Mass Spectrometry in Environmental Toxicology

Ksenia J. Groh and Marc J.-F. Suter

Abstract: In environmental toxicology, mass spectrometry can be applied to evaluate both exposure to chemicals as well as their effects in organisms. Various ultra-trace techniques are employed today to measure pollutants in different environmental compartments. Increasingly, effect-directed analysis is being applied to focus chemical monitoring on sites of ecotoxicological concern. Mass spectrometry is also very instrumental for studying the interactions of chemicals with organisms on the molecular and cellular level, providing new insights into mechanisms of toxicity. In the future, diverse mass spectrometry-based techniques are expected to become even more widely used in this field, contributing to the refinement of currently used environmental risk assessment strategies.

Keywords: Aquatic organisms - Ecotoxicology - Effect-directed analysis - Proteomics - Ultra-trace analysis

Environmental research has for a very long time been focusing on acute toxic effects, trying to find sources of contaminants or dealing with accidental exposure. The negative effects of extensive pesticide use, especially DDT,[1] the Seveso incident in 1976,[2] or ten years later the fire in Schweizerhalle[3] made an incredible impact on human awareness of the detrimental effects of heavy pollution and major accidents. It also boosted sales of high-resolution mass spectrometers needed to sensitively quantify all dioxin isomers and congeners. Since then it has been recognized that it is important to understand the fate and behavior of anthropogenic contaminants, because the exposure situation in an ecosystem is very complex. Not only is the exposure variable, with accidents being the most extreme case, but effects of chronic exposure to low concentrations have to be sensitively quantified as well. Perchloroethylene exposure from contaminated drinking water sources is one classical case where human health is directly affected, resulting in cancer.[4] Furthermore, aquatic organisms are rarely challenged with one pollutant alone, but rather with a chemical mixture resulting from wastewater treatment plant (WWTP) effluent discharge, diffuse inputs such as run-off from urban and agricultural surfaces, dry deposition, remobilization from polluted sediments, or even infiltration of contaminated groundwater into surface waters. Accidents play a minor role in this case and exposure is chronic because inflow is continuous. Therefore today's challenge of environmental toxicology is to understand the effects of chemical mixtures in combination with physical and biological stressors. This requires very sensitive and accurate tools in order to be able to determine the composition and dynamics of the chemical cocktail. For this, mass spectrometry today is the method of choice for determining organic trace contaminants at very low concentrations in various environmental compartments.

Depending on the physico-chemical properties of the analytes, different enrichment and separation techniques are used prior to detection. For instance, GC-MS is still the most widely used method for dioxins, PCBs, PAHs and similarly lipophilic pollutants. Analysis of thermo-labile, polar and ionic compounds, including diverse pesticides, personal care products and pharmaceuticals, became accessible with the advent of electrospray ionization in the 1990s.[5] Many studies have been published since then, investigating the fate and behavior of these compounds in various environmental systems such as WWTPs, which are major point sources of contamination if elimination of a particular chemical is incomplete,[6] or sediments and surface waters.[7]

As of December 2013, more than 72 million products are listed as being commercially available, of which 300'000 plus are inventoried/regulated substances.[8] This is a huge number of compounds that can potentially end up in the environment and cause adverse effects in wildlife. The situation gets even worse when considering the fact that their microbial transformation products might be biologically active as well.[9] Hence, when trying to link an observed effect to causative agents, they chemical, physical or even biological, scientists are faced with the problem of finding the needle in a haystack.[10] Two complementary strategies can be applied to determine the exposure situation in a complex environmental matrix, i) high-throughput chemical multi-residue analysis,[11] or ii) biological effects assessment using either in vitro or in vivo assays, ideally followed by identification of the causative chemical agent.

