Chemosphere 85 (2011) 1211–1219

Contents lists available at ScienceDirect

Chemosphere

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

A new approach to data evaluation in the non-target screening of organic trace substances in water analysis

Alexander Müller [a,b], Wolfgang Schulz [a,*], Wolfgang K.L. Ruck [b], Walter H. Weber [a]

[a] Zweckverband Landeswasserversorgung, Betriebs- und Forschungslaboratorium, Am Spitzigen Berg 1, D-89129 Langenau, Germany
[b] Leuphana Universität Lüneburg, Institut für Umweltchemie, Scharnhorststr. 1, D-21335 Lüneburg, Germany

ARTICLE INFO

Article history:
Received 3 March 2011
Received in revised form 4 July 2011
Accepted 6 July 2011
Available online 5 August 2011

Keywords:
Non-target screening
General Unknown Screening
Transformation products
HPLC–QTOF-MS
Water purification

ABSTRACT

Non-target screening via high performance liquid chromatography–mass spectrometry (HPLC–MS) has gained increasingly in importance for monitoring organic trace substances in water resources targeted for the production of drinking water. In this article a new approach for evaluating the data from non-target HPLC–MS screening in water is introduced and its advantages are demonstrated using the supply of drinking water as an example. The crucial difference between this and other approaches is the comparison of samples based on compounds (features) determined by their full scan data. In so doing, we take advantage of the temporal, spatial, or process-based relationships among the samples by applying the set operators, UNION, INTERSECT, and COMPLEMENT to the features of each sample. This approach regards all compounds, detectable by the used analytical method. That is the fundamental meaning of non-target screening, which includes all analytical information from the applied technique for further data evaluation. In the given example, in just one step, all detected features (1729) of a landfill leachate sample could be examined for their relevant influences on water purification respectively drinking water. This study shows that 1721 out of 1729 features were not relevant for the water purification. Only eight features could be determined in the untreated water and three of them were found in the final drinking water after ozonation. In so doing, it was possible to identify 1-adamantylamine as contamination of the landfill in the drinking water at a concentration in the range of 20 ng L^{-1}. To support the identification of relevant compounds and their transformation products, the DAIOS database (Database-Assisted Identification of Organic Substances) was used. This database concept includes some functions such as product ion search to increase the efficiency of the database query after the screening. To identify related transformation products the database function "transformation tree" was used.

1. Introduction

One of the consequences of the manifold uses of various organic substances in households, agriculture, and industry is that these can eventually end up in the aquatic environment. In addition to pesticides and their metabolites or transformation products (Sancho et al., 2006; Weber et al., 2007; Buttiglieri et al., 2009) as well as pharmaceuticals and X-ray contrast media (Mompelat et al., 2009), industrial chemicals (e.g., benzotriazoles; Giger et al., 2006; Weber et al., 2009) are the most frequently occurring, anthropogenic substances in the environment. Depending on the application of the chemicals, entry occurs directly, for example during the application of pesticides (Hogendoorn et al., 1996), or indirectly, via wastewater treatment plants (Farré et al., 2008) among others, into the environment. The aquifers can also be contaminated via refuse dumps or landfill leachates (Holm et al., 1995) as well as from accidents (Farré et al., 2008). Thus, to protect of the quality of drinking water, a constant and comprehensive monitoring of organic trace substances in the resources used for drinking water treatment is essential.

Hernández et al. (2005) divided the screening methods used for the monitoring into three different categories: (i) Pre-target screening: the particular analyte is selected before the analysis and further contaminants in the sample are not detected. (ii) Post-target screening: all compounds eluted from chromatography columns and ionized in the ion source are detected by mass spectrometry (*full scan mode*). Then post-selected analytes can be identified based on their extracted ion chromatograms (EIC). (iii) Non-target screening (a.k.a. *General Unknown Screening)*: all substances accessible to the particular analysis technique are also detected; however, the focus is not on the pre- or post-target analytes, but on unexpected or unknown substances.

Today, modern screening for medium polar and polar substances is via HPLC–MS. For pre-target screening to detect

* Corresponding author. Tel.: +49 (0)7345 9638 2291; fax: +49 (0)7345 9638 2290.
E-mail address: Schulz.W@lw-online.de (W. Schulz).

0045-6535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2011.07.009

organic trace substances, triple-quadrupole MS in multiple reaction monitoring (MRM) mode is predominantly used, due to its selectivity and high signal-to-noise ratios (S/N) (Barceló and Petrovic, 2007; Krauss et al., 2010), which enable the quantification of target analytes without prior enrichment (direct analysis) down to the ng L^{-1} range (Seitz et al., 2006). However, the S/N of these systems sinks drastically if they are operated in scanning mode (Tolonen et al., 2009; Krauss et al., 2010). Thus, high resolution mass spectrometers, such as time-of-flight (TOF) and Fourier transform MS, are used for the post-target and non-target screening procedures because of their better S/N in scanning mode (Ibáñez et al., 2009; Tolonen et al., 2009). Here, the ion chromatograms of the individual post-target analytes are extracted from the raw data of the entire full scan spectra and then verified.

