Supplementary Information (SI-A)

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

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

[a] Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 157 71 Athens, Greece,

[b] Environmental Institute, Okružná 784/42, 97241 Koš, Slovak Republic

[c] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, L-4367 Belvaux, Luxembourg

[d] Eawag: Swiss Federal Institute for Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland.

[e] Department Effect-Directed Analysis, Helmholtz-Centre for Environmental Research – UFZ, Leipzig, Germany

[f] Research Institute for Pesticides and Water, University Jaume I, Castellón 12071, Spain

[g] 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

[h] 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

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

[j] Toxicological Center, University of Antwerp, 2610 Wilrijk, Belgium

[k] Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA

[l] Institute of Biogeochemistry and Pollutant Dynamics, IBP, ETH Zurich, 8092 Zurich, Switzerland

Table of content

All the supplementary Tables can be found in the Excel file accompanied by the manuscript (SI-B).

SI 1. Materials and Method

SI 1.1. Instrumentation

In RPLC, the chromatographic separation was performed on an Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 μm) from Thermo Fisher Scientific (Driesch, Germany) preceded by a guard column, ACQUITY UPLC BEH C18 1.7 μm, VanGuard Pre-Column, Waters (Ireland), thermostated at 30 ˚C. Mobile phase composition in positive ionization mode (PI) is (A) H_2O:MeOH (90:10) with 5 mM ammonium formate and 0.01% formic acid and (B) MeOH with 5 mM ammonium formate and 0.01% formic acid. For the negative ionization mode (NI), the mobile phase is (A) H_2O:MeOH (90:10) with 5 mM ammonium acetate and (B) MeOH with 5 mM ammonium acetate. The gradient elution program was the same for the two ionization modes and the chromatogram lasts 15.5 min, with 5 min of re-equilibration of the column for the next injection. It starts with 1% B with a flow rate of 0.2 mL min^{-1} for 1 min and it increases to 39 % in 2 min (flow rate 0.2 mL min^{-1}), and then to 99.9 % (flow rate 0.4 mLmin^{-1}) in the following 11 min. Then, it keeps constant for 2 min (flow rate 0.48 mL min^{-1}) and then, initial conditions were restored within 0.1 min and the flow rate decreased to 0.2 mL min^{-1}. The injection volume was set up to 5 μL.

The operating parameters of the electrospray ionization interface (ESI) are for PI mode: capillary voltage, 2500 V; end plate offset, 500 V; nebulizer, 2 bar; drying gas, 8 L min^{-1}; dry temperature, 200 °C; and for NI mode: capillary voltage, 3500 V; end plate offset, 500 V; nebulizer, 2 bar; drying gas, 8 L min^{-1}; dry temperature, 200 °C.

A QToF external calibration was performed daily with a sodium formate solution, and a segment (0.1−0.25 min) in every chromatogram was used for internal calibration, using a calibrant injection at the beginning of each run. The sodium formate calibration mixture consists of 10 mM sodium formate in a mixture of water:isopropanol (1:1). The theoretical exact masses of calibration ions in the range of 50−1000 Da were used for calibration. The instrument provided a typical resolving power of 36000−40000 during calibration (39274 at m/z 226.1593, 36923 at m/z 430.9137, and 36274 at m/z 702.8636).

SI 1.2. Ant Colony Optimization (ACO)

ACO is a swarm intelligence algorithm that is based on the behavior of the ants searching for the food resources by their nest using pheromone deposition without any visual information [1,2]. This enables ants to be adoptable to the environmental changes and by introduction of any changes, they can find a new shortest path to the resources [2]. ACO is preferably a good method to handle optimization or features selection related problems since ants can give the best combination of subsets that has the maximum fitness objective (here is overlap of normal distribution (objective function) between calibrants and the rest of the compounds). Ants solve complex optimization or feature selections problems using an artificial pheromone deposition. Here we tried to select compounds instead of features. For ACO based compounds selection case, the algorithm starts with the generation of certain number of the ants (here we set this at 300 ants) placed randomly on the graph representing the starting compound and with extension the best combinations of

compounds. Thus, each node (in a graph) relates to a compound and each edge shows the traversal of an ant from one compound to another. The number of artificial pheromone [0, 1] for an edge is associated with the popularity of the particular traversal by pervious ants. Therefore, ants could make probabilistic decisions to stay at which node and select which edge based the artificial pheromone and related traversal degree. This will continue until the maximum degree for the objective function has reached otherwise the information in each edge will be updated and a new set of ants will be created and all process will be iterated one more [1,2]. Here we also set the maximum number of iteration to 200 and desired number of calibrants started with 5, 8, 10, 12, 15, 18, 20, 22 and 25 compounds. Evaporation Rate (ER) was also set to 0.05 (this value is being kept constant during performing ACO and generally is small value (0.01-0.05)) [2]. ER causes uniformly decreasing all the pheromone values. From a practical point of view, pheromone evaporation is required to prevent a too rapid convergence of the algorithm toward a sub-optimal space. It presents a useful form of forgetting and cause exploration of new areas in the search space. ACO algorithm was written and performed in MATLAB.

SI 1.3. Intralaboratory validation

SI 1. 3.1. Four different LC conditions used for the RTI system developed in ± ESI

   LC condition 1: the chromatographic separation was done on an Atlantis T3 C18 (2.1×100 mm, 3.0 μm) from Waters (Ireland) preceded by a guard column, Acquity UPLC BEH C18 1.7 μm, VanGuard Pre-Column, Waters (Ireland), thermostated at 30 ˚C. Mobile phase composition in positive ionization mode (PI) is (A) H_2O:MeOH (90:10) with 5 mM ammonium formate and 0.01% formic acid and (B) MeOH with 5 mM ammonium formate and 0.01% formic acid. For the negative ionization mode (NI), the mobile phase is (A) H_2O:MeOH (90:10) with 5 mM ammonium acetate and (B) MeOH with 5 mM ammonium acetate. The gradient elution program was the same for the two ionization modes and the chromatogram lasts 15.5 min, with 5 min of re-equilibration of the column for the next injection. It starts with 1% B with a flow rate of 0.2 mL min^{-1} for 1 min and it increases to 39% in 2 min (flow rate 0.2 mL min^{-1}), and then to 99.9% (flow rate 0.4 mLmin^{-1}) in the following 11 min. Then, it keeps constant for 2 min (flow rate 0.48 mL min^{-1}) and then, initial conditions were restored within 0.1 min and the flow rate decreased to 0.2 mL min^{-1}. The injection volume was set up to 5 μL.

   LC condition 2: the chromatographic separation was done on an Acclaim RSLC C18 column (2.1 × 100 mm, 2.2 μm) from Thermo Fisher Scientific (Driesch, Germany) preceded by a guard column, Acquity UPLC BEH C18 1.7 μm, VanGuard Pre-Column, Waters (Ireland), maintained at 30 ˚C. Mobile phase composition in positive ionization mode (PI) is (A) H_2O with 0.1% formic acid and (B) MeOH with 0.1% formic acid. For the negative ionization mode (NI), the mobile phase is (A) H_2O with 5 mM ammonium acetate and (B) MeOH with 5 mM ammonium acetate. The gradient elution program was the same for the two ionization modes and the chromatogram lasts 25 min, with 5 min of re-equilibration of the column for the next injection. It starts with 10% B for and it increases to 50% in 4 min, and then to 95.0% in 17 min. Then, it keeps constant until 25 min and then, initial conditions were restored within 0.1 min and kept running until 30 min for

re-equilibration of the column. The flow rate was 0.2 mL min^{-1} throughout the run time. The injection volume was set up to 5 µL.

LC condition 3: the chromatographic separation was done on an Acquity UPLC BEH C18 (2.1×100 mm, 1.7 µm) (for +ESI) and XBridge C18 (2.1×150 mm, 2.5 µm) (for -ESI) from Waters (Ireland), preceded by a guard column, Acquity UPLC BEH C18 1.7 µm, VanGuard Pre-Column, Waters (Ireland), maintained at 30 ˚C. Mobile phase composition in positive ionization mode (PI) is (A) H$_2$O with 0.1% formic acid and (B) MeOH with 0.1% formic acid. For the negative ionization mode (NI), the mobile phase is (A) H$_2$O with 5 mM ammonium acetate and (B) MeOH with 5 mM ammonium acetate. The gradient elution program was the same for the two ionization modes and the chromatogram lasts 25 min, with 5 min of re-equilibration of the column for the next injection. It starts with 10% B for and it increases to 50% in 4 min, and then to 95.0% in 17 min. Then, it keeps constant until 25 min and then, initial conditions were restored within 0.1 min and kept running until 30 min for re-equilibration of the column. The flow rate was 0.2 mL min^{-1} throughout the run time. The injection volume was set up to 5 µL.

LC condition 4: the chromatographic separation was done on an Acquity UPLC BEH C18 (2.1×100 mm, 1.7 µm) (for +ESI) and XBridge C18 (2.1×150 mm, 2.5 µm) (for -ESI) from Waters (Ireland), preceded by a guard column, Acquity UPLC BEH C18 1.7 µm, VanGuard Pre-Column, Waters (Ireland), maintained at 30 ˚C. Mobile phase composition in positive ionization mode (PI) is (A) H$_2$O with 0.1% formic acid and (B) ACN with 0.1% formic acid. For the negative ionization mode (NI), the mobile phase is (A) H$_2$O with 5 mM ammonium acetate and (B) ACN. The gradient elution program was the same for the two ionization modes and the chromatogram lasts 25 min, with 5 min of re-equilibration of the column for the next injection. It starts with 20% B for and it increases to 50% in 5 min, and then to 100.0% in 25 min. Then, the initial conditions were restored within 0.2 min and kept running until 30 min for re-equilibration of the column. The flow rate was 0.3 mL min^{-1} throughout the run time. The injection volume was set up to 5 µL.