The limitation of the chemical analysis is, as mentioned above, that no unified method exists which would allow analysis of the universe of chemicals with high sensitivity for all different compound classes simultaneously. Hence, depending on the target analyte, different extraction and enrichment techniques will have to be applied, followed by a suitable separation, in order to reduce the complexity of the sample. Recently, we developed a target-ed approach to simultaneously determine chemicals from different substance classes (ionic, polar, lipophilic) known to interfere with the gluco- and mineralocorticoid signaling pathway, using a single step multi-mode solid-phase cartridge.[12]

The number of analytes measured with a given mass spectrometric technique directly affects sensitivity, because the more compounds are monitored, the longer the
cycle time will be and thus the shorter the
time spent on acquiring one specific target,
resulting in lower sensitivity. Nonetheless,
today’s high resolution and highly ac-
curate mass spectrometers (orbitraps and
time-of-flight instruments), together with
novel approaches such as scan-dependent
and -independent experiments, open up
entirely new possibilities. Moschet et al.
for instance showed that using LC coupled
to high-resolution MS allows the identifi-
cation of known pesticides based on their
exact mass only, an approach termed sus-
pect screening.[13] In order to reduce cycle
times, scan-dependent experiments per-
form single reaction monitoring (SRM) or
MS/MS experiments only when a precur-
sor ion from a target list is found in a full
scan, thus reducing the time lost by moni-
toring uneventful channels in a classical
SRM experiment.

One important drawback of any chemi-
cal analysis is that all samples will have
to be analyzed, and even then their po-
tential to cause adverse effects in organ-
isms is unknown unless ecotoxicological
data is available. Biological effects as-
sessment on the other hand will directly
identify samples with a positive response
in a given in vivo or in vitro assay, allow-
ing the chemical analysis to be focussed on
relevant samples only. The caveat here is
that unless an integrative whole organism
testing is applied, the assay used will only
test for the molecular effect it was designed
for. For example, the well-known estro-
gen screen (YES)[14] will provide the
information only for estrogen receptor
binding (and cytotoxicity, a non-specific
side-effect when using cell-based assays).
This disregards any disturbance of the ste-
roidogenesis, although this pathway may
lead to detrimental estrogen balance-relat-
ed effects as well. Hence, ecotoxicologists
are faced with the challenge of having to
select from a multitude of assays covering
different molecular and physiological end-
point. In the ToxCast program initiated by
the US Environmental Protection Agency,
hundreds of chemicals are currently being
tested in various biological assays cover-
ing diverse molecular and cellular toxicity
pathways, in an effort to build decision
support tools based on in vitro screening
results to help prioritize chemicals for fur-
ther investigation.[15]

The combination of biological assays with
chemical analysis is a very power-
ful tool to reduce the number of samples
having to be analyzed, allowing efforts to
be focussed on ecotoxicologically sig-
nificant cases. The procedure has been
around in various forms since the 1980s.
Effect-Directed Analysis (EDA), the strat-
 egy proposed by Werner Brack (Fig. 1),
starts with the extraction of an environ-
mental sample and testing of the extract in
a biological assay.[16] Active samples are
then fractionated and the resulting frac-
tions again assessed for their biological
activity. Positive fractions are then chemi-
cally analyzed using both target analysis
and screening for unknowns to identify
the active components in the fraction. A
validation of the identified suspects in
the biotest with commercially available
or synthesized reference standards fin-
ishes the procedure that, as one scientist
states in a review of the field, unfortun-
ately often leads to disappointment.[17]
The reason this approach can fail is that
while a multi-residue screening allows the
identification of target analytes, identifi-
cation of an unknown active component
poses a remarkable challenge. Both scan-
dependent and scan-independent methods
will provide accurate mass information
for the precursor and its fragments, but a
multitude of chemical formulas can add up
to the molecular weight of the precursor
ion. For instance, with 10 ppm instrument
accuracy, more than 75 sum formulas are
possible for a mass around 400 Da.[18] This
number goes down if filtering rules are
applied, however, the sum formula alone
does not define the structure. MS/MS in-
formation has to be used to identify func-
tional groups and substructures, followed
by validation of the hypothetical structure,
by use of reference compounds (if avail-
able). Nevertheless, successful identifica-
tion of the active chemical by EDA has
been achieved.

