# Using REACH registration data to rank the environmental emission potential of persistent and mobile organic chemicals

Stefanie Schulze,[1] Daniel Sättler,[2] Michael Neumann,[2] Hans Peter H. Arp,[3] Thorsten Reemtsma,[1] Urs Berger[1,*]

[1] Department of Analytical Chemistry, Helmholtz Centre for Environmental Research – UFZ, Permoserstrasse 15, 04318 Leipzig, Germany

[2] Section IV 2.3 Chemicals, German Environment Agency – UBA, Wörlitzer Platz 1, 06844 Dessau-Roßlau, Germany

[3] Norwegian Geotechnical Institute – NGI, Postboks 3930 Ullevål Stadion, 0806 Oslo, Norway

* Corresponding author e-mail, urs.berger@ufz.de; phone, +49 341 235 4654; fax, +49 341 235 450822

## Abstract

Organic chemicals that are persistent and mobile in the aquatic environment exhibit a hazard to contaminate drinking water resources. In this study an emission score model was developed to rank the potential of substances registered under the REACH legislation to be emitted into the environment. It was applied to a list of 2167 REACH registered substances that were previously identified to be persistent and mobile organic chemicals (PMOCs) in groundwater or to be hydrolyzed to form transformation products fulfilling the PMOC criteria. The emission score model is based on the tonnage placed on the European market and on seven emission-related use characteristics (high release to environment, wide dispersive use, intermediate use, closed system use, professional use, consumer use, and substance in article), reported in the companies' registrations under REACH. Applying the model resulted in a list of 1110 substances (936 PMOCs and 174 precursors to PMOCs) that were estimated to be released into the environment, while 1054 substances had indicators of negligible environmental emissions and 3 substances could not be evaluated due to severe data gaps. The 936 PMOCs and the 174 precursors were ranked in two lists with regard to their emission potential. The model was shown to be fit for purpose in terms of suggesting and prioritizing substances for scientific investigations with a focus on environmental water quality. Though targeted for PMOCs, the presented scoring system is illustrative of how REACH registration data can be used to assess the emission potential of various substances.

Keywords: Prioritization, Environmental emissions, Transformation products, Drinking water, Polar contaminants, Water quality

## 1 Introduction

The number of chemicals produced and used in industrial or in consumer applications worldwide is continuously increasing.[1] Within the ambit of the European chemicals regulation REACH,[2] more than 16,500 substances are currently registered (as of October 2017) with a manufactured or imported volume in the European Union (EU) exceeding one ton per year. Many more substances will be registered by the final registration deadline May 31st, 2018 (according to Article 23(3) of the European Parliament Regulation (EC) 1907/2006).[2] According to Article 10 of the REACH legislation, manufacturers, importers and downstream users of substances in the EU are obliged to collect information on substance properties and uses and to report them in a registration dossier to the European Chemicals Agency (ECHA). The extent of information to be provided depends on the volumes of the substances manufactured in or imported into the EU (including import of substances within products according to Article 7 of the REACH legislation), which have to be reported as well. Besides information on the identity of the registered substance, the dossiers can contain further information, e.g. on persistency, on toxicity and on exposure concentrations within different environmental compartments, depending on the legal requirements. In case a substance is exclusively used for synthesis of another substance under the conditions defined in articles 17 and 18 of REACH, the legislation offers the possibility to register such a substance as an 'isolated intermediate' with reduced requirements regarding the information to be provided for the registration.

Trace-analytical methods to determine contaminants in environmental samples are expensive and time-consuming in their development and application. They are further limited in the number of substances that can be analyzed in a single run. Thus, chemical analytical methods are often restricted to groups of substances with similar physical-chemical properties. Given the vast number of substances in use, it is evident that only a tiny fraction can be monitored

by chemical analysis. It is therefore of utmost importance and relevance to prioritize substances of highest concern for environmental monitoring programs. Prioritization by modeling has proven to be a powerful tool.[3,4] Most prioritization studies reported in literature so far have ranked substances with regard to their human exposure potential, as a prerequisite for risk assessment.[5-12] Other modeling studies attempted to identify emerging contaminants based on substance properties such as persistence in the environment or the potential to bioaccumulate.[13-16] Collectively, such studies model the hazard of the substances, and only few studies so far have explicitly attempted to model the potential of a large set of organic chemicals to be released into the environment, i.e. emissions, which is a key component of assessing risk.[5,12,16-19] Arnot and co-workers[5] ranked about 12,000 organic substances for human exposure (intake rates and internal human concentrations) using quantitative estimates of chemical emissions. The same study included an uncertainty analysis highlighting the greatest source of uncertainty in the model calculations were the estimated emissions of chemicals from manufacture and use. Bitsch et al.[12] used tonnage bands and Environmental Release Categories (ERC) registered under REACH as well as biodegradation and potential bioaccumulation to identify chemicals that are of potential health concern and likely to occur in the food chain. McLachlan et al.[16] ranked substances for both estimates of actual human exposures and concentrations in the environment. Also this study used quantitative estimates of emissions and the rankings were revised with expert judgement. The rankings were further used to prioritize chemicals for target analysis as an evaluation of the results of the model-based screening.[16] Fischer and co-workers[19] developed an 'Emission Index' model that was later evaluated by Undeman et al.[20] for its ability to rank contaminants found in sewage treatment plants. The model was found to be of limited use in its current form, based on only weak correlations between the Emission Indices and the observed levels of the chemicals in the sewage treatment plants. The approach by Breivik et al.[17] was met with difficulties in

application, as access to consistent input data was claimed to be "fragmented or even impossible". The input parameters considered intuitive by Breivik et al. for inclusion in any approach to screen substances for emissions are a) total quantities in commerce, b) chemical function, and c) physical-chemical properties. The same study also demonstrated the importance of having up-to-date and accurate information on quantities for developing reliable emission scenarios.[17]