SI 1.3.2. Interlaboratory validation

The Helmholtz Centre for Environmental Research (UFZ) evaluated the RTI system performed with LCHRMS using an Agilent 1200 series LC system coupled to a hybrid linear ion trap Orbitrap MS (LTQ Orbitrap XL, Thermo Scientific) with an electrospray ionization (ESI) source measuring in both positive and negative modes. Kinetex Core-Shell C18 column (3.0 mm × 100 mm, 2.6 µm; Phenomenex) was also used for LC separation. A gradient elution was carried out with a flow rate of 0.2 mL min^{-1} with water (A) and methanol (B) both containing 0.1% of formic acid. The initial content of mobile phases was 95:5 (A:B) in the beginning and held the same until 1 min, then, it changed to 100% B within a linear gradient at 13 min. A was maintained at 95% for 24 min followed by a re-equilibration for 9 min. The injection volume was 10 µL and the column temperature was 22°C. Full scan MS detection was performed with an LTQ Orbitrap XL (resolution R = 60000 at m/z 400, for m/z = 100 to 1000) from Thermo Fisher Scientific (San Jose,

U.S.) with electrospray ionization (ESI) in positive and negative mode, with a spray voltage of +5.0 and −3.5 kV, respectively, and a capillary temperature of 275 °C.

Swiss Federal Institute of Aquatic Science and Technology (Eawag) performed the evaluation of the RTI system with the following LC system; a PAL Autosampler (CTC Analytics, Zwingen, Switzerland), a Rheos 2200 quaternary low-pressure mixing pump (Flux Instruments, Basel, Switzerland), and an XBridge C18 column (2.1 × 50 mm, 3.5 μm) from Waters (Milford, U.S.) with a 2.1 × 10 mm precolumn of the same material. The gradient (water/methanol, both with 0.1% formic acid) was 90:10 at 0 min, to 50:50 at 4 min, to 5:95 at 17 min, held until 25 min then 90:10 at 25.1 to 30 min at a flow of 0.2 mL min^{-1} and a column temperature of 30 °C. Full scan MS detection was performed with a QExactive Orbitrap XL (resolution R = 60000 at m/z 400, for m/z = 100 to 1000) from Thermo Fisher Scientific (San Jose, U.S.) with electrospray ionization (ESI) in positive and negative mode, with a spray voltage of +4 and −4 kV, respectively, and a capillary temperature of 300 °C.

University Jaume I (UJI) tested RTI system using a Waters Acquity UPLC system (Waters, Milford, MA, USA) interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (XEVO G2 QTOF, Waters Micromass, Manchester, UK). This system was using an orthogonal Z-spray-ESI interface operating in both ± ESI. The chromatographic separation was performed using a Waters Cortecs C18 analytical column (2.1 i.d. × 100 mm length, 2.7μm particle size) at a flow rate of 0.3 mL min^{-1}. The mobile phases used were (A) H2O with 0.01% HCOOH and (B) MeOH with 0.01% HCOOH. The initial percentage of B was 10%, which was linearly increased to 90% in 14 min, followed by a 2 min isocratic period and, then, returned to initial conditions during 2 min. The column temperature was set to 40ºC and the injection volume was 25 μL. MS data were acquired over an m/z range of 50–1000. Capillary voltages of 0.7 and 3.0 kV were used in ± ESI, respectively. A cone voltage of 20 V was used for both ESI modes. Collision gas was argon 99.995% (Praxair, Valencia, Spain). The desolvation temperature was set to 600ºC and the source temperature to 130 ºC.

The United States Environmental Protection Agency (USEPA) performed analysis on the RTI system using an Agilent 1290 Infinity II LC system coupled with an Agilent 6530B Accurate-Mass QTOF/MS utilizing a Dual AJS ionization source (Santa Clara, CA). Chromatographic separation was performed using a Waters Acquity BEH C18 analytical column (2.1 × 50 mm, 1.7 μm). A gradient elution was carried out with a flow rate of 0.2 mL/min with water (A) and acetonitrile (B) both containing 0.1% of formic acid. Mobile phase B was held at 10% for 2 min, then linearly ramped to 100% over 13 min, held for 5 min, and then linearly ramped down to initial conditions for 1 min followed by a 9 min re-equilibration. The column temperature was set to 25ºC and the injection volume was 10 and 15 μL for positive and negative mode, respectively. MS data were acquired with positive and negative electrospray ionization in full scan mode (50 − 1000 m/z) with spray voltages of 3500 V and -3500 V, respectively. Source gas was set to 300ºC and sheath gas was set to 350ºC. Reference solution consisting of purine, HP-0921, and trifluoroacetic acid were directly infused into the electrospray source for correction of mass shift during the run.

Swedish University of Agricultural Sciences (SLU) evaluated the RTI system using an Acquity Ultra-Performance Liquid Chromatography (UPLC) system (Waters Corporation, USA) coupled to a QTOF mass spectrometer (QTOF Xevo G2S, Waters Corporation, Manchester, UK). The chromatographic separation was carried out on an Acquity HSS T3 column (2.1 mm × 100 m, 1.8 μm) in positive ionization (PI) mode. The aqueous phase consisted of 5 mM ammonium formate buffer with 0.01% formic acid and the organic phase with acetonitrile and 0.01% formic acid. The adopted elution gradient for both ionization modes started with 5% of organic phase for 0.5 minutes, increasing to 95% by 16 min, and then to 99% in the following 0.1 min. These almost pure organic conditions were kept constant for 3 min, and then initial conditions were restored and kept for 2 min. The chromatographic flow rate was 0.5 mL min^{-1} and the injection volume was 5 μL. The column temperature was set to 40 °C. The resolution of the TOF mass spectrometer was 30000 at full width and half maximum (FWHM) at m/z 556. MS data were acquired over an m/z range of 100-1200 in a scan time of 0.25s. Capillary voltages of 0.35 kV were used in PI. A cone voltage of 30 V was applied, the desolvation gas flow rate was set at 700 L h^{-1} and the cone gas flow was set to 25 L h^{-1}. The desolvation temperature was set to 450 °C and the source temperature to 120 °C. More details can be found elsewhere. [3]

The University of Antwerp tested the RTI system using an Agilent 1290 Infinity LC system (Agilent Technologies) interfaced to an Agilent 6530 Q-TOF mass spectrometer. This system was using an ESI interface operating in both ± ESI. The chromatographic separation was performed using a Kinetex Biphenyl analytical column (2.1 × 100 mm, 2.6 μm, from Phenomenex) at a flow rate of 0.400 mL min^{-1}. The mobile phases used were (A) H_2O with 0.04% HCOOH and (B) MeOH with 0.04% HCOOH. The initial percentage of B was 2%, held for 2 min, then linearly increased to 40% in 16 min, then linearly increased to 100% in 7.5 min followed by a 4.5 min isocratic period and, then, returned to initial conditions during 10 min. The column temperature was set to 30ºC and the injection volume was 10 μL. MS data were acquired over an m/z range of 50–1000. Capillary voltages of 2.0 kV were used in ± ESI, respectively. Drying gas temperature and flow rate were 250 ˚C and 8 L min^{-1}. Sheet gas was 350 ˚C at 11 L min^{-1}. Capillary and fragmentor voltages were 2,000 V and 150 V, respectively. The nebulizer was set at 45 psi. Collision gas was nitrogen 99.995% (Air Liquide, Liège, Belgium).

The University of California, Davis, Department of Civil and Environmental Engineering (UCD-CEE) performed the evaluation of the RTI system with the following LC condition; an Agilent 1260 Infinity liquid chromatograph with autosampler (Agilent Technologies, Inc.., Santa Clara, CA, USA), and a Zorbax Eclipse Plus column (100 mm length, 2.5 mm ID, 1.8 μm particle size). A gradient elution was carried out with a flow rate of 0.35 mL min^{-1} with water (A) and acetonitrile (B); in +ESI both solvents contained 0.1% formic acid, in -ESI A contained 1 mM ammonium fluoride. The initial content of mobile phases was 98:2 (A:B) held for 1.5 min, changed to 100% B with a linear gradient at 16.5 min. B was maintained at 100% for 5 min followed by a re-equilibration at the initial conditions for 3 min. The injection volume was 10 μL and the column temperature was 30°C. Full scan MS detection was performed with an Agilent 6530 QTOF-MS

(resolution R = 20000 at m/z 1522, for m/z = 50 to 1050) with electrospray ionization in positive and negative modes (±ESI).

SI 1.4. QSRR workflows

After transforming the t_R to RTI linearly through the equation derived based on the calibration data of 18 internal calibrants for each ESI platform, geometries of chemical structures of all compounds were optimized using Balloon [4]. These 3D structures were obtained out of various tautomer forms (the tautomer with the lowest energy was retained to get one structure out of different forms of a duplicate entries) using balloon [5]. These molecular properties were calculated based on the final 3D optimized structures using Padel [6]. In addition, ChemAxon [7] was used to calculate logD at pH of mobile phase used in LC **SI1** which was 3.6 (for positive ESI) and 6.2 (for negative ESI)). The dataset including the molecular features with experimental t_R generated for each platform was pre-treated by removing the constant and near constant features and further checked for existence of collinearity. The remaining molecular features were split into training and test set using affinity propagation method. [8,9] Here, affinity propagation was used as input measures of similarity between pairs of compounds. Real-valued messages were exchanged between compounds until a high-quality set of exemplars and related clusters gradually derived. ACO was used to select the most relevant descriptors that correlate to the RTI. The descriptors selected were linearly and non-linearly correlated RTI by Multiple Linear Regressions (MLR) and Support Vector Machines (SVM) [9,10]. The initial accuracy of the models built to predict t_R were carefully studied using external test set, and cross-validation techniques. The source codes and functions to implement ACO, affinity propagation, MLR and SVM are available in MATLAB. These steps are depicted in **Scheme S1**.

