

In Silico Supported Nontarget Analysis of Contaminants of Emerging Concern: Increasing Confidence in Unknown Identification in Wastewater and Surface Waters

Luisa F. Angeles, Lahiruni M. Halwatura, Jonathan P. Antle, Scott Simpson, Caroline M.B. Jaraula, and Diana S. Aga*



Cite This: *ACS EST Water* 2021, 1, 1765–1775



Read Online

ACCESS |



Metrics & More



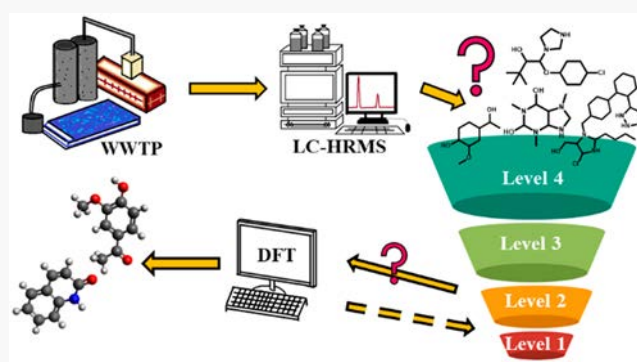
Article Recommendations



Supporting Information

ABSTRACT: Nontarget analysis using liquid chromatography–high resolution mass spectrometry (LC–HRMS) is a valuable approach in characterizing for contaminants of emerging concern (CECs) in the environment. In this study, wastewater and surface water samples from three locations in Manila, Philippines were analyzed for CECs using a nontarget analysis approach with an LC-Orbitrap. A previously optimized semiautomated workflow was used for data processing with Compound Discoverer. A total of 157 compounds were identified, with 21 confirmed with reference standards, 83 confirmed with evidence from a mass spectral library (mzCloud), and 53 tentatively identified using in silico fragmentation (MetFrag). These compounds include pharmaceuticals such as antibiotics, antifungal, and antihypertensive compounds, human metabolites, natural products, pesticides, and industrial chemicals. Compounds confirmed with reference standards include antibiotics ciprofloxacin, clarithromycin, acetyl-sulfamethoxazole, and trimethoprim (2 to 19 ng/L), and antifungal compounds carbendazim and climbazole (3–47 ng/L). The pesticides diethyltoluamide (DEET) and diuron were also detected (37 ng/L). The utility of a preliminary multivariable linear regression quality structure-retention relationship (QSRR) model based on quantum chemical molecular descriptors is demonstrated. This study demonstrates the importance of using tools and software that are helpful for annotating HRMS data and reporting detections according to a standardized classification system. The detection of several CECs in wastewater and surface water samples show the importance of performing nontarget analysis in determining occurrence of CECs in the environment.

KEYWORDS: orbitrap, high resolution mass spectrometry, Philippines, emerging contaminants, liquid chromatography



1. INTRODUCTION

The ability of high resolution mass spectrometry (HRMS) to acquire data with high mass accuracy to facilitate compound identification without limiting the number of target analytes poises it as a revolutionary tool for environmental monitoring applications.¹ The enhanced analytical capability offered by HRMS has led to the discovery of contaminants of emerging concern (CECs) in the environment. CECs include pharmaceuticals and personal care products, pesticides, and industrial chemicals^{2–4} that contaminate water bodies due to the discharge of effluents from wastewater treatment plants (WWTPs), hospital wastes, industrial discharges, as well as agricultural runoff.^{5–9} These compounds may have adverse human health and ecotoxicological effects. Aquatic organisms are constantly being exposed to these chemical contaminants, resulting to the accumulation of these compounds in various organs of different fish species.^{10,11} In addition, the presence of antibiotic residues in the environment has been linked to the

development and spread of antimicrobial resistance (AMR) globally.¹² Low and middle-income countries (LMICs) are major contributors to global AMR because of several factors, one of them being the unregulated use of antibiotics.⁷ The Philippines is an LMIC in Southeast Asia where there is a high prevalence of AMR because of several reasons including the self-medication of antibiotics, antibiotic sharing, and premature termination of antibiotic courses.^{13,14} These practices stem from the insufficient regulations and enforcement of pharmaceutical law in the country.^{13,14} In addition, surface waters in the Philippines are heavily polluted because of the dumping of

Received: March 26, 2021

Revised: July 19, 2021

Accepted: July 21, 2021

Published: August 1, 2021



industrial wastes into the lakes and rivers and the lack of wastewater treatment, with less than 10% of the population having access to piped sewerage systems.^{15,16}

In spite of the high prevalence of AMR and the water pollution crisis in the Philippines, there is limited information on the CECs present in wastewater and surface water in the country. The few studies performed were mostly focused on the analysis of pesticides in surface waters.^{17–20} While it is important to monitor for pesticides, there are several other CECs that may be present in the aquatic environment that remain unreported. Nontarget analysis is a completely unbiased study wherein all the features detected in the analysis are explored and identified. This allows for the analysis of CECs such as metabolites, transformation products, and compounds with expensive or unavailable standards.^{1,3,21–23} The nontarget approach allows for the further expansion of the detection coverage to characterize more chemicals that may be present in the samples. One of the biggest challenges in nontarget analysis is the identification of the unknown compounds detected.

Theoretical methods have been implemented in past studies to accurately confirm the identity of compounds by predicting the chromatographic retention in liquid chromatographic techniques. Among different applications, physicochemical retention models are often studied which utilizes physical and chemical properties of targeted analytes, chromatographic stationary phases, and mobile phases to relate to retention parameters to derive mathematical expressions.²⁴ One study has investigated a range of retention models to predict retention time with chromatographic gradient elution programs, and are able to achieve prediction errors of below 1–2%.^{25,26} Wang et al. have also addressed prediction of retention time by examining different models for a set of sulfonamides with 0.70% prediction error average for multiple gradient conditions and flow rates.²⁷

Quantitative structure-retention relationships (QSRR) models have been used to predict elution order and retention times. The general procedure for QSRR models includes preparing a training set where the retention time/factor of a known analyte is fitted to “molecular descriptors” via a mathematical model. The model is then utilized to predict the retention of an unknown target compound. Oftentimes, these so-called molecular descriptors are the results of quantum chemical calculations.²⁸ For example, researchers utilized molecular descriptors calculated from density functional theory (DFT) to understand the retention mechanism of 30 flavonoids.²⁹ QSRR models as applied to liquid chromatography have been the subject of a number of reviews.^{28,30–34}

Recently, machine learning (ML) has also been used to develop retention time prediction models for small molecules.^{30,35,36} A recent application of ML was utilized to determine new and emerging contaminants in wastewater and surface waters based on predicted retention times.³⁷ Another ML approach assisted in identification of pesticides and pesticide transformation products in strawberry extracts.³⁸ However, the quality of prediction of these ML approaches require large data sets, which is difficult to incorporate experimentally to develop the training set.

Despite these recent advances, it is challenging to predict the identification of a broad spectrum of compounds. Approaches of obtaining physicochemical properties reports computational packages such as EPI Suite, ClogP, and SPARC lack quantum chemical treatment of the studied molecules.^{39–41} Here-in we

employ COSMOtherm, an emerging in-silico model for the validation of chemical compounds based on the “conductor-like screening model for real solvents” (COSMO-RS) theory⁴² and TURBOMOLE to derive a plausible QSRR model to predict the level 2 confirmed compounds of this study. For the prediction of thermodynamic properties of a molecule, COSMO-RS utilizes screening charge density calculated through quantum-chemical methods.⁴³ Therefore, the generation of unique molecular descriptors that are solvent dependent are easily conducted within the framework of COSMO-RS. Hence aforementioned models display much complexity, in this study we introduce a simpler, yet accurate, methodology to facilitate the identification of compounds by relating calculated parameters with experimental retention factor values. In practice, this approach can be beneficial for the prediction of chemical compounds that do not have available analytical standards or when the standards are cost prohibitive.

The objective of the study is to perform a nontarget analysis for CECs in wastewater and receiving surface water samples collected in Manila, Philippines. Compound Discoverer was used as the data processing software, while the mass spectra library mzCloud and in silico fragmentation software MetFrag, were used to tentatively assign structures⁴⁴ to the compounds detected. MetFrag is based on a combinatorial approach that starts by suggesting candidate structures from various databases including ChemSpider,⁴⁴ PubChem,⁴⁵ or KEGG⁴⁶ or by accepting the upload of a structure data file containing candidates. Using a bond dissociation approach, fragmentation of the candidate is predicted to generate theoretical structures that are compared to the *m/z* fragment peaks of the experimental MS/MS spectrum. A detailed description of the latest improvements in MetFrag and its applications in predicting structures of small molecules can be found in a previous publication.⁴⁵ COSMOtherm and TURBOMOLE were also explored as in silico tools to predict the identity of unknown compounds.

