Journal of Chromatography B, 871 (2008) 191–201

Contents lists available at ScienceDirect

Journal of Chromatography B

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

Standardizing GC–MS metabolomics☆

Harin Kanani[a,1], Panagiotis K. Chrysanthopoulos[b,c], Maria I. Klapa[a,b,*]

[a] Metabolic Engineering and Systems Biology Laboratory, Department of Chemical and Biomolecular Engineering, University of Maryland, MD 20742, USA
[b] Metabolic Engineering and Systems Biology Laboratory, Institute of Chemical Engineering and High-Temperature Chemical Processes, Foundation for Research and Technology-Hellas, Patras GR-265 04, Greece
[c] Division of Genetics and Cell and Developmental Biology, Department of Biology, University of Patras, Patras GR-265 00, Greece

ARTICLE INFO

Article history:
Received 18 February 2008
Accepted 30 April 2008
Available online 21 May 2008

Keywords:
Quantitative systems biology
High-throughput "omic" techniques
Data correction and normalization
Data validation
Derivatization biases
TMS-derivatives
Methoximation
Metabolite extraction

ABSTRACT

Metabolomics being the most recently introduced "omic" analytical platform is currently at its development phase. For the metabolomics to be broadly deployed to biological and clinical research and practice, issues regarding data validation and reproducibility need to be resolved. Gas chromatography–mass spectrometry (GC–MS) will remain integral part of the metabolomics laboratory. In this paper, the sources of biases in GC–MS metabolomics are discussed and experimental evidence for their occurrence and impact on the final results is provided. When available, methods to correct or account for these biases are presented towards the standardization of a systematic methodology for quantitative GC–MS metabolomics.

1. Introduction

The post-genomic era is characterized by two major shifts in the way problems in life sciences are now approached. The first refers to what is known as the "high-throughput" revolution triggered by the development of the "omic" technical platforms that allow for the simultaneous measurement of hundreds to thousands of molecular quantities. Thus, rather than examining a small number of genes and/or reactions at any one time, the focus shifts to the analysis of gene expression and protein activity in the context of networks and systems of interacting genes and gene products [1]. The second major shift in biological research concerns the importance that has been attributed to quantitative biology. It is indeed essential to know the structure of a particular gene-, protein- or metabolic-network. However, this alone is insufficient to describe how the *in vivo* state of the cellular function(s) that is(are) described from this network changes depending on the physiological conditions and/or the biological system. Quantitative analysis of the molecular quantities that define the activity of this network, e.g. gene expression, protein concentration, protein activity, metabolite concentration or metabolic flux, is required. Based primarily on these two major shifts, the post-genomic was granted as the era of the quantitative systems biology revolution. To succeed in the challenge of quantitative systems biology, major issues concerning the quantification capabilities of the high-throughput molecular analysis techniques for each level of cellular function need to be resolved. They range from limitations in the available experimental protocols to lack of data analysis techniques for upgrading the information content of the acquired measurements.

Metabolomics is the most recently introduced [2,3], but currently one of the fastest growing, high-throughput molecular analysis platforms. It refers to the simultaneous quantification of the (relative) concentration of the free small metabolite pools of a biological system [4]. It provides thus a comprehensive metabolic fingerprint, correspondent at the metabolic level of the high-throughput transcriptional and proteomic profiles [2]. Considering the role of metabolism in the context of the overall cellular function, it is easily understandable why quantifying a complete and accurate metabolomic profile is among the major goals of quantitative systems biology and metabolic pathway engineering.

☆ This paper is part of a special volume entitled "Hyphenated Techniques for Global Metabolite Profiling", guest edited by Georgios Theodoridis and Ian D. Wilson.

* Corresponding author at: Metabolic Engineering and Systems Biology Laboratory, Institute of Chemical Engineering and High-Temperature Chemical Processes, Foundation for Research and Technology-Hellas, Patras GR-265 04, Greece.
Tel.: +30 2610 965249; fax: +30 2610 965223.
E-mail address: mklapa@iceht.forth.gr (M.I. Klapa).
[1] Present address: Pioneer Hi-Bred International, Inc., IA 50131, USA.

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.04.049

Table 1
Reported applications of metabolomics in life sciences research and practice

Basic research	References
Functional genomics	[2,5,6]
Interaction between metabolome, transcriptome and proteome	[7,8]
Discovery of new biochemical pathways	[9]
Interaction between species	[10]
Analysis of metabolic regulation	[11]
Applied research	**References**
Medical applications	
Understanding of disease pathophysiology	[12,13]
Disease biomarker identification	[14,15]
Early diagnosis	[16]
Personalized medicine	[17–19]
Clinical trial monitoring	[32,33]
Drug discovery	[34,35]
Toxicology—Drug safety	[36]
Agricultural/Nutrition Applications	
Identification of metabolic engineering targets	[20,21]
Understanding of stress response	[22,23]
Classification of special varieties of produce (e.g. tea, ginseng, fish)	[24–26]
Genetically modified (GM) food certification	[27–29]
Human nutrition	[30,31]
Industrial Applications	
Identification of metabolic engineering targets in *Escherichia coli*, yeast, algae	[37]
Fermentation process improvement	[38]
Biologics production and fermentation process optimization	[39]

The selection of the references in this table was mainly based on the time of publication, favoring the most recent (publication date in 2007–2008) in reputed journals of the respective application field. In addition, the table contains the earliest "path-breaking" manuscripts, which were the first to demonstrate the concept and projected impact of metabolomics.

Quantitative metabolomics is foreseen to have a major positive impact in (agri-)biotechnology, disease prognosis and diagnosis, drug design and development, personalized medicine and many other applications (see Table 1) [5–39]. However, the broad deployment of the metabolomic analytical platform to biotechnology and clinical research and practice requires its standardization for accurate, reproducible and validated performance. Failure to achieve this technological status may end up limiting the application of metabolomic analysis, despite its great potential.

Considering that its first application was reported in 2000, metabolomics is presently at its development stage. There are many biological systems and applications to which metabolomics has not been widely or even at all used to-date. Similarly to other "omic" technologies in the past, during this standardization stage the metabolomic data acquisition and analysis protocols need to be optimized and any current limitations regarding data validation, normalization and analysis need to be thoroughly addressed [4]. This paper deals with the standardization of quantitative gas chromatography–mass spectrometry (GC–MS) metabolomics. GC–MS has been to-date the analytical technique of choice for most metabolomic analyses [4]. It is expected to remain integral component of the metabolomics laboratory, used either alone or preferably in combination with other metabolomic analytical platforms, which today include the liquid chromatography (LC) or the capillary electrophoresis (CE)–MS or the nuclear magnetic resonance spectroscopy (NMR) [40,41]. The characteristics that render GC–MS advantageous for metabolomic analysis include (a) high sensitivity that decreases significantly the amount of raw biological material needed for accurate measurements, (b) better separation of compounds in the gas than in the liquid phase, (c) extensive compound databases and experimental protocols, since it has for long been used in other clinical, forensic and biotechnology applications, (d) the lowest purchase, operation and repair cost from all available metabolomic technologies, and (e) its user-friendliness compared to the other technologies. The latter characteristic assists in faster personnel training and less intricate development of new, and adjustment of existing, data acquisition methods to address the needs of a new application and/or biological system.

One could argue that the standardization of GC–MS metabolomics should be a straightforward task, considering that GC–MS has been for long used to a vast number of applications in many disciplines. However, this is not true, because the use of GC–MS in a high-throughput way for the simultaneous quantification of metabolites that belong to a wide range of functional chemical classes, to extract biologically relevant conclusions in the context of a variety of applications and biological systems, poses unique challenges that are not to be encountered in other cases (see e.g. [42]). These challenges can be addressed only after the sources and types of biases in GC–MS metabolomics are identified, and their impact on the multivariate analysis and interpretation of the acquired data is understood. In this paper, the sources of biases in GC–MS metabolomic analysis are discussed in detail and experimental evidence for their occurrence and their impact on the extracted results is provided. Subsequently, when available, methods to correct or account for these biases are presented towards the standardization of a systematic methodology for quantitative GC–MS metabolomics. The ultimate goal is for the final results to be filtered from any experimental biases, ensuring that any observed changes are due only to biological reasons.

2. Experimental

In this paper, data from GC–MS metabolomic experiments is presented to provide evidence for the occurrence of certain of the discussed biases and their impact in the derived conclusions. This data was acquired from two biological systems: *Arabidopsis thaliana* plant liquid cultures and mouse brain (cortex or cerebellum) tissue. If not otherwise specified in the text, the two types of biological samples were, respectively, processed as follows.

