Occurrence and Distribution of Pharmaceuticals and their Transformation Products in Luxembourgish Surface Waters

Randolph R. Singh*, Adelene Lai, Jessy Krier, Todor Kondić, Philippe Diderich, Emma L. Schymanski*

ABSTRACT

Pharmaceuticals and their transformation products (TPs) are continuously released into the aquatic environment via anthropogenic activity. To expand knowledge on the presence of pharmaceuticals and their known TPs in Luxembourgish rivers, several samples collected during routine monitoring events between 2019 and 2020 were investigated using non-target analysis. Water samples were concentrated using solid phase extraction, then analyzed using liquid chromatography coupled to a high resolution mass spectrometer (LC-HRMS). Suspect screening was performed using several open source computational tools and resources including Shinyscreen (https://git-r3lab.uni.lu/eci/shinyscreen/), MetFrag (https://msbi.ipb-halle.de/MetFrag/), PubChemLite (https://zenodo.org/record/4432124), and MassBank (https://massbank.eu/MassBank/). A total of 94 pharmaceuticals: 88 confirmed at a Level 1 confidence (86 of which could be quantified, 2 compounds too low to be quantified), and 6 identified at Level 2a were found to be present in Luxembourg rivers. Pharmaceutical TPs (12) were also found at a Level 2a confidence. The pharmaceuticals were present at median concentrations up to 214 ng/L, with caffeine having a median concentration of 1424 ng/L. Antihypertensive drugs (15), psychoactive drugs (15) and antimicrobials (8) were the most detected groups of pharmaceuticals. A spatio-temporal

analysis of the data revealed areas with higher concentrations of the pharmaceuticals, as well as differences in pharmaceutical concentrations between 2019 and 2020. The results of this work will help guide activities for improving water management in the country and set baseline data for continuous monitoring and screening efforts, as well as for further open data and software developments.

Keywords: pharmaceuticals, surface water, suspect screening, HRMS, transformation products, cheminformatics, open source, non-target screening

INTRODUCTION

The geography and history of Luxembourg have distinct implications on its environment and water quality: it borders Belgium, France, and Germany, and its rivers feed into the Rhine basin. Luxembourg has vineyards lining the Moselle River, agricultural activity in the north of the country, and a population largely centered in the capital, which together brings in a significant and varied chemical load into the environment. Previous studies have reported the presence of analgesics, antimicrobials, and estrogens in Luxembourgish surface water.[1-3] Aside from providing data on the level of xenobiotics in Luxembourgish waters, these studies have also demonstrated that the presence of these chemicals are due to inputs from land use, accidental spillage, wastewater effluent, and long range transport.[1, 3-6] Other studies have reported the measurement of 14 pesticides and their transformation products (TPs) both in surface water and drinking water.[3, 7] The Luxembourg Water Management Agency (Administration de la Gestion de l'Eau,

hereafter AGE), in compliance with the European Union Water Framework Directive (WFD), monitors different organic contaminants in Luxembourgish surface water.[8] Among the 92 compounds included in the targeted analysis performed by AGE, five are pharmaceuticals: carbamazepine, diclofenac, ibuprofen, ketoprofen, and lidocaine, while the rest the targeted organic contaminants are pesticides and related compounds.

As there are conceivably more pharmaceuticals than the 5 included in targeted monitoring that enter into the environment, it is important to determine which other pharmaceuticals may be present, to gain a more holistic idea of the pharmaceutical loading in Luxembourgish surface waters. The presence of pharmaceuticals in the aquatic environment poses a threat to human and environmental health due to exposure to either the pharmaceuticals themselves or their metabolites and TPs, which may still possess bioactivity.[9-11] These chemicals have potential negative impacts on human health and the environment through different routes of exposures.[12, 13]

There are many approaches to account for the presence of xenobiotics in the environment, but recently increasing effort has been in the use of non-targeted analysis (NTA) and/or suspect screening using high resolution mass spectrometry (HRMS) specifically to support risk assessment efforts and regulatory institutions.[14-16] HRMS enables measurement of known pollutants, discovery of contaminants of emerging concern as well as retrospective screening.[17] However, setting up analyses, both experimentally and computationally, is no trivial matter. Despite these challenges, the information that can be obtained from such analyses has a wide breadth of utility, especially for environmental studies. NTA and suspect screening are effective techniques for the monitoring and discovery of xenobiotics in the aquatic environment.[17-20]

Nevertheless, the interpretation of HRMS data presents challenges that highlight the need for computational tools to enable the proper identification and annotation of the chemical components in environmental matrices.[21]

MetFrag (https://ipb-halle.github.io/MetFrag/) [22] is an open source tool for compound identification, including *in silico* fragmentation, mass spectral matching, and metadata functions.[23, 24] MetFrag enables spectral matching with experimental data via the spectral library MassBank of North America (MoNA, https://mona.fiehnlab.ucdavis.edu)[25], and prioritization using metadata from various sources. MetFrag first retrieves candidates by exact mass or molecular formula from one of many available compound databases. PubChem (https://pubchem.ncbi.nlm.nih.gov/) [26] is an open chemistry database at the National Institutes of Health (NIH) containing more than 110M compounds.[27] While such a large database provides access to many chemicals, it can lead to (tens of) thousands of candidates per unknown when performing non-target screening of hundreds of masses.[28] For this work, an early version of PubChemLite was used, which contains ~300,000 compounds selected to be highly relevant for environmental investigations based on annotation content, including information relevant for pharmaceuticals.[28, 29] PubChemLite has been shown to outperform other databases such as the whole of PubChem and CompTox for well-known chemicals [28] and delivers important metadata that can be used during identification with MetFrag. PubChem and PubChemLite also contain information on environmental TPs contributed via the NORMAN Suspect List Exchange (https://www.norman-network.com/nds/SLE/).[28, 30] This information can be exploited programmatically during

the environmental screening of hundreds of compounds, together with their transformation products.

Considering the previously reported presence of chemicals in Luxembourg's environment[2, 4-7] and the widespread use of chemicals in daily life, a large number of compounds could be considered as potential environmental pollutants in Luxembourg. This work focuses on the presence of pharmaceuticals and known pharmaceutical TPs present in Luxembourg surface water systems using a mixture of instrumental measurements and cheminformatics approaches.

MATERIALS AND METHODS

Sample collection and processing. Surface water samples (1L) were collected every four weeks, whenever physically possible, from nine different locations in Luxembourg from April to November 2019 (Figure 1) and eight different locations from April to August in 2020 in accordance with the tri-annual sampling strategy employed at AGE. In this strategy, four locations monitored in compliance with the WFD are consistently sampled every four weeks (locations 1-4, Figure 1) while the other locations throughout Luxembourg are divided into 3 regions and are alternately sampled during a 3-year cycle. The samples were filled in 1000 mL amber glass bottles and stored for up to one week at 5°C ± 3°C in the dark until extraction. A method blank was prepared every month to account for potential contamination from sample handling using ultrapure water. Solid phase extraction (SPE) was performed using Atlantic® HLB SPE Disks from Horizon (Salem, NH, USA) with a 47mm diameter. The disks were conditioned twice for one minute using acetonitrile, and then twice for one minute using milliQ water. The samples

were pumped through each disk at a flow rate of roughly 30 mL/min, using the SPE-DEX 47900 system from Horizon (Salem, NH, USA). Sample loading was followed by washing the disks twice for 1min with milli-Q® water and drying by airflow for 15 min. The analytes were eluted for 1 min with cyclohexane, followed by an acetone elution for 1 min, then 4 times for 1 min with acetonitrile. After each elution step, the disks were air-dried for 1 min. The combined extracts were reduced to dryness under nitrogen flow in a water bath heated to 40°C. The samples were re-suspended in 2ml acetonitrile/water (10:90) by sonication for 5 min. Remaining particles were removed by passing the extracts through a 0.7 um glass-fiber filter (Sartorius, Brussels, BE) into 2mL amber glass LC-MS vials.

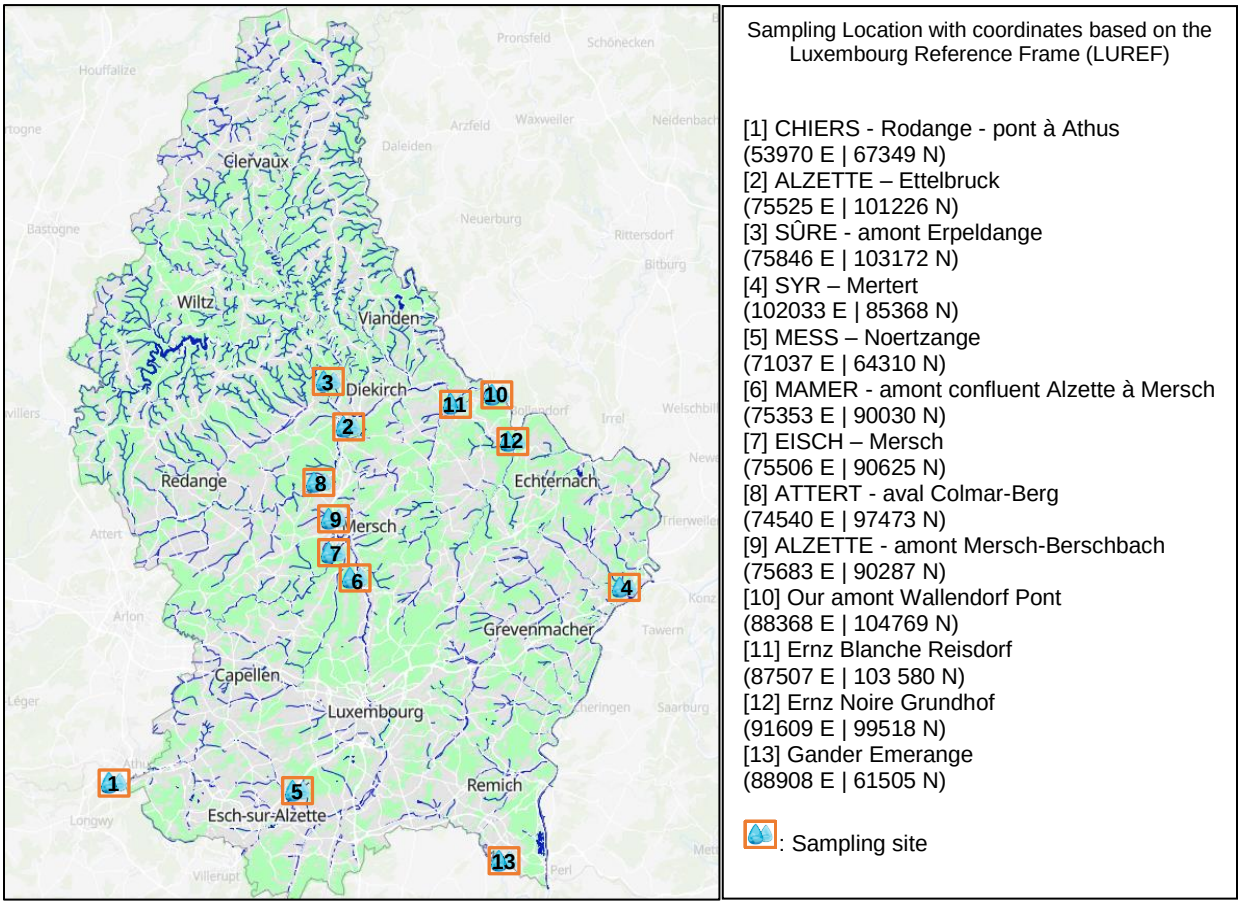


Figure 1. Sampling locations and their respective coordinates. Sampling locations 1-4 were sampled from 2019 to 2020, sampling locations 5-9 were sampled only in 2019, and sampling locations 10-13 were sampled only in 2020. Map generated using https://www.geoportail.lu/en/ . © MapTiler © OpenStreetMap contributors

