

Identification strategy for unknown pollutants using high-resolution mass spectrometry: Androgen-disrupting compounds identified through effect-directed analysis

Jana M. Weiss · Eszter Simon · Gerard J. Stroomborg · Ronald de Boer · Jacob de Boer · Sander C. van der Linden · Pim E. G. Leonards · Marja H. Lamoree

Received: 20 December 2010 / Revised: 11 March 2011 / Accepted: 20 March 2011 / Published online: 21 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Effect-directed analysis has been applied to a river sediment sample of concern to identify the compounds responsible for the observed effects in an *in vitro* (anti-)androgenicity assay. For identification after non-target analysis performed on a high-resolution LTQ-Orbitrap, we developed a *de novo* identification strategy including physico-chemical parameters derived from the effect-directed analysis approach. With this identification strategy, we were able to handle the immense amount of data produced by non-target accurate mass analysis. The effect-directed analysis approach, together with the identification strategy, led to the successful identification of eight androgen-disrupting compounds belonging to very diverse

compound classes: an oxygenated polyaromatic hydrocarbon, organophosphates, musks, and steroids. This is one of the first studies in the field of environmental analysis dealing with the difficult task of handling the large amount of data produced from non-target analysis. The combination of bioassay activity assessment, accurate mass measurement, and the identification and confirmation strategy is a promising approach for future identification of environmental key toxicants that are not included as priority pollutants in monitoring programs.

Keywords LTQ-Orbitrap · Non-target analysis · Identification · (Anti-)androgenic · Effect-directed analysis (EDA) · MODELKEY

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-4939-x) contains supplementary material, which is available to authorized users.

J. M. Weiss
Joint Research Centre, European Commission,
Institute for Environment and Sustainability,
Via Enrico Fermi 2749,
21027 Ispra, Italy

E. Simon · J. de Boer · P. E. G. Leonards · M. H. Lamoree (✉)
Institute for Environmental Studies (IVM),
Faculty of Earth and Life Sciences, VU University,
De Boelelaan 1087,
1081 HV Amsterdam, The Netherlands
e-mail: marja.lamoree@ivm.vu.nl

G. J. Stroomborg · R. de Boer
Rijkswaterstaat, Centre for Water Management,
Zuiderwagenplein 2,
8224 AD Lelystad, The Netherlands

S. C. van der Linden
BioDetection Systems (BDS) B.V.,
Science Park 406,
1098 XH Amsterdam, The Netherlands

Introduction

In the year 2000, the Water Framework Directive (WFD) stated a demand for a good ecological status of European surface waters by 2015. A key issue is to identify the most important active compounds causing a specific effect, i.e., the key toxicants, in the aquatic environment. One of the approaches used to trace key toxicants is the application of effect-directed identification for the establishment of cause-effect relationships. Effect-directed analysis (EDA) studies employ bioassay-directed fractionation techniques to decrease the complexity of the sample matrix before identification of the active compounds. This fractionation approach is a non-selective and non-destructive clean-up methodology that aims to enable the identification of all biologically active compounds in the sample. EDA has successfully been applied to evaluate endocrine potencies in several water systems, such as waste-water treatment plants [1, 2], rivers [2, 3], harbor areas [4], marine sediment [5], and biota [6].

The identification of key toxicants and the final confirmation of toxicity and identity are critical for the success of EDA studies. In the past few years, this aspect of EDA has been lacking due to the immense number of compounds present in the complex samples, even after extensive fractionation. A common approach is target analysis of certain compound groups, based on prior structure-toxicity knowledge [5, 7, 8]. This excludes the possibility to identify compounds with unknown modes of action, or the identification of emerging pollutants. The amounts and purity of the compounds present are often not sufficient for spectrometric analysis using nuclear magnetic resonance spectroscopy or infrared spectroscopy, and thus, more theoretical methods need to be employed in the process, e.g., using classifiers, structural alerts, or even database-screening to limit the number of compounds to be handled [9–11]. Traditionally, in the area of environmental analysis, gas chromatography coupled to mass spectrometry (GC/MS) has been applied extensively. We used GC/MS for the identification of compounds causing the observed androgenic activity in an EDA study in sediment from the river Scheldt basin in Belgium, at the location Schijn Eindhoven. The sample extraction, cleanup, and fractionation strategy as well as the bioassay results have been described in detail, and the GC/MS results of the tentatively identified compounds are discussed elsewhere [12].

Complementary to that study, a parallel analytical strategy using liquid chromatography coupled to mass spectrometry (LC/MS) was developed to identify the more polar and less volatile compounds, which might be overlooked when using GC/MS techniques. Although LC/MS is the method of choice for the analysis of more polar compounds, there is a lack of accessible, easy-to-use spectral libraries or databases for compound identification, in analogy to those available for GC/MS identification (e.g., the NIST library). This demands the development of alternative approaches for structure elucidation exploiting as much information as possible gained from parameters such as chromatographic retention, physical–chemical properties, spectral data, etc.

