Joint Research Programme
BTO 2021.008| Mei 2021

Functional omics techniques for drinking water quality monitoring

Joint Research Programme

KWR

Bridging Science to Practice

Report

Functional omics techniques for drinking water quality monitoring

BTO 2021.008 | Mei 2021

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045/202

Project manager
Patrick Bauerlein

Client
BTO - Verkennend onderzoek

Author(s)
Tessa Pronk, Anne Zwartsen, Nienke Meekel, Milou Dingemans

Quality Assurance
Thomas ter Laak

Sent to
This report is distributed to BTO-participants.

A year after publication it is public.

Keywords

Bioassay, transcriptomics, metabolomics, proteomics, monitoring, water quality, effect based method, micropollutants

Year of publishing
2021

More information
Tessa Pronk
T +31(0)629337806
E tessa.pronk@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

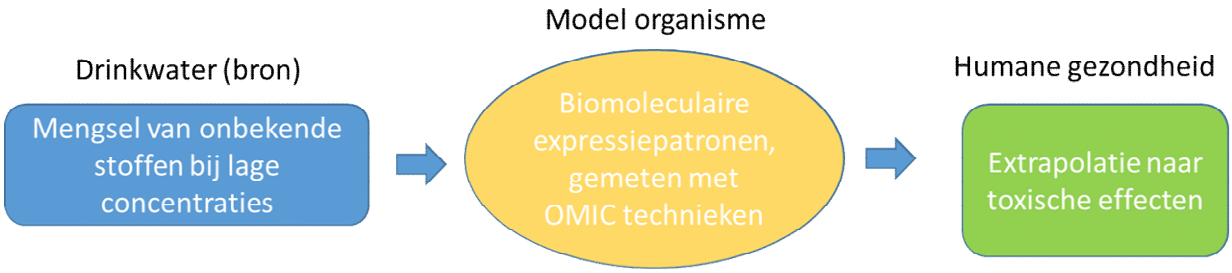
T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

Managementsamenvatting

Bioassays gebaseerd op functionele omics technieken zijn de moeite waard om verder te ontwikkelen voor het gebruik bij het monitoren van drinkwaterkwaliteit

Auteurs Tessa Pronk PhD, Anne Zwartsen PhD, Nienke Meekel MSc, Milou Dingemans PhD

De voordelen van het gebruik van bioassays op basis van functionele omics technieken maken het de moeite waard om een specifiek raamwerk op te gaan bouwen voor de beoordeling van de mengseltoxiciteit in drinkwater(bronnen). Dit blijkt uit een studie naar de stand van zaken rond functionele omics technieken voor het monitoren en duiden van de drinkwaterkwaliteit. Functionele omics is een verzamelnaam voor een aantal biochemische technieken (transcriptomics, proteomics, metabolomics) die veel biomoleculen tegelijk kunnen meten die een rol spelen in de mechanismen waarmee een organisme reageert op veranderende omstandigheden. Hiermee geven deze technieken inzicht in de mogelijke toxische effecten van stoffen op moleculair niveau. Door te analyseren op moleculair niveau zijn -omics technieken zeer gevoelig, zodat effecten na kortdurende blootstelling aan lage stofconcentraties kunnen worden gedetecteerd, zelfs wanneer nog geen fysieke toxische effecten te verwachten zijn. In een modelorganisme kunnen deze technieken de invloed van aanwezige stoffen op veel verschillende biologische mechanismen tegelijk vaststellen. Dit maakt functionele omics technieken een interessante kandidaat om individuele traditionele bioassays – die vaak een enkel eindpunt meten – te vervangen. Daarbij is het voor het inschatten van een risico wel belangrijk de reactie van biomoleculen op een stof te linken is aan het fysieke humaan toxische eindpunt dat met deze stof is geassocieerd. De grootste hindernis voor de toepassing van -omics technieken op de monitoring van de drinkwater(bronnen) kwaliteit is dan ook het gebrek aan onderzoek naar de expressie van biomoleculen in relatie tot het toxisch eindpunt dat wordt geïnduceerd door chemicaliën.



Visualisatie van het concept om na blootstelling van een modelorganisme aan water met een mengsel aan onbekende stoffen bij lage concentraties op basis van de functionele omics data mogelijke risico's voor de humane gezondheid te voorspellen. Met gericht onderzoek kan dit concept worden gerealiseerd.

Belang: breder testen op effecten met gedetailleerde meetresultaten

Vanwege de veelheid van mogelijke effecten die mengsels van stoffen bij lage concentraties op moleculair niveau kunnen veroorzaken, is het niet waarschijnlijk dat deze allemaal in traditionele bioassays kunnen worden gedetecteerd. Functionele omics technieken bieden een meer gedetailleerde, op mechanismen gebaseerde diagnose over toxiciteit.

Methode: een advies voor een aanpak gebaseerd op recente literatuur

In dit rapport onderzoeken we de stand van zaken rond het testen van humaan relevante toxiciteit in water met functionele omics technieken en bevelen

we op onderwerpen (methode, modelorganisme, raamwerk) een aanpak aan.

Resultaten: aanbevelingen voor een bioassay voor waterkwaliteit gebaseerd op functionele omics

Het aanbevolen raamwerk is gebaseerd op een dosis respons curve voor de expressie van biomoleculen die geassocieerd zijn met diverse 'modes of action' die op hun beurt weer indicatief zijn voor toxische eindpunten. Dit wordt gecombineerd met een 'trigger value' om risico's te kunnen beheersen. Het raamwerk zal gestandaardiseerd moeten zijn om variabiliteit in resultaten te voorkomen. De biomoleculen die gemeten worden met metabolomics (metabolieten) zijn met relatief weinig ruis te meten, zijn vaak hetzelfde tussen organismen en staan relatief dicht bij een fenotype waardoor de voorspellende waarde van een raamwerk met deze techniek naar verwachting robuust kan zijn. *Daphnia* en Nematoden zijn veelbelovende modelorganismen omdat ze relatief gevoelig zijn voor toxische stoffen, met weinig middelen in een laboratorium gekweekt kunnen worden, en potentieel veel verschillende reacties op toxiciteit kunnen laten zien.

Implementatie: raamwerk voor het bepalen van het effectpotentieel van stofmengsels

Het onderzoeksveld rond functionele omics ontwikkelt zich nog volop en er is veel potentie voor de inzet hiervan in een universele bioassay. Op dit moment kan er nog geen link gelegd worden tussen gemeten biomoleculen via -omics technieken en relevante toxiciteit maar de mogelijkheid hiertoe zal wel komen, naar verwachting tussen 5-10 jaar. Op dit moment worden in bestaande onderzoeken vooral expressiepatronen van biomoleculen voor bekende effecten vastgesteld, zoals de 'mode of action' van stoffen, en wordt gekeken naar mogelijke verbanden met effecten in ecotoxicologische modelorganismen. Onderzoek naar de voorspellende waarde voor de aanwezigheid van stoffen met humaan relevante toxicologische effecten is nog schaars. Voordat sprake kan zijn van implementatie zijn dan ook gerichte analyses en experimenten nodig die vaststellen wat het diagnosticerend vermogen van de voorgestelde methode is voor stoffenmengsels met humaan relevante toxicologische effecten.

Het Rapport

Dit onderzoek is beschreven in het rapport *Functional omics techniques for drinking water quality monitoring* (BTO-2021.008).

Dit rapport is gerelateerd aan: Schriks M., van der Oost R., Houtman C. (2011). *The application of toxicogenomics for (drinking) water quality assessment*. KWR Water Research Institute (BTO 2011.049)

More information
Anne Zwartsen
T +31(0)652092737
E anne.zwartsen@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

Contents

1 Introduction

Clean and safe drinking water is of the highest importance for the global population. However, water quality is in danger of being affected by the large number of chemicals produced and used which may ultimately end up in the environment (Kools et al., 2019; van Wezel et al., 2017). In addition, long periods of droughts decrease the volume of water bodies, concentrating the chemicals therein (Kümmerer et al., 2018).

To guard the quality of surface water and drinking water, drinking water companies and regional water authorities traditionally monitor waterbodies using targeted chemical analyses. Using these methods, the concentrations of on selected target chemicals are determined and evaluated against (drinking) water quality guidelines. Of the thousands of chemicals in water, targeted analyses only detect a limited number of the chemicals as specific methods have to be run for each chemical (group) of interest. However, the occurrence of chemicals in water sources beyond the regulated list of targeted chemicals needs to be monitored to ensure safe drinking water. To be able to get a better view on water quality, suspect and non-target screening methodologies were developed in which many chemicals can be detected based on comparable chemical properties (known registered chemicals, as well as unknown compounds) using high resolution mass spectrometry (HRMS), even if these chemicals were not identified yet. Identification of the suspects is possible, but it is time and cost consuming to select promising patterns to further identify among the many thousands possible. If suspects are identified, chemicals can be evaluated in terms of target concentrations such as health-based guideline values (hGLVs) (Baken et al., 2018). If these are not available, concentrations can be compared to generic signalling values, and drinking water companies can take action if these values are exceeded. These guideline values are based on the toxicological effects of individual chemicals in official guideline studies. However, because in practice there are many different low-level concentration chemicals present in water it is difficult to detect all with methods for non-target screening (NTS) or target screening (TS). Also, for many emerging chemicals there is a lack of toxicological information, as well as on potential mixture toxicity (EFSA scientific committee, 2019). Therefore it is not possible to monitor, identify and characterize the hazards of all chemicals detected in water.

Water quality managers are challenged to assess the presence and effects of large numbers of chemicals in water. Therefore, chemical analyses are more often combined with effect measurements using *in vitro* bioassays (Brunner et al., 2020; Dingemans et al., 2019). Using such bioassays, the biological effects of all chemicals in an unknown mixture, like in surface water, effluent, or groundwater, can be measured for a specific endpoint without the need for identification of the chemicals. The total effect of the mixture can be interpreted quantitatively (e.g. trends), or can be related to effect-based trigger values, which signal the need for further investigation of the potential risk of the mixture when exceeded (Béen et al., 2020). With the recent development of effect-directed analysis (EDA), chemicals in mixtures exceeding trigger values can be identified by fractionation of the mixture before application to the bioassay, to reduce complexity (Houtman et al., 2020). The fraction that induces an increased bioassay response can be analysed to identify the effect-inducing chemical(s). As chemicals can induce a large variety of effects, a bioassay battery is generally recommended to cover multiple endpoints of interest. Due to cost constraints, usually, a limited number of bioassays are applied (such as genotoxicity, endocrine activity and reactivity assays). Even if all available bioassays would be used, the toxicity of some (classes of) compounds would not be detected. For several (complex) endpoints, like neurotoxicity and reproduction toxicity, there are no assays routinely applicable for water quality. So, it is not (yet) possible to detect the total effect of each chemical mixture in water on each possible endpoints with *in vitro* bioassays.

There are a few model organisms that are applied as whole organism bioassays to monitor the (mixture) effects of (environmental) chemicals, like in water quality monitoring. The concentration or dilution factor at which 50 percent of the animals dies (LC50, lethal concentration for 50% of test population) or other endpoints (EC50, Effect

concentration for 50% of the population) is often used to describe the ecotoxicological effect of single chemicals or mixtures. Other endpoints can be for instance movement, reproduction rate, and development of organs after visual inspection, or overall growth (i.e. Spaan et al., 2019).

The use of a whole organism bioassay ('*in vivo*') has the advantage that a range of physiological functions and organs can be studied in one single model, opposite to the single endpoints that are targeted in *in vitro* bioassays. However, very specific subtle effects are often not tested for, while more prominent endpoints such as growth, reproduction are only affected either at high concentrations of after long exposure that exceed the duration of most tests. Therefore, a potentially very useful model would be a whole organism in which specific effects at more detailed molecular level events that occur can be detected. A possibility is evaluating the response of the exposed organism at the molecular level. This can be achieved by using 'functional omics' techniques that measure the components of such molecular responses. The different functional omics techniques are able to detect, characterize and (semi-)quantify a large set of biomolecules in a single run. In this way they can measure biomolecules that are expressed as a functional response to a change in internal or external circumstances for the organism. Combined with the concept of whole organism bioassays, this can be described as a bioassay based on functional omics.

This study investigates whether current functional omics techniques can be used for quality monitoring of drinking water (sources). This report builds upon earlier work in the BTO of Schriks et al. (2011). This earlier report focused on one particular functional omics technique, transcriptomics, as a technique for use in water quality assessment and assessment of risks of individual chemicals. The current report provides information on several other functional omics techniques, prospective model organisms and which setup would be the most suitable for use of these techniques in risk assessment. Available functional omics data in the literature and from databases are inventoried. Sensitivities of different model organisms to exposure to chemicals with specific toxicological endpoints and toxicity mechanisms are assessed. The report also highlights the current opportunities and shortcomings regarding the use of these techniques for water quality monitoring. It describes the viability of implementing a functional omics assay that will aid the drinking water utilities in surveillance of risk associated with mixtures of low-level chemicals in water.

2 Potential application of -omics techniques in drinking water quality assessment

2.1 Explanation of different -omics techniques

In the cascade from genotype (DNA) to phenotype (Figure 1), there are four common -omics techniques that we will discuss here: genomics, transcriptomics, proteomics and metabolomics. Each technique address a stage in the metabolic response cascade of genotype to phenotype. In short, the genome (all DNA in one organism) is the blueprint for the transcription of RNA. Transcribed RNA, specifically messenger RNA (mRNA) is the RNA that is translated into proteins. Proteins induce metabolic changes.

The first -omics technique, genomics, studies the whole genome of one organism (genomics) or multiple organisms (metagenomics) by identifying and measuring DNA sequences. This technique is not considered a functional omics technique because the genome of an organism itself does not respond to chemicals in the environment, or other influences, it merely encodes the potential of an organism. In water quality monitoring this technique is used to detect changes in the composition of ecological communities and potential functions. An example is detecting the

different microbial communities in different water bodies (Timmers et al., 2020). These differences in compositions of communities can be indicative of the long term chemical status of these water bodies.

The expression of different biomolecules in the cascade following the genome can be analysed with techniques that are used to study functional responses to the environment. That is why these techniques are called 'functional' omics. Three main techniques are used in general. For measuring the expression of biomolecules, transcriptomics is using next-generation sequencing or DNA microarrays to measure RNA, proteomics is using mass spectrometry or two-dimensional electrophoresis to measure proteins and metabolomics is using nuclear magnetic resonance or other chemical techniques to measure biomolecules that are formed as products in normal physiological processes (Martins et al., 2019) (Figure 1). When studies use functional omics techniques to analyse and interpret changes on biomolecular levels as a response to exposure to potentially hazardous chemical substances this is called toxicogenomics (Feussner and Polle 2015). Due to recent technical advances, these functional omics techniques have increased output, sensitivity and accuracy, and require only small sample sizes (Martins, Dreij et al. 2019). Below, we discuss these three different functional omics techniques in more detail.