The environmental hazard potential is generally associated with substances exhibiting persistence, bioaccumulation and toxicity in the environment, so called PBT substances, and having long-range transport potential to reach remote locations. Most modelling studies in the literature have focused on such considerations.[5,11-16] Little attention has been given to highly polar substances that are mobile in the aquatic environment. If such mobile substances are also persistent, they could widely distribute in surface and groundwater (including raw waters used for drinking water production) and therewith present a hazard through threatening the quality of our drinking water resources, as well as pristine freshwater ecosystems. We denote such substances persistent and mobile organic chemicals (PMOCs).[21] PMOCs that are additionally toxic are referred to as PMT (persistent, mobile, and toxic) substances.[22] PMT substances have recently gained the interest of authorities, and there are activities attempting to identify them for potential regulatory measures.[23-25]

In this study we combine two goals. The first is to develop a qualitative emission scoring and ranking system using REACH registration data exclusively, which may serve as a semi-consistent basis for comparing chemicals, and thereby partially addressing the aforementioned concern by Breivik et al.[17] of fragmented input data. The second goal is to apply this system to substances registered under REACH that are PMOCs or PMOC precursors. The purpose of combining these goals is that there is a need to identify PMOCs that may be in the aquatic environment, but are not being monitored. Currently, the research community knows little

about the presence of PMOCs from monitoring studies, due to their intrinsic property to be extremely mobile in water, which causes them to be very challenging to analyze.[21] Only very recently chemical analytical methods specifically targeting at PMOCs were developed[26,27] and a modeling study that identified PMOCs among the REACH registered substances was performed.[28] This modeling study by Arp et al.[28] resulted in a list of more than 2000 substances on the EU market that are suspected to be PMOCs or to hydrolyze to form PMOCs and thus have the potential to be ubiquitous environmental water contaminants. However, in order for a PMOC to be environmentally relevant, it also needs to be released.[23] To address this, the present study expands off of this previous modeling study by Arp et al.,[28] by developing an emission scoring system (E-score) based on information retrieved from dossiers of the substances registered under REACH.

Our approach is distinctively different from published studies[5,12,16-19] in several respects: i) We did not attempt to quantify emissions nor to predict environmental concentrations, but to prioritize (rank) the target substances relatively to each other with respect to their emission potential; ii) we started from a list of substances that were modeled to be PMOCs in groundwater (or PMOC precursors); iii) we included environmental transformation in our study by also estimating the emission potential of substances that were modeled to hydrolyze to PMOCs; iv) we had access to the confidential dossiers from the REACH registration process, giving us accurate figures of marketed volumes.

## 2 Material and methods

### *2.1 PMOC target substances*

As the starting list of substances to be evaluated with respect to their environmental emission potential we used the list of suspected PMOCs and PMOC precursors derived from the

substances registered under REACH (https://echa.europa.eu/information-on-chemicals/registered-substances; as of December 2014) and presented by Arp et al.[28] This list consists of a total of 2167 unique substance identities (including organic and pseudo-organic substances), whereof 1811 have been modeled to be persistent and mobile in the aquatic environment (PMOC score of 4 to 5 in Arp et al.[28]) and 356 have been modeled to be PMOC precursors (i.e. to have the potential to be hydrolyzed to PMOCs with a PMOC score of 4 to 5). These substances typically had a high persistency (>40 days half-life in groundwater, considering biodegradation and hydrolysis), low log $K_{oc}$ (mostly <3; for neutral chemicals), low log $D_{oc}$ (mostly <3; for ionizable and ionic chemicals over a pH range of 4 to 10) and high water solubility (mostly >50 mg/L over a pH range of 4 to 10). EC inventory numbers and CAS numbers were used as identifiers for the unique substances.

*2.2 Environmental emission score (E-score)*

The environmental emission score (E-score) of a substance, i.e. the likelihood of the substance to be emitted into the environment, was calculated for PMOCs and PMOC precursors using the equation

$$\text{E-score} = \log(\text{tonnage} + 1.1) \times \Sigma\text{UCs} \qquad \text{eq. 1}$$

where 'tonnage' is the annual tonnage of the substance placed on the EU market (in t/yr, but for the calculation is considered unitless; see subsection 2.3 below) and ΣUCs is the sum of scores given to the substance for the 7 individual use characteristics (UCs; see subsection 2.4 below). The E-score is thus a unitless figure that allows ranking the qualitative emission potential of the substances relatively to each other, but does not yield quantitative data on the magnitude of estimated emissions.

*2.3 Tonnage*

The information on total tonnage was taken from one of the three databases from ECHA described in Table 1. These data bases were, in order of priority, database A – an aggregated query in early 2015 for all REACH registrations; database B – a similar (but registration-specific) query from May 2014, and database C – the publically available REACH information (as of December 2014). A more detailed description of the databases is given in the Supplementary Material. Exact figures of tonnages placed on the EU market from database A were preferentially used. These figures are expected to reflect potential emissions much better than production volumes or tonnage bands from public databases (e.g. database C). However, the E-score does not take into account the specific tonnage for an individual use, nor any technical or organizational measures to prevent or reduce releases to the environment. Such specific information could not be retrieved automatically from the databases and was thus not feasible to include in a study on thousands of substances. In case no information about tonnage was available from databases A or B, or if the tonnage was given as 0 t, then the upper end of the tonnage band given in database C was used as a worst-case scenario. The logarithm of the tonnage was chosen in the E-score calculation in order to leverage the tonnage data relative to the scoring system used for the ΣUCs, with the range in 'log (tonnage + 1.1)' being from approx. 0.05 to 8.5. A factor of 1.1 was added to the tonnage before calculating the logarithm to avoid negative results for substances with a marketed tonnage <1 t/yr.

**Table 1.** Databases used with availability of data relevant for the present study.

| | **Database A**[a] | **Database B**[a] | **Database C**[b] |
|---|---|---|---|
| Type and source of database | Database query from early 2015 by ECHA for all registrations (aggregated) | Database query from May 2014 by ECHA for all registrations (registration-specific) | Public REACH database on ECHA's website accessed in December 2014 |
| Substance name | +[c] | + | + |
| EC number | + | + | + |

| | | | |
|---|---|---|---|
| CAS number | -[c] | + | + |
| Tonnage placed on the EU market | Exact tonnage | Estimation of maximum tonnage | Tonnage band |
| *Use characteristics* | | | |
| High release to environment | + | - | + |
| Wide dispersive use | + | - | + |
| Intermediate use | + | - | + |
| Closed system use | + | - | + |
| Professional use | + | - | + |
| Consumer use | + | - | + |
| Substance in article | + | - | + |

[a] Databases A and B were compiled by ECHA from confidential business information in the REACH registration dossiers. The databases are available to Member State Competent Authorities for specific regulatory purposes. Access to the databases and the registration dossiers was available through cooperation with the German Environment Agency (UBA). In the context of the present project the data in databases A and B were accessed at UBA's premises and provided by UBA for the 2167 suspected PMOCs and PMOC precursors.