LC-HRMS Analysis. LC-HRMS analysis was performed on a Thermo QExactive HF Mass Spectrometer equipped with a Waters Acquity UPLC BEH C_{18} column (1.7µm, 2.1x150mm) using both positive and negative electrospray ionization with the following spray settings (positive/negative): sheath gas flow rate (45/60 arbitrary units, AU), auxiliary gas flow rate (10/25 AU), sweep gas flow rate (2/2 AU), spray voltage (3.5/3.6 kV), capillary temperature (320/300 °C), S lens RF (50/50 AU), and auxiliary gas temperature (300/370 °C). Mobile phase A (water with 0.1% formic acid) and B (methanol) were mixed using the following LC gradient starting at 90A/10B at 0 min, 90/10 at 2 min, 0/100 at 15 min, 0/100 at 20 min, 90/10 at 21 min, and ending with 90/10 at 30 min at a flow rate of 0.200 mL/min. The following data-dependent (dd-)MS2 settings (in display order of instrumental acquisition method) were used: resolution (120,000 at *m/z* 200), automatic gain control (AGC) target (1.0×10^6), maximum injection time (IT): (70 ms), and scan range (*m/z*=60-900). For the selected ion monitoring dd-MS2/ddSIM the following were used: resolution (30,000 at *m/z* 200), AGC target (5.0×10^5), maximum IT (70 ms), loop count (5), Top N (5), isolation window (1.0 Da), (N)CE (30). Lastly, the following dd settings were used: minimum AGC target (8.0×10^3), intensity threshold (1.1×10^5), apex trigger (4-6 s), exclude isotopes (On), and dynamic exclusion (10.0 s). The instrument was calibrated and optimized every time an analysis was performed using manufacturer settings to ensure consistent performance throughout the 2-year study.

Suspect Screening. Suspect screening was performed using two suspect lists. The first list contains 816 unique pharmaceutical compounds (Supporting Information, Table S1 CNS 'Caisse Nationale de Santé' Suspects, also available on the NORMAN Suspect List Exchange, NORMAN-SLE) [30, 31] that were curated from the Luxembourgish National

Health Fund's "List of marketed medications in Luxembourg".[32] These drugs have marketing authorisation in Luxembourg from the Ministry of Health and are therefore potentially in use domestically. For suspect screening, MS-Ready SMILES of these compounds were obtained via the EPA CompTox Chemistry Dashboard's batch search function.[33, 34] Using MS-ready SMILES as a structural identifier ensures that the structure being used for data analysis is consistent with what is measured by the mass spectrometer, and at the same time remains traceable within online chemical databases.[34]

The second suspect list consists of 82 pharmaceutical TPs. These TPs were derived from two sources: PubChem[28], and a recent study by Anliker *et al*. [18]. From PubChem, TPs were obtained from the Transformations table of a given compound (where available) using R scripts [35] written to programmatically download transformation product information [36]. The TP information in PubChem originates from the NORMAN Suspect List Exchange.[28, 30] Sixty-seven TPs were extracted from PubChem in this way (coming from a total of 53 parents – 44 parents were on the original CNS list of 816 parent compounds, while the remaining 9 parents are actually themselves TPs with reciprocal transformations). The remaining 15 TPs were obtained from Anliker *et al*. [18]. Curation of the final suspect list involved deduplication and multiple steps of interconversion between chemical identifiers (e.g., CAS to PubChem CID, InChIKey to CID) using PubChem's Identifier Exchange Service [37] to facilitate compound comparisons and ensure that the final list of 82 TPs was unique. Then, the final SMILES ("parent SMILES" in PubChem terms, "MS-ready" SMILES in CompTox terms) were retrieved. More information and the full R code is available in the SI and on GitLab as a Jupyter Notebook [38].

Prescreening was performed using Shinyscreen (https://git-r3lab.uni.lu/eci/shinyscreen)[39], an open source and freely available mass spectral processing software developed in house to extract MS1 data and the associated MS2 events and spectra. The following settings for extraction and automatic quality control were used: coarse precursor *m/z* error (±0.5 Da), fine precursor m/z error (±2.5 ppm), extracted ion chromatogram (EIC) *m/z* error (±0.001 Da), retention time tolerance (±0.5 min), MS1 intensity threshold (1.0×10^5), MS2 intensity threshold relative to MS1 peak intensity (0.05), signal-to-noise ratio (3), retention time shift tolerance (±0.5 min). Features that passed QC through manual curation including peak shape, peak width, peak intensity, and alignment of the MS1 and MS2 peaks were then analyzed using MetFrag to achieve tentative identifications. Scripts used for this work are available on GitLab.[38] PubChemLite was used as database, available as a local .csv file [29], to find chemicals that match the exact mass (within 5 ppm) of the suspect pharmaceutical. Both *in silico* fragmentation (mzabs=0.001, frag_ppm=5) and experimental MS/MS matching through MoNA records (built within MetFrag) were performed to obtain the fragmenter (scoring term 1) and MoNA (scoring term 2) scores. Metadata were also collected for the candidates by querying the database for patent count (scoring term 3), number of PubMed references (scoring term 4), PubChem annotation count (scoring term 5), Pharmacology and Biochemistry information (scoring term 6), and Drug and Medication Information (scoring term 7). Candidates were ranked and given a score per category normalized to 1, then added together to obtain the max_Score, with the highest possible score = 7. A more detailed explanation of the parameters used is available elsewhere.[28, 40] Annotation confidence levels were determined using the scheme described by

Schymanski *et al*.[41] Level 2a compounds were assigned when the MoNA score was greater than or equal to 0.9. Level 1 identifications were achieved using authentic standards and the ENTACT mixtures [42], available in-house and analyzed using the same chromatographic method used for sample analysis. The ENTACT mixtures were obtained from participation in the EPA's Non-Targeted Analysis Collaborative Trial.[42] Retention times were considered a match if the difference was less than 0.2 minutes. The compound classification for the compounds identified was obtained by consulting PubChem's 'Drug and Medication Information' section, based on a specific drug's therapeutic use or function. Level 3 confidence was given for compounds with max_Score > 6.0 but with MoNA scores less than 0.9 (103 compounds), however, the scope of the paper has been limited to Level 2a and Level 1 chemicals at this stage due to their higher confidence.

Where reference standards were available, the concentration of the pharmaceuticals were quantified using an external calibration curve ranging from 1ppb to 1000ppb spanning the linear dynamic range for the compounds quantified. Tracefinder (Thermo Scientific, version 5.1) was used for automatic peak integration and generation of the calibration curve. Concentrations below 1ppb were reported to be below quantifiable range. After compound identification and quantification, a spatio-temporal analysis was performed to determine whether there were specific areas with higher pharmaceutical loading and/or monthly variability. The concentration of pharmaceuticals in surface waters is influenced by many factors such as precipitation, volume, wastewater effluent discharge, as well as significant changes in cross-border mobility in 2020 due to the pandemic (a dominating factor in Luxembourg where half of the workforce live outside the country). As a result, the spatial and temporal comparisons are limited to uncorrected

concentration values here and should be interpreted accordingly. For spatial analysis, the median concentration of the identified compound across the different months was calculated and presented by sampling year. For temporal analysis, the median concentration of the identified compound across locations 1-4 were used, as these locations were sampled consistently irrespective of sampling year. A boxplot was also constructed to see which pollutants are consistently high and to show the difference in detected concentrations between 2019 and 2020. Heat maps and boxplots were generated using custom-made, openly-accessible scripts in R.[43] Results were compared to pharmaceuticals found in the Meuse (Belgian and Dutch section) and Rhine (German section) rivers, which all have Luxembourgish rivers as tributaries.

RESULTS AND DISCUSSION

Identification of pharmaceuticals and their TPs. After performing LC-HRMS analysis coupled with cheminformatics tools, 88 compounds were confirmed at Level 1 confidence; 86 of these could be quantified. Amantadine and 8-hydroxyquinoline concentrations were too low to be quantified. A further 6 compounds were identified at Level 2a. These results are summarized in Table 1 and 2. Among the detected compounds, only seven were detected in both positive and negative ionization: diclofenac, fluconazole, irbesartan, losartan, niflumic acid, oxazepam, and valsartan (further identifiers are provided in the SI, Tables S1 & S2). In terms of pharmaceutical class, many of the compounds identified in this work belong to drugs for the management of heart related diseases (15), psychoactive drugs (15), antimicrobials (8), and drugs for the management of pain (8). All 5 chemicals monitored by AGE were also detected in this study.

