COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 130 (2016) 135–141

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Automated resolution of chromatographic signals by independent component analysis–orthogonal signal deconvolution in comprehensive gas chromatography/mass spectrometry-based metabolomics

Xavier Domingo-Almenara[a,c,*], *Alexandre Perera*[b,d], *Noelia Ramírez*[a,c], *Jesus Brezmes*[a,c]

[a] *Metabolomics Platform – IISPV, Department of Electrical and Automation Engineering (DEEEA), Universitat Rovira i Virgili, Tarragona, Catalonia, Spain*

[b] *B2SLAB, Department d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Catalonia, Spain*

[c] *Biomedical Research Networking Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain*

[d] *Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Madrid, Spain*

ARTICLE INFO

Article history:
Received 24 July 2015
Received in revised form
1 March 2016
Accepted 2 March 2016

Keywords:
Comprehensive gas chromatography
Orthogonal signal deconvolution
Multivariate curve resolution
Compound deconvolution
Independent component analysis

ABSTRACT

Comprehensive gas chromatography–mass spectrometry (GC×GC–MS) provides a different perspective in metabolomics profiling of samples. However, algorithms for GC×GC–MS data processing are needed in order to automatically process the data and extract the purest information about the compounds appearing in complex biological samples. This study shows the capability of independent component analysis–orthogonal signal deconvolution (ICA–OSD), an algorithm based on blind source separation and distributed in an R package called *osd*, to extract the spectra of the compounds appearing in GC×GC–MS chromatograms in an automated manner. We studied the performance of ICA–OSD by the quantification of 38 metabolites through a set of 20 Jurkat cell samples analyzed by GC×GC–MS. The quantification by ICA–OSD was compared with a supervised quantification by selective ions, and most of the R^2 coefficients of determination were in good agreement (R^2>0.90) while up to 24 cases exhibited an excellent linear relation (R^2>0.95). We concluded that ICA–OSD can be used to resolve co-eluted compounds in GC×GC–MS.

* *Corresponding author at*: Metabolomics Platform – IISPV, Department of Electrical and Automation Engineering (DEEEA), Universitat Rovira i Virgili, Tarragona, Catalonia, Spain. Tel.: +34 977559619; fax: +34 977559605.
E-mail address: xavier.domingo@urv.cat (X. Domingo-Almenara).

http://dx.doi.org/10.1016/j.cmpb.2016.03.007
0169-2607/

1. Introduction

Metabolomics is the study of low molecular weight compounds in biological systems [1]. Particularly, metabolomics focuses on comparing healthy versus metabolomic disease organisms and, therefore, it attempts to discover predictive biomarkers by detecting early biochemical changes before the appearance of the disease [2]. For that purpose, metabolomics experimental designs include non-targeted analysis of the samples as there is no prior knowledge of the metabolites that may be involved not only in fully developed metabolomic diseases, but also in pre-symptomatic stages.

Analytical techniques to identify and quantify metabolites include gas chromatography–mass spectrometry (GC–MS). Gas chromatography separates the compounds contained in a sample while passing through a chromatographic column. However, when two or more compounds do not completely separate chromatographically, those compounds are known to be co-eluted, and this clearly affects the correct quantification and identification of the metabolites. In that sense, comprehensive gas chromatography–mass spectrometry (GC×GC–MS) [3,4] was devised to minimize co-elution. In GC×GC–MS, the sample passes through two chromatographic columns with orthogonal polarity properties, which improves the compound separation and it leads to an increased compound detection capacity as co-elution is diminished.

However, compounds in the samples usually appear at trace levels and different sources of noise derived from the instrument and the sample biological matrix may interfere with the correct identification of the compounds. In the same way, GC×GC–MS generates large quantity of data and its interpretation cannot be conducted manually. In that sense, GC×GC–MS data processing algorithms are needed to turn the chromatographic signals into interpretable biological information. Besides, GC×GC–MS samples are composed by a large amount of data in comparison with GC–MS samples, and algorithms for GC×GC–MS data processing should be optimized for a fast data processing.

As reviewed in [5], some of the existing data processing algorithms that can be applied to resolve mixtures in comprehensive gas chromatography include PARAFAC [6] and multivariate curve resolution – alternating least squares (MCR-ALS) [7]. Contrarily to MCR, PARAFAC can only be applied to a three-way data set, i.e., PARAFAC cannot resolve a single GC×GC–MS chromatogram.

In recent years, independent component analysis (ICA) [8] has been introduced as an alternative to the traditional MCR for GC–MS data analysis [9–11]. ICA is a blind source separation (BSS) technique used to separate linearly mixed sources, i.e., it is capable of separating and retrieve the original compound sources – elution profile or spectra – from a mass spectra chromatogram. Whereas MCR–ALS resolves a chromatographic mixture by minimizing the residual error between the data and the predicted model, ICA uses another type of measure which is the statistical independence, and it estimates the original compound sources by maximizing the independence between components. ICA is widely applied in biomedical sciences, including data processing in electroencephalography recordings [12–14], and it is also one of the most reported algorithms for resolution of spectroscopy mixtures. More recently, we have developed a new method known as independent component analysis–orthogonal signal deconvolution (ICA–OSD) [15], embedded in an R package, that uses a combination of ICA and principal component analysis (PCA) to identify co-eluted compounds in GC–MS. In ICA–OSD, PCA is proposed as an alternative to the typical use of least squares (LS) in MCR-ALS. The application of LS for spectra extraction has different drawbacks, detailed in [15], which can be summarized in the fact that no correlation or covariance information is taken into account when applying LS, and therefore LS may find difficulties in distinguishing noise and the different compound fragments. This may lead to introducing a bias into the LS regressors especially in situations of co-elution or under undue biological matrix interference. Besides, whereas the current ICA-based methods consider the spectra as the independent source in the chromatograms, in ICA–OSD we implemented a different approach where we assumed that the elution profile was the independent source, as opposite to the spectra. In that sense, we used ICA to extract the elution profiles and then determine the spectra by means of OSD. Finally, ICA–OSD shown itself as a computationally faster alternative to MCR-ALS. As far as we know, the capability of independent component analysis–orthogonal signal deconvolution for compound quantification in chromatographic signals has not been studied.

In this paper we propose an automated method to deconvolve compounds appearing in GC×GC–MS chromatograms by independent component analysis–orthogonal signal deconvolution.

2. Materials and methods

2.1. Materials

The performance of ICA–OSD was evaluated through a set of 38 metabolites appearing in 20 Jurkat cell samples extracted from human acute T cell lymphoblastic leukemia cell line Jurkat. The samples of this experiment were previously used to report the intersection of phosphoethanolamine with menaquinone-triggered apoptosis by Dhakshinamoorthy et al. [16]. More details on the dataset, sample preparation and methods can be found in the original study.

2.2. Data analysis and pre-processing

ICA–OSD was used to automatically extract and deconvolve the compounds concentration profiles and spectra. The GC×GC–MS chromatograms were processed by analyzing each modulation cycle separately. Each modulation cycle was first divided in chromatographic peak features (CPFs) using the same criteria as in [17]. The different CPFs contained several compounds, so the algorithm had to deconvolve them in case of co-elution. The number of factors or components for ICA was determined by evaluation of residual sum of squares (described in Section 3.2).

The chromatograms were automatically processed by ICA–OSD. From the ICA–OSD output we only took into account those metabolites appearing in at least 15 of the 20 samples,