To enable the identification of unknown compounds by LC/MS techniques, the use of equipment that is capable of accurate mass measurements, e.g., the LTQ-Orbitrap, is a prerequisite. At present, there is a wide range of software for processing LC/MS mass scans, from the commercially available Mass Frontier (Thermo Fisher Scientific) and ACD/MS fragmenter (ACD/labs) to freely available MZmine [13], XCMS [14], and FiD [15]. In this study, we used SIEVE software (Thermo Fisher Scientific) to facilitate the identification process of unknown compounds. SIEVE is designed to enable users to analyze large amounts of data generated with LC/MS techniques, especially within the field of -omics, such as metabolomics [16] and

proteomics [17]. Our aim was to develop and implement an integrated strategy for the identification of key toxicants using accurate mass LC/MS. Our approach was demonstrated for a sediment extract by the identification of unknown compounds that are responsible for the observed androgenic and anti-androgenic effects in an EDA study in the river Scheldt basin.

Materials and methods

The sample and the EDA methodology have been described in detail earlier [12]. In short, it included accelerated solvent extraction, gel permeation chromatography (GPC), reversed and normal phase (RP and NP) liquid chromatography (LC) techniques. Androgenic activity was determined in the AR-CALUX[®] bioassay [18]. Dihydrotestosterone (DHT) was used as a reference androgen receptor (AR) agonist. Antagonism (–AR) towards the AR was also determined by testing the extracts in combination with the EC₅₀ of DHT (200 pM). Flutamide (FLU) was used as the reference antagonist. The GPC fraction (whole extract) and all five RP fractions (based on increasing octanol–water coefficient [K_{ow}]) were tested in the bioassays. The RP fractions which were active in the bioassays were selected for further fractionation on an NP column, separating the compounds according to their polarity.

Chemical analysis with the LTQ-Orbitrap

The fractions were analyzed on an Xbridge C18-LC column (Waters, 100×2.1 mm 3.5 μ m) connected to a linear quadrupole ion trap—Orbitrap (LTQ-Orbitrap) mass spectrometer (Thermo Electron) equipped with an electro-spray ionization source (ESI). The eluent flow rate was 0.2 mL/min, and the solvent gradient used was 5% to 95% MeOH (solvent A, 5% MeOH/95% MilliQ water+0.05% formic acid; solvent B, 95% MeOH/5% MilliQ water+0.05% formic acid) in 25 min. The mass spectrometer was operated in the positive ionization mode (the negative ionization mode was evaluated but not reported). The data-dependent mode was activated to automatically switch between Orbitrap-FTMS and LTQ-MS/MS data acquisition (MS2 and MS3 data were not used, but the information is saved for eventual future investigations). Survey full-scan MS spectra (from m/z 50 to 600) were acquired in the Orbitrap with a resolution of 30,000.

Identification strategy

The identification strategy is described in Table 1, including aspects of each step that will be addressed in the [Results and discussion](#) section. The software SIEVE (Thermo Fisher) was used to sieve the active (observed (anti-)

Table 1 The description of the identification strategy of the active compounds present in the fractions, with the software used (method) and the limitations of each step

Workflow	Method	Limitations
1. LTQ-Orbitrap analysis of EDA identified fractions (sample=active fraction, control=non-active fraction)	LTQ-Orbitrap/Xcalibur	Electron spray ionization is a soft ionization technique and could hence exclude many compounds, matrix ion suppression
2. SIEVE procedure, aligns and base-peak frames the masses	SIEVE	No true control samples
3. Excel sheet with extracted information (peak nr, ratio between control and sample, accurate mass and retention time)	Excel	Ratio >100 between peak intensity in active and non-active fraction. Ratio <100 not tested.
4. Manual check of the peak quality (peak shape, isotopes to the molecular ion, base line elevation, etc.	Xcalibur	Time-consuming
5. Generate the chemical formula from the m/z in the elemental composition tool in Xcalibur	Xcalibur elemental composition tool	Selected elements: N, O, C, H, F, Cl, Br, I, Si, P, S
6. NIST search of the suggested chemical formula (Observed m/z is $[M+H]^+$ and hence one hydrogen needs to be deleted in the search)	NIST	Limited representation of polar compounds and other less known products
7. Generate the log K_{ow} , and match with the expected log K_{ow} of the active fraction.	EPI Suite	Theoretical values
8. Purchase and confirm analytically by retention time	LTQ-Orbitrap/Xcalibur	Lack of available standards
9. Confirm the activity in the corresponding bioassay	Bioassay	Lack of available standards
10. List of identified and confirmed active compounds to be target analyzed for concentration determination to determine the contribution to the overall activity of the sample.		

androgenic activity) and the non-active fractions (no observed (anti-)androgenic activity) to discriminate the peaks of interest. SIEVE identifies statistically significant changes in relative signal intensity of m/z values in a predefined experimental design, which typically is a two-group randomized controlled study. SIEVE operates in two steps, a first alignment of the chromatogram employing a ChromalignTM algorithm and, secondly, a recursive base-peak framing. The end product is a list of accurate masses (peaks) and the ratio of the peak intensity between sample and control and the p value of the significant difference.