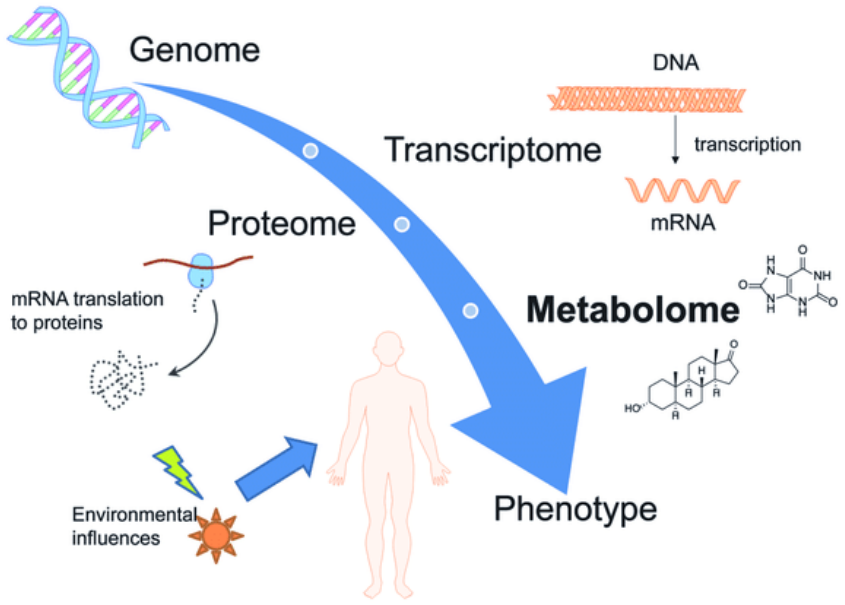


Figure 1. Overview of -omics techniques and their proximity in relation to genotype and phenotype (Steuer et al., 2019).

2.1.1 Transcriptomics

Transcriptomic methods can assess thousands of single ribonucleic acids (RNAs) in a single run, making the technique more cost-effective compared to traditional quantitative reverse transcription polymerase chain reaction (qRT-PCR) techniques if many RNAs have to be characterized. In the past decade, transcription studies have progressed from using microarrays to next generation sequencing of RNA. This technique sequences all RNA present, giving a semi-quantitative measure of expression. Studying transcriptomes can provide indications of initial responses of cellular mechanisms affected by stressors. Effects induced by chemicals can be assessed for all standard model organisms with sequenced genomes (Martins et al., 2019). All RNA that does not code for genes can be digitally removed from the dataset prior to analysis, leaving only gene-coding messenger RNA (mRNA). García-Ortega and Martínez (2015) estimated that transcriptomics could cover 90-95 percent of the genes detected in mammals, which makes transcriptomics stand out in completeness.

One difficulty with transcriptomics is that not all genes that are transcribed are translated into proteins, and for that reason the reaction of a system or organism to chemicals can only be partially understood (Fernie and Stitt 2012).

Translational and post-translational effects affect the resulting composition of the proteome (Canzler et al. 2020). Also, the RNA expression is very sensitive to laboratory circumstances and expression can be upregulated and again downregulated in a matter of minutes to hours. This makes transcriptomics a technique that has a fleeting response that is prone to noise and inter-laboratory variations. Also, what would be a monotonic response to the chemical may produce a non-monotonic dose-response for a given snapshot in time. For example, the time-course of dynamic transcriptional response may differ between doses, such that at higher doses, the transcript abundance may have peaked earlier and have fallen by the time of measurement. Alternatively, there could be a developmental delay in an *in vivo* assay at the higher doses, associated with triggering more and more adverse outcome pathways (AOP) in the organism, thereby causing more and more disruption of normal development. It would seem nothing much happens in a process at a higher dose, however the expected peak in expression is, in that case, delayed.

Nevertheless, it has been shown in many different *in vitro* and *in vivo* bioassays that transcriptomics can provide biomarkers to distinguish the impact of different chemical classes, by transcriptome signatures exhibited in bioassays after exposure to chemicals (e.g. Hermsen et al., 2011; Antczak et al., 2013). Transcriptomics and proteomics may convey similar information on overall pathway responses. However, transcriptomics may include information on regulatory RNAs as well and for that reason can be much more complete than proteomics (Canzler et al., 2020). In the context of drug discovery, many compounds have been tested on model organisms and these transcriptomics responses have also been recorded in publicly available databases like the Comparative toxicogenomics database (http://ctdbase.org/), the DrugMatrix database (can be accessed via various applications), OpenTGgates, Connectivity map (Cmap), iLINCS (https://lincsproject.org/) (Alexander-Dann et al., 2018).

2.1.2 Proteomics

Proteomics studies the set of proteins and peptides (proteome; the final product of gene expression) that have a function in many processes in living organisms, like metabolic signalling, transport and defence (Martins, Dreij et al. 2019). Changes in cell phenotype occur when transcripts are translated into protein. This functional omics technique is often used, especially for assessing potential human health effects. The primary technique applied for proteomics is mass spectrometry using electrospray ionisation (ESI) or matrix-assisted laser desorption ionization (MALDI) or LC-MS. Proteomics can be performed in three different ways; top-down (analysis of intact proteins), bottom-up (analysis of peptides after protein digestion) and middle-down (analysis of larger peptides after limited digestion), depending on the question to be answered. Quantification in proteomics can be done using labelled techniques or label-free techniques. Developments in proteomics are still ongoing. Timp et al. (2020) discuss scalable techniques that could sequence the amino acids in proteins rather than classifying proteins based on spectra.

Compared to transcriptomics, the techniques are somewhat less cost-effective as the number of molecules that are assessed is less (Martins et al., 2019). Proteins do however have a stronger link to the phenotype because their expression directly influences the phenotype. Proteomics groups several regulatory principles that may lead to changes on the protein level, sometimes even without remarkable changes on the transcriptome level (Canzler et al., 2020). Multiple studies stressed that there are limitations with the proteomic approach due to the lack of comprehensive proteomic databases and technical barriers, hindering the identification of differentially expressed proteins and biomarkers (e.g. Trapp et al., 2014; Buesen et al., 2017).

2.1.3 Metabolomics

Metabolomics is still less commonly used than transcriptomics and proteomics to study the effects of chemicals on tissues or organisms. Metabolomics studies biomolecules occurring later in the cascade of events of the toxicological mechanisms affected by chemical exposures, namely the metabolome. The metabolome consists of highly heterogeneous and small (<1500Da) molecules (metabolites) found in cells, tissues, organs and organisms (Wishart 2007). Metabolomics studies use separation techniques like liquid chromatography, gas chromatography or sometimes capillary electrophoresis to separate the metabolites prior to detection with mass spectrometry.

Another commonly applied technique in metabolomics is nuclear magnetic resonance (NMR) which does not require separation of the metabolites but has a much lower sensitivity than mass spectrometry. A challenge in metabolomics is to detect, quantify and identify metabolites present at different concentration levels. In non-targeted metabolomics there is no analytical method yet available that can determine all metabolites due to the current limitations of the analysers and chromatographic separation techniques. A critical point in the use of liquid chromatography mass spectrometry (LC-MS) is the choice of a suitable column to retain certain metabolites. Several studies have stressed the usefulness of applying different types of columns in metabolomics (Booij et al., 2014).

One advantage of metabolomics is that it is less fleeting than transcriptomics and proteomics and can detect molecules at various points in a pathway at the same time, while also detecting instantaneous or slow changes (Canzler, Schor et al. 2020). As in proteomics, the metabolome is not only regulated by gene expression, but also by post-transcriptional and post-translational effects. Changes in the metabolome due to exposure are often amplified compared to changes in the transcriptome. The metabolome is the closest link to the phenotype induced by chemical exposure (Tan et al., 2009). Metabolites are also well conserved across all living organisms. This makes the use of non-model species less challenging (Martyniuk and Simmons, 2016; Sahlin, 2018; Tan et al., 2009). Databases for the identification of small biomolecules from the metabolome are increasing, and the use of metabolomics in toxicogenomic studies are on the rise (Martins et al., 2019). However, interpretation remains a challenge since a significant number of metabolites have not been identified or referenced (Poynton et al., 2008).

2.1.4 Integrated functional omics

Combining functional omics methods can give an even more complete view of the effects of chemical exposure (Martins et al., 2019). Next to interpreting individual transcriptomics, proteomics or metabolomics analyses, the integration of measurements and chemical analysis will allow to explore potential cause-effect relationships (Sahlin, 2018). Even though transcriptomics and proteomics results often show remarkably limited overlap, a study of Simões et al. (2018) found that specific functional categories at protein and transcriptome levels were concordant with each other, despite overall limited correlations between datasets. Combining methods can also give insight in possible interactions between the transcriptome, proteome and metabolome (Martins et al., 2019). There are not many examples yet of successful combinations of the functional omics techniques.

Table 1. Qualitative assessment of different functional omics techniques, based on the text in this paragraph.

	Chemical exposure data available	Number of biomolecules addressed	High throughput (ease of measuring)	Robust patterns	Relation to phenotype
Transcriptomics	++	+++	+++	+	+
Proteomics	+	++	+	++	++
Metabolomics	+	++	+	+++	+++

Explanation of the symbols: +++ optimal, ++ less than optimal, + far less than optimal

2.2 Advantages of the application of functional omics in water quality monitoring

The use of functional omics based bioassays in monitoring of drinking water (sources) quality is relatively new. Functional omics techniques already are being used to assess the toxicological effects of *single* chemicals on health such as in drug development, or to assess adverse effects of environmental mixtures on ecology. The U.S. EPA accepts data from such techniques for establishing mechanisms of toxicity for chemicals (McCarroll et al., 2009).

One advantage of functional omics, in contrast to traditional toxicity testing, is that it allows to understand the mechanisms that cause toxicity (Sahlin, 2018). Functional omics techniques can be applied to establish toxicological pathways and toxicity mechanisms of chemicals by combining molecular with phenotypic responses. Functional omics can also be used to expand detailed knowledge of toxicity mechanisms by identifying novel features (biomolecular responses) induced in response to chemical exposures, for example in metal homeostasis or detoxification (Hägerbäumer et al., 2015). Specific profiles ('signatures') of molecular responses that are associated with chemicals and their effects on specific endpoints can be deducted from an -omics measurement. In potential, such functional omics signatures are predictive, as signatures for specific exposures or toxicity can be identified. These signatures can be used to identify specific effects of unknown chemicals or chemical classes. In that way functional omics techniques can be applied as a diagnostic tool.

Compared to traditional bioassays, which will react in one particular way depending on the tested endpoint, bioassays based on functional omics will thus provide a broader overview of all molecular changes in an organism or tissue (Schriks et al. 2011). This is important because water will typically not contain a single compound but a mixture of chemicals inducing different toxicological effects. There is evidence that chemicals with a similar mode of action can have additive effects. This means such chemicals together could cause an adverse effect, even though their individual concentrations may be under the threshold for this effect (Altenburger et al., 2004; Kar et al., 2019). There is also evidence regarding the additive toxicity of contaminants with *different* mode of action. Chemicals with a similar toxicological endpoint can have additive effects, on the endpoint. This suggests that the combination of different chemical classes could explain the overall toxicity of a water sample (Altenburger et al., 2004). Different combinations of chemicals (acting by different mechanisms and/or interacting with different endpoints) may result in additive effects on gene expression profiles in an organism (Schriks et al., 2011). Omics techniques have the potential to detect effects on the molecular level caused by low-dose interactions of mixtures, they can even provide a picture of the prospective effect after a short exposure (Ankley et al., 2009; Poynton et al. 2008). Because genes, proteins, metabolites are expressed within minutes to hours, insight in which cellular mechanisms are stimulated or inhibited could be gained after a short exposure (Soong et al, 2015).

The insight in mechanisms, the sensitivity, the potential for detection of a broad range of detailed early stage effects of overall toxicity makes functional omics techniques an interesting candidate to replace or complement traditional whole organism bioassays. Moreover, they could potentially replace commonly used batteries of *in vitro* bioassays that each address a specific endpoint, at least for a part (Fang et al., 2020), by detecting simultaneously different effects of low-level chemical mixtures in water. At the same time, functional omics can be used to obtain a more detailed knowledge on the expected effects of chemicals in mixtures. In this way, functional omics approaches can be used for broad-based surveillance monitoring where complex mixtures and unknown stressors may be present (Sahlin, 2018). Functional omics can thus be an informative approach to detect early warning signals of water pollution and help to prioritise when drinking water (sources) need to be further characterised and studied. The use of -omics can on the one hand help formulate hypothesis concerning possible (health) effects and responsible chemicals when evaluating the water on its suitability for the production of drinking water (Schroeder et al. 2017). On the other hand, it may be used to exclude any possible indications of an impairment or risks associated with drinking water quality.

2.3 Assessing water containing chemical mixtures with functional omics data

When a functional omics based assay is applied to assess potential health implications of drinking water sources containing unknown mixtures of chemicals, an important question is whether the investigated response in biomolecules ('features') in a model organism reflects the combined influence of the individual compounds. There is potential to link environmental functional omics studies to current risk assessment procedures if an impact on adverse toxicity endpoints as considered in guideline toxicity studies can be distinguished. Several studies suggest that this is possible.

The ability of any bioassay to adequately detect effects of mixtures can be assessed by comparing observed effects with calculated effects, based on known mixtures and mixture models. In environmental mixtures, concentrations are generally too low to expect any interactive effects such as synergism or antagonism (Cedergreen et al., 2014). Therefore the dose or concentration addition (CA) model for chemicals with a similar working mechanism is generally considered as the most relevant and applicable for complex environmental mixtures. However, as functional omics may reveal an overview of dissimilar working mechanisms that may still result in combined toxicity, this may support the application of the independent action (IA) or response addition model, applying to chemicals causing the same adverse outcome with dissimilar working mechanisms.

Dardenne et al. (2008) investigated the effects of individual substances and binary mixtures on 14 biomarkers for stress. Mixture responses were fitted applying both CA and IA models. In many cases both models were able to predict the mixture response from the individual compound responses. In another study, for two mixtures of 12 compounds, mixture effects per molecular pathway were generally as large as expected if calculated by concentration addition (Xia et al., 2020; Wang et al., 2018). Krasnov et al. (2007) found in their study in which trout were exposed to three compounds at low level exposure, the expression profiles were additive and that two of three compounds could even be distinguished from the mixture profile.

2.4 Linking functional omics responses to diagnose potential adverse outcomes from chemical exposure

When profiles of expressed functional omics features differ between water containing unknown chemical mixtures and unaffected water, an important question is what potential risk this may imply for human or environmental health. An interpretation is needed. Therefore a major challenge with omics for health risk detection in (sources for) drinking water is to define clear links between functional omics responses to phenotypic observations and their relevance to human relevant endpoints. Is there a risk for endocrine disruption, kidney toxicity, reproductive toxicity or another type of toxicity associated with the water? It has been shown with toxicogenomics that such high level endpoints can, for individual chemicals, be distinguished to a reasonable level in functional omics bioassays (e.g Antczak et al., 2013; Hermsen et al., 2011). These methods rely on overlapping expression patterns between chemicals that are associated with a specific endpoint. See Figure 2 for an example to illustrate the necessity for the derivation of a common signature. Based on their complete gene expression pattern the compounds of similar chemical class do not necessarily cluster together (Figure 2). A selection of a common pattern of expressed genes is necessary to achieve this.