2.1. A. thaliana plant samples

2.1.1. Acquisition of plant samples

A. thaliana liquid cultures were grown for 12–13 days on an orbital shaker platform (Barnstead, IL) at 150 rpm, in the ambient air (350 ppm CO_2) of a growth chamber (model M-40, EGC Inc., Chagrin Falls, OH), under constant white light intensity (80–100 $\mu E\,m^{-2}\,s^{-2}$) and temperature (23 °C). The seeds had been cleaned [43] and stored overnight at 4 °C prior to inoculation. The plant cultures grew in 500 mL shake flasks, each containing 200 mL B5 Gamborg media [44] with minimal organics (Sigma, St. Louis), 2% (w/v) sucrose and 0.1% agar and inoculated with ~100 Columbia ecotype seeds. Some of the plants, whose metabolomic profiles are used in this manuscript, were grown during the 13th day either (a) in elevated CO_2 (10 000 ppm) in the air, or (b) 50 mM NaCl in the media, or (c) with 10 mM trehalose in the media, or in combination of (a) and (b) or (a) and (c) conditions. At the time of harvest, all the seedlings of a flask were simultaneously removed from the liquid media using forceps. Subsequently, they were twice dipped in de-ionized water, dried on filter paper and wrapped in aluminum foil before being frozen in liquid nitrogen. The process from harvest to freezing lasted 15–30 s. The experiments took place at the Green House Facility of the University of Maryland, College Park, MD 20742, USA.

2.1.2. Preparation of the metabolite standard mix sample

Standards of 30 metabolites were dissolved in methanol and water (1:1) and mixed to form a metabolite standard mix solution. The 30 metabolites belonged to various functional groups such as: organic acids, alcohols, sugars and amino acids, and mixed in a proportion, which imitated the measured average composition of *A. thaliana* liquid culture samples. The mix solution was subsequently dried in vacuum. Detailed list of the metabolites used and their composition in the mix is provided in [4].

2.1.3. Metabolomic profiling

The polar extracts of the plant liquid cultures were obtained from 125 mg of ground culture using methanol/water extraction [2] and ribitol (0.2 mg/g of fresh weight) as the internal standard. The dried polar extract of each plant or the dried metabolite standard mix was derivatized to its (Meox-)TMS-derivatives through 2 h reaction with 100 μL of 20 mg/mL methoxyamine hydrochloride solution in pyridine, followed by at least 4 h (see [4] and explanation in the text) reaction with 200 μL of *N*-methyl-trimethylsilyl-trifluoroacetamide (MSTFA), both at room temperature. The metabolomic profiles were acquired using the Saturn 2100T gas chromatograph–(ion trap) mass spectrometer (Varian Inc., CA). The peak identification and quantification was carried out as described in [4].

2.2. Mouse brain tissue samples

6 male Balb-c mice were bred in the Animal House Facility of the Laboratory of Human and Animal Physiology, Department of Biology, U. Patras, Patras, Greece, according to the standards of the international statutes on animal housing and handling (86/609/EEC). Specifically, the animals were housed 8 per cage, exposed to regular light–dark cycle (i.e. light period: 7 a.m. to 7 p.m.; dark period: 7 p.m. to 7 a.m.) at $22 \pm 1\,^{\circ}C$ and bred with laboratory chow and water *ad libitum*. On the 60th day of their life (P60), between 9:00 and 11:00 a.m., the animals were decapitated under light ether anaesthesia, according to the international statutes on minimizing animal pain (86/609/EEC). After sacrifice, whole brains were removed in a sterile cooled glass plate and the cerebral cortex (of the two cerebral hemispheres) and cerebellum were immediately isolated. Subsequently, each brain region was rapidly weighed, frozen in liquid Nitrogen and stored at $-80\,^{\circ}C$ until further analysis. The extraction and derivatization were carried out similarly to the plant samples, as described in section 2.1.3, except that (a) 11 mL methanol/0.5 g of fresh tissue was used for extraction, (b) [U-^{13}C]-glucose was used as internal standard, and (c) 150 μL (20 mg/mL) methoxyamine solution in pyridine and 300 μL of derivatizing agent MSTFA were used for derivatization. Each derivatized sample was run at 25:1 split ratio on a Saturn 2200 gas chromatograph-(ion trap)mass spectrometer (Varian Inc.).

3. Results and discussion

3.1. Description of the GC–MS metabolomics analytical platform

The GC–MS metabolomic analysis comprises five (5) steps, as shown in Fig. 1. Among these, three refer to experimental processes up to the acquisition of the peak area profile for each biological sample, and the last two to analytical procedures up to the extraction of biologically relevant conclusions from the acquired datasets. Specifically, the three experimental and two analytical steps of the GC–MS metabolomics analytical platform can be described briefly as follows.

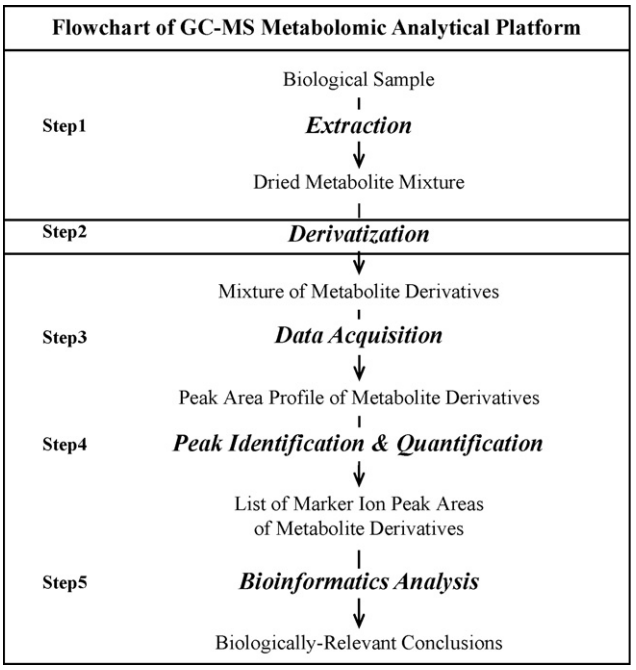

Fig. 1. Schematic diagram of the GC–MS metabolomics analytical platform.

3.1.1. Extraction

This step involves the isolation of the free metabolites from the biological sample. The free polar metabolites are extracted in the supernatant of a water–methanol solution of the homogenized biological sample after the necessary processing and centrifugation for the separation of the macromolecules [2,4]. In the case of the non-polar metabolites, chloroform is also used as the extraction solvent [3]. The free metabolite solution is subsequently dried in vacuum. The acquired dried metabolite mixture could then be used for further analysis.

3.1.2. Derivatization

This is the step that characterizes GC–MS metabolomics (shown in gray background in Fig. 1). Only volatile and thermally stable compounds are to be run through the gas chromatograph. Since most of the metabolites do not possess these qualities, in GC–MS metabolomics it is mandatory for the free metabolite mixture to react with a derivatizing agent prior to the run. There are many derivatizing agents; the selection is based primarily on the investigated problem and the class of molecules that have to be accurately quantified. The derivatization step can introduce a number of biases to the GC–MS metabolomic data, which are not encountered in chemical analyses that do not require derivatization, as it will be described in greater detail in one of the next sections. The difficulty lies on the fact that these biases can affect each metabolite to a different extent, complicating thus the development of methods to correct the data from these errors. At the derivatization step, the dried metabolite mixture reacts with the selected derivatization agent(s) in a particular solvent; attention should be paid for the derivatization agent to be "in excess" to allow for all metabolites to be derivatized. The composition of the derivatized metabolite solution is a function of the derivatization time until the completion of the derivatization reaction for all metabolites. At this time, only metabolite derivatives are present in the solution and their composition will no further change with time, barring degradation effects.

3.1.3. GC–MS data acquisition

A small volume of the metabolite derivative solution with the derivatizing agent is run through the gas chromatograph-

mass spectrometer. As implied from the previous description of the derivatization step, the GC–MS metabolomic profile of a biological sample depends on the derivatization time at which it is acquired. How the relationship between the GC–MS metabolomic profile and the derivatization time is manifested in the acquired data and affects the accuracy of the metabolomic analysis will be discussed in greater detail in a subsequent section.

3.1.4. Peak identification and quantification

The identification of peaks in the GC–MS metabolomic profile is not an easy task. It is based on the comparison with available mass spectrum libraries and samples of standards run by each laboratory at their own equipment. There are software tools that assist in a semi-automated peak identification. However, challenges as (a) the chromatographic drift that will be discussed in a subsequent section, (b) very similar chemical structures of molecules in biologically significant classes, as the mono- or di-saccharides, (c) the vast number of peaks of yet unknown origin, (d) the occurrence of peaks that are products of the GC column bleeding and do not correspond to metabolites, and (e) the multiple derivatives of particular metabolites that are produced at different derivatization times (as will be explained later), render this step the "bottleneck" for the speed and success of the entire analytical process.

Once the peak list has been populated, the peak area quantification is based on the selection of the marker ion for the metabolite derivative to which each peak corresponds. Within the linear range of the equipment's operation, the marker ion peak area (or height) is proportional to the concentration of the metabolite derivative in the derivatized solution at the time of the run. The proportionality coefficient is the response factor (RF) of the metabolite derivative, and depends only on the conditions that determine the detection and quantification capabilities of the GC–MS equipment [45], as it will be further discussed in a subsequent section.