2. MATERIALS AND METHODS

2.1. Sample Collection and Solid Phase Extraction (SPE).

Wastewater influent and effluent and receiving surface water that are located upstream and downstream of WWTPs were collected as grab samples from three locations in Manila, Philippines. The exact locations are not provided in order to keep the identities of the WWTPs. The locations will be referred to as PH1, PH2, and PH3 in this work. For PH1 and PH2, wastewater influent and effluent, and surface waters that are upstream and downstream of the WWTPs were collected. For PH3, only wastewater influent and effluent were collected since the receiving surface waters were not accessible. All samples were collected in triplicate, and each of these samples analyzed individually. The samples were collected in October 2018.

The water samples (500 mL) were collected by grab sampling in amber glass bottles. The samples were filtered using 0.45 μm nylon filters (VWR, Radnor, PA) to remove particulate matter and microorganisms. Then, 40% phosphoric acid (J.T. Baker, Philipsburg, NJ) was added to adjust the pH to 2.5 ± 0.5 and 2 mL of Na_2EDTA (5% w/v in water) was added to the samples. Solid phase extraction (SPE) was performed using the procedure described in our previous work,⁷ which uses OASIS hydrophilic–lipophilic balance (HLB) cartridges (Waters, Milford, MA). The SPE cartridges

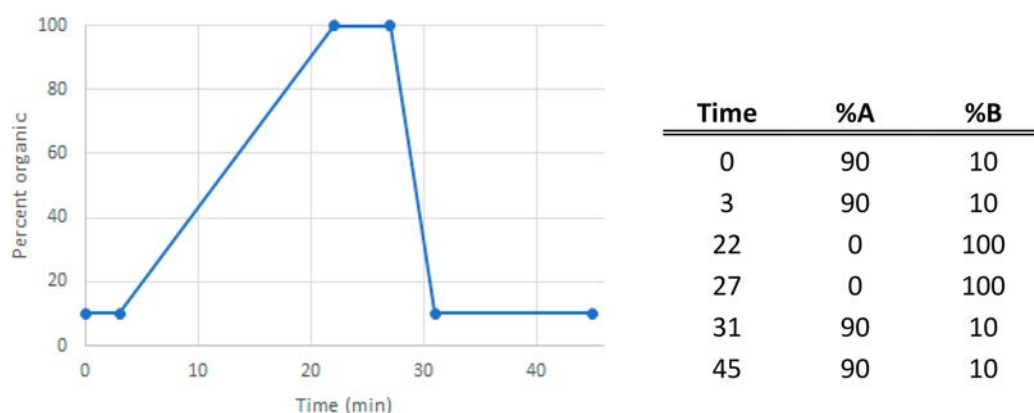


Figure 1. Chromatographic gradient of the LC-Orbitrap method. Mobile phase A contains 0.1% formic acid in water while mobile phase B contains 0.1% formic acid in acetonitrile.

Table 1. Compound Detected Using Non-Target Analysis That Were Confirmed with Reference Standards (Level 1)^a

| name | chemical formula | RT (min) | exact mass | concentration (ng/L) | | | | | | | | | |
|-----------------------------------|------------------|----------|------------|----------------------|-----|-----|-----|-----|----|-----|----|-----|-----|
| | | | | PH1 | | | | PH2 | | | | PH3 | |
| | | | | IN | EF | UP | DN | IN | EF | UP | DN | IN | EF |
| caffeine | C8 H10 N4 O2 | 7.0 | 194.0804 | 143 | 1 | 99 | 110 | 18 | 1 | 13 | 14 | 122 | nd |
| carbamazepine | C15 H12 N2 O | 10.5 | 236.0950 | nd | nd | nd | nd | 2 | nd | 1 | 1 | nd | nd |
| carbendazim | C9 H9 N3 O2 | 3.0 | 191.0695 | 32 | nd | 4 | 6 | nd | nd | nd | nd | 3 | nd |
| ciprofloxacin | C17 H18 F N3 O3 | 6.7 | 331.1332 | 2 | nd | 2 | nd | 3 | nd | nd | nd | 3 | nd |
| clarithromycin | C38 H69 N O13 | 11.8 | 747.4769 | nd | nd | nd | nd | nd | nd | nd | nd | 3 | nd |
| climbazole | C15 H17 Cl N2 O2 | 12.0 | 292.0979 | 3 | nd | 16 | 4 | 3 | 3 | 10 | 7 | nd | 47 |
| DEET | C12 H17 N O | 15.5 | 191.1310 | 37 | 7 | 20 | 11 | 16 | 3 | 16 | 11 | 27 | 11 |
| diethyl phthalate | C12 H14 O4 | 17.3 | 222.0892 | 487 | 191 | 512 | 466 | 36 | 4 | 9 | 9 | 235 | 560 |
| diuron | C9 H10 Cl2 N2 O | 15.8 | 232.0170 | 16 | 5 | 13 | 22 | 17 | 7 | 16 | 20 | 18 | nd |
| gabapentin | C9 H17 N O2 | 2.4 | 171.1259 | 1 | nd | nd | nd | 2 | nd | 2 | 1 | 2 | nd |
| indole-3-acetic acid | C10 H9 N O2 | 11.4 | 175.0633 | 359 | nd | 586 | 17 | 307 | nd | 295 | nd | nd | 480 |
| irbesartan | C25 H28 N6 O | 13.1 | 428.2325 | nd | nd | 2 | nd | nd | nd | 2 | 2 | 18 | nd |
| losartan | C22 H23 Cl N6 O | 14.4 | 422.1622 | nd | nd | 12 | 5 | 3 | nd | 5 | 4 | 26 | nd |
| metoprolol | C15 H25 N O3 | 6.7 | 267.1834 | 2 | nd | 2 | 3 | 1 | nd | 2 | 2 | nd | 9 |
| metronidazole | C6 H9 N3 O3 | 4.0 | 171.0644 | 2 | nd | 3 | 2 | nd | nd | nd | nd | nd | nd |
| mycophenolic acid | C17 H20 O6 | 15.6 | 320.1260 | nd | nd | nd | nd | 3 | nd | 5 | 2 | nd | 3 |
| <i>N</i> -acetyl sulfamethoxazole | C12 H13 N3 O4 S | 11.3 | 295.0627 | 13 | nd | 19 | 8 | nd | nd | nd | nd | 8 | nd |
| oxybenzone | C14 H12 O3 | 20.4 | 228.0786 | 2 | nd | 4 | 3 | 2 | nd | 2 | 2 | nd | 5 |
| piperine | C17 H19 N O3 | 18.3 | 285.1365 | 1 | nd | 2 | 2 | nd | nd | 2 | 1 | nd | nd |
| trimethoprim | C14 H18 N4 O3 | 3.7 | 290.1379 | nd | nd | 3 | nd | nd | nd | nd | nd | nd | nd |
| triphenylphosphine oxide | C18 H15 O P | 16.7 | 278.0861 | 31 | 16 | 4 | 13 | 1 | 2 | 1 | 2 | nd | nd |

^aAll concentrations included in this table are semi-quantitative.

were shipped overnight in ice to the University at Buffalo for analysis using liquid chromatography mass spectrometry (LC-MS). The concentrations of the confirmed targets were semiquantified by one-point external calibration using standards with a concentration of 100 $\mu\text{g/L}$.

2.2. LC-Orbitrap Analysis. The separation method was performed using a reversed-phase Cortecs C_{18}^+ 2.1 \times 150 mm analytical column (Milford, MA), with a 2.7 μm particle size in a Dionex Ultimate 3000 ultra high-performance liquid chromatograph (UHPLC). A full scan data dependent MS/MS (ddMS2) method was performed to analyze the samples using a QExactive Focus Orbitrap LC-MS (Waltham, MA). The separation method as well as the mass spectrometer settings and parameters are described in our previous work.⁴⁷ The elution profile is shown below in Figure 1.

2.3. Nontarget Analysis Workflow. All mass spectral data obtained from the analysis were imported into the

software Compound Discoverer 3.0 for data processing. Supporting Information (SI) Figure S1 shows the workflow utilized for data processing. The details on the specific Compound Discoverer workflow are described in the SI.