Settings during the sieving procedure were a threshold of 500,000 which determines the lowest mass intensity that triggers framing: window of m/z 150–600; 1 min frame time window; 0.02 frame m/z width; retention time 9–17 min; and 30% peak width. The basic requirement to select m/z values of interest was that the difference was significant (p value < 0.05). Secondly, a threshold was set to include only peaks with a ratio of >100 between peak intensity in active and non-active EDA fractions. The peaks with a lower intensity ratio may also be of interest, but these were not addressed in this project to limit the data set volume. A third requirement was that the suggested chemical formulas extracted from the accurate mass (via the elemental composition tool in Xcalibur, Thermo Fisher) should be present in the NIST spectral database and optimally be attributed a CAS number to simplify the purchase of the compounds for further confirmation studies.

To cross-check whether major errors in the identification of unknown toxicants have been made using our identification strategy based on the NIST database, PubChem was selected as the most suitable alternative for the purpose of our study: the identification of unknown pollutants through a combination of biological and chemical characteristics that we have obtained through EDA. PubChem is an example of a shared database that collates an immense amount of data from various existing databases. The focus of PubChem is on biological activity and on compounds that chemically resemble bioactive compounds. The chemical formulas determined in the three most potent fractions (RP3NP2, RP3NP5, and RP3NP6) were searched in PubChem, and results were compared with those obtained searching the NIST library.

Each tentatively identified compound's inherent K_{ow} was correlated to the K_{ow} of the fraction in which it was found (the log K_{ow} range for the RP fractions 1–5 are <2, 2–4, 4–6, 6–9, and >9, ± 0.5 [12]). The logarithmic K_{ow} values were calculated with EPI Suite (KOWWIN v1.67, U.S. EPA) and are hence theoretical values.

Quality assurance/quality control

Analytical The instrumental mass tolerance was between 0.3 and 3 ppm (Table 2), and the typical setting was 2 ppm for the identification of m/z values in Xcalibur. The purchased compounds, indicated in Table 2 and the

Table 2 The name, CAS number, chemical formula, and the mass-to-charge ratio (m/z) as measured in the LTQ-Orbitrap (assumed to be $[M + H]^+$) of the identified and confirmed compounds present in EDA sediment, as well as the limit of detection (LOD, $\mu\text{g/L}$) and mass

tolerance (ppm) established in the sample. Androgenic and anti-androgenic potency are expressed as dihydrotestosterone (DHT) and flutamide (FLU) equivalence factors (EF), respectively, for the AR-CALUX[®] on a molar basis (Reference EC₅₀/Compound EC₅₀)

Compound	CAS number	Molecular formula	$[M + H]^+$ (m/z)	LOD ($\mu\text{g/L}$)	Mass tolerance (ppm)	DHT EF	FLU EF
7 H-Benz[de]anthracen-7-one	82-05-3	C ₁₇ H ₁₀ O	231.0806	20	0.9		1.4E-02
Galaxolide	1222-05-5	C ₁₈ H ₂₆ O	259.2058	20	0.8		3.9E-01
Tonalide	1506-02-1	C ₁₈ H ₂₆ O	259.2058	<20	1.2		3.2E-01
Traseolide	68140-48-7	C ₁₈ H ₂₆ O	259.2058	<20	0.8		2.2E-01
Tris-(2-chloroisopropyl) phosphate	13674-84-5	C ₉ H ₁₈ Cl ₃ O ₄ P	327.0082	<20	0.3		2.0E-04
Tris-(2-ethylhexyl) phosphate	78-42-2	C ₂₄ H ₅₁ O ₄ P	435.3591	<20	3.0		1.7E-02
Nandrolone	434-22-0	C ₁₈ H ₂₆ O ₂	275.2004	<20	0.7	1.2E-01	
5 α -Androst-16-en-3-one	18339-16-7	C ₁₉ H ₂₈ O	273.2215	630	2.2	1.0E-15	7.7E+00

Electronic Supplementary Material Table S1, were analytically confirmed by retention time (± 0.02 s). The instrumental limit of detection (LOD) was set as a minimum peak height of 50,000 relative response units due to bad peak shape below that threshold. The LODs ranged from <20 $\mu\text{g/L}$ up to 630 $\mu\text{g/L}$ (10- μL injection), depending on the ionization efficiency of the identified compound (Table 2). A solvent blank sample (extracted clean sand) was analyzed in parallel to measure background contamination. No bioassay activity was detected in any solvent blank fractions.

SIEVE The performance of the SIEVE program was evaluated with respect to the chromatogram alignment and base-peak framing procedure. The identified compounds ($n=8$) were added to a non-active sediment extract at four concentration levels (0.1–5.0 mg/L) and analyzed on the LTQ-Orbitrap. The obtained data were evaluated using SIEVE. Of the eight compounds, seven were peak-framed successfully and reported a ratio >100 compared with the control sample (non-spiked extract) at the lowest added concentration (0.1 mg/L). The eighth compound was 5 α -androst-16-en-3-one (androstenone) which has a high LOD (Table 2) due to ionization difficulties in ESI mode. Detection of androstenone was not possible below a concentration of 3 mg/L (SIEVE mass threshold setting was 500,000 and mass intensities below that was not selected for base-peak framing), suggesting that the concentration in the active fraction in which this compound was identified was rather high.

