

# Development and Application of Liquid Chromatographic Retention Time Indices in HRMS-based Suspect and Non-target Screening

Reza Aalizadeh <sup>a</sup>, Nikiforos A. Alygizakis <sup>a,b</sup>, Emma L. Schymanski <sup>c,d</sup>, Martin Krauss <sup>e</sup>, Tobias Schulze <sup>e</sup>, María Ibáñez <sup>f</sup>, Andrew D. McEachran <sup>g</sup>, Alex Chao <sup>g</sup>, Antony J Williams <sup>g</sup>, Pablo Gago-Ferrero <sup>h,i</sup>, Adrian Covaci <sup>j</sup>, Christoph Moschet <sup>k</sup>, Thomas M. Young <sup>k</sup>, Juliane Hollender <sup>d,l</sup>, Jaroslav Slobodnik <sup>b</sup>, Nikolaos S. Thomaidis <sup>a,\*</sup>

<sup>a</sup>Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 157 71 Athens, Greece

<sup>b</sup>Environmental Institute, Okružná 784/42, 97241 Koš, Slovak Republic

<sup>c</sup>Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg

<sup>d</sup>Eawag: Swiss Federal Institute for Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland.

<sup>e</sup>Department Effect-Directed Analysis, Helmholtz-Centre for Environmental Research – UFZ, Leipzig, Germany

<sup>f</sup>Research Institute for Pesticides and Water, University Jaume I, Castellón 12071, Spain

<sup>g</sup>Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Mail Drop, D143-02, 109 T.W. Alexander Dr., Research Triangle Park, NC 27711, USA

<sup>h</sup>Institute of Environmental Assessment and Water Research (IDAEA) Severo Ochoa Excellence Center, Spanish Council for Scientific Research (CSIC), Jordi Girona 18-26, E-08034 Barcelona, Spain

<sup>i</sup>Swedish University of Agricultural Sciences (SLU), Department of Aquatic Sciences and Assessment, P. O. Box 7050, SE-750 07 Uppsala, Sweden

<sup>j</sup>Toxicological Center, University of Antwerp, 2610 Wilrijk, Belgium

<sup>k</sup>Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA

<sup>l</sup>Institute of Biogeochemistry and Pollutant Dynamics, IBP, ETH Zurich, 8092 Zurich, Switzerland

**ABSTRACT:** There is an increasing need for comparable and harmonized retention times ( $t_R$ ) in liquid chromatography among different laboratories, to provide supplementary evidence for the identity of compounds in high resolution mass spectrometry (HRMS) based suspect and non-target screening investigations. In this study, a rigorously tested, flexible and less system-dependent unified retention time index (RTI) approach for LC is presented, based on the calibration of the elution pattern. Two sets of 18 calibrants were selected for each of ESI+ and ESI- based on the maximum overlap with the retention times and chemical similarity indices from a total set of 2123 compounds. The resulting calibration set, with RTI set to range between 1 and 1000, was proposed as the most appropriate RTI system after rigorous evaluation, coordinated by the NORMAN network. The validation of the proposed RTI system was done externally on different instrumentation and LC conditions. The RTI can also be used to check the reproducibility and quality of LC conditions. Two QSRR based models were built based on the developed RTI systems, which assist in the removal of false positive annotations. The applicability domains of the QSRR models allowed completing the identification process with higher confidence for substances within the domain, while indicating those substances for which results should be treated with caution. The proposed RTI system was used to improve confidence in suspect and non-target screening and increase the comparability between laboratories as demonstrated for two examples. All RTI related calculations can be performed online at <http://rti.chem.uoa.gr/>.

## INTRODUCTION

Liquid chromatography (LC) coupled to high resolution mass spectrometry (HRMS) has been widely used to analyze complex samples containing thousands of substances with widely varying physicochemical properties. LC-HRMS has a high separation power and increased sensitivity for the measurement of compounds at low concentrations<sup>1</sup> over alternative methods. One of the major bottlenecks in non-target LC-HRMS is to identify the true positive compounds among the pools of plausible candidates with the same molecular formula.<sup>2</sup> This frequently happens in suspect and non-target screening workflows, where the task is to identify the unknowns with a certain level of confidence.<sup>3</sup> Therefore, confirming a structure becomes a task of eliminating false positive candidates based on the available experimental information such as MS fragmentation pattern, retention time ( $t_R$ ), ionization behavior and other evidence.<sup>4-6</sup> Prediction of  $t_R$  for candidate substances can significantly decrease the number of false positive candidates.<sup>4,7-14</sup> Nevertheless, it is often neglected due to the difficulties of accurately predicting and mapping it between different LC conditions.

Identification workflows in gas chromatography (GC) routinely include chromatographic elution information. The Kovats' retention index (KRI)<sup>15</sup> play a major role during identification, enabling cross checking KRI of a tentatively identified compound with a library of KRIs. This and other RIs (e.g. linear<sup>16</sup> and Lee RI<sup>17</sup>) make it possible to compare  $t_R$  values for most GC systems. Despite KRI being long established, a reliable retention index does not exist for LC-based identification workflows, for many reasons including larger differences of LC methods among laboratories. There is also lack of experimental information about the elution of various compounds covering a large chemical space under different LC conditions. There is also not yet sufficient agreement on potential calibrants for LC unlike, for instance, the alkane series used for the KRI.

Since the variety of chemicals amenable to LC methods (and thus the functional groups and properties involved) is large, simple calibrants that would be analogous to the linear alkanes (KRI) or PAHs (Lee RI) used for GC are not applicable. According to Stanstrup and coworkers,<sup>18</sup> most compounds conserve their elution orders for similar LC systems (reversed phase, using a C<sub>18</sub> column)

while this could be extremely variable for different types of chromatographic columns (e.g., hydrophilic interaction liquid chromatography; HILIC). However, PredRet approach<sup>18</sup> introduced the strategy of directly mapping  $t_R$  in various LC conditions, to share  $t_R$  information and include it during identification.<sup>18</sup> This approach, however, requires many compounds with measured  $t_R$  to project the  $t_R$  information with high accuracy. Hall and co-workers also developed an RTI system for non-targeted metabolomics, similar to KRI, using *n*-nitroalkanes.<sup>19</sup> This approach requires detailed descriptions for analyzing the reproducibility of RTI system under different LC conditions (different gradient elution program, mobile phase compositions and stationary phase). The use of a logarithmic scale for the retention times and also an additive function for the number of carbon atoms remain problematic since the number of carbon atoms is not solely, directly and equally correlated to the polarity in LC.<sup>20,21</sup> Moreover, the RTI system was proposed only for a structurally diverse group of 411 small molecules, consisting of endogenous and drug metabolites, which limits its use in environmental analysis. Compounds chosen to test this system were limited to the set of biological elements (C, H, N, O, S, and P) and contained at least one atom capable of accepting protons to facilitate their detection in mass spectrometry. However, there is a limitation in using *n*-nitroalkanes as calibrants in the electrospray ionization mode (ESI) in MS, as they are not ionized properly at reasonable concentration range.

Li and coworkers<sup>22</sup> established a  $t_R$  calibration method for metabolomics studies, which was based on amino acid standards and a chemical labeling strategy. The  $t_R$  values acquired under different LC conditions (with similar mobile phase composition) showed high correlation. However, different columns would require different series of calibrants. In addition, it was necessary to apply dansylation labeling on the samples and individual standards to establish the library. The chemical labeling was extensively used in metabolomics to label metabolites that carried the same moiety to increase their ionization efficiency and LC separation in RPLC.<sup>23</sup> Recently, another RTI approach was developed based on chemical labeling reactions for carboxylated and amine metabolites.<sup>24</sup> In this approach, a series of 2-dimethylaminoethylamine (DMED)-labeled fatty acids were used as calibrants to create an RTI system for DMED-labeled carboxylated compounds. However, the variation of  $t_R$  values between different LC setting was still inevitable. It would be difficult to apply these RTI standards in generalized MS-based screening because of the lack of sufficient homologues in the series and low MS instrumental response factors. It would not be practical in the cases of environmental samples to apply dansylation or other chemically labeling protocol due to the existence of many classes of emerging pollutants and the lack of a similar chemical labeling strategy. In addition, structural differences between chemically-labelled-RTI standards and screened or detected compounds also could lead to errors and false positive results.