[b] https://echa.europa.eu/de/information-on-chemicals/registered-substances

[c] A '+' means that data was available in the respective database, a '–' means that no data was available

### *2.4 Use characteristics*

The seven UCs considered in the present study are listed in Table 1. They outline specific information on operational conditions during uses of the substances related to the likelihood of emissions on a generic level. Each characteristic was individually evaluated for each substance in order to come to the decision if the substance possesses this characteristic (TRUE) or not (FALSE). This was done according to modified criteria (see subsections 2.4.1-2.4.7 below) initially defined by ECHA. The initial criteria by ECHA for the TRUE/FALSE decisions are defined in database B and are based on the generic use descriptors in the REACH registrations. The aggregated information in database A regarding these generic use descriptors was used in the TRUE/FALSE decisions for the UCs (if not stated otherwise below). The scores given to the substances for each of the UCs are summarized in Table 2. The scores (numbers) were given based on the authors' judgement of how strongly a certain UC is expected to correlate with the potential for emissions, due to the absence of consistent, actual, empirical emission rates reported in REACH (or elsewhere). The UC 'high release to

environment' was given the highest priority (highest numerical score), as this amounts directly to environmental emissions. Further, also the UCs 'wide dispersive use', 'intermediate use', and 'closed system use' are directly related to emissions (or the presumed absence of emissions in the latter two cases) and were given second priority, while the remaining characteristics only imply that environmental emissions could (but not necessarily will) occur and were thus given the lowest scores. For 'substance in article' the data availability on the type of article and on the technical function (related to the potential for releases) of the substance is poor. In addition, emissions from articles with treated surfaces, e.g. coatings on construction materials, are already covered by previous life cycle steps of the chemical before application on the article. Therefore a low score of 0.5 was given to avoid false positives impacting the overall assessment. If a UC outcome of either TRUE or FALSE could not exclude emissions, a score greater than zero was assigned to both cases. E.g., a FALSE classification for 'high release to environment' does not mean complete absence of releases according to the ECHA criteria. Likewise, the score of 1 for 'closed system use' TRUE takes into account releases during handling of the substance (e.g. filling/emptying of the system) and possible releases in waste management. In any case, the model is quite insensitive towards changes in the magnitude of these scores, since they are all equally used in a simple summation (ΣUCs). It is emphasized that the model output is not quantitative, but is only a relative ranking of the substances with respect to their emission potential as characterized by the selected UCs. The sum of the scores of all seven UCs can range from 6 to 21. The criteria for evaluation of each UC are described below and two examples of scoring and ranking are given in the Supplementary Material.

**Table 2.** Scores given based on the TRUE/FALSE decision for each of the use characteristics.

| Use characteristic | Score for TRUE | Score for FALSE |
|---|---|---|
| High release to environment | 7 | 3 |
| Wide dispersive use | 4 | 1 |
| Intermediate use | 0 | 3 |
| Closed system use | 1 | 3 |
| Professional use | 1.5 | 0.5 |
| Consumer use | 2 | 0.5 |
| Substance in article | 0.5 | 0 |

**2.4.1 High release to environment.** Following the criteria defined by ECHA a high release is expected if ≥10% of the initial amount of a substance in a process or use is emitted to at least one environmental compartment (air, water, soil). This endpoint was thus evaluated as TRUE if at least one of the Environmental Release Categories (ERC) 2, 5, 8a, 8c, 8d, 8f, 10b, 11b, or 12b, as defined in REACH, was assigned to an individual registration dossier of the substance in the chapter for use description. See table R.16-7 in the respective guidance document[29] for a detailed description of release rates for these ERC. Otherwise the evaluated decision was FALSE.

**2.4.2 Wide dispersive use.** This characteristic was evaluated as TRUE if at least one of the following criteria applied: 'Number of consumer uses (upper bound) >0', 'number of professional uses (upper bound) >0', at least one of the ERC 8-11 was assigned to the substance in an individual registration dossier, or at least one of the process categories (PROC) 10, 11, 13, 15, 17, 18, or 19 was assigned to the substance in an individual registration dossier. See table R.12-11 in the respective guidance document[30] for a detailed description of the PROC. If none of the above criteria applied, then the evaluated decision was FALSE.

**2.4.3 Intermediate use.** The information for a decision on intermediate use (intermediate means a substance that is manufactured for and consumed in or used for chemical processing in order to be transformed into another substance[2]) was taken from database C. It was evaluated as TRUE if there were exclusively registrations of the type 'intermediate' recorded; otherwise it was evaluated as FALSE.

**2.4.4. Closed system use.** The information for a decision on closed system use was taken from database C. It was evaluated as TRUE if 'all identified uses take place in closed system' was answered with 'yes'; otherwise it was evaluated as FALSE.

**2.4.5 Professional use.** This characteristic was evaluated as TRUE, if 'number of professional uses (upper bound) >0' or if this information was ambiguous or lacking. Professional use was thus only evaluated as FALSE if 'number of professional uses (upper bound) = 0'.

**2.4.6 Consumer use.** This characteristic was evaluated as TRUE, if 'number of consumer uses (upper bound) >0' or if this information was ambiguous or lacking. Consumer use was thus only evaluated as FALSE if 'number of consumer uses (upper bound) = 0'.

**2.4.7 Substance in article.** This characteristic was evaluated as TRUE, if 'number of article categories >0' or if this information was ambiguous or lacking. Substance in article was thus only evaluated as FALSE if there was no article category given.

