HBM4EU science and policy for a healthy future

HORIZON2020 Programme
Contract No. 733032 HBM4EU

Emerging chemicals framework

Deliverable Report

AD 16.2

WP 16 – Emerging Substances

Deadline: March 2018

Upload by Coordinator: 04 April 2018

Entity	Name of person responsible	Short name institution	Date [Received]
Coordinator	Marike Kolossa-Gehring	UBA	04 April 2018
Grant Signatory	Robert Barouki	INSERM	30 March 2018

Entity	Name of person responsible	Short name institution	Date [Approved]
Pillar Leader	Robert Barouki	INSERM	30 March 2018
Work Package Leader	Jean-Philippe Antignac	INRA	30 March 2018
Task leader	Roel Vermeulen	IRAS	30 March 2018

Responsible author	Jeroen Meijer	E-mail	j.meijer@uu.nl
Short name of institution	IRAS	Phone	+31-302539474
Co-authors	Jelle Vlaanderen (IRAS), Roel Vermeulen (IRAS), Marja Lamoree (VU), Timo Hamers (VU), Jean-Philippe Antignac (INRA), Herbert Oberacher (Medical university of Innsbruck), Laurent Debrauwer (INRA), Adrian Covaci (UAntwerpen), Martin Krauss (UFZ)		

Table of contents

1 Authors and Acknowledgements

Lead authors

Jeroen Meijer, Jelle Vlaanderen and Roel Vermeulen (IRAS)

Marja Lamoree, Timo Hamers (VU)

Jean-Philippe Antignac, Laurent Debrauwer (INRA)

Herbert Oberacher (Medical university of Innsbruck)

Adrian Covaci (UAntwerpen)

Martin Krauss (UFZ)

2 Abstract/Summary

Thousands of chemicals are present in the environment, a range of which has potential deleterious impact on human health. Some of these chemicals have been present in the environment for years but their toxicity has not yet been described, others have only recently been developed. Collectively, we call these 'emerging chemicals'. As human biomonitoring programmes are typically restricted to inclusion of tens to hundreds of chemicals, a big challenge is to identify which emerging chemicals should be prioritized, based on both exposure data and toxicological profile, for inclusion in biomonitoring programmes. This deliverable describes an approach to develop a framework that can be used to address that challenge.

Due to recent methodological improvements (untargeted mass spectrometry screening), we are able to measure the presence of thousands of small molecules in the environment and in human biological samples. Many of these small molecules reflect (fragments) of chemicals, though the identification of which molecule belong to which chemical is a challenge. A fast, but putative identification can be acquired by comparing the accurate mass of a measured small molecule with the accurate mass of a known compound. The first stage of this framework incorporates a database that contains accurate masses of >60.000 potential emerging chemicals and their metabolites. This database can be used to putatively identify emerging chemicals in newly collected human samples, through a suspect screening approach.

Once there is insight into which chemicals are potentially present in human samples, in stage two of the framework their toxicological profile can be predicted based on their structure using computational tools (e.g. QSARs). The combined information about probability of presence in human samples and probability of toxicity will be used to rank compounds and the identify and toxicity of most highly ranked compounds will be further validated in stage 3 of the framework using standard methodology.

The framework also incorporates Effect Directed Analysis as a complementary approach to detect emerging chemicals. This approach is based on fractionation of a biological sample and identification of the compounds present in the fractions that present toxicity (in a specific biological pathway).

Using the information gathered from suspect screening, toxicity prediction, and Effect Directed Analysis, the output of the framework is a set of chemicals that is prioritized for incorporation in existing and future human biomonitoring programmes.

3 Introduction

The attention of chemical monitoring has shifted over the last decades from the analysis of legacy compounds to a large collection of newly produced and discovered chemicals [1]. These newer compounds are often referred to as 'emerging chemicals' or 'chemicals of emerging concern'. The term 'emerging' is relative, since an emerging chemical of concern a few decades ago might not be classified as an emerging chemical today. Therefore, a clear definition is given to emerging chemicals within HBM4EU following the sentiment of Sauvé & Desrosiers, 2014. In general, three qualifications have been given to what an emerging contaminant is.

The first qualification includes chemicals that are 'truly new'. These chemicals were not previously known or produced and have only recently appeared in scientific literature. A second qualification includes chemicals that were already known to exist, but concerns and issues were not fully apprehended until recently. The final classification includes chemicals of which concerns were raised previously, however, due to new information on behaviour or toxicity, new concerns are raised [2].