Figure 2. From Antczak et al. (2013). Gene expression clusters in *Daphnia* for 36 contaminants of environmental concern of diverse chemical structure. The clustering of non-related compounds based on overall gene expression is exemplary for the need to extract specific, robust signatures that indicate specific similarities between contaminants. While each gene expression profile is unique, Antczak et al. take advantage of similarities in expression profiles between related contaminants in order to develop predictive models that can distinguish between different groups of contaminants (not shown).

While the applicability of these models based on chemical class-based profiling to distinguish between chemical classes does not require any knowledge of the function of these genes, a biological framework to understand differences would increase interpretability of the models (Antczak et al., 2013). This biological framework could be adverse outcome pathways (AOP). Such an approach has an added advantage. If the expressed features are part of an AOP (see Box 1), concentration addition or independent action (Altenburger et al., 2004; Kar et al., 2019) for that specific pathway can be assumed for a mixture. In this way the type of combined adverse effects can be taken into account in the interpretation of the expressed features. Currently, more than 350 AOPs describing endpoints relevant to both human health and the environment are organised and available in the AOP wiki (Ankley and Edwards, 2018) in various states of completeness.

BOX 1. Terms in toxicology.

Toxicity can be defined as the degree or capacity of chemicals to affect physiological processes in humans or other organisms depending on the dose, exposure route(s) and exposure window. An **endpoint** is defined as a physiological function on which a potential **effect** of exposure to a chemical or chemicals is studied in toxicological studies. The organism, organ, tissue, cell or molecule that interacts with a xenobiotic chemical is defined as the **target**. Endpoints and targets are functionally related to physiological processes or organ systems, such as reproduction (reproductive toxicity) or the immune system (immunotoxicity). The **response** is defined as the size of the studied effect in the applied model related to a certain exposure time and exposure concentration or dose. Chemical-induced effects can be in the **adaptive range** if they are counteracted by physiological processes. If the chemical-induced effect is more intense or of longer duration, it can reach the **adverse range**, in which more stress is encountered and irreversible physiological changes occur. The response resulting from exposure to chemicals can be described by the **mode of action** whereas the **(toxic) mechanism of action** describes the more detailed underlying molecular chain of responses. This chain of effects can also be described as **Adverse Outcome Pathways (AOPs)**, which describe the sequence of linked events at different levels of biological organization, i.e. from chemical properties of the toxicant to individual or population responses that may eventually lead to an adverse effect on human or ecosystem health. Generally, all AOP start with a **Molecular Initiating Event**, which consists of a reaction between a xenobiotic molecule and endogenous molecules. Physiological changes that are toxicologically associated with the adverse outcome at an organ, individual or population level are defined in the AOP as **key events**. Much knowledge on AOPs can be found on https://aopwiki.org/ and the OECD's AOP Knowledge Base (https://aopkb.oecd.org/).

Molecular initiating events (MIE) occur at the start of an AOP (see Box 1). Detecting effects that relate to a MIE could mean an earlier detection of potential adverse effects (Bahamonde et al., 2016). Mode of Action (MoA) follows MIE in the chain of events following chemical exposure (see Box 1). MoA has been loosely defined in both human health and ecotoxicology as a functional change at the cellular level, in contrast to the mechanism of action or MIE. MoA represents an intermediate level of complexity between molecular mechanisms and physiological or organismal outcomes (Kienzer et al., 2017). A MoA provides information on underlying toxicity pathways. Groupings across multiple chemical classes can be done based on MoA (Barron et al., 2015; Kienzer et al, 2017). MoA-based model development can be an alternative to chemical class-based profiling. The MoA concept allows for predicting effects caused by chemical mixtures with additive (same MoA) or individual action (different MoA).

Applicability of the assignment of MoA across diverse species is still an area for which more research is needed (Barron et al., 2015; Lee et al., 2015). Interspecies variability may have caused an alteration of the function of certain genes, and some functional pathways may not be conserved between species. On the other hand, MoAs may be similar across taxa, because the basic biochemical systems and molecular targets affected can be generally conserved across many animal species (LaLone et al., 2013).

We can conclude that integrating functional omics with adverse outcome pathways (AOPs) provides a way to predict particular and chronic adverse effects on higher levels of biological complexity, e.g. humans (Sahlin, 2018). Nevertheless, however useful the concept of an adverse outcome pathway for classes of chemicals with a common toxicity endpoint is, chemically induced adverse outcomes are not always reached by a single and defining pathway. In a recent study on chemicals used in agriculture that induced treatment-related non-genotoxic tumours, the majority of the tumour types formed could be linked to six MoA or MoA networks via descriptions in public safety reports (Heusink et al., 2020). Spaan et al. (2019) identified several pathways towards thyroid hormone disruption. Also for flame retardants, different MoA were established (Scanlan et al., 2015). Moreover, single chemicals can have

several MoAs, inducing different adverse effects. This means that it is not always possible to link a single MoA exclusively to one endpoint. Moreover, chemicals inducing different effects will be present in a mixture. These chemicals can induce effects with overlapping pathways. This can have additive effects on *parts of* these pathways. This means it may not be feasible to univocally assign measured biomolecular feature signatures to a single MoA.

There is currently a shortage of information on this issue. In a literature study by Sahlin (2018) it was found that only in very few studies the responses of functional omics assays were connected to possible adverse outcomes at higher levels (e.g. reproduction, impaired growth) in a predictive approach. This means that more work is needed before this link can be firmly established. In Chapter 3 we take a closer look at the possibilities to link the known MoA to toxicological endpoints of chemicals.

2.5 Options for whole organism model species for –omics assays

The bioassay model species is an important determinant of the potential of application for the quality assessment of (sources for) drinking water. The chemical impact can be described more completely if the bioassay covers multiple human-relevant toxicological endpoints and MoAs, thereby providing a linkage between chemical and health effects status (Brack et al., 2018; Sahlin, 2018).

For instance, in a study of Schaap et al. (2015) it was found that *in vitro* primary mouse hepatocytes did not detect all MoA of non-genotoxic carcinogens, and a second *in vitro* test was required to complement this. It is likely that higher and more complex whole organisms exhibit a wider range of effects than a single cell type or tissue *in vitro* model. Even so, an organism that is less similar to humans will have less chance of exhibiting effects that can be used to predict potential effects on human health. Bacteria, even though they are whole organisms and their application would mean an ease of operation, would be arguably less suitable for this aim. The use of an aquatic model organism requiring only a small volume of water has the best potential based on practical considerations. With such organisms, dose response effects of water extracts can be studied, even if only small amounts of extract are available. Invertebrate species can be used as alternatives to vertebrate species for testing chemicals known to act through similar pathways across species (Schriks et al, 2011). Functional omics can provide a mechanistic understanding of toxicity and this can be put to use in cross-species extrapolation in risk assessment studies.

An example of a suitable model organism are nematodes (see Figure 3); they are small but have specific organs and a nerve system. Furthermore, they have been extensively used in scientific fields such as developmental biology and neurobiology because they have neuronal systems and signal pathways together with molecular networks that are comparable to mammals. Moreover, they are easy to keep in a laboratory and suitable for testing of multiple endpoints (Queirós et al., 2019). The overall gene expression in the nematode *C. elegans* upon exposure to various chemicals has been assessed in multiple studies using DNA microarrays (Hägerbäumer et al., 2015). In a study of Menzel et al. (2009), nematode metabolism, functional pathways and developmental processes were found to be differentially regulated in response to contaminated sediment. In general, it has a proven sensitivity to metals and pesticides, similar to that of other model organisms (e.g. *Daphnia magna*). However, most studies are in the biomedical domain. The possibilities to apply *C. elegans* in environmental studies need to be further established (Queirós et al., 2019).

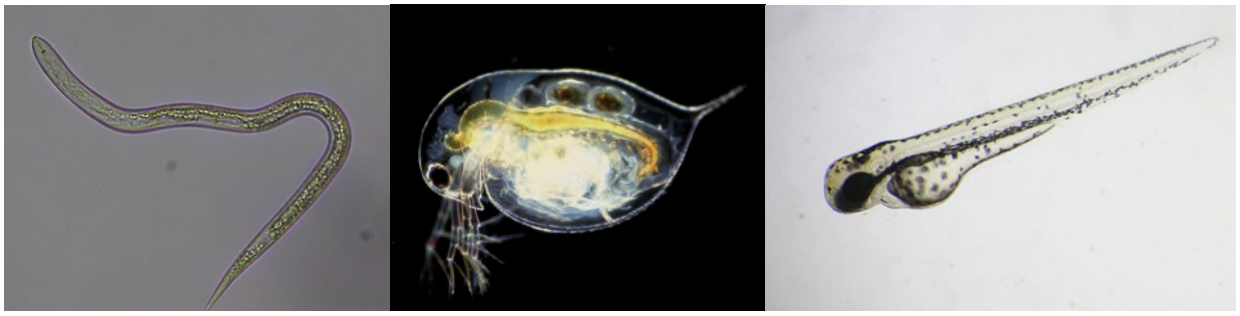


Figure 3. Three promising candidate model species. In order from left to right: Nematode (picture source: https://www.flickr.com/photos/scotnelson/39013404045/in/album-72157635415455383/, CC0), Daphnia (picture source: doi:10.1371/journal.pbio.0030219), Zebrafish embryo (picture source: Ho-Wen Chen, Flickr, zebrafish embryo 50hpf_01, CCBYNC).

Daphnia (see Figure 3) and other small crustacean invertebrates have a small body size and short and cyclic parthenogenetic (reproduction without fertilization) life-cycles, making them suitable as model organisms for studies on large populations in various environmental conditions while maintaining the same genotype. *Daphnia* species are commonly used for the indication of water quality and environmental health (Antczak et al., 2015; Soong et al., 2015). They have a high sensitivity to environmental change (Wagner et al., 2017; Jeong et al., 2019).

Another important model species is the zebrafish (*Danio rerio*) (see Figure 3) whose development is highly similar to humans and other vertebrates. Zebrafish development from fertilization to free swimming larvae takes 2 to 3 days, while morphological development can be monitored since the embryos are transparent (Tufi et al., 2016). The zebrafish are small and have a high reproduction rate making them suitable for high-throughput screening. Wang et al. (2018) have shown that zebrafish embryos can be more sensitive to changes of basic processes such as oxidative phosphorylation, than a single cell-type *in vitro* system. However, zebrafish bioassays require expensive zebrafish laboratory facilities. Also, zebrafish are vertebrates, and 120 hours post fertilisation these tests requires the approval of an ethical committee.

Table 2. Qualitative assessment of model species, based on the literature in the text in this paragraph.

	Chemical exposure data available	Sensitivity to many endpoints	Easy to culture	Whole life cycle endpoints	Remark
Daphnia	++	+	+++	+++	Not all human organ systems are represented in *Daphnia*
Zebrafish	+++	+++	+	+	The use of zebrafish is considered animal free testing only up to 120 hours post fertilisation
Nematode	+	+	+++	+++	Not all human organ systems are represented in nematodes

Explanation of the symbols: +++ optimal, ++ less than optimal, + far less than optimal

2.6 Progress on challenges since 2011

Poynton et al. (2008) noted the main challenges in application of functional omics technologies in 2008, and these were commented on in the report by Schriks et al. (2011). These challenges were in model system complexity, confounding factors, mixtures, limited sequence data, bioinformatics and costs. In this section we discuss how these challenges are still relevant and note possible solutions.

Model system complexity

The model system complexity point of Poynton et al. (2008) concerned the fact that not all chemicals are triggering the same effects in different bioassays. Although this is true, the measured effects can still be meaningful if expressed features are found that can (statistically) be linked to a relevant endpoint. A whole organism assay will have a potential to detect different and more than one adverse effect after exposure to a mixture of different chemicals. It is probable that not all possible endpoints of interest can be detected in a single whole organism bioassay. If this is established, another bioassay may be able to detect this. A combination of bioassays may be necessary to cover all relevant effects.

Another consequence of model system complexity is the variability in the expression of features. Slight differences in exposure conditions can change the expression of features. To mitigate the variability in expression of features, the use of defined groups of features may be more robust than using individual features. Features are expressed as part of a molecular pathway. It was shown earlier that, whereas pathways may, as a whole, be regulated by compounds of a particular MoA, chemical class, or toxicity endpoint, the individual features expressed within these pathways may differ with each individual compound (Hermsen et al., 2013). Therefore, aside from individual biomarkers, pathway regulation may be a useful starting point for defining biomarkers for predicting toxicity endpoints (Hermsen et al., 2013). Metabolomics has the best chance to overcome the sensitivity of expressed features, as this technique integrate a sequence of events rather than provide a snapshot of the situation such as transcriptomics and to a lesser degree proteomics do.

Confounding factors

Confounding factors mainly concern aspects in field experiments, and field water samples. They include nutrient levels, acidity, temperature, salinity, dissolved organic carbon (DOC) and water hardness, which may modulate gene expression in an organism. The application of functional omics to water quality monitoring for drinking water (sources) has most potential as standardized assay. This reduces complexity of interpretation of the expressed biomolecules. In this way results can be compared between waters of different quality. In the laboratory, solid-phase extraction (SPE) can be applied to field water samples to extract only the chemicals. Confounding factors such as mentioned above are no longer an issue in that way.

Mixtures

Another challenge noted by Poynton et al. (2008) related to the question if the features expressed as a result of exposure to a mixture could still yield interpretable signatures for determining risks for different endpoints. The linkage of genomic responses to relevant endpoints (e.g. behaviour, reproduction and development) has to be achieved, by the combined effect of chemicals in the mixture. This will require a better identification of relevant adverse outcome pathways (AOP) (Schriks et al., 2011) and linking molecular responses to these AOP. Several studies have shown promising results to detect additive, independent, and possibly even synergistic effects (Dardenne et al., 2008b; Xia et al., 2020; Wang et al., 2018; Krasnov et al., 2007).

Limited sequence data and bioinformatics

There was a challenge with the limited availability of sequence data for constructing microarrays at the time of Poynton et al. (2008). However with increasingly cheap next generation sequencing techniques this is no longer a problem. With a relatively small investment, complete genomes of different organisms can be sequenced. These

data are also publicly available at NCBI (https://www.ncbi.nlm.nih.gov/genome/) or the European Nucleotide Archive (ENA) https://www.ebi.ac.uk/ena/browser/home.

Also other bioinformatics resources have increased in the last ten years. Among others, open source gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathway databases are available. Also open source initiatives exist to combine such pathway databases. These can be used for transcripts and proteins (as these are the translated products of transcripts). For metabolomics, the bottleneck has been known metabolites in databases. In the meanwhile, progress has been made. The online Metabolomics Databases DrugBank contains information about 1,240 metabolites and 1,550 drugs, including drug–drug interactions. The Human Metabolome Database has details on more than 40,000 quantified, detected, and expected metabolites in humans and their known roles in health and disease. The Toxic Exposome Database describes 3,670 metabolites, including toxic drugs, pesticides, herbicides, endocrine disrupters, carcinogens, solvents, Polychlorinated biphenyls, and furans (SOURCE: Wishart workshop presentation, https://bioinformatics.ca/workshops/2020-informatics-and-statistics-for-metabolomics/).