Studies published to date on the  $t_R$  prediction in LC can be divided into two types: (i) direct experimental  $t_R$  mapping,<sup>18,25,26</sup> which can be used irrespective of the specific LC condition, and (ii)  $t_R$  models established by quantitative structure-retention relationship (QSRR) approach, which works locally for specific LC conditions.<sup>8,10-12,27-29</sup> The common strategy to predict  $t_R$  of any compound focuses primarily on exploring the set of physicochemical descriptors (such as hydrophilicity, polarizability, electronegativity etc.), which yield insight into mechanisms of the elution (interpretability) in contrast to the projection of  $t_R$ . The correlation between hydrophobicity and  $t_R$  is generally the first step towards understanding the elution of a compound in reversed phase liquid chromatography (RPLC).<sup>8,28</sup> These QSRR models, however, rely on  $t_R$  information for a large number of compounds to offer an interpretable elution mechanism and correct prediction results.<sup>8,12</sup> The models can be overfitted, or only applicable for specific groups of compounds within the training set.<sup>9,28</sup> The cause for overfitting is often

an inadequate number of compounds used as training dataset to define the chemical space boundaries during the QSRR modelling, a desire to have a more specific model for a specific set of substances, or uncertainties in the calculation of molecular descriptors.<sup>30</sup> The development of large  $t_R$  databases and addressing the uncertainty in  $t_R$  prediction results would increase the reliability and usability of the models.<sup>31</sup> Specifically for the high throughput screening, e.g., of environmental or exposome samples, method covering a broad chemical application domain (AD) are required as effectively all substance classes amenable to LC-MS can be expected. Recently, a QSRR workflow was developed, which includes an AD function to estimate whether the predicted  $t_R$  should be accepted or rejected for a plausible candidate.<sup>7,8</sup> In addition to the accuracy and interpretability offered by this QSRR model, the approach needed to be extended and coupled to  $t_R$  projection strategies to be applicable under various LC conditions, with the objective of developing a broadly-applicable RTI in LC to facilitate the screening and control LC quality.

This work presents the development and validation of a new RTI system in reversed-phase liquid chromatography (RPLC) that is based on the calibration of the elution patterns demonstrated using two sets of 18 substances (one set for +ESI and one another set for -ESI) selected from 2123 chemicals of emerging contaminants concern (ECs) (303 and 1820 compounds for -ESI and +ESI, respectively). The applicability of this approach is demonstrated in an interlaboratory comparison to harmonize the  $t_R$  information in HRMS screening efforts, predict  $t_R$  across different LC conditions and make  $t_R$  information more useful in suspect and non-target screening workflows.

## EXPERIMENTAL SECTION

**Materials for Core Laboratory (National Kapodistrian University of Athens (NKUA)); Chemicals.** Reference standards for pesticides were donated to the NKUA laboratory by Bruker Daltonics (Bremen, Germany) at a concentration of 1 mg L<sup>-1</sup> in methanol. The rest of the major classes of emerging contaminants included in the study were purchased from Sigma–Aldrich. Individual stock solutions of these compounds were prepared in methanol at a concentration of 1 g L<sup>-1</sup> and stored at -20°C. Working solutions were prepared in methanol at a concentration of 1 mg L<sup>-1</sup>. A list of these chemicals can be found in <sup>7</sup> and **SIF Table S1** and **S2** (the supplementary Excel file accompanying this manuscript). LC-MS grade acetonitrile (ACN) and methanol (MeOH) were purchased from Merck, whereas LC-MS grade 2-propanol was purchased from Fisher Scientific. Sodium hydroxide monohydrate (NaOH) for trace analysis ≥99.9995%, ammonium acetate, ammonium formate and formic acid, all LC-MS grade, were purchased from Fluka, Sigma–Aldrich. Distilled water used for LC–MS analysis was provided by a Milli-Q purification apparatus (Millipore Direct-Q UV). Regenerated cellulose (RC) syringe filters (15 mm diameter, 0.22 μm pore size) were provided from Phenomenex.

**Materials for Core Laboratory (National Kapodistrian University of Athens (NKUA)); Instrumentation and Procedure.** An ultrahigh-performance liquid chromatography (UHPLC) system with a LPG-3400 pump (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific), interfaced to a Quadrupole Time of Flight (Q-ToF) mass spectrometer (Maxis Impact, Bruker Daltonics) was used for the screening analyses. The same chromatographic and instrumental conditions described previously<sup>4,8</sup> were used to record  $t_R$  on a RPLC platform for numerous ECs. More details about the instrumental and chromatographic conditions are described in the **SIF SI-A 1.1**.

**Development of Retention Time Indices.** The aim was to select the most appropriate set of calibrants from a  $t_R$  dataset of 2123 compounds (303 and 1820 compounds for -ESI and +ESI, respectively), considering the  $t_R$  distribution and the chemical similarity of the selected calibrants with the rest of the compounds in the chemical space. This process included seven steps. First, two datasets developed for -ESI and +ESI modes were processed separately to calculate the molecular descriptors.<sup>7</sup> These molecular descriptors were calculated using Padel<sup>32</sup> and E-DRAGON<sup>33</sup> on the final lowest energy 3D conformer of standardized structures achieved by Balloon.<sup>34</sup> In addition, ChemAxon<sup>35</sup> was used to calculate  $\log D^{36}$  at the pH of the LC mobile phases, which was 3.6 (+ESI) and 6.2 (-ESI). In total, 3200 molecular descriptors were calculated for each compound, representing the constitutional, topological, geometric, electrostatic, hydrophobicity, steric effect, quantum related chemical descriptors, chemical fingerprints (PubChem chemical fingerprint<sup>37</sup>) and various 3D molecular descriptors.<sup>38-40</sup> Constant and near constant molecular descriptors were removed from each dataset. The remaining datasets, consisting of 303 (-ESI) and 1820 compounds (+ESI), were further processed with collinearity removal. First, a threshold for the collinearity removal was set to 0.9, where the molecular descriptor that correlated less with the retention time was removed while its collinear pair was retained. This yielded 545 and 607 molecular descriptors for positive and negative ESI, respectively. Second, principal component analysis (PCA) was performed on the retained molecular descriptors (excluding retention time) to project the chemical properties of the compounds based on their covariance in three principal components (PCs, explained variance was over 75%). Third, a matrix containing retention time information and the three PCs was prepared separately for each ESI as input for identifying the potential RTI calibrants.  $t_R$  information along with PCs are the most representative data to address the similarity between compounds as well as the elution in LC. Fourth, several subsets of compounds were created and the overlap of normal distribution (objective function) between these subsets of compounds (RTI calibrants) and the rest of the compounds were calculated at predefined desired number of calibrants for each subset. Fifth, the algorithm seeks to detect the potential subsets of compounds that would increase the overlap of normal distribution of RTI calibrants and the rest of the chemical space. The sixth step combines all compounds between subsets to maximize the objective function. In the seventh step, the prediction accuracy of models are evaluated for a test set. Creating the large number of population of compounds and seeding the subsets of compounds and their combinations required a dynamic algorithm to train itself for selecting the best couple of compounds as calibrants. For such a case, a nature-inspired optimization algorithms was found to be helpful.<sup>41</sup> Here, the Ant Colony Optimization (ACO)<sup>42,43</sup> technique was used to select the optimal number of compounds by testing sets of 5, 8, 10, 12, 15, 18, 20, 22 and 25 compounds as LC calibrants. The optimum number of 18 compounds out of the tested was found for both ESI modes by comparing their overlap values. The external prediction accuracy of the RTI system was evaluated by 30 (+ESI) and 15 (-ESI) external compounds (also selected by ACO-SI). More details about the ACO approach can be found in **SI 1.2 (SI-A)**.

The RTI systems (one for +ESI and one for -ESI) were proposed considering the minimum and maximum elution that was observed for a generic RPLC method and scaled to 1000. This scale system between 1 and 1000 was set to have large RTI units between compounds that elute differently and compare the error more realistically. The RTI system proposed here was formulated as below:

$$RTI = \frac{(t_{Rx} - t_{Rmin})}{(t_{Rmax} - t_{Rmin})} \times 1000 \rightarrow RTI = \alpha(t_{Rc}) + C \quad Eq. 1$$

where  $t_{Rx}$  and  $t_{Rc}$  are the  $t_R$  observed for the calibrants and a compound, respectively,  $t_{Rmin}$  and  $t_{Rmax}$  are the minimum and maximum  $t_R$  observed for the calibrants, respectively. This is needed to define the elution segment for the calibrants in proportion to minimum and maximum  $t_R$  values seen for the calibrants in a LC setting. Once the RTI unit is established, they are set as global values and linear calibration curve was established to get use of these RTI values for a new LC setting.  $\alpha$  and  $C$  are the slope and the intercept at 99% confidence intervals (CIs). The linear correlation forms the RTI calibration equation and is used to transform any experimental  $t_R$  in one system into the RTI for that compound. The RTI values can be used afterwards to harmonize the elution of compounds in various LCs. This could also facilitate the evaluation of LC quality in case of multiclass LC method according to the degree of deviation from linearity (or use of lack-of-fit to examine the residuals after calibration curve development) for RTI calibrants.

**Stability test of RTI calibrants.** Four mixtures (separate mixtures of 18 compounds as the RTI calibrants set for each ESI mode to be tested at two different temperatures) were prepared at a concentration of 2 mg L<sup>-1</sup> in pure methanol (at a final volume of 250  $\mu$ L). 2-hydroxy-simazine was prepared in methanol: phosphoric acid (99:1, v/v) and vancomycin was prepared in MeOH:H<sub>2</sub>O (90:10, v/v) to enhance the solubility. Moreover, guanylurea and coumaphos were soluble in methanol at room temperature, under ultrasonic irradiation for 15 min. The stability test was performed by analyzing each mixture at 0, 6, 18, 24, 36 and 48 hours of storage time at two different temperatures (-18 and +2 °C). The mixtures were returned to the refrigerator and freezer after each analysis time point and stored for the next injection.