3.1.5. Bioinformatic analysis

The significance of the metabolomic analysis lies on the fact that it can contribute to the derivation of significant conclusions about the metabolic activity of a biological system at particular physiological conditions. Thus, for the metabolomic profiling data to be used successfully in the bioinformatics analysis, each profile has to directly reflect the composition of the free metabolite mixture that is extracted from the raw biological sample, corrected from any biases that distort this relationship and are due to reasons that are independent of the biological problem.

3.2. Two types of biases in GC–MS metabolomics

Quantitative GC–MS metabolomics is possible if the change in the marker ion peak area of each metabolite derivative that is quantified in the profiles is proportional to the change in the concentration of the free metabolite in the raw biological samples. Each of the experimental steps described above introduces biases that distort this proportionality. To determine potential solutions to avoid the occurrence of a bias or to account for its impact in the final outcome, one needs to understand how the particular bias disturbs this proportionality. All encountered experimental biases in GC–MS metabolomics can be classified in one of the following two categories.

Type A: When occurring, these biases affect the entire sample uniformly, and hence vary the measured signal of all the metabolites to the same extent.

Type B: When occurring, these biases affect individual metabolites differently, and hence vary their measured signal to a different extent.

Type A biases are common among analyses that extract information about a non-directly measurable physical quantity through its proportional relationship with a measurable one. Type A biases are accounted for by the addition of a known quantity of internal standard to each biological sample before extraction. If only Type A biases occur, the change in the relative with respect to the internal standard concentration of a free metabolite between two biological samples is proportional to the change in the relative with respect to the internal standard marker ion peak area of the metabolite derivative. Type-B biases cannot, however, be corrected and/or accounted for by the use of the internal standard only. Depending on the source of these biases and how they affect the final outcome, specific data correction and/or experimental optimization methodologies have to be developed.

3.3. Standardizing the extraction step

The extraction of particular classes of molecules from a biological sample has for long been the subject of research, independently of metabolomics. Optimization of the free metabolite extraction protocol for a particular biological system requires for two constraints to be satisfied:

- The extracted free metabolite mixture has to correspond to the *in vivo* physiological conditions of the biological system at the time of sample acquisition. This requires for the extraction to be carried out under conditions at which there is no (or practically negligible) enzymatic activity.
- Uniform degree of extraction is achieved for all samples in a batch, which will be analyzed in comparison to each other.

"Freezing" the protein activity of the biological sample is a universal objective of research in life sciences. However, in the case of metabolomics is even more pressing, because the majority of the free metabolite pools have a fast turnover rate. The common practice is for the biological sample to be frozen in liquid nitrogen (or −20 °C ethanol) as quickly as possible and stored for subsequent analysis (for many applications [46,47], fast freezing equipment has been developed). If any further processing is required, e.g. separation of cell pellet from the supernatant, this is carried out while keeping the sample at low temperatures. The cellular samples are subsequently homogenized in the presence of methanol [2–8]; methanol freezes the enzymatic activity and any subsequent steps of the experimental protocol can be carried out at room temperature. For the extraction of free polar or nonpolar metabolites a methanol–water [2] or methanol–chloroform–water extraction protocol is, respectively, used [3,48].

Optimization of the extraction protocol with respect to the second constraint stated above requires mainly the adjustment of (a) the ratio solvent:biomass, and (b) the duration of the extraction. For each investigated biological system [49–51]. The utilized amount of solvent should allow for sufficient degree of extraction for all free metabolites, so that they can be detected in the metabolomic profile, without, however, increasing dramatically the drying time of the samples. Indicative of the need to optimize the solvent:biomass ratio for each biological system was an experiment carried out in our group. We compared the metabolomic profiles of the same *A. thaliana* liquid culture samples (see Section 2.1) after extraction with 14 mL and 28 mL methanol per gram of wet plant tissue. The 14 mL/g solvent:biomass ratio had been reported for the potato tuber tissue in [2]. In the case of the *A. thaliana* samples, the number of metabolite peaks that were detected in the metabolomic profile increased by 25% from the samples extracted with 14 mL/g to the samples extracted with 28 mL/g.

Other experimental variations associated with the extraction step, which could introduced biases to the acquired metabolomic profile, include (a) the varying (between samples) amount of supernatant recovery after centrifugation for the separation of macromolecules, and (b) the varying (between samples) extent of drying. These biases are of type A and can be accounted for through the use of internal standard(s).

3.4. Standardizing the derivatization step

As it was explained earlier, the mandatory derivatization step characterizes the GC–MS metabolomic analytical platform. As the source of both Type A, but mainly Type B errors, it is mainly this step that has so far limited the extensive application of the GC–MS metabolomics, due to lack of validated and reproducible datasets. The derivatization involves the reaction of all the metabolites in the dried metabolite mixture that has been extracted from the raw biological sample (including the internal standard(s)) with the added derivatization agent. Thus, the derivatization step corresponds to hundreds of non-spontaneous, parallel, competing reactions of organic molecules of various chemical classes with the same agent. Therefore, when optimizing an experimental protocol for the derivatization step, the first concern is the amount of the derivatizing agent that needs to be added to the system; it needs to be adequately "in excess" to allow for the complete derivatization of all the free metabolites in the extracted mixture. The second concern refers to the duration of the derivatization at the time the sample is run through the GC–MS equipment. The perfect scenario would be the identification of a duration for the derivatization step, either "universal" for all biological systems or optimized for each biological system separately, that is always longer than the time required for the completion of the derivatization reactions for all the metabolites in the extracted mixture, independent of the composition of this mixture. If the identification of such time is not possible or the particular time is practically long, the selection of the duration for the derivatization step in the optimized experimental protocol needs thorough consideration. If a time at which the derivatization reactions of many metabolites might not have completed yet is finally selected, one needs to validate that the acquired metabolomic profiles will be directly comparable among biological samples. To complicate the situation even further, there is no available "universal" derivatizing agent today that can lead to one derivative for every metabolite, independent of its chemical class. Finally, maybe the major challenge in the optimization of the derivatization protocol is that all these issues should be resolved in a way that does not jeopardize the high-throughput nature of GC–MS metabolomics.

To understand (a) the sources of biases that can be introduced to the metabolomic profiling data at the derivatization step depending on the selected parameters in the protocol, and (b) the proposed ways to either avoid the occurrence of errors by selecting different parameters, or to account for them in the final results, one needs to consider how the selection of a particular derivatization protocol affects the relationship between the following three quantities (see [4]):

1. The amount of a metabolite *M* in the dried metabolite mixture as this was extracted from the raw biological sample, C_M.
2. The amount of the metabolite's *M* derivative in the derivatized metabolite solution at the derivatization time that the solution is run through the gas chromatograph-mass spectrometer C_{MD}.
3. The marker ion peak area of the metabolite's *M* derivative in the measured metabolomic profile, PA_{MD}.

As it was stated in section 3.1, within the linear range of the equipment's operation the quantities 2 (C_{MD}) and 3 (PA_{MD}) are proportional with proportionality coefficient the response factor of the metabolite's M derivative at the detection and quantification conditions of the GC–MS equipment during the particular run (these will be referred as GC–MS equipment conditions in the rest of the text). Thus, for the same GC–MS equipment conditions, the observed change in the quantity 3 (ΔPA_{MD}) between two physiological states of a biological system is directly proportional with the occurred change in the quantity 2 (ΔC_{MD}). This implies that for biologically relevant conclusions to be extracted based on the acquired metabolomic profiles, i.e. the changes in the quantity 3 (ΔPA_{MD}), one needs to ensure that the changes in the quantities 1 (ΔC_M) and 2 (ΔC_{MD}) are directly proportional too. In the rest of this section, it will be demonstrated how the derivatization step could distort this proportionality between ΔC_M and ΔC_{MD} with Type B biases. In addition, ways to optimize the experimental protocol to avoid them or to enable the appropriate correction of the metabolomic profiles will be proposed in the context of the most commonly used derivatization method in GC–MS metabolomics, which converts the metabolites to their (methoxime)-trimethylsilyl [(MeO*x*)-TMS] derivatives (see Fig. 2). As shown in Fig. 2 (for additional details see [4]), this derivatization method involves two reactions:

(1) The methoximation reaction affects only the metabolites in the extracted mixture that contain ketone groups; in [4], these metabolites have been assigned the name of "category 2 metabolites".
(2) Specifically, these metabolites react to form two methoxime derivatives (MeO*x*1 and MeO*x*2) which are geometric isomers. Importantly, the amount ratio between these two methoxime derivatives is constant since the initiation of the methoximation and characterizes the particular metabolite.
(3) The reaction with the silylation agent (MSTFA); the metabolites with hydroxyl (–OH), carboxyl (–COOH) and amine ($-NH_2$) groups react, replacing the active hydrogen atoms with the silylic parts. Importantly, while the (–OH) and (–COOH) groups react simultaneously and (practically) very fast, the hydrogen atoms of the ($-NH_2$) groups react slower than in the other two groups, and also sequentially. Thus, these metabolites are to produce more than one derivatives that appear sequentially and the amount ratio of these derivatives is not expected to be constant while the derivatization reaction lasts. Theoretically, when the derivatization of a ($-NH_2$)-group containing metabolite completes, only the last formed derivative will be present (see Fig. 2). In [4], the metabolites that contain $-NH_2$ groups have been assigned the name "category 3" metabolites, while those containing only (–OH) or (–COOH) groups the "category 1". Category 3 comprises also a small group of metabolites containing (–SH) and (–NH) heterocycle groups. Little [52] provides a comprehensive overview of silylation artifacts and ways to avoid them. Many of these artifacts are currently avoided in the GC–MS metabolomic analysis with the use of BSTFA/MSTFA as silylation agent and the pre-silylation with the methoxyamine-HCl.