Standardization

The issue of a lack of standardization of protocols and the sensitivity of expressed features to experimental conditions between labs remains a challenge. The lack of standardization is a result of the fact that most functional omics studies do not focus on the development of a functional assay, rather they are focused on discovery of mechanisms and effects. A paradigm shift from mechanism discovery towards the development of assays to monitor water quality and their uptake in regulation would aid to further standardize methods and protocols. So, the expectation is that standardization will increase with a change in application to water quality monitoring.

Costs

The costs for applying omics methods are still considerable (Martins et al., 2019). However, these could weigh up to the use of a complete battery of bioassays to test each possible effect that water with an unknown mixture of chemicals could trigger. What can be cost efficient is that the same experiment could provide information on toxicity classes, MoA, and indications for possible effects on human toxic endpoints. This could even be the case for chemicals that remain unknown and unseen. The collection of omics fingerprints in databases (similar to the LC-HRMS chromatograms in the NORMAN Digital Sample Freezing Platform) may even allow retrospective analyses of water quality. The use of a reduced transcriptome or metabolome with only genes or metabolites of established toxicological interest reduces costs of analyses. Although this does not exploit the full capabilities of omics' techniques, particularly not allowing discovery of unexpected changes (Van Aggelen et al., 2010), this is adequate for an interpretable standardized whole organism –omics bioassay (i.e. Xia et al., 2020; Wang et al., 2018).

3 Data analyses to support selections

In this chapter we assess three aspects of currently available data. These are aspects that can aid in the selection of interesting MoA, and in assessing the potential of different organisms and omics techniques to use as a functional omics bioassay.

Firstly, we perform an analysis to establish whether known MoA of chemicals can be statistically linked to known human relevant toxicity endpoints of chemicals. If this is the case, biomarkers linked to the MoA of a chemical can be predictive for its human toxicity endpoint. If this is not the case, specific (other) biomarkers will have to be selected. This would mean we are further removed from easy application of -omics techniques to predict water

quality with respect to human toxicity endpoints. So far, this aspect of linking the MoA of chemicals to high level toxicological endpoints has been underexposed (Sahlin, 2018). We use two datasets, one with MoA of chemicals (Barron et al., 2015) and one dataset with human toxicity endpoints per chemical that we constructed (see Appendix I).

Secondly, we assess for which -omics technique and organism there expectedly is enough data on expressed biomolecular features during exposure to individual chemicals. These data are needed to be able to anchor biomolecular features to a MoA. Such data on biomolecular features can also be used to directly link to toxicological endpoints induced by chemicals. If for a particular -omics technique or model organism chemicals with a particular toxicity endpoint have not been tested, this will mean that applicability of the model organism or technique to provide a prediction of those endpoints based on signatures of features will need more investigation. Organisms for which evidence on their expression of biomolecular features after chemical exposure for different toxicity endpoints is available, will have a preference because less study is needed.

Thirdly, we asses which organism is most sensitive to chemicals with a particular MoA and human toxicity endpoint. For MoA of chemicals we again use the dataset of Barron et al. (2015). For sensitivity we use acute toxicity (expressed as the dose in an aqueous solution at which 50% of a response is exhibited, EC/LC50) for different chemicals for three fish species, fish in general, Daphnia and invertebrates in general, taken from the same dataset (Barron et al., 2015). If an organism is very sensitive, we expect a reaction will be measurable at low concentrations of chemicals.

3.1 Linking assigned MoA and Toxicity Endpoint for chemicals

MoA represents an intermediate level of complexity between molecular mechanisms and physiological or organismal outcomes (Kienzer et al., 2017). A goal of a functional omics assay for drinking water quality monitoring could be to link the signature of expressed biomolecules belonging to a particular MoA to a potential human relevant toxicity endpoint. This requires that the studied MoA has some predictive power for a particular human-relevant toxicity endpoint. We looked at the chemicals having both known MoA and defined toxicity endpoints, and investigated whether chemicals with a particular MoA were significantly overrepresented in the group of chemicals with a particular human relevant toxicity endpoint. Overrepresentation means that chemicals with a particular toxicity endpoint also have a particular MoA, more than can be expected by random overlap. The overrepresentation was limited to be based on at least three chemicals, to avoid chance effects.

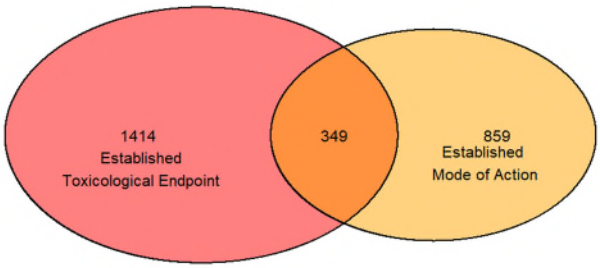


Figure 4. The overlap for chemicals with established MoA (Barron et al., 2015) and human relevant toxicological endpoint (Appendix I list)

Several institutes and organisations have lists of chemicals that are known to induce one or several human relevant toxicological endpoints. From such different publicly available sources a collection was made of 1763 chemicals that induce one or more of eleven different common toxicity endpoints (Appendix I). These toxicological endpoints are: carcinogenicity, cardiotoxicity, developmental/reproductive toxicity, endocrine disruptors, hematotoxicity, immunotoxicity, kidney toxicity, liver toxicity, neurotoxicity, spleen toxicity and thyroid toxicity. For a list of chemicals with a known MoA the list of Barron et al. (2015) was used. This list contains MoAs for 1208 different chemicals. A broad MoA as well as a detailed MoA is listed. Detailed MoA are listed as a subcategory under the broad MoA. The

six broad MoA are: AChE inhibition, Iono/osmoregulatory/circulatory impairment, Narcosis, Neurotoxicity, Reactivity, Electron transport inhibition (Barron et al., 2015).

The U.S. Environmental Protection Agency's ecotoxicity database (https://cfpub.epa.gov/ecotox/) also contains information on chemicals, test animals and endpoints. However, the distinguished endpoints are less mechanism based. We tried extracting data for two noted endpoints, reproduction and intoxication. In the first category, mainly *Daphnia* were tested. The second category is very broad and covers a lot of physical endpoints rather than toxicological mechanisms (anorexia, immobility, convulsion, etc.). For this reason we did not further explore or use the ecotoxicity database for the analysis.

Result

When combining the two lists it was noticeable that the compounds for which *both* toxicological concepts were derived (MoA and Toxicity Endpoint) had relatively little overlap (Figure 4). This means from the chemicals with known human toxicological endpoint often little is known on the MoA, and vice versa. This is probably a result of the selection of the chemicals, based on the relevance for either list. The MoA were derived for chemicals that are relevant by their presence in an ecotoxicology context (Barron et al., 2015). The toxicological endpoints were derived for chemicals collected from diverse sources, probably according to their relevance of exposure risk to humans, e.g. via work, food, drink, pharmaceuticals or usage.

With regard to the link between MoA and toxicity endpoint, only "Hematotoxicity" was overrepresented with chemicals that had broad MoA "Narcosis" (Table 3). This MoA is a non-specific reversible baseline mode of whole organism toxicity (Barron et al., 2015). For the detailed MoAs more overrepresentations were found and these are in Table 3. For a description of the specific MoA, see Barron et al. (2015).

Table 3. Significant associations between toxicity endpoints* and MoA (broad and specific, see Barron et al., 2015 for an explanation on the MoA). If the cell is empty, no MoA was statistically linked to the Toxicity endpoint.

Toxicity endpoint	Broad MoA (Barron et al, 2015)	Specific MoA (Barron et al., 2015)
Liver toxicity		**Acrylate** (broad MoA: Reactivity) **Chromate** (broad MoA: Reactivity) **Arsenical respiratory inhibition** (broad MoA: Electron transport inhibition)
Immunotoxicity		**Alicyclic GABA antagonism** (broad MoA: Neurotoxicity)
Hematotoxicity	**Narcosis**	**Anticoagulation** (broad MoA: Iono/osmoregulatory/circulatory impairment) **Carbonyl** (broad MoA: Reactivity)
Endocrine disruptor		**GABA agonism** (broad MoA: Neurotoxicity) **Carbonyl** (broad MoA: Reactivity)
Neurotoxin		**Diphenyl sodium channel modulation** (broad MoA: Neurotoxicity)

* Other toxicity endpoints that did not have a MoA associated: carcinogenicity, cardiotoxicity, developmental/reproductive toxicity, kidney toxicity, spleen toxicity and thyroid toxicity.

No MoA was found significantly associated for chemicals for six of the eleven considered toxicity endpoints in these datasets. One specific MoA, Carbonyl, was associated to both "Hematoxicity" and "Endocrine disruptor".

The analysis reveals that a clear link between (these) MoA and toxicity endpoint is not readily available or univocal. This means that some toxicity endpoints will require having (new) MoA established for chemicals that induce that toxicity endpoint. Because chemically induced toxicity endpoints have been shown to occur via multiple MoAs (e.g. Heusink et al., 2020; Spaan et al., 2019; Scanlan et al., 2015) this information may be necessary as well in future analyses. In the dataset of Barron et al. (2015) one MoA per chemical is established.

3.2 Historical functional omics data available for individual chemicals

In order to anchor expressed biomolecular features to MoA or toxicity endpoint of a particular type of chemical, data on expressed biomolecular features in a test organism during exposure to individual chemicals are needed. If for a particular omics technique or model organism chemicals with a particular toxicity endpoint have not been tested, this will mean that before the model organism can provide a prediction of those endpoints more investigation is needed. Organisms for which evidence on their effects from chemical exposure for different toxicity endpoints is available will have a preference because less study is needed.

For this report we did a search in literature to find which chemicals with certain toxicity endpoints are tested with different -omics methods and different model organisms. In Appendix II the literature search is explained. The search string applied will not find *all* studies that applied functional omics to assays after chemical exposure. However because we apply the same search string between organisms and techniques, it provides an equal basis to find studies on the different organisms and techniques. In this way results can be compared on a qualitative note. A much more elaborate search would be necessary to uncover all available studies, this is beyond the scope of this report.

3.2.1 Chemicals tested in organisms

Using the search queries listed in Appendix I, 1147 articles were selected for further evaluation. Articles were selected when the following queries were found in the title, abstracts or keywords, namely variants of the terms metabolomics, transcriptomics, proteomics, exposure and chemical, and when a variant of the term monitoring was found. Reviews were excluded if they did not contain original data. Of the selected articles, omics techniques were represented in equal numbers (457, 412 and 377 for metabolomics, proteomics and transcriptomics, respectively). In total 99 articles contained a study with multiple omics techniques.

To establish which organism is frequently used in environmental monitoring studies using omics techniques, the various species were listed (Table 4). Most articles used omics techniques on human tissues (424 articles), followed by fish (293 articles) and rodents (mice 128 articles and rat 114 articles). In total 86 articles were found to refer to bacteria. While bacteria would be relatively easy and cheap to use as bioassays in laboratories, most studies refer to the gut bacteria in other organisms and not to *in vitro* cultured bacteria that could be used for water quality monitoring. In total 69 articles referred to mussels. Mussels were studied mostly in salt water with respect to their economic value and disadvantageous effects of harmful chemicals on growth, their ability to bioaccumulate substances, and their association with producing toxic substances. The main benefit regarding the use of bivalves in exposure studies are due to that they are filter feeders and therefore capable of accumulating potentially high levels of contaminants, thus providing temporally and spatially integrated levels of contaminations (Campos et al., 2012). These are not directly useable for a standardized short term bioassay based on functional omics for fresh water quality monitoring. They could be used though as a functional omics based bioassay to monitor long term water quality. One of the three useful model species identified in paragraph 2.5, nematodes, was not found often in this literature search.

Table 4. Studies resulting from the literature search, divided into species. The number of articles linked to certain species were determined using search queries as listed in Appendix I. The selected species of interest for water quality monitoring are highlighted in blue. For selected organisms (zebrafish, daphnia) it was established how many studies contained data, this is between brackets. Organisms with less than twenty hits are not listed in the table.

Organism	Number of articles
human	424
fish	293
mice	128
rat	114
of fish → zebrafish	104 (95 omics)
in vitro	103
bacteria	86
mussel	69
daphnia	34 (30 omics)
of fish → fathead minnow	30
of fish → shell fish	24
oyster	26
clam	29
algae	24

Since this study focuses on identifying possible whole organism models for water quality monitoring, zebrafish and *Daphnia* were selected as organisms of interest for further investigation. These are organisms that are often used as model species and live in the aqueous environment, giving the opportunity to relate omics to *in situ* responses if required.

A total of 104 zebrafish articles were selected, of which 95 articles included an omics study and chemical exposure. In more detail, in 60, 23 and 24 articles, respectively transcriptomics, proteomics and metabolomics studies were described. In 11 articles multiple omics techniques were used. A total of 198 chemicals, 7 nanomaterials and 1 microplastic component were tested in these articles.

Of the 34 *Daphnia* articles, 30 included an omics study and one or more chemicals to which the organisms were exposed. A total of 18 transcriptomics studies, 3 proteomics studies and 10 metabolomics studies were included (1 article contained multiple omics). In total, 28 articles exposed *Daphnia magna*, whilst two exposed *Daphnia pulex*. Most studies exposed *Daphnia* to several single chemicals as in total 74 chemicals and one type of nanoparticles were included in the 30 articles selected.

3.2.2 Most tested human toxicological endpoint

When linking the chemicals to which the selected organisms were exposed (nanoparticles, microplastics and particulate matter excluded) to CAS numbers, toxicological endpoints could be appointed based on the list of collected human toxicological endpoints described in Appendix I. Using this list, 38% and 50% of all chemicals used in zebrafish and *Daphnia* could be linked to an endpoint. In total 15 chemicals in the zebrafish studies, and 2 in the *Daphnia* studies could not be linked to a CAS number. In Table 5, the most frequently represented endpoints are listed per organism. If an endpoint is relatively frequently tested this could mean that the model species are responsive to these endpoints. Otherwise, such chemicals would not be chosen repeatedly in studies to test effects. The endpoints that were tested more often than could be expected based on their occurrence in the list of human endpoints, are indicated in blue. In Table 6 and 7, the studies in zebrafish and Daphnia, respectively, are divided between omics techniques.

Table 5. Overview of the human toxic endpoints associated with the chemicals used in the omics studies. The amount of chemicals linked to a human endpoint is stated between brackets. The percentage of chemicals capable of affecting certain endpoints are listed for each organism. The percentages do not add up to 100% because most chemicals are associated with multiple endpoints. Endpoints tested >10% more frequently than expected (compared to their occurrence in the human endpoint list) are indicated in blue.