One such example was the identifica-
tion of estrogenic compounds in WWTP
effluents and receiving rivers. In this
study we determined concentrations of es-
trone (E1), estradiol (E2), ethinylestradiol
(E2E), and the xenoestrogen nonylphenol
(NP) using GC-MS and calculated the cor-
responding estrogenicity based on their
relative potencies assessed in the YES.
A good correlation was found between the
measured and calculated estrogenicity.[19]
In another example, estradiol was identi-
fied as predominant estrogen in UK rivers
also using the YES.[20] The zebrafish em-
bryo toxicity assay (zFET) has recently
been proposed as a useful alternative to
rapidly evaluate whole organism responses
to chemical challenge.[21,22] Indeed, using
zFET in an EDA of soil samples collected
from a former municipal landfill site, a
number of previously unknown develop-
mental toxicants were identified.[23]

We have also used EDA in an attempt
to understand the causes for gonad malfor-
mations observed in whitefish (Coregonus
sp.) from Lake Thun, Switzerland.[24] For
this purpose environmental samples from
the food chain (zooplankton) of these fish
were taken and investigated for their abil-
ity to induce gonadal malformations as well
as for their estrogenic potential. In
one experiment where whitefish hatchlings
from Lake Thun were raised during 3-4
years feeding on the collected zooplank-
ton, evidence was obtained that zooplank-
ton from Lake Thun is the key factor in the
development of gonadal malformations.[25]
Zooplankton extracts also gave a positive
signal in the YES, and hence extracts from
Lake Thun and the reference Lake Brienz
were investigated. This lake is located
directly upstream of Lake Thun and has
normal incidences of malformations. The
samples were separated into 10 fractions
on a C18 column. Only fractions 1, 3 and 6
from Lake Thun showed estrogenic ac-
ivity, while none did from Lake Brienz.[26]
In order to rule out the most likely estro-
gens, all samples were analyzed for their
E1, E2, EE2 and NP content using an es-
stablished LC-MS/MS method,[27] but none
contained any traces of these estrogens.
In order to identify the unknown estrogenic
compound(s), differential scan-dependent
MS/MS analyses were performed on the
Lake Thun estrogenic fractions and the
corresponding fractions from the reference
lake. Using this approach, one compound
present only in the Lake Thun extract
was found, with a protonated molecular
ion [M+H]+ of 180.1018. The associated
isotope peaks clearly excluded elements
Cl and Br. With a mass accuracy of 2
ppm, easily achieved on the orbitrap XL
(Thermo Scientific) used in this case, and
the elements H, C, N, O, F, Si, P, S and I se-

---

**Fig. 1. Schematic representation of effect-directed analysis (EDA) adapted from W. Brack.[5]**
lected, only three candidate sum formulas were possible at this low mass, those being $C_{12}H_{18}O_6N$, $C_{14}H_{23}O_6N_2F$ and $C_{14}H_{27}NF_2Si$. Since most estrogenic compounds have a phenolic substructure, the latter two were ruled out because of their low ring and double bond equivalents (1.0 and 0.5 respectively).

Assigning a structure to a sum formula is not trivial since it requires some prior knowledge, e.g. structure of the precursor of a transformation product, or substructures from a structure–activity investigation. When starting with a phenolic substructure, one possible candidate is N-acetyltyramine (CAS 1202-66-0). With an estimated log $K_w$ of 1.02,[28] it would be expected to elute ahead of E2 (log$K_w$ 4.01) which ends up in fraction 6, thus supporting the proposed structure. N-acetyltyramine was then synthesized in-house and tested in the YES. Since it was not estrogenic in the YES, it had to be rejected as potential causative agent for the gonad malformations observed. Nevertheless, the Lake Thun investigation nicely illustrates how EDA helps narrowing down the search for unknown compounds causing adverse effects in the ecosystem.