*2.5 Significant data gaps or ambiguous data*

For some of the PMOCs and PMOC precursors the data that were needed to evaluate the UCs were incomplete, or the information on tonnage and/or UCs was ambiguous. For small data gaps or ambiguity, the worst-case scenario was assumed for the respective UC. However, in some cases significant data gaps or contradictory data existed, which hampered a sound calculation of the E-score. This was the case for the following combinations of data gaps and/or ambiguous data:

a) Tonnage = 0 t and 'intermediate use' FALSE (contradictory data).

b) Tonnage = 0 t, 'intermediate use' not specified, 'closed system use' not specified, and 'substance in article' not specified.

c) Tonnage given, 'high release to environment' not specified, 'intermediate use' not specified, 'closed system use' not specified, and 'substance in article' not specified.

Substances with significant data gaps as specified in a)-c) were not given a numerical E-score based on the information from databases A-C. Tonnages and UCs for these substances were instead evaluated case by case using the information available on ECHA's public website http://echa.europa.eu/de/information-on-chemicals/registered-substances (accessed between July 2015 and December 2016).

*2.6 Evaluation of the E-score model and sensitivity analysis*

The E-score model was evaluated using several approaches, as described in subsection 3.3 below. Correlation analyses based on the Pearson product-moment was conducted between log(tonnage + 1.1) vs. E-score ranking, ΣUCs vs. E-score ranking, and log(tonnage + 1.1) vs. ΣUCs, using Origin Pro 2016. Tonnage proved to be the most influential parameter in the model (see section 3.3); therefore a sensitivity analysis was performed, investigating how the results would change if only tonnage was considered in the E-score, but not UCs. It was also tested how the results would change if only the maximum single UC score was used in eq. 1 instead of ΣUCs. Another evaluation approach was based on a literature search using the Web of Science search engine (www.webofknowledge.com/). For this purpose, the substances with estimated emissions (1110 in total, see section 3.1 'class 2 substances') were grouped in 11 E-score groups with 101 substances in each group. Group I contained the 101 substances with the highest calculated E-score and group XI with the lowest. Roughly every 9$^{th}$ substance in each group was randomly picked (11 per group, resulting in a total of 121 substances) and

294 searched for using the following keywords in the search category 'topic': 'substance name'
295 AND (*environment* OR *water* OR *soil* OR *effluent*). The same search was also done
296 for 30 (from a total of 1054) randomly selected substances with no predicted emissions (group
297 XII, see section 3.1 'class 3 substances'). The median of the number of 'hits' was calculated
298 for the 11 substances per group (30 substances for group XII) and correlated with the E-score
299 ranking of the groups (i.e. the Roman numeral group numbering). The grouping and
300 calculation of medians were done to smoothen the results of the correlation. The assumption
301 in this evaluation was that the more of a substance is emitted into (and consequently for
302 persistent substances occurring in) the environment, the more reports exist in the scientific
303 literature containing the name of the substance together with any of the searched keywords,
304 i.e. the more 'hits' one would get when performing such a search.

305

306 **3 Results and discussion**

307 *3.1 Prioritized substances*

308 Applying our E-score calculation approach (eq. 1) to the 2167 modeled PMOCs (1811
309 substances) and PMOC precursors (356 substances) resulted in three classes of substances
310 based on emission potential, as follows:

311 Class 1 - substances for which an E-score could not be calculated due to incomplete
312 information. Initially, this class of substances with significant data gaps or ambiguous data
313 comprised a total of 29 substances. After case by case evaluation using the information
314 available on ECHA's public website, 14 and 12 of these substances could be classified into
315 class 2 and class 3, respectively. The 3 substances remaining in class 1 were all REACH
316 registered PMOCs (not precursors).

Class 2 - substances with indicators of environmental emissions: A total of 1110 substances (including the 14 cases from class 1) had tonnage and UCs indicating emissions (i.e. not fulfilling both a tonnage of 0 t and 'intermediate use' TRUE). The calculated E-score for class 2 substances is assumed to be positively correlated with the likelihood of the substance being emitted into the environment.

Class 3 - substances with indicators of negligible environmental emissions: A total of 1054 substances (including the 12 cases from class 1) had indicators of no (or minor) environmental emissions (i.e. a tonnage of 0 t and an 'intermediate use' TRUE). Class 3 substances were not considered further in the present study.

The final distribution of PMOCs and PMOC precursors between the three classes is shown in Figure 1.