QSRR Workflow
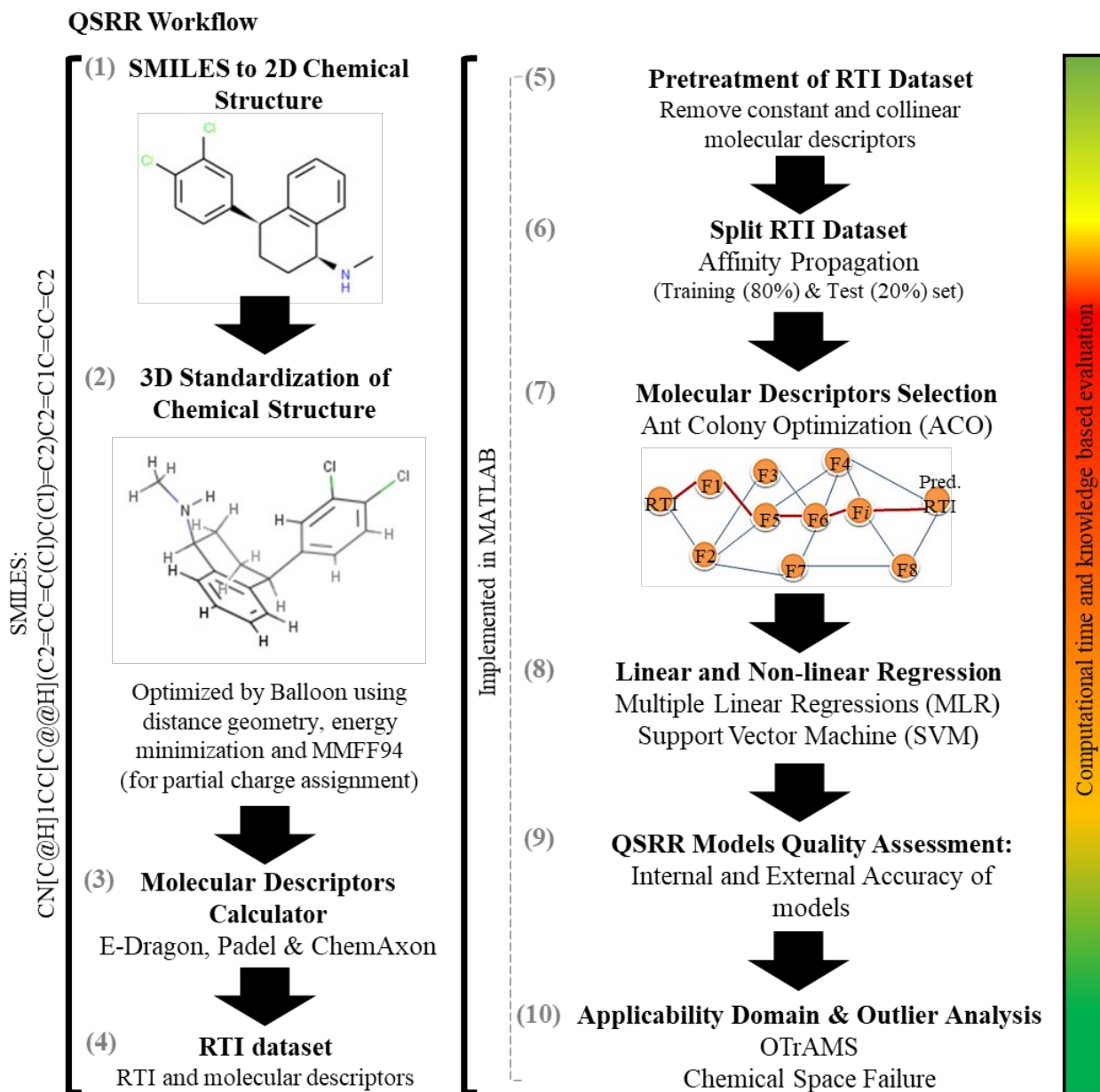
Scheme S1. The modelling workflow for predicting RTI values

SI 1.4.1. Regression Analysis

Multiple linear regression (MLR) is used to develop a simple and linear model to understand mechanism of actions (MOA) (molecular descriptors correlation with RTI values) associated with RTI. To derive a MLR model, the number of compounds in the training set should be five times higher than the number of selected molecular descriptors (the descriptors should be orthogonal). Lower number of molecular descriptors can help minimize the information overlap in the molecular descriptors (intercorrelation). The MLR model provides a linear equation which is combining the physicochemical properties of a chemical structural to estimate and interpret the RTI value.

$$RTI = a_0 + b_1x_1 + \dots + b_nx_n \qquad\qquad (Eq.S1)$$

where a_0 is the intercept and the b_i is regression coefficients of a selected molecular descriptors x_i. In addition to MLR, Support Vector Machines (SVM) are used for building nonlinear models for RTI values. This technique is based on the structure risk minimization (SRM) principle by introducing *<epsilon>* (ε) *insensitive loss function* to solve a regression problem. This constitutes a trade-off between the complexity of a QSRR model and its ability to estimate the RTI values closer to its experimental values. The accuracy of a QSRR model based on SVM depends mainly on a good selection of three internal parameters in SVM: C, ε, the kernel type, and corresponding kernel parameters. The Parameter C is a regularization constant calculating the trade-off between the model complexity and the degree in which deviations larger than ε are tolerated in optimization formulation. The selection of the kernel function and its related parameters are one of the important steps in building a SVM model, because it can affect the hyperspace of SVM (transformation of data). In this study, the radial basis function (RBF) was used as kernel function in which the most significant parameter is the width (γ) of the radial basis function. The radial basis function (RBF) is defined as follows:

$$k(\bar{x}_i, \bar{x}_j) = \exp\left(-\gamma\|\bar{x}_i - \bar{x}_j\|^2\right) \qquad\qquad (Eq.S2)$$

where k refers to the kernel function and γ is a parameter of kernel, $\bar{x}_i$ and $\bar{x}_j$ are independent variables. All calculations regarding implementation of SVM and its analyses are done within MATLAB.

SI 1.4.2. QSRR models internal and external validation

Here only three main parameters that are first indication of an acceptable QSRR model are discussed in details (**Table 1**). The full details about the validation criteria for a QSRR-based retention time model can be found in our pervious works. [9,10] R^2 value calculates the proportion of the variation in the RTI values and it is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{I}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{I}(y_i - \bar{y})^2} \qquad\qquad (Eq.S3)$$

where y_i is the experimental RTI value, $\bar{y}$ is the mean value of the experimental RTI values and $\hat{y}_i$ is the predicted RTI value. A R^2 value which is higher than 0.5 and near 1.0 is the first indication to accept the predictive ability of the model. The error of the RTI models is given by the root mean square error (RMSE). A lower RMSE value is one of the parameters to be used to select a QSRR model and it indicates that a model can be accepted for predicting RTI values. The RMSE value is calculated as below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad\qquad (Eq.S4)$$

n is the number of compounds in the dataset. The most important statistical parameter used in the internal validation of MLR and SVM models is the leave-one-out cross-validation correlation coefficient denoted as Q_{LOO}^2. It is calculated by excluding one of the compounds in the training set

and by rebuilding the model. Afterwards, the RTI value of the excluded compound is predicted from the newly rebuilt model. This process is continued until all the compounds in the training set are excluded once, and their RTI values are predicted by the rebuilt model. This technique is a good indicator of the strength of the derived models and shows the dependency of a QSRR model on each individual compound in the training set. The acceptable threshold for Q_{LOO}^2 value is above 0.7. Q_{LOO}^2 value is calculated similar to *Eq. S3* after exclusion of all compounds from dataset for once.

SI 1.5. Applicability domain study

The applicability domain is part of any QSRR workflow to define the chemical space that a model is capable of covering [11]. The Williams plot is considered to be a robust method to define the boundaries of the models [12]. It is based on the leverage and on standardized residual values that the leverages are being estimated from the molecular descriptors used to build the QSRR model and can be calculated as follows:

$$h_i = x_i^T (X^T X)^{-1} x_i \qquad where \qquad h^* = 3(p+1)/n \qquad\qquad (Eq. S5)$$

where X is the matrix of molecular descriptors, T is the indicator of the training set, x_i is the descriptor vector of each compound, n is the number of training set compounds, p is the number of molecular descriptors used as modeling variables, and h* is the warning leverage value. The chemical structures exceeding this threshold are outliers due to their high chemical structures dissimilarity [12]. The commonly used cut-off value for standardized residuals is ±3δ that covers 99% of the normally distributed data. Compounds which fall outside of this threshold would be considered outliers due to the abnormal t_R/RTI observed, however, compounds outside of the leverage cut-off value, but inside the standardized residual limits are considered as good leverages.

In addition to this method, Euclidean distance can be measured for training and test set, and then the mean distance for the test set compounds can be normalized based on mean distance of training set *versus* observed t_R/RTI. This shows how the diversity of chemical structures behaves toward the target t_R/RTI [13]. A test set compound outside the cut-off value of 1.0 (derived from normalization of mean distance of the training set), are considered to be outside of applicability domain of the model, and the training set is not representative for this test set compound.

In our previous work concerning the uncertainty in prediction results via QSRR method, we coupled leverage, standardized residuals (relative t_R error window) and normalized mean distance in a single 3D *bubble plot* so called OTrAMS [10]. The general idea behind the OTrAMS method was to define the error window for prediction results of t_R relatively by considering their chemical similarity effect [10]. For instance, for a compound with a high chemical similarity comparing to the training set used to model t_R/RTI, the error of (or higher than) ±2 min would be relatively high value, while for a compound with low chemical structure similarity, this threshold can be applied. Another application aspect was to define the chemical space boundaries and to define whether the observed residual is due to the chemical space failure (leading to a wrong prediction of t_R and RTI) or the structure tested has abnormal elution behavior. In this plot, the bubbles size is proportional to the leverage values and the larger it gets the more dissimilar the chemical structure of suspected

compound becomes. Standardized residuals (SR) were also used to code the compounds based on their relative residual with color. Four regions are obvious in this plot corresponding to different levels of acceptance for the predicted t_R or RTI. Box 1 (very similar to the training set and the error (in terms of SR) is less than 1.0 SR) and box 2 (the structure is diverse and the observed error is mostly due to the chemical dissimilarity, in this case, the error (in terms of SR) is less than 2.0) are the areas of acceptance of predicted t_R or RTI. However, box 3 shows the region where the SRs are high and the predicted t_R or RTI are questioned. The last region (box 4) shows the area where the model is not applicable for a given compound when the bubble size is huge otherwise the compound is false positive (very small bubble size) and it should not be corresponded to the observed t_R or RTI. Moreover, normalized mean distance shows if the used training set is representative for any suspected compounds tested. This helps to distinguish between compounds that are of good leverage and those that are outside of applicability domain.

For a compound without an experimental t_R/RTI measured, the only way to address the applicability domain of the QSRR based models 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 [5]. This method offers several conditions where chemical space failure of a compound with unknown t_R or RTI could be addressed: (a) The chemical space failure zone is the area above the normalized mean distance (cut off value of 1.0) and warning leverage cut off. Any compounds found to be in this area are outside of the applicability domain and the models should not be used to predict t_R or RTI; (b) The safe zone is the area where compounds are within the warning leverage and normalized mean distance limit. These predictions are accepted because they are highly similar to the compounds in the training set and thus the error is not function of chemical space failure; (c) Last but not least, for the compounds exceeding the warning leverage cut-off limit, while falling inside the limit of normalized mean distance and the maximum leverage value (calculated from the training set), the prediction results are less reliable and should be verified experimentally.