Endpoint	*Danio rerio* (200 of 402 chemicals could be linked to a human relevant endpoint)	*Daphnia* (all) (124 of 183 chemicals could be linked to a human relevant endpoint)	Percentage in human endpoint list
Carcinogenic	28%	37%	36%
Cardiotoxicity	4%	6%	1%
Developmental/reproductive toxicity	30%	42%	15%
Endocrine disruptor	42%	27%	16%
Hematotoxicity	5%	14%	4%
Immunotoxicity	9%	25%	3%
Kidney toxicity	22%	26%	12%
Liver toxicity	35%	27%	14%
Neurotoxicity	28%	43%	33%
Spleen toxicity	3%	2%	1%
Thyroid toxicity	6%	2%	2%

Table 6. Overview of the human toxic endpoints associated with the chemicals used in *Danio rerio* studies. The amount of chemicals linked to a human endpoint is stated between brackets. The percentage of chemicals capable of affecting certain endpoints are listed for omics technique. For the transcriptomics, proteomics and metabolomics, respectively 47%, 68% and 50% of all chemicals investigated are linked to endpoint using the human endpoint list. Endpoints tested >10% more frequently than expected (compared to their occurrence in the human endpoint list) are indicated in blue.

Endpoint	Transcriptomics (162 of 343 chemicals linked to a human relevant endpoint)	Proteomics (19 of 28 chemicals linked to a human relevant endpoint)	Metabolomics (15 of 30 chemicals linked to a human relevant endpoint)	Percentage in human endpoint list
Carcinogenic	27%	42%	27%	36%
Cardiotoxicity	3%	11%	0%	1%
Developmental/reproductive toxicity	27%	47%	47%	15%
Endocrine disruptor	43%	47%	33%	16%
Hematotoxicity	5%	5%	0%	4%
Immunotoxicity	9%	11%	13%	3%
Kidney toxicity	22%	26%	20%	12%
Liver toxicity	35%	42%	33%	14%
Neurotoxicity	28%	21%	40%	33%
Spleen toxicity	4%	0%	0%	1%
Thyroid toxicity	6%	5%	13%	2%

Table 7. Overview of the human toxic endpoints associated with the chemicals used in *Daphnia* studies. The amount of chemicals linked to a human endpoint is stated between brackets. The percentage of chemicals capable of affecting certain endpoints are listed for each omics technique. For the transcriptomics, proteomics and metabolomics studies, respectively 66%, 54% and 64% of all chemicals investigated are linked to endpoint using the human endpoint list. Endpoints tested >10% more frequently than expected (compared to their occurrence in the human endpoint list) are indicated in blue.

Endpoint	Transcriptomics (104 of 157 chemicals linked to a human relevant endpoint)	Proteomics (7 of 13 chemicals linked to a human relevant endpoint)	Metabolomics (18 of 28 chemicals linked to a human relevant endpoint)	Percentage of chemicals in human endpoint list
Carcinogenic	42%	14%	6%	36%
Cardiotoxicity	5%	14%	6%	1%
Developmental/reproductive toxicity	41%	57%	50%	15%
Endocrine disruptor	27%	29%	28%	16%
Hematotoxicity	16%	0%	0%	4%
Immunotoxicity	28%	14%	6%	3%
Kidney toxicity	29%	14%	6%	12%
Liver toxicity	25%	29%	50%	14%
Neurotoxicity	48%	43%	22%	33%
Spleen toxicity	0%	0%	0%	1%
Thyroid toxicity	0%	0%	17%	2%

Remarkably, more than 50% of tested chemicals had multiple toxicity endpoints (not shown). Of all chemicals with endpoints in the list only 22% had multiple endpoints, so it seems that chemicals with multiple endpoints are favoured in -omics experiments. This would confound the assignment of MoA to toxicity endpoints. Additionally, results indicate that preferably chemicals with human relevance for developmental/reproductive toxicity, endocrine disruption, and liver toxicity are studied (endpoints in Table 5, 6, 7 indicated in blue). In *Daphnia*, in addition to those, immunotoxicity and kidney toxicity were relatively often studied (Table 5). In both zebrafish and Daphnia, from all chemicals that could be used in the study, around 60% had an endpoint associated in the human toxicity endpoint list (see Table 6 and 7 headers). There is no remarkable difference in the use of chemicals with a particular human endpoint between proteomics, metabolomics, and transcriptomics in either *Daphnia* or *Danio* (Table 6 and 7). Because of the relative low amount of chemicals tested in proteomics and metabolomics, these numbers are also more vulnerable for chance effects and these numbers should therefore not be overinterpreted.

3.2.3 Experimental setups used in studies

As it is known that the experimental setup can influence results, such differences were determined for all metabolomics experiments for the two selected organisms, as an example. These were differences in the age/life stage of exposure, the duration of exposure, and the tissues used for extraction of the biomarkers. In *Danio rerio* studies, 37.5% of the metabolomics studies were performed on larvae or embryos at ≤120 hours post fertilization (hpf), this is the cut-off beneath which the organisms are not considered experimental animals. Exposure durations varied heavily from 6 h to 118 h. Studies using ≤120 hpf *Danio rerio*, all extracted the biomarkers from the whole body. This is most probably due the difference in size (3.9 mm at 120 hpf vs. 4-5 cm for adults). Of the studies using >120 hpf *Danio rerio*, 60% extracted specific tissues; liver (78%), muscles (11%), intestines (11%), head (11%), or blood serum (11%). Exposure durations in studies investigating the metabolomics profiles in specific tissues are usually 7 days, but 4 and 90 days were also seen. The studies using the whole zebrafish body (40%), often extract at 7 days post fertilization (dpf), and expose for either the full period or 2-3 days.

Almost all *Daphnia magna* studies extracted metabolites using the whole body. 70% of the studies used adult (10-28 day old) *Daphnia*'s versus 30% of the studies using <24 h old neonates at the start of the experiment). Several studies stated that, depending on the chemical studied, diverging responses were seen for neonates and adults (Wagner et al., 2017). More specifically, *Daphnia* over 8 days of age were deemed more appropriate for monitoring purposes as they had lower metabolomics variation compared to younger *Daphnia*'s (Jeong et al., 2019). Almost all studies, followed an acute exposure protocol with exposure durations between 24-48 h. One neonate study exposed their *Daphnia*'s for 21 days to determine developmental effects (Wang et al., 2018a), and one study explicitly shortened the exposure duration to 6 h for monitoring purposes (Jeong et al., 2019).

This implies that the use of model species in experimental set-ups is not standardized as of yet. Results on expressed features cannot be directly compared for most studies. This is confirmed by a study of Schüttler et al. (2017) for transcriptomics. In their re-analysis of zebrafish embryo data they found that overlap of differentially transcribed genes in response to chemical stress across independent studies is generally low. The most commonly differentially transcribed genes appear in less than 50% of all treatments across studies.

3.3 Sensitivity of model organisms to MoA and Toxicological endpoint

A sensitive species will be more likely to respond to the low concentrations of chemicals in (sources for) drinking water. In addition to providing a list with chemicals and their established MoA, the MoAtox database also lists acute toxicity. This is expressed as the dose in an aqueous solution at which 50% of a response is exhibited, EC/LC50. The acute toxicity is listed for different chemicals for three fish species, fish in general, *Daphnia* and invertebrates in general (Barron et al., 2015). Zebrafish is not present. Therefore, instead of comparing zebrafish and *Daphnia,* we make a general statement whether different fish species were more, or less sensitive than *Daphnia*. We linked established toxicological endpoints (Appendix I) and MoA of chemicals (Barron et al., 2015) to the acute toxicity data of Barron et al. (2015) of those chemicals. For nematodes, there is no data available in these databases. Queirós et al. (2019) showed that *C. elegans* was more sensitive to Mercury (Hg) than *Daphnia* but otherwise for several other metals less sensitive. We did not find a comprehensive database in which chemicals are linked to formal MIE (see Box 1).

Figure 5 and 6 show that for the MoA and toxicity endpoints, *Daphnia* shows in general higher and in some cases significantly higher sensitivity than (one of) the fish species. In other words, *Daphnia* is affected at lower concentrations of chemicals for several MoAs and toxicity endpoints. It can be observed that the fathead minnow is least sensitive and this coincides with the fact that this species is robust to pollution and is therefore regularly used to test heavily polluted water like wastewater. For some MoAs all species showed low sensitivity (e.g. narcosis chemicals, spleen toxicity) and for others all species showed high sensitivity (e.g. neurotoxicity, cardiotoxicity, liver toxicity related chemicals). It is noticeable that all median LC/EC50 effects took place above the 10 µg/l concentration level. From these results we cannot deduct if the sensitivity for chemicals at low concentrations is similar. This is relevant because concentrations in water may be lower.

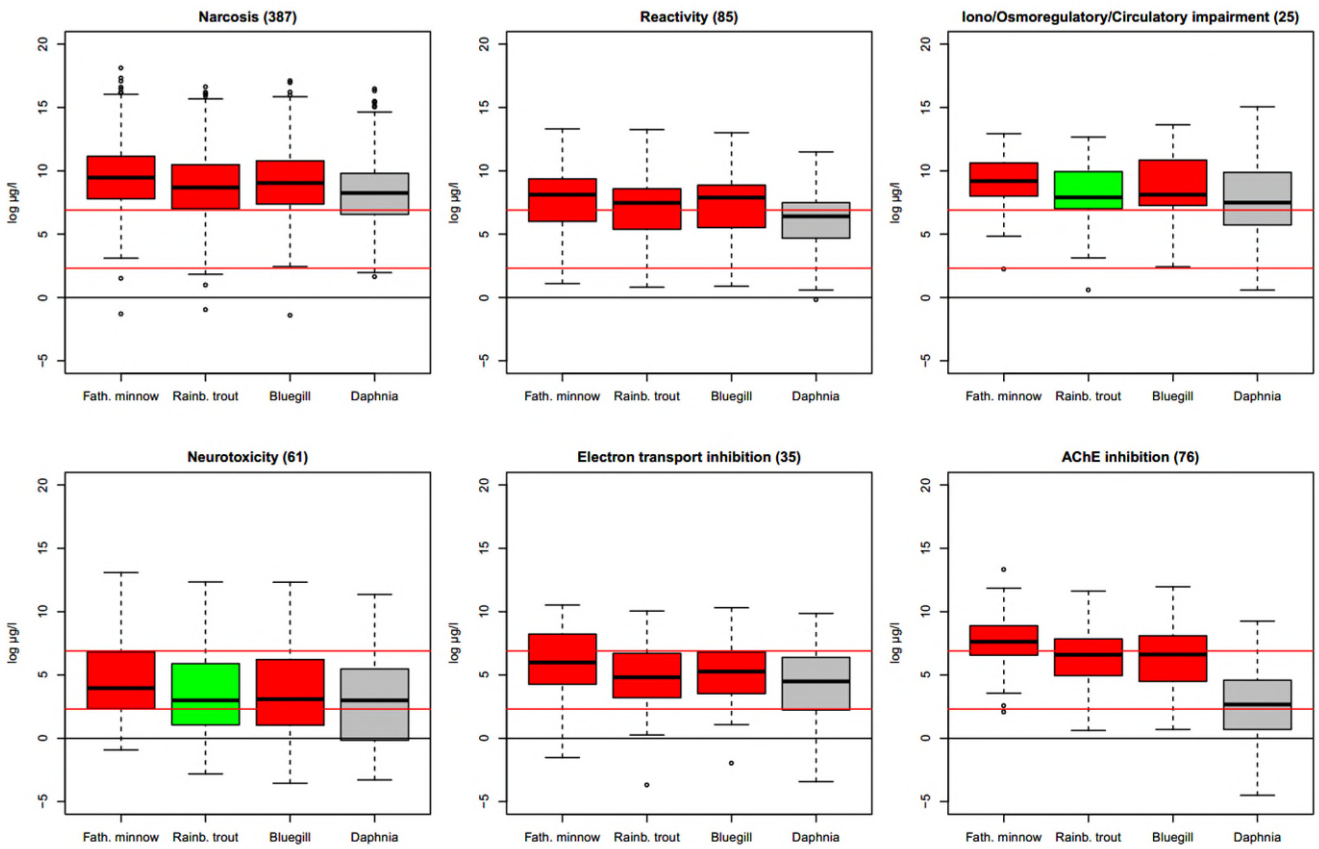


Figure 5. The sensitivity of three different fish species and Daphnia for chemicals associated to broad mode of action. Between brackets is the number of chemicals that the data is based on. On the y-axis are reported EC/LC50 concentrations in natural logarithm (µg/l). The black line is 1 µg/l. The red lines are respectively 10 µg/l and 1000 µg/l. A red box indicates a significant difference in sensitivity from *Daphnia*. Green boxes are not significantly different.

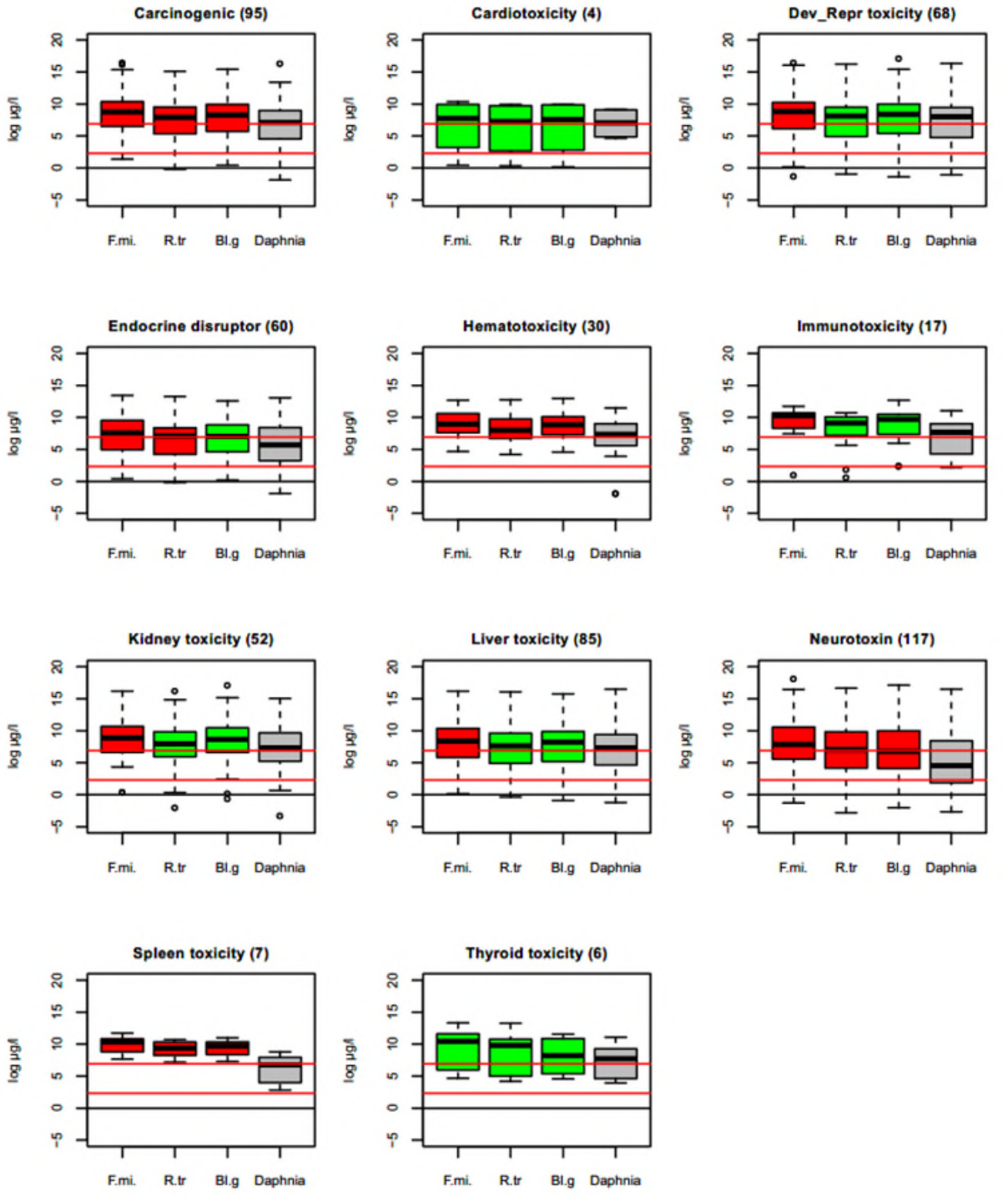


Figure 6. The sensitivity of fish species and *Daphnia* for chemicals associated to human toxicity endpoints. Between brackets is the number of chemicals that the data is based on. On the y-axis are reported EC/LC50 concentrations in natural logarithm (µg/l). The black line is 1 µg/l. The red lines are respectively 10 µg/l and 1000 µg/l. A red box indicates a significant difference in sensitivity from *Daphnia*. Green boxes are not significant.

