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.