The considerable number of emerging chemicals poses a challenge for risk assessors and policy makers. In addition, the list of emerging chemicals continues to expand and change because of both the production and discovery of new chemicals each day and the improved knowledge on current and past chemicals. Furthermore, the lack of knowledge on the behaviour and toxic effect of these chemicals in humans is also problematic [2]. To deal with these issues regarding, the NORMAN network was founded. The aim of that project was to set up a network to promote the exchange of information on emerging chemicals and to propose a methodology to prioritise emerging chemicals using environmental samples [3]. HBM4EU WP16 will capitalise on these efforts to produce a sustainable framework for the detection of emerging chemicals in human biological samples.

To achieve this goal, several steps need to be undertaken, that can be seen as an aggregation of the different parts of the workflow expected to be developed within HBM4EU WP16. These include:

- The development of a suspect screening database of emerging chemicals for which to date no HBM data is available.
- The development of a MS reference library susceptible to annotate a maximum of these markers,
- The generation of untargeted high resolution mass spectrometry (HRMS) profiles from human samples
- Linking of occurrence data with toxicological and hazard data through computational approaches and effect-directed-analysis (EDA).

While other actions defined in WP16 are globally designed for short to medium term results obtained from low scale data, the present action is positioned as a more long term view including a pre-validation based on larger scale data. In this context, the present document contains a global scheme addressing these steps and aims to provide a basis to eventually achieve the WP16 goal in term of integration and sustainability.

4 Emerging chemicals framework

The goal of this document is to provide a global framework and work scheme to identify and prioritize emerging chemicals based on their related exposure levels and toxicological profile. The wide range and the considerable number of possible structures in combination with large knowledge gaps pose a major challenge to monitor emerging chemicals. In addition, for human biomonitoring, phase I and/or phase II (conjugates) metabolites are commonly more relevant biomarkers than the parent compounds. Yet, these biotransformation products are often not adequately considered in the identification process [4]. Several steps and considerations are necessary to effectively deal with the emerging chemical problem. These will be discussed in more detail in the following chapters. As a global synoptic, the proposed emerging chemicals framework is illustrated in *Figure 1*.

The first step in the emerging chemicals framework is to create a database of *a priori* known/suspected emerging chemicals. The composition of database will be primarily based on already existing databases focussing on emerging chemicals in environmental and food compartments. Subsequently, untargeted HRMS profiles (already existing data) generated from human biological samples will be screened to look for the selected chemicals and metabolites in the suspect database. Due to the sheer number of possible suspects, a putative identification approach will be used. This screening will be based as first on already existing HRMS data and will aim to provide tentative identifications as well as to improve the overall identification capabilities available. These improvements include the development and validation of new software and to advance and enlarge spectral databases as described in actions two and three. Furthermore, the advanced data mining workflows will be validated. Finally, this action aims to detect the most abundant metabolites which can represent relevant exposure markers of their parent compound. This focusing process results in a more specified suspect database which will be used as guidance for the identification in new biological samples.

Compounds which are most frequently detected in the new samples will be subjected to toxicological profiling using computational tools. Furthermore, the identification of frequently detected compounds must be validated with reference compounds in future efforts. By doing this, suspect screening will be converted to target screening providing level 2 identifications which is a minimum inclusion criteria for compounds to be included in future HBM programs [5]. Finally, EDAs will be performed as an alternative method to identify emerging chemicals that show some kind of biological activity. This integrated framework will result in a refined and prioritized database of chemicals of emerging concern based on their presence in the human body and their toxicology to be included in future HBM programs.