**Internal validation.** An intra-laboratory evaluation followed considering various C<sub>18</sub> columns (Acclaim RSLC C<sub>18</sub>, Atlantis T3 C<sub>18</sub>, XBridge C<sub>18</sub> Waters, and Acquity UPLC BEH C<sub>18</sub> column), mobile phase compositions (MeOH, H<sub>2</sub>O and ACN with and without buffer system) and elution gradient. In addition to our previous method<sup>8</sup>, two other gradient elution program were adopted from literature<sup>1,44</sup> to evaluate the reproducibility of the proposed RTI system. More details about the instrumental and chromatographic conditions are described in the **SIF SI-A 1.3.1 and SI-B Table S3**. The performance of the proposed RTI systems was studied in terms of root mean square error (RMSE), square correlation coefficient ( $R^2$ ), distribution of residuals derived from both experimental and predicted RTI/ $t_R$  and the true positive harmonization rate of the elution of compounds in different LC conditions. The true positive harmonization rate (TPHR) is derived by the following equation:

$$TPHR = \frac{\text{No. True Harmonised RTIs within 99\% CIs}}{\text{No. detected compounds}} \times 100 \quad Eq. 2$$

TPHR indicates the percentage of RTI values for the compounds that have the overlap with the experimental RTIs/its 99% CI under various LC conditions in contrast to the RTI values derived from the main LC condition. TPHR was also visualized as a cloud plot using multiple comparison procedures. In this cloud plot, the bubble size is proportional to the CIs of the experimental RTI (at 99% CI from calibration curve) in which the overlap between these bubbles correspond to the successful harmonization of elution of compounds from one LC condition to another. To compare the LC conditions based on experimental  $t_R$  and RTI values statistically, the student t-test was used.

**External validation; Reference Laboratories for Initial Evaluation.** In addition, inter-laboratory comparisons were organized with the aim of evaluating the accuracy of the RTI system from one lab to another with completely different instrumentation. The proposed RTI system was evaluated initially by three laboratories (the

Swiss Federal Institute for Aquatic Science and Technology (Eawag), the Helmholtz Centre for Environmental Research (UFZ) and the University of Jaume I (UJI). Eawag used LC condition reported in <sup>1</sup>, UFZ conducted the evaluation under the LC condition reported in <sup>45</sup> and UJI applied the LC condition reported in <sup>46</sup> to evaluate the RTI system. More details about the instrumental and chromatographic conditions for all the participants are described in the **SI-A 1.3.2**. The  $t_R$  information for the calibrants as well as a list of compounds as a blind set (for evaluating the external prediction capability of proposed RTI system) were reported by each laboratory.

**External validation; External Application and Validation through Collaborative Trials.** Four laboratories evaluated the proposed RTI system externally<sup>47</sup> within the NORMAN network in addition to the core laboratories. These external laboratories were the United States Environmental Protection Agency (EPA), the Swedish University of Agricultural Sciences (SLU),<sup>48</sup> the University of Antwerp and the University of California (UC Davis)<sup>49</sup>. The details about the LC conditions can be found in **Table S4 (SI-B)**. This was done over a four-years-period during various joint collaborative trials such as analysis of indoor house dust initiatives<sup>47</sup> and EPA's ongoing Non-Targeted Analysis Collaborative Trial (ENTACT).<sup>50</sup>

**Uncertainty estimation and modelling.** The modelling workflow introduced in our previous study<sup>7,8</sup> was applied to predict and model RTI values. The details about each step of the modelling workflow can be found in **SI 1.4 (SI-A)**.

There were two layers of uncertainties, one arising from the QSRR models and one from the calibration curve of RTI versus  $t_R$ . Compounds falling within the 99% CI of the experimental RTI accuracy among different LC conditions were considered to be identical throughout the various LC conditions. Rigorous statistical comparisons included the student t-test, ANOVA, least significant difference (LSD)<sup>51</sup> and multiple comparison procedures.<sup>52,53</sup> For single RTI values between two labs, lower and upper CIs from the calibration curve were used to perform these tests and thus the difference between two experimental RTIs measured for a compound by two labs is significant when

$$|\overline{RTI}_{lab1} - \overline{RTI}_{lab2}| \geq t_{N-k,1-\frac{\alpha}{2}} \times \sqrt{S_l^2 \left( \frac{1}{n_{lab1}} + \frac{1}{n_{lab2}} \right)} \quad Eq. 3$$

where  $t_{N-k,1-\frac{\alpha}{2}}$  is the critical student t value at  $N-k$  degrees of freedom for a significance level set to  $\alpha$ .  $N$  is the total number of observations,  $k$  is the number of labs and  $S_l^2$  is the estimation of the variance within the labs.

To represent the uncertainty in prediction results *via* the QSRR method, leverage, standardized residuals (relative  $t_R$  error window) and normalized mean distance were combined into a single 3D *bubble plot*, the so called OTrAMS.<sup>8</sup> This is a method based on QSRR error diagnosis to decrease the chance of false positives in those cases where several plausible candidates were present for a single  $t_R$ /RTI value. For a compound without an experimental  $t_R$ /RTI measured, the only way to address the applicability domain of the QSRR is to use the chemical space boundaries. Therefore, if the error is the function of chemical space failure, warning leverage values versus normalized mean distance can be used to define the applicability domain.<sup>54</sup> These two methods (OTrAMS and chemical space boundaries) were used here during RTI modelling and applied across the prediction results of RTI in the inter-laboratory study. The details about OTrAMS and chemical space boundaries can be found in **SI 1.5 (SI-A)**.

**Suspect and Non-target Screening with the DSFP.** In contrast to retention time prediction, matches between experimental retention time indices is a valuable indicator for a candidate to be true positive.<sup>55</sup> This is done more routinely for GC platforms and only recently in LC-HRMS.<sup>47</sup>

The NORMAN Digital Sample Freezing Platform (DSFP) was used to demonstrate the application of the RTI system in suspect and non-target screening.<sup>56</sup> NORMAN DSFP is a digital archive for environmental samples that accepts chromatograms as mzML files accompanied by the appropriate metadata (contributor details, instrumental settings, matrix-specific metadata and the retention time of the calibrant substances). The system contains an integrated workflow for generation of the component list (list of  $m/z$  of all detected compounds) and the experimental RTI of all features, calculated using the experimental retention time of the components *via* **Eq. 1**. A query for the investigation of the occurrence of any substance in digitally archived LC-HRMS/MS data involves the match of (i) the accurate mass of the selected adduct and the experimental  $m/z$  of the component list, (ii) the plausibility of the retention time based on the observed and predicted RTI (in case the compound has proven to be inside the applicability domain of the RTI models), (iii) isotopic fit and (iv) the presence of fragment ions.

LC-HRMS data from two laboratories equipped with LC-Orbitrap and LC-QTOF (Laboratory A and B, respectively) were used. The instrumental setup for the laboratories is provided in **Table S5**. The candidate substances were retrieved from PubChem compound database<sup>57</sup> *via* MetFrag, once identifications reached the level of unequivocal molecular formula. MetFrag was used to fragment the candidate structures.<sup>58</sup> RTI was predicted for the candidate structures and a RTI score was derived using the error between predicted and experimental RTI values (1000 minus error between predicted and experimental RTIs divided by 1000 where 1000 is the RTI range). The sum of the fragmentation score and the RTI score was used to rank the candidate substances.

**RTI platform.** All RTI related calculations of suspected compounds were performed using the online tool 'Development and Prediction of Retention Time Indices for LC-HRMS', accessible at <http://rti.chem.uoa.gr/>. The tool is provided with a detailed user guide and describes how to submit suspect lists in the required format, the procedure for establishment of the calibration curve for RTI calibrants, the RTI prediction and multiple comparison procedure for the experimental RTIs.

## RESULTS AND DISCUSSION

**Selection of RTI calibrants.** **Figure S1 (SI-A)** shows the final overlap that is achieved based on 18 RTI calibrants (18 set of compounds for +ESI and another 18 compounds for -ESI), shown separately for experimental  $t_R$  and chemical space. The optimal number of calibrants was achieved using the lowest residuals derived between the predicted and experimental RTI and  $t_R$  for 30 and 15 external compounds as validation set in positive and negative ESI, respectively. These compounds were selected with the same algorithm as the calibrants, using ACO-SI. The errors observed for these compounds, after using various number of calibrants, are shown in **Figure S2 (SI-A)**.