Following data processing in Compound Discoverer, data filters were applied in order to reduce the number of hits and eliminate features that may be insignificant. The workflow used in this study is a modified version of the suspect screening workflow that was used in our previous work.⁴⁸ Features that were present in the blank, not present in three technical replicates, and those that had areas less than 5×10^4 were eliminated from the list of compounds to be identified. The unknown compounds that remained after the application of the filters were then either confirmed with reference standards or tentatively identified according to the Schymanski scale using mass spectra library matching with mzCloud or in silico fragmentation using MetFrag.⁴⁹

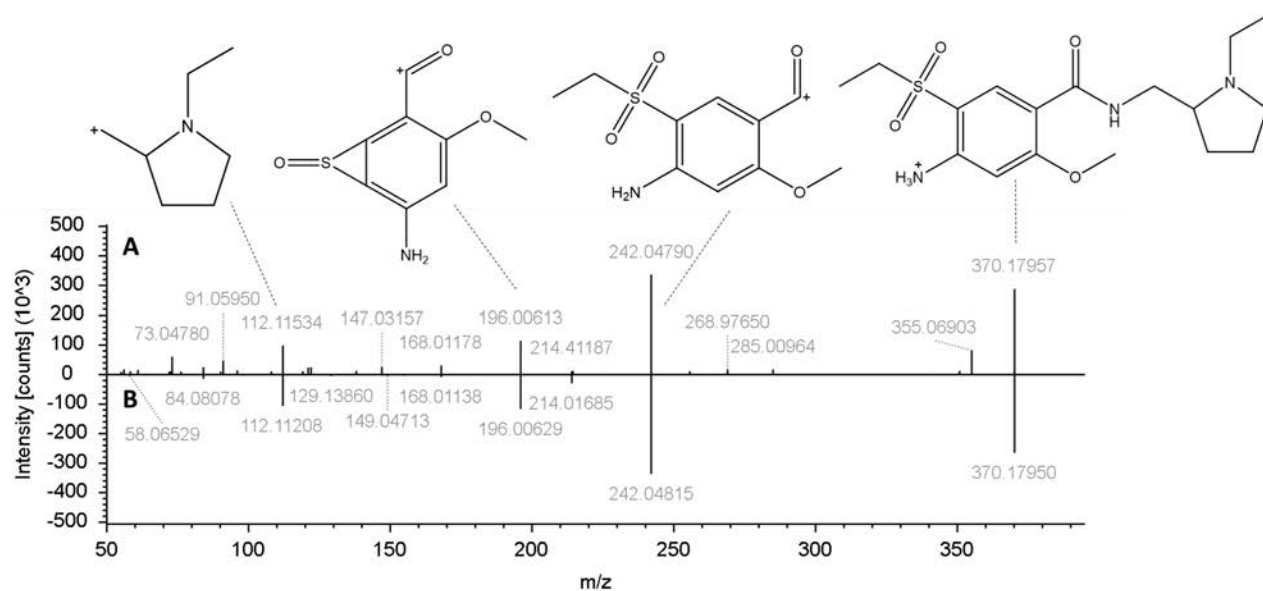


Figure 2. Identification of amisulpride (level 2) including fragment annotations in a surface water sample in the Philippines. (A) shows the mass spectra for amisulpride in the surface water sample and (B) shows the mass spectra of amisulpride obtained from the mzCloud library.

2.4. Procedures for In Silico Retention Modeling: TURBOMOLE and COMSO-RS Methodologies.

Determination, optimization of lowest energy conformers, and calculation of polarizability values along with other COSMO-RS derived thermodynamic properties of each of the compounds were performed according to one of our previously published method.^{41,43} The quantum chemical software package, TURBOMOLE was utilized in all Density functional theory (DFT) calculations throughout this study.⁵⁰ The BP86 GGA functional, and a def2-TZVPD basis set were used for each element for the optimization of all conformer geometries.^{51–54} Solubility calculations were carried based on solvents used in chromatographic mobile phases of the LC-Orbitrap separation methodology. COSMO-RS derived solubilities were calculated at different points of the chromatographic gradient program with various ratios of the mobile phases at (a) 0–3 min (S_{start}), (b) 22–27 min (S_{end}), and (c) ratio of mobile phases at the point of compound elution (S_{elution}).⁵⁵ Solvents used in mobile phases described in a previous work of ours.⁴⁷

3. RESULTS AND DISCUSSION

3.1. Occurrence of CECs in Wastewater and Surface Water by Nontarget Analysis. A nontarget analysis workflow following the data processing steps shown in SI Figure S1 was performed. The features that remained after the application of the data filters in the workflow were identified and classified into different levels according to the proposed levels of the Schymanski scale.⁴⁹ A compound can be assigned a level 1 if it has been confirmed with a reference standard that was analyzed in the same LC-MS method used to analyze the samples. The LC retention time and MS/MS data between the sample and the standard must match. Level 2 can be assigned to a compound when a mass spectra library match exists or when diagnostic evidence shows no other possible structure. Level 3 is assigned for a tentative structure when there is evidence for a compound class but the exact structure is unknown. Levels 4 and 5 can be assigned for compounds with

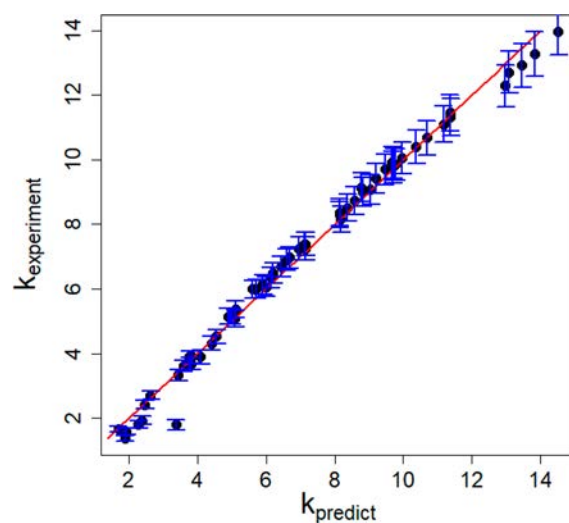


Figure 3. Predictive ability of the selected regression model for the level 1 training set with $\pm 5\%$ error bars.

evidence for a chemical formula and exact structure, respectively.

The identity of the features where the reference standards were available in the Aga laboratory were confirmed. The standards were analyzed in the same method that the samples were analyzed in the LC-Orbitrap method. The retention times and mass spectra of the compounds in the sample were compared to those of the corresponding reference standards, and a level 1 was assigned when a match was found. See SI Section S2 for the confirmation of carbendazim in the PH1 downstream sample, where the chromatographic peaks in both the sample and the carbendazim standard were at 2.9 min. In addition, the mass spectra of the feature in the sample and the carbendazim standard also showed the same fragments, confirming the identity of the unknown compound.

A total of 21 compounds were identified as level 1 as shown in Table 1. The concentrations for these compounds were estimated using a one-point external calibration. The 21