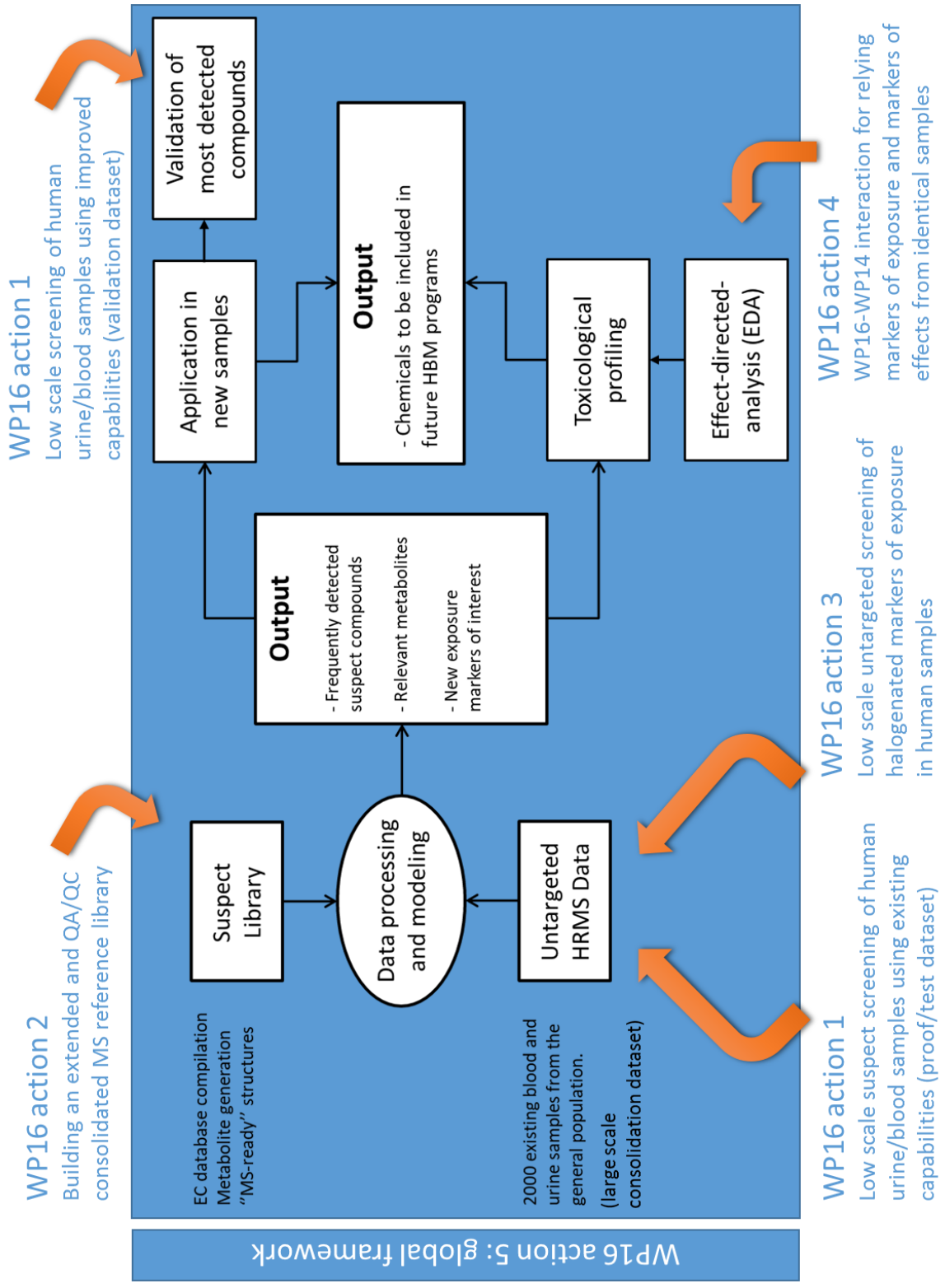

Figure 1: The emerging chemical framework and its relations with the current sub-actions developed in WP16.

5 Prioritized emerging chemicals

5.1 Suspect database

The first aspect of the emerging chemicals framework is to establish a database of chemicals that can serve as a marker of exposure to emerging chemicals. The amount of possible emerging chemicals is immense, even more so if metabolites are also considered. Therefore, a comprehensive harmonized database with the focus on emerging chemicals and their expected ion masses would be greatly beneficial as a guidance for chemical identification [4]. A detailed description of the composition and the methodology on the development of the prioritized database of emerging chemicals was published by WP16 as deliverable 16.1 in December 2017 [6].

A number of initiatives and sources have published databases of emerging chemicals previously. Many of these initiatives focused on environmental exposure and some included also toxicological information. However, no global database that focuses on human exposure to emerging chemicals has been published as of yet. In order to produce an all-inclusive database of emerging chemicals in humans, existing data sources of emerging chemicals were combined. In addition, several considerations were taken into account. The sources used and the considerations are extensively described in Deliverable 16.1 [6].