**Table 1.** Prediction performance of RTI models.

	Training			Test	
	R <sup>2</sup>	RMSE	Q <sub>LoO</sub> <sup>2</sup>	R <sup>2</sup>	RMSE
<i>RTI for (-ESI):</i>					
ACO-MLR	0.835	89.221	0.827	0.801	84.606
ACO-SVM	0.984	27.709	0.813	0.833	75.703
<i>RTI for (+ESI):</i>					
ACO-MLR	0.847	89.187	0.846	0.835	92.630
ACO-SVM	0.945	53.605	0.864	0.868	82.642

The lowest distribution of error (between  $\pm 1$ ) is derived using 18 calibrants (in both +ESI and -ESI). However, a lower error is observed for +ESI due to the larger data set available (1820 vs 303 compounds). The stability test of each individual calibrant in the prepared mixture is provided in **Figure S3 (SI-A)**. The RTI calibrants are stable within 48 hours of consecutive analysis, in the worst case, a drop in their response factor up to 60% is observed for some individual calibrants.

**Modeling Retention Time Indices.** Two models were built, after calibrating the  $t_R$  values of large number of emerging contaminants to RTI ( $\pm$ ESI) values, using ACO to select the most representative molecular descriptors and Support Vector Machine (ACO-SVM) to non-linearly correlate these molecular descriptors with RTI. The performance of the models for the proposed RTI system can be found in **Table 1**. Both models show high correlation coefficients and leave one out cross-validation with low root-mean-square-error (RMSE). The linear models (ACO-Multiple Linear Regression), however, showed less prediction accuracy than ACO-SVM. It is important to note that although the compounds detecting in -ESI mode have often unique chemical structure and may have distinct and different elution pattern than rest of chemical space, the models developed here for two ESI modes are irrespective their detection mode and mainly due to small changes in the LC setting of -ESI mode.

The main molecular descriptor used in the RTI model (for -ESI) was  $\log D$  ( $\text{pH}=6.2$ ) (with a relative importance of 67.7%).  $\log D$  (the pH-dependent partition coefficient of all forms of a compound (both neutral and ionized) in each of two phases, generally octanol and water) is a parameterized representation of the hydrophobicity displayed in chromatography. Further descriptors were the largest absolute eigenvalue of Burden modified matrix - n4 /weighted by relative mass (SpMax4\_Bhm, which is eigenvalues of a modified connectivity matrix and can represent the steric effect in the chemical graph) (with an importance of 19.6%), electronic profile of the molecule relative to molecular size (ETA\_BetaP) (with an importance of 4.5%) and the chemical fingerprint of atom paired (C-S) (with an importance of 8.2%). The linear equation to derive experimental RTIs from the  $t_R$  values of calibrants as well as QSRR-based predicted RTIs in -ESI are formulated below:

$$\text{Exp. RTI}_{(-\text{ESI})} = 76.90 (\pm 0.120)t_{R(-\text{ESI})} - 128.2 (\pm 0.015) \quad \text{Eq. 4}$$

$$\begin{aligned} \text{Pred. RTI}_{(-\text{ESI})} = & + 90.268 (\pm 78.778) \\ & + 68.692 (\pm 2.955) \log D_{\text{pH}=6.2} \\ & + 173.29 (\pm 20.568) \text{SpMax4}_{\text{Bhm}} \\ & - 264.78 (\pm 46.864) \text{ETA}_{\text{BetaP}} \\ & - 97.937 (\pm 14.594) \text{PubchemFP293}[\text{C} \\ & - \text{S}] \quad \text{Eq. 5} \end{aligned}$$

For +ESI, the descriptors used to model the RTI were  $\log D$  ( $\text{pH}=3.6$ ) (with a relative importance of 68.3%), and the charged partial surface area (THSA)<sup>59</sup> with a relative importance of 22.0% (which is sum of solvent accessible surface areas of atoms with absolute value of partial charges less than 0.2), 3D topological distance-based autocorrelation lag 5 / weighted by covalent radius (TDB5r) with a relative importance of 7.9% and hybridization ratio (HybRatio) with a relative importance of 1.8%. The linear equation to derive experimental RTIs from  $t_R$  values of calibrants as well as QSRR based predicted RTIs in +ESI are formulated below:

$$\text{Exp. RTI}_{(+\text{ESI})} = 76.38 (\pm 0.012)t_{R(+\text{ESI})} - 99.91 (\pm 0.100) \quad \text{Eq. 6}$$

$$\begin{aligned} \text{Pred. RTI}_{(+\text{ESI})} = & - 57.003 (\pm 19.619) \\ & + 70.903 (\pm 1.119) \log D_{\text{pH}=3.6} \\ & + 159.88 (\pm 11.555) \text{HybRatio} \\ & + 62.219 (\pm 10.285) \text{TDB5r} \\ & + 0.5516 (\pm 0.0226) \text{THSA} \quad \text{Eq. 7} \end{aligned}$$

The experimental and predicted RTIs for  $\pm$ ESI modes can be found in **Table S1** and **S2 (SI-B)**. **Figure S4 A&B (SI-A)** show the results of OTrAMS for the predicted RTI using **Eq. 5 and 7**. All the compounds are within the acceptance threshold of  $\pm 3\text{SR}$  (Standardized Residuals, box3) and within the chemical space boundaries (**Figure S4 C&D (SI-A)**). This helps to identify the potential source of inaccuracies in the prediction of RTI when the error is the subject of chemical space failure.

**Intra-laboratory accuracy of RTI.** The  $t_R$  values observed in +ESI for RTI calibrants in each LC condition described in **SI-A 1.3.1** can be found in **Table S6**. The internal and external performance of the proposed RTI system in +ESI can be found in **Table 2**. The RTI proposed for +ESI generally has better outcomes in contrast to RTI for -ESI. The internal and external accuracy of the proposed RTI system in +ESI is found to be reliable and work well in a mobile phase composition of MeOH:H<sub>2</sub>O, performed with any column type and gradient elution program, with a TPHR of 100%, as shown in **Figure 1A**. The result of the multiple comparison procedure for evaluating the internal accuracy of RTI in +ESI can be found in **Table S7 (SI-B)**. The proposed RTI system was evaluated internally at four different LC conditions, changing the mobile phase composition, gradient elution program and column type. The retention time values observed for RTI calibrants in -ESI and each LC condition described in **SI-A 1.3.1** can be found in **Table S8**. Dinoterb and valproic acid were outside the prediction limits while building the RTI calibration equation in LC condition 4. LC condition 4 remained less accurate than the other LC conditions because ACN:H<sub>2</sub>O was used as

**Table 2.** RTI and  $t_R$  linear equations for (+ESI) in different LC conditions

LC conditions	RTI versus $t_R$ equation	Standard Error	Internal accuracy	External accuracy (n=30)
LC1	$\text{RTI} = 72.40(t_R) - 98.67$	Intercept: $\pm 10.21$ Slope: $\pm 1.151$	$R^2=0.996$	$R^2=0.960$ , RMSE=46.61, TPHR = 100%
LC2	$\text{RTI} = 59.60(t_R) - 66.91$	Intercept: $\pm 20.82$ Slope: $\pm 2.020$	$R^2=0.982$	$R^2=0.930$ , RMSE=63.78, TPHR = 100%
LC3	$\text{RTI} = 60.02(t_R) - 47.39$	Intercept: $\pm 19.04$ Slope: $\pm 1.909$	$R^2=0.984$	$R^2=0.930$ , RMSE=64.78, TPHR = 100%
LC4	$\text{RTI} = 35.62(t_R) + 12.29$	Intercept: $\pm 29.74$ Slope: $\pm 1.942$	$R^2=0.952$	$R^2=0.779$ , RMSE=119.59 TPHR = 84.61%

LC1: Atlantis T3 C18; Mobile Phase: H<sub>2</sub>O/MeOH with 5 mM ammonium acetate and 0.01% formic acid; Flow rate: multi-flow-rate gradient; Run time: 15 min