3.4.1. Optimizing the duration of the derivatization step

Based on the earlier discussion, determining the duration of the derivatization in the experimental protocol is of great significance for quantitative metabolomics. In the currently available protocols for different biological systems, the duration of the methoximation and silylation reactions varies significantly [2–4,6,23,42,46,47]. It is apparent that allowing for complete methoximation and silylation reaction for all metabolites in every biological sample of an

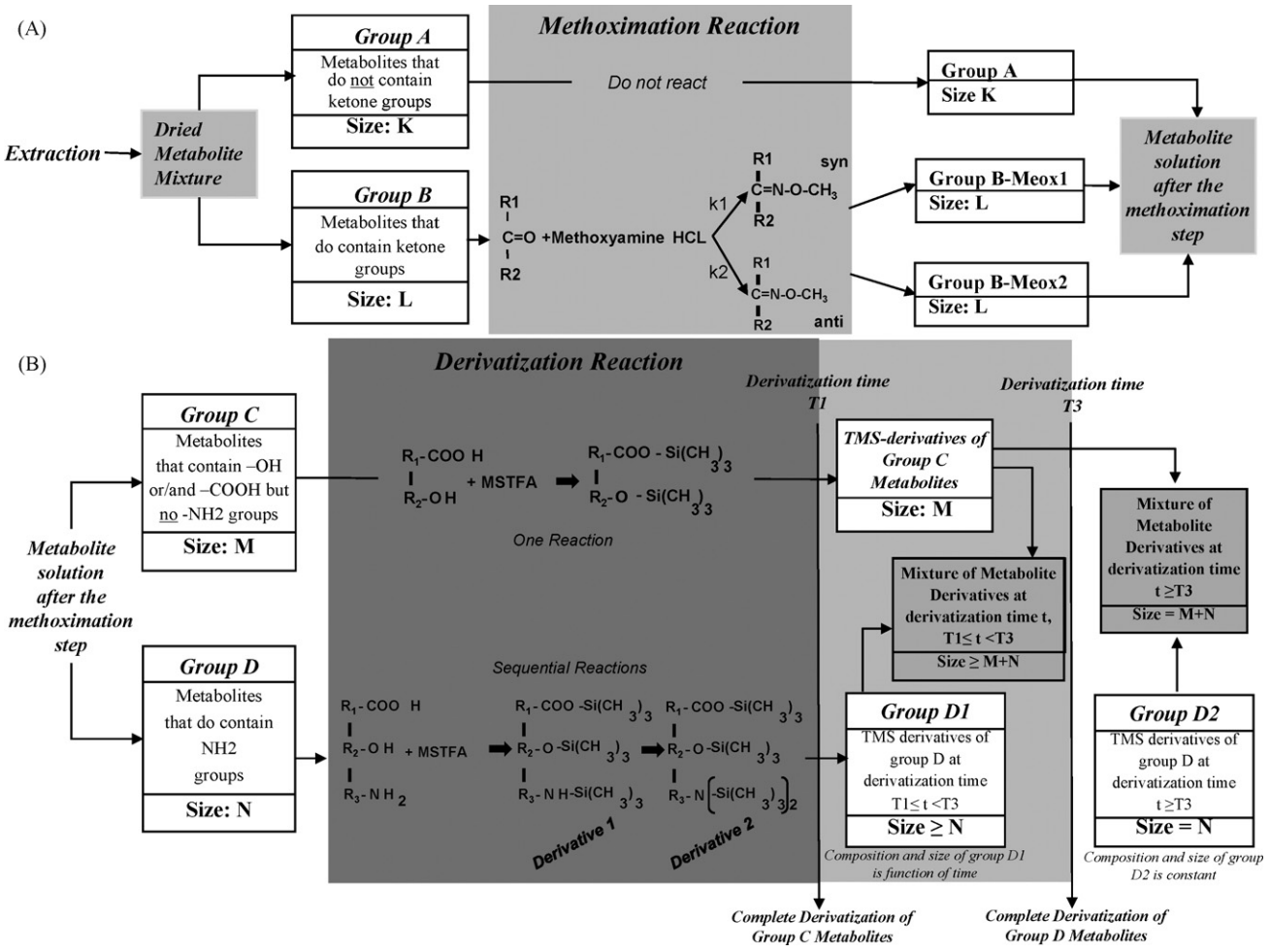

Fig. 2. Schematic diagram of the derivatization process towards the production of the (methoxime)-trimethylsilyl derivatives of the free metabolites in the extracted from the biological sample mixture.

experimental set would guarantee comparability among the samples. In this case, Type B biases that might be present at shorter derivatization times, as it will be shown later in the section, are avoided. According to the classification of the metabolites that was discussed earlier, at times longer than the full completion of the derivatization for all metabolites, the derivative solution will contain one derivative for the category 1 and the category 3 with no ketone groups metabolites, and two for the category 2 metabolites, independently of them containing ($-NH_2$) groups or not. The amount ratio of the two derivatives for each category 2 metabolite is constant and the amount of each is directly proportional to the amount of the category 2 metabolite in the extracted mixture. Thus, at these derivatization times, the metabolomic profile comprising the marker ion peak areas of the category 1 and 3 metabolites, and one of the two marker ion peak areas of the category 2, is directly proportional to the amount profile of the extracted metabolite mixture.

In an effort to identify this derivatization time for *A. thaliana* liquid culture polar extracts and decide whether it is practically convenient for research purposes, we acquired the metabolomic profile of the same sample at different derivatization times ranging from 30 min to 450 min after addition of the silylating agent (MSFTA). Fig. 3 shows the silylation time profile of the marker ion peak area of the derivatives of selected metabolites normalized with the same peak area at 30 min of the silylation. Firstly, it is clear from Fig. 3 that the silylation reactions of the various metabolites are not spontaneous, including the internal standard (ribitol). In the case of this biological system, for the category 1 and 2 metabolites the plateau in the marker ion peak area of their derivatives (indicating completion of the silylation) is expected to have been reached by 6 h after the addition of MSTFA (additional data of our laboratory (not shown) further support this conclusion). However, the derivative peak areas of certain category 3 metabolites continue varying after this derivatization time. In addition, their time profile indicates the sequential production of multiple metabolite derivatives; when the peak area of the first to appear derivative decreases, the peak area of the subsequent increases. Fig. 3 supports the schematic representation of the derivatization process provided in Fig. 2. This is even clearer in Fig. 4 that shows the time profiles of the marker ion peak areas of selected category 1, 2 and 3 metabolites of another *A. thaliana* liquid culture sample, as these were measured in metabolomic profiles acquired at silylation times ranging from 5 h to 29 h. According to Fig. 3, at 5 h of silylation, the derivatization of both category 1 and 2 metabolites is expected to have completed. Indeed, the acquired profiles indicate constant derivative peak areas for the category 1 and 2 metabolites throughout the particular time period (Fig. 5). However, the silylation reaction of the category 3 metabolites continues even after a day of silylation. These observations indicate that the duration of the silylation to be selected in the experimental protocol cannot correspond to complete silylation for all metabolites. 30 h or longer silylation times are not practically convenient. Moreover,

