MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis

ARTICLE *in* ANALYTICAL CHEMISTRY · MAY 2009

Impact Factor: 5.64 · DOI: 10.1021/ac802689c · Source: PubMed

6 AUTHORS, INCLUDING:

Karsten Hiller
University of Luxembourg
45 PUBLICATIONS **1,822** CITATIONS
SEE PROFILE

Christian Jäger
University of Luxembourg
6 PUBLICATIONS **143** CITATIONS
SEE PROFILE

Jana Spura
Technische Universität Braunschweig
5 PUBLICATIONS **263** CITATIONS
SEE PROFILE

Dietmar Schomburg
Technische Universität Braunschweig
589 PUBLICATIONS **11,521** CITATIONS
SEE PROFILE

Anal. Chem. **2009,** *81,* 3429–3439

MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis

Karsten Hiller, Jasper Hangebrauk, Christian Jäger, Jana Spura, Kerstin Schreiber, and Dietmar Schomburg*

Department of Bioinformatics and Biochemistry, Technische Universität Braunschweig, Langer Kamp 19b, D-38106 Braunschweig, Germany

We have developed a new software, MetaboliteDetector, for the efficient and automatic analysis of GC/MS-based metabolomics data. Starting with raw MS data, the program detects and subsequently identifies potential metabolites. Moreover, a comparative analysis of a large number of chromatograms can be performed in either a targeted or nontargeted approach. MetaboliteDetector automatically determines appropriate quantification ions and performs an integration of single ion peaks. The analysis results can directly be visualized with a principal component analysis. Since the manual input is limited to absolutely necessary parameters, the program is also usable for the analysis of high-throughput data. However, the intuitive graphical user interface of MetaboliteDetector additionally allows for a detailed examination of a single GC/MS chromatogram including single ion chromatograms, recorded mass spectra, and identified metabolite spectra in combination with the corresponding reference spectra obtained from a reference library. MetaboliteDetector offers the ability to operate with highly resolved profile mass data. Finally, all analysis results can be exported to tab delimited tables. The features of MetaboliteDetector are demonstrated by the analysis of two experimental metabolomics data sets. MetaboliteDetector is freely available under the GNU public license (GPL) at http://metabolitedetector.tu-bs.de.

Technical advances in the field of systems biology in combination with the collection of huge amounts of high-throughput-based data including the completion of many genome sequences have opened up new views on living systems. In addition to transcriptomics and proteomics, metabolomics allows the simultaneous measurement of the concentration of a large number of metabolites. During the last years, metabolomics has developed to a routine process in matters of sample throughput and robustness and evolved to an essential part of quantitative biology. Furthermore, metabolomics is increasingly applied for biomarker discovery and phenotype characterization.[1,2] Currently, several experimental techniques are routinely applied to study the set of metabolites, including nuclear magnetic resonance spectroscopy (NMR), gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE) coupled to mass spectrometry (MS).[3,4] In contrast to NMR, LC/MS, CE/MS, and GC/MS are able to separate different chemical compounds into individual peaks. Improvements and adoption of GC/MS based techniques allow the analysis of the endometabolome of many organisms like plants,[5] prokaryotes,[6] and fungi,[7] as well as the metabolomics analysis of complex fluids like human blood plasma. With dependence on the instrument and the analysis method, several hundred substances can be detected.[8] Modern GC/MS based methods are able to perform accurate measurements in less than 20 min thereby covering the main parts of the metabolome.[9] When applied in a high-throughput manner, these techniques produce a huge amount of high dimensional data which require an efficient and accurate way of mathematical analysis.

Typically, the different types of metabolomics analysis can be separated into two major groups: targeted and nontargeted analyses. In targeted analyses, the metabolomics data are scanned for specific compounds normally collected in a reference library. In contrast, during a nontargeted approach, the compounds are not identified and the spectroscopic features of all potential compounds are considered for further analyses.[10] One major drawback of the first method is the relatively limited size of most current reference libraries, thus preventing the use of the whole amount of information present in the spectral data. Currently, several free programs for both types of data analysis exist. Automated Mass Spectral Deconvolution and Identification Software (AMDIS) is a free Windows based software featuring a powerful algorithm for the deconvolution and subsequent iden-

* To whom correspondence should be addressed. Phone: +49 531 3918300. Fax: +49 531 3918302. E-mail: d.schomburg@tu-bs.de.

(1) Sabatine, M. S.; Liu, E.; Morrow, D. A.; Heller, E.; McCarroll, R.; Wiegand, R.; Berriz, G. F.; Roth, F. P.; Gerszten, R. E. *Circulation* **2005**, *112*, 3868–3875.
(2) Pauling, L.; Robinson, A. B.; Teranishi, R.; Cary, P. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 2374–2376.
(3) Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606–625.
(4) Weckwerth, W.; Morgenthal, K. *Drug Discovery Today* **2005**, *10*, 1551–1558.
(5) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157–1161.
(6) Strelkov, S.; von Elstermann, M.; Schomburg, D. *Biol. Chem.* **2004**, *385*, 853–861.
(7) Villas-Bôas, S. G.; Højer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Yeast* **2005**, *22*, 1155–1169.
(8) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.
(9) Börner, J.; Buchinger, S.; Schomburg, D. *Anal. Biochem.* **2007**, *367*, 143–151.
(10) Wishart, D. S. *Briefings Bioinf.* **2007**, *8*, 279–293.

10.1021/ac802689c CCC: $40.75
Published on Web 04/09/2009

tification of chemical compounds.[11] Initially, AMDIS was developed for the verification of the chemical weapons convention and is, therefore, best suited for the accurate analysis of mixtures consisting of only a few compounds. If, on the other hand, AMDIS is used for the analysis of complex compound mixtures as they occur during metabolomics analysis, time-consuming manual intervention is necessary. Moreover, if the integration of peak areas is requested, the use of additional programs (e.g., Xcalibur) is required. Another open source program for targeted metabolomics analysis is MetaQuant.[12] Once a compound library is (manually) created, MetaQuant searches GC/MS chromatograms for these compounds and allows for an absolute metabolite quantification. Met-Idea[13] is a powerful software for the targeted extraction of individual single ion chromatographic peak areas and the subsequent determination of relative metabolite abundances. However, before this program is used, an input list of ion/retention time pairs has to be generated either manually or by the AMDIS software. In addition to these programs which are primarily intended for targeted metabolome analyses, there are further software packages for nontargeted analyses available including MathDAMP,[14] XCMS,[15,16] MSFACTs,[17] and MetAlign.[18] MathDAMP is implemented in the language Mathematica and visually highlights differences between complex metabolite profiles. The advantage of MathDAMP is that the whole chromatographic data are used for comparison and peak detection steps are not necessary. XCMS takes advantages of the statistical language R and is part of the Bioconductor project. It features a nonlinear retention time alignment of several (LC/MS based) chromatograms and a limited compound identification ability. MSFACTs is able to import chromatographic data in ASCII format, perform a retention time alignment, and export the adjusted and binned data into a two-dimensional matrix thus making it accessible for further statistical analysis. Finally, the commercial program MetAlign incorporates peak detection and several multivariate clustering tools in order to filter out statistically significant differences between GC- or LC/MS data sets.