The question remains open whether this sensitivity to chemicals in terms of EC/LC50 is also reflected in the expression of relevant and distinguishing features. Such features are needed as biomarkers to predict MoA, chemical class, or human toxicological endpoints at environmentally relevant concentrations.

4 Options for use of -omics techniques in risk assessment

4.1 Possibilities in a regulatory context

Both human health and ecological risk assessment are needed to determine the environmental safety of chemicals. In Europe, the Registration, Evaluation, Authorization and restriction of Chemicals (REACH) regulation assures that substances, mixtures of substances and articles containing substances placed on the EU market do not cause adverse effects on human health and on the environment (Queirós et al. 2019). This regulation however concerns individual products and chemicals. In the aquatic environment including drinking water sources, countless possible combinations of chemicals can be present, with possible additive adverse effects. Environmental policy makers have recognised mixture toxicity as a major issue in environmental risk assessment of chemicals, also realizing that it is not feasible to assess the risk of each possible combination separately (Ankley and Edwards 2018; Sahlin 2018). The Water Framework Directive, Europe's most comprehensive instrument for water policy, stresses the necessity of a 'good ecological and chemical status' of water. The chemical status, however, is monitored by comparing concentrations of individual priority chemicals with environmental quality standards (EQS). Considering this, effect-based monitoring should be advocated and considered when the Water Framework Directive will be revised (Brack et al. 2017).

Regulatory bodies increasingly see the advantages of whole-organism assays to do an effect-based evaluation of risk. The U.S. EPA accepts toxicogenomics data for establishing mechanisms of toxicity for chemicals (Queirós et al., 2019). The EFSA panel recommends the use of specific aquatic and sediment-dwelling organisms. Additionally, the nematode species *Caenorhabditis elegans* has been recognized as a meaningful add-on to the ecotoxicological test battery in pesticides environmental risk assessment at least in the EU (EFSA, 2017), considering the availability of standard test guidelines to assess both acute (ASTM International 2014) and chronic (ISO 2010) toxicity. Still, *C. elegans* has yet to be routinely included in environmental risk assessment test batteries (Queirós et al., 2019). *Daphnia* and fish already are a part of test batteries, although not routinely included.

To be used as risk assessment instrument for water quality related to human health, -omics signatures that link to impacts on human toxicity endpoints will be very important. It is reasonable that proteomics and metabolomics will receive increased regulatory acceptance compared to transcriptomics as it measures impacts on a higher level closer to the phenotype (Sahlin, 2018). Metabolomics has the added advantage that metabolites are relatively often not species-specific (Martyniuk and Simmons, 2016) and this will benefit the inter-species extrapolation of results (Sahlin, 2018).

MoA can be considered as an intermediary link between chemical exposure and final effect. Several ways to define MoA exist and these have potential to strengthen risk assessment with functional omics methods. However, these ways lack harmonization and therefore often give contradictory results (Kienzer et al., 2017). The use of functional omics data in AOPs in general is not well studied, leading to a lack of clear linkage and relevance to such adverse outcomes. The difficulty to establish a clear link between established MoA in model organisms and chemically induced human toxic endpoints was shown in Chapter 3 as well. Hence, it is necessary to further establish pathways to connect functional omics and phenotype, in the form of MoA or AOPs (Buesen et al. 2017; Sahlin, 2018; Schriks et al., 2011) and to see which of these MoA or AOPs in model organisms have relevance and predictive value for chemically induced human toxicity endpoints. Future research and developments of omics as well as a standardized way to define MoAs are necessary before gaining regulatory acceptance (Sahlin, 2018).

A concrete promising framework towards assessing possible risks for a mixture of chemicals in water is recently provided by Xia et al. (2020) and Wang et al. (2018) (Figure 7 and 8). This method works with groups of biomarkers (the 'signature' features) that represent a specific pathway that can be triggered following exposure to chemicals. With a concentration range of exposure a 'Point of Departure' (POD) it is established, relating to transcriptional potency. This POD was set for individual genes to the dose where the gene expression surpassed 1 standard deviation from the control. For pathways the POD was set to the dose where 20% of the genes of the pathway were expressed (EC20). This means the potency of a particular chemical (mixture) to trigger responses can be established for the 'signature' features that indicate a specific pathway, or for a single biomarker feature. For a chemical or a mixture, the size of the effect on different pathways can be compared. For individual compounds, a typical pattern can be established of the potency of chemicals to trigger particular pathways (see Figure 7 and 8). For two mixtures of 12 compounds with different MoA, it was shown that mixture effects per pathway were as expected if these were calculated by compound addition (Xia et al., 2020; Wang et al., 2018).

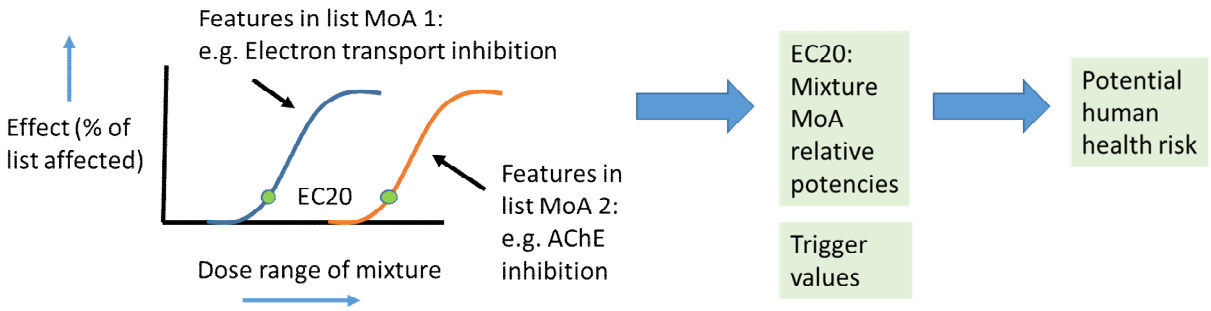


Figure 7. Point of departure (POD) analysis of pathways (Xia et al., 2020). This leads to relative potencies based on EC20 of the unknown mixture to affect different MoA (see Figure 8 for a visualization from Xia et al., 2020) in a model species. As an addition, these MoA can be linked to human relevant health effects, and trigger values can be derived.

To increase applicability of such a method for risk assessment, pathways or other (novel) established expressed 'signatures' should be (statistically) linked to the likelihood of presence of chemicals that are linked to a particular human relevant toxicological endpoint. The more pathways can be identified that link to a human toxicological endpoint, the wider the range of possible adverse mixture effects is that can be signaled by the assay.

Routine implementation of new technologies is supported if data can be easily interpreted. This approach can also be expanded with 'trigger values' as indication for a potential risk associated with the environmental mixture studied. Trigger values are generally derived based on human relevant guideline concentrations (Béen et al., 2020). Functional omics data for a sufficient number of individual compounds, preferably under similar experimental setup, is required to derive trigger values for these types of bioassays. For each of the effects (pathways representing MoA) of interest, a reference compound can be tested in the model species alongside the unknown mixtures. In this way, the relative potency of that particular effect in the mixture can be established more precisely than by using separate control studies. It might also be possible to derive -omics trigger values for specific toxicity endpoints.

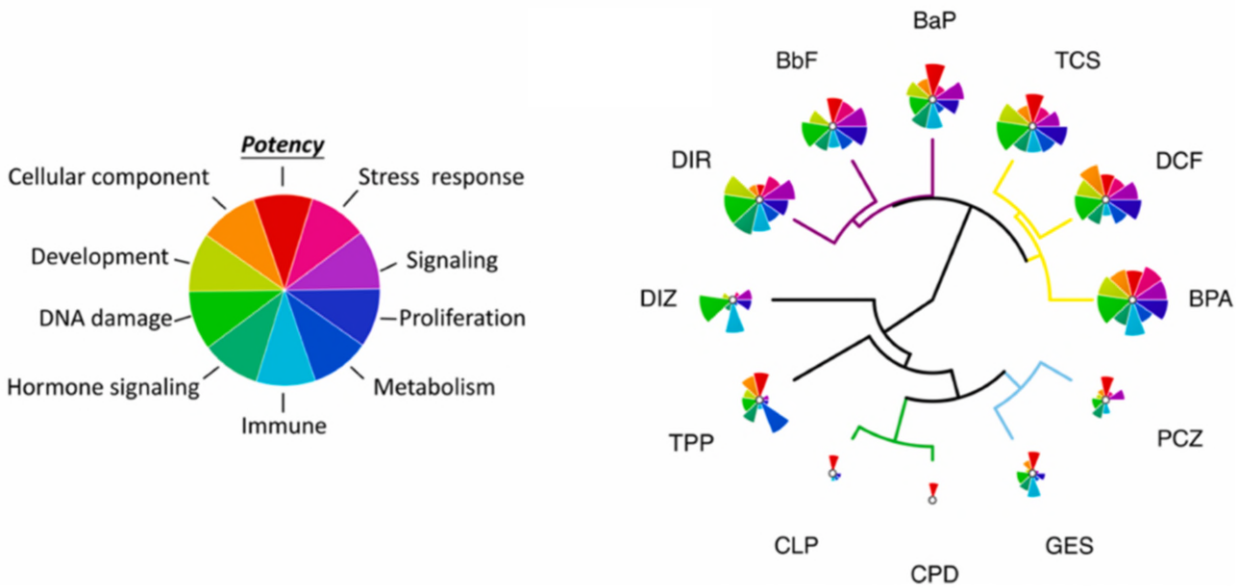


Figure 8. Clustering of twelve chemicals based on established PODs (figure from Xia et al, 2020). Similarly, an unknown mixture can be assessed based on the measured effects on these pathways. Trigger values could signal possible risks per detected MoA. Abbreviations: TCS: Triclosan; BaP: Benzo(a)pyrene; BbF: Benzo(b)fluoranthene; BPA: Bisphenol A; CLP: Chlorophene; CPD: Cyprodinil; DFC: Diclofenac; DIR: Diuron; GES: Genistein; PCZ: Propiconazole; TPP: Triphenylphosphate, DIZ: Diazinon.

If the concentration response curve gives an unreliable identification of the MoA, a tiered approach can be considered. Relevant in vitro bioassays can target suspected MoA more precise *after* the initial whole organism omics assay (Fang et al. 2020).

4.2 Recommended steps for implementation

Although omics technologies are being regarded as the future generation of effect-based monitoring tools (Leung, 2018) they are currently not yet incorporated in any environmental monitoring scheme (Sahlin, 2018). An uptake in water quality regulation will boost the use of omics as an effect based monitoring tool.

Several issues need to be solved before an omics technique can be implemented as an early warning for risks in (sources for) drinking water.

It is essential to develop standardized frameworks for consistency and to reduce methodological (see paragraph 3.2.3 in this report) and technical uncertainties. With sufficient standardization, -omics data can be reliably interpreted and integrated into regulatory frameworks across borders (Buesen et al., 2017). A standard model organism, with a standard study design and standard data processing per omics- technique, such as proposed in the previous paragraph 4.1. will help interpretability of an effect-based monitoring tool based on omics.

Standard data processing possibilities will also depend on established and verified 'signatures' of features in –omics data that are indicative and predictive for the presence of a mixture that may trigger a human relevant toxicological endpoint (either a statistical link or via an established AOP), for presence of chemicals of particular chemical classes, or for presence of chemicals with particular MoA. Such signatures can be identified in further scientific research with functional omics techniques. Since functional omics methods are still relatively novel, especially in assessing drinking water quality, still more experiments with exposures of chemicals with known mechanisms and adverse outcomes are needed to reliably establish these mechanistic links. Only in very few studies the responses

of functional omics assays were connected to possible adverse outcomes at higher levels (e.g. reproduction, impaired growth) in a predictive approach (Sahlin, 2018). Proteomics and metabolomics are arguably the best candidates for an omics- based risk assessment tool because they have closer link to phenotype of an adverse effect (such as impaired growth or reproduction) (Sahlin, 2018). These are however also the techniques that are least developed and tested.

To bridge the gap between expression of features in model organisms and human health risks of low-level mixtures of chemicals in water, we will first have to establish MoA of individual chemicals, to be able to link biomolecular expression (-omics) patterns common for chemicals with that MoA. In paragraph 3.1 a poor linkage between MoA and human toxicological endpoints in chemicals was demonstrated, and this also needs improvement. For this purpose public and non-public data on effects of individual chemicals in *in vitro* tests can be analysed. If a chemical triggers a biomolecular response in an *in vitro* test, the MoA that is addressed in that *in vitro* test can be assigned to that chemical. In addition or as an alternative, QSARS can be developed that predict the MoA. Another way to establish MoA of chemicals is to look at existing information on MoA of chemicals in (REACH) risk assessment reports (Heusink et al., 2020) and literature (Lee et al., 2015; Spaan et al., 2019). These MoAs in turn will be linked via a predictive model to the relevant human endpoints established in literature of these chemicals. The resulting toxicological knowledge of the meaning/relevance of expressed biomolecular pathways after relative short exposures in model organisms can be used in risk assessment.

Lastly, Leung (2018) stressed that there is a lack of expertise with omics technology. Hence, there is a great need for increased communication and exchange of knowledge and experiences between researchers and laboratories to enhance and improve the use of omics (Sahlin, 2018), and to support their implementation for water quality monitoring.