Table 1. Summary of pharmaceuticals and pharmaceutical transformation products in positive mode found in Luxembourgish river water. An extended version with structural information is available in Table S2 Pharma IDs.

m/z, $[M+H]^+$	Name	t_r, min	Level	MetFrag Score	MoNA score	PubChem CID
253.097	3-Hydroxycarbamazepine	14.9	1	-	-	135290
152.0706	Acetaminophen	8	1	6.54	0.9998	1983
152.1434	Amantadine	12.1	1	6.96	0.9980	2130
370.1795	Amisulpride	10.7	1	7	0.9993	2159
278.1903	Amitriptyline	14.3	1	5.83	0.9876	2160
267.1703	Atenolol	8.1	1	6.94	0.9363	2249
119.0604	Benzimidazole	2.5	1	6	0.9974	5798
326.2326	Bisoprolol	13.7	1	7	0.9997	2405
195.0877	Caffeine	11.2	1	6.92	0.9970	2519
237.1023	Carbamazepine	15.7	1	7	0.9999	2554
192.0768	Carbendazim	10.2	1	6.97	0.9999	25429
380.2544	Celiprolol	13	1	6.13	0.9934	2663
389.1627	Cetirizine	16.4	1	7	0.9999	2678
748.4842	Clarithromycin	16.2	1	7	0.9985	4663848
425.1871	Clindamycin	14.5	1	7	0.9985	2786
315.1623	Clomipramine	16.3	1	7	0.9993	2801
300.1594	Codeine	9.1	2a	6.81	0.9509	2828
177.1023	Cotinine	2.4	1	5.05	0.9896	408
296.024	Diclofenac*	18.6	1	7	0.9995	3033
415.1686	Diltiazem	14.9	1	7	0.9990	3076
271.1805	Doxylamine	10.4	1	7	0.9986	3162
330.0804	Epoxiconazole	18.0	1	-	-	3317081
415.1451	Flecainide	14.3	1	6.85	0.9984	3356
307.1114	Fluconazole*	13	1	6.99	0.9901	3365
821.8876	Iohexol	5.7	1	6.9	0.9062	3730
429.2397	Irbesartan*	16.6	1	7	0.9993	3749
255.1016	Ketoprofen	16.9	1	6.93	0.9546	3825
256.0151	Lamotrigine	12.9	1	6.97	0.9693	3878
235.1805	Lidocaine	11.3	1	6.99	1.0000	3676
407.221	Lincomycin	10.6	1	7	0.9993	3928
321.0192	Lorazepam	16.5	2a	6.77	0.9952	3958
423.1695	Losartan*	16.5	1	7	0.9999	3961
180.1747	Memantine	1		-	-	4054
130.1087	Metformin	1.9	1	6.99	0.9856	4091
310.2166	Methadone	15.6	1	6.85	0.975	4095
300.1473	Metoclopramide	11.5	1	7	0.9999	4168
268.1907	Metoprolol	12.4	1	6.94	0.9357	4171
172.0717	Metronidazole	6.9	1	7	0.9983	4173
266.1652	Mirtazapine	12.2	1	6.69	0.9831	4205
269.1051	Moclobemide	11.8	1	7	0.9999	4235
286.1438	Morphine	17.8	2a	5.43	0.9943	4253

m/z, $[M+H]^+$	Name	t_r, min	Level	MetFrag Score	MoNA score	PubChem CID
231.1016	Naproxen	17.3	1	6.48	0.8966	1302
123.0553	Niacinamide	2.5	1	6.94	0.9871	936
163.123	Nicotine	2.2	1	6.74	0.9952	942
124.0393	Nicotinic acid	2.5	1	5.9	0.9035	938
283.0689	Niflumic acid*	18.8	2a	7	0.9989	4488
264.1958	O-desmethylvenlafaxine	12.2	1	-	-	125017
287.0582	Oxazepam*	16.7	1	6.62	0.9172	4616
384.0824	Pantoprazole	14.6	1	6.98	0.9783	4679
369.2384	Perindopril	15	1	7	0.9996	4169159
189.1023	Phenazone	12.1	1	6.7	0.9578	2206
166.0863	Phenylalanine	6.3	1	6.95	0.9997	994
253.0977	Phenytoin	15.5	1	-	-	1775
286.1438	Piperine	17.8	1	5.43	0.9943	4840
260.1645	Propranolol	14.4	1	6.99	0.9932	4946
325.1911	Quinine	11.5	1	6.74	0.8115	1065
315.1486	Ranitidine	7.9	1	6.99	0.9944	3001055
304.1543	Scopolamine	10.1	1	-	-	5184
408.1254	Sitagliptin	12.6	1	6.99	0.9889	11306691
273.1267	Sotalol	6.1	1	6.92	0.9188	5253
251.0597	Sulfadiazine	8	1	6.98	0.9819	5215
254.0594	Sulfamethoxazole	12.1	1	6.94	0.9766	5329
342.1482	Sulpiride	7.1	1	7	0.9996	5355
515.2442	Telmisartan	16.3	1	6.99	0.9929	65999
202.0434	Thiabendazole	11.2	1	6.59	0.9743	5430
329.153	Tiapride	9	1	7	0.9999	5467
317.1642	Timolol	12.4	1	6.98	0.9840	5478
264.1958	Tramadol	12.3	1	7	0.9999	5523
291.1452	Trimethoprim	10.7	1	6.78	0.9619	5578
436.2343	Valsartan*	17.4	1	6.97	0.9717	5650
278.2115	Venlafaxine	14	1	6.9	0.9925	5656
130.0863	Vigabatrin	1.7	2a	5.32	0.9998	5665
304.202	Vildagliptin	8.1	1	7	0.9999	5251896
309.1122	Warfarin	17.4	1	6.94	0.9410	54678486

Note: Table S2 Pharma IDs in the Supporting Information provides the same information but with the corresponding SMILES strings. t_r = retention time

*(found in both positive and negative mode)

Table 2. Summary of pharmaceuticals and pharmaceutical transformation products in negative mode found in Luxembourgish river water. An extended version with structural information is available in SI.

m/z, [M-H]-	Name	t_r, min	level	MetFrag Score	MoNA score	PubChem CID
144.0455	8-Hydroxyquinoline	13.2	1	5.55	0.4714	1923
180.0334	Acamprosate	2.33	1	-	-	71158
220.9809	Acetazolamide	8.5	1	6.99	0.9888	1986
135.0310	Allopurinol	3.46	1	-	-	135401907
179.035	Aspirin	13.7	2a	5.99	0.9964	2244
429.0538	Bicalutamide	16.7	1	6.98	0.9868	2375
287.0247	Ciprofibrate	18.1	1	5.7	0.0000	2763
294.0094	Diclofenac*	18.6	1	7	0.9972	3033
423.1384	Eprosartan	14.2	1	6.83	0.8289	5281037
288.1594	Etodolac	18.5	1	-	-	3308
280.0591	Flufenamic acid	19.2	1	6	0.9998	3371
329.0004	Furosemide	14.7	1	6.94	0.9370	3440
295.9572	Hydrochlorothiazide	8.2	1	7	0.9972	3639
427.2252	Irbesartan*	16.6	1	7	0.9992	3749
269.0543	Leflunomide	17.6	1	5.79	0.0000	3899
421.1549	Losartan*	16.4	1	6.98	0.9844	3961
270.2075	N-Dodecanoyl-N-methylglycine	17.7	1	4.47	0.0000	7348
281.0543	Niflumic acid*	18.8	2a	6.99	0.9944	4488
187.0976	Nonanedioic acid	14.8	1	6	0.0000	2266
285.0436	Oxazepam*	16.6	1	5.27	0.0000	4616
151.0261	Oxypurinol	3.2	1	5.1	0.0246	1188
204.1241	(dex)Panthenol	8.2	1	6.75	0.7626	4678
137.0244	Salicylic acid	14.9	1	7	0.9997	338
117.0193	Succinic acid	3.4	1	6.08	0.9995	1110
179.0574	Theophylline	10.1	1	6.67	0.8883	2153
434.2198	Valsartan*	17.4	1	6.98	0.9830	5650

Note: Table S2 Pharma IDs in the Supporting Information provides the same information but with the corresponding SMILES strings. tr = retention time

*(found in both positive and negative mode)

Two TPs (3-hydroxycarbamazepine and O-desmethylvenlafaxine) were identified with Level 1 confidence, while twelve TPs were identified at Level 2a confidence and are listed including their parent compounds in parentheses: 4-acetamidoantipyrine (metamizole), 4-aminoantipyrine (metamizole), clopidogrel carboxylic acid (clopidrogel), cotinine (nicotine), D617 (verapamil), ritalinic acid (methylphenydate), fenofibric acid

(fenofibrate), flucytosine (emtricitabine), guanylurea (metformin), morphine (codeine), N4-acetylsulfamethoxazole (sulfamethoxazole), 4-hydroxydiclofenac (diclofenac). Flucytosine on its own is used as an antifungal agent, while morphine can be used as the parent compound for pain management. In addition, two TPs (2-hydroxycarbamazepine and 10,11-dihydroxycarbamazepine) were tentatively identified (Level 3) during the parent pharmaceutical screening because they were isobaric with some parent pharmaceuticals.

Spatio-temporal distribution of pharmaceuticals in Luxembourg. The median concentrations of the different compounds identified in this work, irrespective of ionization polarity, were plotted to generate the spatial (N = 6 timepoints for 2019, N = 5 timepoints for 2020) and temporal (N = 4 sampling points) heat maps presented in Figure 2, 3, and 4 respectively. Tables S3 (negative mode) and S4 (positive mode) in the SI summarize the individual concentration of each pharmaceutical quantified from 2019 to 2020 from each location. The spatial heat maps (Figures 2 and 3) for both 2019 and 2020 consistently show that *Chiers-Rodange-pont à Athus* [location 1, Figure 1], followed by *Alzette-Ettelbruck* [location 2, Figure 1] and *Alzette-Mersch-Berschbach* [location 9, Figure 1] have higher levels of pharmaceutical contamination.

The Chiers river receives effluent from the Petange wastewater treatment plant (capacity: 70,000 population equivalents), which is close to the *Chiers-Rodange-pont à Athus* sampling point. This proximity is likely one of the reasons why *Chiers-Rodange-pont à Athus* was found to have the highest concentration of pharmaceuticals within this study. In comparison, both *Alzette-Ettelbruck* and *Alzette-Mersch-Berschbach* are downstream of the Beggen wastewater treatment plant [44] (capacity: 210,000 population

293 equivalents), which receives sewage from Luxembourg City, the biggest and most
294 populated city in Luxembourg. Despite the bigger capacity, both sampling points are not
295 as close to the source as the *Chiers* location and thus may experience dilution. The lowest
296 median concentrations for the pharmaceuticals quantified in this study were found at
297 *Eisch-Mersch* (2019, location 7 in Figure 1), *SÛRE - amont Erpeldange* (2020, location
298 3, Figure 1), and *Our amont Wallendorf Pont* (2020, location 10, Figure 1).