We have developed MetaboliteDetector, an open source and C++ based software for a comprehensive and comparative analysis of GC/MS data of complex metabolite mixtures. MetaboliteDetector is especially suitable for the automatic analysis of a large series of GC/MS based spectrometric data. The software is able to perform a comparative investigation of many chromatograms either in a targeted or a nontargeted approach. Therefore, the chromatograms of interest are aligned, potential compounds are extracted, and peak areas are calculated for all of these compounds. Moreover, if a reference library is provided, MetaboliteDetector allows the identification of compounds based on the determined retention indices and the extracted mass spectra. Finally, a principal component analysis can be performed thus allowing the classification of interesting samples.

In order to demonstrate the main features of MetaboliteDetector, two GC/MS data sets have been analyzed. The first example shows a targeted analysis of a defined mixture of 10 amino acids measured in different concentrations. For the second case study, MetaboliteDetector has been applied to perform a nontargeted comparative metabolome analysis of 18 GC/MS chromatograms recorded for *Saccharomyces cerevisiae* cells sampled at three different time points during the diauxic shift, a well-known phenomenon in yeast cells: In the presence of high glucose concentration, *S. cerevisiae* ferments glucose and produces ethanol as a byproduct. When glucose is depleted, cell growth ceases and gene expression changes dramatically.[19] The cells undergo a diauxic shift by performing aerobic ethanol respiration and restart growing but at lower rates than during glucose fermentation.

MATERIAL AND METHODS

Growth Conditions. *Saccharomyces cerevisiae* CEN.PK 113-7D was kindly provided by P. Kötter (University of Frankfurt, Germany). The strain was streaked out from glycerol stocks on YPD agar plates (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) and incubated at 30 °C for 3 days. Several colonies were inoculated in 20 mL of YPD in buffled shaking flasks and incubated at 30 °C and 200 rpm for 9 h before being diluted 1:10 in 20 mL of fresh minimal medium[20] with 12.5 g/L glucose. This overnight culture was incubated for 14−16 h. The main culture (50 mL of minimal medium) was inoculated with the overnight culture to an OD_{600nm} of 1 and incubated until the desired growth phase (high glucose, low glucose, ethanol consumption) was reached.

Sample Preparation and Metabolite Extraction. For the metabolome analysis, cells were harvested at the desired growth phase with the method described before.[6] Specific amounts of cells (high glucose 10 mL, low glucose 6.5 mL, ethanol consumption 6 mL) were harvested by centrifugation at 3904*g* and 4 °C for 5 min to ensure comparable amounts of cells per growth phase. Afterward, the cells were washed twice with cold (4 °C) 0.9% (w/v) sodium chloride solution. Although quenching of the metabolism is certainly essential for a reliable and detailed analysis of the metabolic state of the cells, this was not essential for the purpose of the current paper as the differences between the three measured states were significantly larger than those between quenched and unquenched samples.[21,22] The extraction of intracellular metabolites was done according to the previously described extraction method followed by a two-step derivatization.[6,9] For this, cell pellets were resuspended in 0.75 mL of ethanol, containing 30 μL of a 0.2 mg/mL ribitol solution, and incubated in an ultrasonic bath for 15 min at 70 °C for cell lysis and metabolite extraction. Samples were cooled on ice before 750 μL of water were added and mixed followed by the addition of 1 mL

(11) Stein, S. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770–781.
(12) Bunk, B.; Kucklick, M.; Jonas, R.; Münch, R.; Schobert, M.; Jahn, D.; Hiller, K. *Bioinformatics* **2006**, *22*, 2962–2965.
(13) Broeckling, C. D.; Reddy, I. R.; Duran, A. L.; Zhao, X.; Sumner, L. W. *Anal. Chem.* **2006**, *78*, 4334–4341.
(14) Baran, R.; Kochi, H.; Saito, N.; Suematsu, M.; Soga, T.; Nishioka, T.; Robert, M.; Tomita, M. *BMC Bioinf.* **2006**, *7*, 530.
(15) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
(16) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 6382–6389.
(17) Duran, A. L.; Yang, J.; Wang, L.; Sumner, L. W. *Bioinformatics* **2003**, *19*, 2283–2293.
(18) America, A. H. P.; Cordewener, J. H. G.; van Geffen, M. H. A.; Lommen, A.; Vissers, J. P. C.; Bino, R. J.; Hall, R. D. *Proteomics* **2006**, *6*, 641–653.
(19) Brauer, M. J.; Saldanha, A. J.; Dolinski, K.; Botstein, D. *Mol. Biol. Cell* **2005**, *16*, 2503–2517.
(20) Saldanha, A. J.; Brauer, M. J.; Botstein, D. *Mol. Biol. Cell* **2004**, *15*, 4089–4104.
(21) de Koning, W.; van Dam, K. *Anal. Biochem.* **1992**, *204*, 118–23.
(22) Hans, M. A.; Heinzle, E.; Wittmann, C. *Appl. Microbiol. Biotechnol.* **2001**, *56*, 776–779.

of chloroform. The samples were shaken vigorously for chloroform extraction of the apolar phase. After phase separation by centrifugation (3904*g*, 5 min), 800 μL of the polar phase were taken, transferred to a conically shaped glass vial, and afterward dried in a vacuum concentrator overnight. The two-step derivatization procedure (methoxymation using a methoxyamine hydrochloride solution with a concentration of 20 mg/mL in pyridine followed by silylation applying MSTFA) was done automatically using a MPS2 Twister autosampler (Gerstel GmbH & Co.KG, Mühlheim an der Ruhr, Germany) equipped with a 10 μL syringe, ensuring no differences in endurance until measurement as described before.[9]

Standard Measurements. For absolute quantification, calibration measurements were performed. Each substance was measured in a number of concentrations ranging from 1 to 0.1 μmol/L. For each single substance, a standard solution with a concentration of 0.1 mol/L was prepared. Several substances were pooled in one stock solution with a concentration of 1 mmol/L each, before being diluted to the final concentrations. The samples were dried in a vacuum concentrator as described, before the two-step derivatization was performed. For this, 20 μL of methoxyamine/pyridine (20 mg/mL methoxyamine hydrochloride in pyridine) were added to the dried sample and incubated for 90 min at 30 °C with constant mixing. Afterward, 32 μL of MSTFA were added and again incubated for 30 min at 37 °C followed by 2 h at 25 °C with constant agitation. The samples were centrifuged at 14 000*g* for 5 min, and the supernatant was transferred into glass vials for GC/MS analysis.

GC/MS Measurements. Samples were analyzed using an AccuTOF MS (JEOL (Germany) GmbH, Eching, Germany) coupled to an Agilent 6890N fast GC (Agilent Technologies Sales & Services GmbH & Co.KG, Waldbronn, Germany) equipped with a DB-5MS column (Agilent (J&W Scientific), Folsom), a PTV-injector (Gerstel GmbH & Co.KG, Mühlheim an der Ruhr, Germany), and a MPS2 Twister autosampler (Gerstel GmbH & Co.KG, Mühlheim an der Ruhr, Germany). A total of 2 μL of the derivatized samples were injected (split ratio 1:25). Measurements were performed over 18 min. All parameters for sample injection, gas chromatography, and mass spectrometry as well as metabolite identification were described earlier.[9] As a time standard, for every 20 measurements, an *n*-alkane mix was analyzed to allow reproducibility of the measurements.