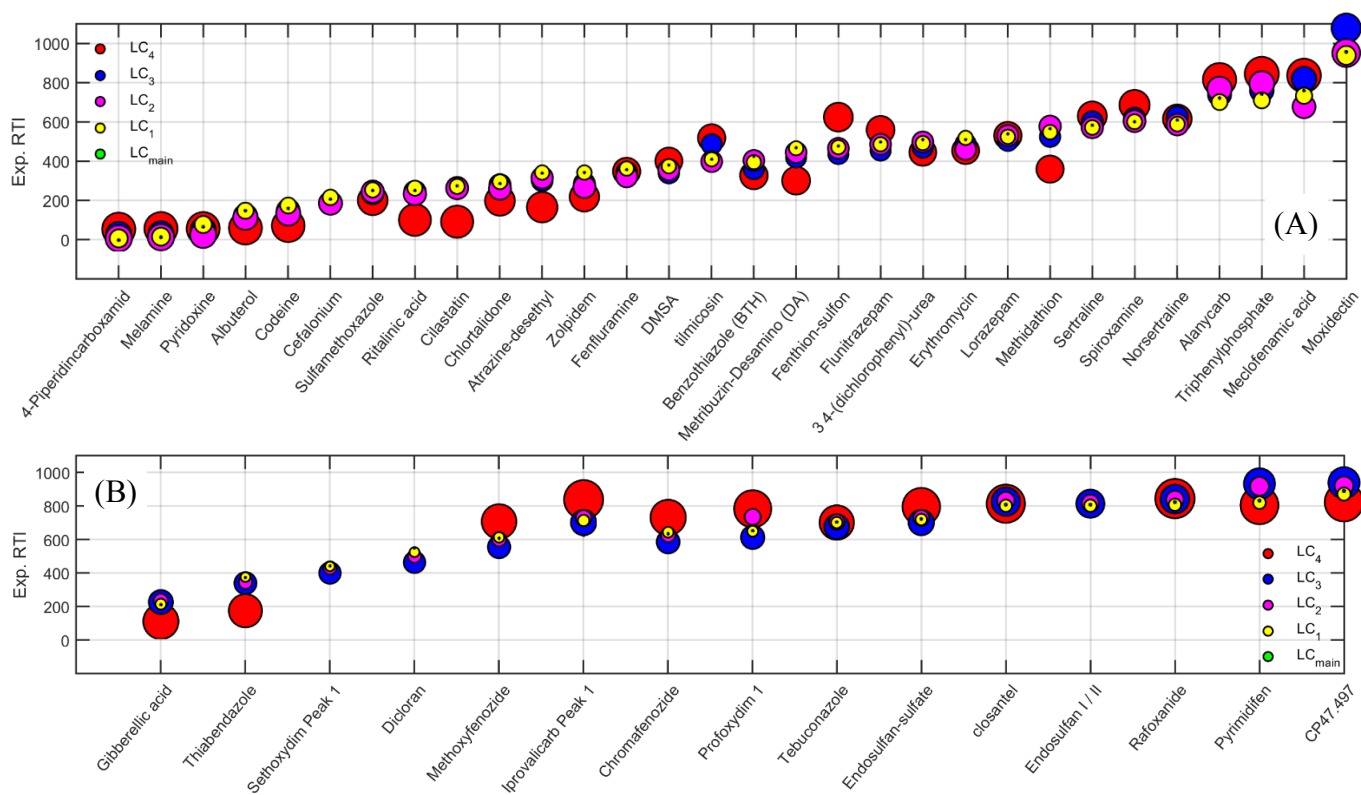
LC2: Acclaim RSLC C18; Mobile Phase: H<sub>2</sub>O/MeOH with 0.1% formic acid; Flow rate: 0.200 mL/min; Run time: 25 min

LC3: Acquity UPLC BEH C18; Mobile Phase: H<sub>2</sub>O/MeOH with 0.1% formic acid; Flow rate: 0.200 mL/min; Run time: 25 min

LC4: Acquity UPLC BEH C18; Mobile Phase: H<sub>2</sub>O/ACN with 0.1% formic acid; Flow rate: 0.300 mL/min; Run time: 25 min

mobile phase and the LC method was not optimized considering the gradient time and column particle size after method transfer. Nevertheless, the use of quality parameters associated with the RTI calibration curve could be an added value during the quality control of LC setting for a multi-group ECs LC-MS method. The outcomes of evaluating the internal and external accuracy for the proposed RTI system in -ESI can be found in **Table 3**. The cloud plot of the experimental RTIs for the -ESI platform is shown in **Figure 1B**. The result of multiple comparison procedure for evaluating the internal accuracy of RTI for the -ESI platform can be found in **Table S9 (SI-B)**. Overall, the internal and external accuracy of the proposed RTI system in -ESI was found to be reliable and work well in mobile phase composition of MeOH:H<sub>2</sub>O with a T<sub>PHR</sub> of 100%. **Table S10 (SI-B)** shows the results of the student t-test between

main LC condition and the LC 4 condition (as the most diverse LC setting). The student t-test indicates that two LC conditions are identical at 99% CIs based on experimental RTI values, whereas they are significantly different based on t<sub>R</sub> values. For instance, CP47.497 gives a t<sub>R</sub> value of 13.24 min (LC main) and 24.29 min (LC 4) while the RTI values are steady (RTI value in LC main=869.18 and LC 4= 868.04). Despite the error observed for LC4, the harmonization of t<sub>R</sub> values was achieved for over 30 compounds, indicating the use of this approach in most difficult scenarios. Overall, the proposed RTI system can harmonize the t<sub>R</sub> information throughout different LC conditions in the +ESI mode better than the -ESI mode. The plot of experimental *versus* predicted RTI based on QSRR models for these four LC conditions are presented in **Figure S5 (SI-A)**.



**Figure 1.** The cloud plot of experimental RTIs measured in various LC conditions with their acceptance CIs in (A) +ESI and (B) -ESI

**Table 3.** RTI and t<sub>R</sub> calibration curve for (-ESI) under different LC conditions

LC conditions	RTI versus t <sub>R</sub> equation	Standard Error	Internal accuracy	External accuracy (n=15)
LC1	RTI = 73.12 (t <sub>R</sub> ) -121.6	Intercept: ±5.800 Slope: ±0.6835	R <sup>2</sup> =0.999	R <sup>2</sup> =0.970, RMSE= 36.26, TPHR= 100%
LC2	RTI = 79.50 (t <sub>R</sub> ) -30.98	Intercept: ±9.424 Slope: ±1.386	R <sup>2</sup> =0.995	R <sup>2</sup> =0.953, RMSE=50.14, TPHR= 100 %
LC3	RTI = 58.15 (t <sub>R</sub> ) + 67.66	Intercept: ±21.16 Slope: ±2.698	R <sup>2</sup> =0.967	R <sup>2</sup> =0.932, RMSE=54.74, TPHR= 100 %
LC4	RTI = 32.42(t <sub>R</sub> ) + 67.00	Intercept: ±29.33 Slope: ±2.056	R <sup>2</sup> =0.947	R <sup>2</sup> =0.872, RMSE=110.608, TPHR= 66.66 %

LC1: Atlantis T3 C18; Mobile Phase: H<sub>2</sub>O/MeOH with 5 mM ammonium acetate; Flow rate: multi-flow-rate gradient; Run time: 15 min

LC2: Acclaim RSLC C18; Mobile Phase: H<sub>2</sub>O/MeOH with 5 mM ammonium acetate; Flow rate: 0.200 mL/min; Run time: 25 min

LC3: XBridge C18; Mobile Phase: H<sub>2</sub>O/MeOH with 5 mM ammonium acetate; Flow rate: 0.200 mL/min; Run time: 25 min

LC4: XBridge C18; Mobile Phase: H<sub>2</sub>O (with 5 mM ammonium acetate)/ACN; Flow rate: 0.300 mL/min; Run time: 25 min

<sup>a</sup> Dinoterb and Valproic acid were outside of the prediction limit

**External Accuracy of RTI; Evaluation of RTI proposed in +ESI.** All seven labs showed the internal accuracy for the RTI calibration curve with  $R^2$  above 0.95 for +ESI. For the external accuracy, TPHR values were above 89% for all the labs and  $R^2$  was greater than 0.820. The data from EPA, SLU and UC Davis, which used completely different elution gradient programs and mobile phases, showed similar internal and external accuracy results in contrast to the LC settings used by NKUA, Eawag, UJI and UFZ. The results from the University of Antwerp, which used a different stationary phase chemistry (Kinetex Biphenyl core-shell particle rather than  $C_{18}$ ), show that the RTI calibrants still follow the elution order pattern initially designed for  $C_{18}$  columns. Only small and negligible deviations (within 99% CIs) were observed for three non-polar compounds (Tylosin, TCMTB and Rifaximin) in the calibration curve. The results are shown in **Table 4**. UFZ and EPA evaluated the external accuracy with a set of 607 and 443 compounds, respectively. Among these compounds, 100 and 85 compounds were matched to the NKUA +ESI database and TPHR values were calculated. 93 (for UFZ, TPHR= 93%) and 76 (for EPA, TPHR= 89%) compounds overlapped and fell inside the RTI CIs. Other labs also showed high TPHR values (from 89% to 100%) for the range of 29-44 compounds. More details about the multiple comparison process of individual compounds in each lab can be found in **Table S11**.