Apart from the eight parent compounds used for the spiking experiment, 51 m/z values were selected with a ratio >100 between spiked (at a level of 0.1 mg/L) and non-spiked extract. In total, 24 of these m/z values were present in the pure standard mixture added to the sample, while

6 m/z values were the isotopes of another mass ($[M+H+1]^+$ or $[M+H+2]^+$ ions). The identity of the remaining 14 m/z values could not be explained, but the possibility that these signals belong to adducts containing alkali metal ions and solvent molecules should be taken into account.

Results and discussion

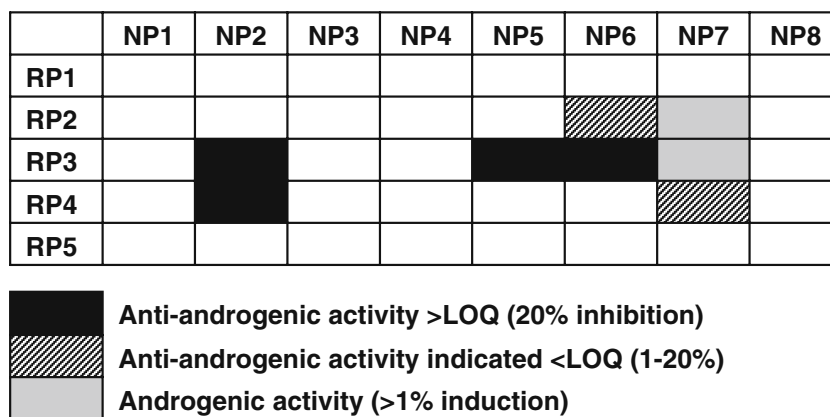
Bioassay results as starting point for identification

The bioassay results of the sediment sample from Schijn Eindhoven are given in Fig. 1, revealing the active and non-active fractions. The sediment extract was first reversed phase (RP) fractionated separating compounds according to their log K_{ow} value and secondly with normal phase (NP) LC giving fractions with increasing polarity. In total, three distinct clusters of active fractions could be observed. There were two potent anti-androgenic clusters consisting of one non-polar group (RP3NP2 and RP4NP2) and one slightly more polar group (RP3NP5 and RP3NP6), and one cluster of polar androgenic compounds (RP3NP7 and RP4NP7). The bioassay results are discussed in detail elsewhere [12].

Identification strategy

The data handling strategy represented in Table 1 consists of ten steps (described in detail below) to be performed in order to obtain a list of identified and confirmed biologically active compounds in the sample of interest. The method is not complicated, but since it incorporates several different software programs, it is laborious and time-consuming. Automatic data processing could drastically decrease the work load. However, in this study, there was no possibility to investigate the compatibility of the used

Fig. 1 Overview of bioassay results integrating both androgenic as well as anti-androgenic activity from the reversed phase (RP) and normal phase (NP) fractions. Active (*filled*) and non-active (*empty*) fractions are indicated. These bioassay results have been reported in detail earlier [12]



software programs. In Table 3, the results of the various steps in the procedure are given, represented by numbers of peaks remaining after a specific step.

The first step in Table 1 is the analysis of the fractions on the LTQ-Orbitrap to obtain accurate mass measurements of the compounds present in the fractions. The soft ionization technique (ESI) used to create molecular ions may not be optimal for all compound groups. Therefore, one should realize that the choice of LC/MS interface inherently influences the identification results. In general, the choice for a specific analytical technique for identification of unknowns inevitably narrows the scope of the results that can be obtained. All fractions were analyzed both in positive and negative ionization mode, but since the *m/z* signals in the negative ionization mode did not contribute to the results list, the data are not presented here.

The second step involves the SIEVE program, used to align and frame the mass chromatographic peaks that should be evaluated for identification. A comparison with a control sample is part of this procedure. In EDA studies,

the fractions that show activity in the bioassay are the fractions of interest in which the active compounds need to be identified. In our study design, we used two control samples for the SIEVE procedure. The first control is the non-active fraction adjacent (on the RP fractionation axis or the NP fractionation axis) to the active fraction, while the second is the identical fraction of a solvent blank sample. One limitation in our application is that SIEVE is designed for controlled studies, with exposure (sample) and non-exposure (control) scenarios in the same matrix. In the current study design, there is no “true” control sample: a “true” control would be the same fraction as the active fraction in the fractionation scheme presented (Fig. 1), originating from a sample with no activity from an identical environment, i.e., containing all the non-active compounds. However, such a control sample/extract was obviously not feasible.

In the third step, a ratio of >100 of the mass intensities between active fraction and control fractions, was adopted as a threshold. The lack of true control samples led to the

Table 3 The number of *m/z* values discriminated in the active androgenic fractions (agonistic or antagonistic and corresponding reversed phase [RP] and normal phase [NP] fractions) corresponding to identifications steps 1–10, separated into: the number of base-peak-framed *m/z* values from SIEVE; the number of *m/z* values with a ratio

between 10–100 and >100 between control and sample; tentatively identified compound (some *m/z* values have multiple possibilities); available compounds for the confirmation steps; the analytically confirmed and the bioassay-confirmed compounds that represent the key toxicants in the sample