ALGORITHMS AND IMPLEMENTATIONS

MetaboliteDetector Description. Currently, MetaboliteDetector is able to import raw GC/MS data in centroid or profile netCDF format. The software of nearly all GC/MS instruments should be able to convert the recorded raw data into this format. Additionally, MetaboliteDetector imports raw profile data in FastFlight2 format as it is provided for example by JEOL Instruments. Since these data contain time-of-flight instead of mass to charge values, a calibration of the TOF dimension is necessary. Therefore, a suitable TOF calibration function was included in the software package. During a preprocessing step, the baseline of each recorded mass spectrum is analyzed and shifted if necessary. Moreover, noise present in the data is determined, and a data smoothing is performed. In the case of highly resolved profile data, a data compression can be performed.

Afterward, MetaboliteDetector detects single ion peaks present in the chromatogram, performs a deconvolution step, and finally extracts the mass spectra of potential compounds. For the purpose of compound identification or chromatogram comparison, it is strongly recommended to calculate the Kovats retention index (RI)[23] for each compound. In combination with a recorded GC/MS chromatogram of an appropriate calibration mixture (e.g., *n*-alkane mix), this step can be performed automatically for a number of chromatograms. Furthermore, if an internal standard (e.g. ribitol) was added to each sample, a constant time shift of the chromatographic dimension can be applied, yielding to a RI determination with a higher accuracy. If a reference compound library is provided, an identification of the detected compounds on the basis of their retention indices and mass spectra is feasible. This library can either directly be created with MetaboliteDetector or an existent library in NIST format can be imported. The results of the analyses are graphically presented. The user interface allows for the easy examination of the total ion chromatogram (TIC), the single ion chromatograms, the mass spectra at a particular time point, and the mass spectra of the detected compounds in comparison to spectra of the best matching library compounds. Additionally, the single ion chromatographic peak area of specific quantification ions is integrated and displayed in a table.

Although MetaboliteDetector offers a feature for the automatic determination of appropriate quantification ions, these ions can be set manually. Furthermore, MetaboliteDetector provides a function for either a targeted or a nontargeted analysis of a large series of GC/MS chromatograms in parallel resulting in a table presenting the detected compounds and their relative amounts across the different measurements. For this purpose, the chromatograms are aligned in their retention time dimension and single ion peaks appropriate for quantification are determined for each detected compound.

In addition to other programs, MetaboliteDetector allows for the extraction and further processing of highly resolved mass profiles. The strength of the novel software is the mainly automatic driven data analysis pipeline especially designed for high-throughput experiments. The analysis results are exported in tab delimited format and can be imported into spreadsheet programs, like OpenOffice or Microsoft Excel, or statistical programs, like R or MatLab. Furthermore, all diagrams can be exported into several image formats including png and jpeg. Finally, MetaboliteDetector offers the ability to perform a principal component analysis (PCA) with the analysis results. The application uses QT4 as basis for the intuitive and easy-to-use graphical user interface thus making it compatible with most modern operating systems.

The implementation of the MetaboliteDetector software package is divided into two main parts. The first part consists of the algorithmic implementations and is available as a C++ library named libgcms. libgcms provides all proposed algorithmic features and can be used in other C++ based projects. The second part of the implementation is the QT4 based graphical user interface of MetaboliteDetector and includes the main program which depends on the algorithms of libgcms. QT4 based programs can be compiled for Linux, Windows, and MacOS operation systems thus covering the majority of available systems. Figure 1 depicts a flowchart of the general operation of MetaboliteDetector. The

(23) Kovats, E. *Helv. Chim. Acta* **1958**, *41*, 1915–1932.

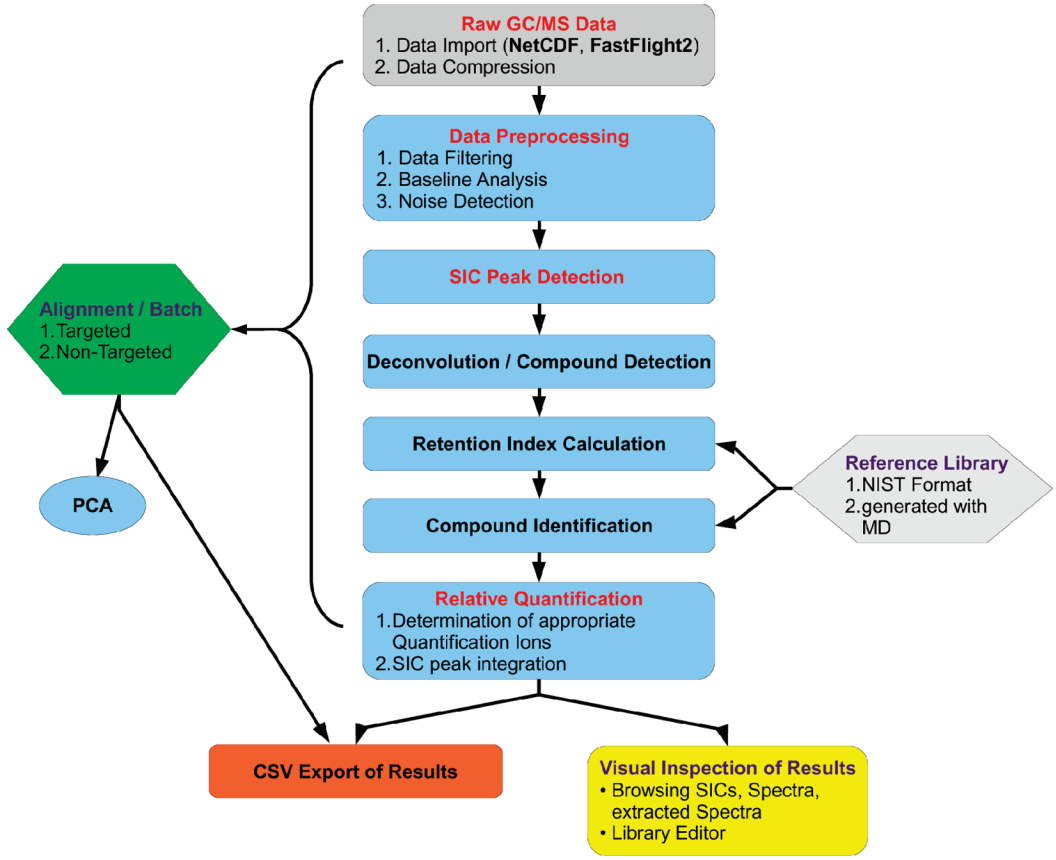

Figure 1. General analysis strategy of MetaboliteDetector. Raw GC/MS data can be imported in netCDF or FastFlight2 format. The analysis results are exported as CSV files.

main parts and algorithm which have been implemented are described below.

Data Preprocessing and Smoothing. Data retrieved by mass spectrometry are typically noisy. In order to smooth these data and improve the signal-to-noise ratio, each recorded spectrum can be filtered by applying a convolution filter. In the context of this work, a five point cubic Savitzky−Golay filter[24] has been applied to smooth both the spectral as well as the retention time (chromatographic) dimension of the data (eq 1).

$$f(x) = \frac{-3(x_{-2} + x_{+2}) + 12(x_{-1} + x_{x+1}) + 17x}{35} \qquad (1)$$

where $f(x)$ is the filtered value for the intensity x, x_{-2}, x_{-1}, x_{+1}, and x_{+2} are the neighboring intensities of x. Since we observed a positive baseline shift of highly resolved mass spectra containing strong signals, a baseline correction is performed for each recorded mass spectrum. For this purpose, the mean, median, and standard deviation of each baseline are calculated. According to Stein,[11] the spectrum of interest is divided into segments consisting each of 25 data points. These segments are then scanned for potential signal peaks, and only segments without detected signals are used for the baseline analysis. Finally, each value of a spectrum is subtracted by the determined mean of the baseline and thus made comparable in the chromatographic dimension.