5 Conclusions and recommendations

- The advantages of using bioassays based on functional omics techniques merit making the effort to build such a framework for the assessment of mixture toxicity in drinking water or in sources for drinking water.
- Progress in technology is continuously being made, the field of functional omics remains relevant. Most efforts are, however, made in ecotoxicology (no human relevance) and drug development (no mixtures).
- Although there is a lot of data on functional omics following chemical exposure of model organisms or *in vitro* bioassays, the lack of standardization of experiments cause high variability in the expression of features. For uptake in any regulatory framework a standardized approach needs to be developed.
- The biggest hurdle for application of omics to drinking water quality monitoring is the lack of study towards the expression of biomolecules in relation to AOP of chemicals. The MoA is a useful concept to act as an intermediary between the initially expressed biomolecules and the adverse outcome.
- Proteomics and metabolomics are the best candidates to use in water quality monitoring of adverse effects because they have a closer link to phenotype, however existing data is still overrepresented by transcriptomics.
- The ease of operation, their intermediary position in trophic levels and the high sensitivity to chemicals with different MoA and toxicological endpoint suggests that *Daphnia* is an appropriate model organism. It remains to be seen if signatures for all relevant toxicity endpoints could be established in this model organism. Nematodes could be a good candidate as well for the same reasons, however these are less frequently studied in a water quality context.
- The development of a concrete framework for risk assessment of chemical mixtures in water for drinking water (sources) was proposed, based upon the work of Xia et al. (2020).

6 References

Van Aggelen G., Ankley G.T., Baldwin W.S., et al., 2010. Integrating Omic technologies into aquatic ecological risk assessment and environmental monitoring: Hurdles, achievements, and future outlook. Environ Health Perspect 118:1–5.

Alexander-Dann B., Pruteanu L.L., Oerton E., et al. Developments in toxicogenomics: understanding and predicting compound-induced toxicity from gene expression data. Mol Omics. 2018;14(4):218-236. doi:10.1039/c8mo00042e

Altenburger R., Walter H., Grote M. (2004) What contributes to the combined effect of a complex mixture? Environ Sci Technol 38:6353–6362. doi: 10.1021/es049528k

Ankley G.T., Edwards S.W. (2018) The adverse outcome pathway: A multifaceted framework supporting 21st century toxicology. Curr Opin Toxicol 9:1–7. doi: 10.1016/j.cotox.2018.03.004

Ankley G.T., Bencic D.C., Breen M.S., et al (2009) Endocrine disrupting chemicals in fish: Developing exposure indicators and predictive models of effects based on mechanism of action. Aquat Toxicol 92:168–178. doi: 10.1016/j.aquatox.2009.01.013

Antczak P., Jo H.J., Woo S., Scanlan L., Poynton H., Loguinov A., Chan S., Falciani F., Vulpe C. (2013) Molecular toxicity identification evaluation (mTIE) approach predicts chemical exposure in Daphnia magna. Environ Sci Technol. 47(20):11747-56. doi: 10.1021/es402819c. Epub 2013 Oct 7. PMID: 23875995.

Bahamonde, P.A., Feswick, A. Isaacs, M.A. Munkittrick K.R. and Martyniuk C.J. (2016). "Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program." Environmental Toxicology and Chemistry 35(1): 20-35.

Baken, K.A., Sjerps, R.M.A., Schriks, M., van Wezel, A.P. (2018) Toxicological risk assessment and prioritization of drinking water relevant contaminants of emerging concern. Env. International, 118 (2018), pp. 293-303, 10.1016/j.envint.2018.05.006

Barron M.G., Lilavois C.R., Martin T.M. (2015) MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. Aquatic Toxicology 161 102–107.

Béen, F., Pronk, T.E., Louisse, J., Houtman, C., van der Velden-Slootweg, T., van der Oost, R., Milou M.L. Dingemans, M.L. (2020) Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality. BTO report, in preparation.

Brack W, Escher BI, Müller E, et al. (2018) Towards a holistic and solution-oriented monitoring of chemical status of European water bodies: how to support the EU strategy for a non-toxic environment? Environ Sci Eur 30:. doi: 10.1186/s12302018-0161-1

Brunner AM, Bertelkamp C, Dingemans MML, Kolkman A, Wols B, Harmsen D, Siegers W, Martijn BJ, Oorthuizen WA, Ter Laak TL. (2020) Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment. Sci Total Environ. 705:135779. doi: 10.1016/j.scitotenv.2019.135779.

Booij, P., Lamoree M.H., Sjollema, S.B., de Voogt, P., Schollée J.E., Vethaak A.D., Leonards, P.E.G. (2014) Non-target Metabolomic Profiling of the Marine Microalgae Dunaliella tertiolecta After Exposure to Diuron using Complementary High Resolution Analytical Techniques. Current Metabolomics, 2014, 2, 000-000

Buesen R, Chorley BN, da Silva Lima B, et al (2017) Applying 'omics technologies in chemicals risk assessment: Report of an ECETOC workshop. Regul Toxicol Pharmacol 91:S3–S13. doi: 10.1016/j.yrtph.2017.09.002

Canzler, S., J. Schor, W. Busch, K. Schubert, U. E. Rolle-Kampczyk, H. Seitz, H. Kamp, M. von Bergen, R. Buesen and J. Hackermüller (2020). "Prospects and challenges of multi-omics data integration in toxicology." Archives of Toxicology 94(2): 371-388.

Campos A., Tedesco S., Vasconcelos V., Cristobal S. (2012) Proteomic research in bivalves. Towards the identification of molecular markers of aquatic pollution. J Proteomics 75:4346–4359. doi: 10.1016/j.jprot.2012.04.027

Cedergreen N., 2014. Quantifying Synergy: a systematic review of mixture toxicity studies within Environmental Toxicology. PLoS ONE 9(5), e96580

Dingemans MM, Baken KA, van der Oost R, Schriks M, van Wezel AP. (2018) Risk-based approach in the revised European Union drinking water legislation: Opportunities for bioanalytical tools. Integr Environ Assess Manag. 15(1):126-134. doi: 10.1002/ieam.4096. PMID: 30144268.

Dardenne F., Nobels I., De Coen W., Blust R., 2008. Mixture toxicity and gene inductions: Can we predict the outcome? Environmental Toxicology and Chemistry, 27 (3), 509518.

EFSA Scientific Committee (2019) Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals https://doi.org/10.2903/j.efsa.2019.5634

Fang, W. Peng, Y., Yan, L, Xia P., Zhang,X. (2020) A Tiered Approach for Screening and Assessment of Environmental Mixtures by Omics and In Vitro Assays. Environ. Sci. Technol. 54, 12, 7430–7439

Fernie, A.R. and Stitt M. (2012). On the discordance of metabolomics with proteomics and transcriptomics: Coping with increasing complexity in logic, chemistry, and network interactions. Plant Physiology 158(3): 1139-1145.

Feussner, I. and A. Polle (2015). "What the transcriptome does not tell — proteomics and metabolomics are closer to the plants' patho-phenotype." Current Opinion in Plant Biology 26: 26-31.

García-Ortega, L. F. and O. Martínez (2015). "How Many Genes Are Expressed in a Transcriptome? Estimation and Results for RNA-Seq." PLOS ONE 10(6): e0130262.

Jeong, T.Y., Simpson,M.J. (2019) Daphnia magna metabolic profiling as a promising water quality parameter for the biological early warning system,Water Research 166, 115033. https://doi.org/10.1016/j.watres.2019.115033.

Hägerbäumer A, Höss S, Heininger P, Traunspurger W. Experimental studies with nematodes in ecotoxicology: an overview. J Nematol. 2015;47(1):11-27.

Heusinkveld, H., Braakhuis, H., Gommans, R., Botham, P., Corvaro, M., van der Laan, J.W., Lewis, D., Madia, F., Manou, I., Schorsch, F., Wolterink, G., Woutersen, R., Corvi, R., Mehta, J., and Luijten,M. (2020) Towards a mechanism-based approach for the prediction of nongenotoxic carcinogenic potential of agrochemicals, Critical Reviews in Toxicology, 50:9, 725-739, DOI: 10.1080/10408444.2020.1841732

Hermsen S.A.B., Pronk T.E., van den Brandhof E.J., van der Ven L.T., Piersma A.H. (2011) Chemical class-specific gene expression changes in the zebrafish embryo after exposure to glycol ether alkoxy acids and 1,2,4-triazole antifungals. Reprod Toxicol. 32(2):245-52. doi: 10.1016/j.reprotox.2011.05.010.

Hermsen, S.A.B., Pronk, T.E., van den Brandhof, E.J., van der Ven, L.T.M., Piersma, A.H., 2013. Transcriptomic analysis in the developing zebrafish embryo after compound exposure: individual gene expression and pathway regulation. Toxicol. Appl. Pharmacol. 272, 161–171. https://doi.org/10.1016/j.taap.2013.05.037.

Houtman CJ, Ten Broek R, van Oorschot Y, Kloes D, van der Oost R, Rosielle M, Lamoree MH. (2020). High resolution effect-directed analysis of steroid hormone (ant)agonists in surface and wastewater quality monitoring. Environ Toxicol Pharmacol. 80:103460. doi: 10.1016/j.etap.2020.103460.

Kienzler A., Barron M.G., Belanger S.E., Beasley A., Embry M.R. (2017) Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches Environ. Sci. Technol. 2017, 51, 10203−10211

Kools, S.A.E., van Loon, A.H., Sjerps, R.M.A., Rosenthal, L.P.M. (2019) De kwaliteit van bronnen van drinkwater in Nederland KWR rapport - KWR 2019.072

Kar S, Leszczynski J. Exploration of Computational Approaches to Predict the Toxicity of Chemical Mixtures. Toxics. 2019 Mar 19;7(1):15. doi: 10.3390/toxics7010015. PMID: 30893892; PMCID: PMC6468900.

Krasnov, A., Afanasyev, S., Oikari, A. (2007). Hepatic Responses of Gene Expression in Juvenile Brown Trout (Salmo trutta lacustris) Exposed to Three Model Contaminants Applied Singly and in Combination. Environ Toxicol Chem 26(1),100-109.

Kümmerer, K., Dionysiou, D.D., Olsson, O., Fatta-Kassinos, D. (2018) A path to clean water. Science: Vol. 361, Issue 6399, pp. 222-224 DOI: 10.1126/science.aau2405

Lalone C.A., Villeneuve D.L., Burgoon L.D., Russom C.L., Helgen H.W., Berninger J.P., Tietge J.E., Severson M.N., Cavallin J.E., Ankley G.T. (2013) Molecular target sequence similarity as a basis for species extrapolation to

assess the ecological risk of chemicals with known modes of action. Aquat Toxicol. 144-145:141-54. doi: 10.1016/j.aquatox.2013.09.004.

Lee J.W., Won E.-J., Raisuddin S., Lee J.-S. (2015) Significance of adverse outcome pathways in biomarker-based environmental risk assessment in aquatic organisms. J Environ Sci (China) 35:115–127. doi: 10.1016/j.jes.2015.05.002

Leung K.M.Y. (2018) Joining the dots between omics and environmental management. Integr Environ Assess Manag 14:169–173. doi: 10.1002/ieam.2007

McCarroll, N., E., Reaves, M. Manibusan, D. C.,Wolf, S. D. Hester, et al. (2009). Use of toxicogenomic data by the US EPA Office of Pesticide Programs Cancer Assessment Reivew Committee. Presented at Society of Toxicology, Baltimore, MD, March 15 - 19, 2009. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NHEERL&dirEntryId=199726

Martyniuk C.J., Simmons D.B. (2016) Spotlight on environmental omics and toxicology: a long way in a short time. Comp Biochem Physiol - Part D Genomics Proteomics 19:97–101. doi: 10.1016/j.cbd.2016.06.010

Martins, C., Dreij K. and Costa P. M. (2019). The State-of-the Art of Environmental Toxicogenomics: Challenges and Perspectives of "Omics" Approaches Directed to Toxicant Mixtures. International journal of environmental research and public health 16(23): 4718.

Menzel R, Swain SC, Hoess S, et al. Gene expression profiling to characterize sediment toxicity--a pilot study using Caenorhabditis elegans whole genome microarrays. BMC Genomics. 2009;10:160. Published 2009 Apr 14. doi:10.1186/1471-2164-10-160

Poynton, C.H., Wintz, H., Vulpe, C.D. (2008). Progress in ecotoxicogenomics for environmental monitoring, mode of action and toxicant identification. In Comparative toxicogenomics, Hogstrand, C. and Kille, P., Eds, Elsevier Science: Amsterdam, The Netherlands.

Queirós L, Pereira JL, Gonçalves FJM, Pacheco M, Aschner M, Pereira P. (2019) Caenorhabditis elegans as a tool for environmental risk assessment: emerging and promising applications for a "nobelized worm". Crit Rev Toxicol. 49(5):411-429. doi:10.1080/10408444.2019.1626801

Simões, T., Novais, S.C., Natal-da-Luz, T. et al. (2018) An integrative omics approach to unravel toxicity mechanisms of environmental chemicals: effects of a formulated herbicide. Sci Rep 8, 11376. https://doi.org/10.1038/s41598-018-29662-6

Sahlin, S. (2018) The next-generation of aquatic effect-based monitoring? - A critical review about the application, challenges and barriers with omics in field. Examination Project Work, Faculty of Health and Life Sciences.

Scanlan, L. D., Loguinov, A. V., Teng, Q., Antczak, P., Dailey, K. P., Nowinski, D. T., Kornbluh, J., Lin, X. X., Lachenauer, E., Arai, A., Douglas, N. K., Falciani, F., Stapleton, H. M., & Vulpe, C. D. (2015). Gene transcription, metabolite and lipid profiling in eco-indicator daphnia magna indicate diverse mechanisms of toxicity by legacy and emerging flame-retardants. Environmental science & technology, 49(12), 7400–7410. https://doi.org/10.1021/acs.est.5b00977

Schaap, M.M., Wackers, P.F.K., Zwart, E.P. et al. A novel toxicogenomics-based approach to categorize (non-)genotoxic carcinogens. Arch Toxicol 89, 2413–2427 (2015). https://doi.org/10.1007/s00204-014-1368-6

Schriks, M., R. van der Oost and C. Houtman (2011). The application of toxicogenomics for (drinking) water quality assessment, KWR Water Research Institute.

Schroeder AL, Martinović-Weigelt D, Ankley GT, et al (2017) Prior knowledge-based approach for associating contaminants with biological effects: A case study in the St. Croix River basin, MN, WI, USA. Environ Pollut 221:427–436. doi: 10.1016/j.envpol.2016.12.005

Schüttler A, Reiche K, Altenburger R, Busch W. The Transcriptome of the Zebrafish Embryo After Chemical Exposure: A Meta-Analysis. Toxicol Sci. 2017;157(2):291-304. doi:10.1093/toxsci/kfx045

Soong, R.a, Nagato, E.a, Sutrisno, A.a, Fortier-Mcgill, B.a, Akhter, M.a,b, Schmidt, S.c, Heumann, H.c, Simpson, A.J.(2015) In vivo NMR spectroscopy: Toward real time monitoring of environmental stress. Magnetic Resonance in Chemistry, 53(9): 774-779.

Spaan K., Haigis A.-C., Weiss J., Legradi J. (2019) Effects of 25 thyroid hormone disruptors on zebrafish embryos: A literature review of potential biomarkers. Science of the Total Environment, 656 , pp. 1238-1249.

Steuer, A., L. Brockbals and T. Kraemer (2019). "Metabolomic Strategies in Biomarker Research–New Approach for Indirect Identification of Drug Consumption and Sample Manipulation in Clinical and Forensic Toxicology?" Frontiers in Chemistry 7.