Bioassay activity	Fraction	SIEVE <i>m/z</i> values	Ratio >100 <i>m/z</i> values	Tentatively identified	Available compounds	Analytical confirmed	Bioassay-confirmed
Identification step		2	3–4	7	8	8	9
Agonistic	RP2NP7	1,002	38	25	8	2	1
	RP3NP7	2,915	62	7	1	0	0
Antagonistic	RP2NP6	1,206	35	6	2	1	1
	RP3NP2	3,309	26	2	0	0	0
	RP3NP5	1,141	42	42	7	5	4
	RP3NP6	2,940	26	4	2	1	1
	RP4NP2	291	1	1	0	0	0
	RP4NP7	2,003	29	8	2	1	1
Total		14,807	259	95	22	10	8

inclusion of a relatively large number of compounds that were in fact not related to the observed biological activity. Because of the labor-intensive identification strategy, there was a need to reduce the amount of data, and therefore the relatively arbitrary threshold of a ratio of >100 between active and non-active fractions was implemented. The reduction of data is illustrated in Table 3: after the application of the SIEVE program, a number of signals in the order of some ten thousands was obtained for most of the active fractions, while after implementation of a threshold of an intensity ratio of >100 between active fractions and control fractions to a few 100 peaks remained for all active fractions.

A general aspect of SIEVE is that the program is base-peak framing m/z values with a reasonable peak shape, although errors occurred where an elevated baseline could be selected, or isotope masses $[M+H+1]^+$ and $[M+H+2]^+$ could be included if the intensity of the peak was enough for selection. Although very time-consuming, it was therefore necessary to check the peaks one by one after selection, both the shape and the m/z values with the same retention time (step 4).

In the Xcalibur, elemental composition tool (step 5) the elements known to be present in environmental toxicants were pre-selected, including nitrogen, oxygen, carbon, hydrogen, fluorine, chlorine, bromine, iodine, silicon, phosphorous, and sulfur. The tool calculates which elemental compositions would fit the accurate masses of the unknown compounds determined by the Orbitrap. The accurate mass is the principal parameter in our identification pipeline. We have designed our strategy for protonated molecular ions because these represent the most dominant ions in the mass chromatograms, but there is a possibility that we are dealing with solvent adducts or adducts containing alkali metals such as Na^+ and K^+ . This phenomenon could lead to erroneous results in the identification of unknown compounds [19].

In step 6, the suggested molecular formulas (minus one $[\text{H}]^+$) are checked for presence in the NIST spectral database, that contains around 200,000 compounds and is an electron impact and highly GC/MS-oriented database. In addition, within the NIST database, there is a smaller LC/MS library available (nist_msms) that contains ~4,000 positive ion spectra, which was also used.

The last theoretical step (step 7) in filtering the interesting compounds from the bulk of non-active compounds was to check whether the compound's K_{ow} matched with the K_{ow} window of the RP fraction. The $\log K_{ow}$ of all candidates was determined with EPI Suite, which has been reported to have a correlation coefficient (r^2) of 0.94 between experimental and calculated values for $\log K_{ow}$'s. A source of error could be the deviation of the K_{ow} window of the fractions (± 0.5) [12]. In order to minimize such

errors, prior to each RP fractionation, a standard solution (known RT and $\log K_{ow}$) was injected to the preparative LC column to check the stability of the system. Still, the presence of matrix could alter the elution and the RT of the compounds on the column during fractionation.

After the first seven steps discussed above, the number of m/z values has decreased from a total of 14,807 to 59 peaks of interest. Since some m/z values can consist of isomers or have multiple possible identities, 95 possible compounds are presented in the Table S1. In conclusion, of the selected 259 m/z values of interest (ratio >100) 59 m/z values were tentatively identified, i.e., 23%.

For confirmation purposes, only 22 of the tentatively identified compounds could rapidly be purchased via major providers. Half of the purchased compounds could be analytically confirmed (by RT, step 8), but none of them was present at such intensity that they were MS/MS-fragmented for identification purposes. Finally, the activity of eight of those compounds could be confirmed in the bioassay (step 9).

When using the PubChem database instead of the NIST library (in step 6) to search for compounds matching the molecular formulas determined in the three selected, most potent fractions (RP3NP2, RP3NP5, and RP3NP6), hardly any significant new results were obtained. Quite unexpectedly, the search function in PubChem showed some practical limitations when using it for identification purposes. In fact, a number of compounds that we identified through NIST searching that could be analytically and biologically confirmed were not found via PubChem searching at all. For mass spectral searching under the umbrella of the PubChem database, only the NIST Chemistry WebBook, which is a separate product from the overarching NIST Spectral Libraries, is covered in PubChem. The NIST Chemistry Webbook contains information on less compounds than the NIST library itself, thereby limiting the applicability identification of unknown toxicants. However, one compound was exclusively found in PubChem and added to the supplementary Table (S1).

Identified compounds

The identified and confirmed compounds described in Table 2 are three polycyclic musks, two organophosphates, two steroids, and one oxygenated polycyclic aromatic compound. Three of them, galaxolide, androstenone, and Tris-(2-chloroisopropyl) phosphate (TCPP), were estimated to be present in a rather high concentration ($>1 \mu\text{g/g}$ sediment) based on peak intensity and detection limit. The compounds androstenone and galaxolide exhibit strong antagonistic AR binding potency in the AR-CALUX[®], having equivalent factor (EF) values of 7.7 and 0.39, respectively. Androstenone is a steroidal pheromone found

in both male and female sweat and urine. It is the active ingredient in Boarmate™, a commercial product sold to pig farmers to test sows for timing of artificial insemination [20]. The simultaneous weak androgenic and strong anti-androgenic potency in the AR-CALUX® observed for androstenedione could most probably be explained by partially agonistic behavior, i.e., despite the activation of the AR there is an additional, dominant pathway leading to anti-androgenic activity [18]. The androstenedione tested here was only one suggested structure from nine (Table S1). It is likely that other isomers are present in the active fraction.