A high degree of mass spectrometric noise is contributed by the ion counting detector. The contribution of this kind of noise is proportional to the square root of the measured signal. On the basis of the determined median and standard deviations of the mass spectrometric baselines, the noise proportionality factor n_f is calculated as described by Stein.[11] However, in contrast to Stein, here the proportionality factor is calculated on the basis of mass spectra instead of single ion chromatograms. In agreement with Stein, we found that this value shows a good run-to-run consistency on the same instrument. The proportionality factor is used for the calculation of the signal-to-noise ratio of detected peaks.

Peak Detection. Spectrometric as well as chromatographic peaks are detected by the analysis of the first derivative of the smoothed recorded intensity values. A special form of the Savitzky−Golay filter is applied to determine the first derivative and the smoothing of the data within one computational step. For this purpose, the following formula is used (eq 2):

$$f'(x) = \frac{-2x_{-2} - x_{-1} + x_{+1} + 2x_{+2}}{10} \qquad (2)$$

with $f'(x)$ as the first derivative of eq 1. In order to detect peak borders, the first derivative is examined sequentially. Once the slope excesses a predefined positive threshold, a potential peak beginning is set. Then, the peak is extended until the slope turns to negative values thus passing the peak maximum. The peak ending is recognized once two subsequent slopes pass a predefined negative threshold. A peak is counted as valid if the following two criteria are achieved: The peak must consist of more

(24) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627–1639.

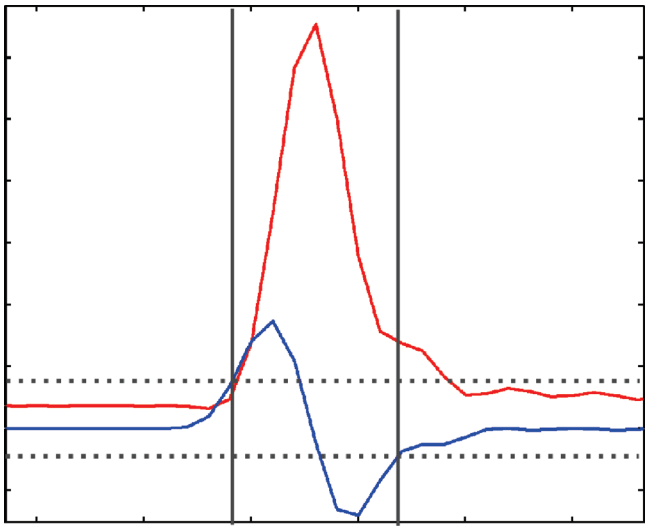

Figure 2. Single ion chromatographic peak detection. The peak borders are determined based on the first derivative of the intensity values. The red line represents the intensity values in dependence of the retention time. The blue line depicts the first derivative of the intensity values. If the values of the first derivative cross the peak threshold, a peak begin or end is set (dotted lines).

than three values and the height above the baseline in signal-to-noise units of the maximum peak value must exceed a predefined threshold (Figure 2). For all chromatographic peaks, a precise maximization time is interpolated by least-squares fitting a parabola to the three points surrounding the peak maximum as described by Dromey et al.[25] Finally, the quality of the peak shape is estimated by the application of a simple but effective algorithm. For ideal nonoverlapping single peaks, it can be assumed that all values of the first derivative must be positive inside the interval from peak begin to peak maximum; however, inside the interval from peak maximum to peak end they have to be negative. For this reason, the absolute values of the first derivative of peak intensity that follow this assumption are summed as ideal slopes (ideal $= \sum(\mathrm{d}I/\mathrm{d}t)$). Furthermore, all absolute values of the first derivative that disagree with this assumption are summed as nonideal slopes (nideal $= \sum(\mathrm{d}I/\mathrm{d}t)$). The discrepancy index of the peak shape (q_p) is defined by the percentage of the ratio of nonideal to ideal slopes (eq 3):

$$q_\mathrm{p} = \frac{\text{nideal}}{\text{ideal}} \times 100 \qquad (3)$$

Reasonable peak discrepancy values are in the range between 0% and 10%.

Deconvolution/Compound Detection. Frequently, compounds of highly complex metabolite mixtures elute at similar time points during GC/MS. If the retention times of two or more components differ in only a few scans, they often tend to form a single peak in the total ion current. Without further mathematical analysis, it is impossible to extract pure mass spectra for these compounds. An example for such a case is the coelution of glycerol-3TMS, phosphate-3TMS, and leucine-2TMS with retention times of 6.615, 6.620, and 6.632 min. Although these compounds

(25) Dromey, R. G.; Stefik, M. J.; Rindfleisch, T. C.; Duffield, A. M. *Anal. Chem.* **1976**, *48*, 1368–1375.

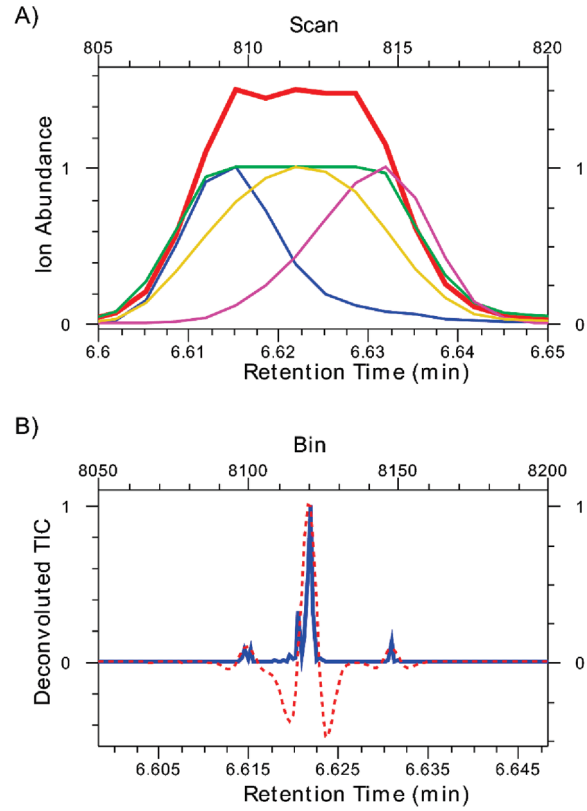

Figure 3. Ion chromatographic deconvolution. During GC/MS of highly complex mixtures, it is common that several compounds elute at similar retention times. Part A depicts single ion chromatograms of closely coeluting compounds in combination with the total ion chromatogram (TIC). The single ion chromatograms (158, magenta; 205, blue; 299, green; 314, orange) are base-peak normalized to a maximal value of 1 while the total ion current (red) is normalized to a maximal value of 1.5. It is obvious that these compounds could not be detected only based on total ion current intensity. In order to exactly detect such compounds and to extract pure mass spectra, it is essential to perform an ion chromatographic deconvolution step. The result of the applied deconvolution procedure is shown in part B of the figure. Single ion peaks are assigned to subintervals or bins based on their interpolated maximization times. In this case, each scan of the mass spectrometer was divided into 10 subintervals. The deconvoluted total ion current (blue graph) clearly shows the separation of the coeluting compounds. Finally the compounds are detected by the application of a second derivative Gaussian filter (dotted red line). All single ion peaks assigned to bins that are lying under peaks of the Gaussian filter are used to form the mass spectra of the detected compounds. The compounds were identified as glycerol-3TMS (6.613 min), phosphoric acid-3TMS (6.623 min), and leucine-2TMS (6.635 min).

appear as a single compound within the total ion current, the peaks of the single ion chromatograms can be separated based on their local maxima (Figure 3A). In order to detect these compounds as independent compounds and to extract pure mass spectra, it is essential to perform an ion chromatographic deconvolution step.