Chemical data sources, as used for the development of the prioritization database of emerging chemicals, often include diverse identifiers. For example, sources include different names to describe a compound (systematic, trivial or product names), CAS numbers (active, alternate or deleted) and database identifiers (SMILES, InChI Strings, InChIKeys, MOL files) [6]. For suspect screening of emerging chemicals, a molecular formula is sufficient. However, toxicity prediction and the calculation of other properties require more extensive data on molecular structure. In addition, commercial forms of chemicals sometimes comprise salts or mixtures, which have to be converted in the free, MS-detectable forms. Therefore, to efficiently screen for suspect compounds, prioritization databases should provide a ‘’MS-ready’’ form of the compound as described by Schymanski and Williams, 2017. In these MS-ready structures, counter ions, water, neutral forms of bases and acids etc. are removed while still retaining the link with the ‘’non-MS-ready’’ form which often contain more information. The ‘’MS-ready’’ form is therefore a direct input for the expected ion mass to be identified in HRMS data [7].

The majority of data sources addressing emerging chemicals are based on the presence in the external environment. However, in the human body, most compounds undergo biotransformation which can result in several metabolites. Therefore, in the light of human biomonitoring, these metabolites should be considered in addition to their parent compounds. Metabolites resulting from biotransformation in humans are in most cases not included in environmental based data sources. Consequently, possible metabolites should be generated to be included in the database of prioritized emerging compounds. As a result the database will expand considerably because in some cases more than ten metabolites can be produced from one parent compound [6]. In case they are not yet known or identified, these metabolites can be generated using for example computational tools in addition to bioreactor, analytical techniques and field studies [8].

5.2 Suspect screening

The produced suspect database will be used as a guiding rail for suspect screening. As this list consists of ‘’MS-ready’’ structures, it should be directly applicable to untargeted HRMS data. Initially, the produced chemical database will be applied to HRMS data of approximately 2000 existing samples. These include blood, urine, breast milk and cord blood samples collected at different times

from several cohorts. This action aims to identify frequently occurring suspect compounds including metabolites and to identify temporal trends in emerging chemical occurrence. Eventually, based on the result of the suspect screening, a database of new exposure markers of interest will be developed as possible HBM targets.

The definite identification of chemicals in HRMS involves the use of reference standards. These standards should be measured under the same conditions and methodology as the suspected compounds [9]. This is a very laborious and costly process. The next best approach is making use of tandem mass spectral information stored in libraries. Accordingly, a spectral reference library (collection of MS and MS/MS spectra with appropriate consolidated QA/QC dispositions and criteria) will have to be produced in parallel to the elaborated inventory of potential markers of interest, in the scope of identifying unambiguously a maximal number of those markers from experimental HRMS profiles. This MS reference library, ideally with sufficient flexibility to integrate various HRMS platforms and sufficient opening capabilities for shared use (e.g. compliant with the Massbank requirements), appears as a corner stone of the proposed framework, and a main strategical element for its successful implementation. Besides the conceptual and scientific effort necessary for structuring this library (taking into account some already existing international standards for instance in the metabolomics community), this action will require a huge technical investment for characterizing and indexing such high number of chemicals (those for which a pure standard will be available) to feed this collection of reference spectra. Mutualisation of already existing standard collections and analytical capabilities will be then necessary to achieve this goal, as well as a good coordination between different actors for generating the significant volume of not yet existing reference data.

Now, this approach is not realistically feasible for the short term efficient suspect screening of an extensive number of possible chemicals. On the other hand, searching only the mass-to-charge (m/z) ratio of the monoisotopic mass of a suspect compound will result in a high number of false-positive matches either for compounds with similar m/z ratios, but different molecular formulas that are not resolved by the MS instrument used or for structural isomers for the correct molecular formula . Consequently, this ideal strategy has to be completed by other approaches for short term results. In this respect, a computational approach will be considered to establish a confidence in possible chemical identification [10].

One example of such computational approach was published by Uppal et., 2017. In this approach, additional data besides monoisotopic mass is used to decrease the number of false positives and increase the confidence level of a possible chemical identification. Additional data includes elution characteristics, adducts and isotopic patterns. Furthermore, abundance ratios of isotopes, multiply charged form and multimers are used to assign a confidence score to database matches. These database matches can then be categorized based on their confidence level and used to prioritize suspect chemicals for further validation. This approach is included in an R package called ‘’xMSannotator’’ [10]. Comparable approaches exist and will be compared.

5.3 Application and validation

The newly identified exposure markers of interest will be applied and validated in new samples. Comparable to the existing samples, the new samples are taken from the general population and consist mainly of blood and urine. Subsequently, toxicological information will be collected on these chemicals and used to suggest a number of emerging chemicals to be included in future HBM programs. This toxicological information can be gained using computation tools like literature search and Quantitative Structure-Activity Relationships (QSARs) approaches. In addition, chemicals that are detected in a large number of samples and have high detection confidence, will be validated by

using reference standards. The validation of all detected compound will not be feasible due to a lack of resources and for most metabolites, no reference standards are available. However, validation of a set of detected compounds is desirable to evaluate the putative identification approach and to reach level 2 identification confidence which is a requisite for compounds to be classified as an ‘’identified exposure marker’’ and submitted to the next steps.