The EDA approach described above relies on established bioassays measuring specific responses. However, many toxicity pathways, especially those activated through low level chronic exposures, are currently incompletely understood and consequently there are no suitable bioassays available yet. To develop novel biomarkers one strategy is to focus on molecular modes of action. Exposure of organisms to chemical stressors results in the activation of defense mechanisms. When these are overwhelmed, damage of cells and organs ensues as a final manifestation of chemical toxicity. Any cellular and physiological changes observed in response to exposure are accompanied by molecular and biochemical alterations in the organism. Gene transcription rates can be directly affected by various compounds. Such is the case for many endocrine disrupting chemicals that act through binding to nuclear receptors, e.g. estrogen receptor or aryl hydrocarbon receptor, which function as transcription factors regulating the expression of responsive genes.[29–31] Complex interactions can also occur at post-translational level, leading, for example, to a change in the enzyme activity.[32,33] Moreover, some chemicals can also directly interact with cellular constituents and metabolites, thus affecting the biochemical pathways and cellular functions by non-genomic mechanisms.[34–36]

During the last decade, diverse ‘-omics’ technologies that allow a high-throughput characterization of gene expression and cellular metabolites, have been developed. Application of ‘-omics’ in environmental toxicology holds a great promise, because detailed knowledge on the molecular changes induced by chemical exposure may provide valuable insights into the mechanisms of toxicity and support the search for biomarkers of exposure and effects.[37–39]

Transcriptomics characterizes the composition and abundance of mRNA transcripts in the specimen. Such analyses are conventionally performed with use of microarray technology and, more recently, by diverse next-generation sequencing-based approaches.[40] Vast amounts of information on the gene expression patterns characteristic of normal development, as well as evolving in response to diverse stressors, have been obtained with use of transcriptomics.[40–43] However, it is the protein and not the transcript, that carries out the cellular functions of a particular gene in most of the cases. Thus, only the information on the protein expression alterations may be regarded as a reliable link between the chemical-induced gene expression changes and perturbed biological functions. Conventionally, mRNA levels have been taken as a proxy for the expression of the corresponding proteins. However, it has been convincingly demonstrated that transcript levels are rather poor indicators of the abundance of the corresponding proteins,[44–46] since protein expression can also be regulated on post-transcriptional, translational and post-translational levels.[47,48] Therefore, in order to understand the meaning and significance of gene expression changes induced by pollutants for the overall functioning of the organism under investigation, the information obtained by transcriptomics needs to be complemented by the characterization of the respective proteome. This is addressed by proteomics.[49,50] Going yet one level further, metabolomics deals with characterizing all low-molecular-weight (<1000 Da) primary and secondary metabolites found in the organism, which ultimately represent the most ‘functional’ among all ‘-omics’ targets, since metabolites are the end products of the cellular regulatory processes.[51] Metabolomics data in combination with proteomics may also help to unravel the roles played by the genes with yet unknown functions.[52] Both proteomics and metabolomics analyses largely rely on diverse MS-based approaches.[53,54] Below, we will provide a few specific examples of the diverse MS applications employed in our group in this regard.

Two main approaches to study the proteome can be distinguished, gel-based and gel-free. In the first one, proteins are usually separated and quantified by two-dimensional gel electrophoresis (2D-GE), followed by excision of spots of interest and MS-based identification of respective proteins, either intact or through peptides generated by digestion.[55] In an alternative, gel-free, approach, also called shotgun proteomics, proteins are directly digested and the resulting peptides are separated by LC, followed by MS analysis. Peptide separation can be performed either offline or online. We have optimized an online setup, Multidimensional Protein Identification Technology (MudPIT), initially developed for yeast,[56] to study the proteome of zebrafish, a model teleost species. Here, peptides are separated by 2D-LC and directly sprayed into an orbitrap mass spectrometer performing MS/MS analysis.