SI 2. Results and Discussion
SI 2.1. Selection of RTI calibrants
Figure S1 shows the final overlap that is achieved based on the set of 18 compounds as RTI calibrant. The overlap of normal distribution between calibrants and rest of dataset are shown separately for experimental t_R and chemical space. **Figure S2** shows various calibrants and the corresponding errors observed for various sets of 5 to 25 compounds as evaluation set. The lowest distribution of error (between ±1) is derived using 18 calibrants.
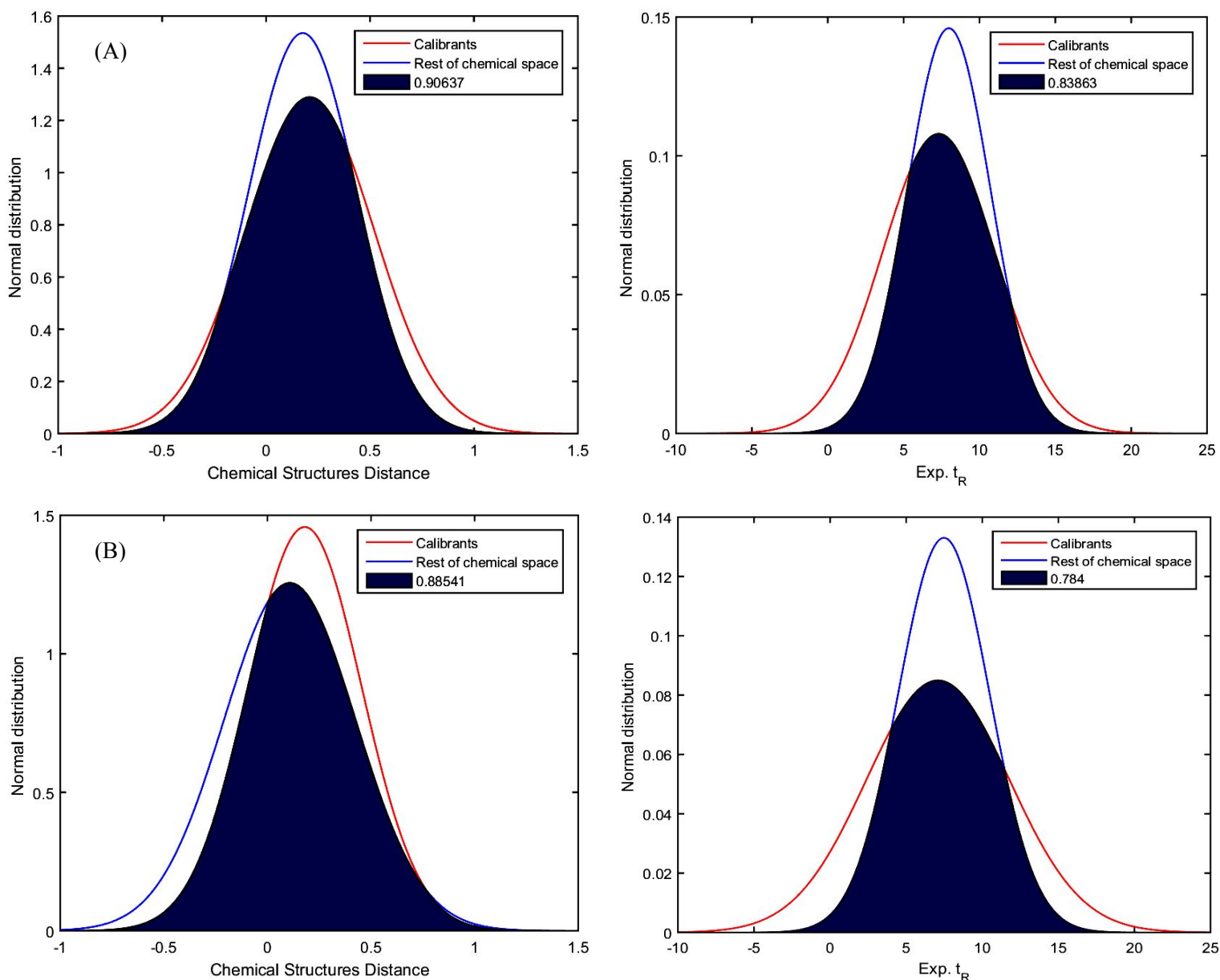
Figure S1. Overlap of normal distribution of chemical space and t_R between selected 18 calibrants and rest of compounds; (A) for negative and (B) for positive ESI
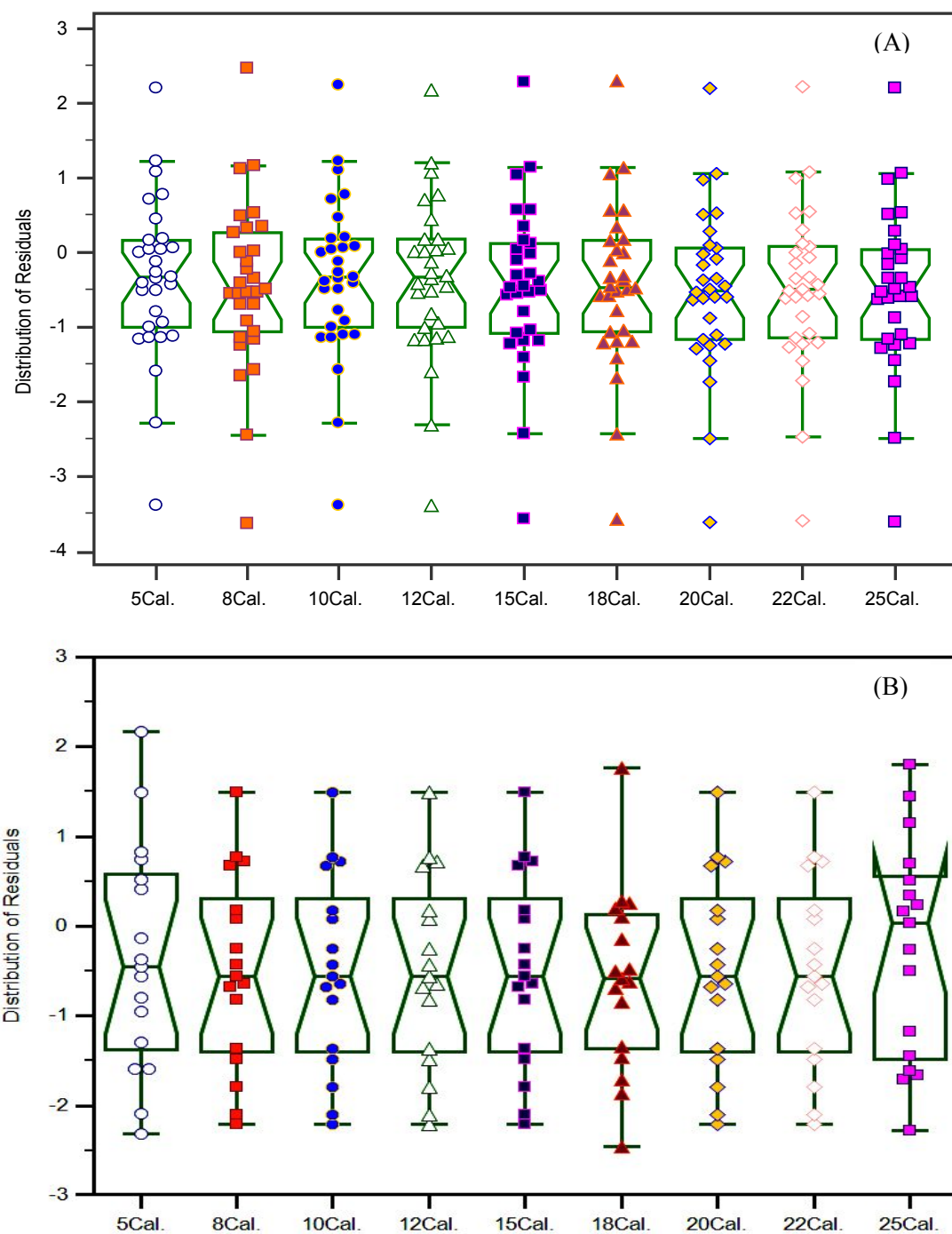
Figure S2. Distribution of errors observed by different combinations of calibrants used as validation set; (A) for positive and (B) for negative ESI

SI 2.2. RTI Stability Test

The stability test of each individual calibrant in the prepared mixture is provided in **Figure S3 (SI-A)**. Stability test was performed by analyzing each mixture at 0, 6, 18, 24, 36 and 48 hours of storage time at two different temperature (-18 and +2 °C). The mixtures were returned to refrigerator and freezer after each analysis time point and stored for the next injection. The RTI calibrants are stable within 48 hours of consecutive analysis whereas in worst case, having dropped in their response factor to 60% for some individual calibrants. This drop is most probably due to putting the RTI mix outside of storage condition and in lab environment before analysis, remaining in the tray during analysis for each time point.