**External Accuracy of RTI; Evaluation of RTI proposed in –ESI.** Only five of the labs evaluated the RTI in –ESI mode, as shown in **Table 4**. This was partially because during NORMAN collaborative trials,<sup>47</sup> the screening efforts were allocated mainly to the evaluation of +ESI mode data, as fewer compounds were expected to be detected in –ESI. Overall, the internal accuracy in terms of  $R^2$  values are above 0.740 in –ESI. The RMSE is also higher in –ESI mode than +ESI mode. This maximizes the confidence intervals around the RTI calibration curve, with increasing probabilities for accidental overlap. Although this might cause some concerns during direct experimental RTI comparison, it could reflect the LC setting quality and demands for QC evaluation of the

LC setting. For instance, the UC Davis evaluated RTI under two different LC settings specifically designed for each ESI mode. Under such practice, the mobile phase for –ESI is not fortified with formic acid in contrast to other labs and the internal accuracy of the RTI calibration curve is  $R^2=0.948$ . The fact that the RTI calibrants in –ESI deviates from linearity when using the common LC setting for both polarities indicates that distinct groups of chemicals (like acids, phenols or amines) have different elution patterns due to different speciation under different pH conditions. Therefore, using the respective LC setting would be the appropriate action to preserve the elution pattern. TPHR values (from 94% to 100%) for a range of 11-33 compounds are provided for each lab in **Table S12**. The evaluation results of the RTI systems in all seven laboratories are briefly explained in **Table S13 (SI-B)**. This table summarizes the internal accuracy, external accuracy, application domain for predicted RTIs and the overlap of the RTI data with NKUA data. Moreover, the distribution of error is also depicted for each laboratory. The error is generally below a value of 180 RTI units (absolute). While this could be a rough threshold for screening workflows to adopt in screening workflows,<sup>56</sup> the correct RTI uncertainty range can be obtained every time from calibration curve CIs and QSRR. The pairwise harmonization and overlap of experimental RTI values are also depicted in **Table S13**. In addition, the overall harmonization ability of RTI across different LC settings for matched common compounds is demonstrated in **Table S14**. Regardless of the LC conditions used by each laboratory, the RTI value for each compound are not significantly different, while the  $t_R$  values are significantly different. For instance, the University of Antwerp reported experimental retention time of 21.51 min for triphenyl phosphate whereas other labs reported experimental retention time ranging from 11.00 to 14.00 mins. However, the experimental RTI values are harmonized between 658 and 808 which are within the uncertainty of RTI units and matching to the predicted RTI value (predicted RTI= 741). More examples of harmonization are provided in **Table S14 (SI-B)**.

**Table 4.** RTI and  $t_R$  calibration curve for ( $\pm$ ESI) reported by different laboratories

Labs	RTI versus $t_R$ equation	Standard Error	Internal accuracy	External accuracy
<i>For +ESI:</i>				
Lab01	RTI = 62.17 ( $t_R$ ) – 23.53	Intercept: $\pm 31.30$ ; Slope: $\pm 3.535$	$R^2=0.960$	$R^2=0.846$ , RMSE=105.8193, TPHR=93.00% (n=100)
Lab02	RTI = 60.03 ( $t_R$ ) + 18.00	Intercept: $\pm 27.27$ ; Slope: $\pm 3.068$	$R^2=0.967$	$R^2=0.884$ , RMSE=90.4343, TPHR=97.14% (n=35)
Lab03	RTI = 63.34 ( $t_R$ ) – 18.72	Intercept: $\pm 13.12$ ; Slope: $\pm 1.408$	$R^2=0.993$	$R^2=0.876$ , RMSE=81.8010, TPHR=96.67% (n=30)
Lab04	RTI = 66.35 ( $t_R$ ) – 10.65	Intercept: $\pm 26.70$ ; Slope: $\pm 3.130$	$R^2=0.966$	$R^2=0.831$ , RMSE=104.64, TPHR= 89.41% (n=85)
Lab05	RTI = 63.10 ( $t_R$ ) – 1.460	Intercept: $\pm 16.94$ ; Slope: $\pm 1.850$	$R^2=0.990$	$R^2=0.879$ , RMSE=60.79, TPHR= 100% (n=44)
Lab06	RTI = 38.21 ( $t_R$ ) – 13.25	Intercept: $\pm 33.85$ ; Slope: $\pm 2.226$	$R^2=0.952$	$R^2=0.821$ , RMSE=79.87, TPHR= 100% (n=2)
Lab07	RTI = 64.34 ( $t_R$ ) – 87.53	Intercept: $\pm 42.35$ ; Slope: $\pm 4.340$	$R^2=0.944$	$R^2=0.862$ , RMSE=77.72, TPHR= 100% (n=29)
<i>For –ESI:</i>				
Lab01	RTI = 54.28 ( $t_R$ ) – 60.61	Intercept: $\pm 85.95$ ; Slope: $\pm 8.691$	$R^2=0.736$	$R^2=0.710$ , RMSE=122.405, TPHR= 100 % (n=33)
Lab02	RTI = 36.95 ( $t_R$ ) – 65.01	Intercept: $\pm 86.26$ ; Slope: $\pm 5.761$	$R^2=0.760$	$R^2=0.681$ , RMSE=146.358, TPHR=100 % (n=11)
Lab03	RTI = 62.98 ( $t_R$ ) – 124.4	Intercept: $\pm 63.13$ ; Slope: $\pm 6.310$	$R^2=0.877$	$R^2=0.687$ , RMSE=123.628, TPHR= 94.12 % (n=17)
Lab04	RTI = 65.74 ( $t_R$ ) – 86.49	Intercept: $\pm 74.65$ ; Slope: $\pm 7.900$	$R^2=0.863$	$R^2=0.862$ , RMSE=115.690, TPHR= 100% (n=24)
Lab07	RTI = 58.07 ( $t_R$ ) – 99.31	Intercept: $\pm 38.04$ ; Slope: $\pm 3.770$	$R^2=0.948$	$R^2=0.882$ , RMSE=76.0654, TPHR= 100% (n=19)

Lab01: Department Effect-Directed Analysis, Helmholtz-Centre for Environmental Research –UFZ

Lab02: Swiss Federal Institute for Aquatic Science and Technology (Eawag)

Lab03: Research Institute for Pesticides and Water, University Jaume I (UJI)

Lab04: United States Environmental Protection Agency (EPA)

Lab05: Swedish University of Agricultural Sciences (SLU)

Lab06: University of Antwerp

Lab07: University of California (UC Davis)



**Application of RTI in Suspect and Non-target Screening.** The RTI system has undergone rigorous testing within DSFP and is being used routinely for the reduction of false positive results and to support suspect screening. As illustrated in **Figure S6 (SI-A)**, the applicability of the system in suspect screening relies on the fact that it allows searching for a suspected compound in a specific chromatographic region, where the compound is expected to elute. As proof-of-concept example presented in **Figure S6 (SI-A)**, ibuprofen is retrospectively searched in the data of laboratory A. The extracted ion chromatogram (EIC) of the  $[M+H]^+$  adduct ( $m/z$  207.1378  $\pm$  0.002) resulted in four peaks with retention times 9.09, 9.79, 12.21 and 14.02 min. The predicted RTI was estimated to be 718, which is translated to a retention time of 13.34 min for the laboratory A based on **Eq. 1**. Therefore, the closest peak (12.21 min) was recognized to be ibuprofen, which was later verified by HRMS/MS library spectrum match. The peaks eluting at the retention times of 9.09, 9.79 and 14.02 min corresponded to molecules with unequivocal molecular formula  $C_{13}H_{18}O_2$ . However, it was not possible to propose further tentative structures at the time because of the high number of candidate substances retrieved from PubChem. In such cases, extrapolation to tentatively identified molecules is extremely challenging, despite the exclusion of non-plausible structures based on the RTI prediction. Continuing community developments including the development of smaller, relevant databases such as the CompTox Chemicals Dashboard<sup>60</sup> and PubChemLite<sup>61</sup> will further help reduce the number of candidates (e.g. 350 and 83 candidates respectively for this formula, compared with 9608 candidates from PubChem; all numbers calculated *via* MetFrag accessed on 16-12-2020).

Moreover, the use of the proposed RTI systems has been demonstrated in non-target screening identification workflows and especially in ranking candidate compounds. An indicative example is given in **Figure S7 (SI-A)**, in which an unknown compound was detected at retention time 15.69 min in data of Laboratory B. The peak provided a characteristic isotopic pattern indicative of the presence of a chlorinated atom in its molecular formula. The unequivocal molecular formula ( $C_{16}H_{22}ClNO_2$ ) was obtained based on the mass accuracy and the necessity to have at least one chlorine in the molecular formula. The HRMS/MS spectrum yielded four fragment peaks (**Figure S7 (SI-A)**). 249 candidates were able to explain the four HRMS/MS fragments based on MetFrag analysis, and the RTI score provided additional information in ranking the candidate structures. The top ten ranked compounds are presented in **Table S15 (SI-B)**. Four out of the top ten ranked compounds were positional isomers. Finally, the compound was identified at level 2A (pethoxamid, after matching with reference spectra in MassBank (AU253506), spectrum similarity score of 0.989 (dot product)), which was ranked eighth among the top ten candidates. Many similar examples have demonstrated the utility of the proposed RTI approach in the context of the dust collaborative trial (DCT) organized by the NORMAN Association.<sup>47</sup> In the DCT, 12 out of 21 participants, analyzing the dust samples with LC-HRMS, utilized the RTI system. Through the RTI system, some compounds such as diethylene glycol and morphine proved to be false positives and were excluded from the results.<sup>3</sup> Some RTI calibrants are not detected under some LC conditions (such as coumaphos and buto-carboxim which show a low ionization efficiency),  $m/z$  range settings (for instance amitrole  $m/z$  85.05089 will not be detected if MS operates between  $m/z$  100-1000 Da) or LC condition and void time (e.g., very polar compound such as guanyurea). Since the overlap between RTI calibrants and the chemical space of ECs is high, there are compounds in the calibrants list to compensate the missing ones and establish calibration curve. The proposed RTI

system is now used by members of the NORMAN Association in various activities related to suspect and non-target screening.<sup>62</sup>

Several  $t_R$  prediction approaches have recently been evaluated by Chaker et al.,<sup>13</sup> including PredRet,<sup>18</sup> Retip,<sup>10</sup> the present work and a generic logP model with the aim to incorporate these into suspect screening and aid identification. Although the RTI was developed for environmental analysis, in Chaker et al.,<sup>13</sup> the utility of RTI was independently tested in metabolomics. RMSE values were reported to be 11.5, 12.6 and 13.7% of run time for the RTI models, Retip and logP model, respectively. This suggests a more precise prediction of  $t_R$  values using the RTI model ( $R^2=0.767$  for xenobiotics ( $n=99$ )), than Retip, and the log P model. The PredRet approach could not be followed, as the number of common compounds required to establish a projection model was insufficient.