Using MudPIT to study the proteome of mature zebrafish gonads, much more populated protein lists than those previously delivered by gel-based approaches could be obtained. Several novel protein groups expressed in a sexually dimorphic pattern in zebrafish gonads were identified,[46] demonstrating that global proteome characterization can be a useful source of initial information for defining the molecular pathways in a specific context and for identifying the genes that carry out particular functions. However, one drawback of global proteomics is its general bias towards higher abundance proteins. This happens because in global proteomics one attempts to measure all peptides present in the mixture. Due to the stochastic nature of scan-dependent acquisition, more abundant peptide species generated from higher abundance proteins, are more likely to be selected and analyzed during the run. As a consequence, less abundant peptides are only rarely identified in a particular MudPIT run, and thus low abundance proteins cannot be reliably studied by global profiling. This may limit the application of this technique to the study of molecular mechanisms of toxicity in a complex organism such as zebrafish, since only a small proportion of all the proteins whose expression is altered in response to a stressor would be identified. The interpretation of the results of differential proteomics studies performed with global approaches should be done with caution to avoid misleading conclusions based on the observed variations in high abundance proteins. Many changes commonly observed on the proteome levels could in fact represent only general stress responses, while for the elucidation of toxicant-specific responses the degree of proteome coverage currently delivered by global proteomics may be insufficient.[57,58] Nonetheless, a number of useful insights into the mechanisms of toxicity can be obtained with MudPIT. For example, it was found that exposure to ionic silver and silver nanoparticles (AgNP) induces distinct protein profile changes in zebrafish em-
bryos, demonstrating the similarities and differences between the actions of these two toxic agents.[69]

The sensitivity of an MS-based analysis can be improved by using the data acquisition approach aimed at detecting specific proteins of interest, so-called targeted proteomics.[60] SRM, a technique initially developed for ultra-trace chemical analysis, can be adapted to proteomics. Here, only specific fragmentation transitions for chosen peptides representative of targeted proteins are monitored,[61] which lowers the sensitivity of MS detection to be enhanced.[62] In this workflow, first a set of proteins to be monitored is selected and proteotypic peptides for each protein are chosen based on global profiling data and/or in silico prediction. SRM conditions for detection of each peptide species can then be optimized either directly for the endogenous peptides or by using synthesized peptides first. The developed SRM protocols are then used to study respective endogenous peptide signal changes. Since SRM analysis typically takes one hour and up to several hundred targets can be monitored within one run in scheduled acquisition mode, this approach allows for much higher throughput compared to MudPIT, where one technical replicate requires up to 22 hours instrument time. The quantification in SRM can be done by using isotopically labeled counterparts for each peptide target.[63] This method, although providing more accurate quantification data, is at the same time more expensive, which may limit its broader application in environmental toxicology research during the initial screening stages in biomarker search projects. We have recently demonstrated that semi-quantitative data can be obtained with SRM by using a few isotopically labeled standards to evaluate MS conditions in different runs and performing a comparison of relative abundance of target proteins across samples by normalizing them to the levels of ‘housekeeping’ proteins measured in the same run, similar to a normalization approach used for PCR-based mRNA quantification. This setup provides a relatively inexpensive means to perform initial screening of proteins of interest across many different samples obtained from specimens exposed to different chemical concentrations at different time points. Selected proteins demonstrating promising expression patterns, e.g. large changes in response to a particular stressor, can then be studied in more detail and precisely quantified using labeled standards. Using the described semi-quantitative SRM approach, time-resolved profiles of candidate sex-related genes in zebrafish during the period of gonad differentiation were generated.[64] As expected, SRM increased the sensitivity of MS detection, allowing several low-abundance proteins unseen during global profiling in adult gonads to be monitored.[65] This allowed the confirmation of protein expression for several genes previously characterized on the mRNA level or by bioinformatic prediction only. For several of the studied proteins, the patterns of protein expression during gonad differentiation confirmed or suggested a possible role in sexual differentiation processes in zebrafish, pinpointing them as candidates for further studies.[66] Currently, targeted proteomics is being used to study the expression of zebrafish proteins potentially responsive to glucocorticoid-like compounds, which may allow identification and establishment of sensitive biomarkers of glucocorticoid exposure for application in EDA of environmental samples.

Even using SRM, several low-abundance proteins of particular interest for sexual differentiation in zebrafish still could not be observed, what was likely due to a very low abundance of these proteins, typically expressed in only a few specific cell types.[67] Monitoring such proteins in the complex samples generated from several organs or whole body extracts indeed presents a serious challenge at the moment. To further improve the detection of such proteins, the complexity of analyzed protein mixtures can be reduced by using diverse fractionation or enrichment techniques.[68] However, such methods are associated with extensive sample processing steps, introducing increased variation between technical replicates and usually requiring increased amounts of starting material, which may present a challenge for ecotoxicological studies. Alternatively, specific cell types where the expression of a protein of interest is expected can be isolated, for example by laser capture micro-dissection.[69] The latter method, however, still needs further development to establish its compatibility with downstream MS analysis.