The applied deconvolution algorithm is based on the algorithms presented by Colby et al.[26] and Stein;[11] however, some improvements were made. For a detailed description of the original algorithms, we refer to the cited publications. At first, the peaks of all single ion chromatograms are detected based on the first derivatives of the single ion intensities as described

(26) Colby, B. N. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 558–562.

in the previous section. A sharpness value sp(n) is computed for each peak as proposed by Stein[11] (eq 4):

$$\mathrm{sp}(n) = \frac{A_{\max} - A_n}{nN_f\sqrt{(A_{\max})}} \quad (4)$$

with n as the distance from the peak maximum in scans, $A_{\max}$ the maximal peak intensity, A_n the intensity at position n, and N_f the global noise factor. The maximal sharpness values on both sides of the peak maximum are averaged and used during ion chromatographic deconvolution. Once these values have been calculated for all detected peaks, they are used for compound identification as follows: Each scan on the retention time dimension is divided into a predefined number of subintervals (bins). Afterward, each peak sharpness value is added to the bin representing its precise maximum (e.g., if the predefined number of bins is 10 and a peak has an interpolated maximum of 120.34 scans, its sharpness value is added to bin 1203). Metabolites are identified based on the maximization of these bin values (Figure 3B). In contrast to Stein, this procedure is performed by the application of a matched second derivative Gaussian filter.[27] The filter width is user-selectable and should be in the range between 0.8 and 3 scans depending on the type of analysis. Each peak of the Gaussian filter represents a detected compound (dotted red line in Figure 3B). All single ions that have a maximum within this range are assigned to this particular compound.

The next step comprises the determination of a model peak shape for each extracted compound. Therefore, all single ion peaks of a compound having a discrepancy index lower than 10% are sorted by their sharpness values. In the rare case that no peak passes this discrepancy index criterion, all single ion peaks of this compound are included for further processing. The intensities of those 25% of the peaks with the highest sharpness values are summed to form the model peak for this compound.

Whereas Dromey et al. only used the peak with the largest intensity value as a model, Stein extended this procedure and used all peaks that have sharpness values within 75% of the maximum value[11] for modeling. We agree with Stein that the use of additional ions increases the model accuracy especially for weak signals. However, in many cases only one peak fulfills the criterion defined by Stein. Therefore, we decided to use always 25% of all peaks of a compound for model peak construction. In order to prevent badly shaped peaks to be included in the model peak, the peak discrepancy index is taken into account. The determined model peak is then used for fitting all single ion peaks of this particular compound under consideration of a linear baseline[25] (eq 5):

$$I(n) = cM(n) + nb + a \quad (5)$$

with $I(n)$ as the measured peak intensity at scan n, $M(n)$ the intensity of the model peak at scan n, and c the calculated proportionality factor. $nb + a$ represents the linear baseline of the peak. $cM(n)$ denotes the peak abundance above background at scan n.[25,11] The values of this product ($cM(n)$) are used for the determination of the peak integral.

These deconvolution steps are performed for all single ion chromatograms contained in a GC/MS chromatogram. In the case of highly resolved GC/MS chromatographic data, about 80 000 single ion chromatograms are usually processed.

Spectrum Compression. Although modern GC/MS instruments record the whole spectrum profile for each chromatographic time point, not all of these data contain valuable information. Most parts of these mass spectra are only consisting of the background noise. Because a typical highly resolved GC/MS chromatogram needs about 1.5 GB of hard disk space, only information-containing parts of the mass spectra are extracted and stored by MetaboliteDetector. For this purpose, the spectral profile data is smoothed by a Savitzky−Golay filter (eqs 1 and 2) followed by the detection of all peaks based on the first derivative of the recorded ion counts. For each detected peak, an interval of 3 times the peak size around its maximum is kept. Thereby, parts of the baseline that are located next to the peak are included into the interval too. Only these parts of the spectrum are kept for further processing.

Similarity Measures/Library Matching. In the case of GC/MS, a detected compound is characterized by its mass spectrum and its retention time. The retention time is dependent on the used instrument, the GC-capillary, or the applied temperature program, etc. and is, therefore, not comparable once one parameter is changed. To overcome these limitations, the derived retention index as introduced by Kovats[23] is used for all retention time based similarity measures throughout this work. Since in most cases a linear temperature ramp is used for the gas chromatographic compound separation, the retention index is calculated in a linear way.[28,29]

Assuming that the determined retention indices for a certain compound are distributed in a Gaussian manner across different chromatograms, a Gaussian function is used for the RI based similarity index calculation (eq 6):

$$S_{RI} = e^{-\left(\frac{(RI(s_1) - RI(s_2))^2}{2\sigma^2}\right)} \quad (6)$$

with S_{RI} as the similarity score, RI(S_1) the RI of the first compound, and RI(S_2) the RI of the second compound. σ represents the user selectable RI window. This calculation results in a score ranging from 0 (no similarity) to 1 (identical retention indices). With dependence on the reproducibility of the applied experimental GC conditions, σ is typically set to a value between 2 and 10.

In addition to their retention indices, GC/MS derived compounds are further characterized by their recorded mass spectra. In the context of this work mass profiles are represented as vectors of intensity in defined mass bins. The cosine value (normalized dot product) of two profile vectors is used as a similarity measure for two mass spectra (eq 7).

$$S_{Spec} = \frac{P_1 P_2}{|P_1| \cdot |P_2|} \quad (7)$$

with S_{Spec} as the spectrum similarity score, P_1 the vector representing the mass profile of the first compound, and P_2 of

(27) Danielsson, R.; Bylund, D.; Markides, K. E. *Anal. Chim. Acta* **2002**, *454*, 167–184.
(28) Castello, G. *J. Chromatogr., A* **1999**, *842*, 51–64.
(29) Vandendool, H.; Kratz, P. D. *J. Chromatogr.* **1963**, *11*, 463–471.

the second compound. Before the calculation of this value is performed, the extracted mass spectra are centroided and the obtained intensities are transferred into vectors of a user selectable bin width (resolution). With dependence on the resolution of the recorded mass spectra, this value should be located within the range of 0.01−1 m/z units. Additionally, according to Stein,[11] each dimension of these vectors is weighted to decrease the relative importance of the larger peaks. The spectrum similarity score is located within the interval zero (no identity) to one (identical spectra). Finally, based on the spectral similarity S_{Spec} and the retention index similarity S_{RI}, a total similarity score is calculated. Since the spectral profile of a compound contains more information than the retention index, it is weighted stronger (eq 8).

$$S_{total} = \sqrt[3]{S_{Spec}{}^2 S_{RI}} \tag{8}$$

The values of S_{total} are in the range between 0 (no identity) and 1 (identical compounds).