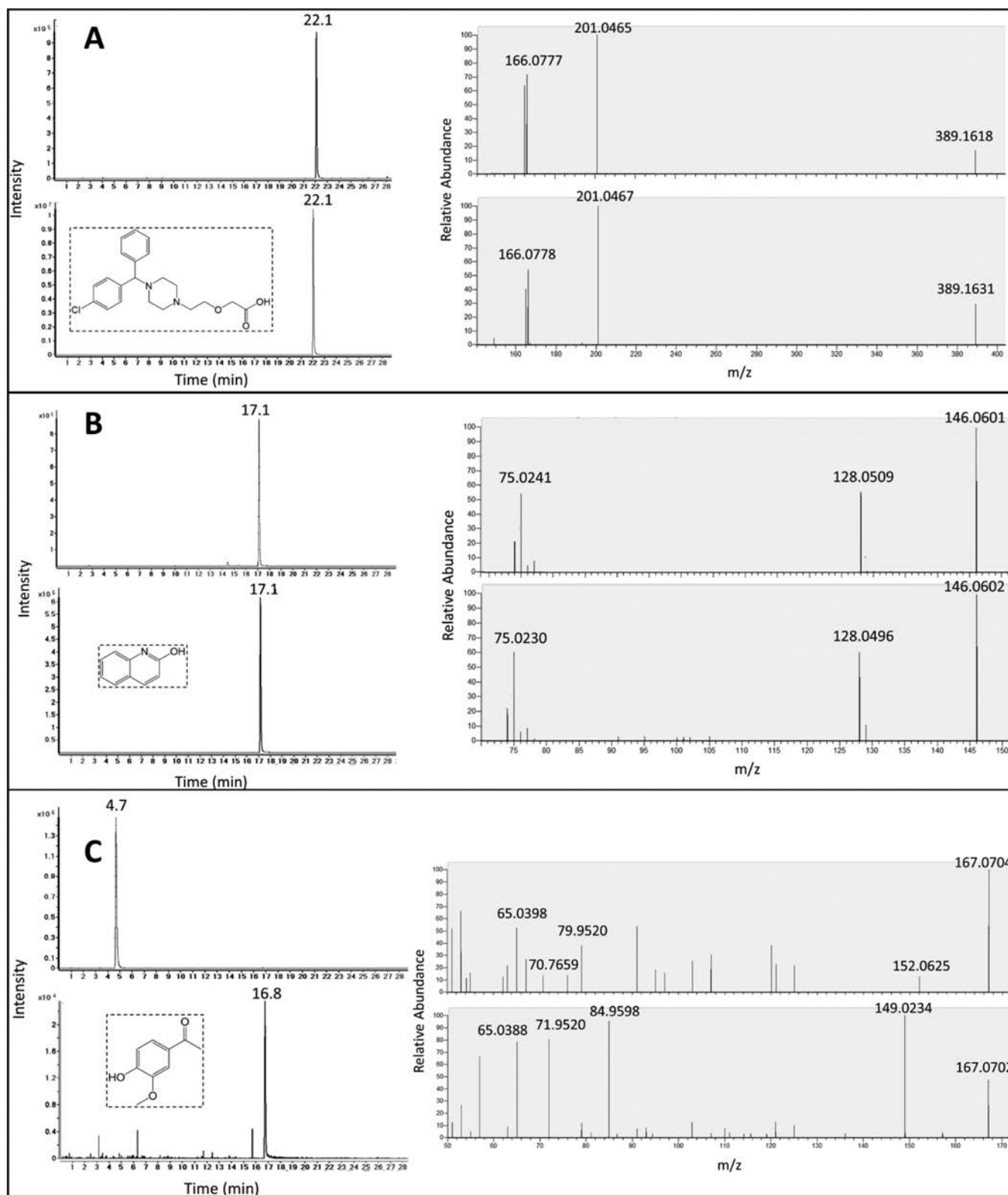


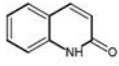
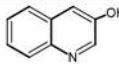
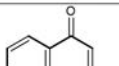

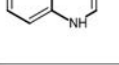
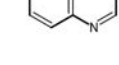
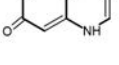
Figure 4. Extracted ion chromatograms and mass spectra of test unknowns from level 2 identified compounds in an influent water sample (top) with reference standards (bottom). (A) Cetirizine, (B) 2-hydroxyquinoline, and (C) apocynin.

compounds confirmed with level 1 confidence were antibiotics, antifungal compounds, anticonvulsants, antihypertensive drugs, pesticides, and industrial chemicals. The antibiotics detected were ciprofloxacin, clarithromycin, acetyl-sulfamethoxazole, and trimethoprim at concentrations of up to 19 ng/L. All of

these antibiotics were also detected in wastewater from a different WWTP in the Philippines, as reported in our previous work.⁷

Carbendazim and climbazole, which are antifungal compounds, were also detected in the samples. Carbendazim is

Table 2. Predicted Retention Factor (k_{predict}) of Different Isomers of 2-Hydroxyquinoline Reported in Minutes

| Molecule | Structure | k_{predict} | % Error |
|--------------------|---|----------------------|---------|
| 2-hydroxyquinoline |  | 5.647 | 3.5 |
| 3-hydroxyquinoline |  | 3.687 | 37.0 |
| 4-hydroxyquinoline |  | 3.526 | 39.8 |
| 5-hydroxyquinoline |  | 3.700 | 36.8 |
| 6-hydroxyquinoline |  | 3.675 | 37.2 |
| 7-hydroxyquinoline |  | 3.681 | 37.1 |
| 8-hydroxyquinoline |  | 3.600 | 38.5 |

used in agriculture while climbazole is used in human medicine as a treatment for fungal infections such as dandruff, and is therefore present in personal care products.⁵⁶ Carbendazim and climbazole were both detected in all WWTPs at concentrations ranging from 3 to 47 ng/L. Antifungal compounds are now being considered as emerging environmental contaminants because of their use in both agriculture and personal care products. These compounds are found to be ubiquitous in the aquatic environment due to their resistance to biodegradation and adsorption into sludge.^{47,57}

The antihypertensive drugs irbesartan, losartan, and metoprolol, were also detected in the samples at concentrations ranging from 1 to 26 ng/L. These three compounds were detected in all the locations sampled. The high prevalence of hypertension⁵⁸ in the Philippines and the low biodegradability of these antihypertensive compounds contribute to their presence in surface waters. The pesticides diethyltoluamide (DEET) and diuron were detected in all three locations with concentrations of up to 37 ng/L and 22 ng/L, respectively. DEET is an insecticide that is the main ingredient in insect repellents such as bug sprays and lotions. The high usage of these products in the Philippines may be contributing to the high levels of this compound in the aquatic environment.⁵⁹ DEET is neither persistent nor bioaccumulative, but is commonly detected in wastewater and surface water worldwide. It reaches wastewater and surface water through the washing of skin and clothing and other human activities.⁶⁰ In another study,⁵⁹ DEET was detected in three lakes in the Philippines at concentrations ranging from 12–240 ng/L. Diuron, on the other hand, is an herbicide that is used for weed

control. Diuron is known to be persistent in the environment and has been found to reach surface water and groundwater.⁶¹

Generally, it was observed from the data that the concentrations of the contaminants were lower in the effluent samples compared to the influent samples. Interestingly, the concentrations of the compounds in the surface water samples for PH 1 and PH 2 locations were higher than those in wastewater, suggesting input from untreated sewage. The contaminant levels in surface water samples are typically lower in surface water because of removal from WWTPs, and dilution after the effluent is released into the environment. However, the higher contaminant concentrations in surface waters is not surprising because many of the sampling sites were receiving domestic wastes from nearby residential homes. There were residential homes around the sampling sites which are direct sources of contamination. These areas are also known to be dumping sites of industrial wastes, which means that the wastewater effluent being discharged into these surface waters may not be the only source of contamination.

Next, for the features where reference standards were not available, structures were proposed based on library spectrum matching with mzCloudTM, which is enriched with more than 7 million mass spectral data (<https://www.mzcloud.org/>) of different compounds, and were assigned level 2. Figure 2 shows the comparison of the mass spectra of the unknown compound with an m/z of 370.1796 in the sample and the mass spectra of the compound amisulpride obtained from mzCloud. Three fragments, with m/z values of 242.0479, 196.0061, and 112.1153 were found in both mass spectra and were assigned to structures shown in Figure 2. Amisulpride is an antipsychotic drug that is used for the treatment of schizophrenia.⁶² A total of 83 compounds were assigned a level 2 confirmation with the details shown in SI Table S1. These compounds include natural products, human metabolites, pharmaceuticals, personal care products, and industrial chemicals. To provide further verification of the assigned level 2 compounds, we utilize COSMO-RS based calculations to aid in our identifications.

