Journal of Chromatography A, 1384 (2015) 133–141

Contents lists available at ScienceDirect

Journal of Chromatography A

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

An untargeted gas chromatography mass spectrometry metabolomics platform for marine polychaetes

R. Fernández-Varela[a,b,*], G. Tomasi[a], J.H. Christensen[a]

[a] Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark
[b] Assays Department, Chr Hansen A/S, Boege Alle 10, DK-2970 Hoersholm, Denmark

ARTICLE INFO

Article history:
Received 31 July 2014
Received in revised form 6 January 2015
Accepted 11 January 2015
Available online 19 January 2015

Keywords:
Marine polychaetes
Nereis virens
Derivatization
GC–MS
XCMS
Environmental metabolomics

ABSTRACT

The development of an appropriate extraction method for untargeted environmental metabolomic analysis of marine polychaetes could promote their use for environmental monitoring purposes. To this end, we compared four extraction methods on the marine polychaete *Nereis virens* both exposed to crude oil and non-exposed. XCMS was used for feature detection and preprocessing; different normalization and scaling approaches were tested; and principal component analysis (PCA) was used together with basic statistical tests to ascertain common metabolic patterns and determine the most suitable extraction method. We conclude that a two-step extraction procedure with 80:20 (v/v) methanol:water on freeze dried polychaete tissue provides the best trade-off between analysis time, and extraction efficiency and intermediate reproducibility. No definitive conclusions could be drawn about the ability of the method to discriminate controls and crude oils in actual biological replicates because the experiment was carried out by design on analytical replicates only. We show that the normalization to the sum of all the common features, and the use of a weighted least squares criterion to fit the PCA by means of scaling to the median absolute deviation (MAD) of the pooled quality control samples significantly improved the clustering of controls and crude oil exposed samples. The scaling alone led to an increase of 19% in explained variance compared to ordinary PCA.

1. Introduction

Metabolomics is the systematic study of the set of metabolites present within an organism, cell, or tissue. Its application to environmental monitoring has been sparse but is receiving increasing attention. The role of a wide range of metabolites in the adaptation of biological species to a variety of stressors such as contaminant exposure has been the subject of much research: inter alia, amino acids, sugars, fatty acids, carbohydrates, and organic acids [1–3]. One advantage of untargeted metabolomic approaches is that a wide range of compounds involved with diverse traits, pathways and stress responses can be detected; the varying physicochemical properties pose a considerable challenge when choosing extraction protocols and analytical techniques, as the optimum extraction conditions may vary considerably between classes of compounds [1,4,5].

Sample preparation and extraction of metabolites from tissues is often the most labour-intensive and rate-limiting step in metabolomics [3]. Prior to extraction, different methods can be used to break up tissues and cells. To have reproducible measurements, the conditions of the biological material should be as homogeneous as possible. When working with tissue samples, it is important to keep the tissue in liquid state for as short time as possible to avoid enzymatic activity and degradation processes and to obtain an accurate reflection of the in vivo metabolism at the time of sampling. The metabolism therefore has to be quenched (e.g., by freezing) as rapidly as possible to prevent any enzymatic changes (e.g., labile phosphates and glycolytic intermediates can change on a millisecond timescale) and loss of volatiles [6].

Infaunal deposit-feeding invertebrates such as *Nereis virens* can be extensively exposed to oil hydrocarbons due to feeding activity and surface contact. The degree of systemic accumulation of oil hydrocarbons will depend not only on intake but also on the efficiency of biotransformation and excretion by the organism. The wide distribution and importance as food source to higher organisms of *N. virens* is known to extensively biotransform PAHs, making it important to understand the mechanisms underlying oil hydrocarbons biotransformation in this species [7–9].

* Corresponding author at: Assays Department, Chr Hansen A/S, Boege Alle 10, DK-2970 Hoersholm, Denmark. Tel.: +45 35332456.
E-mail addresses: dkrafe@chr-hansen.com, raquelfernandezv@gmail.com (R. Fernández-Varela).

http://dx.doi.org/10.1016/j.chroma.2015.01.025
0021-9673/

Grinding the individual worms in liquid nitrogen – cooled mortar until a fine powder is obtained and prepare a single homogenate sample is the 'gold-standard' method, but it is highly labour intensive [10–12] and requires considerable care to transfer the frozen powder. The method is also susceptible to sample carryover unless the mortar is warmed up and thoroughly cleaned between extractions. Furthermore, the method is challenging when small amounts of tissue are analyzed, and it cannot be used to extract multiple tissue samples simultaneously in a high-throughput manner [3,12].

To obtain relevant metabolic profiles, the number and amounts of metabolites extracted must be maximized, while the analytical variation is kept to a minimum [5]. Therefore, an appropriate extraction technique for each organism should be established that reduces experimental variability by increasing extraction efficiency and reproducibility [3]. Especially when new organisms are tested this entails the optimization of quenching and extraction methods [4,13–17].