5.4 Effect-directed-analysis

An additional method of detecting chemicals of toxicological concern is by using EDA. EDA combines toxicological and analytical methods to identify the chemicals that cause toxicity in a mixture. Samples are fractionated and subjected to bioassays in order to detect biological activity. The biological active fractions are fractionated again and the process is repeated. Eventually, chemical analysis is performed on the bio-active fractions [11]. As a result only the chemicals that are active in the bioassay are measured and chemicals that show no biological activity are ignored as opposed to suspect screening. EDA approach is toxicity driven whereas this is not primarily the case with suspect screening. Therefore, EDA might be able to identify emerging chemicals related to the used specific toxicological endpoints which were not considered primarily with suspect screening and vice versa.

A wide range of EDA approaches have been developed over the years using single or combinations of different in vitro bioassays. One of the WP16 tasks was to create an inventory of these different *in vitro* bioassays and to make a selection of the most suitable ones for EDA work in human samples for the identification of emerging chemicals. A detailed description of this action including the completed inventory and selected approaches was published by WP16 as additional deliverable 16.3 in December 2017. The best methodologically established *in vitro* bioassays used in EDA for the identification of emerging chemicals in human samples are the arylhydrocarbon receptor (AhR) bioassay and the transthyretin (TTR) binding assay. The methodologies of these assays are well established and frequently used in human samples [12].

6 Conclusion

The proposed long term view and sustainable framework aims to integrate different pieces of workflow for finally addressing the need for large scale screening of emerging chemicals in human matrices. This global framework is facing a number of significant challenges and difficulties, among which the huge resources necessary to build an extended spectral reference library for unambiguous identification of the considered markers of exposure, or the real integration of chemical and toxicological profiling considering in addition their related high *versus* low throughput capabilities.

The objective of WP16 is to deal with this complexity by combining several focused sub-actions as a contribution to this implementation of a global framework, and to use multiple stage sampling (for test/proof of concept/feasibility, consolidation and validation) also with various sample types for enlarged applicability and robustness of the proposed framework.

7 References

1. Daughton, C.G., *Pharmaceuticals and Personal Care Products in the Environment: Overarching Issues and Overview*, in *Pharmaceuticals and Care Products in the Environment*. 2001, American Chemical Society. p. 2-38.
2. Sauvé, S. and M. Desrosiers, *A review of what is an emerging contaminant.* Chemistry Central Journal, 2014. **8**(1): p. 15.
3. Brack, W., V. Dulio, and J. Slobodnik, *The NORMAN Network and its activities on emerging environmental substances with a focus on effect-directed analysis of complex environmental contamination.* Environmental Sciences Europe, 2012. **24**(1): p. 29.
4. Zedda, M. and C. Zwiener, *Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools.* Analytical and Bioanalytical Chemistry, 2012. **403**(9): p. 2493-2502.
5. Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics : Official journal of the Metabolomic Society, 2007. **3**(3): p. 211-221.
6. Vlaanderen, J., et al., *Prioritised list of known emerging chemicals*. 2017, IRAS.
7. Schymanski, E.L. and A.J. Williams, *Open Science for Identifying "Known Unknown" Chemicals.* Environmental Science & Technology, 2017. **51**(10): p. 5357-5359.
8. Tyrkkö, E., *In silico methods in prediction of drug metabolism, mass fragmentation, and chromatographic behavior : Application to toxicological drug screening by liquid chromatography/time-of-flight mass spectrometry*, in *Department of Forensic Medicine*. 2014, University of Helsinki: Helsinki.
9. Schymanski, E.L., et al., *Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis.* Analytical and Bioanalytical Chemistry, 2015. **407**(21): p. 6237-6255.
10. Uppal, K., D.I. Walker, and D.P. Jones, *xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data.* Analytical Chemistry, 2017. **89**(2): p. 1063-1067.
11. Brack, W., et al., *Effect-directed analysis supporting monitoring of aquatic environments — An in-depth overview.* Science of The Total Environment, 2016. **544**: p. 1073-1118.
12. Lamoree, M., et al., *Direct effect based approaches applied to the screening of emerging substances*. 2017, VU-IVM DTU.