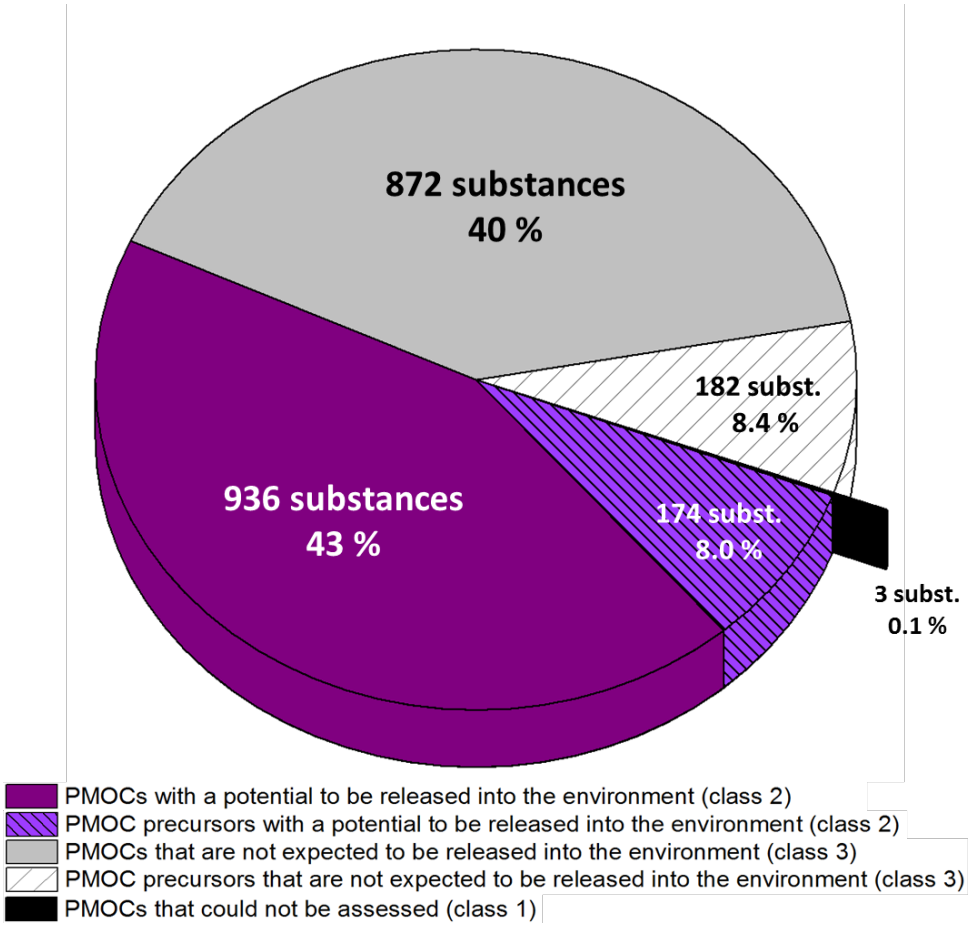

**Figure 1.** Distribution of PMOCs and PMOC precursors between the E-score classes.

**3.1.1 PMOCs registered under REACH.** Of the 1110 class 2 substances 936 were REACH registered substances predicted to be PMOCs (Figure 1). These 936 substances are ranked in Table S1 in the Supplementary Material in order of their calculated E-score, with rank 1 (carbonate/carbonic acid) representing the highest E-score. The values of the E-scores themselves cannot be disclosed, as they might allow back-calculation of confidential data from the REACH registration dossiers (especially tonnages) used as input data. The 3 remaining PMOCs from class 1 (no E-score calculable, see Figure 1) are listed at the end of Table S1. Some individual cases of substances are discussed in section 3.4.

**3.1.2 Precursors of PMOC hydrolysis products.** The remaining 174 of the 1110 class 2 substances are REACH registered substances that were predicted to hydrolyze under environmental conditions to form PMOCs[28] (Figure 1). These 174 precursors are listed in Table S2 in the Supplementary Material in order of their calculated E-score (rank 1 represents the highest E-score). The predicted hydrolysis products that were modeled to be PMOCs are shown in Table S2 with their Simplified Molecular Input Line Entry Specification (SMILES) codes. As can be seen from Table S2, one precursor substance can hydrolyze to form several PMOC hydrolysis products. Vice versa, one PMOC hydrolysis product can also be formed from different precursor substances. Some of the PMOC hydrolysis products are also the same structure as other REACH registered substances, including some of the PMOCs already considered. This has the implication that a subset of the REACH registered PMOC substances can be released directly, or as transformation products of other substances.[28] Selected highly ranked precursors are briefly discussed in subsection 3.4.

*3.2 Uncertainties in the prioritization*

354 The aim of the study was to prioritize PMOCs with regard to their environmental emission
355 potential. The list of suspected PMOCs and PMOC precursors published by Arp et al.[28] was
356 used as a starting point. As discussed in detail by Arp and co-workers, the persistency and
357 mobility modeling as well as the modeling of hydrolysis under environmental conditions are
358 associated with uncertainties, which are thus transferred into our study. The E-score model
359 itself also contains uncertainties. As mentioned in subsection 2.3 above, specific tonnages for
360 individual uses or 'end-of-pipe' measures to reduce emissions to the environment from the
361 individual use processes were not taken into account in the E-score calculation. The applied
362 UCs do not contain sufficient empirical information to predict actual substance emissions.
363 The TRUE/FALSE decisions with regard to the different UCs were further based on data
364 submitted by registrants within the REACH registration process, and these data were not
365 independently checked. A recent compliance check by order of the UBA demonstrated that
366 only 4 to 45% of the investigated dossiers were compliant with the requests from the REACH
367 regulation (information requirements referred to in article 10; Annexes VI-XI) with respect to
368 information provided for five different environmental endpoints. A large number of dossiers
369 (43 to 82%) were generally classified as 'complex', i.e. a classification in 'compliant' or 'non-
370 compliant' was not possible due to poor documentation.[31] Taken together, all these
371 uncertainties will undoubtedly lead to both false negatives as well as false positives in our
372 estimation of the likelihood of a substance to be emitted in significant amounts. Thus, some of
373 the highly ranked substances in Table S1 may not necessarily be present in environmental
374 water samples; whereas, REACH registered substances missing from Table S1 may be
375 currently contaminating water resources. The prioritization should be seen as qualitative
376 hypotheses of substances that could potentially threaten raw water bodies, but this has to be
377 confirmed (or disproved) case by case. On the other hand, the tonnage information from the
378 confidential sections of the registration dossiers we used as input data for our E-score model

379 is certainly more accurate than publically available tonnage band data. Therefore, we expect
380 our E-score estimation model to perform at least as well as or better than models earlier
381 published in literature.[5,16-19] Assuming that reporting in REACH will become more accurate
382 and more comprehensive in future also with respect to UCs, it would be of interest to repeat
383 this E-scoring at a later time.

384 *3.3 Evaluation of the E-score model and sensitivity analysis*

385 The relative sensitivity of the model output (the E-score ranking) towards the two factors in
386 the model equation (eq. 1) was tested by correlating the E-score ranking with both factors
387 individually. A strong positive correlation ($r = 0.92$) was found between log(tonnage +1.1)
388 and ranking and a weaker positive correlation ($r = 0.55$) between ΣUCs and ranking. This
389 shows that in our model both factors significantly contributed to the output, whereby the
390 marketed tonnage had the strongest influence on the final rank of a substance. However, using
391 only tonnage as ranking criterion would result in 21% of the substances changing their
392 position in Table S1 or S2 with more than 100 ranks. This demonstrates that ΣUCs is also an
393 important parameter in the model. ΣUCs and log(tonnage +1.1) correlated only very weakly
394 with each other ($r = 0.28$), confirming that the TRUE/FALSE decision criteria for the UCs
395 were not (markedly) influenced by the tonnage of the substance, i.e. that the two factors in the
396 E-score calculation were not strongly co-dependent of each other. If only the maximum single
397 UC score was used in the E-score calculation instead of ΣUCs, less than 10% of the
398 substances would change their ranking position with more than 100 ranks. This further
399 corroborates that the model is relatively insensitive towards the values of the scores for the
400 different UCs.

