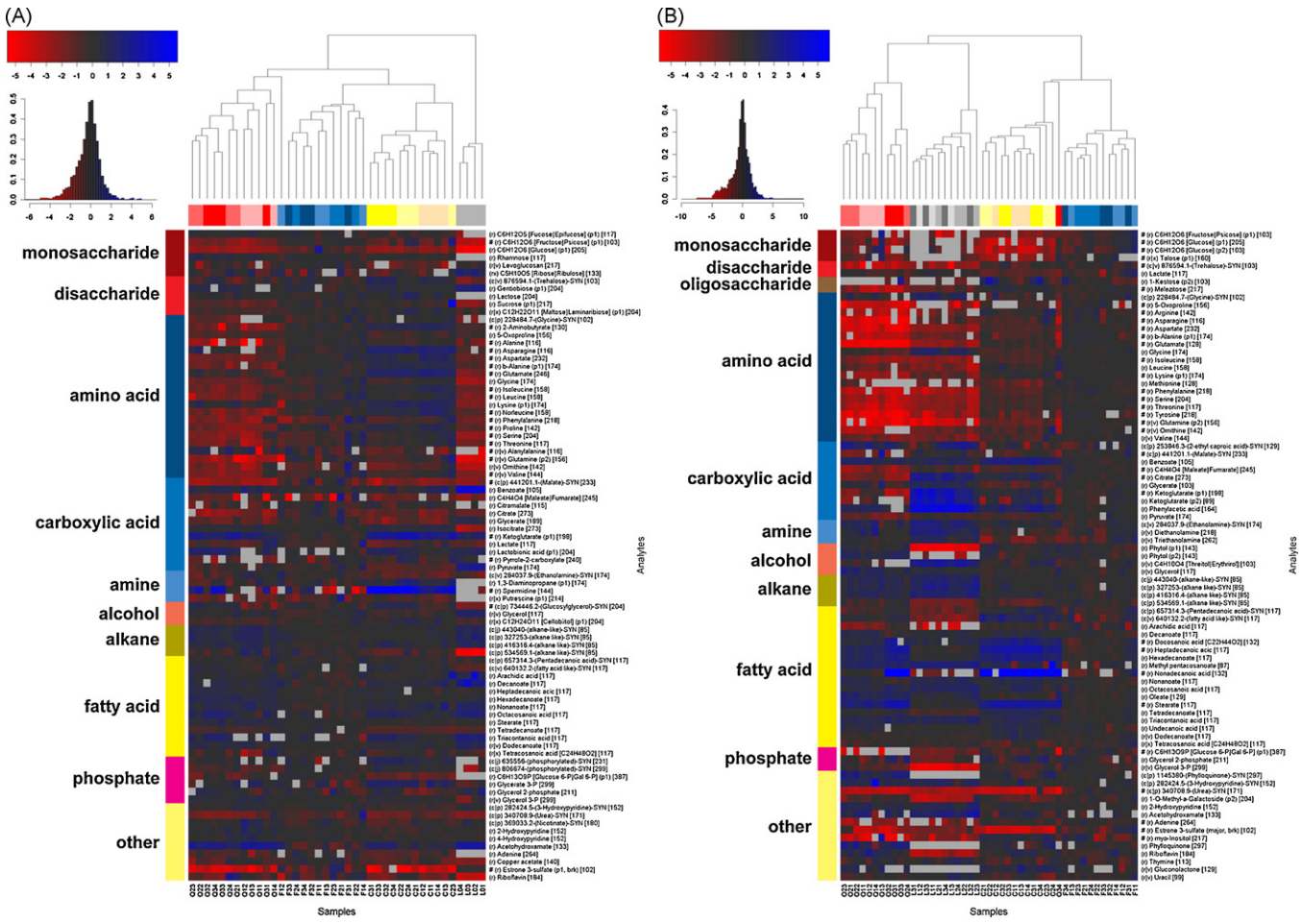


**Fig. 5.** Heat map visualization and cluster tree representation of the log 2-fold changes, compared to the average of the filtering group, for the (A) Synechocystis and (B) Nostoc dataset. The left top-side color bar represents the color code ranging from −5 to 5 of the observed log 2-fold changes. Values above or below this range are either bright red or dark blue color coded. The distribution of the log 2-fold changes are depicted as a histogram. The sampling and replicate group assignments are the color coded columns on top: blue = filtering, yellow = centrifugation, red = quenching, grey = leakage. The independent biological replicates of each sampling group are represented by different shades of the respective sampling group color. The left sided color column represents a color code for each of the used analyte classes with the given analyte class name. Outlier analytes driving the group separation are preceded with a hash mark (#). All analyte names are abbreviated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

