Analytica Chimica Acta 854 (2015) 20–33

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Chemometric strategy for untargeted lipidomics: Biomarker detection and identification in stressed human placental cells

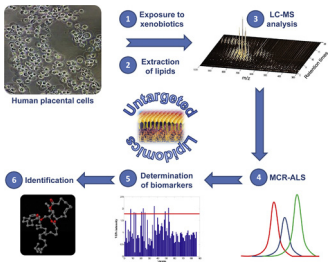
Eva Gorrochategui [a], Josefina Casas [b], Cinta Porte [a], Sílvia Lacorte [a], Romà Tauler [a,*]

[a] Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona 08034, Catalonia, Spain
[b] Department of Biomedicinal Chemistry, Institute of Advanced Chemistry of Catalonia (IQAC), Jordi Girona 18-26, Barcelona 08034, Catalonia, Spain

HIGHLIGHTS

- The lipidome of human placental cells exposed to xenobiotics is studied.
- Multiple liquid chromatographic coelutions are resolved by MCR-ALS.
- Two untargeted strategies are proposed to discover lipid disruption biomarkers.
- Biomarker identification allows interpretation of lipid changes in stressed cells.
- The proposed method is a powerful alternative to targeted lipidomic strategies.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Available online 11 November 2014

Keywords:
Untargeted lipidomics
Liquid chromatography–mass spectrometry
Multivariate curve resolution-alternating least squares
Partial least squares-discriminant analysis
Biomarkers
Obesogens

ABSTRACT

A lipidomic study was developed in a human placental choriocarcinoma cell line (JEG-3) exposed to tributyltin (TBT) and to a mixture of perfluorinated chemicals (PFCs). The method was based on the application of multivariate curve resolution-alternating least squares (MCR-ALS) to data sets obtained by ultra-high performance liquid chromatography coupled to time-of-flight mass spectrometry (UHPLC–TOF-MS) using an untargeted approach. Lipids from exposed JEG-3 cells were solid–liquid extracted and analyzed by UHPLC–TOF-MS in full scan mode, together with control samples. Raw UHPLC–TOF-MS data of the different cell samples were subdivided into 20 distinct chromatographic windows and each window was further organized in a column-wise augmented data matrix, where data from every sample was in an individual data matrix. Then, the 20 new augmented data matrices were modeled by MCR-ALS. A total number of 86 components were resolved and a statistical comparative study of their elution profiles showed distinct responses for the lipids of exposed versus control cells, evidencing a lipidome disruption attributed to the presence of the xenobiotics. Results from one-way ANOVA followed by a multiple comparisons test and from discriminant partial least squares (PLS-DA) analysis were compared as usual strategies for the determination of potential biomarkers. Identification of 24 out of the 33 proposed biomarkers contributed to the better understanding of the effects of PFCs and TBT in the lipidome of human placental cells. Overall, this study proposes an innovative untargeted LC–MS MCR-ALS approach valid for -omic sciences such as lipidomics.

* Corresponding author. Tel.: +34 934006140; fax: +34 93 204 59 04.
E-mail address: roma.tauler@cid.csic.es (R. Tauler).

http://dx.doi.org/10.1016/j.aca.2014.11.010
0003-2670/

1. Introduction

Lipids play essential roles in energy production and storage, structure and signaling in the human body [1,2]. Information about lipids and their individual classes can aid in understanding the pathogenesis of many disease states. Among others, obesity is characterized by a series of lipid disturbances [3], producing far-ranging effects on human health. Despite generally accepted causes for obesity are the consumption of calorie-dense food and diminished physical activity, the environmental obesogen hypothesis is raising acceptance in recent years. This hypothesis proposes that chemical exposure to molecules called "obesogens" during critical developmental stages influences subsequent adipogenesis, lipid balance and obesity [4]. Tributyltin (TBT) is a well-known endocrine disruptor previously used as a biocide in anti-fouling paints. TBT is a high-affinity agonist of the retinoic X receptor (RXR) and peroxisome proliferator activated receptor gamma (PPARγ), which are key components in adipogenesis and the function of adipocytes. TBT exposure has reported to promote inappropriate activation of RXR–PPARγ, causing a direct alteration of adipose tissue homeostasis [4]. Perfluorinated chemicals (PFCs) are potential obesogens used for many years in numerous industrial products, such as Teflon and have emerged as global environmental pollutants [5]. Among the variety of PFCs, perfluorooctanesulfonate (PFOS) and perfluorooctanoic acid (PFOA) are the most deeply investigated compounds in the last decades, whereas, other PFC homologs have been rarely studied. Recently, shorter chain length PFCs have been proposed as safe substitutes for the longer chain length ones, since they are expected to be less bioaccumulative and less toxic [6]. PFCs have been reported to alter lipid levels in some animal species and humans [7,8], but their mechanism of action is still unknown.