401 The results of the E-scoring were further evaluated using the Web of Science search approach
402 described in subsection 2.6. The obtained histogram between the ranges of E-score ranks and
403 the Web of Science 'hits' is shown in Figure 2. The very strong positive relationship in Figure

2 between increasing E-score range and 'hits' suggests that our model in general identifies substances of interest to the environmental and chemical community, and is fit for the purpose of qualitatively ranking emissions. This is further confirmed by a glance at the top ranked PMOCs in Table S1. Many of these are common salts or solvents; though these may not be the most interesting substances for an environmental chemist looking for emerging contaminants, they are expected to qualify as PMOCs with a high emission potential.

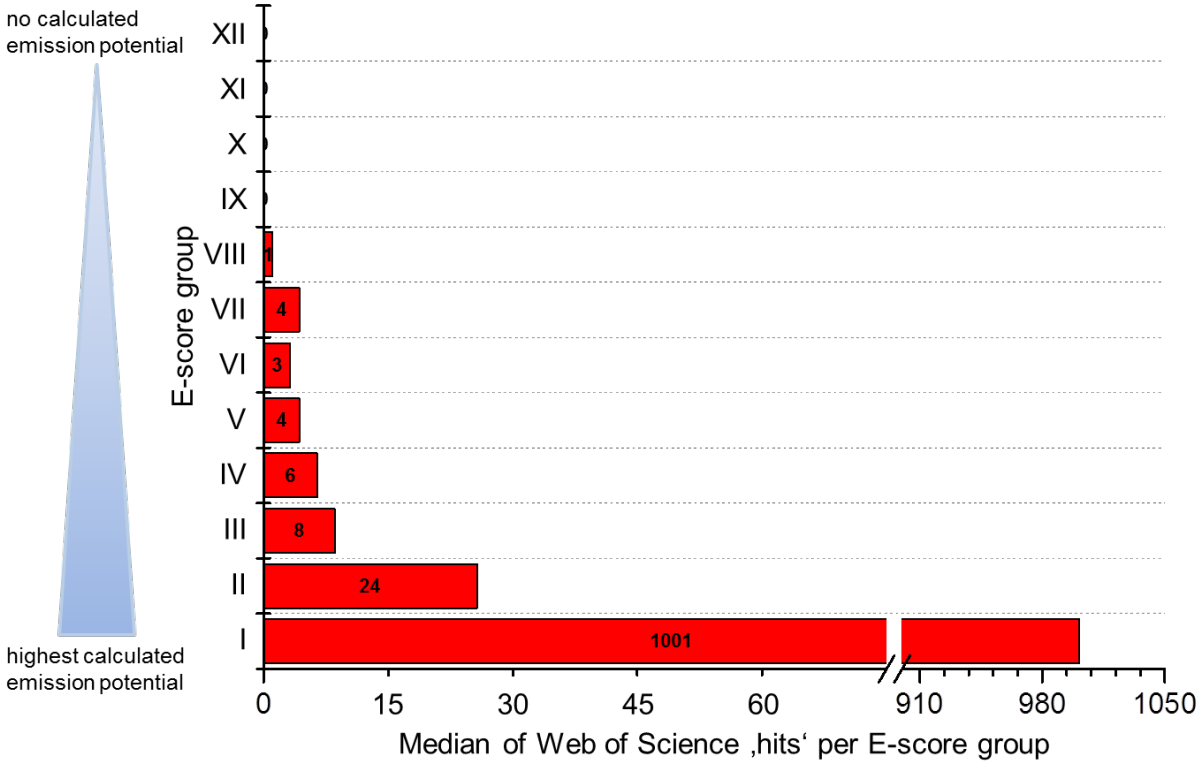


**Figure 2.** Evaluation of the E-score results using a Web of Science literature search. The x-axis shows the median of the 'hits' of the 11 investigated substances per group (30 for group XII). The y-axis shows the groups consisting of 101 substances per group (1054 for group XII). Group I is the group of substances with the highest E-scores and group XI with the lowest E-scores (class 2). Group XII are the class 3 substances without predicted emissions.

*3.4 Identifying PMOCs of concern*

Amongst the highly ranked PMOCs (Table S1) there are organic water contaminants known from the literature, such as melamine (rank 8, reported in river water, groundwater, and tap water[32-34]), bisphenol S (rank 132, reported in river water[35]), sulphanilic acid (rank 159, reported in groundwater[36,37]), acesulfame (rank 277, reported in wastewater, surface water, groundwater, and tap water[38]), dapsone (rank 324[39,40]), and saccharine (rank 498, reported in wastewater, surface water, and groundwater[38,41]). It is likely that many less well-known or hitherto unreported PMOCs that are problematic to raw water are also within these highly ranked substances. To this end, Table S1 has already been used successfully by Montes and co-workers[27] in a first chemical analytical approach to screen environmental water samples for novel and emerging PMOCs. PMOCs detected in this screening study included toluene-4-sulfonic acid (rank 50), 1,3-di-*o*-tolylguanidine (rank 427), and trifluoromethane sulfonic acid (rank 429), which has recently been detected in raw water and drinking water sources for the first time.[26,27] These results, as well as future observations of other PMOCs in this prioritized list in raw water sources, represent the ultimate evaluation of our modelling approach, and its utility. The list of prioritized PMOCs presented in Table S1 can thus serve as a starting point for suspect screening of further, yet unknown raw water contaminants. This is particularly the case for the highly ranked substances that did not occur commonly in the literature evaluation exercise, as the reason for this gap in the literature may be a general lack of available analytical techniques for these substances, and therefore a lack of screening and monitoring data.[21] Further considerations for prioritization of the highly ranked PMOCs presented here would be information on exposure, toxicity[24,25] and risk and more detailed information on areas of application, such as in household products, which would increase the risk of wide-spread emissions and water contamination.

The top-ranking precursors that were predicted to hydrolyze into PMOCs (Table S2) include several aromatic isocyanates (rank 1-3), the brominated flame retardant

hexabromocyclododecane (rank 4), as well as large molecules that resulted in a multitude of potential PMOC hydrolysis products (such as propoxylated trimethylolpropane, rank 7). However, it needs to be acknowledged that the yields and accuracy of the predicted hydrolysis are uncertain, as these were all based on QSARs;[28] therefore, the likelihood of detecting these transformation products in the environment is less than for PMOCs from Table S1. The high ranking of hexabromocyclododecane hydrolysis products is indicative of this uncertainty, as this compound is not known to readily hydrolyze under environmentally relevant conditions. Nevertheless, this list can be used as a starting point to prioritize which substances should be investigated for their ability to hydrolyze or transform into potentially problematic PMOCs.