The other compound exhibiting androgenic potency was nandrolone (17 β -19-nortestosterone), an anabolic steroid naturally present in the human body. Nandrolone is administered as an anabolic agent for fattening veal calves and cattle, although the use of hormones for growth-promotion or fattening is prohibited in the EU [21, 22]. The retention time of the purchased nandrolone corresponded to one low-intensity peak within an isomer cluster of the accurate mass. It was not possible to confirm the presence of other androgenic compounds (due to lack of standards) in this study, but indications of several steroids in the tentatively identified compounds list (Table S1) suggest that natural and synthetic steroids are candidates that may be partially responsible for the effects found in the fractions.

Musks, such as the polycyclic musks identified here, are widely used synthetic fragrances. They can be found in almost all consumer products, e.g., perfumes, deodorants, cosmetics, soaps, shampoos, laundry detergents, fabric softeners, household cleaners, and air fresheners [23]. Polycyclic musks are highly lipophilic (log K_{ow} ~6) and are well-known to be emitted into waste-water, reach freshwater, and the marine environment and finally accumulate in sediment, sludge, and biota [24, 25]. High concentrations of galaxolide and tonalide have been reported in sediment from the Berlin area (Germany) strongly polluted with sewage sludge (median concentration of each compound was 0.9 mg/kg dw), whereas non-contaminated sediment contained levels below <0.02 mg/kg [26]. Schreurs and coworkers have shown that galaxolide and tonalide are exhibiting mainly anti-estrogenic activity in both in vitro and in vivo assays (zebra fish) [27]. Both antagonistic ER β and AR antagonistic potency for galaxolide and tonalide (–AR EF 0.25 and 0.20) have been described using the same AR-CALUX® assay as reported here [28].

The oxy-polycyclic aromatic hydrocarbon (PAH) 7 H-benz[*de*]anthracen-7-one (benzanthrone) is widely found in the environment. An important source is its use as a dye intermediate for the synthesis of vat and disperses dyes [29]. Benzanthrone is also detected in atmospheric aerosols and combustion-related particulate emissions [30]. The mutagenic and acute toxic properties of benzanthrone are well reported

due to the fact that it is an industrial chemical and tested in an array of biological and toxicological assays [30]. Only one study has reported benzanthrone levels in sediment from the river Elbe basin at several locations. Benzanthrone was found to be present in one sediment fraction with high algal toxicity at a concentration of 1.2 μ g/g dw [7].

Due to its increased use, the European Commission published an updated risk assessment report on the organophosphate tris-(2-chloroisopropyl) phosphate (TCPP) [31]. Chlorinated alkylphosphate esters, particularly TCPP, have been identified as possible substitutes for the flame retardant pentabromodiphenyl. TCPP has been detected as one of the most abundant compounds in effluents of municipal waste-water treatment plants in Austria and the TCPP levels from the surrounding river sediments were reported to <LOD up to 1,300 ng/g dw [32]. Several studies have shown that the wastewater treatment plants are not able to sufficiently decrease the TCPP levels of the effluent compared with influent [31, 33]. The identified and bioassay-confirmed organophosphates in this study, TCPP, and tris-(ethylhexyl) phosphate are only exhibiting moderately anti-androgenic activity (FLU EF 2E-04 and 1.7E-02). It is likely that other compounds of this class are present in these fractions, based on GC/MS data evaluation [12], where several peaks indicated organophosphates.

Some of the tentatively identified compounds that were not tested here could be suspected to cause androgenic disturbances, such as steroids, pheromones, and fragrances (Table S1), but no literature was found to confirm neither the presence nor the activity of these compounds. Information was searched via the search tool Google, and hits are indicated in the Table S1 for each tentatively identified compound.

The list of androgen-disrupting compounds identified in this study, together with earlier identified (GC/MS) key toxicants in the same fractions [12], consists of PAHs, oxy-polycyclic aromatic hydrocarbon (PAH), nonylphenol isomers, phthalates, organophosphates, musks, and steroids. Interestingly, only a very limited number of compounds were identified by both techniques, which shows that the combination of GC- and LC-based identification techniques is a powerful and complementary approach for the identification of unknown compounds. Of the compounds identified in the same fractions using these complementary techniques, only PAHs and nonylphenol are on the WFD priority list. One phthalate, dibutyl phthalate, is a candidate substance for the priority pollutants list as well as musk xylene, but not the polycyclic musks as identified here [34]. There is currently no regulation regarding the organophosphates in the EU [31]. In the supplement data list (Table S1), there are several candidates presented that may be included in future investigations.