The study of lipids emerges as a complicated area of research due to their structural diversity and the considerable technical challenges associated with their identification within complex samples [9]. "The large-scale analysis of lipid profiles in cells and tissues" [10] was made possible by the dawn of lipidomics [11,12]. Lipidomics is a field that aims at the study of lipids and their interaction with other biochemicals [13]. While lipid-, and metabolome research in general, over past decades was overshadowed by the progress of genomics, recently revived and burgeoning interest in lipids illustrates their critical biological importance [14].

Mass spectrometry-based analytical methods come out as powerful strategies for lipidomics, since, they offer high sensitivity and resolution for the characterization of global lipid profiles in cells or organisms. Moreover, recent advances in mass spectrometry techniques have allowed the screening of many lipid molecular species in parallel [14]. These advances, however, pose a greater challenge for researchers to handle massive amounts of information-rich MS data from modern analytical instruments in order to understand the complexity of lipid systems. The application of modern chemometric methods to these complex megavariate data systems is opening new ways in bio and environmental sciences, facilitating a shift from the concept of studying one chemical compound or process at a single experiment. Thus, there is an urgent need to improve and automate all the steps involved in analyzing the data generated in -omic studies such as lipidomics by means of chemometric and multivariate data analysis methods.

Chemometrics is presently a well established field in chemical data analysis and has recently been proven to be valuable in the analysis of -omic data [15–17]. There is a considerable number of techniques especially suited for the study of complex megavariate -omic data sets, meant for exploratory or modelling purposes, such as principal component analysis (PCA), partial least squares (PLS) and its orthogonal variant (OPLS). Moreover, other less explored chemometric methodologies in -omic studies, such as multivariate curve resolution (MCR) methods evolve as powerful tools to properly resolve the profiling problem in -omic data sets [17,18]. In addition, chemometric methods can be used for biomarker detection in the context of finding sample descriptors which show systematic differences between normal and environmentally injured organisms in an untargeted approach. However, in areas of biology and toxicology, chemometric methodologies are still largely overlooked in favor of traditional statistical methods, which are generally focused on targeted evaluation of specific classes of compounds.

The aim of this study was to elucidate the lipidomic disruption produced in JEG-3 cells exposed to TBT and a mixture of PFCs, using MCR-ALS resolved profiles. Determination of biomarkers was based on the use of a traditional statistical approach, one-way ANOVA followed by a multiple comparisons test, versus the analysis by PLS-DA. The untargeted chemometric strategy presented in this study was designed as a novel alternative to the classical targeted approach generally used in lipidomics.

2. Theory

A brief description of the chemometric and statistical methods used in this study is shown below.

2.1. Multivariate curve resolution-alternating least squares (MCR-ALS)

Multivariate curve resolution methods [19] are based on the same bilinear decomposition of original data sets used by PCA, but under completely different constraints and with a different goal. The mathematical basis of the bilinear model used by MCR is shown in Eq. (1):

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

In this equation, matrix $\mathbf{D}$ ($I \times J$) represents the data output of a second-order instrument. In the case of LC–MS data, $\mathbf{D}$ matrix contains the MS spectra at all retention times ($i = 1, \ldots I$) in its rows, and the chromatograms at all spectra m/z channels ($j = 1, \ldots J$) in its columns. This matrix is decomposed in the product of two small factor matrices, $\mathbf{C}$ and $\mathbf{S}^{\mathbf{T}}$. The $\mathbf{C}$ ($I \times N$) matrix contains column vectors which correspond to the elution profiles of the N ($n = 1, \ldots N$) pure components of matrix $\mathbf{D}$. In $\mathbf{S}^{\mathbf{T}}$ ($N \times J$) matrix, row vectors correspond to the spectra of the N pure components. The part of $\mathbf{D}$ that is not explained by the model forms the residual matrix, $\mathbf{E}$ ($I \times J$).

MCR-ALS methods assume that the variation measured in all samples in the original data set can be described by a combination of a small number of chemically meaningful profiles. In the case of LC–MS data sets, information of the data table can be reproduced by the combination of a small number of pure mass spectra (row profiles in the $\mathbf{S}^{\mathbf{T}}$ matrix) weighted by the concentration of each of them along the elution direction (the related chromatographic elution peaks, column profiles in $\mathbf{C}$).

2.1.1. Column-wise augmented data matrices

In second order data, MCR-ALS can be implemented through different sample types simultaneously, conforming column-wise augmented data matrices ($\mathbf{D}_{\mathbf{aug}}$) containing distinct matrices correlated to different processes attached one at the top of each other. Thus, spectral direction is equal for all of them and the data matrix length is augmented in the process direction. Resolved pure mass spectra are equivalent to all experiments ($\mathbf{S}^{\mathbf{T}}$) whereas concentration profiles can differ from experiment to experiment, conforming $\mathbf{C}_{\mathbf{aug}}$, as shown in Eq. (2):