and B). This consistency is confirmed by the cohesion within each group (Table S-1) and supported by high coefficients of determination (Fig. 2A and B; Table S-2). Based on this finding it is clear that reproducible data can be obtained from each technique. While previous studies have clearly shown loss of intracellular metabolites in a number of different organisms through quenching [23,33,36,37], here we show, by using three independent cultures of cyanobacteria grown under standard conditions, that the loss of metabolites by quenching is consistent between these three cultures. This is true for both Synechocystis and Nostoc. This observation has so far not been shown for other microbial quenching studies. Thus, a preference for any of the employed sampling techniques in terms of reproducibility and precision cannot be supported for the cyanobacterial strains used.

### *4.2. Advantages and disadvantages of the employed sampling techniques*

It is clear, however, that the methanol–water quenching does have severe influences, due to metabolite leakage of amino and carboxylic acids in both cyanobacterial strains (Fig. 4A and B). Moreover, a number of slight changes are observed among various analyte classes demonstrating that quenching leads to a general loss and thus to decreased peak heights (Fig. 5A and B). While the amino acids are robust metabolites readily seen in GC–MS measurements their significant loss through quenching may drive the pool sizes to below detection levels. Furthermore, the quenching loss is not only limited to small molecules but can also include large molecules, such as trisaccharides, as illustrated for Nostoc (Fig. 5B).

Altered metabolite levels can also pose a large problem for downstream modeling of metabolism, which is an integral component of systems biology driven approaches [38,39]. As some analyte classes are more likely to be lost by quenching, the use of quenching data for modeling can lead to wrong model assumptions regarding enzyme kinetics and perceived flux rates. This problem might be greatly magnified if data from different quenching (or extraction) [40] techniques are combined for holistic overviews of coordinated metabolic changes, as it has been done with transcript datasets [41,42].

Interestingly, while PCA can clearly separate the filtered from the centrifuged samples (Fig. 1A and B), closer analysis reveals that this separation is driven by numerous small changes throughout the dataset. We have shown that both fast filtering and centrifugation produce highly similar results (Figs. 2A–3B). This indicates that the measured pool sizes are relatively stable over the course of the centrifugation, a time of a few minutes (Fig. 3A and B). It is believed that the fast turnover rates of enzymes, such as phosphatases which can rapidly cleave phosphate groups of sugar phosphates, would cause

the loss of phosphorylated analytes on the order of seconds [19]. However, we only found one significant change of a phosphorylated analyte (C6H13O9P [Glucose 6-P|Galactose 6-P] (major)) in Nostoc. All other detected phosphorylated analytes are either unchanged or only slightly changed (Fig. 5A and B).

It seems likely that for normally growing cyanobacterial cells small perturbations of the growth conditions do not influence metabolism, approachable with GC–MS, over a short time frame based on the similarity of the biological replicates (Fig. 3A and B). This might be as well due to the growth conditions employed. For these experiments cyanobacterial strains were grown under conditions resulting in a doubling time of around 24 h. This inherently slow growth may be less sensitive to short term environmental fluctuations on metabolite level as has been shown with rapidly growing *E. coli* or yeast cells [43].

Overall, the advantages of fast filtering and centrifugation are multifold over cold methanol–water quenching. By using either fast filtering or centrifugation, the media is completely removed from the sample, and can be separately collected for further analysis. A wash step may also be easily added. A large number of biological replicates can be harvested in parallel using a filtering manifold for the filtered samples. One main advantage of filtering over the other two techniques is that morphologically complex strains can be easily harvested. For some filamentous cyanobacterial strains, centrifugation, even for long periods, is not sufficient to produce stable cell pellets. In such situations the pellet, or a substantial fraction, is usually lost during supernatant removal, or insufficient removal of the remaining supernatant may lead to extracellular contaminating metabolites. As well, fast filtering is extremely quick, with individual samples taking about 10 s to filter and freeze. However, one drawback of the fast filtering method is that it is not easily applicable to all growth conditions. Strains growing in an anoxic state cannot be easily or inexpensively filtered.

### *4.3. Stable isotope labeling can facilitate the identification of biological peaks*

Basically, one of the major limitations of GC–MS, or metabolomics in general, is that researchers are limited to metabolites for which a known standard has already been processed. Novel peaks without a clear matching to standards are usually ignored. Usually these unknown analytes remain so until matched with arbitrary confidence to a reference standard. Therefore efficient strategies are needed to uncover biologically relevant unknowns from contaminants. Stable isotope labeling with $^{13}C$, but as well as with other stable isotopes, such as $^{15}N$ for nitrogen-fixing organisms, can facilitate the discrimination between biological and potentially contaminating peaks [44,45].