The RTI predictions were so far calculated for >65,000 substances and their transformation products listed in the NORMAN Substance Database,<sup>63</sup> using an online tool accessible at <http://rti.chem.uoa.gr/>. It is expected that this number will continue to increase rapidly. In the future, incorporation of the RTI system in routine analysis could promote the creation of experimental RTI database based on available reference standards, such as is already established for GC. This would provide even higher confidence for feature annotation and compound identification in LC-HRMS. The RTI system, initially tested within the NORMAN network, is nowadays being used by a wide scientific community for suspect and non-target screening of environmental samples involving analysis of surface, sea, ground, waste or drinking water, sediments, biota and indoor dust samples.

## CONCLUSIONS

In this work, a new approach was proposed to harmonize retention times ( $t_R$ ) obtained for the same compounds under various liquid chromatographic (LC) conditions. The harmonization of  $t_R$  is based on the calibration of LC methods *via* 18 calibrants into predefined retention time indices (RTI: 1-1000). These calibrants were selected from a total set of 2123 compounds (303 and 1820 compounds for -ESI and +ESI, respectively) using ant colony optimization similarity indices. This method was trained to identify the potential subsets of compounds in a way that their selection increases the overlap of the normal distribution of calibrants and rest of chemical space. It dynamically assessed the overlap of normal distribution of various combinations of compounds (calibrants) with the rest of the compounds (>2080). Intra- and inter-laboratory trials rigorously evaluated the harmonization ability of  $t_R$  in both  $\pm$ ESI modes. Quality assurance of the proposed RTI system was assisted by checking the cloud plot, multiple comparison procedure and confidence intervals for the calibrants. The main use of the RTI predictions *via* QSRR models is to reduce potential false positives resulting from suspect and non-target screening workflows. It has been demonstrated that the RTI predictions are LC system independent and can be applied under any LC condition, showing high internal accuracy and an acceptable degree of linearity for calibration curve. The uncertainty of correctly harmonizing the  $t_R$  values was found to be lower for +ESI than -ESI. This was partially due to lack of optimal LC settings for compounds that are being detected in -ESI. The applicability of this study might be limited if there is a poor linearity observed for the calibrants under the testing LC setting. In addition, the models are restricted to predict the compounds that fall inside the applicability domain and the predicted RTI values should be used with caution.

## ASSOCIATED CONTENT Supplementary Information (SI)



**SI-A:** Appendix A, contains a document discussing the description of LC conditions used in this work as well as the development and validation of all the QSRR models. **SI-B:** Appendix B contains a MS Excel file. The compounds with their corresponding RTI values and some additional information about the internal and external evaluations of RTI system are presented.

## AUTHOR INFORMATION

### Corresponding Author

**Nikolaos S. Thomaidis** – Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 157 71 Athens, Greece; Tel.: +302107274317 – Fax: +302107274750  
E-mail address: [ntho@chem.uoa.gr](mailto:ntho@chem.uoa.gr)

### Author Contributions

Conceptualization: RA and NST. Writing, editing and validation: RA, NA and ES. DSFP analysis: NA. Editing and external validation: MK, TS, MI, AM, ACH, AJW, PG, AC, CM, TY, JH, JS and NST. Supervision: NST.

## ACKNOWLEDGMENT

Authors acknowledge the contributions of Birgit Beck (Eawag) during this work. Authors also would like to acknowledge the partial financial support from the NORMAN Association. RA acknowledges the scholarship and financial support from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI Ph.D. Fellowship grant (GA. no. 14484). ELS is supported by the Luxembourg National Research Fund (FNR, Grant A18/BM/12341006).

## REFERENCES

- (1) Schymanski, E. L.; Singer, H. P.; Longrée, P.; Loos, M.; Ruff, M.; Stravs, M. A.; Ripollés Vidal, C.; Hollender, J. Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2014**, *48*, 1811-1818.
- (2) Krauss, M.; Singer, H.; Hollender, J. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* **2010**, *397*, 943-951.
- (3) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097-2098.
- (4) Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* **2015**, *49*, 12333-12341.
- (5) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51*, 11505-11512.
- (6) Guo, J.; Shi, W.; Chen, Q.; Deng, D.; Zhang, X.; Wei, S.; Yu, N.; Giesy, J. P.; Yu, H. Extended Virtual Screening Strategies To Link Antiandrogenic Activities and Detected Organic Contaminants in Soils. *Environ. Sci. Technol.* **2017**, *51*, 12528-12536.
- (7) Aalizadeh, R.; Nika, M.-C.; Thomaidis, N. S. Development and application of retention time prediction models in the suspect and nontarget screening of emerging contaminants. *J. Hazard. Mater.* **2019**, *363*, 277-285.
- (8) Aalizadeh, R.; Thomaidis, N. S.; Bletsou, A. A.; Gago-Ferrero, P. Quantitative Structure–Retention Relationship Models To Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples. *J. Chem. Inf. Model.* **2016**, *56*, 1384-1398.

- (9) Bade, R.; Bijlsma, L.; Sancho, J. V.; Hernandez, F. Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water. *Talanta* **2015**, *139*, 143-149.
- (10) Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D. K.; Fiehn, O. Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. *Anal. Chem.* **2020**, *92*, 7515-7522.
- (11) Eugster, P. J.; Boccard, J.; Debrus, B.; Breant, L.; Wolfender, J. L.; Martel, S.; Carrupt, P. A. Retention time prediction for dereplication of natural products (CxHyOz) in LC–MS metabolite profiling. *Phytochemistry* **2014**, *108*, 196-207.
- (12) Falchi, F.; Bertozzi, S. M.; Ottonello, G.; Ruda, G. F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A.; Bandiera, T.; Armirotti, A. Kernel-Based, Partial Least Squares Quantitative Structure-Retention Relationship Model for UPLC Retention Time Prediction: A Useful Tool for Metabolite Identification. *Anal. Chem.* **2016**, *88*, 9510-9517.
- (13) Chaker, J.; Gilles, E.; Léger, T.; Jégou, B.; David, A. From Metabolomics to HRMS-Based Exposomics: Adapting Peak Picking and Developing Scoring for MS1 Suspect Screening. *Anal. Chem.* **2021**, *93*, 1792-1800.
- (14) Kern, S.; Fenner, K.; Singer, H. P.; Schwarzenbach, R. P.; Hollender, J. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* **2009**, *43*, 7039-7046.
- (15) Kováts, E. *Helv. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone.* *Chim. Acta* **1958**, *41*, 1915-1932.
- (16) van Den Dool, H.; Dec. Kratz, P. A generalization of the retention index system including linear temperature programmed gas–liquid partition chromatography. *J. Chromatogr. A* **1963**, *11*, 463-471.
- (17) Lee, M. L.; Vassilaros, D. L.; White, C. M. Retention indices for programmed-temperature capillary-column gas chromatography of polycyclic aromatic hydrocarbons. *Anal. Chem.* **1979**, *51*, 768-773.
- (18) Stanstrup, J.; Neumann, S.; Vrhovsek, U. PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems. *Anal. Chem.* **2015**, *87*, 9421-9428.
- (19) Hall, L. M.; Hall, L. H.; Kertesz, T. M.; Hill, D. W.; Sharp, T. R.; Oblak, E. Z.; Dong, Y. W.; Wishart, D. S.; Chen, M. H.; Grant, D. F. Development of Ecom50 and Retention Index Models for Nontargeted Metabolomics: Identification of 1,3-Dicyclohexylurea in Human Serum by HPLC/Mass Spectrometry. *J. Chem. Inf. Model.* **2012**, *52*, 1222-1237.
- (20) Adolfo Teñe; Martí-Rosés; Bosch, E. Modeling the Retention of Neutral Compounds in Gradient Elution RP-HPLC by Means of Polarity Parameter Models. *Anal. Chem.* **2009**, *81*, 9135-9145.
- (21) Albaugh, D. R.; Hall, L. M.; Hill, D. W.; Kertesz, T. M.; Parham, M.; Hall, L. H.; Grant, D. F. Prediction of HPLC Retention Index Using Artificial Neural Networks and IGroup E-State Indices. *J. Chem. Inf. Model.* **2009**, *49*, 788-799.
- (22) Huan, T.; Wu, Y.; Tang, C.; Lin, G.; Li, L. DnsID in MyCompoundID for Rapid Identification of Dansylated Amine- and Phenol-Containing Metabolites in LC–MS-Based Metabolomics. *Anal. Chem.* **2015**, *87*, 9838-9845.
- (23) Guo, K.; Li, L. Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome. *Anal. Chem.* **2009**, *81*, 3919-3932.
- (24) Zheng, S.-J.; Liu, S.-J.; Zhu, Q.-F.; Guo, N.; Wang, Y.-L.; Yuan, B.-F.; Feng, Y.-Q. Establishment of Liquid Chromatography Retention Index Based on Chemical Labeling for Metabolomic Analysis. *Anal. Chem.* **2018**, *90*, 8412-8420.
- (25) Abate-Pella, D.; Freund, D. M.; Ma, Y.; Simón-Manso, Y.; Hollender, J.; Broeckling, C. D.; Huhman, D. V.; Krokshin, O. V.; Stoll, D. R.; Hegeman, A. D.; Kind, T.; Fiehn, O.; Schymanski, E. L.; Prenni, J. E.; Sumner, L. W.; Boswell, P. G. Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. *J. Chromatogr. A* **2015**, *1412*, 43-51.
- (26) Boswell, P. G.; Schellenberg, J. R.; Carr, P. W.; Cohen, J. D.; Hegeman, A. D. A study on retention “projection” as a supplementary