Identification of unknowns: state-of-the-art, limitations, and future aspects

The identification pipeline that we have developed is a sequence of data treatment steps that aims at the reduction of mass signals and results in a list of identified compounds with biologically confirmed activity that is present in the analyzed sediment sample. With regard to the difficulties encountered in the identification of unknown compounds, there is a striking analogy between the fields of EDA and (environmental) metabolomics. Concepts already developed for the identification of unknown compounds from metabolomic origin [19, 35] such as the filtering of molecular formulas obtained by accurate mass spectrometry and procedures for correct deduction of molecular formulas from molecular ions that are not simply protonated parent compounds, could possibly be applied for the improvement of the de novo identification of unknown key toxicants.

In our strategy, several arbitrary thresholds, settings, and assumptions have been included that, without any doubt, lead to the exclusion of valuable information. In other words, we have only studied the metaphorical tip of the iceberg. However, in order to be able to mine the huge amount of available data from non-target analysis and to identify the most abundant key toxicants, these choices were unavoidable. In order to become more widely applicable, the development of an automated workflow instead of the time-consuming step by step treatment of the data is an absolute requirement.

In our work, the use of the NIST database for the identification of unknown compounds may to a greater or lesser extent have biased the outcome of our studies. The NIST database is originally compiled as a GC/MS database and therefore intrinsically not fully suitable as a tool in the de novo identification of environmental pollutants by LC/MS. It seems inevitable that all databases are biased towards a target research area or group(s) of compounds, thereby automatically influencing the outcome of the database searching. For a realistic and feasible, reliable identification of unknown compounds, it is essential to choose the database that includes the information that is most relevant for the field of study or the application area.

Currently, there seems to be a common need in various fields of research including genomics, metabolomics, and environmental analysis for databases that can be consulted for the identification of unknown compounds, varying from DNA fragments to mammalian and environmental metabolites and lipids to environmental key toxicants [36, 37]. Although, apart from the NIST and PubChem, various databases are available, e.g., MassBank and more biologically oriented databases such as KEGG and HMDB, there is still an enormous need for development, improvement, and streamlining of databases for the identification of unknown compounds from various origins.

Here, we present an identification strategy of compounds having an (anti-) androgenic effect using a non-target analysis approach. The next step will be to perform target analysis of the identified compounds to estimate whether the compounds could qualitatively and quantitatively explain the measured effects. This step also includes mixture toxicity issues when adding up the identified compounds in concentrations reflecting the environmental exposure. Finally, the hazard confirmation addresses the question whether the identified compounds pose a risk for the ecosystem. Hazard confirmation takes parameters into account that describe the exposure conditions in the field, e.g., bioavailability, individual and organism differences, and mixture effects. The hazard confirmation, going beyond the analytical and effect confirmation, will result in a realistic picture of the environmental exposure and is highly relevant for risk assessments.

Acknowledgment This work was done in the framework of the EU-supported program MODELKEY (Contract-No. 511237 (GOCE)) and the Marie Curie Research Training Network KEYBIOEFFECTS (MRTN-CT-2006-035695).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Grung M, Lichtenthaler R, Ahel M, Tollefsen KE, Langford K, Thomas KV (2007) Effects-directed analysis of organic toxicants in wastewater effluent from Zagreb, Croatia. *Chemosphere* 67:108–120
2. Thomas KV, Hurst MR, Matthiessen P, McHugh M, Smith A, Waldock MJ (2002) An assessment of in vitro androgenic activity and the identification of environmental androgens in United Kingdom estuaries. *Environ Toxicol Chem* 21(7):1456–1461
3. Urbatzka R, Van Cauwenberge A, Maggioni S, Vigan L, Mandich A, Benfanati E, Lutz I, Kloas W (2007) Androgenic and antiandrogenic activities in water and sediment samples from the river Lambro, Italy, detected by yeast androgen screen and chemical analyses. *Chemosphere* 67:1080–1087
4. Houtman CJ, Booij P, Jover E, Pascual del Rio D, Swart K, van Velzen M, Vreuls R, Legler J, Brouwer A, Lamoree MH (2006) Estrogenic and dioxin-like compounds in sediment from Zierikzee harbour identified with CALUX assay-directed fractionation combined with one and two dimensional gas chromatography analyses. *Chemosphere* 65:2244–2252
5. Grote M, Brack W, Altenburger R (2005) Identification of toxicants from marine sediment using effect-directed analysis. *Environ Toxicol* 20(5):475–486
6. Houtman CJ, Van Oostveen AM, Brouwer A, Lamoree MH, Legler J (2004) Identification of estrogenic compounds in fish bile using bioassay-directed fractionation. *Environ Sci Technol* 38:6415–6423
7. Schwab K, Altenburger R, Luecke-von Varel U, Streck G, Brack W (2009) Effect-directed analysis of sediment-associated algal toxicants at selected hotspots in the river Elbe basin with a special focus on bioaccessibility. *Environ Toxicol Chem* 28(7):1506–1517