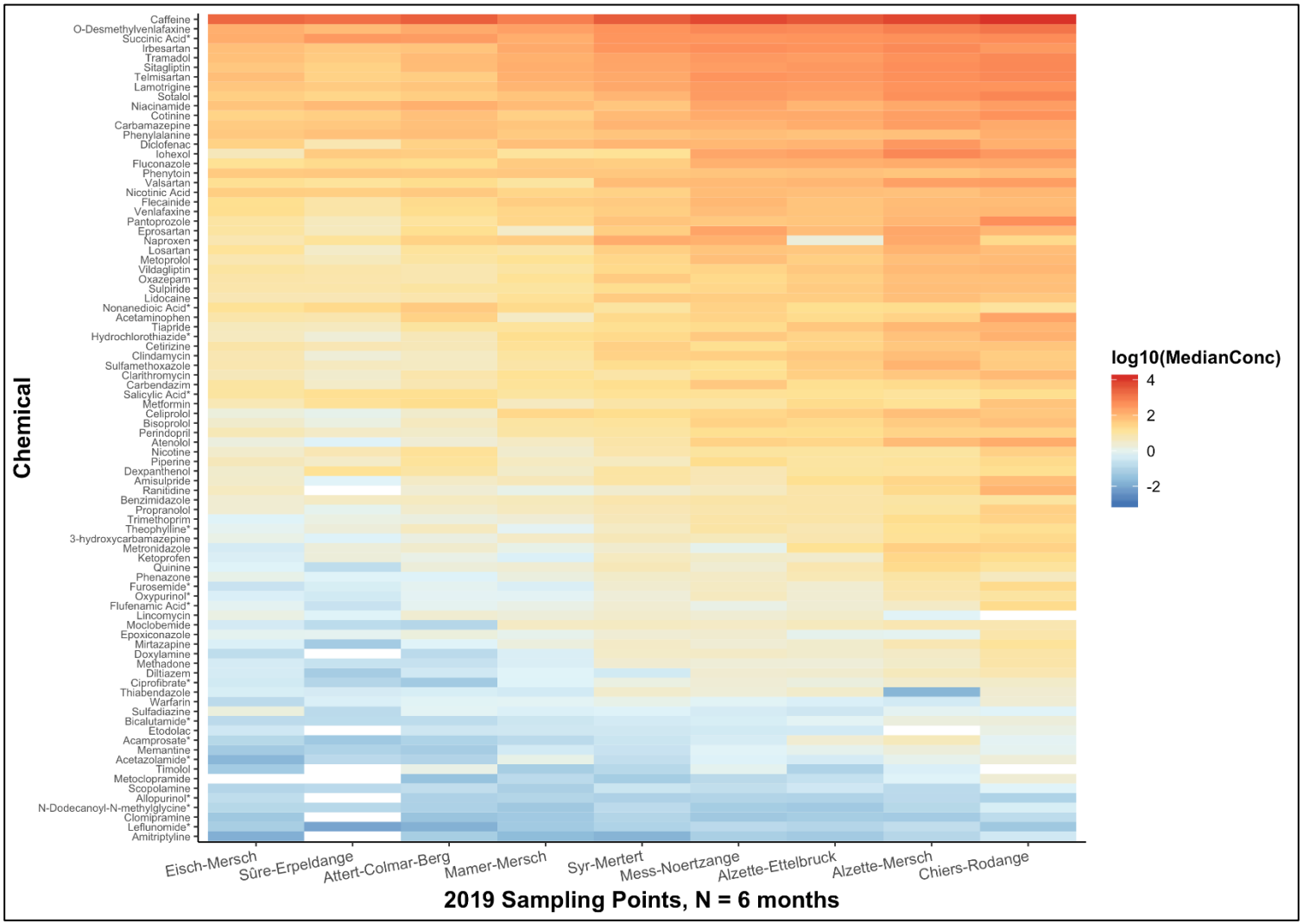


Figure 2. Spatial heat map showing median concentration values (original units: ng/L) per compound measured per sampling location over 6 months in 2019, plotted using a base-10 logarithmic scale. Median values were calculated across the concentrations measured over the relevant months of sampling for the respective compound and location. Zero-value median concentrations are indicated by grey-shaded boxes. White boxes indicate that there were no concentration values within the quantification range. All compounds were measured in positive mode except for those marked with an asterisk, which were measured in negative mode.

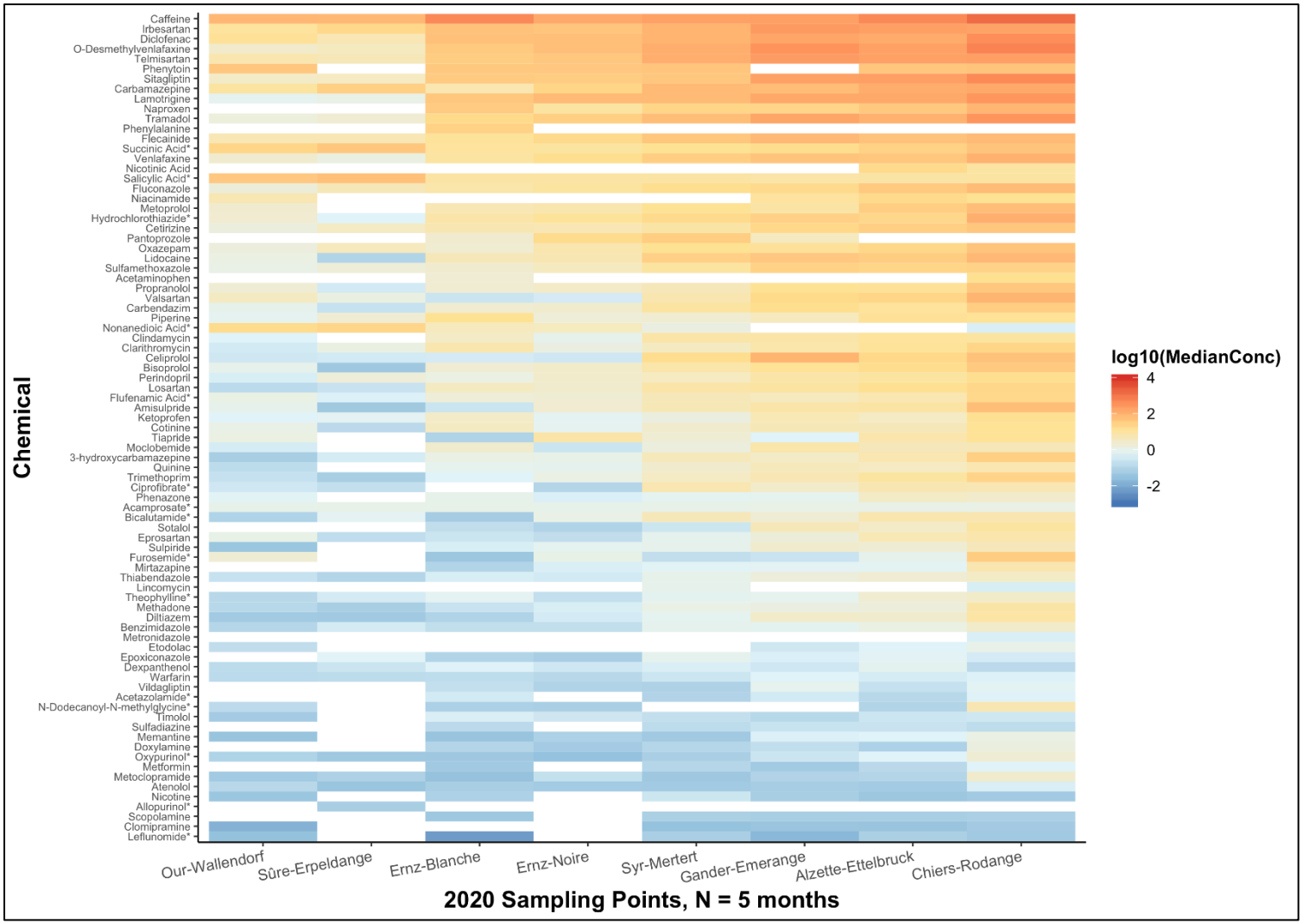


Figure 3. Spatial heat map showing median concentration values (original units: ng/L) per compound measured per sampling location over 5 months in 2020, plotted using a base-10 logarithmic scale. Median values were calculated across the concentrations measured over the relevant months of sampling for the respective compound and location. Zero-value median concentrations are indicated by grey-shaded boxes. White boxes indicate that there were no concentration values within the quantification range. All compounds were measured in positive mode except for those marked with an asterisk, which were measured in negative mode.