Determination of Quantification Ions. Typically, the integrals of appropriate single ion chromatographic peaks are used for the determination of relative compound amount in comparative metabolomics experiments. The selection of these ions is not trivial and should meet at least these criteria: First, the selected ions should be representative for the particular compound, thus its single ion peak should not be superimposed by another peak. Second, the selected single ion peak should exhibit a high relative intensity in order to be present in most of the extracted spectra for this specific compound even if measured in low concentrations. Although these quantification ions could manually be selected for each detected compound, this procedure would be time-consuming. Therefore, we developed the following procedure to automatically detect these ions: During the first step, all single ion chromatographic peaks of a particular compound are ordered according to their relative intensities. Subsequently, those ions whose single ion peak discrepancy index (q_p, eq 3) does not exceed a predefined threshold are chosen as quantification ions for this particular compound. In many cases, some of the major relative intensities of a compound spectrum are contributed by the derivatization reagent and these ions are hence not suited for quantification. For this purpose, specific ions can be defined that should not be used for quantification. Furthermore, a minimal distance between two subsequent quantification ions can be adjusted.

TOF Calibration. The raw data of our JEOL GC−TOF instrument are saved in a special binary format called FastFlight2. Since the mass spectrometric unit used in this format is time-of-flight, all data had to be transferred to a mass to charge (m/z) based scale. In theory, the m/z ratio is proportional to the square of the time-of-flight (eq 9):

$$m/z = at^2 + b \tag{9}$$

A recorded mass spectrum in combination with accurate reference mass to charge ratios for each peak present in the spectrum can be applied for solving eq 9, thus obtaining parameters a and b. Preferably, this spectrum should consist of mass fragments equally distributed over the whole m/z range of interest. To meet these points, we used the mass spectrum of perfluorokerosene (pfk) for TOF calibration. However, in practice, the dependency between time-of-flight and m/z is further influenced by instrument specific factors. For this purpose, a polynomial of sixth degree is used for calibration. The parameter of the polynomial are determined as follows: (1) The centroids of all peaks detected in the recorded mass spectrum are determined. (2) A table of a few initial pairs connecting the recorded time-of-flight (centroid) with the corresponding m/z value has to be defined manually. (3) As a first approximation, a parabola of second degree is least-squares fitted to these defined data pairs. Therefore, the definition of at least three calibration pairs is necessary. (4) The determined function is then applied to calculate the m/z value for all peak centroids of the measured mass spectrum. (5) In order to obtain additional calibration pairs, these approximated m/z values are assigned to the theoretical m/z values of the reference spectrum. (6) The resulted table of calibration pairs is used for fitting a sixth degree polynomial, and the parameters of this polynomial are used for converting TOF values into m/z values.

Chromatogram Alignment/Batch Quantification. If a number of GC/MS chromatograms are to be analyzed comparatively, it is necessary to align their chromatographic dimension to match similar mass spectra. In the context of this work, a set of reference compounds for the alignment of the retention time dimension has been applied. Specifically, a set of *n*-alkanes has been used for the calculation of the retention index. In order to avoid interference with other metabolites, this mix has not been added to all samples but has been measured separately. However, because of rarely occurring synchronization inaccuracies of the system during the injection process, the whole chromatogram could be time-shifted by a few seconds. To overcome this problem, a known amount of ribitol has been added to all samples. After the calculation of the retention indices for all compounds contained in a chromatogram, the whole chromatogram has been time-shifted such that the determined retention index of ribitol exactly matches the retention index stored in the reference library. Another benefit of the ribitol addition is the capability of peak integral normalization. Because of the fact that the amount of added ribitol is constant throughout all measurements, all peak integrals can be normalized by the division of the ribitol peak area thus obtaining peak integrals that are comparable among different samples.

In order to compare GC/MS chromatograms recorded by different instruments or chromatograms from different organisms, the following strategies are performed by MetaboliteDetector. In the case of a targeted analysis, a reference compound library has to be provided by the user. This library can either be created with MetaboliteDetector or an existing library in NIST format can be imported. Subsequent to the compound detection and RI calibration process, the compounds of all chromatograms are consecutively matched against the provided reference library. If an extracted compound matches a compound of the library, the unknown compound is assigned to this reference compound. When all chromatograms have been processed, appropriate quantification ions for each reference are automatically detected based on the assigned compounds. Instead of the automatic quantification ions, it is possible to use manually defined ones which have been deposited in the reference library.

The next step consists of the peak area calculation. For that purpose, the area under those single ion peaks that have been determined as quantification ions during the previous step are integrated. The results are presented in table format, in which each row represents a certain metabolite and each column a specific sample. These results can be exported to CSV format for further processing in other programs.

In the case of a nontargeted analysis, the steps are similar to the targeted analysis with the exception that the reference library is automatically produced based on the GC/MS chromatograms of interest. For this purpose, the mass profiles and RIs of all extracted compounds of the first chromatogram are added to an empty library. Afterward, all remaining chromatograms are consecutively matched against this initial library. If a compound has been found in the library, the associated reference spectrum and RI are updated inside the library. However, if a compound has not been found within the library, this compound is added to the library as a new reference compound. This way, the initial library is continuously growing until all compounds of all chromatograms produce hits against the automatic library. The generated library is then applied for the described steps of the targeted analysis. It should be noted that until now the compounds of the generated library are only characterized by their RIs and mass spectra. However, the nontargeted analysis detects that a particular compound is present throughout multiple chromatograms and additionally performs a relative quantification based on the single ion peak integrals of the automatically detected quantification ions. If a real reference library is provided, MetaboliteDetector is able to identify parts of the automatic library.

RESULTS AND DISCUSSION

In order to demonstrate the main features and the performance of MetaboliteDetector, we analyzed two different metabolomics data sets. The first example presents a targeted analysis of a mixture of known metabolites with different concentrations. This analysis is intended to demonstrate the deconvolution (metabolite detection) process, the automatic quantification ion assignment, and finally the peak integration capabilities of MetaboliteDetector. The second analysis presents a nontargeted approach and analyzes the metabolome of *S. cerevisiae* cells harvested at different time points during the diauxic shift. These samples contain complex cell extracts with a large number of metabolites with different chemical features. The data set was used to demonstrate the capability of MetaboliteDetector in regard to peak differentiation, compound separation, and identification in highly complex metabolite mixtures.

In order to evaluate the performance of MetaboliteDetector, the analysis results of both data sets were compared with the results of more traditional mainly manually driven and, therefore, time-consuming quantification procedures. For the manual analysis of a group of samples, first the *n*-alkane-mix is set as calibration to calculate the retention indices. The chromatograms of the samples are then analyzed in AMDIS, where the peak deconvolution and identification is performed based on the *n*-alkane mix and our reference library. The AMDIS results are processed to be compatible with Xcalibur, a software in which the integration of the peak areas can be controlled manually. The final step is the summarization of the metabolite derivatives with an ExcelMacro and the output in a spreadsheet.[6,9]

Table 1. Comparison between Analysis Results of MetaboliteDetector vs Manual Processing[a]

	quantification ions		correlation concn vs. peak area	
compound	MD	manual	MD	manual
L-asparagine	116; 132; 231	333; 348	0.93	0.96
glycine	174; 175; 248	276	0.99	0.99
L-leucine	45; 158; 159	158	0.98	0.98
L-lysine	128; 156; 174	174; 317	0.99	0.99
L-methionine	61; 128; 176	176	0.99	0.99
L-phenylalanine	75; 91; 120	120	0.98	0.99
L-serine	75; 116; 132	132	0.99	0.96
L-threonine	117; 218; 219	218; 291	0.99	0.99
L-tryptophan	202; 203; 204	291	0.98	0.99
L-valine	144; 145; 218	144; 218	0.98	0.99

[a] Shown are the automatically determined quantification ions for the 10 amino acids and the quantification ions used for the manual integration. Furthermore, the correlation coefficient between the obtained peak integrals and the amino acid concentration are presented. The peaks of the following amino acid derivates were used: asparagine-3TMS, glycine-3TMS, leucine-2TMS, lysine-4TMS, methionine-2TMS, phenylalanine-1TMS, serine-2TMS, threonine-3TMS, tryptophan-3TMS, and valine-2TMS. MD, MetaboliteDetector.