RTI calibrants for -ESI:

Coumaphos

- Stored at -18°C
- Stored at +2°C

Triclosan

- Stored at -18°C
- Stored at +2°C

Avermectin B1a

- Stored at -18°C
- Stored at +2°C

Salinomycin

- Stored at -18°C
- Stored at +2°C

RTI calibrants for +ESI:

Guanylurea

- Stored at -18°C
- Stored at +2°C

Amitrole

- Stored at -18°C
- Stored at +2°C

Histamine

- Stored at -18°C
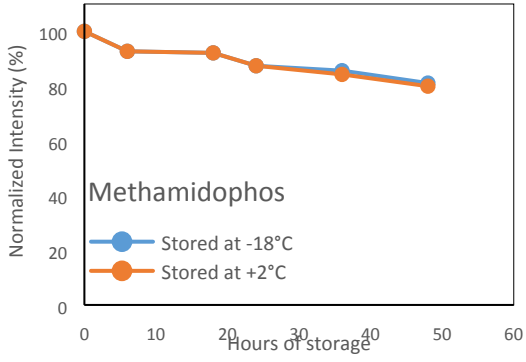- Stored at +2°C

Chlormequate

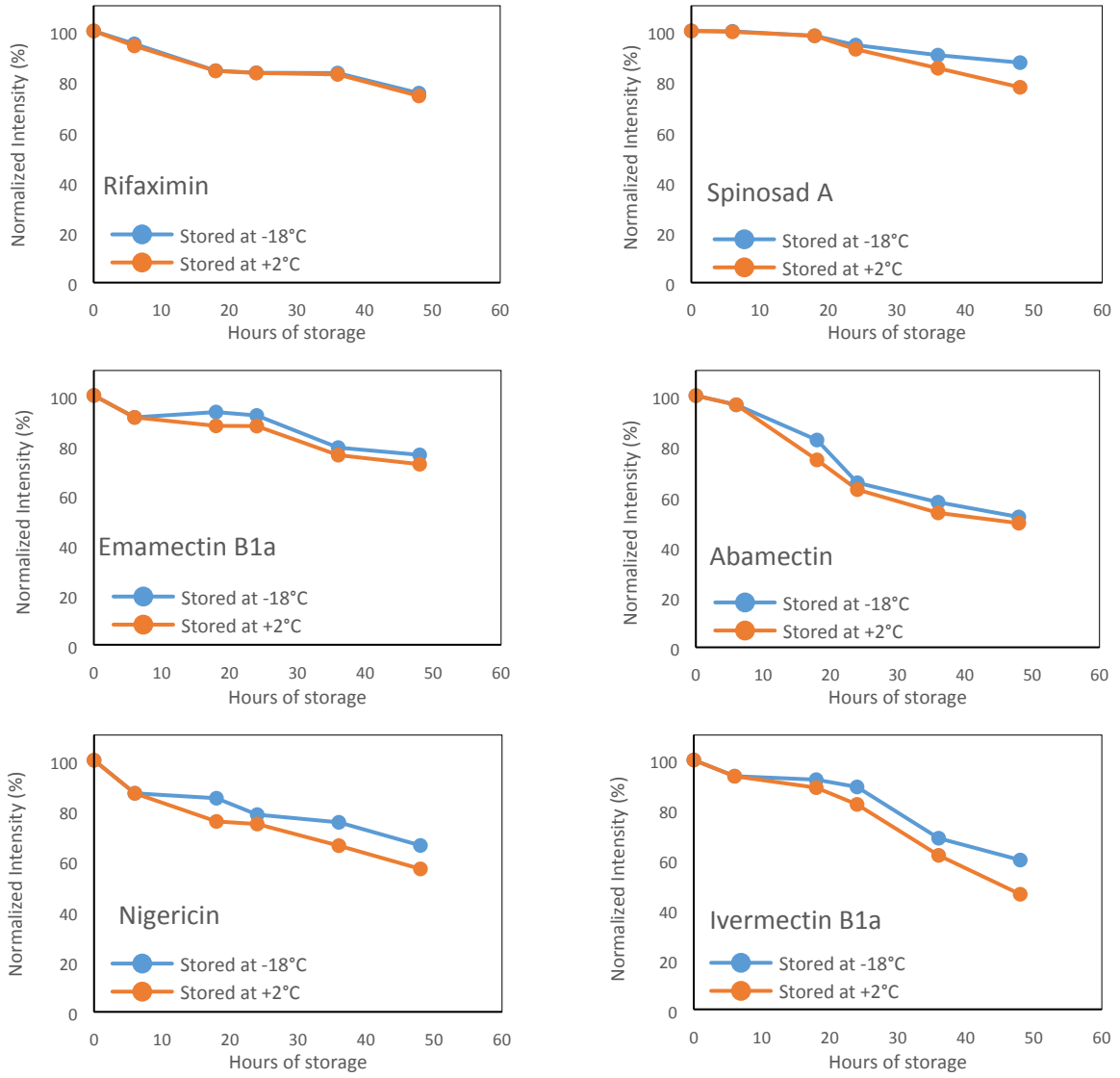- Stored at -18°C
- Stored at +2°C
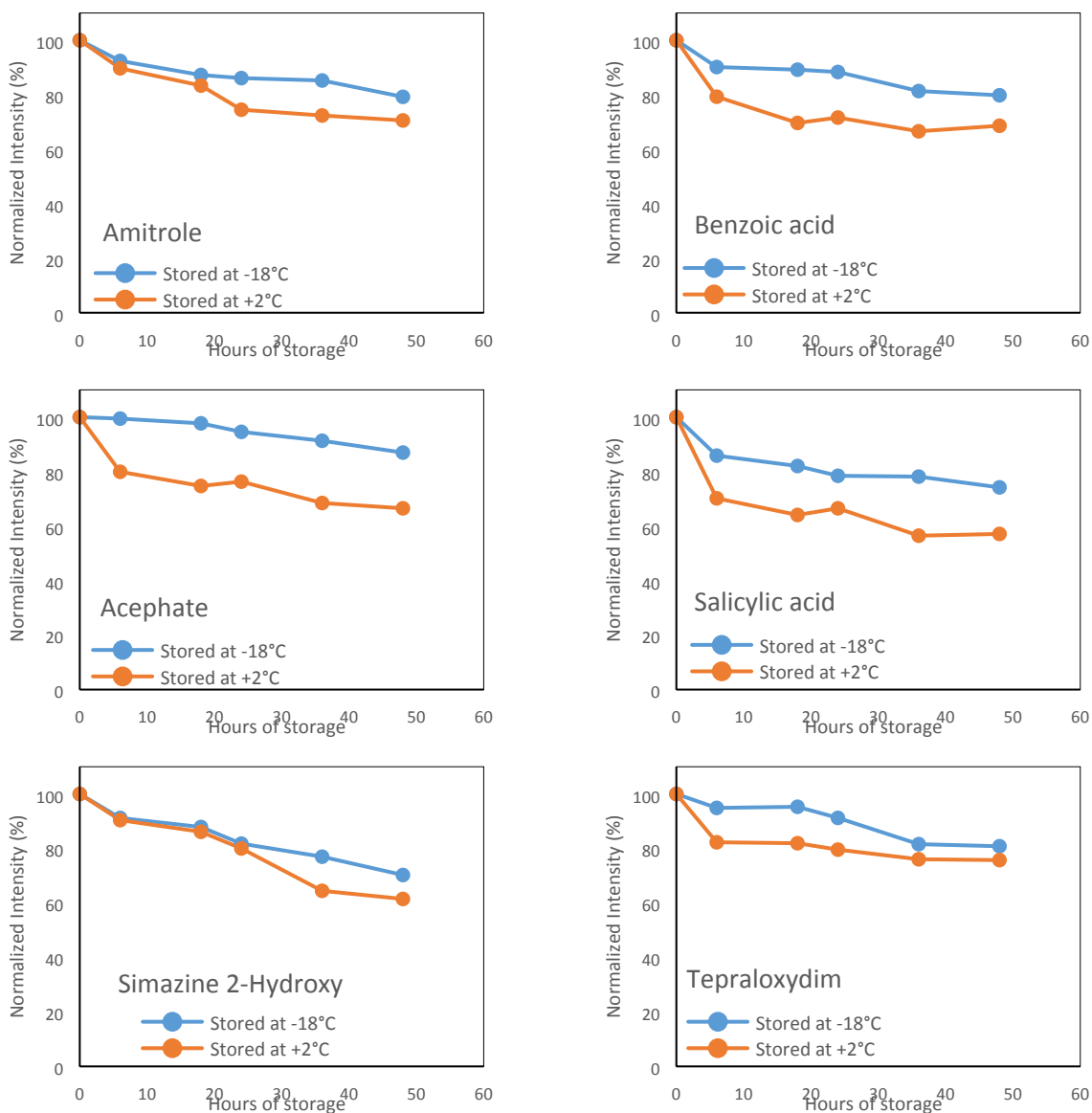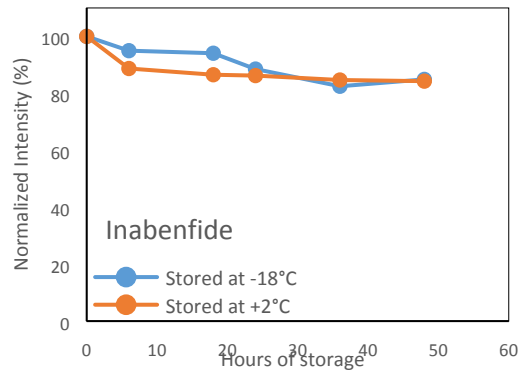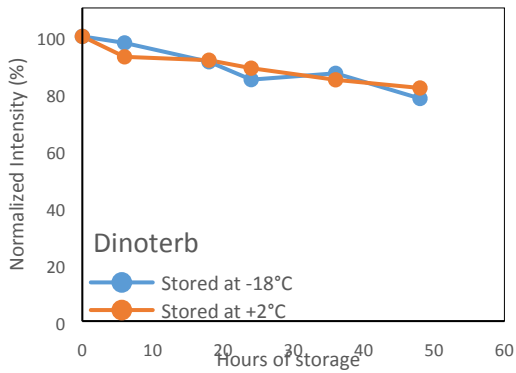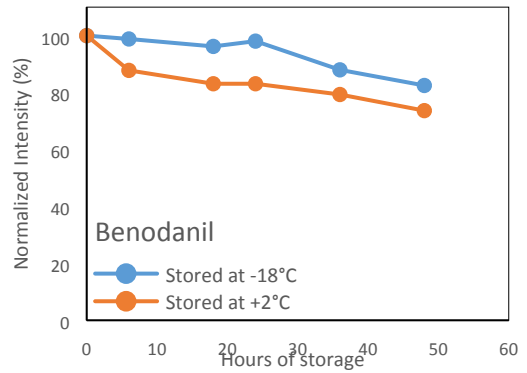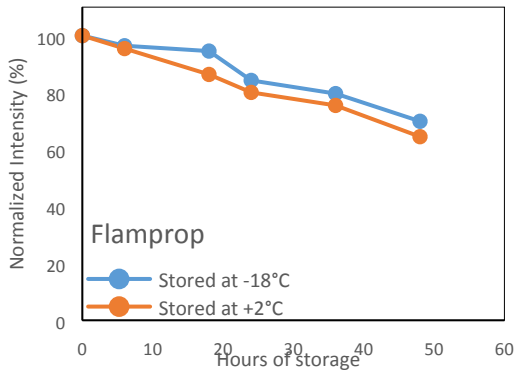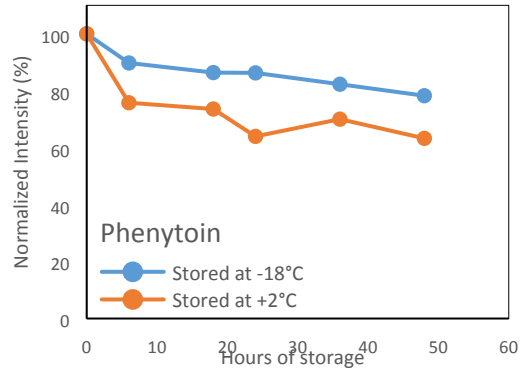