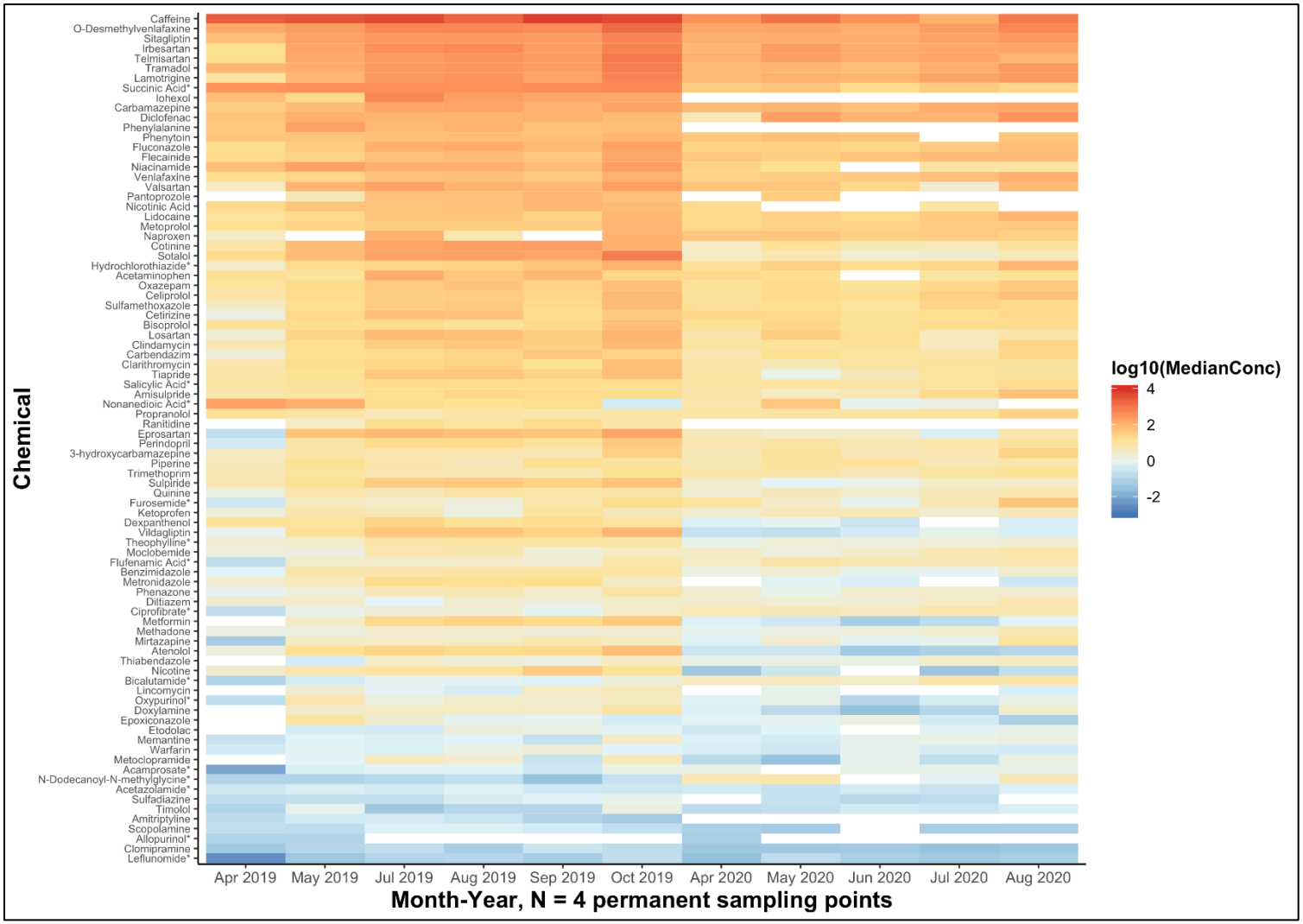


Figure 4. Temporal heat map showing median concentration values (original units: ng/L) per compound measured per sampling month-year plotted using a base-10 logarithmic scale. Median values were calculated across the concentrations measured at the four permanent sampling locations for the respective compound and month-year. Zero-value median concentrations are indicated by grey-shaded boxes. White boxes indicate concentration values that were below the respective quantification range, which were therefore discarded from median calculation. All compounds were measured in positive mode except for those marked with an asterisk, which were measured in negative mode.

The stimulant caffeine, antidepressant metabolite O-desmethylvenlafaxine, antihypertensive drugs irbesartan and telmisartan, the anti-diabetic drug sitagliptin, and the opioid analgesic tramadol were among the most concentrated pharmaceuticals found in Luxembourgish surface waters (Figure 2 and 3) in both 2019 and 2020. From a temporal point of view (Figure 4), the highest median concentrations of the pharmaceuticals were detected in September and October of 2019, and are consistently lower during the spring. The most visually obvious differences between the two sampling years include: 1) amytriptyline, iohexol, phenylalanine, and ranitidine were only detected at quantifiable levels in 2019 and 2) decreases in the median concentrations of dexpanthenol, metformin, nicotine, sotalol, and vildagliptin. As an example, metformin had median concentrations of 3.0 ng/L (May) to 39 ng/L (October) in 2019, much higher than the highest detected median concentration of metformin in 2020 (0.62 ng/L in August 2020). Dexpanthenol is a drug used for prophylactic purposes, both metformin and vildagliptin are drugs used for managing diabetes, sotalol is for the management of arrhythmia, while nicotine relates to smoking. A juxtaposition of data from 2019 and 2020 are presented as boxplots in Figure 5, showing the general decrease in many pharmaceutical concentrations in 2020 (green boxes). For simplicity, only the top 50 pharmaceuticals ranked by median concentration are presented. Some of the most notable drops in detected concentration was observed for dexpanthenol, nicotine, metformin, and sotalol.

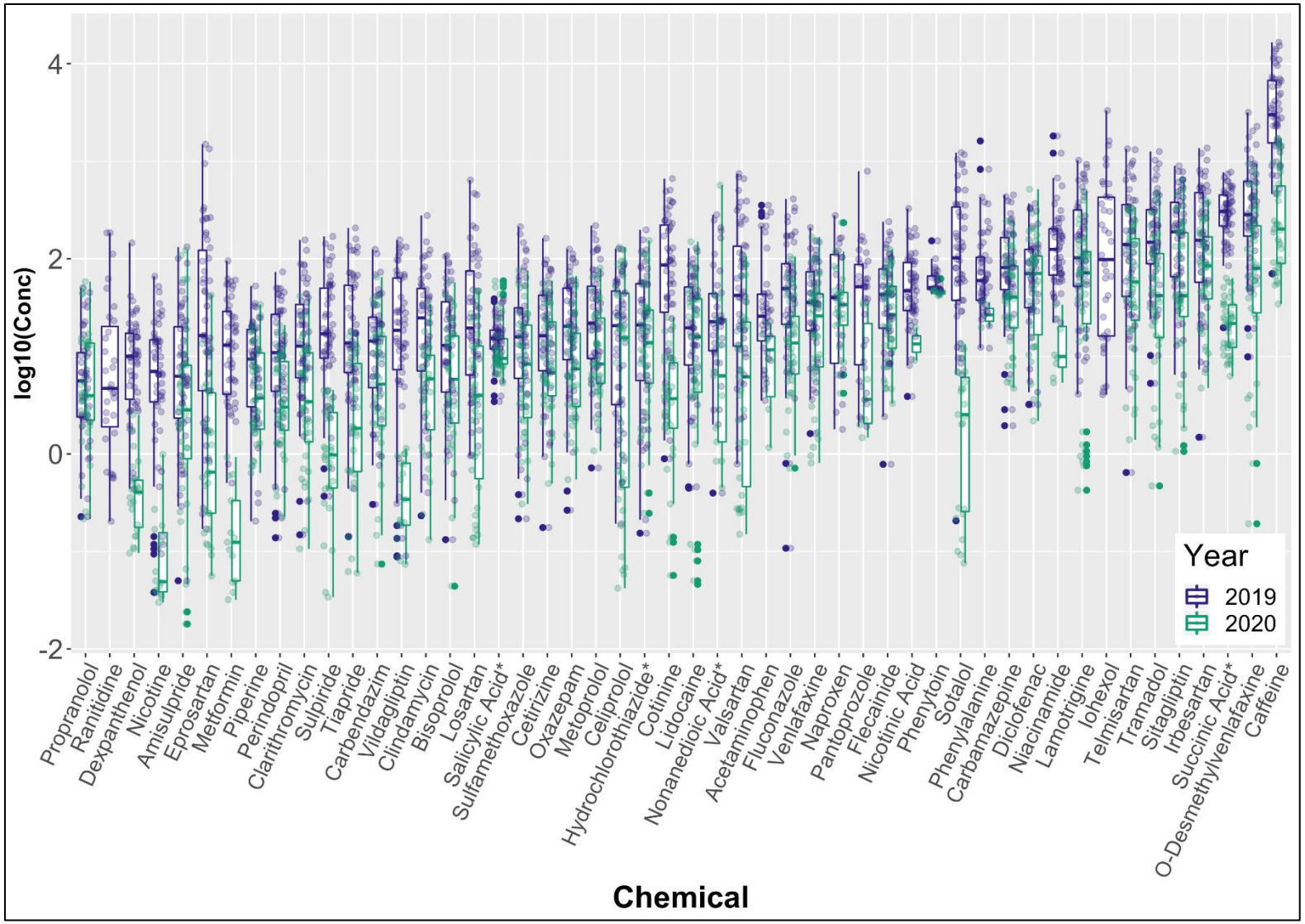


Figure 5. Boxplots showing the range of concentrations (original units: ng/L) measured for the top-50 highest concentration pharmaceutical chemicals across all months and sampling locations in 2019 and 2020, plotted using a base-10 logarithmic scale. Concentration values that were below the respective quantification ranges were excluded. All chemicals were measured in positive mode.

Factors that affected pharmaceutical concentrations in Luxembourg. Interestingly, lower median concentrations of the pharmaceuticals were measured in 2020 compared to 2019 (as shown in Figure 5), which may be partially due to the reduced presence of cross-border workers during the pandemic. COVID-19 has brought on a major shift in working practices, as more people were advised and allowed to work remotely. In Luxembourg, a major part of the workforce is comprised of cross-border workers (approximately 206,000 people in 2019).[45] This translates to an approximately 25% decrease in the daytime population, which may translate to reduced pharmaceutical loading in the sewage system. Two interesting features in Figure 5, also apparent in Figure 4, are the detections of iohexol and ranitidine in 2019 but not in 2020. Iohexol is a radiocontrast agent used for medical imaging. Due to the COVID-19 pandemic, there was a significant decrease in medical procedures for non-communicable diseases, including radio imaging.[46] This decrease may explain why iohexol was not detected at a quantifiable level in 2020 despite having the 6th highest median concentration in 2019. Ranitidine use in the EU, on the other hand, was discontinued in 2020 because of the suspected carcinogen N-nitrosodimethylamine, an impurity present in ranitidine drugs.[47] It is interesting to see how changes in drug usage are abruptly reflected in their detection in the environment.

Changes in precipitation had been reported to affect contaminant levels in water, generally increasing with increased precipitation due to factors such as runoff and combined sewer overflow.[48] Compared to the long term average (1981 to 2010), both 2019 and 2020 experienced a decrease in the annual precipitation (Table 3). For the samplings months that were studied in both 2019 and 2020 (April, May, July, and August),

2020 showed the lowest amount of precipitation, which may have contributed to the lower concentration of pharmaceuticals detected. However, the authors acknowledge that more factors are responsible for the observed differences in pharmaceutical contamination in rivers and is not limited to population and precipitation and can be a subject for further investigation.