3.2. Using COSMOtherm Prediction of Retention Factor for Identification of Level 2 Compounds.

In order to construct the QSRR regression models to predict the retention factor (k), different molecular descriptors were taken into consideration to facilitate the identification of level 2 categorized compounds. Solubility of each of the molecules were calculated using COSMO-RS based on the solvent mixture used in the separation method in the LC-Orbitrap to accompany the changes in solubility of compounds with varying states of the chromatographic gradient program. These points include mobile phase ratio (a) at 0–3 min, (b) at 22–27 min, and (c) ratio of mobile phases at the point of compound elution, respectively abbreviated as S_{start} , S_{end} , and S_{elution} . Solubility calculations were taken in two different representations, the log value of the mole fraction based solubility (S_{-x}) and the log value of the volume based solubility (S_{-v}). Solubilities were selected as a molecular descriptor based upon the performance of the Snyder-Soczewiński Equation, which relates retention factor to the change in mobile phase composition.^{30,63} It is well-known that mobile phase composition has direct consequences for analyte elution. Therefore, our COSMO-RS predicted solubility at different compositions of the mobile phase should be an acceptable molecular descriptor. Polarizability calculations were utilized as another parameter thus it is well-known that

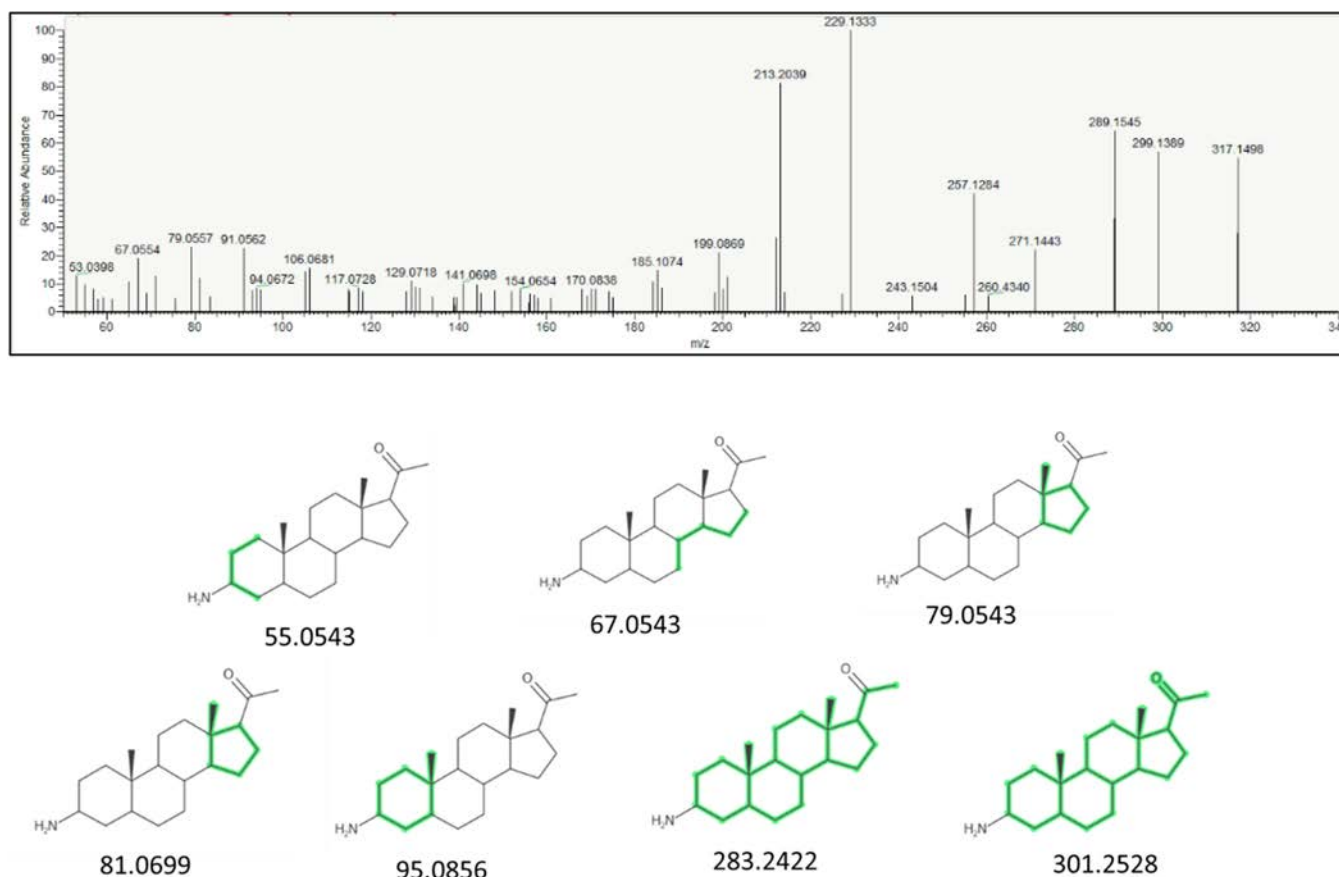


Figure 5. Mass spectrum for the unknown compound with an m/z of 317.2720 tentatively identified as funtumine. The portion of the compounds highlighted in green represent the structures of the fragments generated in silico that corresponds to green peaks in the experimental mass spectrum. The blue peaks are fragments in the experimental mass spectrum that did not match any in silico-generated fragments.

polarizability of compounds govern the interaction of molecules with the chromatographic stationary phase.^{30,64} Other parameters taken into account consist of octanol–water partition coefficients (K_{ow}), vapor pressure, and solvation energy ($\Delta G_{solvation}$). The K_{ow} parameters ascertain the hydrophobic nature of a molecule, hence it is a valuable parameter in predicting compounds. Previous studies have reported $\log K_{ow}$ to be convenient to represent retention of molecules on C18 columns under methanol based organic mobile phases.⁶⁵ Water saturated octanol (wet) was used in calculations of $\log K_{ow}$ values, which implies as another good candidate parameter.⁶⁶ Vapor pressure and $\Delta G_{solvation}$, were also utilized in previous studies as physicochemical properties that govern the interaction of molecules with the chromatographic mobile phases and stationary phase.⁴⁰

Utilizing these parameters, 61 molecules from the level 1 confirmed compounds from our previous work⁴⁸ using the same LC-Orbitrap method, which contained a gamut of structurally different compounds, were applied to establish multivariable regression models. Molecules with a retention factor lower than 1.10 were excluded from the regressions as the performance of the regression metrics were seen to drastically improve.^{30,31} With the intention of constructing the best fit model with minimized standard error values and maximum linearity, experimentally determined retention factor values were pooled with calculated parameters to establish a convincing relationship.

Experimental retention factor (k) values were initially paired with above-mentioned calculated parameters to develop each

regression. Coefficient of determination (R^2), residual standard error (RSE), mean absolute error (MAE), root-mean-square error (RMSE), Akaike information criterion (AIC), and Bayesian information criterion (BIC) were used for the evaluation of models.^{67,68} Technical details and statistical properties of these regressions can be found in the SI. The top regression model of the 1023 constructed models were determined to be with 4- variables: K_{ow} , $S_{start-V}$, $S_{elution-X}$, and $S_{elution-V}$. The individual correlation of the previously described molecular descriptors with k was determined to have poor performance, as demonstrated in the SI. We would like to note that numerous regressions could have been selected that were shown to perform adequately, and a variety of different regressions could be selected as molecular descriptors. However, we selected the regression with the lowest percent error for its prediction of level 2 compounds that had the lowest AIC/BIC values.

For the selected regression, the average percent error (% error) in the retention factor was determined to be 3.2% for the 3 level 2 compounds, with an adjusted R^2 of 0.989, RSE of 0.3408, %MAE of 5.6102, and %RMSEP of 2.9412 with lower AIC and BIC values. The model was determined to be

$$k = 0.1179 \cdot K_{OW} + 0.4311 \cdot S_{start-V} + 48.7952 \cdot S_{elution-X} - 49.1647 \cdot S_{elution-V} + 85.1051$$

Figure 3 contains the experimentally measured retention time ($k_{\text{experiment}}$) versus the predicted retention time (k_{predict}). It is clear to see that the performance of the regression wanes at the

upper/lower values of k . This is most likely due to a concentration of data points in the intermediate values of k that were used in the construction of the model. The major outlier was determined to be metronidazole.

Three level 2 confirmed compounds, cetirizine, 2-hydroxyquinoline, and apocynin (structures shown in Figure 4) were randomly chosen as unknowns to examine the best fit mathematical model for the possibility to correctly identify the chosen compounds. Respectively for cetirizine, 2-hydroxyquinoline, and apocynin % error values were 3.5%, 4.84%, and 1.15%. This is well within the window of error deemed acceptable by quantitative structure-retention relationships (QSRR) models.^{28,30,69} Figure 4 shows extracted ion chromatograms and mass spectra of the three chosen unknowns in an influent water sample. However, samples analyzed with the reference standards ascertain the presence of cetirizine and 2-hydroxyquinoline and lack of apocynin in the influent sample as shown in Figure 4. Performance criteria of the model for the prediction of three compounds can be found in the SI. Figure 4(c) shows that relying only on molecular ion to confirm the identity of an unknown could lead to false positive identification. It has been suggested in an earlier study that in order to achieve a low false positive rate two isotopes should be considered; the use of three isotopes provides higher confidence at the expense of losing the ability to detect low abundant ions. These isotopic peaks pertain to ¹³C isotopes, and isotopes from other elements (e.g., nitrogen, oxygen, and sulfur) that may be present in the mass spectra, depending on the heteroatoms in the compounds.⁷⁰

To demonstrate the utility of our regression, we calculated the retention times of several isomers of 2-hydroxyquinoline (Table 2). Should a researcher have an idea of the molecular formula but be unsure of the bond topology, one could simply calculate the COSMO-RS derived properties for each of the isomers. Having done this for 2-hydroxyquinoline, it is clear to see that the experimentally confirmed structure has the lowest percent error. A strength of this methodology is that it can accurately predict the k/RT of the correct structure if an RT is observed. However, a shortcoming of this approach is if a researcher is searching for an analyte without an observed RT, such as a peak that coelutes or a detection peak is too small, this methodology may not be suitable as the regression depends on S_{elution} . In the future, we plan to investigate avenues to compensate for these weaknesses, along with verifying the identities of the remaining 80 level 2 compounds.