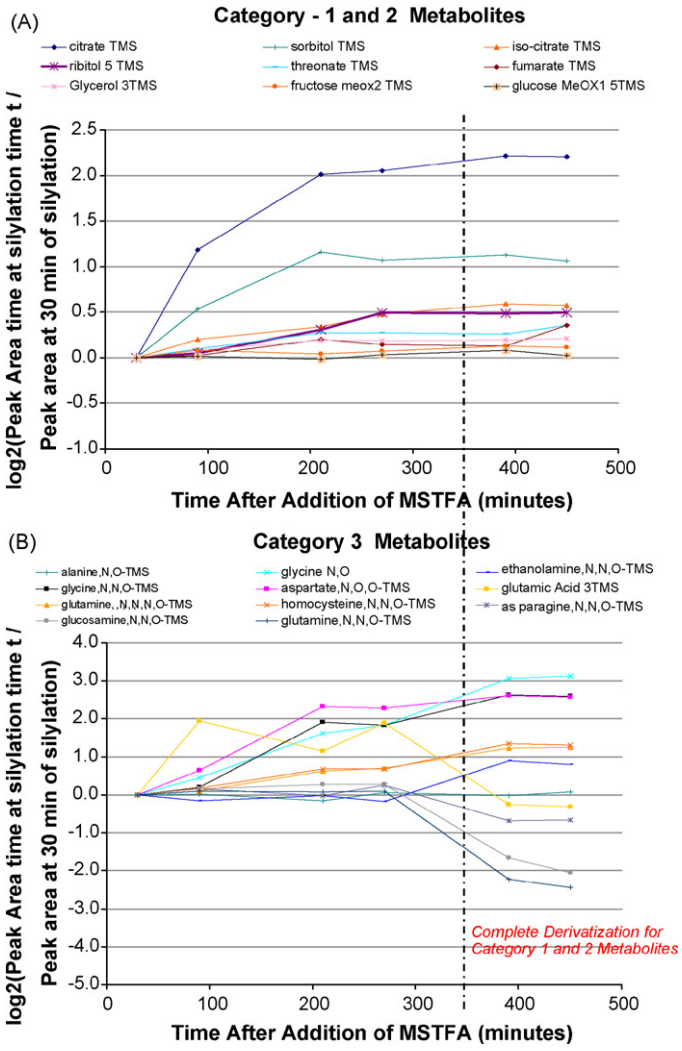

Fig. 3. The profile over MSTFA silylation time (between 30 min and 450 min of silylation) of the derivative marker ion peak area of (A) category 1 and category 2, and (B) category 3 metabolites in a 12-days old *A. thaliana* liquid culture normalized with the same peak area at 30 min of silylation. The first two metabolomic profiles (at 30 min and 90 min of derivatization) were acquired at split ratio 25:1, the next two (at 210 min and 270 min) at 35:1 and the last two (390 min and 450 min) at 50:1. The 25:1 peak areas for each metabolite were used to plot the graphs; the metabolite peak areas of split ratio 35:1 and 50:1 were, respectively, multiplied by 35/25 and 50/25. The silylation time profile of the category 1 and 2 peak areas, including the internal standard ribitol, indicates that the derivatization of these metabolites may continue considerably beyond the 90 min. 6 h of silylation are indicated as the time at which the derivatization reaction of all category-1 and category 2 metabolites has completed and at least one derivative of category 3 metabolites has been formed. Clearly, the silylation of category 3 metabolites continues beyond this derivatization time and the formation of multiple derivatives is apparent for the metabolites whose more than one derivative peak areas were monitored.

they may coincide with degradation of certain metabolite derivatives. According to this discussion, the duration of silylation for the *A. thaliana* liquid culture samples of our project could be determined at 6 h after the addition of MSTFA, ensuring at least that the silylation of the category 1 and 2 metabolites will be complete and category 3 metabolites will have been converted to at least one of their derivative forms. At these silylation times, the bioinformatics analysis among samples could be based on the marker ion peak areas of known and unknown category 1 metabolite derivatives and one of the two peak areas of the known category 2 derivatives. Peak areas of unknown category 2 derivatives could still be used in the multivariate statistical analysis. However, one needs to keep in mind that inclusion of both peak areas of a category 2 metabolite introduces bias to the analysis as they are not independent; the larger the number of the unknown category 2 metabolites, the stronger the bias. However, at these silylation times, the peak areas of the category 3 metabolites' derivatives cannot be used as such to extract quantitative conclusions for the physiological state of the investigated systems. Their inclusion in the bioinformatics analysis is expected to significantly distort the final results, because the observed changes in their peak area profiles among samples might be due to experimental and not to biological reasons.

To avoid time-dependent derivatization biases, the use of automated in-line derivatization has been proposed (see e.g. [53]); these devices allow for samples to run at the same derivatization time. There is a perception in the GC–MS metabolomic community that the same quantitative biological results can be extracted even at silylation times shorter than the time at which the silylation reaction is complete, as long as the duration of the silylation is the same among the compared biological samples. Figs. 5 and 6 demonstrate that this is not the case due to the, so-called, "matrix effects". For the current perception to be correct, the kinetics of silylation of each metabolite should be the same among biological samples independent of the composition of their free metabolite pool. In other words, the ratio of the amount of a metabolite's derivative at silylation time *t* to the amount of the metabolite derivative at the

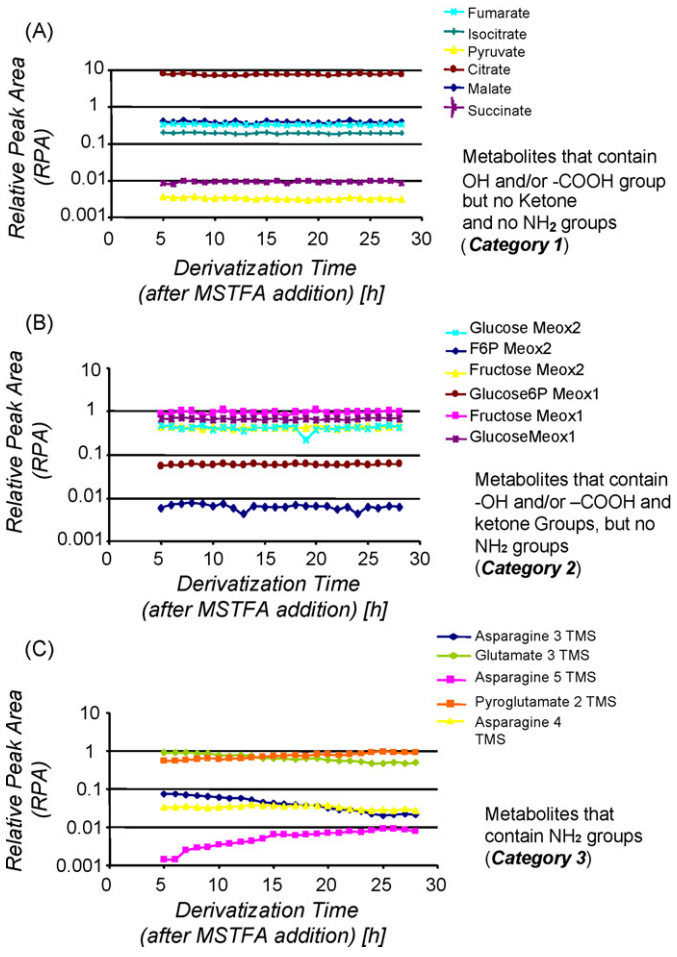

Fig. 4. The time profile between 5 h and 28 h of MSTFA silylation of the relative with respect to the internal standard (ribitol) marker ion peak area of the derivatives of selected (A) category 1, (B) category 2, and (C) category 3 metabolites in a single *A. thaliana* liquid culture polar extract sample.

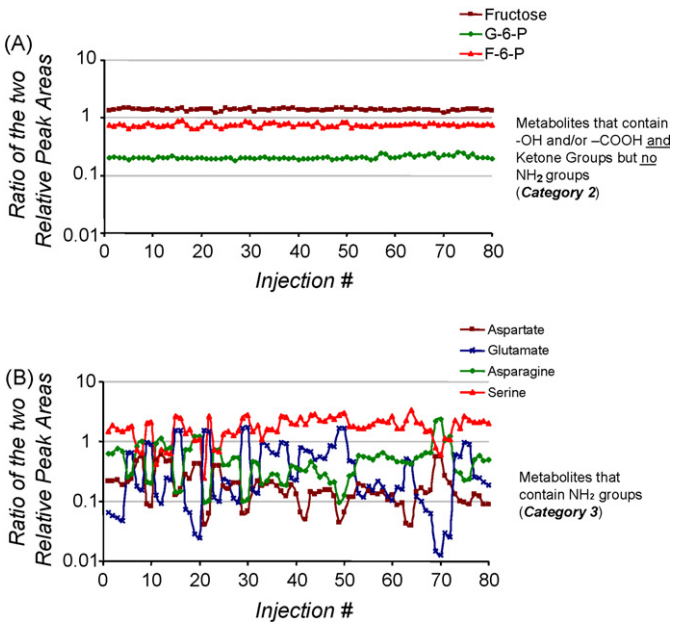

Fig. 5. The ratio of the two derivative peak areas of (A) category 2, and (B) category 3 metabolites in 80 *A. thaliana* plant liquid culture metabolomic profiles acquired at the same GC–MS operational conditions. The profiles correspond to biological samples of 40 different physiological conditions run in two replicates (the replicate metabolomic profiles are shown in sequence). The replicates of the same sample were run with an hour of silylation time difference. The biological samples were derivatized and run in four batches (10 samples each × 2 replicates per sample); each batch was run at silylation times between 6 h and 30 h. The ratio of the peak areas of the category 2 metabolite derivative forms remains constant, while there is a clear variation in the corresponding ratio of the category 3 forms. In this set, the metabolomic profiles are subject of differences in silylation time at which each sample was run and matrix effects.

completion of the silylation should be the same function of time among biological samples, independent of the composition of their free metabolite pool. In this case, if the marker ion peak area of a particular metabolite derivative were measured in all samples at a certain silylation time t shorter than the time of the silylation completion, then, in all samples, it would be equal to the same fraction of the marker ion peak area that this metabolite derivative will finally reach at the plateau of the silylation. This condition, however, is not expected to be true. As stated earlier, the silylation reactions of all metabolites in a biological sample compete for the same silylating agent. Thus, the kinetics of the individual reactions are expected to depend on the composition of the metabolite mixture at the initiation of the silylation. Figs. 5 and 6 indicate the impact of the

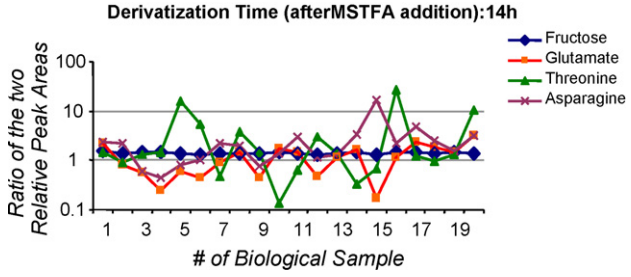

Fig. 6. The ratio of the two derivative peak areas of four metabolites in the metabolomic profiles of 20 *A. thaliana* liquid culture samples that were acquired at the same silylation time, i.e. 14 h after addition of MSTFA. The biological samples correspond to different biological conditions. The figure shows that there is a big variation in this ratio for category 3 metabolites even though all samples were measured at the same derivatization time.

matrix effects in the context of category 3 metabolites at silylation times at which the silylation of category 1 and 2 metabolites has completed.