In this study we used fully photoautotrophic conditions to achieve isotopic labeling, as has been performed for plants [16]. With the application of this technique we were only able to confirm by manual curation of isotopic mass shifts the biological origin of 70% of the knowns and 32% of the unknown analytes examined. Despite the fact that a large number of known analytes revealed no isotopic mass shifts, we are confident that they are of biological origin, such as some of the organic acids (Supplementary data S-1). Reasons for the lack of detectable labeling for some metabolites is most likely due to stable metabolite pools that were only partially labeled by our relatively short incubation time, or low peak sizes in the chromatograms produced through insufficient material from the labeled cultures. Future studies will address these issues in detail, as well as focusing on the $^{13}C$ flux through metabolism. As well, the difficulty to manually identify specific mass shifts in a complex spectrum is sometimes great. Regardless, emerging methodologies will hopefully be able to put a name and a chemical formula to these labeled or unlabeled analytes in the near future [45]. As well, the usage of RI standards [46] and the increasing efforts in sharing curated mass spectral libraries within the metabolic community [34,35] will aid to solve these challenges.

## Supplementary material

Supplementary data S-1 (Excel sheet) contains the raw dataset for both strains under investigation including additional analyte information. The supporting Fig. S-1 shows the result of the PCA analyses on the two differentially pre-normalized Synechocystis datasets. The supporting Fig. S-2 shows the result of the PCA analyses without the leakage samples. Supporting Table S-1 contains the results of the cohesion analysis. Supporting Table S-2 describes the averaged regression coefficients. Supporting Table S-3 contains the relative standard deviations (RSD). The supporting Table S-4 contains the result of analytes driving the sampling group separation of the (a) Synechocystis and (b) the Nostoc dataset and the *t*-test analyses for those analytes.

## Acknowledgements

The authors wish to thanks Aenne Eckardt, Gudrun Wolter, and Antje Bolze for excellent technical assistance regarding analytical measurements and Ina Sabrina Schultz for strain maintenance. We are grateful to Dr. Rosmarie Rippka and Prof. Dr. Nicole Tandeau de Marsac (Institut Pasteur, Paris, France) for access to the cyanobacterial strains. We would like to thank Dr. Patrick Giavalisco for critical reading of the manuscript. This work was supported by the Max Planck Society.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2009.07.006.

## References

[1] R. Rippka, J. Deruelles, J.B. Waterbury, M. Herdman, R.Y. Stanier, J. Gen. Microbiol. 111 (1979) 1.
[2] B.A. Whitton, M. Potts, in: B.A. Whitton, M. Potts (Eds.), Ecology of Cyanobacteria: Their Diversity in Time and Space, vol. 1, The Kluwer Academic Publishers, Dordrecht, 2000.
[3] N. Nelson, C.F. Yocum, Annu. Rev. Plant Biol. 57 (2006) 521.
[4] J.L. Roose, K.M. Wegener, H.B. Pakrasi, Photosynth. Res. 92 (2007) 369.
[5] K.C. Yeh, S.H. Wu, J.T. Murphy, J.C. Lagarias, Science 277 (1997) 1505.
[6] A.R. Fernie, R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Nat. Rev. Mol. Cell Biol. 5 (2004) 763.
[7] D. Steinhauser, J. Kopka, Exs 97 (2007) 171.
[8] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Nat. Biotechnol. 18 (2000) 1157.
[9] U. Roessner, A. Luedemann, D. Brust, O. Fiehn, T. Linke, L. Willmitzer, A. Fernie, Plant Cell 13 (2001) 11.
[10] R.J. Bino, C.H. Ric de Vos, M. Lieberman, R.D. Hall, A. Bovy, H.H. Jonker, Y. Tikunov, A. Lommen, S. Moco, I. Levin, New Phytol. 166 (2005) 427.
[11] G.S. Catchpole, M. Beckmann, D.P. Enot, M. Mondhe, B. Zywicki, J. Taylor, N. Hardy, A. Smith, R.D. King, D.B. Kell, O. Fiehn, J. Draper, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 14458.
[12] S. Strelkov, M. von Elstermann, D. Schomburg, Biol. Chem. 385 (2004) 853.
[13] H.L. Wang, B.L. Postier, R.L. Burnap, J. Biol. Chem. 279 (2004) 5739.
[14] C.J. Bolten, P. Kiefer, F. Letisse, J.C. Portais, C. Wittmann, Anal. Chem. 79 (2007) 3843.
[15] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, A.R. Fernie, Nat. Protoc. 1 (2006) 387.
[16] J. Huege, R. Sulpice, Y. Gibon, J. Lisec, K. Koehl, J. Kopka, Phytochemistry 68 (2007) 2258.
[17] C. Yang, Q. Hua, K. Shimizu, Metab. Eng. 4 (2002) 202.
[18] J. Zhao, K. Shimizu, J. Biotechnol. 101 (2003) 101.
[19] W. de Koning, K. van Dam, Anal. Biochem. 204 (1992) 118.
[20] S.G. Villas-Bôas, in: S.G. Villas-Bôas, U. Roessner, M.A.E. Hansen, J. Smedsgaard, J. Nielsen (Eds.), Metabolome Analysis: An Introduction, Wiley-Interscience, Hoboken, 2007, p. 39.
[21] A. Buchholz, R. Takors, C. Wandrey, Anal. Biochem. 295 (2001) 129.
[22] Y. Lee do, O. Fiehn, Plant Methods 4 (2008) 7.
[23] C. Wittmann, J.O. Kromer, P. Kiefer, T. Binz, E. Heinzle, Anal. Biochem. 327 (2004) 135.