Table 3. Precipitation data for Luxembourg. (Source: https://www.meteolux.lu)

Luxembourg	Precipitation, mm												
	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Year
Long term average (1981 to 2010)	77	63	69	58	79	80	71	75	76	87	76	87	898
2019	51	43	83	57	61	55	17	51	59	129	88	87	781
2020	47	148	66	20	36	114	8	30	54	113	33	119	788

While the Chiers flows into the Meuse River and the Alzette flows into the Sauer River (eventually leading into the Rhine), both rivers contribute to the chemical load that eventually ends up in the North Sea. Several studies have determined the presence of pharmaceuticals in the Meuse and Rhine rivers. A 2010 study by ter Laak *et al*. reported compounds such as caffeine, carbamazepine, lidocaine, and iohexol as some of the more concentrated pharmaceuticals in their study of the Rhine, with sulfamethoxazole as the most abundant antimicrobial.[49] The same study also found antihypertensive drugs such as atenolol, metoprolol, and sotalol. Despite being apart by almost a decade, similar trends can be observed in Luxembourgish waters. Later studies of different parts of the Rhine and Meuse rivers reported similar pharmaceuticals [50, 51], however, in some studies the anti-diabetic drug metformin and its TP guanylurea were found to be the most abundant pharmaceutical in surface water samples.[50,52,53] While metformin was also

quantified in this study, the median concentration only ranks 44th over both years among the pharmaceuticals found. Higher levels of the anti-diabetic drug sitagliptin, 5th most abundant, were detected in Luxembourg. The two drugs differ in their mode of regulating sugar in the body.

Challenges in Compound identification. The presence of isobars, isomers, and in source fragments complicate the identification of chemicals in HRMS data, sometimes even leading to these analytes to be excluded from HRMS analysis.[54, 55] Several cases of isobars were encountered in this work including: a) acetaminophen and 1,2,3,6-tetrahydrophthalimide, b) salicylic acid, 3-hydroxybenzoic acid, and 4-hydroxybenzoic acid, c) piperine, morphine, and etodolac, d) cocaine and scopolamine, e) tramadol and O-desmethylvenlafaxine, and f) phenytoin, 2-hydroxycarbamazepine, and 3-hydroxycarbamazepine. While cases a-d were easily resolved using authentic standards, case e and f introduced specific challenges. Tramadol (parent compound) and O-desmethylvenlafaxine (TP of venlafaxine) are constitutional isomers whose extracted ion chromatogram show two unresolved peaks that are both annotated by MetFrag as tramadol (due to tramadol's higher metadata scores). Using standards, the first peak (12.2 min) was ultimately assigned to be O-desmethylvenlafaxine, while the second peak (12.4 min) was tramadol. In order to quantify both compounds, the peaks had to be manually integrated to avoid integrating the two peaks as one compound.

For the suspect screening of phenytoin, three prominent peaks (retention times: 13.95, 14.31, and 14.85 min, respectively) were observed in the positive mode extracted ion chromatogram of *m/z* 253.0972 within 5 ppm error (Figure 6A). Looking at the structure of phenytoin, the absence of chiral carbons renders the possibility of

385 diastereomers, which could explain the presence of multiple peaks, invalid. Analysis of
386 the phenytoin standard showed that this compound elutes at 15.53 min, thus not matching
387 any of the three peaks being investigated. Further inspection using MetFrag and database
388 matching suggested that the second and third peaks belong to the positional isomers 2-
389 hydroxycarbamazepine and 3-hydroxycarbamazepine, metabolites of the anticonvulsant
390 carbamazepine. Retention time matching using a standard confirmed that the peak at
391 14.85 min is indeed 3-hydroxycarbamazepine while the peak at 14.31 min can be
392 assigned as 2-hydroxycarbamzepine (Level 3), despite the lack of standards, due to the
393 similarity of its mass spectrum with 3-hydroxycarbamazepine. However, the first and
394 biggest peak proved to be challenging. Inspection of the MS1 spectrum at 13.95 min
395 shows another peak with *m/z* 271.1075 (mass difference equivalent to the loss of water,
396 Figure 6B) can be found whose MS2 spectrum is very similar to the 253.0972 peaks at
397 14.31 and 14.85 min (Figure 6C and 6D). Using these pieces of information, it can be
398 suggested that the 253.0972 peak is potentially an in source fragment of 271.1075. Using
399 271.1075 as the precursor ion, MetFrag suggests that the peak is potentially 10,11-
400 dihydroxycarbamazepine (MoNA score: 0.8340) or phenytoin acid (MoNA score: 0.8076),
401 which are TPs of carbamazepine and phenytoin, respectively. The presence of the
402 210.0915 and 180.0811 fragments, which match fragments of other carbamazepine
403 metabolites, and the earlier elution suggesting that the molecule is more polar than the
404 mono-hydroxylated analogs, support the tentative identification of the 13.95 min peak as
405 10,11-dihydroxycarbamazepine (Level 3).

406 One case that needs further inspection are the stereoisomers vidarabine and
407 adenosine, which are impossible to separate using the chromatographic method

employed in this study. While there are reports on the utility of ion mobility to discriminate between stereoisomers, it is still to be tested whether such resolution is practically achievable.[56-58] Published collisional cross sections of vidarabine (156.4 Å² for $[M+H]^+$) and adenosine (156.9 Å² for $[M+H]^+$) measured on the same instrument are available, revealing a difference of only 0.5 Å² or 0.3%, which is too close to distinguish currently within the typical resolving power of ion mobility spectrometers.[59, 60]

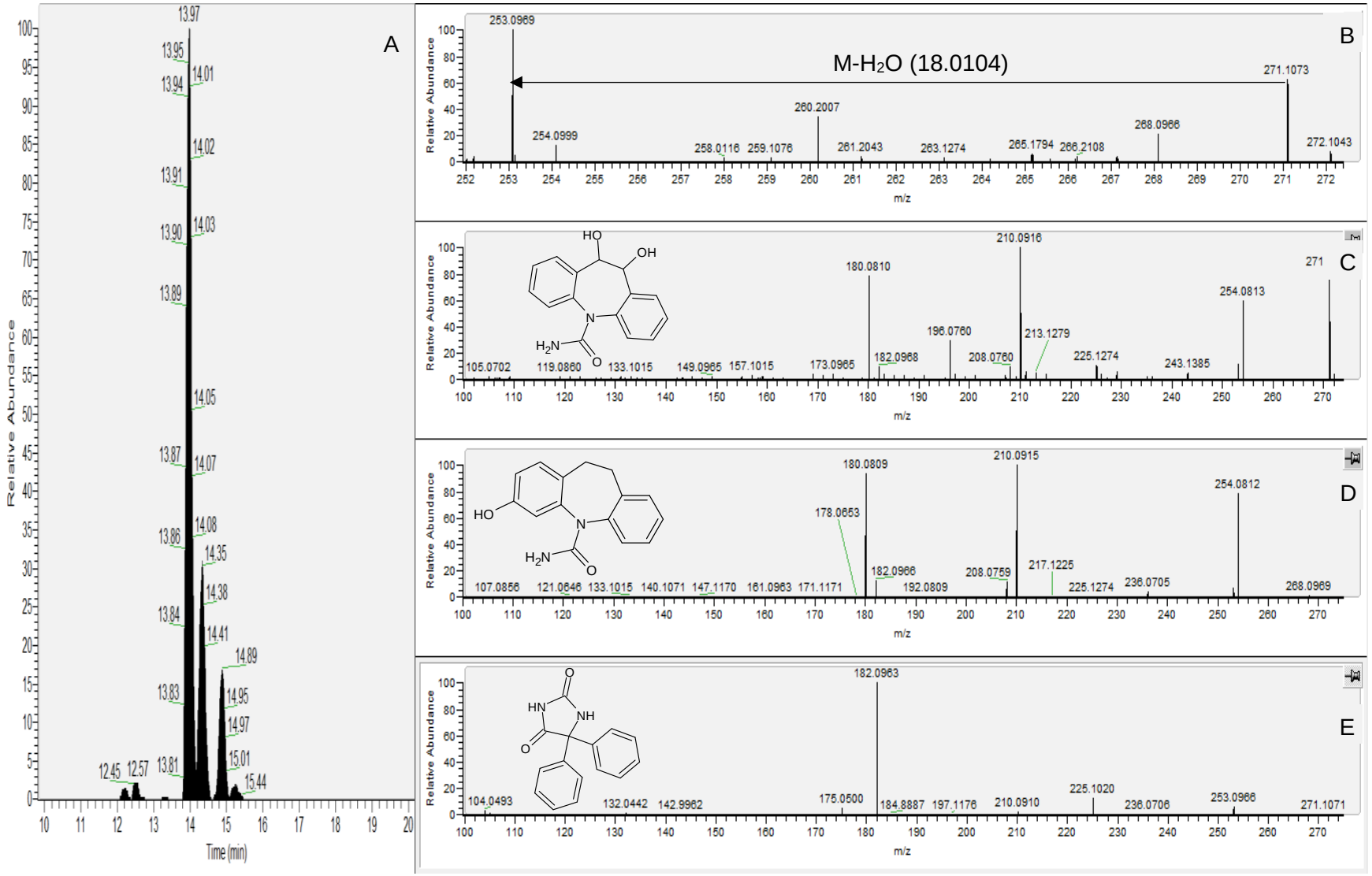


Figure 6. A) Extracted ion chromatogram of *m/z* = 253.0969 in a surface water sample showing three distinct peaks. B) MS1 spectrum of the 13.97 peak showing a higher peak that may have lost water to produce the 253.0969 peak. C) MS2 spectrum of *m/z* = 271.1073 (potentially 10,11-dihydroxycarbamazepine, structure on the same pane) showing similar fragments to the MS2 fragments of 3-hydroxycarbamazepine standard (structure on the same pane) see D). E) MS2 spectrum of the phenytoin standard (structure on the same pane).

This study documents suspect screening efforts thus far for pharmaceuticals and their known TPs as a starting point for further understanding pharmaceutical levels in Luxembourgish surface waters. Other activities looking into different chemical classes such as pesticides [65], industrial chemicals, and other emerging pollutants are ongoing. The continuous analysis of surface water using HRMS as part of the routine monitoring efforts will enable retrospective screening [61, 62] for newly identified contaminants that may impact local surface water quality and biota, such as the effect observed by city runoff on coho salmon.[63] This study reports primarily Level 1 and 2a identifications due to the hard filter of MoNA score >0.9 applied during the MetFrag analysis. Other tentative identifications have been communicated with AGE and these, along with more detailed trend analysis as more temporal data points are collected, can be investigated in future works as resources allow. Quantification efforts could be further improved by using the list of pharmaceuticals identified in this work as a target list, as well as investing in isotopically labeled standards (which was beyond the scope of the current works, as target analysis is performed by AGE). Finally, as experimental databases increase in size and coverage, the ability to screen for more compounds with higher confidence with these open source methods such as the one presented here will also increase, highlighting the need for the community at large to continue to contribute to publicly available databases.