The use of chemically synthesized fully labeled metabolite standards or extracts from organisms that have been grown with fully labeled carbon source(s) has been proposed as an approach to address the problem of derivatization limitations in GC–MS metabolomics [54–56]. The fully labeled extracts/standards are added to the target biological sample to serve as an internal reference sample for the normalization of the acquired data. Following similar strategy to quantification methods in other "omics", this approach could address derivatization issues in GC–MS metabolomics. However, (a) the cost to prepare this reference sample is high, (b) including this reference sample in the metabolomic process increases the complexity of peak area deconvolution in the metabolomic profiles, while it does not address the problem of multiple derivative peaks for a metabolite, and (c) it is difficult for this approach to be applied to biological systems that cannot easily grow *in vitro*. In the next section, we describe a GC–MS metabolomic data correction and normalization strategy, which does not jeopardize the high-throughput nature of the analysis and does not require the use of fully labeled reference.

3.4.2. *Correcting the metabolomic profile for the multiple derivatives*

In the previous section, we indicated as optimal duration of silylation in the experimental protocol the time at which the silylation reactions of all category 1 and 2 metabolites are expected to be complete. We also demonstrated that at this, and even up to 30 h of, silylation time, the silylation of many category 3 metabolites is not expected to have been completed. Thereby, the inclusion of their derivative peak areas in the quantitative analysis is going to negatively affect the final results. However, category 3 comprises metabolites that are of great biological significance for many life sciences applications, including amino acids, which have been widely used as markers of physiological activity. Thus, it is of high interest for the quantitative profile of the category 3 metabolites to be accurately determined in the context of GC–MS metabolomics. A solution, originated in the mass spectrometrists' works of the 60–70s [57,58], proposes the use of different derivatizing agents for every class of molecules in the biological sample to ensure one derivative per metabolite. This solution, however, is in contradiction with the high-throughput nature of the metabolomics platform that dictates the simultaneous measurement of all metabolites in one run.

Kanani and Klapa [4] have proposed a methodology for quantitative metabolomics of all three metabolite categories. This methodology allows for the weighted summation of the marker ion peak areas of all derivatives of a category 3 metabolite to one effective peak area that is directly proportional to the amount of the metabolite in the extracted metabolite mixture. Fig. 7 shows the results of this methodology's application on the metabolomic profiles of category 3 metabolites in *A. thaliana* liquid culture samples, which were acquired at silylation times equal to or longer than 6 h of silylation, As determined earlier, at this time the silylation for the category 1 and 2 metabolites should have completed. As expected, the effective peak area of the category 3 metabolites was constant throughout the measured silylation period. Table 2 shows in quantitative terms the improvement in the variation in the derivative peak areas of category 3 metabolites among injections of the same bio-replicate and among biological replicates of the same sample after application of the particular methodology.

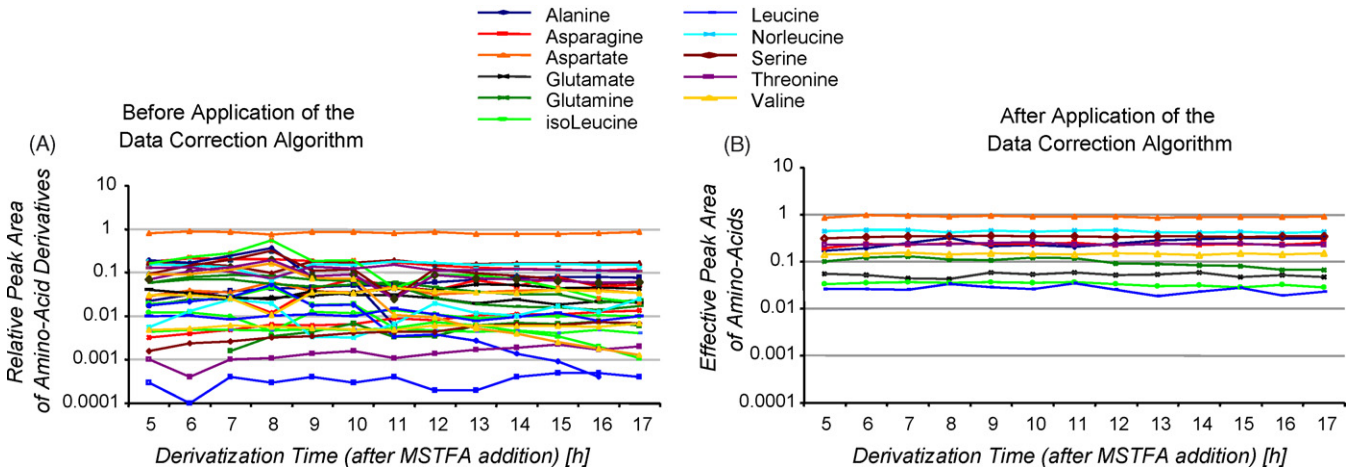

Fig. 7. The relative with respect to the internal standard (ribitol) (A) measured marker ion peak areas of the derivatives, and (B) estimated effective peak areas (based on the algorithm described in [4]), of 10 category 3 metabolites in a single *A. thaliana* liquid culture sample run at different derivatization times. The effective and derivative peak areas of a metabolite are shown in the same color. As expected, the effective peak areas of the metabolites, which are proportional to the concentration of the metabolite in the biological sample, do not vary with the MSTFA silylation time (within the margin of experimental error).

In conclusion, barring changes in the GC–MS equipment conditions (as it will be discussed in the following section), (a) running the biological samples at silylation times equal to or longer than the identified for a particular biological system time. At which the silylation of the category 1 and 2 metabolites is expected to have completed, and (b) using the methodology described in [4] for the estimation of the effective peak area of category 3 metabolites, ensures direct proportionality between the amount of each metabolite in the extracted metabolite mixture and the measured marker ion peak area of the metabolite derivative(s), for the category 1 and 2 metabolites, or the effective peak area, for the category 3 metabolites. In the case of the unknown category 3 metabolite peak areas that cannot be combined into one effective peak area through the use of the available data correction methodology, it is still valid that their inclusion in the bioinformatics analysis will distort the quantitative results. Hence, these peaks should not be considered in the quantitative analysis (see Table 2 and [4]). As the identification of unknown peaks progresses, the number of category 3 metabolite peaks to be included in the quantitative analysis will be gradually increasing, enhancing thus the resolution of the observable metabolic fingerprint.

3.5. Standardizing the GC–MS data acquisition process

Since the GC–MS equipment has been used for the analysis of chemical compounds for quite a long time, the GC–MS data acquisition procedure has been fairly standardized. In this paper, we will focus on three issues that need to be addressed to ensure quantitative GC–MS metabolomics. If not addressed, they can introduce Type B biases that cannot be corrected through the use of an internal standard. Specifically, quantitative GC–MS metabolomics requires that (a) the metabolomic profiles are acquired within the linear range of the GC–MS equipment's operation, (b) the metabolomic profiles of the biological samples that are to be compared to extract biologically relevant conclusions need to have been acquired at the same GC–MS operating conditions, and (c) the metabolomic profiles need to be filtered from chromatography artifacts and adducts.

3.5.1. Ensuring linear range of GC–MS operation

In a biological system, the concentrations of the various free metabolite pools can differ by three to four orders of magnitude. Thus, the simultaneous measurement of all free metabolites' concentration through GC–MS metabolomic analysis requires that all (or at least most) of these concentrations fall within the linear

Table 2
Positive impact of the application of the data correction methodology for category 3 metabolites that is described in [4] on the acquired peak area profiles shown in Figs. 5 and 7

	Before correction	After correction	% Reduction
Derivatization time effects (Fig. 7)			
# Known category 3 peaks	32	12	63
Average over all category 3 metabolite derivative peaks coefficient of variation (CoV %) of the relative with respect to the internal standard peak area between 12 injections of the same sample (acquired at different derivatization times as shown in Fig. 7)	44	10	77
Derivatization time and Matrix effects (Fig. 5)			
# Known category 3 peaks	45	18	60
Average over all category 3 metabolite derivative peaks and over 40 samples CoV (%) of the relative with respect to the internal standard peak area between replicate injections of the same biological sample	12	6	50
Average over all category 3 metabolite derivative peaks and over 18 physiological states CoV (%) of the relative with respect to the internal standard peak area between injections of biological replicates (i.e. samples that represent the same physiological state)	40	26	35
# Unknown category 3 metabolite derivatives used in the analysis	42	0	100

There is a significant decrease in the category 3 metabolites' peak area variation between injections of the same biological sample at different derivatization times and between injections of biological samples representing the same physiological state at different states (the latter are subject to matrix effects). To quantify the peak area variation before the application of the data correction methodology, the peak areas of all observed derivatives of the category 3 metabolites (known and unknown) were used. After the application of the methodology, only the estimated effective peak areas of the known category 3 metabolites were used.
% Reduction = ("after correction" value − "before correction" value)/("before correction" value) × 100%.