The first data set consists of a mixture of the following 10 amino acids: glycine, L-asparagine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-serine, L-threonine, L-tryptophan, and L-valine. As internal standard for retention time correction and normalization purposes, a specific amount of ribitol was added to all samples. This standard mix has been measured in six different dilutions considering three replicates for each dilution: 1, 10, 50, 100, 250, and 500 mM. Additionally, to allow the RI calculation, a data set of an *n*-alkane mix has been recorded. In order to perform a targeted analysis, a library containing reference mass spectra and RIs of appropriate amino acid derivatives has been constructed with MetaboliteDetector. During this procedure, particular quantification ions have been automatically chosen by the software. Afterward, the retention time dimensions of all chromatograms have been shifted to align the peaks of the internal standard ribitol-5TMS. Finally, MetaboliteDetector was used to perform a targeted batch quantification of all chromatograms. The correlation coefficient between the amino acid concentration and the obtained peak area has been calculated for the automatically driven and the manual analysis (Table 1). Additionally, Table 1 shows the applied quantification ions which have been chosen automatically in the case of MetaboliteDetector. Typically, many ions of a specific compound pass the previously defined criteria to function as a quantification ion. The single ion peak integrals are in most cases directly correlated to each other and could be exchanged, with the automatic procedure allowing a more rigorous approach. Therefore, the manually and automatically chosen quantification ions differ in the case of L-asparagine, glycine, and L-tryptophan (Table 1).

The correlation coefficients between amino acid concentration and peak area indicate a high degree of correlation ($R > 0.98$) for most amino acids and both analysis methods. The only exception is L-asparagine with a correlation of 0.93 in the case of MetaboliteDetector and 0.96 in the case of the manual integration. In most cases, both the automatic and the manual analysis performed nearly equally well. These results render both methods valid for an accurate peak integral determination; however,

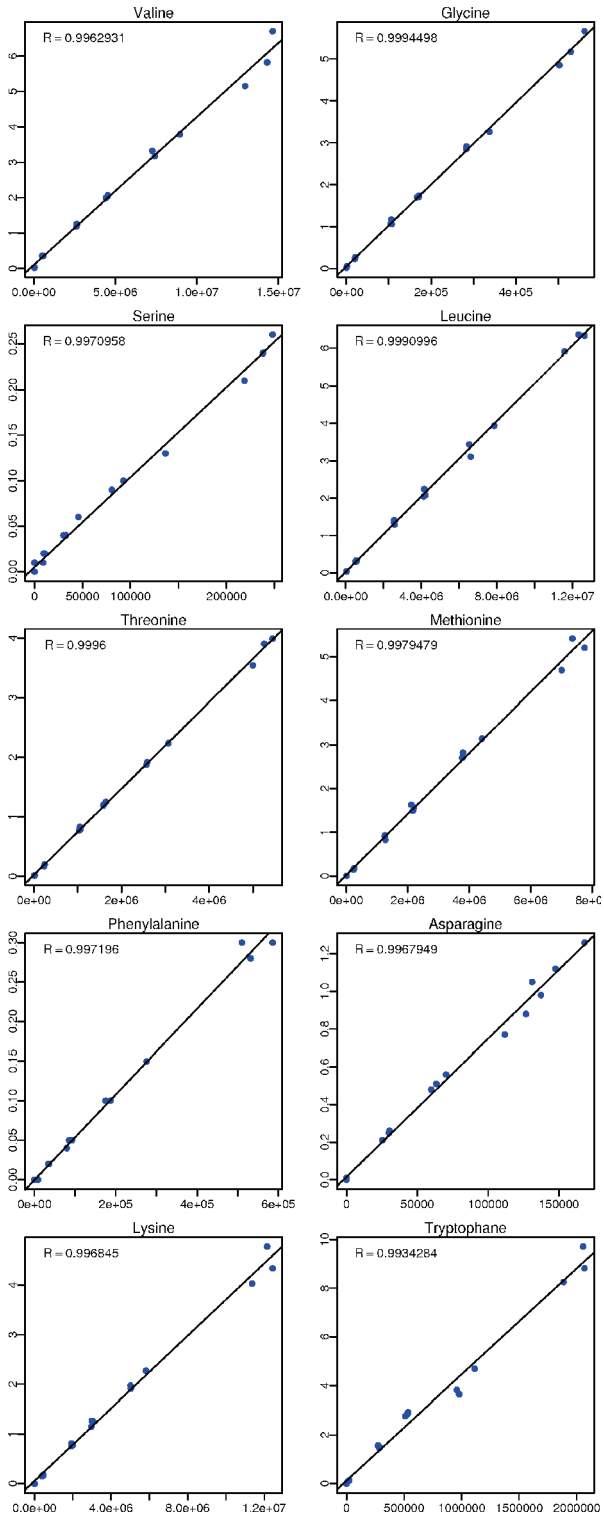

Figure 4. Scatterplots of MetaboliteDetector based quantification values vs manually integrated values for 10 amino acid derivatives. The *X*-axis depicts the manually obtained peak areas and the *Y*-axis the automatically derived values of single ion chromatographic peaks used for quantification (Table 1). All peak areas (manual and automatic) have been normalized by the single ion peak area of ribitol5TMS.

MetaboliteDetector does this analysis in approximately 10 min without manual intervention.

Furthermore, scatter plots of the values obtained by MetaboliteDetector vs the manually obtained values were generated for all measured amino acids (Figure 4). The adjusted correlation coefficients (R) indicate in all cases that the automatically obtained values are highly correlated to the manually determined values ($R > 0.99$). For this reason, it is obvious that the novel developed algorithms provide data equivalent to the traditional manual method, however, in an automatic time saving and objective fashion.

The second data set consists of metabolomics data of *S. cerevisiae* measured at three different time points during the diauxic shift. The samples were taken during the exponential phase, in which glucose is fermented and ethanol is produced (high glucose), after glucose depletion (low glucose), and during aerobic ethanol respiration (ethanol consumption). Since each time point was measured in six replicates, the whole data set contains 18 samples, named A1−6 for high glucose, B1−6 for low glucose and C1−6 for ethanol consumption. In order to determine the RIs of the detected compounds, a GC/MS chromatogram of an *n*-alkane mix has been recorded additionally. Furthermore, a specific amount of ribitol-5TMS was added as internal standard to each sample allowing a normalization of the extracted peak integrals as well as the compensation of constant retention time shifts. Before a nontargeted analysis was performed, all potential compounds have been detected for all GC/MS chromatograms and the chromatograms were time shifted in order to align all ribitol-5TMS peaks, thus enhancing the subsequent RI calculation step.