To identify unexpected or unknown compounds, full scan MS spectra must be examined by molecular feature algorithm for masses whose abundances build chromatographic and mass spectrometric peak profiles (Gómez et al., 2010). Given the exact mass, along with the isotope pattern, it is then possible to use a formula generator to compute a molecular formula of the feature (Hogenboom et al., 2009). The allocation of the molecular structure of the compound (feature) can be determined with the help of databases. Database hits are then verified by means of reference substances and MS/MS experiments. For compounds, not being identified this way, the structure elucidation is commonly done by sophisticated MS scan techniques (e.g., MS^n experiments; Liu and Hop, 2005) supplemented by derivatization reactions (Werner et al., 2008) and H/D exchange (Wolff and Laures, 2006; Liu et al., 2007) combined with information from other analysis methods (Müller et al., 2010).

The selection of those compounds recognized by non-target screening to be submitted for further identification is usually based on signal intensity (e.g., Bobeldijk et al., 2001). The signal intensity represents the product of the concentration and the ionization efficiency of the compound, which is perhaps why relevant compounds with high concentration but small ionization efficiency are often not considered. Due to the multiplicity of features that are usually detected in most environmental samples, however, it is necessary to focus on the identification of only relevant compounds (Gómez et al., 2011).

In this work a new approach for focusing this selection on relevant features is described. Here the sample is not regarded as an isolated specimen, but rather it is evaluated in relation to a set of other samples based on considerations of e.g., their temporal, spatial, or process-related connections. (i) All detected features of the different samples are used for the following data evaluation. The detection of substances or substance groups is only limited by the applied analytical technique. That is classical meaning of non-target screening. (ii) The features of a sample are considered as mathematical sets. This enables the comparison of all features in the various samples by applying the set operators, UNION, INTERSECT, and COMPLEMENT to determine the compounds of interest. This mathematical procedure allows solving complex analytical questions. (iii) All features above a given intensity threshold are equally treated regarding their intensities during data evaluation. In contrast, for further investigation, e.g., principle component analysis (PCA), the intensities are included.

To effectively support the identification of compounds from aquatic environments, in this article the use of the DAIOS (Database-Assisted Identification of Organic Substances) database is introduced. Finally, the possibilities of this new approach are demonstrated using the example of the influence of contaminated groundwater on a drinking water supply.

2. Material and methods

2.1. Chemicals

Acetone, methanol, *n*-hexane, concentrated sulfuric acid (96–97%), formic acid (purity >98%), ammonium acetate (purity >98%), and hyper-grade methanol for HPLC–MS analysis were purchased from Merck (Darmstadt, Germany). HPLC water was prepared from deionised water using a Millipore Milli-Q system (Billerica, MA, USA). The following reference substances were obtained from various suppliers (Sigma Aldrich, Steinheim; VWR International GmbH, Darmstadt; and Dr. Ehrenstorfer, Augsburg, all from Germany): 1-adamantylamine, 2-hydroxybenzothiazole, 4-aminoantipyrine, acetylaminoantipyrine, carbamazepine, clofibrinic acid, crotamiton, dimethylaminoantipyrine, N,N-diethyl-meta-toluamide (DEET), formylaminoantipyrine, phenazone, propyphenazone, sulfadiazine, and sulfamerazine.

2.2. Sampling

The samples of landfill leachates as well as ground and process waters were collected in 2-liter amber glass bottles and stored, without further additives, in the dark at 4–6 °C.

2.3. Sample preparation (solid phase extraction)

Solid phase extraction (SPE) of the water samples was carried out on resin-based sorbent Isolute ENV + cartridges (200 mg, Biotage AB, Uppsala, Sweden) at pH 7 and pH 3; pH was adjusted by the addition of a 1:4 dilution of sulfuric acid. The Isolute ENV + cartridge was preconditioned with successive additions of 6 mL each, *n*-hexane, acetone, methanol, and HPLC water (pH 7 or pH 3). Subsequently, a liter of the sample was applied via a peristaltic pump at a constant flow of approximately 3 mL min^{-1}. After air-drying the resin for 30 min, the sample was eluted with a total of 6 mL methanol in 1 mL aliquots. The eluate was concentrated to near dryness with 50 °C nitrogen, taken up in 1 mL methanol, and filtered over a cellulose acetate membrane filter (0.2 μm, Restek GmbH, Bad Homburg, Germany). A sample blank for background was generated by processing a liter of HPLC water through the same procedure as for the samples. Additionally, the conditioned solid phase was immediately eluted with methanol to differentiate sample blank and solid phase cartridge blank. Methanolic extracts were stored in the dark at 4–6 °C.

2.4. HPLC–QTOF-MS analysis

The extracts were analyzed via high performance liquid chromatography, quadrupole time-of-flight mass spectrometry (HPLC–QTOF-MS). The HPLC (1100 Series, Agilent Technologies, Waldbronn, Germany) consisted of a degassing unit (G1322A), a binary pump (G1312A), an autosampler (G1313A), and a thermostated column compartment (G1316A). A reversed-phase column with a flow rate of 0.6 mL min^{-1} was used for chromatography (Zorbax Eclipse XDB-C18, 1.8 μm, 4.6 × 50 mm). As ionization modifiers, both eluents, HPLC water (A) and methanol (B), contained 0.1% (v/v) formic acid in the positive electrospray ionization (ESI+) mode and 5 mM ammonium acetate in the negative (ESI−) mode. The gradient was as follows: 1 min isocratic with 10% B in A, an 8 min linear increase to 90% B, 5 min isocratic with 90% B, and 5 min equilibration with 10% B. The injection volume was 10 μL. QTOF-MS (Accurate-Mass 6520 QTOF, Agilent Technologies, Santa Clara, CA, USA) was conducted with the settings shown in Supplementary Information (Table S1). The stability of the