8. Thomas KV, Balaam JL, Barnard N, Dyer R, Jones C, Lavender J, McHugh M (2002) Characterisation of potentially genotoxic compounds in sediments collected from United Kingdom estuaries. *Chemosphere* 49:247–258
9. Muir DCG, Howard PH (2006) Are there other persistent organic pollutants? A challenge for environmental chemists. *Environ Sci Technol* 40(23):7157–7066
10. Schymanski EL, Meinert C, Meringer M, Brack W (2008) The use of MS classifiers and structure generation to assist in the identification of unknowns in effect-directed analysis. *Anal Chim Acta* 615:136–147
11. Von der Ohe PC, Kuhne R, Ebert R-U, Altenburger R, Liess M, Schüürmann G (2005) Structural alerts—a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute Daphnid assay. *Chem Res Toxicol* 18:536–555
12. Weiss JM, Hamers T, Thomas KV, Van der Linden S, Leonards PEG, Lamoree MH (2009) Masking effect of anti-androgens on androgenic activity in European river sediment unveiled by effect-directed analysis. *Anal Bioanal Chem* 394:1385–1397
13. Katajamaa M, Miettinen J, Oreai M (2006) Toolbox for processing and visualization of mass spectrometry based molecular profile data. *BMC Bioinform* 22(5):634–363
14. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779–787
15. Heinonen M, Rantanen A, Mielikäinen T, Kokkonen J, Kiuru J, Ketola RA, Rousu J (2008) FiD: a software for *ab initio* structural identification of product ions from tandem mass spectrometric data. *Rapid Commun Mass Spectrom* 22:3043–3052
16. Kamleh A, Barret MP, Wildridge D, Burchmore RJS, Scheltema RA, Watson DG (2008) Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. *Rapid Commun Mass Spectrom* 22:1912–1918
17. Listgarten J, Emili A (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatograph-tandem mass spectrometry. *Mol Cell Proteom* 4(4):419–434
18. Sonneveld E, Jansen HJ, Riteco JAC, Brouwer A, Van der Burg B (2005) Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. *Toxicol Sci* 83:136–148
19. Draper J, Enot DP, Parker D, Beckmann M, Snowdon S, Lin W, Zubair H (2009) Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive *m/z* annotation tool utilising predicted ionisation behaviour 'rules'. *BMC Bioinform* 10(227):1–16
20. DuPont. http://www2.dupont.com/DAHS_EMEA/en_GB/products/spec_prod/boarmate.html Accessed 25/8 2010
21. Cacciatore G, Eisenberg SWF, Situ C, Mooney MH, Delahaut P, Klarenbeek S, Huet A-C, Bergwerff AA, Elliott CT (2009) Effect of growth-promoting 17 β -estradiol, 19-nortestosterone and dexamethasone on circulating levels of nine potential biomarker candidates in veal calves. *Anal Chim Acta* 637:351–359
22. Van der Merwe PJ, Pieterse JW (1994) Stability of zeranol, nandrolone and trenbolone in bovine urine. *Analyst* 119:2651–2653
23. Reiner JL, Kannan K (2006) A survey of polycyclic musks in selected household commodities from the United States. *Chemosphere* 62:867–873
24. Bester K (2009) Analysis of musk fragrances in environmental samples. *J Chromatogr A* 1216:470–480
25. Rüdell H, Böhmer W, Schröter-Kermani C (2006) Retrospective monitoring of synthetic musk compounds in aquatic biota from German rivers and coastal areas. *JEM* 8:812–823
26. Fromme H, Otto T, Pilz K (2001) Polycyclic musk fragrances in different environmental compartments in Berlin (Germany). *Water Res* 35(1):121–128
27. Schreurs RHMM, Legler J, Artola-Garicano E, Sinnige TL, Lanser PH, Seinen W, Van der Burg B (2004) *In vitro* and *in vivo* antiestrogenic effects of polycyclic musks in Zebrafish. *Environ Sci Technol* 38:997–1002
28. Van der Burg B, Schreurs RHMM, Van der Linden S, Seinen W, Brouwer A, Sonneveld E (2008) Endocrine effects of polycyclic musks: do we smell a rat? *Int J Androl* 31:188–193
29. Dwivedi N, Das M, Khanna SK (2001) Role of biological anti-oxidants in benzanthrone toxicity. *Arch Toxicol* 75:221–226
30. Spitzer T, Takeuchi T (1995) Determination of benzanthrone in environmental samples. *J Chromatogr A* 710:109–116
31. EC (2008) European Union Risk Assessment Report—tris(2-chloro-1-methylethyl) phosphate (TCPP).
32. Martinez-Carballo E, Gonzales-Barreiro C, Sitka A, Scharf S, Gans O (2007) Determination of selected organophosphate esters in the aquatic environment of Austria. *Sci Total Environ* 388:290–299
33. Meyer JA, Bester K (2004) Organophosphate flame retardants and plasticisers in wastewater treatment plants. *J Environ Monit* 6:599–605
34. EC (2008) European Union Risk Assessment Report - 1,3,4,6,7,8-Hexahydro-4,6,6,7,8,8-Hexamethylcyclopenta- γ -2-Benzopyran (HHCB).
35. Kind T, Fiehn O (2007) Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform* 8(105):1–20
36. Bundy JG, Davey MP, Viant RM (2009) Environmental metabolomics: a critical review and future perspectives. *Metabolomics* 5:3–21
37. Schymanski E, Bataineh M, Goss KU, Brack W (2009) Integrated analytical and computer tools for structure elucidation in effect-directed analysis. *TrAC* 28(5):550–561