- means for compound identification by liquid chromatography–mass spectrometry capable of predicting retention with different gradients, flow rates, and instruments. *J. Chromatogr. A* **2011**, *1218*, 6732-6741.
- (27) Aichele, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Anal. Chem.* **2015**, *87*, 7698-7704.
- (28) Gorynski, K.; Bojko, B.; Nowaczyk, A.; Bucinski, A.; Pawliszyn, J.; Kalisz, R. Quantitative structure–retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: Endogenous metabolites and banned compounds. *Anal. Chim. Acta* **2013**, *797*, 13-19.
- (29) Haddad, P. R.; Taraji, M.; Szűcs, R. Prediction of Analyte Retention Time in Liquid Chromatography. *Anal. Chem.* **2021**, *93* (1), 228-256.
- (30) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762-3772.
- (31) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977-5010.
- (32) Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466-1474.
- (33) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual Computational Chemistry Laboratory – Design and Description. *J. Comput. Aided Mol. Des.* **2005**, *19*, 453-463.
- (34) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462-2474.
- (35) Partitioning(logD) Marvin 6.3.1, ChemAxon, <http://www.chemaxon.com>. Accessed date: 15-04-2018.
- (36) Xing, L.; Glen, R. C. Novel Methods for the Prediction of logP, pKa, and logD. *J. Chem. Inform. Comput. Sci.* **2002**, *42*, 796-805.
- (37) PubChem fingerprints, [ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem\\_fingerprints.txt](ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt).
- (38) Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693-705.
- (39) Todeschini, R.; Consonni, V. In *Handbook of Molecular Descriptors*; Wiley-VCH Verlag GmbH, 2008, pp 1-523.
- (40) Todeschini, R.; Gramatica, P. In *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds.; Springer Netherlands: Dordrecht, 1998, pp 355-380.
- (41) Žuvela, P.; Liu, J. J.; Macur, K.; Bączek, T. Molecular Descriptor Subset Selection in Theoretical Peptide Quantitative Structure–Retention Relationship Model Development Using Nature-Inspired Optimization Algorithms. *Anal. Chem.* **2015**, *87*, 9876-9883.
- (42) Dorigo, M.; Birattari, M.; Stützle, T. Ant colony optimization. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28-39.
- (43) Dorigo, M.; Blum, C. Ant colony optimization theory: A survey. *Theor. Comput. Sci.* **2005**, *344*, 243-278.
- (44) Moulard, Y.; Bailly-Chouriberry, L.; Boyer, S.; Garcia, P.; Popot, M.-A.; Bonnaire, Y. Use of benchtop exactive high resolution and high mass accuracy orbitrap mass spectrometer for screening in horse doping control. *Anal. Chim. Acta* **2011**, *700*, 126-136.
- (45) Beckers, L.-M.; Brack, W.; Dann, J. P.; Krauss, M.; Müller, E.; Schulze, T. Unraveling longitudinal pollution patterns of organic micropollutants in a river by non-target screening and cluster analysis. *Sci. Total Environ.* **2020**, *727*, 138388.
- (46) Fabregat-Safont, D.; Barneo-Muñoz, M.; Martínez-García, F.; Sancho, J. V.; Hernández, F.; Ibáñez, M. Proposal of 5-methoxy-N-methyl-N-isopropyltryptamine consumption biomarkers through identification of in vivo metabolites from mice. *J. Chromatogr. A* **2017**, *1508*, 95-105.
- (47) Rostkowski, P.; Haglund, P.; Aalizadeh, R.; Alygizakis, N.; Thomaidis, N.; Arandes, J. B.; Nizzetto, P. B.; Booij, P.; Budzinski, H.; Brunswick, P.; Covaci, A.; Gallampos, C.; Grosse, S.; Hindle, R.; Ipolyi, I.; Jobst, K.; Kaserzon, S. L.; Leonards, P.; Lestremou, F.; Letzel, T., et al. The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques. *Anal. Bioanal. Chem.* **2019**, *411*, 1957-1977.
- (48) Gago-Ferrero, P.; Krettek, A.; Fischer, S.; Wiberg, K.; Ahrens, L. Suspect Screening and Regulatory Databases: A Powerful Combination to Identify Emerging Micropollutants. *Environ. Sci. Technol.* **2018**, *52*, 6881-6894.
- (49) Moschet, C.; Anumol, T.; Lew, B. M.; Bennett, D. H.; Young, T. M. Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. *Environ. Sci. Technol.* **2018**, *52*, 2878-2887.
- (50) Ulrich, E. M.; Sobus, J. R.; Grulke, C. M.; Richard, A. M.; Newton, S. R.; Strynar, M. J.; Mansouri, K.; Williams, A. J. EPA's nontargeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. *Anal. Bioanal. Chem.* **2019**, *411*, 853-866.
- (51) Dodge, Y. In *The Concise Encyclopedia of Statistics*; Springer New York: New York, NY, 2008, pp 302-304.
- (52) Pizarro, J. n.; Guerrero, E.; Galindo, P. L. Multiple comparison procedures applied to model selection. *Neurocomputing* **2002**, *48*, 155-173.
- (53) Hartmann, C.; Smeyers-Verbeke, J.; Penninckx, W.; Vander Heyden, Y.; Vankeerberghen, P.; Massart, D. L. Reappraisal of Hypothesis Testing for Method Validation: Detection of Systematic Error by Comparing the Means of Two Methods or of Two Laboratories. *Anal. Chem.* **1995**, *67*, 4491-4499.
- (54) Aalizadeh, R.; von der Ohe, P. C.; Thomaidis, N. S. Prediction of acute toxicity of emerging contaminants on the water flea *Daphnia magna* by Ant Colony Optimization–Support Vector Machine QSTR models. *Environ. Sci.: Process. Impacts* **2017**, *19*, 438-448.
- (55) Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibáñez, M.; Goshawk, J.; Barknowitz, G.; Hernández, F.; Bijlsma, L. Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environ. Sci. Technol.* **2020**, *54*, 15120-15131.
- (56) Alygizakis, N. A.; Oswald, P.; Thomaidis, N. S.; Schymanski, E. L.; Aalizadeh, R.; Schulze, T.; Oswaldova, M.; Slobodnik, J. NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in “digitally frozen” environmental samples. *TrAC Trends Anal. Chem.* **2019**, *115*, 129-137.
- (57) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **2020**, *49*, D1388-D1395.
- (58) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminform.* **2016**, *8*, 3.
- (59) Stanton, D. T.; Jurs, P. C. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Anal. Chem.* **1990**, *62*, 2323-2329.
- (60) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminform.* **2017**, *9*, 61.
- (61) Schymanski, E. L.; Kondić, T.; Neumann, S.; Thiessen, P. A.; Zhang, J.; Bolton, E. E. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *J. Cheminform.* **2021**, *13*, 19.
- (62) Dulio, V.; Koschorreck, J.; van Bavel, B.; van den Brink, P.; Hollender, J.; Munthe, J.; Schlabach, M.; Aalizadeh, R.; Agerstrand, M.; Ahrens, L.; Allan, I.; Alygizakis, N.; Barcelo, D.; Bohlin-Nizzetto, P.; Boutrou, S.; Brack, W.; Bressy, A.; Christensen, J. H.; Cirka, L.; Co-

vaci, A., et al. The NORMAN Association and the European Partnership for Chemicals Risk Assessment (PARC): let's cooperate! *Environ. Sci. Eur.* **2020**, *32*, 100.

(63) Network, N.; Aalizadeh, R.; Alygizakis, N.; Schymanski, E.; Slobodnik, J.; Fischer, S.; Cirka, L., Exchange, N. S. L., Ed., 2020. DOI: 10.5281/zenodo.2664077

For TOC only