The following parameter have been set for the automatic nontargeted analysis of all samples: peak threshold, 8; peak height, 5; deconvolution width, 1; RI window, 10; required similarity for batch analysis, 0.7. A known problem of the applied deconvolution algorithm is the extraction of artificial mass spectra. Many of these spectra are characterized by a small number of peaks and a poor peak quality. In the case of highly resolved mass profiles, an extracted and well separated spectrum typically consists of more than 100 peaks. Therefore, a filter was applied that eliminates all mass spectra with less than 100 single ion peaks and all spectra with an estimated average peak discrepancy index greater than 15. On the basis of these settings, MetaboliteDetector was able to detect 239 compounds that were verifiable in at least 6 samples and 205 of them were traceable in all 18 samples. In combination with our in-house reference library (containing RIs and reference spectra of 513 compounds), MetaboliteDetector was able to identify 135 of these compounds. The more traditional manual analysis revealed 152 compounds that were found in at least 6 samples with 139 of these detectable in all 18 samples. Although MetaboliteDetector was able to detect more compounds than the manually driven analysis, not all of them caused hits with our reference library and hence were not identifiable. Since the applied scoring procedures differ between the automatic and manual analysis pipeline, the number of identifiable compounds differs between both methods. The time needed for the whole analysis has been ~45 min for MetaboliteDetector and ~20 h for the manually driven analysis.

In order to get an overview of the automatically determined data, a heat map has been calculated (Figure 5). All growth phases could be clearly separated from each other. It is obvious from the figure that the metabolic profile of the more stationary-like growth phases “low glucose” (B1−6) and “ethanol consumption” (C1−6) are more closely related to each other than

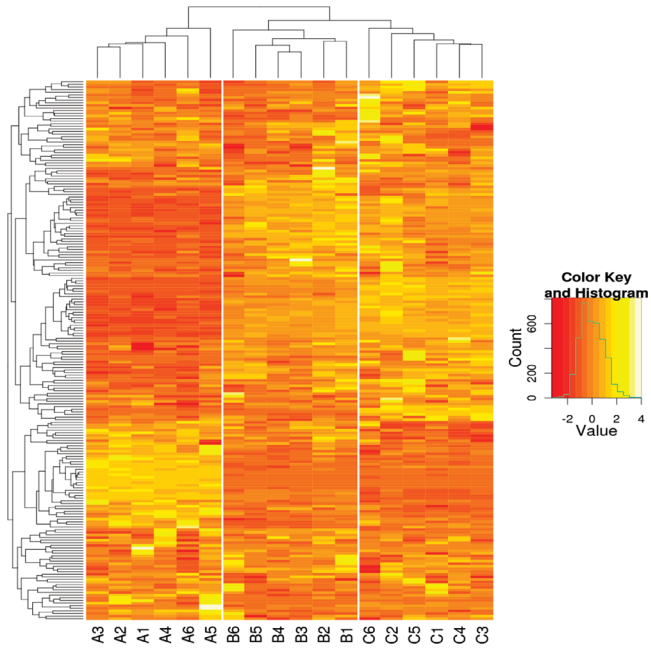

Figure 5. Automatically obtained nontargeted metabolome data of three different growth phases of *Saccharomyces cerevisiae*. The heat map compares all metabolites that were traceable in the GC/MS chromatograms of all samples. Three different growth phases of the diauxic shift of *S. cerevisiae* were analyzed in six replicates: "high glucose" (A1−6), "low glucose" (B1−6), and "ethanol consumption" (C1−6). Each row of the plot represents a detected metabolite, each column the metabolome of one sample. The cells were colored according to the *Z*-score normalized semiquantitative metabolite amount, with yellow meaning increased amounts and red meaning decreased amounts in relation to the mean. The histogram depicts the range and frequency of changes in units of standard deviations. The heat map has been generated with the "gplots" package of *R* using the function "heatmap.2". It is obvious that the three growth phases could clearly be separated based on the data obtained by MetaboliteDetector.

to the exponential growth phase "high glucose" (A1−6). In these phases, the cells are energy-depleted because of low availability of the carbon sources ("low glucose") and the lower energy generation during ethanol respiration. Although similar patterns have been observed for the analysis of the manual data, the separation pattern of the automatically derived data clearly distinguishes between the different growth phases.

Since the metabolic profile of each growth phase was measured in six replicates, it was possible to calculate the relative standard error of the mean for each compound and each set of replicates. The relative standard error of the mean is a combined measure of both the biological variance and the technical variance and is, therefore, capable to evaluate the analysis accuracy. The mean of all relative standard errors has been calculated for both analysis types. The average estimated means of the standard error have been 13% for MetaboliteDetector and 12% for the manual integration. In order to further prove the quantification function of the program, a correlation analysis has been performed (Figure 6). For this purpose, the mean integral for each metabolite and sampling point has been calculated for both analysis types. The logarithms of the values have then been applied for the calculation of the correlation coefficients between the three different sampling points. Figure

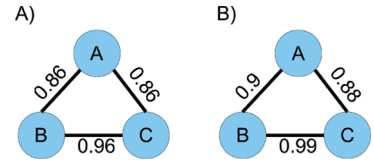

Figure 6. Correlation analysis between the metabolomes of the three different sampling points of the diauxic shift. (A) Correlation coefficients have been calculated on the basis of automatically derived single ion chromatographic peak integrals for 239 metabolites. (B) Same analysis, however, based on manually obtained peak integrals of 151 compounds. Results obtained with MetaboliteDetector reveal a similar structure compared to the results of the traditional manual analysis. A, "high glucose"; B, "low glucose"; and C, "ethanol consumption".

6A depicts the results for the data obtained by MetaboliteDetector. Samples B and C exhibit the highest correlation, A and C the lowest. These results clearly support the results obtained by the heat map visualization. The correlation coefficients obtained from the manual analysis are shown in Figure 6B and are nearly identical to those automatically derived by MetaboliteDetector. Considering the short analysis time in combination with an accuracy comparable to a manually driven analysis, the main advantage of MetaboliteDetector lies in the analysis of high-throughput based metabolomics data. Moreover, in contrast to the manual method, MetaboliteDetector is able to operate without a reference compound library, if running in nontargeted mode.

CONCLUSIONS

MetaboliteDetector is a novel software especially suited for the analysis of highly resolved and high-throughput based GC/MS data as they accumulate during modern experiments. The automatical analysis takes GC/MS raw data as input and exports the analysis results to CSV tables. In addition to the browsing and visual inspection capabilities of single GC/MS chromatograms, MetaboliteDetector allows for the targeted as well as nontargeted analysis of a large number of chromatograms. All detected compounds are presented in rows with the corresponding single ion chromatographic peak integrals of particular quantification ions located in columns of the result table. In contrast to time-consuming manually driven analysis pipelines, the application of our novel software delivers reliable and objective results. The algorithms used in MetaboliteDetector are also applicable for chromatographic data originating from LC/MS. It is planned to extend the program for this type of analysis in the near future. Furthermore, we would like to extend the import capabilities of MetaboliteDetector. As soon as information about further (native) file formats will be available, an import filter for these particular formats will be added. In order to make the algorithmic implementation of MetaboliteDetector reusable in other projects, all described algorithms are implemented in an object orientated C++ library named libgcms. The authors would like to encourage other developers to participate in this project. MetaboliteDetector is freely available under the GNU public license (GPL) at http://metabolitedetector.tu-bs.de. Currently, installable packages for Linux (Debian, Red Hat packages) and Windows are provided.

ACKNOWLEDGMENT

The authors would like to thank Boyke Bunk for useful discussions about the netCDF format and the peak detection algorithm. Timo Lühr provided the logos for MetaboliteDetector and the project's website. Furthermore, the authors would like to thank Till Beuerle and Ulrich Papke for fruitful discussions about mass spectrometry. The authors thank the German ministry for education and research (BMBF) for funding.

Received for review December 19, 2008. Accepted March 22, 2009.

AC802689C