## Acknowledgments

We gratefully acknowledge the European Union Joint Programming Initiative 'Water Challenges for a Changing World' (Water JPI) with financial support by the 'Bundesministerium für Bildung und Forschung' (Germany, project no. 02WU1347A) and 'Forskningsrådet' (Norway, project no. 241358/E50).

## References

(1) Chemical Abstracts Service (2008) *Statistical Summary 1907-2007*, February 2008, CAS2475-0208, Chemical Abstracts Service, Columbus, USA, available from www.shinwon.co.kr/cas/ASSETS/casstats.pdf

(2) European Parliament (2006) *Regulation (EC) 1907/2006 of the European parliament and of the council of 18 December 2006 concerning the registration, evaluation, authorisation and restriction of chemicals (REACH), establishing a European chemicals agency, amending directive 1999/45/ec and repealing council regulation (eec) 793/93 and commission regulation (ec) 1488/94 as well as council directive 76/769/eec and commission directives 91/155/eec, 93/67/eec, 93/105/ec and 2000/21/ec*, Official Journal of the European Union, 30.12.2006, L 396/1-849.

(3) Bu, Q.; Wang, D.; Wang Z. (2013) *Review of screening systems for prioritizing chemical substances*, Crit. Rev. Env. Sci. Tec., *43*, 1011-1041.

(4) Muir, D.C.G.; Howard, P.H. (2006) *Are there other persistent organic pollutants? A challenge for environmental chemists*, Environ. Sci. Technol., *40*, 7157-7166.

(5) Arnot, J.A.; Brown, T.N.; Wania, F.; Breivik, K.; McLachlan, M.S. (2012) *Prioritizing chemicals and data requirements for screening-level exposure and risk assessment*, Environ. Health Perspect., *120*, 1565-1570.

(6) Cohen Hubal, E.A.; Richard, A.; Aylward, L.; Edwards, S.; Gallagher, J.; Goldsmith, M.R.; Isukapalli, S.; Tornero-Velez, R.; Weber, E.; Kavlock, R. (2010) *Advancing exposure characterization for chemical evaluation and risk assessment*, J. Toxicol. Environ. Health, Part B, *13*, 299-313.

(7) Egeghy, P.P.; Vallero, D.A.; Cohen Hubal, E.A. (2011) *Exposure-based prioritization of chemicals for risk assessment*, Environ. Sci. Pol., *14*, 950-964.

(8) Guillén, D.; Ginebreda, A.; Farré, M.; Darbra, R.M.; Petrovic, M.; Gros, M.; Barceló, D. (2012) *Prioritization of chemicals in the aquatic environment based on risk assessment: Analytical, modeling and regulatory perspective*, Sci. Tot. Environ., *440*, 236-252.

(9) Hansen, B.G.; van Haelst, A.G.; van Leeuwen, K.; van der Zandt, P. (1999) *Priority setting for existing chemicals: European Union risk ranking method*, Environ. Toxicol. Chem., *18*, 772-779.

(10) Jayjock, M.A.; Chaisson, C.F.; Franklin, C.A.; Arnold, S.; Price, P.S. (2009) *Using publicly available information to create exposure and risk-based ranking of chemicals used in the workplace and consumer products*, JESEE, *19*, 515-524.

(11) Mitchell, J.; Arnot, J.A.; Jolliet, O.; Georgopoulos, P.G.; Isukapalli, S.; Dasgupta, S.; Pandian, M.; Wambaugh, J.; Egeghy, P.; Cohen Hubal, E.A.; Vallero, D.A. (2013) *Comparison of modeling approaches to prioritize chemicals based on estimates of exposure and exposure potential*, Sci. Total Environ., *458-460*, 555–567.

(12) Bitsch, A.; Bohlen, M.L.; Escher, S.; Licht, O.; Oltmanns, J.; Schneider, K.; Wibbertmann, A. (2016) *Final report: Testing a procedure for the identification of emerging chemical risks in the food chain*, External Scientific Report, OC/EFSA/SCER/2014/03, EFSA Supporting publication 2016: EN-1050, http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2016.EN-1050/epdf.

(13) Brown, T.N.; Wania, F. (2008) *Screening chemicals for the potential to be persistent organic pollutants: A case study of Arctic contaminants*, Environ. Sci. Technol. *42*, 5202-5209.

(14) Howard, P.H.; Muir, D.C.G. (2010) *Identifying new persistent and bioaccumulative organics among chemicals in commerce*, Environ. Sci. Technol., *44*, 2277-2285.

(15) Zarfl, C.; Matthies, M. (2013) *PBT borderline chemicals under REACH*, Environ. Sci. Europe, *25*, 1-11.

(16) McLachlan, M.S.; Kierkegaard, A.; Radke, M., Sobek, A.; Malmvärn, A.; Alsberg, T.; Arnot, J.A.; Brown, T.N.; Wania, F.; Breivik, K.; Xu, S. (2014) *Using Model-Based Screening to Help Discover Unknown Environmental Contaminants*, Environ. Sci. Technol., *48*, 7264–7271.

(17) Breivik, K.; Arnot, J.A.; Brown, T.N.; McLachlan, M.S.; Wania, F. (2012) *Screening organic chemicals in commerce for emissions in the context of environmental and human exposure*, J. Environ. Monit., *14*, 2028-2037.

(18) Fauser, P.; Thomsen, M.; Pistocchi, A.; Sanderson, H. (2010) *Using multiple regression in estimating (semi) VOC emissions and concentrations at the European scale*, Atmos. Pollut. Res., *1*, 132-140.

(19) Fischer, S.; Almkvist, Å.; Karlsson, E.; Åkerblom, M. (2006) *Preparation of a Product Register based Exposure Index*, Swedish Chemicals Agency, Sweden.

(20) Undeman, E.; Fischer, S.; McLachlan, M.S. (2011) *Evaluation of a novel high throughput screening tool for relative emissions of industrial chemicals used in chemical products*, Chemosphere, *82*, 996-1001.