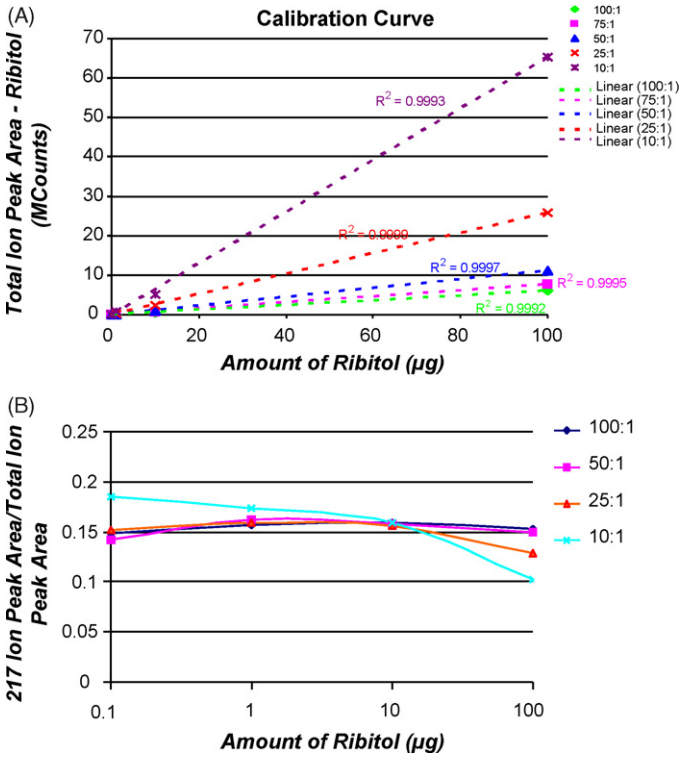

Fig. 8. (A) Calibration curves of the internal standard (ribitol) at different split ratios; (B) ratio of the 217 marker ion peak area to the total peak area of ribitol in the profiles used for the calibration curves in (A). Interestingly, both 10:1 and 25:1 split ratios could be considered valid, if the total ribitol peak area is used for the identification of the linear range of the equipment's operation. However, calibration with the marker ion peak area indicates the need to use larger than 10:1 split ratios to ensure linear range of operation and avoid saturation biases.

range of operation of the equipment. To determine the minimum and maximum order of magnitude of a metabolite's concentration that falls within the linear range of operation of the equipment, a number of concentrations of the internal standard, metabolite standards, but also derivatized biological samples are initially run through the equipment at different split ratios. The smallest split ratio that avoids saturation effects for the metabolites that are expected in highest amounts in the biological sample is selected as the optimal split ratio. It needs to be underlined, however, that the selection of the optimal split ratio should be based on calibration curves for these metabolites that are made based on the metabolite marker ion peak areas. An indicative example is shown in Fig. 8. Specifically, based on the calibration curve of the total peak area of the internal standard ribitol, Fig. 8A points out 1:10 as the optimal split ratio. However, Fig. 8B indicates that at the particular split ratio saturation effects influence the peak area of the ribitol marker ion 217, pointing thus out 1:25 as the optimal split ratio. If the lowest metabolite concentration that can be detected at this split ratio excludes a number of significant free metabolites in the biological sample, the same sample may be run at two split ratios to accurately quantify each end of the range of metabolite concentrations in the biological sample, and/or, if possible, the amount of the raw material that is used in the analysis should be increased.

3.5.2. Ensuring constant mass spectrometer operational conditions among samples

As it was discussed in the introductory part of Section 3.4, quantitative GC–MS metabolomics is based on the proportional relationship between the measured change in the marker ion peak areas of a metabolite derivative (ΔPA_{MD}) between two physiological states of a biological system and the occurred change in the metabolite derivative's concentration in the derivatized metabolite solution (ΔC_{MD}). The proportionality coefficient that connects the two quantities is the response factor (RF) of the metabolite derivative at the particular conditions of the mass spectrometer. Change of GC–MS equipment or common variations in the mass spectrometer conditions of the same equipment are expected to affect the RFs of all metabolites, including the internal standard, by the same extent, corresponding thus to Type A biases. However, there are changes in the mass spectrometer conditions that can give rise to Type B biases by affecting the RF of individual metabolites to a different extent. These changes include variations in the ionization assembly, the electron multiplier or the alignment of the mass spectrometer throughout the various runs. They can significantly modify the fragmentation pattern of certain metabolite derivative forms. In addition, the methodology that has been proposed [4] for the estimation of the effective peak area of category 3 metabolites can only apply to metabolomic profiles that were acquired at the same equipment's conditions, i.e. in the absence of these critical variations. Because it is difficult to quantitatively correct the metabolomic profiles from the biases introduced from these sources of error, the only way to avoid for them to affect the final results is (a) to regularly monitor the conditions of the mass spectrometer operability, and (b) in the case that these operational variations have indeed occurred, to filter into the quantitative analysis only the samples that were acquired at the same operational conditions. Kanani and Klapa in [4] introduced a criterion that enables the researchers to identify whether all profiles of a batch were acquired at the same mass spectrometer operational conditions. Specifically, if the latter had remained the same throughout all sample runs, then the ratio of the marker ion peak areas of the two derivatives of any category 2 metabolite would have remained constant within the margin of the experimental error. The profiles that correspond to different ratios should be filtered out of the analysis. Fig. 9 shows a set of metabolomic profiles (see Supplementary Table S1), which, based on the above criterion, corresponds to large variations of the mass spectrometer operational conditions. On the other hand, the ratio of the same category 2 metabolite in the samples of Figs. 3–5 remains constant.

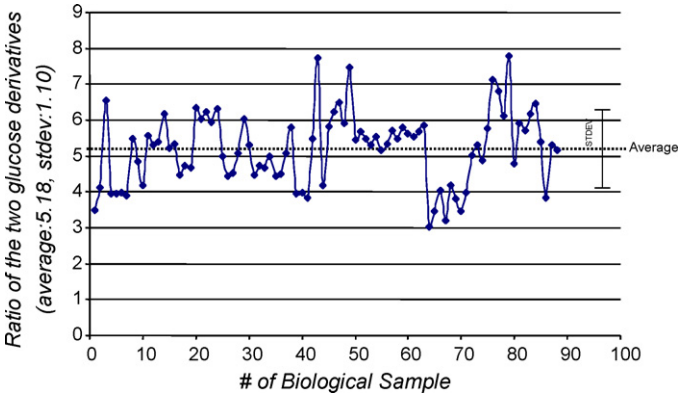

Fig. 9. The ratio of the marker ion peak areas of the two derivatives of the internal standard [^{13}U-C] glucose in the metabolomic profiles of brain tissue samples. The samples were run on the same GC–MS equipment in batches; each batch was run on different days (differing by a day to two months); the biological samples run on the same day were injected multiple times at different MSTFA silylation times (see Supplementary Table S1).

3.5.3. Chromatographic variations

A typical issue in the GC–MS chemical analysis is the drift in the metabolite retention times between metabolomic profiles acquired mainly from different equipments, but from the same equipment too. This drift is due to variations in the gas chromatograph operational conditions. Retention time indices have been traditionally used to align the metabolomic profiles from the chromatographic drift. Publicly available [59,60] and commercial peak alignment software tools have been developed (for comprehensive review, see http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/Peak_Alignment/). Further research to enhance the peak alignment algorithms towards a systematic and automated peak identification and quantification is under way. Taking into consideration that the peak identification is the "bottleneck" in the speed of the GC–MS metabolomic analysis, these algorithms are needed to ensure extensive use of GC–MS metabolomics in large-scale applications in the near future.

Acknowledgments

We would like to gratefully acknowledge the financial support of (a) the US NSF (QSB-0331312), (b) the UMD Minta-Martin Foundation, (c) the UMD Chemical and Biomolecular Engineering Department, (d) the FORTH/ICE-HT, and (e) the Nano2Life "Bringing Nanotechnologies to Life" FP6-funded European Network of Excellence. In addition, we would like to thank Dr. M. Margarity, Dept. of Biology, U. Patras, for providing the brain tissues and Mr. Nicholas Linn, then NSF-funded REU student at the UMD Chemical and Biomolecular Engineering Department, for running the GC–MS calibration samples.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2008.04.049.