Tan, K.C., Ipcho, S. V. S. Trengove R. D., Oliver R. P.and Solomon P. S. (2009). "Assessing the impact of transcriptomics, proteomics and metabolomics on fungal phytopathology." Molecular plant pathology 10(5): 703-715.

Timmers, P.H.A., van den Bulk, J., Heijnen, L., Kardinaal, W.E.A., Sollie, S., Medema, G.J. – Microbiële vingerafdruk voor bronopsporing in oppervlaktewaterbeheer – Water Matters (2020) juni, p.4-7)

Timp, W., TImp, G. (2020) Beyond mass spectrometry, the next step in proteomics Science Advances Vol. 6, no. 2, eaax8978

Trapp J, Armengaud J, Salvador A, et al (2014) Next-generation proteomics: Toward customized biomarkers for environmental biomonitoring. Environ Sci Technol 48:13560–13572. doi: 10.1021/es501673s

Tufi, S. Leonards, P., Lamoree,M.,de Boer,J., Legler, J., Legradi J. (2016a) Changes in Neurotransmitter Profiles during Early Zebrafish (Danio rerio) Development and after Pesticide Exposure Environ. Sci. Technol. 2016, 50, 3222−3230

Tufi, S., Wassenaar,P.N.H., Osorio V., de Boer, J., Leonards, P.E.G., Lamoree M.H. (2016b) Pesticide Mixture Toxicity in Surface Water Extracts in Snails (Lymnaea stagnalis) by anin Vitro Acetylcholinesterase Inhibition Assay and Metabolomics Environ. Sci. Technol. 50(7):3937-44.

Wagner, N.D., Simpson, A.J. and Simpson, M.J. (2017), Metabolomic responses to sublethal contaminant exposure in neonate and adult Daphnia magna. Environ Toxicol Chem, 36: 938-946. https://doi.org/10.1002/etc.3604

Wang P, Ng QX, Zhang H, Zhang B, Ong CN, He Y. (2018a) Metabolite changes behind faster growth and less reproduction of Daphnia similis exposed to low-dose silver nanoparticles. Ecotoxicol Environ Saf. 2018 Nov 15;163:266-273. doi: 10.1016/j.ecoenv.2018.07.080.

Wang, P., Xia, P. Yang, J., Wang, Z., Peng, Y., Shi, W. Villeneuve, D.L.,Yu, H. and Zhang, X. (2018) A Reduced Transcriptome Approach to Assess Environmental Toxicants Using Zebrafish Embryo Test Environmental Science & Technology 2018 52 (2), 821-830 DOI: 10.1021/acs.est.7b04073

Wishart, D. S. (2007). "Human Metabolome Database: completing the 'human parts list'." Pharmacogenomics 8(7): 683-686.

van Wezel, A. P., ter Laak, T. L., Fischer, A., Bäuerlein, P. S., Munthe, J., & Posthuma, L. (2017). Mitigation options for chemicals of emerging concern in surface waters; operationalising solutions-focused risk assessment. Environmental Science: Water Research & Technology, 3(3), 403-414.

Xia, P. Zhang,H. Peng,Y. Shi, W., Zhang, X. (2020) Pathway-based assessment of single chemicals and mixtures by a high-throughput transcriptomics approach. Environment International Volume 136, 105455

I Appendix I: Methods

I.I List of human relevant toxicological endpoints of chemicals

The list of endpoints was constructed with the downloaded chemicals from online sources in Table 1 Appendix I with help of a custom in-house R-script. From the list of 1763 chemicals, 1379 chemicals had a single endpoint, 231 had two endpoints, 96 had three endpoints, 65 had four or more endpoints.

Table 1 Appendix I. Overview and short description of the sources for uptake of chemicals in the list of chemicals with a human relevant endpoint.

Weblink	Source number	Source Description
https://oehha.ca.gov/proposition-65/crnr/chemicals-meeting-critieria-listing-developmental-and-reproductive-toxicants	S1	California office of environmental health hazard Assessment: source U.S. EPA
https://saferchemicals.org/wp-content/uploads/2014/11/mindthestore.org-full-list-toxic-chemicals.pdf	S2	Organization Safer chemicals, healthier families: source diverse
https://www.health.state.mn.us/communities/environment/childenvhealth/docs/chlist/mdhchc2019.pdf OR https://www.health.state.mn.us/communities/environment/childenvhealth/tfka/highconcern.html#list	S3	Minnesota Department of Health Toxic Free Kids Act Chemicals of High Concern June 2019
https://echa.europa.eu/nl/ed-assessment	S4	ECHA European Chemicals Agency
https://www.chemsafetypro.com/Topics/Restriction/UN_list_identified_endocrine_disrupting_chemicals_EDCs.html	S5	published by the International Panel on Chemical Pollution (IPCP) commission by the UN Environment
https://sinlist.chemsec.org/endocrine-disruptors/	S6	ChemSec Sinlist (REACH compliant information)
https://edlists.org/the-ed-lists/list-i-substances-identified-as-endocrine-disruptors-by-the-eu?page=0	S7	Recognised by EU in PPPR, BPR or REACH (2020)
https://cfpub.epa.gov/ncea/iris/search/index.cfm Search filters: Route of exposure: oral. Selection of endpoint of interest in 'Organ/System Affected'. For most endpoints, all chemicals were selected. Except for spleen toxicity, thyroid toxicity and kidney toxicity. For spleen toxicity, 'immune' was selected and only the chemicals of which the 'Critical Effect or Tumor Type' contained spleen, splenic or splenomegaly. For thyroid toxicity, 'endocrine' was selected and only the chemicals of which the 'Critical Effect or Tumor Type' contained thyroid or TSH. For kidney toxicity, 'urinary' was selected and only the chemicals of which the 'Critical Effect or Tumor Type' contained kidney, renal, nephro, glomerulonephritis.	S8	Integrated Risk Information System (IRIS) from EPA (United States Environmental Protection Agency)

https://www.uspharmacist.com/article/drug-induced-acute-renal-failure	S9	Howell HR (2007) Drug-Induced Acute Renal Failure. US Pharm. 32(3):45-50.
https://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/genotox/ECVAM_Ames_positives_DB.xls	S10	JOINT RESEARCH CENTRE big data platform: ECVAM European Centre for the Validation of Alternative Methods data
https://doi.org/10.5281/zenodo.2648768	S11	NORMAN List of Human Neurotoxins
https://www.env-health.org/IMG/pdf/06tl9094page.pdf (DOI:10.1016/S01406736(06)69665-7)	S12	Grandjean P, Landrigan PJ (2006) Developmental neurotoxicity of industrial chemicals, The Lancet Chemicals (n=201) known to be neurotoxic in man
http://192.82.104.231/documents/ehc/ehc/ehc180.htm	S13	WHO International program on chemical safety (IPCS) INCHEM Internationally peer reviewed chemical safety information
https://haz-map.com/heptox1.htm	S14	NCI National Library of Medicine: HAZ-MAP Information on Hazardous Chemicals and Occupational Diseases: Industrial Chemicals Associated with Acute Liver Injury as the Primary Toxic Effect: Illness or Death
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4783956/	S15	Björnsson ES (2016) Hepatotoxicity by Drugs: The Most Common Implicated Agents Int J Mol Sci. 17(2): 224
https://comptox.epa.gov/dashboard/chemical_lists/neurotoxins	S16	EPA NEURO: Neurotoxicants Collection from Public Resources
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6598532/	S17	Petejova et al. (2019) Acute toxic kidney injury Ren Fail. 41(1): 576–594.
https://www.dovepress.com/drug-induced-impairment-of-renal-function-peer-reviewed-fulltext-article-IJNRD	S18	Pazhayattil GS, Shirali A (2014) Drug-induced impairment of renal function Int J Nephrol Renovasc Dis. 2014;7:457-468 https://doi.org/10.2147/IJNRD.S39747
http://www.inchem.org/documents/ehc/ehc/ehc119.htm#PartNumber:5	S19	WHO international programme on chemical safety, Environmental health criteria

		119. Principles and methods for the assessment of nephrotoxicity associated with exposure to chemicals

Table 2 Appendix I. Overview of the number of chemicals per human relevant endpoint, and the sources for the chemicals.

Human Endpoint	Identified Substances (1763)	Source
Developmental / Reproductive toxicity	267	1,2,3,8
Endocrine disruptor	278	2,3,4,5,6,7,8
Carcinogen	637	2,3,8
Neurotoxin	579	3,8,11,12,16
Immunotoxicity	48	3,8,13
Liver toxicity	249	3,8,14,15
Hematotoxicity	74	3,8
Kidney toxicity	210	3,8,9,17,18,19
Thyroid toxicity	40	3,8
Spleen toxicity	17	3,8
Cardiotoxicity	14	3,8

II Appendix II: Comparative overview of functional omics studies in literature and databases

II.I Overview of functional omics studies in literature

To obtain insight into the availability of omics data in literature, a literature search was performed using Scopus.com. Out of 1.296.725 (11-08-2020) "*omics" articles, 15.011 articles were found using the following search term: (TITLE-ABS-KEY (metabolomic*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*)) AND TITLE-ABS-KEY (expos*). 'TITLE-ABS-KEY (expos*)' was added to select the articles describing changes in physiology due to exposure only. The number of articles was reduced to 3.629 articles following selection of organisms exposed to a chemical using 'AND TITLE-ABS-KEY (chemical*)', to filter out organisms exposed to environmental conditions. Limiting the search to articles solely, 3156 articles of potential interest were found. To select articles which use organisms as bioassays for monitoring purposes 'AND (*monitor*) was added. In total 1147 articles were selected using the following search query:

(TITLE-ABS-KEY (metabolomics*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*)) AND TITLE-ABS-KEY (expos*) AND TITLE-ABS-KEY (chemical*) AND (*monitor*) AND (LIMIT-TO (DOCTYPE , "ar")).

The 1147 selected articles were used to determine the amount of transcriptomic, proteomic and metabolomics articles published at that point (11-08-2020). In addition, the organisms most represented in the selected articles were looked into to further specify the search for omics data relevant to this project. The following model search query was used:

(TITLE-ABS-KEY (metabolomics*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*)) AND TITLE-ABS-KEY (expos*) AND TITLE-ABS-KEY (chemical*) AND (*monitor*) AND (TITLE-ABS-KEY (zebrafish*) OR TITLE-ABS-KEY ("danio rerio")) AND (LIMIT-TO (DOCTYPE , "ar")).

Following selection of the organisms of interest, the following three search queries were used to select the final batch of articles:

(TITLE-ABS-KEY (metabolomic*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*) OR TITLE-ABS-KEY (genomic*)) AND TITLE-ABS-KEY (expos*) AND TITLE-ABS-KEY (chemical*) AND (*monitor*) AND (TITLE-ABS-KEY ("danio rerio") OR TITLE-ABS-KEY (zebrafish)) AND (LIMIT-TO (DOCTYPE , "ar")),

(TITLE-ABS-KEY (metabolomic*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*) OR TITLE-ABS-KEY (genomic*)) AND TITLE-ABS-KEY (expos*) AND TITLE-ABS-KEY (chemical*) AND (*monitor*) AND (TITLE-ABS-KEY (*mussel*) OR TITLE-ABS-KEY (mytilus) OR TITLE-ABS-KEY (unionidae) OR TITLE-ABS-KEY (mytilidae)) AND (LIMIT-TO (DOCTYPE , "ar")), and

(TITLE-ABS-KEY (metabolomic*) OR TITLE-ABS-KEY (transcriptomic*) OR TITLE-ABS-KEY (proteomic*) OR TITLE-ABS-KEY (genomic*)) AND TITLE-ABS-KEY (expos*) AND TITLE-ABS-KEY (chemical*) AND (*monitor*) AND (TITLE-ABS-KEY (daphnia)) AND (LIMIT-TO (DOCTYPE , "ar")).

The analyses of the results of this literature search can be found in Chapter 3.

II.II Overview of functional omics studies in databases

Omics data is often shared in repositories online. Per omics technique, several data repositories exist. Most of them are integrated into the open source platform Omics Discovery Index (https://www.omicsdi.org/). As 23 repositories are integrated in OmicsDI, this platform was used for this project. OmicsDI combines the following data repositories: EGA (European Genome-Phenome Archive; genomics), MetaboLights (metabolomics), PeptideAtlas (proteomics), GPMDB (Global Proteome Machine Database; proteomics), MetabolomeExpress (metabolomics), GNPS (Global Natural Products Social Molecular Networking; metabolomics), ArrayExpress of Functional Genomics Data (genomics), LINCS (Library of Integrated Network-based Cellular Signatures; multiomics), PAXDB (Protein Abundance Database; proteomics), JPOST Repository (Japan ProteOme Standard Repository; proteomics), ExpressionAtlas (genomics), Pride (proteomics), MetabolomicsWorkbench (metabolomics), dbGaP (database of Genotypes and Phenotypes; genomics), GEO (NCBI Gene Expression Omnibus; genomics), ENA (Eurpean Nucleotide Archive; transcriptomics/genomics), BioModels (models), EVA (European Variation Archive (genomics), MassIVE (proteomics), NODE (National Omics Data Encyclopedia; transcriptomics/genomics), Physiome Model Repository (models), Cell Collective (models), FAIRDOMHub (models). The OmicsDI database is a convenient platform combining omics data from various omics platforms. It is possible to search specific chemicals, yet not exposure information. In addition, by clustering many databases, overlap between hits were regularly seen. In several cases, both the dataset behind the article and the article were up-loaded separately. In addition, often datasets of multiple chemicals were uploaded per single chemical, making the database less convenient for the selection of additional databases/articles with ≥6 chemicals. To determine if additional omics datasets could be gathered from OmicsDI, the open source platform integrating datasets coming from multiple omics studies, was searched. From the total of 504 *Danio rerio* datasets, 5 additional datasets containing ≥ 6 chemicals could be found. One dataset was included in the literature search. For *Daphnia* (*magna* + *pulex*), 2 additional datasets containing ≥ 6 chemicals could be found, whilst 3 were also found during the literature search. The additional datasets only contained genomics/transcriptomics data and belonged to articles which were not found during the literature search due to the addition of the search term '*monitor*'. The addition of this term to the literature search query reduced the total hits from 3156 to 1147, which made the search more manageable.

To determine if additional datasets were available through OmicsDI, the following search query was used:

expos* AND ((zebrafish OR (Danio rerio)) OR (mytilus OR perna OR mussel) OR ((Daphnia magna) OR (Daphnia pulex)))

At the time of the search (07-09-2020), this query found 923 hits, which could be broken down into 504 *Danio rerio* datasets, 99 *Daphnia magna* datasets, 25 *Daphnia pulex* datasets, 29 *Mytilus galloprovincialis* datasets, 6 *Mytilus edulis* datasets and 2 *Mytilus californianus* datasets. As the addition of 'AND chemical*', to remove exposure to environmental factors (as with the literature search query), reduced the number of hits by 41%. The search term above was taken as a starting point. Per organism, a list of articles containing ≥ 6 different chemicals were selected. The datasets on the list were integrated with the list of literature articles selected in Chapter 3.