[24] S.G. Villas-Boas, P. Bruheim, Anal. Biochem. 370 (2007) 87.
[25] M. Eisenhut, J. Huege, D. Schwarz, H. Bauwe, J. Kopka, M. Hagemann, Plant Physiol. 148 (2008) 2109.
[26] W. Huber, A. von Heydebreck, H. Sultmann, A. Poustka, M. Vingron, Bioinformatics 18 (Suppl. 1) (2002) S96.
[27] R Development Core Team, Vienna, 2007.
[28] R.R. Sokal, F.J. Rohlf, Biometry: The Principles and Practice of Statistics in Biological Research, W.H. Freeman and Company, New York, 1995.
[29] H.O. Hartley, Biometrika 37 (1950) 308.
[30] W. Stacklies, H. Redestig, M. Scholz, D. Walther, J. Selbig, Bioinformatics 23 (2007) 1164.
[31] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley-VCH, 2003.
[32] B. Mirkin, in: B. Mirkin (Ed.), Nonconvex Optimisation and its Application, Mathematical Classification and Clustering, Kluwer Academic Publishers, 1996, p. 109.
[33] S.G. Villas-Boas, J. Hojer-Pedersen, M. Akesson, J. Smedsgaard, J. Nielsen, Yeast 22 (2005) 1155.
[34] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A.R. Fernie, D. Steinhauser, Bioinformatics 21 (2005) 1635.
[35] N. Schauer, D. Steinhauser, S. Strelkov, D. Schomburg, G. Allison, T. Moritz, K. Lundgren, U. Roessner-Tunali, M.G. Forbes, L. Willmitzer, A.R. Fernie, J. Kopka, FEBS Lett. 579 (2005) 1332.
[36] C.L. Winder, W.B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G.M. Stephens, R. Goodacre, Anal. Chem. 80 (2008) 2939.
[37] H. Taymaz-Nikerel, M. de Mey, C. Ras, A. ten Pierick, R.M. Seifar, J.C. van Dam, J.J. Heijnen, W.M. van Gulik, Anal. Biochem. 386 (2009) 9.
[38] Z. Nikoloski, S. Grimbs, P. May, J. Selbig, J. Theor. Biol. 254 (2008) 807.
[39] T. Handorf, N. Christian, O. Ebenhoh, D. Kahn, J. Theor. Biol. 252 (2008) 530.
[40] R.P. Maharjan, T. Ferenci, Anal. Biochem. 313 (2003) 145.
[41] D. Steinhauser, B. Usadel, A. Luedemann, O. Thimm, J. Kopka, Bioinformatics 20 (2004) 3647.
[42] M. Schmid, T.S. Davison, S.R. Henz, U.J. Pape, M. Demar, M. Vingron, B. Scholkopf, D. Weigel, J.U. Lohmann, Nat. Genet. 37 (2005) 501.
[43] L. Lopez-Maury, S. Marguerat, J. Bahler, Nat. Rev. Genet. 9 (2008) 583.
[44] C. Birkemeyer, A. Luedemann, C. Wagner, A. Erban, J. Kopka, Trends Biotechnol. 23 (2005) 28.
[45] P. Giavalisco, J. Hummel, J. Lisec, A.C. Inostroza, G. Catchpole, L. Willmitzer, Anal. Chem. 80 (2008) 9417.
[46] N. Strehmel, J. Hummel, A. Erban, K. Strassburg, J. Kopka, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 182.