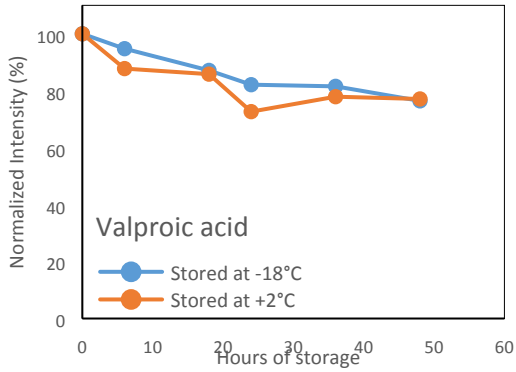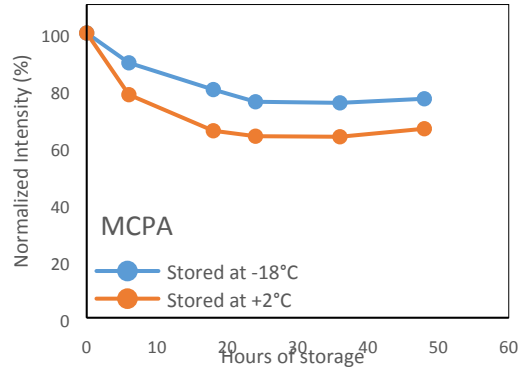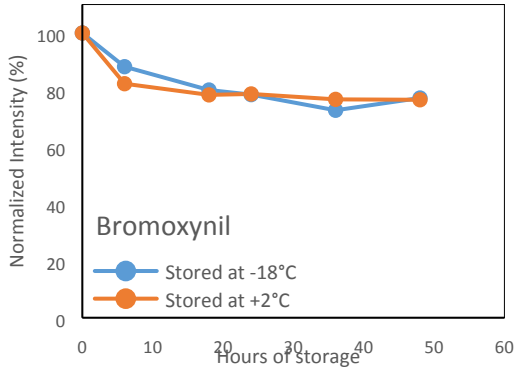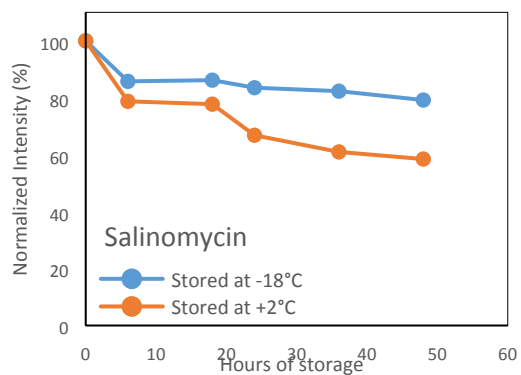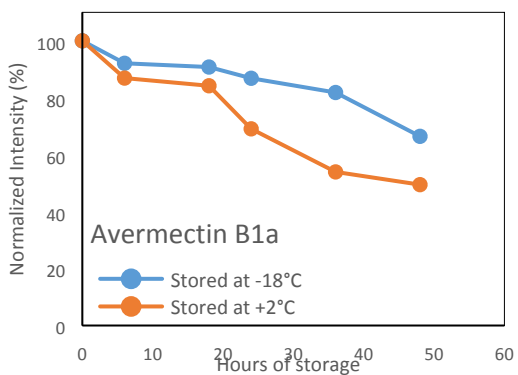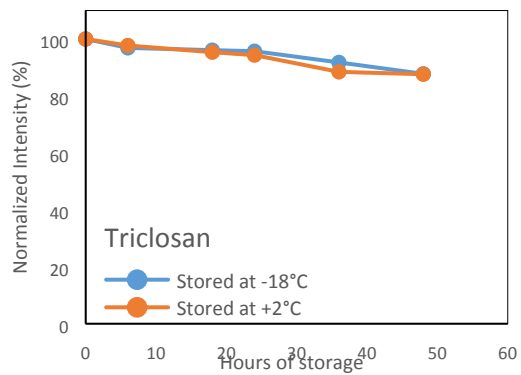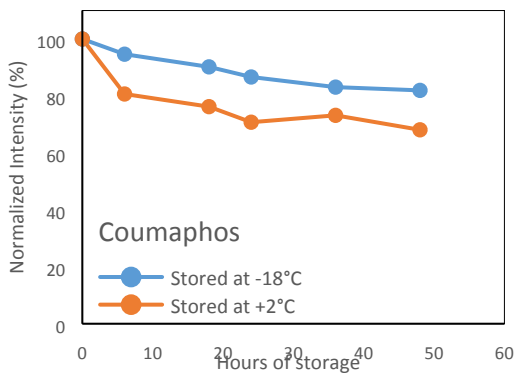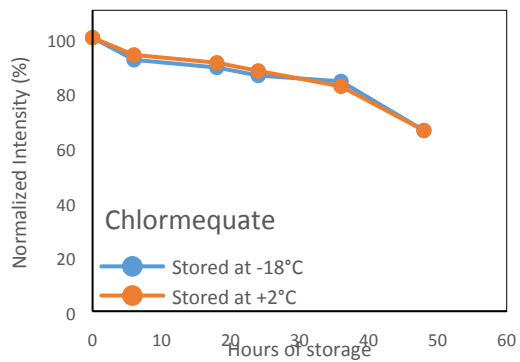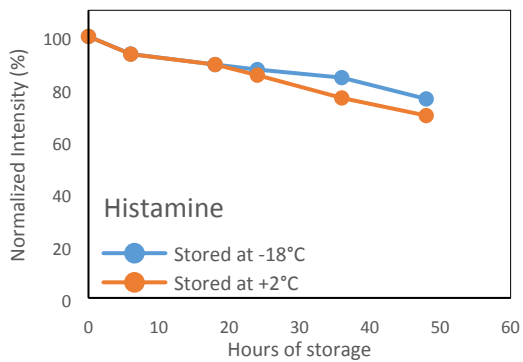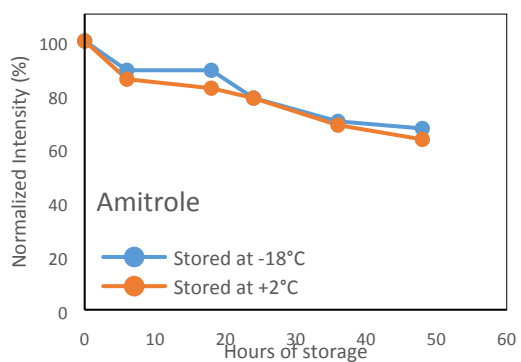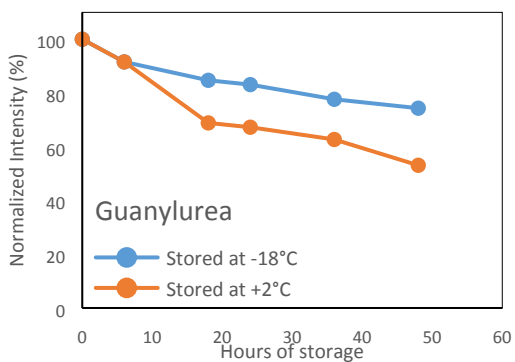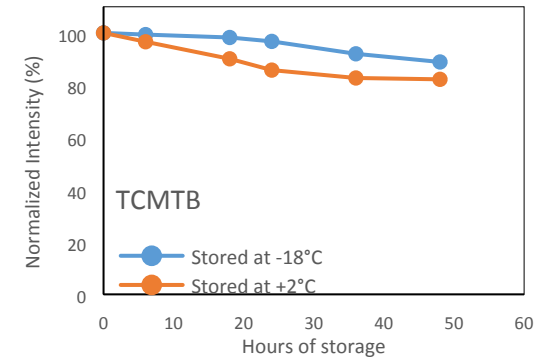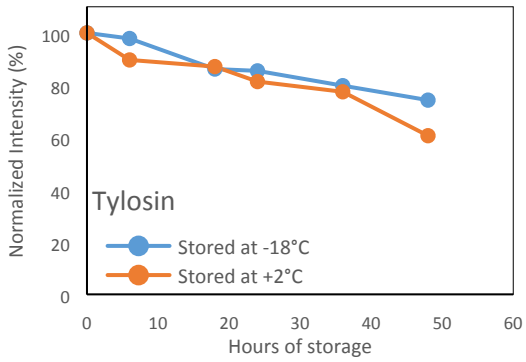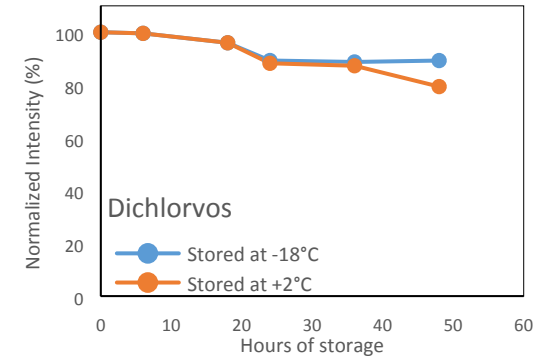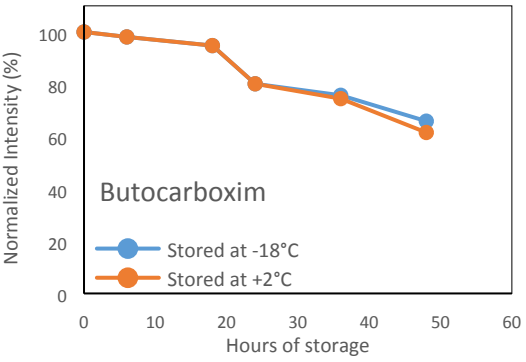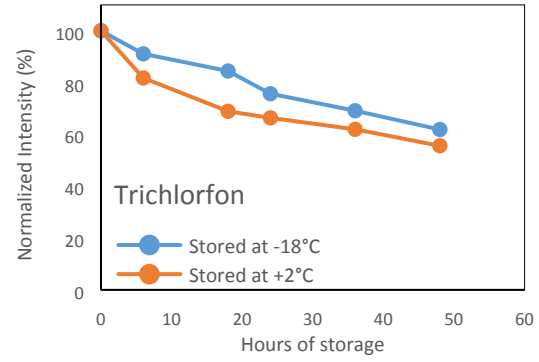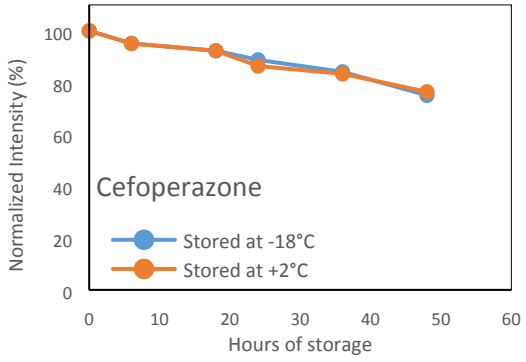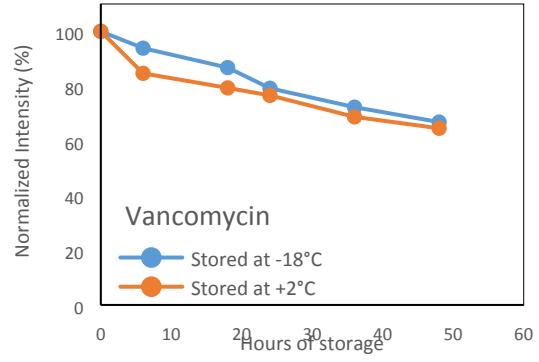
Figure S3. Stability test for the RTI calibrants