3.3. Using In Silico Fragmentation to Tentatively Identify Level 3 Compounds. The annotation of HRMS data remains one of the biggest challenges in nontarget analysis. Mass spectral libraries such as mzCloud are available but with limited coverage that does not encompass the complete chemical space. In silico fragmentation is a technique that has been used in order to tentatively characterize unknown compounds using computer-based algorithms. MetFrag is a freely available software that provides in silico fragmentation of unknown compounds using different compound libraries.

The mass spectral data of features that did not match compounds in the mzCloud library were imported into MetFrag for in silico fragmentation and PubChemLite was the database used to extract the candidate compounds. In silico fragmentation using MetFrag was performed for a total of 96 compounds. The top hit that had the highest score from the candidate list was the tentative structure assigned to the

corresponding feature and was assigned a level 3. An example of a level 3 identification is presented in Figure 5, which shows the comparison of the mass spectra of the feature with an m/z of 317.2720 and the fragments generated from in silico fragmentation using MetFrag. This unknown was assigned a level 3, and tentatively identified as funtumine, a steroid alkaloid, with the structure shown in Figure 5. The other compounds that were identified with level 3 confidence are shown in SI Table S2. Most of these compounds include pharmaceuticals and personal care products and industrial chemicals.

In cases where there was more than one compound assigned as the top hit, only a chemical formula (level 4) was assigned. For instance, the feature with an m/z of 143.1310 was assigned a level 4 since multiple compounds generated the same score and were assigned as the top hit in MetFrag. SI Figure S3 shows the fragmentation patterns of three compounds that all had a score of 1.0 in MetFrag. Only one fragment matched between the experimental data and the in silico generated fragments. Since a tentative structure cannot be identified from the results of MetFrag, only a molecular formula of $C_8H_{17}NO$ and level 4 was assigned to this feature. Details for all features that were assigned level 4 are outlined in SI Table S3.

4. CONCLUSIONS

A previously optimized semiautomated workflow was used to process data obtained from the analysis of wastewater and surface water samples from Manila, Philippines. The features detected in the samples were reported according to the Schymanski levels of confidence. A total of 19 compounds were confirmed with reference standards and were assigned level 1. These compounds included several pharmaceuticals such as antibiotics, antifungal compounds, and antihypertensive compounds. In addition, 83 compounds that were not confirmed with reference standards but were confirmed with experimental data from the mass spectra matching library, mzCloud were assigned a level 2. The use of COSMO-RS as an in silico tool to provide additional confidence in predicting level 2 compounds using retention time was demonstrated. Predictions based on retention factors would especially be useful after full investigation of all mass spectral data obtained from both experiments and databases. The COSMOtherm model can be used when one has an idea or a hypothesis of what the compound structure is depending on the collected mass spectral data.

Features that did not have a match with experimental data from mzCloud were fragmented in silico using MetFrag. Tentative structures for 96 compounds were assigned using in silico fragmentation results. A total of 53 compounds were classified as level 3, where the top hit in the candidate list from MetFrag was the structure assigned, while 43 compounds were only assigned a molecular formula (level 4). Nontarget analysis using LC-HRMS is a valuable approach in determining the occurrence of several CECs in environmental samples. The availability of data from mass spectral libraries like mzCloud and tools such as the in silico fragmenter MetFrag, significantly increases confidence in reporting detections in the environment.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsestwater.1c00105>.

Figure S1, Compound Discoverer data processing workflow, Figure S2 Mass spectrum of apocynin from the mzCloud database. Figure S3, Mass spectrum of unknown compound with m/z 143.1310. Figure S4, Extracted ion chromatograms and mass spectra of carbendazim. Table S1, Level 1 compounds detected in wastewater and surface water. Table S2, Level 2 compounds detected in wastewater and surface water. Table S3, Level 3 compounds detected in wastewater and surface water. Table S4, Level 4 compounds detected in wastewater and surface water. Table S5, List of variables used in the construction of the regressions. Table S6, Statistical information for multi-variable regressions. Table S7, Percent error of level 2 compounds as predicted by regressions. Optimized DFR coordinates (PDF)

AUTHOR INFORMATION

Corresponding Author

Diana S. Aga – Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, New York 14260, United States; orcid.org/0000-0001-6512-7713; Email: dianaaga@buffalo.edu

Authors

Luisa F. Angeles – Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, New York 14260, United States; orcid.org/0000-0002-5326-5710

Lahiruni M. Halwatura – Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, New York 14260, United States

Jonathan P. Antle – Department of Chemistry, St. Bonaventure University, St. Bonaventure, New York 14778, United States

Scott Simpson – Department of Chemistry, St. Bonaventure University, St. Bonaventure, New York 14778, United States; orcid.org/0000-0001-8093-1814

Caroline M.B. Jaraula – The Marine Science Institute, University of the Philippines, Diliman, Quezon City 1101, Philippines

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsestwater.1c00105>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge support from the National Science Foundation PIRE-HEARD award number 1545756 and USAID PEER subaward number 2000009924. S.S. thanks the National Science Foundation (Award #1904825) for support of this research, and the Donors of the American Chemical Society Petroleum Research Fund (PRF-58954-UNIS). We also thank Shyrill Mae Mariano from the Marine Science Institute in the University of the Philippines – Diliman who helped with the sample collection and Rebecca Dickman from University at Buffalo for her help on technical work.

REFERENCES

(1) Moschet, C.; Piazzoli, A.; Singer, H.; Hollender, J. Alleviating the Reference Standard Dilemma Using a Systematic Exact Mass Suspect Screening Approach with Liquid Chromatography-High Resolution Mass Spectrometry. *Anal. Chem.* **2013**, *85* (21), 10312–10320.

(2) Kolpin, D. W.; Furlong, E. T.; Meyer, M. T.; Thurman, E. M.; Zaugg, S. D.; Barber, L. B.; Buxton, H. T. Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999–2000: A National Reconnaissance. *Environ. Sci. Technol.* **2002**, *36* (6), 1202–1211.

(3) Pochodylo, A. L.; Helbling, D. E. Emerging investigators series: prioritization of suspect hits in a sensitive suspect screening workflow for comprehensive micropollutant characterization in environmental samples. *Environmental Science: Water Research & Technology* **2017**, *3* (1), 54–65.

(4) Luo, Y.; Guo, W.; Ngo, H. H.; Nghiem, L. D.; Hai, F. I.; Zhang, J.; Liang, S.; Wang, X. C. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci. Total Environ.* **2014**, 473–474, 619–641.

(5) Noguera-Oviedo, K.; Aga, D. S. Lessons learned from more than two decades of research on emerging contaminants in the environment. *J. Hazard. Mater.* **2016**, *316*, 242–251.

(6) Scott, T.-M.; Phillips, P. J.; Kolpin, D. W.; Colella, K. M.; Furlong, E. T.; Foreman, W. T.; Gray, J. L. Pharmaceutical manufacturing facility discharges can substantially increase the pharmaceutical load to U.S. wastewaters. *Sci. Total Environ.* **2018**, *636*, 69–79.

(7) Singh, R. R.; Angeles, L. F.; Butryn, D. M.; Metch, J. W.; Garner, E.; Vikesland, P. J.; Aga, D. S. Towards a harmonized method for the global reconnaissance of multi-class antimicrobials and other pharmaceuticals in wastewater and receiving surface waters. *Environ. Int.* **2019**, *124*, 361–369.

(8) Casado, J.; Santillo, D.; Johnston, P. Multi-residue analysis of pesticides in surface water by liquid chromatography quadrupole-Orbitrap high resolution tandem mass spectrometry. *Anal. Chim. Acta* **2018**, *1024*, 1–17.