(21) Reemtsma, T.; Berger, U.; Arp, H.P.H.; Gallard, H.; Knepper, T.P.; Neumann, M. Quintana, J.B.; de Voogt P. (2016) *Mind the gap: Persistent and mobile organic compounds – water contaminants that slip through*, Environ. Sci. Technol., *50*, 10308-10315.

(22) Neumann, M. (2017) *Vorschlag für Kriterien und ein Bewertungskonzept zur Identifizierung von persistenten, mobile und toxischen (PMT-) Stoffen zum Schutz des Rohwassers zur Trinkwassergewinnung unter der EU-Verordnung REACH*, Zbl. Geol. Paläont. Teil I, *Jg. 2017*, Heft 1, 91-101.

(23) Kalberlah, F.; Oltmanns, J.; Schwarz, M.; Baumeister, J.; Striffler, A. (2014) *Guidance for the precautionary protection of raw water destined for drinking water extraction from contaminants regulated under REACH*. Environmental Research of the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety. Project Report (UFOPLAN) FKZ 371265416. Umweltbundesamt, Dessau, Germany.

(24) Berger, U.; Ost, N.; Sättler, D.; Schliebner, I., Kühne, R.; Schüürmann, Neumann, M.; G.; Reemtsma, T. (2017) *Assessment of persistence, mobility and toxicity (PMT) of 167 REACH registered substances*, Environmental Research of the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety. Project No. 74925, Umweltbundesamt, Dessau, Germany.

(25) Neumann, M.; Schliebner, I. (2017) *Protecting the sources of our drinking water – A revised proposal for implementing criteria and an assessment procedure to identify Persistent, Mobile and Toxic (PMT) and very Persistent, very Mobile (vPvM) substances registered under REACH*, German Environment Agency, Dessau-Rosslau, 16 pages, ISSN 2363-8273.

(26) Zahn, D.; Frömel, T.; Knepper, T.P. (2016) *Halogenated methanesulfonic acids: A new class of organic micropollutants in the water cycle*, Water Res., *101*, 292-299.

(27) Montes, R.; Aguirre, J.; Vidal, X.; Rodil, R.; Cela, R.; Quintana, J.B. (2017) *Screening for polar chemicals in water by trifunctional mixed-mode liquid chromatography-high resolution mass spectrometry*, Environ. Sci. Technol., *51*, 6250-6259.

(28) Arp, H.P.H.; Brown, T.N.; Berger, U.; Hale, S.E. (2017) *Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility*, Environ. Sci. Process Impacts, *19*, 939-955.

(29) ECHA (2016) *Guidance on information requirements and chemical safety assessment, Chapter R.16: Environmental exposure estimation*, Version 3.0, February 2016, ECHA-16-G-03-EN, European Chemicals Agency, Helsinki, Finland.

(30) ECHA (2015) *Guidance on information requirements and chemical safety assessment, Chapter R.12: Use descriptor system*, Version 3.0, December 2015, ECHA-15-G-11-EN, European Chemicals Agency, Helsinki, Finland.

(31) Springer, A.; Herrmann, H.; Sittner, D.; Herbst, U.; Schulte, A. (2015) *REACH Compliance: Data Availability of REACH Registrations, Part 1: Screening of chemicals > 1000 tpa*, Umweltbundesamt (German Environment Agency), Texte 43/2015, FKZ 3714 67 4200.

(32) Qin, Y.; Lv, X.; Li, J.; Qi, G.; Diao, Q.; Liu, G.; Xue, M.; Wang, J.; Tong, J.; Zhang, L.; Zhang, K. (2010) *Assessment of melamine contamination in crop, soil and water in China and risks of melamine accumulation in animal tissues and products*, Environ. Int., *36*, 446-452.

(33) Stroomberg, G.J.; Neefjes, R.E.M.; Haar van de, G.; Bannink, A.; Zwamborn, C.C. (2016) *Jahresbericht 2015 Der Rhein*, RIWA – Verband der Flusswasserwerke, ISBN/EAN 978-90-6683-163-6.

(34) Winzenbacher, R.; Seitz, W.; Schulz, W. (2015) *Bedeutung der Spurenstoffanalytik für die Wasserversorgung – Was, Wie, Warum?*, LW-Schriftenreihe 2015, *Beitrag 6*, 66-74.

(35) Yang, Y.; Lu, L.; Zhang, J.; Yang, Y.; Wu, Y.; Shao, B. (2014) *Simultanous determination of seven bisphenols in environmental water and solid samples by liquid chromatography-electrospray tandem mass spectrometry*, J. Chrom. A, *1328*, 26-34.

(36) Holm, J.; Bügge, K.; Bierg, P.L.; Christensen, T.H. (1995) *Occurrence and Distribution of Pharmaceutical Organic Compounds in the Groundwater Downgradient of a Landfill*, Environ. Sci. Technol., *29*, 1415-1420.

(37) Jones, O.A.H.; Voulvoulis, N.; Lester, J.N. (2001) *Human pharmaceuticals in the aquatic environment – a review*, Environ. Technol., *12*, 1383-1394.

(38) Buerge, I.J.; Buser, H.-R.; Kahle, M.; Müller, M.D.; Poiger, T. (2009) *Ubiquitous Occurence of the Artificial Sweetener Acesulfame in the Aquatic Envrionment: An Ideal Chemical Marker of Domestic Wastewater in Groundwater*, Environ. Sci. Technol., *43*, 4381-4385.

(39) Derksen, J.G.M.; Rijs, G.B.J.; Jongbloed, R.H.; (2004) *Diffuse pollution of surface water by pharmaceuticals*, Water Sci. Technol., *49*, 213-221.

(40) Sacher, F.; Lange F.T.; Brauch, H.-J.; Blankenhorn, I. (2001) *Pharmaceuticals in groundwaters – analytical methods and results of a monitoring program in Baden-Württemberg*, Germany, J. Chrom. A, *938*, 199-210.

(41) Stempvoort, D.R.; Roy, J.W.; Brown, S.J.; Bickerton, G. (2011) *Artificial sweeteners as potential tracers in groundwater in urban environments*, J. Hydrol., *401*, 126-133.