(A)

SI 2.3. Applicability Domain

Figure S4 A&B show the results of OTrAMS for the predicted RTI in both ±ESI mode. As it can be seen, all the compounds are within the acceptance threshold of ±3SR (box3). **Figure S4 C&D** show the chemical space boundaries for the compounds used in each ESI platform. It is found that all chemicals are inside the chemical space boundaries.
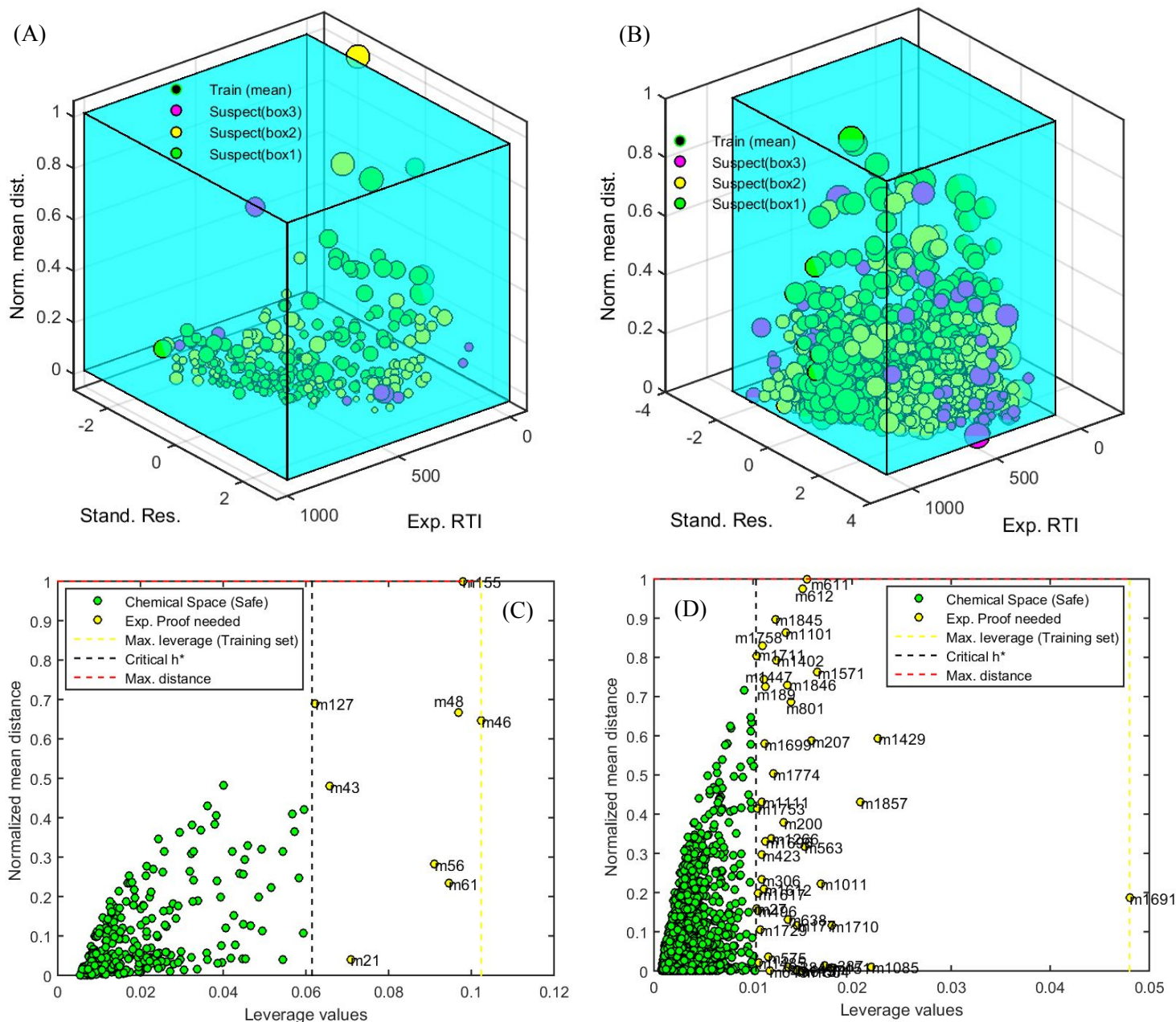

Figure S4. Applicability domain of the proposed RTI models for (A) using OTrAMS in −ESI; (B) using OTrAMS in +ESI; (C) chemical space boundaries in −ESI and (D) chemical space boundaries in +ESI.
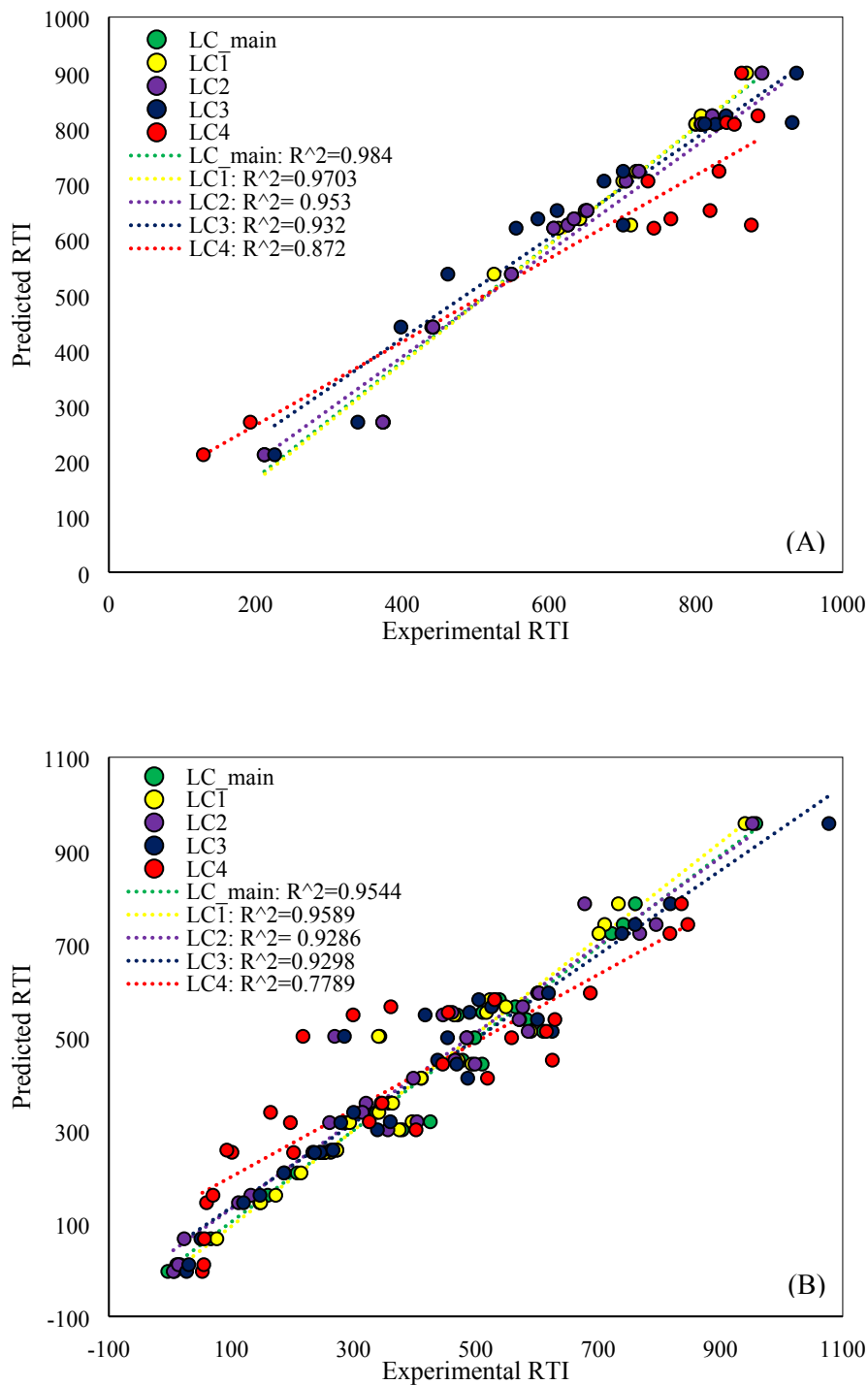
SI 2.4. Intra-laboratory evaluation