GC–MS has been described as one gold standard for metabolomic studies, although it is biased against non-volatile, higher-MW metabolites [18]. The simultaneous determination of organic acid, sugars, fatty acids, carbohydrates and amino acids by GC methods requires the conversion of these compounds to volatile and thermal stables derivatives such as methyl-, acyl-, silyl-, oxime-, or oxime-silyl-derivatives. Due to the range of chemical functionality of metabolites, a two-stage derivatization procedure has been employed to remove hydrogen bond formations to increase volatility and reduce the interactions with the column phase, which can cause tailing peaks, poor sensitivity and poor chromatographic separation [19]. A popular approach is based on the work of Fiehn et al. [19], where carbonyl functional groups (e.g., sugars) are converted to oximes by reaction with hydroxylamine in pyridine, followed by silylation using N-methyl-N-trimethylsilylation (MSTFA) with TMCS as catalyzing reagent.

All metabolomic studies generate complex multivariate data sets that require visualization software and chemometric and bioinformatics methods for analysis and interpretation [6]. The XCMS R-package, integrated with some in-house routines, was used here for the preprocessing prior to the actual data analysis; the package is mature and it has been successfully employed in many metabolomic studies [20–22]. XCMS was used for feature detection, alignment and matching across samples to generate a consensus table of features common to a majority of the samples.

Normalization and scaling boil down to similar mathematical operations (viz., pre- or post-multiplication of the data matrix by a diagonal matrix): normalization is applied within samples whereas the scaling is applied across samples and within variables [23]. These two operations are not independent of one another and can have a significant impact on the interpretation of the model. The normalization of the analytical profiles (i.e., the set of area/intensity values obtained for each sample) is a central step in the preprocessing of metabolic profiling data [24]. The difference in concentrations between extracts can be large and may depend on a variety of random factors like fluctuations in the sample preparation and the extraction efficiency. If this variability were not taken into account by the models or corrected by appropriate preprocessing, multivariate analyses of unprocessed signals of biological samples are bound to be biased by the samples with the most intense signals. Scaling prior to data analysis most often serves one of two goals: to bring all variables into the same range and to correct for a non-constant signal variance. It can also be employed, as is the case here, to fit the PCA model in a weighted least squares (WLS) sense [23]. Different normalization and scaling approaches were tested, an aspect which is not dealt with in XCMS by design [22]. Basic statistical tests together with multivariate clustering approaches in the principal component space were used to ascertain common metabolic patterns and establish the most suitable analytical platform.

The main objective of this study was therefore to develop and validate an untargeted GC–MS metabolomics platform for marine polychaetes. Four extraction methods were tested for the extraction of metabolites from *N. virens* and compared in terms of intermediate reproducibility (both for individual peaks and for the entire metabolite fingerprint), analyte coverage, and sensitivity; simplicity, and speed of analyses were also taken into account. XCMS was used for feature detection and extraction and WLS-PCA was used to assess the ability of the method to discriminate analytical replicates of two homogenized samples of *N. virens*: one from individuals exposed to crude oil and one of non-exposed worms in whole body homogenates. The four extraction protocols are based on 80:20 (v/v) methanol:water: (a) extraction of wet homogenized tissue, (b) extraction of freeze dried tissue, (c) pressurized liquid extraction (PLE) with C18 solid phase extraction (SPE) material, and (d) PLE without the C18-SPE material. The extraction procedure used in (a) and (b) involves multiple steps such as, solvent additions, mixing, vortexing, centrifugation and filtration [3]. Methods (c) and (d) required freeze-drying to avoid enzymatic activity during the extraction. 80:20 (v/v) methanol:water was selected as often recommended in the literature and seemingly able to provide a good analyte coverage, extraction efficiency, and reproducibility for various tissue metabolomes [3,10,25–27].

This study constitutes only a stepping stone to the ultimate goal of identifying the metabolites that discriminate between crude oil exposed and non-exposed marine polychaetes, and therefore the discovery of a potentially complex set of biomarkers that define the biological context and help explain the mechanisms related to tissue response. The experimental set up did not include biological replicates, since, in general, the analytical/technical variance is considerably smaller than metabolite variance from biological sources [3] even under very controlled conditions [18], and the objective of this study was to determine the best sample preparation, extraction and data preprocessing method for untargeted GC–MS metabolomics for marine polychaetes.

2. Experimental

2.1. Chemicals and reagents

All the chemicals used were of analytical grade. Methanol was purchased from Fisher Scientific (Loughborough, UK), dichloromethane from Raphburn Chemicals (Walkerburn, UK), ultrapure water (18.2 Ω) from Millipore (Molsheim, France), formic acid and N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) with 1% Trimethylchlorosilane (TMCS) from Fluka, Sigma-Aldrich (Broendby, Denmark), Methoxyamine from Supelco (Supelco Park, PA, USA), pyridine from Merck (Darmstadt, Germany). Standard compounds used for metabolite identification were obtained from Sigma Aldrich (Broendby, Denmark). 2-Hydroxy-2-methylpropionic acid (in 20:80 (v/v) methanol:water), L-arabinose (in 20:80 (v/v) methanol:water) and Myristic acid D27 (in pure methanol) (all from Sigma-Aldrich, Broendby, Denmark) at 1 mM concentrations were used as internal standards. Ottawa sand (particle size 20–30 mesh, Fisher Chemicals) was burned in a muffle oven at 400 °C for 24 h.

2.2. Animals

Twelve *N. virens* individuals (2.6 ± 1.6 g; mean ± std) were introduced into two contaminated microcosms with crude oil to reach a density of around 200 individuals m^{-2} (6 worms in each); and ten *N. virens* (2.5 ± 1.5 g; mean ± std) were then