A recently developed scan-independent method called SWATH combines the sensitivity of SRM with the screening capabilities of global proteomics.[70] Here, all analytes coming off the LC column are activated and all fragment ions acquired afterwards. This is done by increasing the precursor ion selection window (swath) to e.g. 25 Da instead of typically one for SRM, activating all compounds found in this window, then acquiring all fragments from 400–1200 m/z. The 25 Da window is then moved to higher mass and the process repeated until the full mass range has been covered. The fragments of a precursor ion are then found by matching their chromatographic profile to the precursor. One very important aspect of this approach is that fragment ions from all precursors are available, because there is no bias caused by thresholds or the stochastic nature of scan-dependent selection of precursors. SWATH data can thus be mined at later times for other compounds that were not in the focus of the initial study! So far, SWATH has been successfully applied to biomarker discovery in clinical research.[71] This technique certainly holds a great potential for environmental research, when applied to both pollutant screening and molecular effects characterization.

The workflows for global proteomics analysis by MudPIT and targeted proteomics analysis by SRM established for zebrafish have been applied to study the proteome of other species relevant for ecotoxicology. The use of global proteomics analysis for studies of stress response mechanisms in unicellular species, such as bacteria, cell lines or green algae, is a particularly promising approach, since in such organisms typically a higher proteome coverage can be obtained as compared to multicellular ones.[50] For example, we have investigated the proteome changes induced by the herbicides paraquat, diuron, and norflurazon in the green alga Chlamydomonas reinhardtii.[69] Here, protein profiles were generated by MudPIT and label-free quantification based on spectral counting in combination with G-test statistics was performed to identify proteins exhibiting significant abundance changes. With this approach, 149–254 differentially expressed proteins could be identified, involved in a variety of metabolic pathways. This work demonstrated that proteome responses to toxicants are more sensitive than the physiological and biochemical endpoints characterized previously,[72] establishing global proteome profiling as a valuable method for studying mechanisms of toxicity in this species.[69] Recently, combined transcriptomics and proteomics analyses have allowed detailed insights into the mechanisms of silver toxicity in C. reinhardtii, demonstrating that the initial damage caused by silver through its effects on ATP and photosynthesis, leading to promotion of oxidative stress, is counteracted by the cells through the activation of antioxidant defenses as well as possible elimination of silver via efflux transporters.[71] MudPIT can also be used to obtain precise information on protein synthesis and degradation dynamics.[72] This is done by pulse-labeling with isotopically labeled substrate. It was observed that the time course of diuron-induced degradation of the D1 protein, which is an important photosystem II component, can be accurately followed by performing MudPIT on samples of proteins labeled when algae are growing in 15N-enriched nutrient solution.[73] This method constitutes a valuable addition to label-free quan-
tification approaches, expanding the array of proteome characterization methods. Presently, the analysis of the proteome of non-sequenced or partially sequenced species presents a particular challenge. Nonetheless, MudPIT can be applied to obtain the initial proteome characterization in non-target species as well, by using databases from related species. For example, to decipher the mechanisms of acid resistance in frogs, we have studied the proteome in the egg shells of the frog *Rana arvalis* using the Xenopus laevis (fully sequenced species) database as a reference [35, 84].

Proteins sequenced in the public databases for a given species are also a useful source of reference information for initial characterization. For the mussel *Mytilus galloprovincialis*, by using databases from related species, we have studied the changes in the protein and chemical-protein interactions, systems biology models integrating transcriptional, proteomic and metabolomic information, as well as data on protein-protein and chemical-protein interactions, will provide an improved understanding of molecular mechanisms underlying the adverse outcomes resulting from chemical exposure. These advancements will strengthen the current strategies applied in environmental risk assessment to evaluate and predict the hazards and risks posed by chemicals in the environment.

### References

Received: February 4, 2014