Figure S5. The correlation between the experimental and predicted RTIs measured in various LC conditions for validation set used in (A) −ESI and (B) +ESI

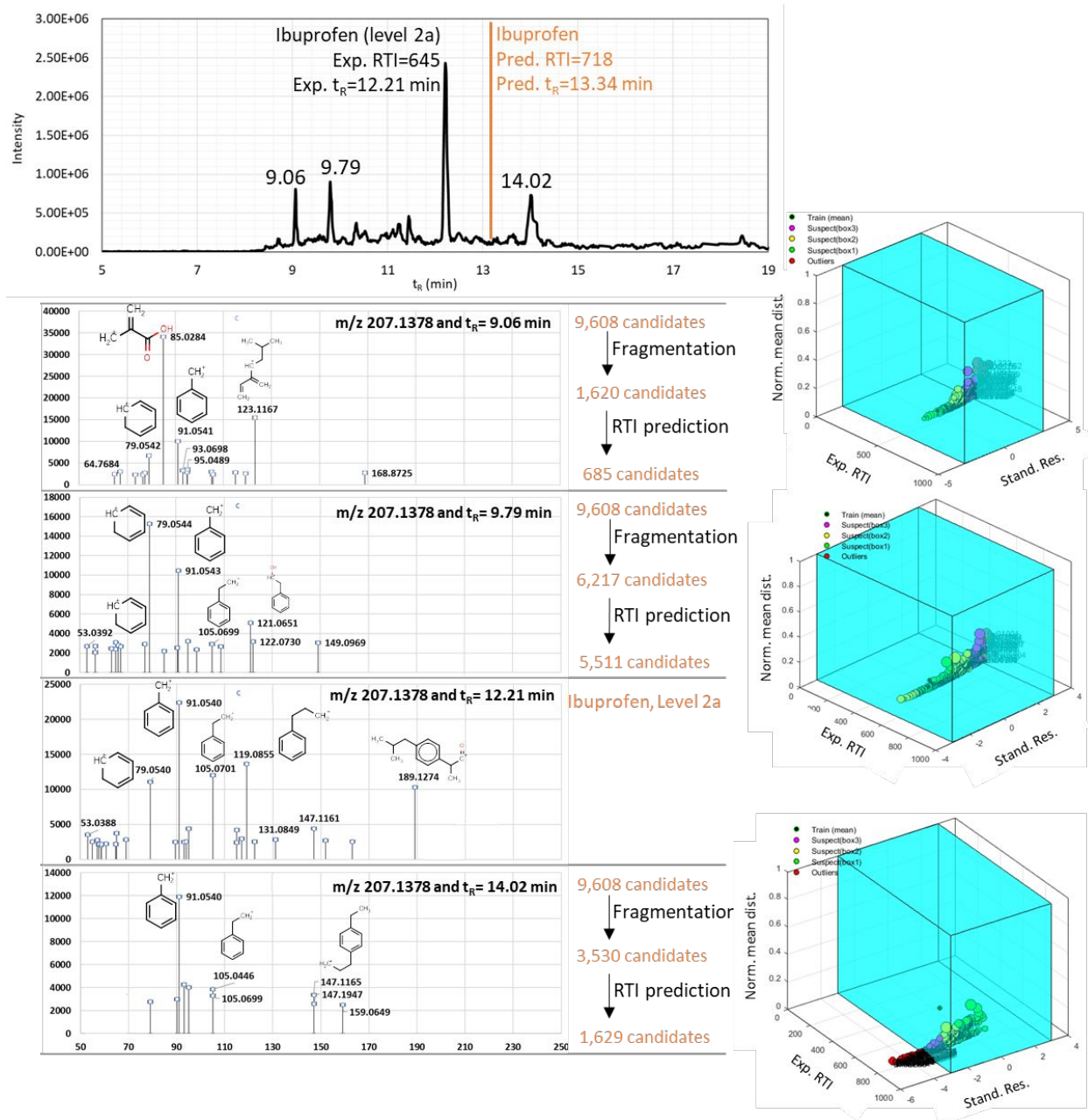## SI 2.5. Application of RTI in Suspect and Non-target Screening.

Figure S6. Extracted ion chromatogram (EIC) of m/z 207.1378 ± 0.002 in river water from Joint Danube Survey 4 (JDS4) conducted in 2019. Four chromatographic peaks were detected in the EIC. RTI prediction successfully excluded the peaks with retention time 9.06 min and 9.79 min and indicated that the peak at 14.02 min belongs most likely to ibuprofen. The identification example highlights the potential reduction of false positive detections and misannotations through the application of the RTI system. 3D bubble plots show the error window for the prediction results. Red bubbles represent candidates that are out of the applicability domain.
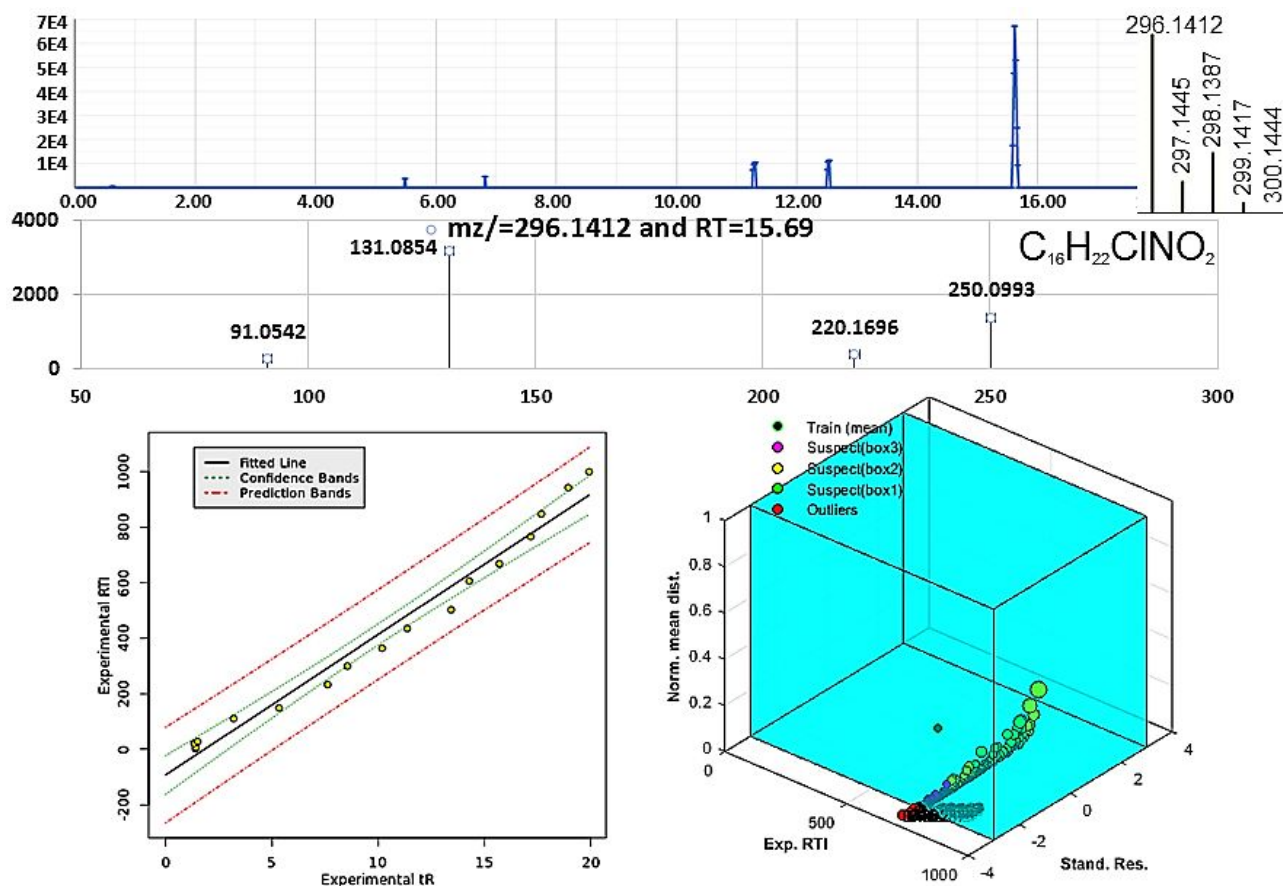
Figure S7. Extracted ion chromatogram of 296.1412±0.002, isotopic pattern, HRMS/MS spectrum, RTI=f(tR) and applicability domain plot for candidate substances. The data were extracted from a river water sample collected in context of Joint Danube Survey 4 (JDS4) and analysed by Laboratory B. 3D bubble plots show the error window for the prediction results. Box 1 and box 2 are the areas of acceptance of predicted RTI, while boxes 3 and 4 are areas with unacceptable predicted RTI.

REFERENCES

(1) Dorigo, M.; Birattari, M.; Stützle, T. Ant colony optimization artificial ants as a computational intelligence technique. *IEEE Comput. Intell. Mag.* **2006**, *1*, 28-39.

(2) Dorigo, M.; Blum, C. Ant colony optimization theory: A survey. *Theor. Comput. Sci.* **2005**, *344*, 243-278.

(3) 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.

(4) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462-2474.

(5) 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.

(6) Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466-1474.

(7) Partitioning(logD) Marvin 6.3.1; ChemAxon, 2014; http://www.chemaxon.com/. (Accessed date: 15-04-2018).

(8) Frey, B. J.; Dueck, D. Clustering by Passing Messages Between Data Points. *Science* **2007**, *315*, 972-976.

(9) Aalizadeh, R.; Nika, M.-C.; Thomaidis, N. S. Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. *J. Hazard. Mater.* **2019**, *363*, 277-285.

(10) 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.

(11) Weaver, S.; Gleeson, M. P. The importance of the domain of applicability in QSAR modeling. *J. Mol. Graph. Model.* **2008**, *26*, 1315-1326.

(12) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; van de Sandt, J. J. M.; Tong, W.; Veith, G., et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships, *ATLA, Altern. Lab. Anim.* **2005**, *33*, 1-19.

(13) Golmohammadi, H.; Dashtbozorgi, Z.; Acree Jr, W. E. Quantitative structure–activity relationship prediction of blood-to-brain partitioning behavior using support vector machine. *Eur. J. Pharm. Sci.* **2012**, *47*, 421-429.