One main factor limiting TP suspect screening is the lack of available information in open databases that is standardized and thus suitable to be extracted consistently and reproducibly to form meaningful suspect lists. Of the 816 parent compounds on the CNS list, only 44 had associated TP information (*i.e.,* one or more TPs) that could be extracted from PubChem as performed in this study. Certainly, there are far more pharmaceutical

metabolites/TPs than those that are identified here, but this information is not yet available in a readily extractable form suitable for an automated workflow within PubChem (the efforts within the NORMAN Suspect List Exchange have just commenced recently).[28, 65] As more information is added and as more environmental transformation studies are performed and deposited in a FAIR (Findable, Accessible, Interoperable and Reusable) manner [64], the ability to screen for TPs in an automated fashion would also increase. Contributions to this effort are welcome and discussions with select journals are currently ongoing to improve this in the future via deposition templates.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at: https://pubs.acs.org/doi/xx.xxx. The suspect list used in this work is available online as LUXPHARMA (S76) on Zenodo (DOI: 10.5281/zenodo.4587356), CompTox (https://comptox.epa.gov/dashboard/chemical_lists/LUXPHARMA, PubChem (https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101) and NORMAN-SLE (https://www.norman-network.com/nds/SLE/).

The data (as .mzML files) are available as dataset MSV000087190 from the GNPS MassIVE repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp), citable under DOI: 10.25345/C5D81C and accessible via ftp://massive.ucsd.edu/MSV000087190/ and https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000087190. Table S5 in the SI lists the original file names and their corresponding names in this paper.

The original file names were kept to allow traceability to the original sample files stored locally at the University of Luxembourg. Both Shinyscreen (https://git-r3lab.uni.lu/eci/shinyscreen/) and MetFrag (http://ipb-halle.github.io/MetFrag/) are open source; additional support scripts mentioned are available from the ECI GitLab repository (https://git-r3lab.uni.lu/eci/pubchem). All code used to run MetFrag in the command line using R, generate the Transformation Products suspect list, and plot Figures 2-5 is available via https://git-r3lab.uni.lu/adelene.lai/additional_si_luxpharma_singh_et_al. All other code and databases used as part of MetFrag identification are likewise openly available (links inline throughout this manuscript).

AUTHOR INFORMATION

Corresponding Authors

Randolph Singh- *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg; Current affiliation: IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer), Laboratoire Biogéochimie des Contaminants Organiques, Rue de l'Ile d'Yeu, BP 21105, Nantes Cedex 3, 44311, France ;* orcid.org/0000-0003-4500-3400 Email: randolph.singh@ifremer.fr

Emma Schymanski- *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg;* orcid.org/0000-0001-6868-8145 Email: emma.schymanski@uni.lu

Authors

Adelene Lai- *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg; Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller University, Lessing Strasse 8, 07743, Jena, Germany;* orcid.org/0000-0002-2985-6473

Jessy Krier- *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg;* orcid.org/0000-0001-6986-5545

Todor Kondic- *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg;* orcid.org/0000-0001-6662-4375

Philippe Diderich- *Administration de la gestion de l'eau, Ministère de l'Environnement, du Climat et du Développement durable, Luxembourg;* orcid.org/0000-0001-6969-2162

Author Contributions

ELS, PD, and RRS designed the study. PD prepared the samples. JK and RRS performed instrumental analysis of samples and standards, and suspect screening. AL, ELS, and TK wrote the code/ developed the computational pipeline used. AL and RRS generated the figures. RRS drafted the manuscript with contributions from all authors. All authors revised and approved the submitted version.

ACKNOWLEDGMENTS

AL, ELS, RRS, and TK acknowledge supported by the Luxembourg National Research Fund (FNR) for project A18/BM/12341006. The authors acknowledge the people in the

background who tirelessly collect samples and have contributed to this work through healthy discussions in the office, as well as all those contributing to the many open science efforts mentioned. ELS gratefully acknowledges the contributions of several of the PubChem team to the TP and PubChemLite efforts, including Evan Bolton, Jeff Zhang and Paul Thiessen (all NIH/NLM/NCBI).

REFERENCES

1. Pailler, J.-Y.; Krein, A.; Pfister, L.; Hoffmann, L.; Guignard, C., Solid phase extraction coupled to liquid chromatography-tandem mass spectrometry analysis of sulfonamides, tetracyclines, analgesics and hormones in surface water and wastewater in Luxembourg. *Science of the Total Environment* **2009,** *407*, (16), 4736-4743.

2. Pailler, J.-Y.; Guignard, C.; Meyer, B.; Iffly, J.-F.; Pfister, L.; Hoffmann, L.; Krein, A., Behaviour and fluxes of dissolved antibiotics, analgesics and hormones during flood events in a small heterogeneous catchment in the Grand Duchy of Luxembourg. *Water, air, and soil pollution* **2009,** *203*, (1-4), 79-98.

3. Meyer, B.; Pailler, J.-Y.; Guignard, C.; Hoffmann, L.; Krein, A., Concentrations of dissolved herbicides and pharmaceuticals in a small river in Luxembourg. *Environmental Monitoring and Assessment* **2011,** *180*, (1), 127-146.

4. Krein, A.; Keßler, S.; Meyer, B.; Pailler, J.-Y.; Guignard, C.; Hoffmann, L., Concentrations and loads of dissolved xenobiotics and hormones in two small river catchments of different land use in Luxembourg. *Hydrological Processes* **2013,** *27*, (2), 284-296.

5. Karier, P.; Kraus, G.; Kolber, I., Metazachlor traces in the main drinking water reservoir in Luxembourg: a scientific and political discussion. *Environmental Sciences Europe* **2017,** *29*, (1), 25.

6. Schummer, C.; Tuduri, L.; Briand, O.; Appenzeller, B. M.; Millet, M., Application of XAD-2 resin-based passive samplers and SPME–GC–MS/MS analysis for the monitoring of spatial and temporal variations of atmospheric pesticides in Luxembourg. *Environmental Pollution* **2012,** *170*, 88-94.

7. Bohn, T.; Cocco, E.; Gourdol, L.; Guignard, C.; Hoffmann, L., Determination of atrazine and degradation products in Luxembourgish drinking water: origin and fate of potential endocrine-disrupting pesticides. *Food Additives & Contaminants: Part A* **2011,** *28*, (8), 1041-1054.

8. Commission, E., Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. *Off. J. Eur. Union* **2013,** *226*, 1-17.

9. Richardson, S. D.; Fasano, F.; Ellington, J. J.; Crumley, F. G.; Buettner, K. M.; Evans, J. J.; Blount, B. C.; Silva, L. K.; Waite, T. J.; Luther, G. W., Occurrence and mammalian cell toxicity of iodinated disinfection byproducts in drinking water. *Environmental Science & Technology* **2008,** *42*, (22), 8330-8338.

10. Escher, B. I.; Fenner, K., Recent advances in environmental risk assessment of transformation products. *Environmental science & technology* **2011,** *45*, (9), 3835-3847.

11. Fenner, K.; Kooijman, C.; Scheringer, M.; Hungerbühler, K., Including transformation products into the risk assessment for chemicals: The case of nonylphenol ethoxylate usage in Switzerland. In ACS Publications: 2002.

12. Hollender, J.; Rothardt, J.; Radny, D.; Loos, M.; Epting, J.; Huggenberger, P.; Borer, P.; Singer, H., Comprehensive micropollutant screening using LC-HRMS/MS at three riverbank filtration sites to assess natural attenuation and potential implications for human health. *Water Research X* **2018,** *1*, 100007.

13. 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. *Environmental Science & Technology* **2017,** *51*, (18), 10652-10662.

14. Hollender, J.; van Bavel, B.; Dulio, V.; Farmen, E.; Furtmann, K.; Koschorreck, J.; Kunkel, U.; Krauss, M.; Munthe, J.; Schlabach, M.; Slobodnik, J.; Stroomberg, G.; Ternes, T.; Thomaidis, N. S.; Togola, A.; Tornero, V., High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. *Environmental Sciences Europe* **2019,** *31*, (1), 42.

15. Sobus, J. R.; Wambaugh, J. F.; Isaacs, K. K.; Williams, A. J.; McEachran, A. D.; Richard, A. M.; Grulke, C. M.; Ulrich, E. M.; Rager, J. E.; Strynar, M. J., Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. *Journal of exposure science & environmental epidemiology* **2018,** *28*, (5), 411-426.

16. Pourchet, M.; Debrauwer, L.; Klanova, J.; Price, E. J.; Covaci, A.; Caballero-Casero, N.; Oberacher, H.; Lamoree, M.; Damont, A.; Fenaille, F.; Vlaanderen, J.; Meijer, J.; Krauss, M.; Sarigiannis, D.; Barouki, R.; Le Bizec, B.; Antignac, J.-P., Suspect and non-targeted screening of chemicals of emerging concern for human biomonitoring, environmental health studies and support to risk assessment: From promises to challenges and harmonisation issues. *Environment International* **2020,** *139*, 105545.

17. Brack, W.; Hollender, J.; de Alda, M. L.; Müller, C.; Schulze, T.; Schymanski, E.; Slobodnik, J.; Krauss, M., High-resolution mass spectrometry to complement monitoring and track emerging chemicals and pollution trends in European water resources. *Environmental Sciences Europe* **2019,** *31*, (1), 62.

18. Anliker, S.; Loos, M.; Comte, R.; Ruff, M.; Fenner, K.; Singer, H., Assessing Emissions from Pharmaceutical Manufacturing Based on Temporal High-Resolution Mass Spectrometry Data. *Environmental Science & Technology* **2020,** *54*, (7), 4110-4120.

19. Jernberg, J.; Pellinen, J.; Rantalainen, A.-L., Identification of organic xenobiotics in urban aquatic environments using time-of-flight mass spectrometry. *Science of The Total Environment* **2013,** *450-451*, 1-6.

20. Carpenter, C. M. G.; Wong, L. Y. J.; Johnson, C. A.; Helbling, D. E., Fall Creek Monitoring Station: Highly Resolved Temporal Sampling to Prioritize the Identification of Nontarget Micropollutants in a Small Stream. *Environmental Science & Technology* **2019,** *53*, (1), 77-87.

21. Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O., Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* **2018,** *8*, (2), 31.

22. MetFrag (https://msbi.ipb-halle.de/MetFrag/) (May 2020)

23. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S., MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of Cheminformatics* **2016,** *8*, (1), 3.

24. Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S., In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **2010,** *11*, (1), 148.

25. MassBank of North America (https://mona.fiehnlab.ucdavis.edu/). (May 2020)

26. PubChem (https://pubchem.ncbi.nlm.nih.gov/). (May 2020)

27. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E., PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **2020,** *49*, (D1), D1388-D1395.

28. Schymanski, E. L.; Kondić, T.; Neumann, S.; Thiessen, P. A.; Zhang, J.; Bolton, E. E., Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *Journal of Cheminformatics* **2021,** *13*, (1), 19.

29. Bolton, E.; Schymanski, E., PubChemLite tier0 and tier1. 10.5281/zenodo.3611238. PubChemLite.0.2.0; Zenodo, 2020. http://doi.org/10.5281/zenodo.3611238 (May 2020)

30. NORMAN Network (2021) NORMAN Suspect List Exchange (NORMAN-SLE, https://www.norman-network.com/nds/SLE/) (May 2020)

31. Singh, R. R. (2021), S76 | LUXPHARMA | Pharmaceuticals Marketed in Luxembourg (Version NORMAN-SLE-S76.0.1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4587356 (May 2020)

32. Caisse Nationale de Santé (https://cns.public.lu/en/professionnels-sante/medicaments/medicaments-commercialises.html) (November 2019)

33. Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M., The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics* **2017,** *9*, (1), 61.

34. McEachran, A. D.; Mansouri, K.; Grulke, C.; Schymanski, E. L.; Ruttkies, C.; Williams, A. J., "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies. *Journal of Cheminformatics* **2018,** *10*, (1), 45.

35. Schymanski, E. (2020), https://git-r3lab.uni.lu/eci/pubchem/-/blob/master/annotations/tps/extractAnnotations.R Accessed October 2020

36. PubChem Query Syntax (https://pubchemdocs.ncbi.nlm.nih.gov/sdq-query-syntax) (October 2020)

37. PubChem Identifier Exchange Service (https://pubchem.ncbi.nlm.nih.gov/idexchange/idexchange.cgi) (May 2020)

38. Lai, A. https://git-r3lab.uni.lu/adelene.lai/additional_si_luxpharma_singh_et_al. (May 2020)

39. Shinyscreen (https://git-r3lab.uni.lu/eci/shinyscreen) (May 2020)

40. Lai, A.; Singh, R.; Kovalova, L.; Jaeggi, O.; Kondic, T.; Schymanski, E., Retrospective Non-target Analysis to Support Regulatory Water Monitoring: From Masses of Interest to Recommendations via in silico workflows. *Environmental Sciences Europe* **2021**, 33, 43.

41. Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J., Identifying Small Molecules via High Resolution Mass Spectrometry:

638 Communicating Confidence. *Environmental Science & Technology* **2014,** *48*, (4), 2097-
639 2098.
640 42. Ulrich, E. M.; Sobus, J. R.; Grulke, C. M.; Richard, A. M.; Newton, S. R.; Strynar,
641 M. J.; Mansouri, K.; Williams, A. J., EPA's non-targeted analysis collaborative trial
642 (ENTACT): genesis, design, and initial findings. *Analytical and bioanalytical chemistry*
643 **2019,** *411*, (4), 853-866.
644 43. Singh, R.R. ; Lai, A. ; Krier, J. ; et al. Supplemental Information for "Occurrence
645 and Distribution of Pharmaceuticals and their Transformation Products in Luxembourgish
646 Surface Waters". https://git-
647 r3lab.uni.lu/adelene.lai/additional_si_luxpharma_singh_et_al/-/tree/master/figures.
648 (Accessed 23 March 2021),
649 44. Beggen Wastewater Treatment Plant
650 ((https://map.geoportail.lu/theme/eau?lang=en&version=3&zoom=12&X=681318&Y=63
651 86989&rotation=0&layers=645&opacities=1&bgLayer=topo_bw_jpeg)). (October 2020)
652 45. STATEC Vue d'ensemble du marche du travail (en 1000 personnes) 2000-2019
653 (https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12951&IF_Lan
654 guage=fra&MainTheme=2&FldrName=3)(October 2020)
655 46. Cavallo, J. J.; Forman, H. P., The Economic Impact of the COVID-19 Pandemic
656 on Radiology Practices. *Radiology* **2020,** *296*, (3), E141-E144.
657 47. Agency, E. M. Suspension of ranitidine medicines in the EU
658 (https://www.ema.europa.eu/en/news/suspension-ranitidine-medicines-eua) (October
659 2020)

660 48. Impact of high precipitation and temperature events on the distribution of emerging
661 contaminants in surface water in the Mid-Atlantic, United States. *Science of The Total*
662 *Environment* **2021,** *755*, 142552.
663 49. ter Laak, T. L.; van der Aa, M.; Houtman, C. J.; Stoks, P. G.; van Wezel, A. P.,
664 Relating environmental concentrations of pharmaceuticals to consumption: A mass
665 balance approach for the river Rhine. *Environment International* **2010,** *36*, (5), 403-409.
666 50. Ruff, M.; Mueller, M. S.; Loos, M.; Singer, H. P., Quantitative target and systematic
667 non-target analysis of polar organic micro-pollutants along the river Rhine using high-
668 resolution mass-spectrometry – Identification of unknown sources and compounds.
669 *Water Research* **2015,** *87*, 145-154.
670 51. de Jongh, C. M.; Kooij, P. J. F.; de Voogt, P.; ter Laak, T. L., Screening and human
671 health risk assessment of pharmaceuticals and their transformation products in Dutch
672 surface waters and drinking water. *Science of The Total Environment* **2012,** *427-428*, 70-
673 77.
674 52. ter Laak, T. L.; Kooij, P. J. F.; Tolkamp, H.; Hofman, J., Different compositions of
675 pharmaceuticals in Dutch and Belgian rivers explained by consumption patterns and
676 treatment efficiency. *Environmental Science and Pollution Research* **2014,** *21*, (22),
677 12843-12855.
678 53. Houtman, C. J.; ten Broek, R.; de Jong, K.; Pieterse, B.; Kroesbergen, J., A
679 multicomponent snapshot of pharmaceuticals and pesticides in the river Meuse basin.
680 *Environmental Toxicology and Chemistry* **2013,** *32*, (11), 2449-2459.
681

54. Sobus, J. R.; Grossman, J. N.; Chao, A.; Singh, R.; Williams, A. J.; Grulke, C. M.; Richard, A. M.; Newton, S. R.; McEachran, A. D.; Ulrich, E. M., Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. *Analytical and bioanalytical chemistry* **2019,** *411*, (4), 835-851.

55. Singh, R. R.; Chao, A.; Phillips, K. A.; Xia, X. R.; Shea, D.; Sobus, J. R.; Schymanski, E. L.; Ulrich, E. M., Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: a case study with ENTACT mixtures. *Analytical and Bioanalytical Chemistry* **2020**.

56. Colson, E.; Decroo, C.; Cooper-Shepherd, D.; Caulier, G.; Henoumont, C.; Laurent, S.; De Winter, J.; Flammang, P.; Palmer, M.; Claereboudt, J., Discrimination of regioisomeric and stereoisomeric saponins from Aesculus hippocastanum seeds by ion mobility mass spectrometry. *Journal of The American Society for Mass Spectrometry* **2019,** *30*, (11), 2228-2237.

57. McCooeye, M.; Ding, L.; Gardner, G. J.; Fraser, C. A.; Lam, J.; Sturgeon, R. E.; Mester, Z., Separation and quantitation of the stereoisomers of ephedra alkaloids in natural health products using flow injection-electrospray ionization-high field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Analytical chemistry* **2003,** *75*, (11), 2538-2542.

58. Hofmann, J.; Hahm, H. S.; Seeberger, P. H.; Pagel, K., Identification of carbohydrate anomers using ion mobility–mass spectrometry. *Nature* **2015,** *526*, (7572), 241-244.

59. Hines, K. M.; Ross, D. H.; Davidson, K. L.; Bush, M. F.; Xu, L., Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Analytical Chemistry* **2017,** *89*, (17), 9023-9030.

60. Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibáñez, M.; Goshawk, J.; Barknowitz, G.; Hernández, F.; Bijlsma, L., Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environmental Science & Technology* **2020,** *54*, (23), 15120-15131.

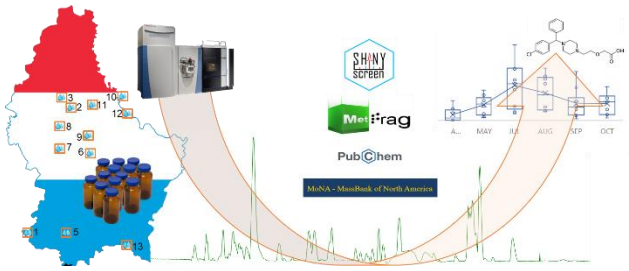
61. Creusot, N.; Casado-Martinez, C.; Chiaia-Hernandez, A.; Kiefer, K.; Ferrari, B. J. D.; Fu, Q.; Munz, N.; Stamm, C.; Tlili, A.; Hollender, J., Retrospective screening of high-resolution mass spectrometry archived digital samples can improve environmental risk assessment of emerging contaminants: A case study on antifungal azoles. *Environment International* **2020,** *139*, 105708.

62. Alygizakis, N. A.; Samanipour, S.; Hollender, J.; Ibáñez, M.; Kaserzon, S.; Kokkali, V.; Van Leerdam, J. A.; Mueller, J. F.; Pijnappels, M.; Reid, M. J., Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. *Environmental science & technology* **2018,** *52*, (9), 5135-5144.

63. Tian, Z.; Zhao, H.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X.; Prat, J.; Mudrock, E.; Hettinger, R., A ubiquitous tire rubber–derived chemical induces acute mortality in coho salmon. *Science* **2020,** *371*, (6525), 185-189.

64. Fair Principles (https://www.go-fair.org/fair-principles/). (October 2020)

65. Krier, J.; Singh, R.R.; Kondic, T. ; et al. Discovering Pesticides and their Transformation Products in Luxembourg Waters using Open Cheminformatics Approaches, 30 April 2021, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-478324/v1]



TOC Graphic