References

[1] M.I. Klapa, J. Quackenbush, Biotechnol. Bioeng. 84 (2003) 739.
[2] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Nat. Biotechnol. 18 (2000) 1157.
[3] U. Roessner, C. Wagner, J. Kopka, R. Trethewey, L. Willmitzer, Plant J. 23 (2000) 131.
[4] H. Kanani, M.I. Klapa, Metab. Eng. 9 (2007) 39.
[5] T. Tohge, Y. Nishiyama, M.Y. Hirai, M. Yano, J. Nakajima, M. Awazuhara, E. Inoue, H. Takahashi, D.B. Goodenowe, M. Kitayama, M. Noji, M. Yamazaki, K. Saito, Plant J. 42 (2005) 218.
[6] W. Weckwerth, M.E. Loureiro, K. Wenzel, O. Fiehn, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 7809.
[7] M.Y. Hirai, M. Klein, Y. Fujikawa, M. Yano, D.B. Goodenowe, Y. Yamazaki, S. Kanaya, Y. Nakamura, M. Kitayama, H. Suzuki, N. Sakurai, D. Shibata, J. Tokuhisa, M. Reichelt, J. Gershenzon, J. Papenbrock, K. Saito, J. Biol. Chem. 280 (2005) 25590.
[8] T.W. Fan, R.M. Higashi, A.N. Lane, Drug. Metab. Rev. 38 (2006) 707.
[9] R. Steuer, J. Kurths, O. Fiehn, W. Weckwerth, Bioinformatics 19 (2003) 1019.
[10] Y. Nakamura, A. Kimura, H. Saga, A. Oikawa, Y. Shinbo, K. Kai, N. Sakurai, N. Suzuki, M. Kitayama, D. Shibata, S. Kanaya, D. Ohta, Planta 227 (2007) 57.
[11] K. Narasimhan, C. Basheer, V.B. Bajic, S. Swarup, Plant Physiol. 132 (2003) 146.
[12] D. Morvan, A. Demidem, Cancer Res. 67 (2007) 2150.
[13] A. Di Leo, W. Claudino, D. Colangiuli, S. Bessi, M. Pestrin, L. Biganzoli, Ann. Oncol. 12 (2007) (Suppl. xii8).
[14] W.M. Claudino, A. Quattrone, L. Biganzoli, M. Pestrin, I. Bertini, A. Di Leo, J. Clin. Oncol. 25 (2007) 2840.
[15] H.C. Keun, J.J. Athersuch, Pharmacogenomics 8 (2007) 731.
[16] J.C. Lindon, E. Holmes, J.K. Nicholson, FEBS J. 274 (2007) 1140.
[17] W. Tang, Expert Opin. Drug. Metab. Toxicol. (2007) 407.
[18] C.J. Borgert, Toxicol. Appl. Pharmacol. 223 (2007) 114.
[19] C.U. Niemann, N.J. Serkova, Expert Opin. Drug Metab. Toxicol. 3 (2007) 527.
[20] K.M. Oksman-Caldentey, K. Saito, Curr. Opin. Biotechnol. 16 (2005) 174.
[21] F. Carrari, E. Urbanczyk-Wochniak, L. Willmitzer, A.R. Fernie, Metab. Eng. 5 (2003) 191.
[22] S. Vij, A.K. Tyagi, Plant Biotechnol. J. 5 (2007) 361.
[23] D. Cook, S. Fowler, O. Fiehn, M.F. Thomashow, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 10583.
[24] W. Pongsuwan, E. Fukusaki, T. Bamba, T. Yonetani, T. Yamahara, A. Kobayashi, J. Agric. Food Chem. 55 (2007) 231.
[25] S.A. Overy, H.J. Walker, S. Malone, T.P. Howard, C.J. Baxter, L.J. Sweetlove, S.A. Hill, W.P. Quick, J. Exp. Bot. 56 (2005) 287.
[26] N. Angelova, H.W. Kong, R. van der Heijden, S.Y. Yang, Y.H. Choi, H.K. Kim, M. Wang, T. Hankemeier, J. van der Greef, G. Xu, R. Verpoorte, Phytochem. Anal. 19 (2008) 2.
[27] 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.
[28] H.A. Kuiper, E.J. Kok, K.H. Engel, Curr. Opin. Biotechnol. 14 (2003) 238.
[29] H. Rischer, K.M. Oksman-Caldentey, Trends Biotechnol. 24 (2006) 102.
[30] J.B. German, M.A. Roberts, L. Fay, S.M. Watkins, J. Nutr. 132 (2002) 2486.
[31] R.D. Hall, I.D. Brouwer, M.A. Fitzgerald, Physiol. Plant. 132 (2008) 162.
[32] Y.S. Kim, P. Maruvada, J.A. Milner, Future Oncol. 4 (2008) 93.
[33] L.C. Kenny, et al., Metabolomics 1 (2005) 227.
[34] K.K. Jain, Curr. Opin. Mol. Ther. 9 (2007) 563.
[35] B. van Ravenzwaay, G.C. Cunha, E. Leibold, R. Looser, W. Mellert, A. Prokoudine, T. Walk, J. Wiemer, Toxicol. Lett. 172 (2007) 21.
[36] J. van der Greef, T. Hankemeier, R.N. McBurney, Pharmacogenomics 7 (2006) 1087.
[37] S.J. Park, S.Y. Lee, J. Cho, T.Y. Kim, J.W. Lee, J.H. Park, M.J. Han, Appl. Microbiol. Biotechnol. 68 (2005) 567.
[38] A. Kern, E. Tilley, I.S. Hunter, M. Legisa, A. Glieder, J. Biotechnol. 129 (2007) 6.
[39] S.H. Khoo, M. Al-Rubeai, Biotechnol. Appl. Biochem. (2007) (published online).
[40] J. Kopka, J. Biotechnol. 124 (2006) 312.
[41] J. Kopka, A. Fernie, W. Weckwerth, Y. Gibon, M. Stitt, Genome Biol. 5 (2004) 109.
[42] G. Noctor, G. Bergot, C. Mauve1, D. Thominet, C. Lelarge-Trouverie, J. Prioul Metabolomics, 3 (2007) 161.
[43] D.C. Boyes, A.M. Zayed, R. Ascenzi, A.J. McCaskill, N.E. Hoffman, K.R. Davis, J. Gorlach, Plant Cell 13 (2001) 1499.
[44] O.L. Gamborg, T. Murashige, T.A. Thorpe, I.K. Vasil, In Vitro 12 (1976) 473.
[45] L.A. Colón, L.J. Baird, in: R.L. Grob, E.F. Barry (Eds.), Modern Practice of Gas Chromatography, 4th edition, John Wiley & Sons Inc., Hoboken, New Jersey, USA, 2004, p. 277.
[46] A. Buchholz, J. Hurlebaus, C. Wandrey, R. Takors, Biomol. Eng. 19 (2002) 5.
[47] M.R. Mashego, K. Rumbold, M.M. De Mey, E. Vandamme, W. Soetaert, J.J. Heijnen, Biotechnol. Lett. 29 (2007) 1.
[48] W. Weckwerth, K. Wenzel, O. Fiehn, Proteomics 4 (2004) 78.
[49] J. Gullberg, P. Jonsson, A. Nordström, M. Sjöström, T. Moritz, Anal. Biochem. 331 (2004) 283.
[50] E.J. Want, G. O'Maille, C.A. Smith, T.R. Brandon, W. Uritboonthai, C. Qin, S.A. Trauger, G. Siuzdak, Anal. Chem. 78 (2006) 743.
[51] H. Wu, A.D. Southam, A. Hines, M.R. Viant, Anal. Biochem. 372 (2008) 204.
[52] J.L. Little, J. Chromatogr. A 844 (1999) 1.
[53] A. Erban, N. Schauer, A.R. Fernie, J. Kopka, in: W. Weckwerth (Ed.), Metabolomics: Methods and Protocols, Humana Press, Totowa, 2007, p. 19.
[54] M.R. Mashego, L. Wu, J.C. van Dam, C. Ras, J.L. Vinke, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, Biotechnol. Bioeng. 85 (2004) 620.
[55] C. Birkemeyer, A. Luedemann, C. Wagner, A. Erban, J. Kopka, Trends Biotechnol. 23 (2005) 28.
[56] L. Wu, M.R. Mashego, J.C. van Dam, A.M. Porell, J.L. Vinke, C. Ras, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, Anal. Biochem. 336 (2005) 164.
[57] R.P. Evershed, in: K. Blau, J. Halket (Eds.), Handbook of Derivatives for Chromatography, 2nd edition, John Wiley & Sons Ltd, West Sussex, England, 1993, p. 51.
[58] C.F. Poole, in: K. Blau, G. King (Eds.), Handbook of Derivatives for Chromatography, Heydon & Son Inc., Philadelphia, USA, 1978, p. 152.
[59] A.L. Duran, J. Yang, L. Wang, L.W. Sumner, Bioinformatics 19 (2003) 2283.
[60] Y. Tikunov, A. Lommen, C.H. de Vos, H.A. Verhoeven, R.J. Bino, R.D. Hall, A.G. Bovy, Plant Physiol. 139 (2005) 1125.