(9) Belmonte Vega, A.; Garrido Frenich, A.; Martínez Vidal, J. L. Monitoring of pesticides in agricultural water and soil samples from Andalusia by liquid chromatography coupled to mass spectrometry. *Anal. Chim. Acta* **2005**, *538* (1), 117–127.

(10) Arnnok, P.; Singh, R. R.; Burakham, R.; Pérez-Fuentetaja, A.; Aga, D. S. Selective Uptake and Bioaccumulation of Antidepressants in Fish from Effluent-Impacted Niagara River. *Environ. Sci. Technol.* **2017**, *51* (18), 10652–10662.

(11) Melvin, S. D. Short-term exposure to municipal wastewater influences energy, growth, and swimming performance in juvenile Empire Gudgeons (*Hypseleotris compressa*). *Aquat. Toxicol.* **2016**, *170*, 271–278.

(12) Hendriksen, R. S.; Munk, P.; Njage, P.; van Bunnik, B.; McNally, L.; Lukjancenko, O.; Röder, T.; Nieuwenhuijse, D.; Pedersen, S. K.; Kjeldgaard, J.; Kaas, R. S.; Clausen, P. T. L. C.; Vogt, J. K.; Leekitcharoenphon, P.; van de Schans, M. G. M.; Zuidema, T.; de Roda Husman, A. M.; Rasmussen, S.; Petersen, B.; Bego, A.; Rees, C.; Cassar, S.; Coventry, K.; Collignon, P.; Allerberger, F.; Rahube, T. O.; Oliveira, G.; Ivanov, I.; Vuthy, Y.; Sopheak, T.; Yost, C. K.; Ke, C.; Zheng, H.; Baisheng, L.; Jiao, X.; Donado-Godoy, P.; Coulibaly, K. J.; Jergović, M.; Hrenovic, J.; Karpíšková, R.; Villacis, J. E.; Legesse, M.; Egualé, T.; Heikinheimo, A.; Malania, L.; Nitsche, A.; Brinkmann, A.; Saba, C. K. S.; Kocsis, B.; Solymosi, N.; Thorsteinsdottir, T. R.; Hatha, A. M.; Alebouyeh, M.; Morris, D.; Cormican, M.; O'Connor, L.; Moran-Gilad, J.; Alba, P.; Battisti, A.; Shakenova, Z.; Kiyukia, C.; Ng'eno, E.; Raka, L.; Avsejenko, J.; Běziš, A.; Bartkevics, V.; Penny, C.; Rajandas, H.; Parimannan, S.; Haber, M. V.; Pal, P.; Jeunen, G.-J.; Gemmel, N.; Fashae, K.; Holmstad, R.; Hasan, R.; Shakoob, S.; Rojas, M. L. Z.; Wasyl, D.; Bosevska, G.; Kochubovskii, M.; Radu, C.; Gassama, A.; Radosavljevic, V.; Wuertz, S.; Zuniga-Montanez, R.; Tay, M. Y. F.; Gavačová, D.; Pastuchova, K.; Truska, P.; Trkov, M.; Esterhuysen, K.; Keddy, K.; Cerdà-Cuellar, M.; Pathirage, S.; Norrgren, L.; Örn, S.; Larsson, D. G. J.; Heijden, T. V. d.; Kumburu, H. H.; Sanneh, B.; Bidjada, P.; Njanpop-Lafourcade, B.-M.; Nikiema-Pessinaba, S. C.; Levent, B.; Meschke, J. S.; Beck, N. K.; Van, C. D.; Phuc, N. D.; Tran, D. M. N.; Kwenda, G.; Tabo, D.-a.; Wester, A. L.; Cuadros-Orellana,

- S.; Amid, C.; Cochrane, G.; Sicheritz-Ponten, T.; Schmitt, H.; Alvarez, J. R. M.; Aidara-Kane, A.; Pamp, S. J.; Lund, O.; Hald, T.; Woolhouse, M.; Koopmans, M. P.; Vigre, H.; Petersen, T. N.; Aarestrup, F. M.; The Global Sewage Surveillance project, c. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **2019**, *10* (1), 1124.
- (13) Barber, D. A.; Casquejo, E.; Ybañez, P. L.; Pinote, M. T.; Casquejo, L.; Pinote, L. S.; Estorgio, M.; Young, A. M. Prevalence and correlates of antibiotic sharing in the Philippines: antibiotic misconceptions and community-level access to non-medical sources of antibiotics. *Trop. Med. Int. Health* **2017**, *22* (5), 567–575.
- (14) Saito, N.; Takamura, N.; Retuerma, G. P.; Frayco, C. H.; Solano, P. S.; Ubas, C. D.; Lintag, A. V.; Ribo, M. R.; Solante, R. M.; Dimapilis, A. Q.; Telan, E. O.; Go, W. S.; Suzuki, M.; Ariyoshi, K.; Parry, C. M. Frequent Community Use of Antibiotics among a Low-Economic Status Population in Manila, the Philippines: A Prospective Assessment Using a Urine Antibiotic Bioassay. *Am. J. Trop. Med. Hyg.* **2018**, *98* (5), 1512–1519.
- (15) Andrews, G., Resolving the Water Pollution Crisis in the Philippines: the Implications of Water Pollution on Public Health and the Economy. *Pepperdine Policy Review* **2018**, *10*.
- (16) *Many at Risk of Contracting Diseases from the Poorly Managed Wastewater of 26 Million Filipinos*; World Health Organization, 2017.
- (17) Elfman, L.; Tooke, N. E.; Patring, J. D. M. Detection of pesticides used in rice cultivation in streams on the island of Leyte in the Philippines. *Agricultural Water Management* **2011**, *101* (1), 81–87.
- (18) Navarrete, I. A.; Tee, K. A. M.; Unson, J. R. S.; Hallare, A. V. Organochlorine pesticide residues in surface water and groundwater along Pampanga River, Philippines. *Environ. Monit. Assess.* **2018**, *190* (5), 289.
- (19) Varca, L. M. Pesticide residues in surface waters of Pagsanjan-Lumban catchment of Laguna de Bay, Philippines. *Agricultural Water Management* **2012**, *106*, 35–41.
- (20) *Monitoring and Governance of Persistent Organic Pollutants in Asia*; United Nations University: Tokyo, 2016.
- (21) Singer, H. P.; Wössner, A. E.; Mc Ardell, C. S.; Fenner, K. Rapid Screening for Exposure to “Non-Target” Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling. *Environ. Sci. Technol.* **2016**, *50* (13), 6698–6707.
- (22) Rajsiki, L.; Gómez-Ramos, M. d. M.; Fernández-Alba, A. R. Simultaneous combination of MS2 workflows for pesticide multi-residue analysis with LC-QOrbitrap. *Anal. Methods* **2017**, *9* (15), 2256–2264.
- (23) 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* (20), 12333–12341.
- (24) Reinisch, J.; Klamt, A.; Eckert, F.; Diedenhofen, M. Prediction of the temperature dependence of a polyether-water mixture using COSMOtherm. *Fluid Phase Equilib.* **2011**, *310* (1), 7–10.
- (25) López-Ureña, S.; Torres-Lapasíó, J. R.; Donat, R.; García-Alvarez-Coque, M. C. Gradient design for liquid chromatography using multi-scale optimization. *J. Chromatogr A* **2018**, *1534*, 32–42.
- (26) Baeza-Baeza, J. J.; Ortiz-Bolsico, C.; Torres-Lapasíó, J. R.; García-Alvarez-Coque, M. C. Approaches to model the retention and peak profile in linear gradient reversed-phase liquid chromatography. *J. Chromatogr A* **2013**, *1284*, 28–35.
- (27) Wei, X.; Pang, Z.; Fan, G.; Xu, X.; Wang, L. Simultaneous Prediction of Retention Times and Peak Shapes of Sulfonamides in Reversed-Phase High-Performance Liquid Chromatography. *Trans. Tianjin Univ.* **2018**, *24* (3), 256–262.
- (28) Amos, R. I. J.; Haddad, P. R.; Szucs, R.; Dolan, J. W.; Pohl, C. A. Molecular modeling and prediction accuracy in Quantitative Structure-Retention Relationship calculations for chromatography. *TrAC, Trends Anal. Chem.* **2018**, *105*, 352–359.
- (29) Buszewski, B.; Žuvela, P.; Sagandykova, G.; Walczak-Skierska, J.; Pomastowski, P.; David, J.; Wong, M. W. Mechanistic Chromatographic Column Characterization for the Analysis of Flavonoids Using Quantitative Structure-Retention Relationships Based on Density Functional Theory. *Int. J. Mol. Sci.* **2020**, *21* (6), 2053.
- (30) Haddad, P. R.; Taraji, M.; Szücs, R. Prediction of Analyte Retention Time in Liquid Chromatography. *Anal. Chem.* **2021**, *93* (1), 228–256.
- (31) Sagandykova, G. Perspectives and recent advances in quantitative structure-retention relationships for high performance liquid chromatography. How far are we? *TrAC, Trends Anal. Chem.* **2021**, *141*, 141 2021 v..
- (32) Žuvela, P.; Liu, J. J.; Wong, M. W.; Baczek, T. Prediction of Chromatographic Elution Order of Analytical Mixtures Based on Quantitative Structure-Retention Relationships and Multi-Objective Optimization. *Molecules* **2020**, *25* (13), 3085.
- (33) Kaliszán, R. QSRR: Quantitative Structure-(Chromatographic) Retention Relationships. *Chem. Rev.* **2007**, *107* (7), 3212–3246.
- (34) Taraji, M.; Haddad, P. R.; Amos, R. I. J.; Talebi, M.; Szucs, R.; Dolan, J. W.; Pohl, C. A. Chemometric-assisted method development in hydrophilic interaction liquid chromatography: A review. *Anal. Chim. Acta* **2018**, *1000*, 20–40.
- (35) Domingo-Almenara, X.; Guijas, C.; Billings, E.; Montenegro-Burke, J. R.; Uritboonthai, W.; Aisporna, A. E.; Chen, E.; Benton, H. P.; Siuzdak, G. The METLIN small molecule dataset for machine learning-based retention time prediction. *Nat. Commun.* **2019**, *10* (1), 5811.
- (36) Yang, Q.; Ji, H.; Lu, H.; Zhang, Z. Prediction of Liquid Chromatographic Retention Time with Graph Neural Networks to Assist in Small Molecule Identification. *Anal. Chem.* **2021**, *93* (4), 2200–2206.
- (37) Bade, R.; Bijlsma, L.; Miller, T. H.; Barron, L. P.; Sancho, J. V.; Hernández, F. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis. *Sci. Total Environ.* **2015**, *538*, 934–41.
- (38) Feng, C.; Xu, Q.; Qiu, X.; Jin, Y.; Ji, J.; Lin, Y.; Le, S.; She, J.; Lu, D.; Wang, G. Evaluation and application of machine learning-based retention time prediction for suspect screening of pesticides and pesticide transformation products in LC-HRMS. *Chemosphere* **2021**, *271*, 129447.
- (39) Arp, H. P. H.; Niederer, C.; Goss, K.-U. Predicting the Partitioning Behavior of Various Highly Fluorinated Compounds. *Environ. Sci. Technol.* **2006**, *40* (23), 7298–7304.
- (40) Wang, Z.; MacLeod, M.; Cousins, I. T.; Scheringer, M.; Hungerbühler, K. Using COSMOtherm to predict physicochemical properties of poly- and perfluorinated alkyl substances (PFASs). *Environmental Chemistry* **2011**, *8* (4), 389–398.
- (41) Guardian, M. G. E.; Antle, J. P.; Vexelman, P. A.; Aga, D. S.; Simpson, S. M. Resolving unknown isomers of emerging per- and polyfluoroalkyl substances (PFASs) in environmental samples using COSMO-RS-derived retention factor and mass fragmentation patterns. *J. Hazard. Mater.* **2021**, *402*, 123478.
- (42) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1* (1), 101–122.
- (43) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–2235.
- (44) Chemistry, R. S. o. ChemSpider. <http://www.chemspider.com/>.
- (45) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminf.* **2016**, *8* (1), 3.
- (46) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **2000**, *28* (1), 27–30.
- (47) Angeles, L. F.; Islam, S.; Aldstadt, J.; Saqeeb, K. N.; Alam, M.; Khan, M. A.; Johura, F.-T.; Ahmed, S. I.; Aga, D. S. Retrospective

suspect screening reveals previously ignored antibiotics, antifungal compounds, and metabolites in Bangladesh surface waters. *Sci. Total Environ.* **2020**, 712, 136285.

(48) Angeles, L. F.; Singh, R. R.; Vikesland, P. J.; Aga, D. S. Increased Coverage and High Confidence in Suspect Screening of Emerging Contaminants in Global Environmental Samples. *J. Hazard. Mater.* **2021**, 414, 125369.

(49) 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 (4), 2097–2098.

(50) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, 162 (3), 165–169.

(51) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, 33 (12), 8822–8824.

(52) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, 38 (6), 3098–3100.

(53) Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010**, 133 (13), 134105.

(54) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, 7 (18), 3297–3305.

(55) Michailoudi, G.; Hyttinen, N.; Kurtén, T.; Prisle, N. L. Solubility and Activity Coefficients of Atmospheric Surfactants in Aqueous Solution Evaluated Using COSMOtherm. *J. Phys. Chem. A* **2020**, 124 (2), 430–443.

(56) Youn, H. J.; Kim, S. Y.; Park, M.; Jung, W. H.; Lee, Y. W.; Choe, Y. B.; Ahn, K. J. Efficacy and Safety of Cream Containing Climbazole/Piroctone Olamine for Facial Seborrheic Dermatitis: A Single-Center, Open-Label Split-Face Clinical Study. *Ann. Dermatol.* **2016**, 28 (6), 733–739.

(57) Chen, Z.-F.; Ying, G.-G. Occurrence, fate and ecological risk of five typical azole fungicides as therapeutic and personal care products in the environment: A review. *Environ. Int.* **2015**, 84, 142–153.

(58) Geroy, L. S. A. Economic evaluation for first-line anti-hypertensive medicines: applications for the Philippines. *Cost Eff Resour Alloc* **2012**, 10 (1), 14–14.

(59) Dimzon, I. K. D.; Morata, A. S.; Müller, J.; Yanela, R. K.; Lebertz, S.; Weil, H.; Perez, T. R.; Müller, J.; Dayrit, F. M.; Knepper, T. P. Trace organic chemical pollutants from the lake waters of San Pablo City, Philippines by targeted and non-targeted analysis. *Sci. Total Environ.* **2018**, 639, 588–595.

(60) Weeks, J.; Guiney, P.; Nikiforov, A. Assessment of the environmental fate and ecotoxicity of N,N-diethyl-m-toluamide (DEET). *Integr. Environ. Assess. Manage.* **2012**, 8 (1), 120–134.

(61) Giacomazzi, S.; Cochet, N. Environmental impact of diuron transformation: a review. *Chemosphere* **2004**, 56 (11), 1021–1032.

(62) Curran, M. P.; Perry, C. M. Amisulpride. *Drugs* **2001**, 61 (14), 2123–2150.

(63) Snyder, L.; Dolan, J. THE LINEAR-SOLVENT-STRENGTH MODEL OF GRADIENT ELUTION. *Advances in Chromatography* **1998**, 38, 115–187.

(64) Izydorczak, A. M.; Gross, M. S.; Aga, D. S.; Simpson, S. Accurate Prediction of Gas Chromatographic Retention Times via Density Functional Theory Calculations: A Case Study Using Brominated Flame Retardants. *ChemistrySelect* **2020**, 5 (8), 2476–2481.

(65) Poole, C. F.; Gunatilleka, A. D.; Poole, S. K. In search of a chromatographic model for biopartitioning. *Adv. Chromatogr* **2000**, 40, 159–230.

(66) Chen, B.; Siepmann, J. I. Microscopic Structure and Solvation in Dry and Wet Octanol. *J. Phys. Chem. B* **2006**, 110 (8), 3555–3563.

(67) Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, 19 (6), 716–723.

(68) Stone, M. Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)* **1979**, 41 (2), 276–278.

(69) Bączek, T.; Kaliszan, R. Predictive approaches to gradient retention based on analyte structural descriptors from calculation chemistry. *Journal of Chromatography A* **2003**, 987 (1), 29–37.

(70) Vergeynst, L.; Van Langenhove, H.; Demeestere, K. Balancing the False Negative and Positive Rates in Suspect Screening with High-Resolution Orbitrap Mass Spectrometry Using Multivariate Statistics. *Anal. Chem.* **2015**, 87 (4), 2170–2177.